



Voorspelmodellen voor toerismedrukte

Literatuurstudie

Eerste oplage

Onderdeel van het Data & Development Lab (DDL)
& European Tourism Future Institute (ETFI)



Centraal Bureau
voor de Statistiek



NBTC
Nederlands Bureau
voor Toerisme &
Congressen



CELTH Centre of Expertise
leisure, tourism & hospitality

Auteurs

Noël Middelhoek
Literatuuronderzoek en hoofdauteur

Jasper Heslinga
Co-auteur

25 mei 2021





Inhoudsopgave

1. Voorspelmodellen	4
1.1 Introductie	4
1.2 Toerismedrukte voorspellen	4
1.3 De potentie van voorspelmodellen	5
1.4 Een korte terugblik op toeristische voorspelmodellen	5
1.5 De benodigdheden van voorspelmodellen	5
1.6 Voorspelmethodieken	7
2. Causale modellen	8
2.1 Inleiding	8
2.2 Enkelvoudig-en meervoudige regressiemodellen	9
2.3 Vector Auto Regressie (VAR)	10
3. Extrapolatie modellen	11
3.1 Inleiding	11
3.2 Autoregressief bewegende gemiddeldes: SMA, ARMA, ARIMA, SARIMA en ARFIMA	11
3.2.1 Simple Moving Average	11
3.2.2 Auto Regressive Moving Average	11
3.2.3 Auto Regressive Integrated Moving Average	11
3.2.1 Seasonal Auto Regressive Integrated Moving Average	11
3.3 Exponentiële afvlakking: SES, DES, TES	12
3.3.1 Simple Exponential Smoothing	12
3.3.2 Double Exponential Smoothing	13
3.3.3 Triple Exponential Smoothing	13
4. Kunstmatige intelligentie	14
4.1 Inleiding	14
4.2 Artificial Neural Networks	15
4.3 Support Vector Machines, Regression and Classification	16
4.4 Recurrent Neural Network	17
4.5 Long-Short Term Memory	18
5. Hybride modellen	19
6. Flowchart & modellenoverzicht	20
7. Bronvermeldingen	24

1. Voorspelmodellen

1.1 Introductie

Dit werkpakket heeft als doel om te bepalen wat er al bekend is over voorspelmodellen voor toerisme(drukke) en hoe hierin bezoekersstromen als indicator worden gehanteerd, met speciale aandacht voor de (on)mogelijkheden, de onderzoekstechnische en methodologische uitdagingen (en de benodigde databronnen). De verkenning kan zo (op metaniveau) ook een bijdrage leveren aan de opzet van (onderstaande) experimenten (als het gaat om de voorwaarden en condities voor de ontwikkeling van een voorspelmodel).

1.2 Toerismedrukke voorspellen

Het maken van voorspelmodellen met betrekking tot toerismedrukke lijkt gehanteerd te kunnen worden als middel om aankomend toerisme nauwkeuriger in te schatten. Allereerst is het belangrijk om te benoemen dat gezien de aard en doel van dit onderzoek alleen aandacht besteed zal worden aan kwantitatieve modellen en staat het begrip voorspelmodellen in deze literatuurstudie gelijk aan **kwantitatieve voorspelmodellen**. In tegenstelling tot veelal subjectieve voorspellingen op basis van expertise, marktinformatie en scenarioplanning, zijn kwantitatieve modellen constant, objectief en instaat om meer informatie te verwerken (Snellen, 2019).

Het gebruik van kwantitatieve modellen komt echter samen met de sterke afhankelijkheid van beschikbare data (Li en Song, 2008). Het maken van voorspellingen wordt namelijk gedaan op basis van numerieke data waarbij wiskundige calculaties trachten te berekenen wat het meest waarschijnlijke is om zich in de toekomst af te spelen.

Een dergelijke voorspelling kan gemaakt worden op basis van historische en huidige data. Ondanks dat voorspelmodellen ver van ons bed lijken maakt bijna iedereen dagelijks gebruik van verschillende voorspelmodellen. Ze zijn het werkende brein achter webshops die producten aanbevelen op basis van eerdere gebruikersinteractie en ander persoonsgegevens, maar ook de reden dat Google een “bedoelde u misschien” kan aanbieden na het maken van een typefout. Daarnaast gebruiken onze smartphones een voorspelmodel om het eerstvolgende woord van uw zin te suggereren en zijn zelfs de stoplichten op een kruispunt instaat om keuzes te maken op basis van scenario's die zich nog af moeten spelen (Kooistra, 2012).

Deze voorbeelden geven maar aan dat voorspelmodellen vrijwel niet meer weg te denken zijn uit onze maatschappij en toegepast worden voor het maken van simpele correcties of meer ingewikkelde voorspellingen, afhankelijk van het aantal meegenomen elementen. Zo kan er op basis van historische bezoekersaantallen, seizoensinvloeden en online zoekvolumes een voorspelling over bezoekersaantallen op een specifieke bestemming gemaakt worden.

Het maken van deze voorspelling wordt berekend door uit te zoeken hoe verschillende datapunten met elkaar verbonden zijn en hier de kleinste correlaties in te vinden. De uitkomst hiervan kan vervolgens geanalyseerd en gebruikt worden voor het maken van voorspellingen. Al deze stappen worden in de meeste gevallen tot stand gebracht door algoritmes die data omzetten in statistieken en deze analyseren. Sommige algoritmes zijn in staat om zichzelf te evalueren en te leren van het doorlopen proces zodat zij in het vervolg nog effectiever en nauwkeuriger kunnen worden.



1.3 De potentie van voorspelmodellen voor toerisme

De steeds sneller toenemende aanvoer en beschikbaarheid van data heeft ervoor gezorgd dat big data beschikbaar is voor bijna alle bedrijven, organisaties en instanties. Echter, is data vaak ongestructureerd, complex en tijdrovend voor mensen om te structureren, analyseren en/of visualiseren. Met een voorspelmodel wordt het mogelijk om patronen te identificeren en op basis hiervan conclusies te trekken. Dit maakt een onderbouwde anticipatie op wat er mogelijk komen gaat mogelijk. In combinatie met big data kan er een meer gepersonaliseerd aanbod geboden worden aan consumenten, zoals het bij voorbaat plaatsen van een glutenvrij menu in hotelkamers of het online aanbieden van huwelijks pakketten aan een net getrouwd stel. Op deze manier kan de industrie zijn benadering naar consumenten transformeren door het aanbod te laten anticiperen op de vraag. Desondanks de commerciële waarde die dit toegevoegd aan het aanbod, zijn er zaken omtrent toerismedrukke waarvan geopperd kan worden dat ze minstens net zoveel aandacht verdienen. Bijvoorbeeld ontwikkelingen en bedreigingen op het gebied van massatoerisme, stedelijke knelpunten, prijsstijgingen in de woningmarkt en/of een combinatie van deze factoren tijdens de Covid-19 epidemie. Het correct hanteren van flexibel en oplossingsgericht beleid met betrekking tot deze kwestie is lastig en dit is dan ook juist waar zich aantrekkelijke mogelijkheden bevinden voor voorspelmodellen die het mogelijk maken om te anticiperen op een breed scala aan zaken zoals bezettingsmanagement, druk op infrastructuur, maatschappij en de natuurlijke omgeving van een bestemming.

1.4 Een korte terugblik op toeristische voorspelmodellen

Voorspelmodellen zijn zeker niet nieuw. Al vanaf de zestiger jaren tot nu heeft er onderzoek plaatsgevonden met betrekking tot het toepassen en ontwikkelen van voorspelmodellen binnen de toeristische sector. Met name de belangstelling om fundamentele factoren achter het adopteren van nieuwe technologieën te identificeren zorgde voor een toenemend verloop van onderzoek (Gharibi, 2020). Het eerste decennia aan onderzoek vormde de grondslag voor de eerste beoordelingsonderzoeken uitgevoerd door Brand (1973), Chan (1979), en Witt en Witt (1995). De laatstgenoemde auteurs concludeerde dat er destijds al dat er niet één specifieke methode bestond die andere methodieken op alle vlakken oversteeg. Maar dat de accuraatheid van een model in verband stond met de te behandelen casus. Crouch (1995), deed verslag over 300 onderzoeken tussen 1961 en 1993. Hierna zijn er volgens Li, Song & Witt (2005) nogmaals 120 publicaties met betrekking tot hetzelfde onderwerp uitgebracht. Kort daarna brachten Li & Song (2008), een nieuw beoordelingsrapport uit waarin zij de eigen publicatie (2005) bekritiseerde. Deze kritiek was gericht op het feit dat de beoordeling enkel gefocust was op econometrisch perspectief en vormde hiermee de grondslag voor diversiteit in methodiek analyses. Li & Song beoordeelde gelijkwaardige onderzoeken tussen 2005 en 2007 waarin gebruik werd gemaakt van andere methodieken zoals: tijdreeks-, kustmatige intelligentie- (AI) en (non)statistisch-modellen. Een half decennia later onderzochten Goh en Law (2011) de methodologische ontwikkeling van het voorspellen van toerisme drukte tussen 1995 en 2009. Recentelijk hebben Wu et al. (2017) een beoordelingsstudie uitbracht met betrekking voorspelmodellen omtrent bezoekersdrukke bij hotels tussen 2007 en 2015. Volgens Jiao en Chen (2018), is er over de afgelopen tien jaar een duidelijke trend ontstaan in het gebruiken van kunstmatige intelligentie. De meest recentelijk geïdentificeerde trend is hieraan gerelateerd en bestaat uit 'Hybride modellen'.

