

Meerwaarde van data science voor het openbaarvervoerbeleid, een verkenning

Tjeu Corbeij – Provincie Noord-Brabant – mcorbeij@brabant.nl

Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 13 en 14 oktober 2022, Utrecht

Samenvatting

De sector openbaar vervoer staat aan de vooravond van belangrijke vernieuwingen op het gebied van data en informatie. Mogelijk al in 2023 komt – eindelijk – de herkomstbestemmingsmatrix (HB-matrix) voor alle reizen per OV in Nederland als tabel tot stand en beschikbaar voor onderzoek en beleid. Een belangrijke belofte van de chipkaart wordt ingelost, kort voordat de chipkaart zelf wordt vervangen.

Dit paper verkent de mogelijkheden van data science, machine learning en algoritmen voor het openbaarvervoerbeleid van concessieverleners en overheden die investeren in infrastructuur.

De drie belangrijke beleidsonderdelen en datatoepassingen in het openbaar vervoer zijn:

- Doet de vervoerder wat is afgesproken (monitoring van concessies). Hier heeft data science weinig meerwaarde. Wel kunnen data over stiptheid, rijnsnelheid en oponthoud relevant zijn voor andere beleidsvragen.
- Wat is het optimale netwerk qua diensten en infra, gemeten naar geld, reisweerstand en betrouwbaarheid. Hier heeft data science een grote meerwaarde: via een herkomstbestemmingsmatrix als model kan gewerkt worden aan het optimaliseren van netwerken van diensten en van infrastructuur. Hiervoor kan een grotere 'dieptescherpte' worden bereikt dan met de nu gebruikte modellen.
- De positie van OV binnen mobiliteit en maatschappelijke ontwikkelingen. Ook hier heeft data science een grote meerwaarde: via prognoses en te benaderen potentie van het openbaar vervoer, en als basis voor transparante(re) bestuurlijke besluitvorming.

Het lijkt verstandig om zoveel mogelijk aan te sluiten bij en voort te bouwen op al bestaande verkeers- en rekenmodellen en bij de lopende ontwikkelingen rond die modellen.

Kunnen nieuwe data en bewerkingen worden aangesloten op bestaande modellen of is een radicaal andere modelarchitectuur nodig? Die vraag zal in het veld moeten worden beantwoord. Dat is een voortschrijdend proces, waarvan dit paper een momentopname is. Als eerste stap lijkt een assessment van de data, in combinatie met het prepareren van het informatiepakket voor een komende aanbesteding, mogelijk. Vervolgstep kan dan een experiment zijn om de HB-matrix als model te ontwikkelen, waarbij de feitelijke data uit de eerste stap mede functioneren als controle op het model.

Op het colloquium van 13 en 14 oktober 2022 verwacht ik ook een deel van het vervolg te kunnen vertellen.

1. Inleiding

1.1 Aanleiding(en)

- De data uit de OV-chipkaart biedt de mogelijkheid om een herkomstbestemmingsmatrix (HB-matrix) samen te stellen voor het hele openbaar vervoer, op de enkele reiziger nauwkeurig en in de vorm van een (enorme) tabel. Zonder problemen is dat niet. O.m. is de toedeling van losse kaarten en het toevoegen van data uit MAAS en andere registreerbare last mile vervoersbewegingen mogelijk, maar niet op voorhand eenvoudig of volledig.
- Wetgeving en concessievoorwaarden gaan ervoor zorgen dat deze data beschikbaar komt voor onderzoek en beleid en voor het informatiepakket bij het aanbesteden van een volgende concessie. Ook nu is het al mogelijk om via Trans Link Systems met deze data te werken, onder voorwaarden, kosten en met een beperkte bewaartermijn.
- Mijn werkgever, de provincie Noord-Brabant, wil datagedreven werken. Dat is natuurlijk mede een modewoord, maar het betekent in de praktijk extra aandacht voor de mogelijkheden van data en data science voor het verbeteren van beleid. Ik heb deelgenomen aan een intensieve cursus Data science voor overheden bij JADS in 's-Hertogenbosch van ca. 6 maanden. Mijn aandachtsgebied was ook mij eigen werkterrein: het openbaarvervoerbeleid. De resultaten en mijn ideeën over een vervolg deel ik in dit paper.
- De belangen zijn groot. De omzet in de exploitatie van het OV ligt pré-corona landelijk in de orde grootte van vijf miljard euro. Dan hebben ook kleine verbeteringen al een relevant financieel effect. Ca. 50% van de kosten van de concessies bus/tram/metro (BTM) wordt door de provincies en vervoerregio's gesubsidieerd. Met de concessieverlening is in Noord-Brabant ca. € 95 mln. structureel per jaar gemoeid, ca. 15% van de provinciale middelen. Op een gemiddelde werkdag werden in Noord-Brabant pré-corona ca. 190.000 busreizigers vervoerd. De gemiddelde bezetting van een busrit is ca. 8 personen. De provincie werkt de komende jaren aan de verlening van drie regionale OV-concessies voor 8 tot 10 jaar. Het is van belang om maximaal voordeel te halen uit de beschikbare data, ten behoeve van de uitvraag, het dienstregelingsontwerp en/of de beoordeling van offertes. Daarnaast wordt een jaarlijks wisselend bedrag geïnvesteerd in infrastructurele voorzieningen ten behoeve van het OV. Optimalisering binnen dit veld en dit budget geeft al snel belangrijke structurele voordelen voor reiziger en samenleving. Een vergelijkbaar verhaal geldt voor de andere concessieverleners, inclusief het Rijk voor het hoofdrailnet.

2. Het gebruik van data in het OV-beleid.

Het Openbaarvervoerbeleid kent drie gebieden waar data relevant zijn: 1) de monitoring van concessies, 2) de verbetering of optimalisering van netwerken, en 3) het integrale ruimtelijke en sociaaleconomische beleid. De mogelijkheden van data science verschillen per beleidsgebied.

2.1 De monitoringsfunctie

Concessieverleners monitoren de nakoming van de verplichtingen uit de verleende concessies. Deze monitoring is de afgelopen 10 jaar per concessieverlener ingevuld, maar gelang hun beleidsdoelen. Overheden krijgen hun data van de vervoerder geleverd en verwerken die in dashboards en rapportages aan hun besturen, of zij hebben het hele verantwoordings- en rapportageproces uitbesteed aan een bedrijf of aan de concessiehouder, vergezeld van contractuele garanties over juistheid en kwaliteit.

Per OV-autoriteit gelden andere verplichtingen en criteria. De samenwerkingsorganisaties DOVA en CROW werken voor de concessieverleners aan een 'ladenkast' en eenduidige definities om deze data en de informatie daaruit breed en open toegankelijk te maken. Daarmee is deze functie m.i. adequaat ingevuld en is er voor data science t.a.v. de monitoring geen echte meerwaarde. Wel kunnen data over stiptheid, rijsnelheid en oponthoud relevant zijn voor andere beleidsvragen. Zij kunnen gebruikt worden in andere bewerkingen ten bate van optimalisering van netwerken of infrastructuur.

2.2 De ontwikkeling van voorzieningen en netwerken

Dit is het traditionele werkgebied van vervoerkundigen bij vervoerbedrijven, consultants en overheden. Zij werken met verkeersmodellen, business cases, maatschappelijke kosten-batenanalyses, vervoersprognoses enz. Daarbij is steeds de complexe afweging aan de orde hoe de drie belangrijkste waarden: kosten, reisweerstand en betrouwbaarheid het beste kunnen worden gediend en welke voorzieningen en investeringen daarvoor nodig zijn.

Met data science kan het mogelijk worden om deze verhouding te optimaliseren voor te kiezen waarden (bvb een maximaal budget of kortste totale reistijd).

2.3 Integraal beleid

Mobiliteit is een gevolg van ruimtelijke, economische en sociale ontwikkeling en kan belangrijk profiteren van datagedreven beleid. Brabant moet tot 2030 ruim 10% meer woningen bouwen boven op het huidige bestand, en tot 2050 bijna 20%, naar huidige prognoses. Voortgaande verstedelijking betekent een versterkt toenemende rol voor OV en gedeelde mobiliteit. Met een data science model kan gedetailleerd verband gelegd worden tussen de te verwachten vervoersvraag naar OV en de ruimtelijke, economische en sociale veranderingen van steden, dorpen en wijken. Ook kan gesignaleerd worden wanneer de vraag achterblijft bij de verwachting en welke factoren daaraan mogelijk bijdragen.

3. Concrete beleidsvragen

In het vorige hoofdstuk heb ik de belangrijkste datatoepassingen binnen het OV aangegeven. In een dialoog met betrokkenen binnen en buiten mijn eigen organisatie zijn meer specifiek vragen naar voren gekomen waarvoor behoefte bestaat aan betere modeloplossingen en die met data science kunnen worden versterkt. Het gaat om de volgende beeld van behoeften aan instrumenten voor beleid, waarbij niet iedereen steeds alle onderdelen benoemt.