1.5 De benodigheden van voorspelmodellen

Om voorspellingen te kunnen maken zijn er (on)afhankelijke variabelen nodig om modellen mee te 'voeden'. Een afhankelijk variabele kan gezien worden als een waarde die constant en niet direct **beïnvloedbaar** is. Bijvoorbeeld het aantal bezoekers in Amsterdam tijdens de zomer van 2018, 2019, en 2020. De waarde van dit **afhankelijke variabele**, in dit geval het aantal bezoekers, wordt bepaald door verschillende **onafhankelijke variabelen**. Denk hierbij aan weersomstandigheden, georganiseerde evenementen, sentiment, zoekvolumes, aanbiedingen van luchtvaartmaatschappijen, etc. De benodigde **input data** van het model is dus anders per casus en wordt bepaald door zowel invloed hebbende factoren als het type model dat gebruikt wordt. Ondanks de casegebondenheid zijn er over de afgelopen jaren degelijke trends geweest in het vergaren en gebruiken van data. Om de chronologische ontwikkelingen van deze trends in stand te houden zal dit onderwerp in het volgende hoofdstukken verder belicht worden.

Allereerste kan gezegd worden dat de oorspronkelijk gebruikte variabelen in de meeste gevallen gebaseerd waren op economische variabelen in relatie tot zowel de plek van herkomst als bestemming van een toerist. Vooral de bezoeker zijn niveau van inkomen en de prijs van toerisme gebaseerd op het indexcijfer van consumptieprijzen waren de meest gebruikte variabelen tussen 2008 en 2017 (Jiao & Chen, 2018). Deze variabelen werden het meest gebruikt omdat men hiermee de koopkracht van toeristen kan berekenen. Echter, waren dit niet de enige variabelen die ingezet werden om toerisme drukte te voorspellen, tussen 2008 en 2017 werden de volgende variabelen gebruikt (Li en Song, 2008; Ayeh en Lin, 2011; Lee, 2011; Song et al., 2011; Vanegas, 2013; Gunter en Önder, 2015; Petrevska, 2015):

Gebruikte variabelen: 2008-2017
Toeristische uitgaven per vakantie
Toeristische uitgaven per bestemming
Toeristische uitgaven per toeristisch product
Toeristische uitgaven per bezienswaardigheden
Toeristische uitgaven per winkeluitgaven
Het inkomen van de bezoeker
De prijs van toerisme gebaseerd op het indexcijfer van consumentenprijzen
De prijs van toerisme in naastgelegen en/of concurrerende bestemmingen
De inkomsten uit toerisme
De werkgelegenheid in toerisme
Volgroeidheid en ontwikkeling van een bestemming
De ontwikkeling van een bestemming
Demografische variabelen
Sociale variabelen
Politieke variabelen
Toeristisch import
Toeristisch export
Reiskosten
Koopkracht
Het seizoen

Tabel 1: gebruikte variabelen tussen 2008-2017

Samen met de toenemende beschikbaarheid van gegevens met betrekking tot onlineactiviteiten, is er recentelijk een overvloed aan onderzoek gedaan naar het voorspellen van economisch gedrag met behulp van onlinegegevens. Deze onlinedata verwijst naar sporen van consumentengedrag, waarbij interesses en aankopen direct onthuld kunnen worden (Hubbard, 2011). Law et al (2019) omschrijft het internetverkeer als een duidelijke indicator voor de behoeftes van toeristen en kunnen daarom geïntegreerd worden in toerisme drukte voorspelmodellen. De onlinedata zijn beschikbaar in de vorm van **zoekvolumes** of **geposte berichten op sociale media**. Bangwayo-Skeete en Skeete (2015) waren een van de eerste auteurs die met onlinedata een nieuw perspectief boden op het voorspellen van toerisme drukte. In het onderzoek werden zoekopdrachten gericht aan accommodatie en vluchten in- en naar het Caribische gebied gebruikt en werd een verhoogde accuraatheid dan gebruikelijk vastgesteld. Ook daaropvolgende onderzoeken op basis van internetverkeer kwamen tot een soortgelijke conclusie. Pan en Yang (2014) omschreven hoe het integreren van online zoekopdracht gerelateerde data een positief effect heeft op de accuraatheid van korte-termijn voorspellingen. Sinds het onderzoek van Pan en Yang zijn er meerdere succesvolle onderzoeken uitgevoerd waarbij een groot aantal landen Google Trend data gebruiken om toerisme

te voorspellen (Bangwayo-Skeete en Skeete, 2015; Yang et al, 2015, Zewnalov, 2017; Bokelmann en Lessman, 2019). Yang et al (2015), onderstreepte de behoefte naar een algoritme dat de intensiteit van verschillende zoektermen naar een bestemming verzameld, om zoveel mogelijk bijpassende zoektermen te identificeren en de accuraatheid van voorspelmodellen te verbeteren. In sommige gevallen worden deze zoektermen gelinkt aan specifieke zoekregio's. Dat wil zeggen, wanneer een individu in Bremen zoekt op "Höhepunkte der Niederländischen Watteninseln", hij of zij hoogstwaarschijnlijk van plan is de Waddeneilanden te bezoeken. Echter, kan het ook voorkomen dat een groot deel van de bezoekers uit eigen land afkomstig zijn. Dit houdt in dat bepaalde gebieden of zelfs het hele land onder de zoekregio komt te vallen (Bokelmann & Lessman, 2019). Tot slot, is het volgens Gunter en Önder (2016) belangrijk om bij soortgelijke data ook de accuraatheid van lange-termijn voorspellingen verbeteren. De laatstgenoemde auteurs gebruikte tien op websiteverkeer gebaseerde indicatoren onttrokken van Google Analytics. Dergiades et al. (2018) benadrukt de waarde van zoekopdracht correcties met betrekking tot taalbarrières en cultuur- en platform gebonden vooroordelen. Dit onderzoek sluit goed aan bij Li, et al (2020) zijn stelling dat het gebruik van verschillende beoordelingsplatformen '**Multi-platform**' sterk bijdraagt aan toenemende accuraatheid.

Over het algemeen kan gezegd worden dat de meest gebruikte online bronnen voor voorspelmodellen onttrokken wordt van Google en Baidu (de Chinese variant van Google). Echter, kan op basis van de bovengenoemde onderzoeken gezegd worden dat de volgende variabelen gebruikt zijn voor het voorspellen van toerisme drukte tussen 2014 en 2020:

Variabelen onttrokken uit online data: 2014-2020
Zoekopdrachten naar een geheel land
Zoekopdrachten naar een gebied waarin én of meerdere bestemmingen zich begeven
Zoekopdrachten naar een specifieke regio
Zoekopdrachten gerelateerd aan accommodaties
Bijpassende zoekopdrachten van een bestemming
Sentiment analyses op sociale media of beoordelingsplatformen

Tabel 2: variabelen onttrokken uit online data: 2014-2020

1.6 Voorspelmethodieken

Hoewel kwantitatieve voorspelmodellen voorkomen in verschillende vormen en maten, is geen model hetzelfde en worden ze op verschillende manieren gebruikt. Om overzicht te krijgen en duidelijk onderscheid te kunnen maken, houden de volgende hoofdstukken zich bezig met het beschrijven van verschillende type voorspelmethodies- modellen en bijbehorende karakteristieken. Het is vanzelfsprekend dat niet alle voorspelmodellen belicht kunnen worden en daarom zal dit verslag alleen de meest gebruikte en relevant modellen van de toeristische sector benoemen. De volgende hoofdstukken zullen één voor één verschillende methodieken belichten, omschrijven en beoordelen. De volgende drie categorieën zullen behandeld worden:

- Causale modellen (hoofdstuk 2)
- Tijdsreeksmodellen (hoofdstuk 3)
- Kunstmatige Intelligentie (hoofdstuk 4)
- Hybride modellen (hoofdstuk 6) (onderdeel 2e oplevering)

Voor een helder overzicht van alle methodieken en een werkwijze wanneer deze te gebruiken zijn wordt in hoofdstuk 6 de belangrijke bevindingen samengevat in een tabel en flowchart.

2. Causale modellen

2.1 Inleiding

Causale modellen gaan ervan uit dat er een causale relatie bestaat tussen de (on)afhankelijke variabelen. Het toeschrijven van een oorzaak/gevolg relatie wordt dus aangenomen en hoeft niet altijd op waarheid te berusten (ETC, 2021). Bijvoorbeeld, bij het voorspellen van het aantal bezoekers van een stad (A), kan men stellen dat het aantal afhankelijk is van verschillende factoren zoals demografiek (W), koopkracht (X), seizoen (Y), promotiebudget (Z), etc. Een causaal model omschrijft in deze setting de wiskunde verhouding van de variabelen tot elkaar en kan deze wiskundige verhouding gebruiken om een voorspelling te maken (Snellen, 2019).

Ondanks de inzichten die causale modellen kunnen brengen, is het uitermate belangrijk om causaliteit niet te verwarren met **correlatie**. Hiermee wordt bedoeld dat de onafhankelijke variabelen handige tools kunnen zijn om het aantal toekomstige bezoekers te berekenen, het niet daadwerkelijk de oorzaak hoeft te zijn voor een verandering. Zo kan het er in de praktijk op lijken dat het ene variabele de oorzaak is voor een verandering in het aantal toeristen. Terwijl in de werkelijkheid, de variabelen enkel met elkaar in verbinding staan terwijl een derde variabele invloed op beide uitoefent (ETC, 2021).

Desondanks de complexiteit die correlaties en causaliteit met zich meebrengen, zijn causale modellen makkelijker en sneller inzetbaar dan tijdreeksmodellen, waarvoor modelbouw vereist is (Snellen, 2019).

Methodieken:

De volgende methodieken zullen in dit hoofdstuk aan bod komen:

- Simple Linear Regression (enkelvoudig)
- Multiple Linear Regression (meervoudig)
- Vector Auto Regression
- TVP *(onderdeel van 2e oplevering)*
- TVP-STSM *(onderdeel van 2e oplevering)*



2.2 Enkelvoudig-en meervoudige regressiemodellen

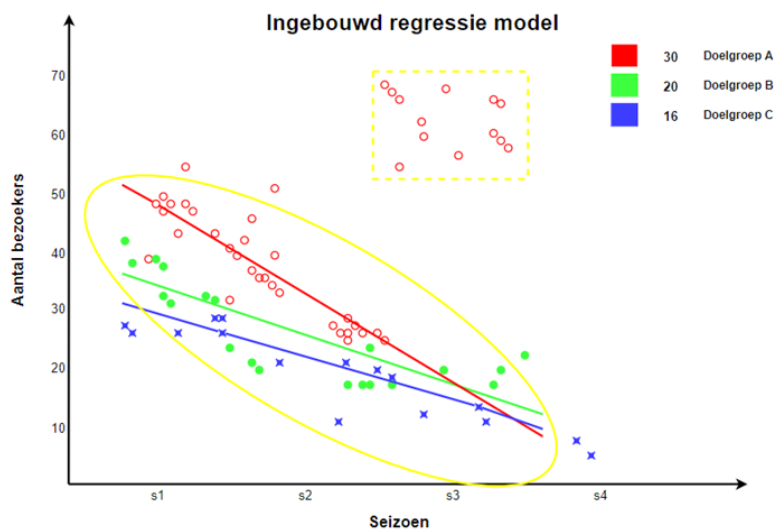
Regressieve modellen zijn tijdreeksmodellen die de wiskundige waarde van **afhankelijke** variabelen berekenen op basis van één of meer **onafhankelijke** variabelen (Ghalekhondabi et al., 2019). Vervolgens wordt een voorspelling gemaakt op basis van de relationele sterkte tot elkaar en de sterkte van het onafhankelijke variabele.

Zo kan een regressief model bijvoorbeeld de correlatie van een specifiek seizoen op inkomend toerisme weergeven (figuur 1). Het model houdt rekening met de correlatie tussen de afgelopen seizoenen en het aantal toeristen dat er in die periode aanwezig waren (Hyndman & Athanasopoulos, 2018). In dit voorbeeld wordt dat weergegeven doormiddel van een spreidingsdiagram. Aan de hand van de hierop geplote datapunten wordt vervolgens een lineaire trendlijn aangeduid

Voor enkelvoudige modellen wordt er net als in figuur 1 enkel één onafhankelijk variabele in acht genomen (seizoenen). In meervoudige lineaire regressiemodellen worden meerdere onafhankelijke variabelen in acht genomen waardoor het mogelijk wordt om met meer precisie voorspellingen te doen (Webfocus Rstat, z.d.). Bijvoorbeeld, wanneer men de prijs van een vliegticket wil voorspellen is het onvoldoende om alleen de lengte van de vlucht als onafhankelijke variabele te nemen. Echter, wanneer men het seizoen, de luchtvaartmaatschappij en bijvoorbeeld de huidige olieprijs toevoegt, kan de nauwkeurigheid van een model in grote mate versterkt worden.

Het doel van een regressiemodel is om trends in tijdreeksdata te bepalen en op basis hiervan trendlijn te bepalen die het beste bij (een) onafhankelijke variabele(n) past. Dit houdt in dat regressie methodes zich goed lenen voor het aanduiden van trends en tegelijkertijd betekent dit dat de nauwkeurigheid van het model sterk afhankelijk is van de beschikbare data. Regressiemodellen maken gebruik van relatief simpele berekeningen die ook uit te voeren zijn in programma's zoals Excel. Regressiemodellen vereisen een lage verwerkingstijd en zijn dan ook snel uit te voeren. Deze factoren samen zorgen ervoor dat het model een lage toegankelijkheidsdrempel heeft.

Echter, komt de eenvoudigheid van deze modellen ook gepaard met een aantal beperkingen. Toerismedrukke wordt sterk beïnvloed door externe factoren. Door slechts een aantal van deze externe factoren mee te nemen in de voorspelling wordt de daadwerkelijke complexiteit van de sector niet voldoende gerepresenteerd. Daarnaast, is het in de gebruikte formule niet mogelijk om de waarde van variabelen tot elkaar bepalen. Voorspellingen worden gemaakt op basis van aannames met betrekking tot relationele verbanden tussen (on)afhankelijke variabelen. Dit betekent dat er aannames worden gemaakt m.b.t. tot causaliteit, maar in de werkelijkheid kan deze veronderstelling niet hard gemaakt worden met enkel- en meervoudige regressiemodellen (Hyndman en Athanasopoulos, 2018; Martínez, 2018). Tot slot, zorgt het in acht nemen van een minimaal aantal variabelen tot een hoge sensitiviteit voor uitschieters en kan hiermee de accuraatheid van een voorspelling sterk beïnvloed worden (ETC, 2021).



Figuur 1: visualisatie van een regressief model, de trendlijnen en ruis.

2.3 Vector Auto Regressie (VAR)

Daar waar enkele-en meervoudige autoregressie en ARMA modellen uitgaan van eenzijdig verkeer van onafhankelijke variabelen naar afhankelijke variabelen, gaat het VAR model ervan uit dat beide variabelen elkaar beïnvloeden. Zo kan bijvoorbeeld het persoonlijk beschikbare inkomen van een persoon zijn uitgave patroon beïnvloeden en vice versa (Hyndman en Athanasopoulos, 2018).

Het Vector AutoRegressie model is voornamelijk interessant voor het voorspellen van verschillende variabelen uit een tijdreeks. Praktisch kan men het model zien als een uitwerking van het meervoudige regressiemodel (Hanck et al., 2020). Het model kan voor het volgende gebruikt worden:

- Het voorspellen van een set aan variabelen zonder verdere interpretatie
- Het toeschrijven van effecten van het ene variabele op het ander
- Testen of variabele A behulpzaam is in het voorspellen van variabele B
- Het analyseren wat plotse en korte impulsen doen met een variabele

Naast deze mogelijkheden komt het model ook met een aantal beperkingen. Zo kan het gebruikt worden voor voorspellingen, echter blijkt de accuraatheid van deze voorspellingen lager uit te vallen dan het ARIMA model (Hanck et al., 2020). Daarnaast heeft het model in het verleden kritiek ontvangen vanwege de aanname dat elk variabele impact heeft op het andere variabele wat het interpreteren van coëfficiënten bemoeilijkt (Hyndman en Athanasopoulos, 2018).



3. Extrapolatie modellen

3.1 Inleiding

Extrapolatie voorspellingsmodellen zijn in plaats van het zoeken naar oorzaak-gevolg correlaties, op zoek naar **terugkerende patronen van waarden in een specifieke periode**. Deze periodes worden tijdreeksen genoemd en komen tot stand door het stelselmatig verzamelen van informatie over een specifieke periode (CBS, 2021). Nadat het extrapolatiemodel in de tijdreeks de benodigde patronen heeft gevonden zal het daar vervolgens een voorspelling mee maken (Reid et al, 2011). Aangezien de methode gebaseerd wordt op terugkerende patronen wordt er altijd vanuit gegaan dat de benodigde informatie voor een voorspelling altijd in de tijdreeks te vinden is. Echter, betekent dit ook dat de gebruikte tijdreeksen altijd over gelijkwaardige periodes verzameld moeten worden en dat de voorspelling vooral betrekking heeft tot de nabije toekomst (Geer, 2012). Het visualiseren van extrapolatie methodes wordt in de meeste gevallen vormgegeven door spreidingsdiagrammen waarbij gebruik gemaakt van vier basiselementen (Snellen, 2019):

- *Niveau*: weerlegt het patroon van waarden rond een constant gemiddelde
- *Trend lijn*: weerlegt het patroon van toe- en afnemende waarden
- *Seizoen*: weerlegt een herhaaldelijk patroon met constante lengtes
- *Cycli*: weerlegt de impact van economische fluctuaties

Doordat extrapolatie methodieken gebruik maken van tijdreeksen kennen zij ook een gezamenlijke limitatie. Namelijk het bevatten van **niet voorspelbare ruis**. Deze ruis kan gezien worden als een waargenomen waarde die als niet veel zeggend beschouwd dient te worden. In kleine aantallen vormt de ruis geen probleem. Echter kan het in grotere aantallen de analyse beïnvloeden en daarmee impact hebben op de accuraatheid van het model.

Methodieken:

De volgende methodieken zullen in dit hoofdstuk aan bod komen:

- Enkelvoudige- en meervoudige regressietechnieken
- ES
- DES
- TES
- SMA
- ARMA
- ARIMA
- SARIMA
- ARFIMA *(onderdeel van 2e oplevering)*
- ETS *(onderdeel van 2e oplevering)*
- STSM *(onderdeel van 2e oplevering)*

3.2 Autoregressief bewegende gemiddeldes: SMA, ARMA, ARIMA, SARIMA en ARFIMA

3.2.1 Simple Moving Average

Het simpel bewegende gemiddelde, ook bekend als **'Simple Moving Average'** (SMA) is een veelgebruikte methode die relatief eenvoudig toe te passen is. De techniek neemt het gemiddelde van het aantal waarnemingen, over de meest recente periodes. Het algoritme houdt balans door verouderde data te verwijderen terwijl het de nieuwste data opslaat. Hierdoor beweegt het gemiddelde mee met de trend en dat is dan ook waar het zijn naam aan heeft verdiend (Snellen, 2019). De methode is zeer gebruiksvriendelijk en kan uitgevoerd worden op toereikende programma's zoals Excel. Deze simpliciteit zorgt er tegelijkertijd ook voor dat methode niet in staat is om de complexiteit van toerisme in acht te nemen. Echter, zijn er methodes die voortborduren op deze theorie.

3.2.2 Auto Regressive Moving Average

Wanneer men gebruikt maakt van tijdreeksen merkt men op dat datapunten die dicht bij elkaar staan meer overeenkomen dan datapunten die over een grotere tijdspan verspreid zijn. Dit fenomeen wordt **autocorrelatie** genoemd. Om accurate voorspellingen te blijven maken is het

belangrijk voor regressiemodellen om rekening te houden met dit fenomeen. Het ARMA model gebruik van twee verschillende technieken; de eerder beschreven autoregressie techniek (AR) en tijdreeks (ETC, 2021).

Het ARMA model komt gepaard met twee kernbegrippen die veel invloed hebben en uiteindelijk zullen leiden tot de (door)ontwikkeling van het model. Deze begrippen zijn **stationairiteit** en **differentiatie**. Stationairiteit verwijst naar het onderdeel van tijdreeks die niet afhankelijk is van de tijd en waarin de dataserie die geobserveerd wordt een constant gemiddelde bevat. Dit zijn bijvoorbeeld tijdreeksen die niet beïnvloed worden door trends. Ongeacht de periode waarin men de data analyseert zou het sterk overeen moeten komen met data uit een andere periode (Hyndman en Athanasopoulos). Het tegenovergestelde, de afwezigheid van trends indiceert het stationair zijn van de geobserveerde waardes. Vele voorspelmodellen, ARMA inbegrepen, zijn niet opgewassen tegen het verwerken van **niet stationaire** data (Sangarshanan, 2018; ETC, 2021). Dit is waar **differentiatie** zijn intrede maakt. Door gebruik te maken van stationaire data in dezelfde tijdreeks, kan men de door trends beïnvloede mediaan stabiliseren door zelf stationaire eigenschappen te creëren (Snellen, 2019; Hyndman en Athanasopoulos, 2018).

3.2.3 Auto Regressive Integrated Moving Average

Dit brengt ons bij de opvolger van de ARMA methode, het ARIMA model. Hierbij staat de 'I' voor **integratie**, wat inhoudt dat originele datapunten meerdere keren vervangen zijn door gedifferentieerde datapunten (Hyndman en Athanasopoulos, 2018). De manier waarop men een ARIMA model toepast wordt beschreven in de naam die men er aan geeft. Zo kan het model als volgt omschreven worden; ARIMA (1.2.0). Hierbij staat de 1 voor het aantal parameters dat is gebruikt, de 2 voor het aantal keren dat differentiatie is toegepast en de 0 het aantal parameters gebruikt door het bewegende gemiddelde (ETC, 2021).

3.2.1 Seasonal Auto Regressive Integrated Moving Average

Als extensie op het ARIMA model is de SARIMA methode ontwikkeld. Hierbij staat de S voor Seasonal. Het concept is grotendeels hetzelfde, echter neemt dit model in staat om de **non-stationaire seizoengebonden** impact op de dataset te verwerken (Hyndman en Athanasopoulos, 2018). Zo kunnen er bijvoorbeeld bij een bestemming vier verschillende seizoenen geïdentificeerd worden die vervolgens meegenomen worden in de voorspelling van het model (Snellen, 2019). Het aanpassen naar seizoengebondenheid wordt verzorgd door een extra auto regressieve polynoom en een polynoom voor het bewegend gemiddelde. Net als bij de ARIMA methode wordt het gebruikte model beschreven met numerieke waardes achter de naam. De eerste drie waardes blijven staan en drie nieuwe waardes worden als volgt toegevoegd: SARIMA (1.2.0) (1.0.1)4. De op het laatst vermelde "4" indiceert het aantal seizoen dat in acht genomen wordt (ETC, 2021).

Tot slot, het maken van voorspellingen op basis van het bewegend gemiddelde en autoregressie kan op verschillende niveaus uitgevoerd worden. Het enkel toepassen van een bewegend gemiddelde kan zonder professionele hulp toegepast worden, echter leidt deze gebruiksvriendelijkheid tot een te versimpelt model dat de complexiteit van bezoekersdrukke niet geheel kan bevatten. De extensies ARMA, ARIMA en SARIMA komen dichterbij het realistisch weerleggen van bezoekersdrukke. Hierin dienen echter wel de non-stationairiteit en seizoengebondenheid van de tijdreeks nadrukkelijk in acht genomen te worden.

3.3 Exponentiële afvlakking: SES, DES, TES

3.3.1 Simple Exponential Smoothing

De meest toegankelijke methode in de reeks van exponentiële afvlakking 's modellen is het '**Simple Exponential Smoothing**' model. Dit model berust zich op het feit dat de meest recente datapunten in tijdreeksen het beste zijn om een voorspelling te maken en des te ouder een data punt is, hoe minder geschikt het is voor een voorspelling. Het model speelt in op deze kennis door gebruikt te maken van een bewegend gemiddelde dat exponentieel zijn voorgaande waardes vermindert (Hyndman en Athanasopoulos, 2018).

Het model is gebruiksvriendelijk en relatief makkelijk om toe te passen. Echter zorgt het gebruiken van gemiddeldes en lineaire trendlijnen er voor dat het model minder geschikt is voor het aanduiden en verwerken van variaties in data (Hyndman en Athanasopoulos, 2018). Hierdoor is het model ongeschikt om voorspellingen te maken op basis van tijdreeksen die trends en/of seizoenen bevatten. Als antwoord op deze beperking is Holt's trend-corrected double exponential smoothing model ontwikkeld.

3.3.2 Double Exponential Smoothing

Holt's Linear Exponential Smoothing model, ook wel het '**Double Exponential Smoothing**' model genoemd. Dit model is gecompliceerder dan zijn voorganger en berekent het afvlakken van waardes en trends. Vervolgens wordt de voorspelling gemaakt door de laatst bewaarde data te linken aan twee datapunten in de toekomst. Dit zorgt er voor dat het model in staat is om tijdreeksen die trends bevatten te verwerken.

Net als zijn voorganger is dit model relatief gebruiksvriendelijk en kan uitgevoerd worden in verschillende statistische software zoals SPSS, R en SAS JMP. Het gebruik van Excel wordt afgeraden omdat het de accuraatheid van de voorspelling zou beperken (ETC, 2021). Daarnaast is het belangrijk om te realiseren dat al hoewel dit model de mogelijkheid biedt om trends te voorspellen, het niet in staat is om seizoengebondenheid in te calculeren. Gezien de intentie om toerismedrukke te kunnen voorspellen zullen extensies van dit model verder belicht worden.

3.3.3 Triple Exponential Smoothing

Dat brengt ons bij het '**Triple Exponential Smoothing**' (TES) model, ook wel bekend als het 'Multiplicative Holt-Winters' model. Dit model bestaat uit een dezelfde elementen als de voorgaande modellen, echter maakt het gebruik van drie afvlakkende formules in plaats van één of twee. Met andere woorden, een extra formule voor het afvlakken van seizoensgebonden data is verwerkt in het model.

Er twee verschillende manieren op het model kan worden toegepast: de additieve methode en de multiplicatie methode. De additieve methode wordt voornamelijk toegepast wanneer seizoensgebonden variaties over het algemeen constant zijn in de tijdreeks. De multiplicatie methode wordt in de meeste gevallen gebruikt wanneer de seizoensgebonden variaties proportioneel veranderen m.b.t. de tijdreeks (Brownlee, 2018).

Ondanks de verhoogde complexiteit van het TES model is het nog steeds mogelijk het model toe te passen in statistische software zoals SPSS, R en SAS JMP.

4. Kunstmatige intelligentie

4.1 Inleiding

Kunstmatige intelligentie (AI) kan beschreven worden als een systeem of machine die de intelligentie van een mens nabootst. Met het nabootsen van menselijke intelligentie wordt bedoeld dat het algoritme leert van eerdere ervaringen en voorbeelden. Hierdoor ontwikkelt het systeem zich en wordt het mogelijk om begrip te ontwikkelen, verbaal te kunnen antwoorden, beslissingen te maken en problemen oplossende acties te ondernemen (IBM, 2021).

Wat is dan doelmatig het verschil met traditionele methodes? Het antwoord ligt in de manier waarin de accuraatheid van econometrische en extrapolatiemodellen beïnvloed wordt door het aantal externe en complexe krachten die meewegen in het ontstaan van bezoekersaantallen. De mogelijkheid van AI om zichzelf te trainen, ontwikkelen en verbeteren op basis van voorgaande prestaties en geïdentificeerde relaties tussen invloed hebbende factoren creëert potentie. Dit verschil is dan ook de waar AI zijn intrede maakt en waar het zich kan onderscheiden van traditionele methodes (Law et al., 2019).

Echter, is het wel belangrijk om te vermelden dat de accuraatheid van AI een veelbesproken onderwerp blijven. Zo prijzen verschillende onderzoeken de modellen als veelbelovend en wijst benchmarking van specifieke onderzoeken op betere performance van AI ten opzichte van traditionele modelering (Lin en Lee, 2013; Joao en Chen, 2018; Law et al., 2019). Dit gezegd hebbende, moet er ook terug verwezen worden methodologische reviews van Li et al. (2005), en Jiao en Chen (2018), “...no single model outperforms other forecasting models under all circumstances”.

(Un)supervised Machine Learning

Kunstmatige intelligentie kan worden toegepast op drie verschillende manieren: ‘Supervised machine learning’, ‘Unsupervised machine learning’ of een combinatie van beide. Men spreekt van Supervised Machine Learning wanneer bestaande variabelen gebruikt worden voor het voorspellen van andere variabelen. Deze methode wordt vaak toegepast wanneer men aan de hand van seizoen, het weer en de promotiecampagne, het aantal bezoekers(X) wil berekenen. Dit is vergelijkbaar met wat er bij autoregressie plaatsvindt (Mailund, 2017). De modellen worden ‘getraind’ aan de hand van bestaande datasets. Het ene onderdeel van deze set wordt gebruikt voor het trainen en een ander onderdeel om de accuraatheid van het model te testen. Hiermee kan van te voren worden aangetoond of het model bruikbaar en effectief is (ETC, 2021).

Unsupervised machine learning staat daar haaks tegenover en wordt gebruikt met het ontdekken van patronen in data wanneer het doel van het algoritme nog onbekend is. Hierbij is het dus nog onbekend wat variabele (X) en variabele (Y) is (Mailund, 2017). Ondanks de losbandigheid van deze methode, kan het een zeer effectieve manier zijn om overzicht te creëren in grote datasets. Een Unsupervised AI is in staat om patronen te leggen tussen bijvoorbeeld demografische kenmerken zoals, uitgave patronen, inkomen, educatie, leeftijd, interesses, reis gedrag en oneindig meer combinaties. Door deze aanpak toe te passen wordt het mogelijk om een random dataset in te delen in verschillend categorieën. Die eventueel weer gebruikt kunnen worden in Supervised AI. Tot slot is het mogelijk om gebruik te maken van een het Random Forest (RF) algoritme. Dit algoritme creëert een groot aantal modellen en bepaald zelfstandig welke van deze modellen de beste oplossing voor het gestelde probleem is (ETC, 2021).

Methodieken

De volgende methodieken zullen in dit hoofdstuk aan bod komen:

- ANN
- SVM
- SVC
- SVR
- RNN
- LSTM

4.2 Artificial Neural Networks

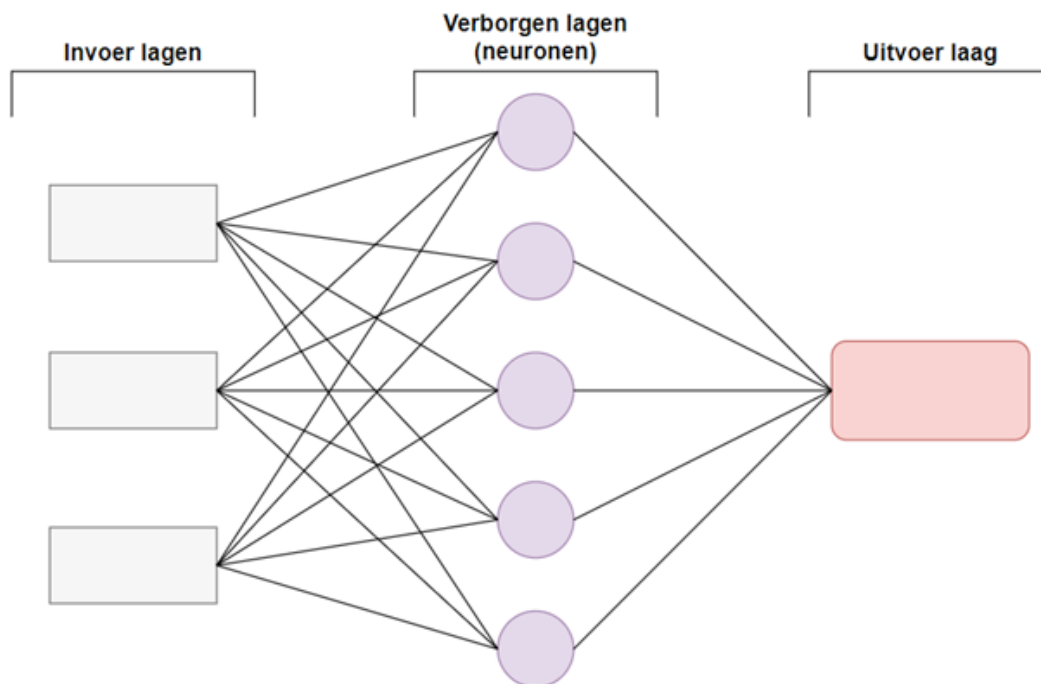
Een van de methodes die regelmatig wordt toegepast als voorspelmethode is het Kunstmatig Neuraal Network, ook wel bekend als een 'Artificial Neural Network' (**ANN**). Dit netwerk gebruikt een non-lineair statistische aanpak en simuleert doormiddel van wiskundige calculaties processen van het menselijk brein en creëert aan de hand hiervan algemene correlaties tussen verschillende variabelen (Li et al., 2019; Snellen, 2019).

Het netwerk bestaat uit drie verschillende lagen (figuur 2) en in elke laag bevinden zich kunstmatige 'neuronen' die individueel communiceren met neuronen in de volgende laag. Het aantal aanwezige neuronen in de in- en output lagen geeft aan hoeveel variabelen er worden meegewogen in het maken van de voorspelling. Elk neuron wordt gevoed met zijn eigen kenmerken en toegerekend gewicht. Dit houdt in dat het model anticipeert op het feit dat sommige variabelen zwaarder meegewogen dienen te worden (Sadatiseyedmahalleh et al., 2016).

Een kunstmatig neuraal netwerk wordt voornamelijk toegepast tot datasets waarin het lastig is om correlaties te identificeren en waarin veel ruis aanwezig is. Daarnaast is het ook mogelijk om multidimensionale data en data waarvan er geen relationele historische data aanwezig is te gebruiken voor het maken van voorspellingen (Karlaftis & Vlahogianni, 2011).

Het nadeel van een ANN is dat het niet mogelijk is om het effect en de invloed van variabelen op de neuronen te meten, waar dat met andere modellen wel tot de mogelijkheden behoort. Hierdoor is het lastig causaliteit aan te tonen tussen twee variabelen. Daarnaast is internsief voor een computer om de berekeningen van een ANN uit te voeren. De accuraatheid van een ANN loopt ironisch genoeg op bij de grootte van de dataset, echter zorgt deze grootte ook voor een lange rekentijd die voor real-time voorspellingen onacceptabel is (Snellen, 2019).

Over het algemeen worden ANN's toegepast voor vraagstellingen waarin zowel de in- en output variabelen analoog zijn (ETC, 2021). Denk hierbij aan afbeeldingen, teksten, documenten of audiofragmenten zoals de fotoherkenning van Google of Audio herkenning als Shazam.



Figuur 2: De 3-laagse structuur van een kunstmatig neuraal netwerk

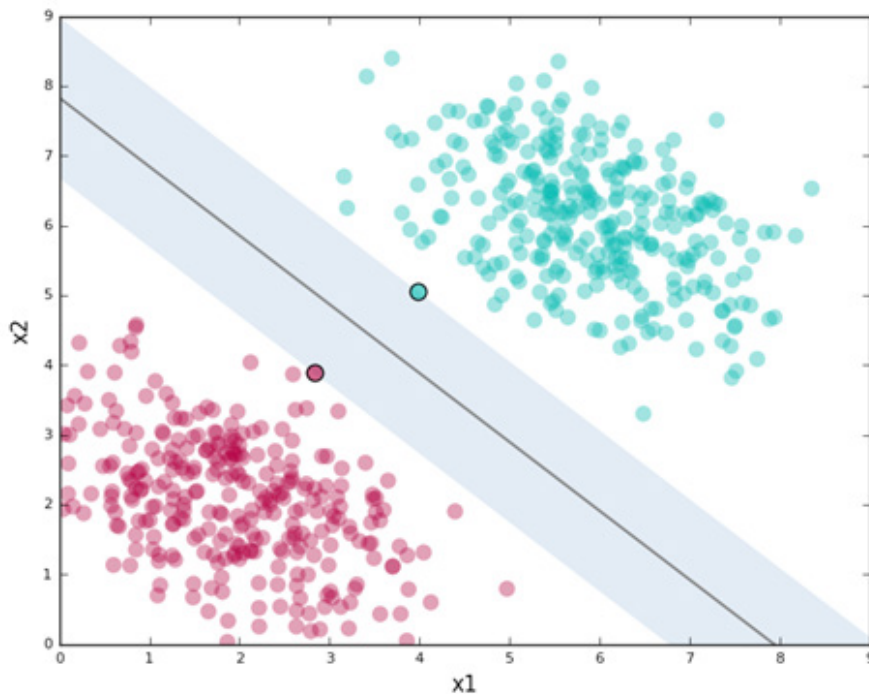
4.3 Support Vector Machines, Regression en Classification

Een Support Vector Machine (SVM) is een zelflerend algoritme dat non-lineaire trends op in-voerniveau omzet naar lineaire trends in een multidimensionale omgeving. Doormiddel van de zelflerende eigenschappen kan het kleine correlaties identificeren in een brei van complexe data (Ghalekhondabi et al., 2019). SVM's kunnen verdeeld worden in twee verschillende categorieën. De eerste categorie, **support vector classificatie (SVC)**, richt zich op het omzetten van data in specifieke classificaties en groepen. De tweede categorie, **support vector regressie (SVR)**, focust op het beperken van de foutmarge om de accuraatheid zo systematische als mogelijk te maken (Basak et al., 2007).

De SVC of SVM wordt gevisualiseerd doormiddel van een scheidingslijn of vlak, afhankelijk van of de data in 2D of 3D gepresenteerd wordt. Deze lijnen of vlakken worden bepaald door zogenoemde 'support vectors'. Dit zijn verschillende datagroepen die ontstaan na het trekken van twee parallelle lijnen tussen verzamelde datapunten. Het gebied dat tussen deze lijnen ontstaat wordt een scheidingslijn of vlak genoemd (Figuur 3).

Het voorspellen van toeristische drukte aan de hand van een SVM werd voor het eerst voorgesteld door Pai en Hong (2005). Hierbij maakte zij gebruik van een SVM bestaande uit meerdere factoren en combineerde dit met een neurale netwerk om de nauwkeurigheid van voorspellingen te vergroten. Hong et al. (2011) ontwikkelde een hybride chaotisch genetisch algoritme (GA) om een voorspelling van toerisme drukte te maken. Hierbij verklaarde zij dat de geminimaliseerde generaliseer fouten een zorgde voor minder training 's fouten voor neurale netwerken. SVM's staan erom bekend beter te presteren dan regressiemodellen en op tijdreeks gebaseerde modellen (ETC, 2021). Daarnaast toont het onderzoek van Lin en Lee (2013), aan dat het SVR model ook accurater presteert dan het ANN model.

Tot slot hebben Support Vector Machines voornamelijk beperkingen op het gebied van ruis. Dit heeft te maken met het feit dat SVM's verschillende groepen proberen te identificeren door datapunten van elkaar te onderscheiden.



Figuur 3: De verdeling van classificaties die ontstaan door de support vectors en hyperlanes

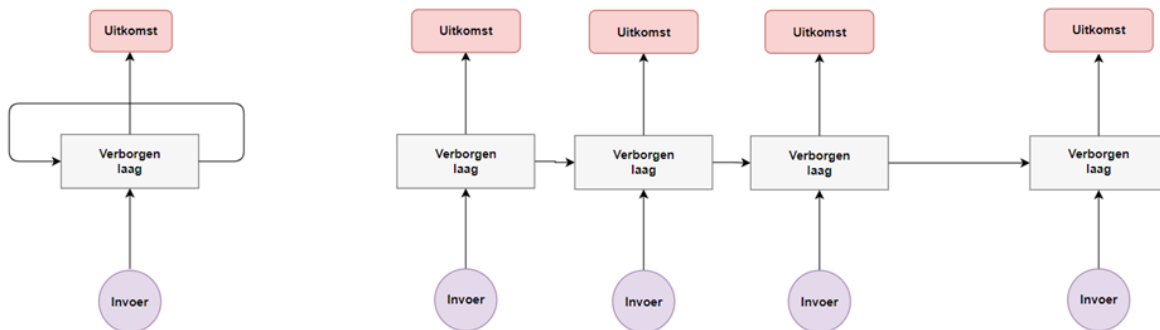
4.4 Recurrent Neural Network

Het RNN is een algoritme dat gebruik maakt van opeenvolgende informatie om voorspellingen te maken. Net zoals een ANN is het een drie-laags model: (1) input laag, (2) verborgen laag en (3) output laag. In elk van deze lagen bevinden zich kunstmatige 'neuronen' die individueel communiceren met neuronen in de volgende laag.

Zoals gevisualiseerd wordt in Figuur 4 is het mogelijk voor het algoritme om zowel van de invoer laag naar de verborgen laag te communiceren als binnen de verborgen laag van neuron tot neuron te communiceren. Dit in combinatie met de potentie om de doorgegeven informatie te onthouden creëert grote waarde voor het model omdat het data uit de invoer laag, indien gewenst, kan combineren met eerder ingevoerde data. Dit maakt het model geschikt voor tijdsgebonden, meervoudige invoerdata en dus het voorspellen van toerisme drukte (Law et al., 2019).

Het verschil tussen de algoritmes van het ANN en RNN-model is dat neuronen in de verborgen laag, specifieke, tijdsgebonden en door zichzelf geselecteerde data aan elkaar doorgeven. Oftewel, de neuronen van het model onthouden de informatie die het heeft overgedragen en kan op basis van gegeven en de huidige data een uitkomst genereren.

Zoals beschreven, is één van de aantrekkelijke kenmerken van een RNN model dat het voorspellingen kan maken op basis van voorgaande informatie. Echter, zijn deze eigenschappen niet zonder beperkingen. De RNN's zijn theoretisch wel in staat om te voorspellen op basis van voorgaande informatie, echter, is de accuraatheid van het model afhankelijk van hoe recent voorgaande informatie is. In de praktijk betekent dit dat RNN's in staat zijn accurate voorspellingen te geven met een input van 5 tot 10 tijdstappen. Dit wordt **Lange-termijn afhankelijkheid** genoemd. Als antwoord op deze afhankelijkheid is het LSTM-model ontwikkeld.

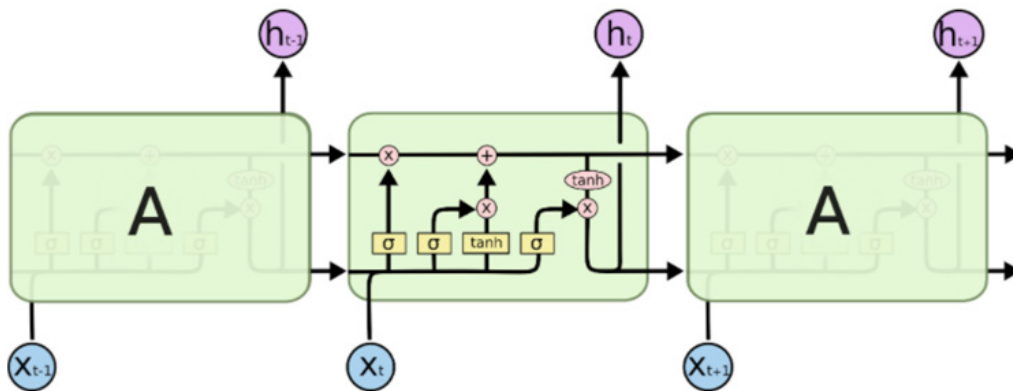


Figuur 4: links; RNN, rechts: een uitgeklapte RNN

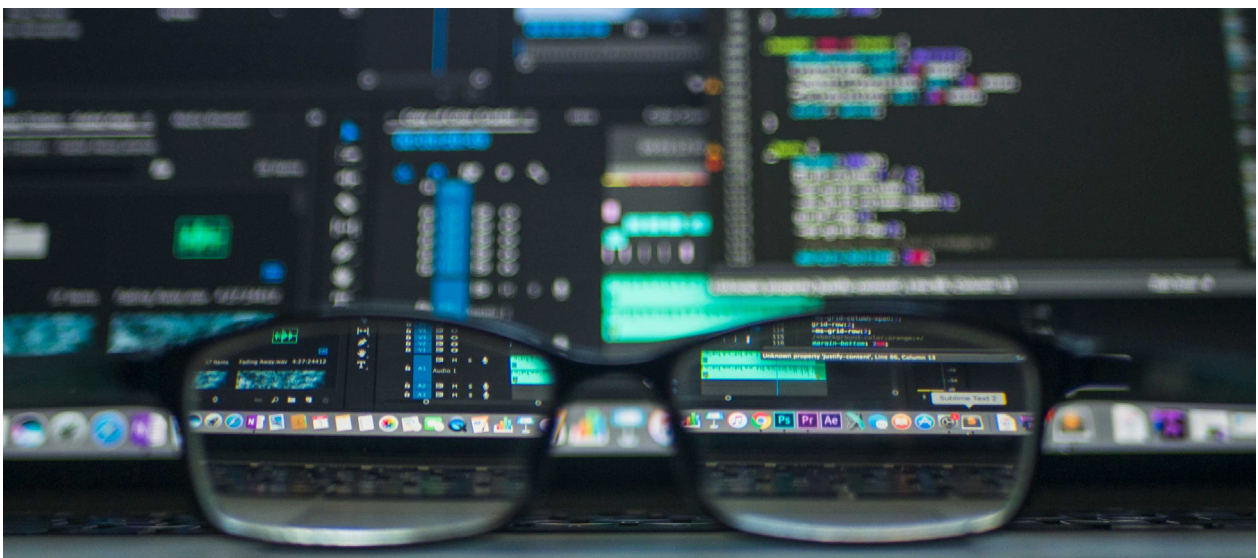
4.5 Long-Short Term Memory

Het LSTM model is een extensie op het RNN model en staat voor **Long-Short Term Memory**. Het doel en de functionaliteit van het model zijn voornamelijk hetzelfde. Het model bevat meerdere lagen waarin zich meerdere neuronen bevinden. De neuronen zijn geprogrammeerd om eerder ontvangen informatie te onthouden en mee te wegen in het doorvoeren van voorspellingen naar de output laag.

Het model heeft net als het RNN model een kettingvormige celstructuur. Echter bevindt zich in de verborgen laag een specifieke functie die kan bepalen welke informatie er wel en niet wordt doorgelaten. Hierdoor wordt niet alle data opgeslagen en kan het model op een effectieve manier de meest relevante informatie opslaan. RNN modellen een lange-termijn afhankelijkheid kennen, zijn LSTM modellen hier onafhankelijk voor. Het model is specifiek ontworpen voor het onthouden van informatie over langere periodes en is hier dan ook zeer effectief in. Tot slot is het vast te stellen dat het LSTM model in functionaliteit sterk overkomt met het RNN model en in de meeste gevallen beter presteert (Olah, 2015).



Figuur 5: blauw: invoer, groen: celstructuur met vergeetfunctie, paars; uitkomst (Olah, 2015)



5. Hybride modellen

5.1 Inleiding

Zoals eerder vermeld, functioneert elk model verschillend per situatie. Dit omdat de voor- en nadelen van de methodieken erg specifiek en opdracht 'gebonden' zijn. Als antwoord op deze nadelen, zijn er over de afgelopen jaren zijn er steeds meer hybride modellen te vinden in methodologieën. Deze trend is het meest aanwezig in het veld van kunstmatige intelligentie maar wordt ook toegepast voor extrapolatie en regressietechnieken.

Over de afgelopen jaren heeft de toenemende beschikbaarheid van data gezorgd voor de inzetbaarheid van steeds complexere en grotere modellen. Het kunnen verwerken van complexe datasets kan ervaren worden als een groot voordeel, maar kan ook leiden tot niet-transparante resultaten. Ook al kan het model correct zijn in de voorspelling, het is in vele gevallen niet in staat om te verklaren hoe het tot deze voorspelling is gekomen, ook wel '**black-box models**' genoemd. In vele industrieën wordt een 'onbegrijpelijk' model dan ook gezien als lastig te vertrouwen en zorgt dit voor zowel moeizame implementatie als acceptatie.

De huidige generatie, komt voornamelijk sterk uit de voeten met sterk gedefinieerde opdrachten. Daarentegen is het lastig voor de algoritmes om getrainde eigenschappen effectief toe te passen in verschillende contexten. Zo is het gemakkelijk voor de algoritmes om voorspellingen te maken in een gesloten omgeving zonder onzekerheid, bijvoorbeeld een schaakspel. In deze omgeving kan het algoritme zich blijven ontwikkelen terwijl alles om hem heen gelijk blijft. Echter, bestaat onze Wereld uit onzekerheden zullen de in het vorige hoofdstuk beschreven modellen altijd onderworpen worden aan de groeiende veranderingen en complexiteit van de menselijke beschaving en zijn omgeving. De nieuwe hybride algoritmes bezitten verbeteringen op meerdere vlakken. Zo zijn ze beter in staat om de context te begrijpen, te beredeneren en vervolgens eerder geleerde eigenschappen toe te passen in een nieuwe omgeving. Dit verhoogt niet alleen de intelligentie van de modellen, maar ook de **transparantie, verklaarbaarheid en inzetbaarheid**.

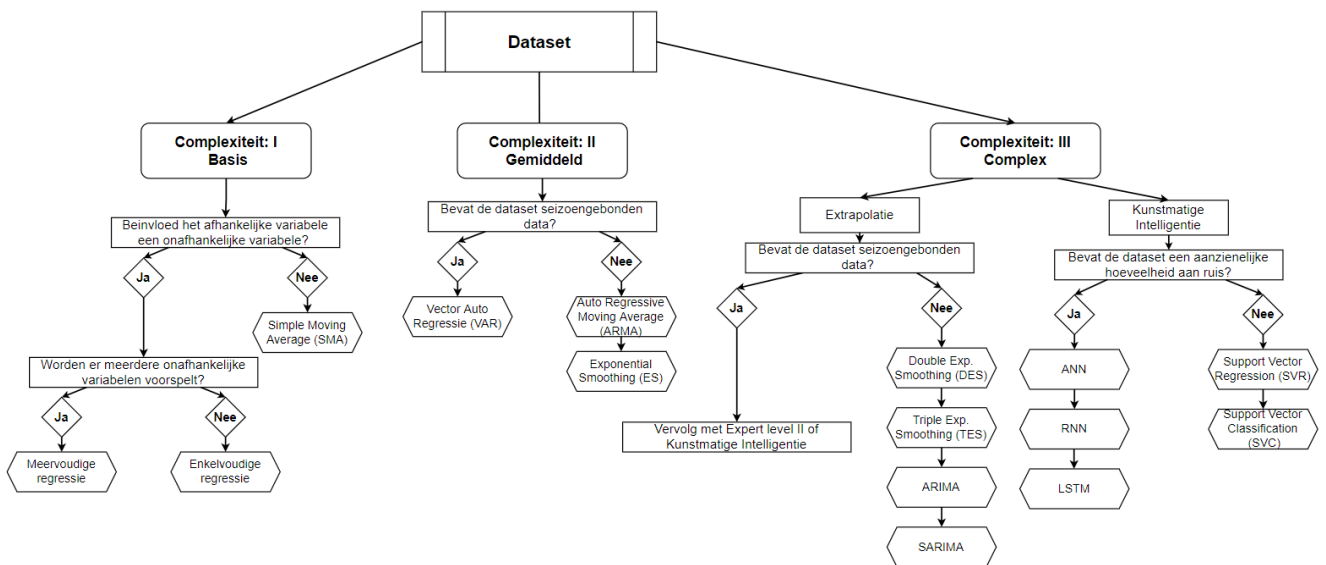
Methodieken:

De volgende methodieken zullen in **tweede oplevering** aan bod komen:

- AGA (SVM + Adaptive GA)
- SVR + FOA
- SVR + GA
- MGFFS (Genetic Fuzzy expert system + Data preprocess)
- ES + Naive + ARIMA



6. Flowchart & modellenoverzicht



Method	Model	Mogelijkheden	Beperkingen	Vereiste software en computer-vermogen	Complexiteit
Causaal	Enkelvoudige regressie	Voorspellen o.b.v. autoregressie	Ongeschikt door simplicitéit	Minimaal	Basis
Causaal	Meervoudige regressie	Voorspellen o.b.v. autoregressie	Ongeschikt door simplicitéit	Minimaal	Basis
Causaal	Vector Auto Regressie (VAR)	Voorspellen zonder interpretatie Toeschrijven effect van variabele op variabele Test	Voorspellingen kennen lage accuraatheid Voorspellingen minder accuraat dan ARIMA en SARIMA Kan geen coëfficiënten interpreteren	Gemiddeld	Gemiddeld
Extrapolatie	Simple Moving Average (SMA)	Voorspelt op basis van het bewegend gemiddelde	Ongeschikt door simplicitéit	Minimaal	Basis

Extrapolatie	Auto Regressive Moving Average (ARMA)	Voorspelt op basis van auto regressie en het bewegend gemiddelde Kan dienen als Benchmark voor ingewikkeldere modellen	Ongeschikt door incapabiliteit om trends en seizoenen te verwerken Accuraatheid wordt overtroffen door meeste AI modellen	Minimaal	Gemiddeld
Extrapolatie	Auto Regressive Integrated Moving Average (ARIMA)	Voorspelt op basis van auto regressie en het bewegend gemiddelde Kan trend gebonden datasets verwerken Kan dienen als benchmark voor ingewikkeldere modellen.	Ongeschikt door incapabiliteit om	Specialistische software	Complex
Extrapolatie	Seasonal Auto Regressive Integrated Moving Average (ARIMA)	Voorspelt op basis van auto regressie en het bewegend gemiddelde Kan trend en	Accuraatheid kan overtroffen worden door AI	Specialistische software	Complex
Extrapolatie	Exponential Smoothing (ES)	Het maken van effectieve korte-termijn voorspellingen	Kan geen trend en	R, SPSS, SAS JMP	Gemiddeld
Extrapolatie	Double Exponential Smoothing (DES)	Het maken van effectieve korte-termijn voorspellingen uit een trend gebonden dataset	Kan geen	R, SPSS, SAS JMP	Complex
Extrapolatie	Triple Exponential Smoothing (ETS)	Het maken van effectieve korte-termijn voorspellingen uit trend en	Lage accuraatheid voor lange-termijn voorspellingen	R, SPSS, SAS JMP	Complex

Kunstmatige intelligentie	Artificial Neural Networks (ANN)	<p>Kan kleine correlaties identificeren</p> <p>Kan datasets verwerken waarin zich een hoge hoeveelheid van ruis bevindt.</p> <p>Voorspellen op basis van zowel</p>	<p>In niet in staat om de relationele sfeer tussen variabelen te meten.</p> <p>Om de accuraatheid te kunnen verbeteren moet men de dataset vergroten</p> <p>Een grote dataset leidt tot onrealistisch vereist rekenvermogen.</p>	<p>Professionele software</p> <p>Zeer intensieve verwerking</p>	Complex
Kunstmatige intelligentie	Support Vector Machine (SVM)	<p>Kan kleine correlaties identificeren</p> <p>Het maken van een aanpassing is relatief gemakkelijk</p> <p>Relatief eenvoudige implementatie</p> <p>Hoge accuraatheid t.o.v. op tijdreeks en autoregressie gebaseerde modellen</p>	<p>Functioneert niet optimaal wanneer de dataset veel ruis bevat</p> <p>Heeft een lange trainingstijd bij grotere datasets</p>	<p>Professionele software</p> <p>Intensieve verwerking</p>	Complex
Kunstmatige intelligentie	Support Vector Classification (SVC)	<p>Kan kleine correlaties identificeren</p> <p>Het maken van een aanpassing is relatief gemakkelijk</p> <p>Relatief eenvoudige implementatie</p> <p>Hoge accuraatheid t.o.v. op tijdreeks en autoregressie gebaseerde modellen</p>	<p>Functioneert niet optimaal wanneer de dataset veel ruis bevat</p> <p>Heeft een lange trainingstijd bij grotere datasets</p>	<p>Professionele software</p> <p>Intensieve verwerking</p>	Complex

Kunstmatige intelligentie	Support Vector Regression (SVR)	<p>Kan kleine correlaties identificeren</p> <p>Het maken van een aanpassing is relatief gemakkelijk</p> <p>Relatief eenvoudige implementatie</p> <p>Hoge accuraatheid t.o.v. op tijdreeks en autoregressie gebaseerde modellen en ANN's</p> <p>Bevat beperkte foutmarges</p>	<p>Functioneert het beste in combinatie met een andere AI model (Hybride samenstelling).</p> <p>Functioneert niet optimaal wanneer de dataset veel ruis bevat</p> <p>Heeft een lange trainingstijd bij grotere datasets</p>	<p>Professionele software</p> <p>Intensieve verwerking</p>	Complex
Kunstmatige intelligentie	Recurrent Neural Network (RNN)	<p>Neuronen onthouden eerder verwerkte informatie</p> <p>Genereert de uitkomst op basis van huidige en historische waarden</p>	Lange-termijn afhankelijkheid	<p>Professionele software</p> <p>Intensieve verwerking</p>	Complex
Kunstmatige intelligentie	Long-Short Term Memory (LSTM)	<p>Neuronen onthouden eerder verwerkte informatie</p> <p>Genereert uitkomst op basis van huidige en historische waarden</p>	Recentelijk pas geïntroduceerd voor het voorspellen van toerismedrukke	<p>Professionele software</p> <p>Intensieve verwerking</p>	Complex

7. Bronvermeldingen

- Amadeus. (2020), five example of predictive analytics in the tourism industry, Insights. Retrieved From: <https://amadeus.com/en/insights/blog/5-examples-predictive-analytics-travel-industry>
- Peng, B (2014) "A meta-analysis of international tourism demand forecasting and implications for practice", *Tourism Management*, vol 45 [2014] pp. 181-193.
- Ayeh, J, K., Lin, S.,. (2011) 'Estimating tomorrow's tourist arrivals': forecasting the demand for China's tourism using the general-to-specific approach. *Tourism and Hospitality Research* 11(3): 197- 206.
- Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454–464.
- Basak, D., Pal, S. and Patranabis, D.C. (2007), "Support vector regression", *Neural Information Processing-Letters and Reviews*, Vol. 11 No. 10, pp. 203-224.
- Brand, D. (1973), "Travel demand forecasting: some foundations and a review", *Highway Research Board Special Report*, No. 143, Williamsburg, VA.
- Brownlee, J. (2018). Exponential Smoothing for time series forecasting in python. *Machine learning mastery*. Onttrokken van:<https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/#:~:text=Triple%20Exponential%20Smoothing%20is%20an,to%20the%20univariate%20time%20series.&text=As%20with%20the%20trend%2C%20the,exponential%20change%20in%20the%20seasonality>.
- Centraal Bureau van Statistiek (2021) Tijdsreeksen: keuzemenu programmeertaal. Onttrokken uit: <https://www.cbs.nl/nl-nl/onze-diensten/open-data/statline-als-open-data/tijdsreeksen#:~:text=Wanneer%20dezelfde%20variabele%20op%20verschillende,kwartaal%20of%20elk%20jaar%20geactualiseerd>.
- Data Science Partners (2020). Support Vector Machine in Python: Uitleg & Tutorial. Onttrokken uit: <https://pythoncursus.nl/support-vector-machine>
- European Travel Commission, Insights & Silverbullet. (2021). *Handbook on Tourism Forecasting Methodologies* (2021). Brussel, Belgium. ISBN no: 978-92-95107-45-8
- Gharibi, N. (2020), *Journal of Tourism Futures*, "The evolution of predictive models and tourism." Emerald Publishing Limited., ISSN 2055-5911,
- Crouch, G. I. (1994). "The Study of International Tourism Demand: A Survey of Practice." *Journal of Travel Research*, 33: 41-54.
- Geer, F. (2012). Wageningen University and Research. *Stromingen* 18 (2012), nummer 2. Onttrokken van: <https://edepot.wur.nl/315307>
- Ghysels, E. (2016) *Macroeconomics and the reality of mixed frequency data*. *Journal of Econometrics*
- Elsevier, <http://dx.doi.org/10.1016/j.jeconom.2016.04.008>
- Grady, J. S., Her, M., Moreno, G., Perez, C., & Yelinek, J. (2019). Emotions in storybooks: A comparison of storybooks that represent ethnic and racial groups in the United States. *Psychology of Popular Media Culture*, 8(3), 207–217. <https://doi.org/10.1037/ppm0000185>
- Gunter, U., & Onder, I. (2016). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research*, 61, 199–212. Hong, W.-C., Dong, Y., Chen, L.-Y. and Wei, S.-Y. (2011), "SVR with hybrid chaotic genetic algorithms for tourism demand forecasting", *Applied Soft Computing*, Vol. 11 No. 2, pp. 1881-90.
- Hubbard, D. W. (2011). *Pulse: The New Science of Harnessing Internet Buzz To Track Threats and Opportunities*. Hoboken, NJ: Wiley.

- Karlaftis, M., & Vlahogianni, E. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. Elsevier.
- Kooistra, W. (2012) Vrije Universiteit Amsterdam, simulation of coordinated traffic light control using model predictive control.
- Hyndman, R, J., & Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd edition. Monash University, Australia
- Hanck, C., Arnold, M., Gerber, A., Schmelzer, M. (2020). Introduction to Econometrics with R. Chair of Econometrics, Department of Business administration and Economics. Duisberg-Essen, Germany
- Li, H., Hu, M., Li, G (2020). Forecasting tourism demand with multisource big data. Elsevier, *Annals of tourism research*, 83,. <https://doi.org/10.1016/j.annals.2020.102912>
- Li, G., Song H., (2008). Tourism demand modelling and forecasting; a review of recent research. Elsevier, *Tourism Management*, 29, 2, 22-29
- Li, G., Song, H., and Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research*, 44, 82-99.
- Li, G., Song, H., and Witt, S. F. (2010). Tourism demand modelling and forecasting: How should demand be measured. *Journal of Travel Economics*, DOI: 10.5367/000000010790872213
- Martínez, R. (2018). Gradient talks, “If correlation does not imply causation then what does?”. Ontrokken van: <https://medium.com/gradient-talks/if-correlation-does-not-imply-causation-then-what-does-8fa462943b84>
- Olah, C. (2015). Understanding LSTM networks. Ontrokken van: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56(7), 957–970.
- Petrevska, B. (2015). Economic research, predicting tourism demand by A.R.I.M.A models. Pages 939-950. Ontrokken van: <https://doi.org/10.1080/1331677X.2017.1314822>
- R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, 75, 410–423.
- Samsudin, R., Saad, P. and Shabri, A. (2010), “Hybridizing GMDH and least squares SVM support vector machine for forecasting tourism demand”, *International Journal of Research and Reviews in Applied Sciences*, Vol. 3 No. 3, pp. 274-9.
- Sangarshanan. (2019, 7 april). Time series Forecasting — ARIMA models - Towards Data Science. Medium. <https://towardsdatascience.com/time-series-forecasting-arima-models-7f221e9eee06>
- Sgaier, S. K., Huang, V., Charles, G. (2020). The Case for Causal AI (SSIR). *Stanford Social Innovation Review*: https://ssir.org/articles/entry/the_case_for_causal_ai.
- Snellen, L., (2019). Een instrument voor het voorspellen van de bezettingsgraad van parkeergarages. Universiteit Twente. Retrieved from:https://essay.utwente.nl/77336/1/Snellen_BA_BMS.pdf
- Vanegas M (2013) Co-integration and error correction estimation to forecast tourism in El Salvador. *Journal of Travel and Tourism Marketing* 30(6): 523-537
- Vapnik, V. (2013), *The Nature of Statistical Learning Theory*, Springer Science & Business Media, New York, NY.

Vragen over deze publicatie? Neem contact op met:

Centre of Expertise Leisure, Tourism & Hospitality

Mgr. Hopmansstraat 2
4817 JS Breda

Email: info@celth.nl
Website: www.celth.nl