- Een bewerkbare en voor ieder toegankelijke herkomstbestemmingsmatrix voor trein, metro, tram bus en overige gedeelde mobiliteit, waarbij de problemen met vertrouwelijkheid, privacy, en korte beschikbaarheid van data zijn opgelost, en waarmee flexibel gerekend en voorspeld kan worden in een modelomgeving.
- Een grotere nauwkeurigheid en detaillering van modeluitkomsten voor openbaar vervoer.
- Een methode om per verbinding en netwerk de omvang van de potentiële reizigers en de mogelijkheden voor vervoersgroei te benaderen.
- Een overzicht van de reisbehoefte binnen een netwerk en over de hele vervoersketen, de optimale inrichting van het netwerk en de investeringen daarin. O.a. ten behoeve van het voorbereiden van een aanbestedingsprocedure of bij de beoordeling van offertes.
- Afweging en voortdurende verbetering of optimalisering van de maatschappelijke waarden reistijd/reisweerstand, kosten en betrouwbaarheid/robuustheid binnen een model.

- Benaderen van de omvang en de locatie van vervoersgroei als gevolg van maatschappelijke ontwikkelingen (TBOV 2040; woningbouwopgave; schaa sprong spoor) en de daarvoor benodigde (infrastructurale) maatregelen voor het faciliteren van de (groei van de) toekomstige ruimtelijke en sociaal economische ontwikkeling van de samenleving.
- Bereiken van een helder afwegingskader voor investeringen in infrastructuur, zowel HUB's als doorstromingsmaatregelen, alsook de 'zachte' maatregelen als reisinfo en beleving, e.e.a. om de reisweerstand verder te verlagen.
- Transparantie van bestuurlijke besluitvorming t.a.v. het OV en de faciliteiten daarvoor.

De kwaliteit van beleidsbeslissingen is van directe invloed op de maatschappelijke functie en bruikbaarheid van het OV. Die betreft de mobiliteit van mensen: reistijd, reisweerstand en betrouwbaarheid van het vervoer, maar ook de kosten van het vervoer: de faciliteiten (infra), de subsidies en de prijs voor de reiziger.

Het openbaar vervoer staat niet op zichzelf, maar heeft zijn plaats in het geheel van de mobiliteit. Binnen het palet van modaliteiten verschilt de rol en het gebruik sterk naar gelang de ruimtelijke inrichting van een gebied of regio. Er vindt een mate van concurrentie plaats tussen OV en andere modaliteiten, afhankelijk van de voorkeur van de reizigers van het moment of naar gelang de omstandigheden. Zo gaan er bij regen of sneeuw ineens veel fietsers met de bus. Bij treinstoringen slaagt een groot deel van de passagiers erin een andere vervoerwijze te regelen. Hoge parkeertarieven en beperkte parkeerruimte maken stadscentra onaantrekkelijk voor de auto.

Deze factoren komen tot uitdrukking in de factor reisweerstand per modaliteit. Het is van belang om adequaat te sturen op vermindering van de reisweerstand ten bate van de reiziger in het OV.

Als voorbeeld zijn hieronder de vijf speerpunten in het Brabantse OV-beleid weergegeven met de aangrijpingspunten voor data science. Voor collega-provincies geldt m.m. hetzelfde.

Fig. 1. Vijf speerpunten van het Brabantse OV-beleid en opties van data science.

Speerpunten binnen OV	Toepassing data science
1) Doorontwikkelen snelle busverbindingen HOV en BRT	Optimalisering van infra en samenhangend netwerk van BTM, HOV en Spoor; integraal beleid
2) Inzet op knooppunten (HUBs)	Optimalisering van infra en samenhangend netwerk van BTM, HOV en Spoor; integraal beleid
3) Versterken kleine vervoersstromen	Optimalisering; beoordelen offertes; voorspellen betrouwbaarheid
4) Relatie met en informatie aan de reiziger (o.a. MAAS)	samenhangend netwerk, voorspellen betrouwbaarheid
5) Spoor: 'basis op orde' en 'schaalsprong'	Optimalisering van infra en samenhangend netwerk van BTM, HOV, Spoor; integraal beleid

4. Sturing en Data science.

Binnen het openbaar vervoer vinden belangrijke modernisering plaats. De denkstap wordt gemaakt van dienstregelingen die zo goed mogelijk op de vraag zijn afgestemd naar gedeelde mobiliteit van deur tot deur. Tevens bestaat een grote behoefte om beter inzicht te hebben in de potentiële reiziger en de effecten van integraal ruimtelijk en sociaal-economisch beleid en investeringen t.b.v. de reiziger, de kwaliteit en de kosten.

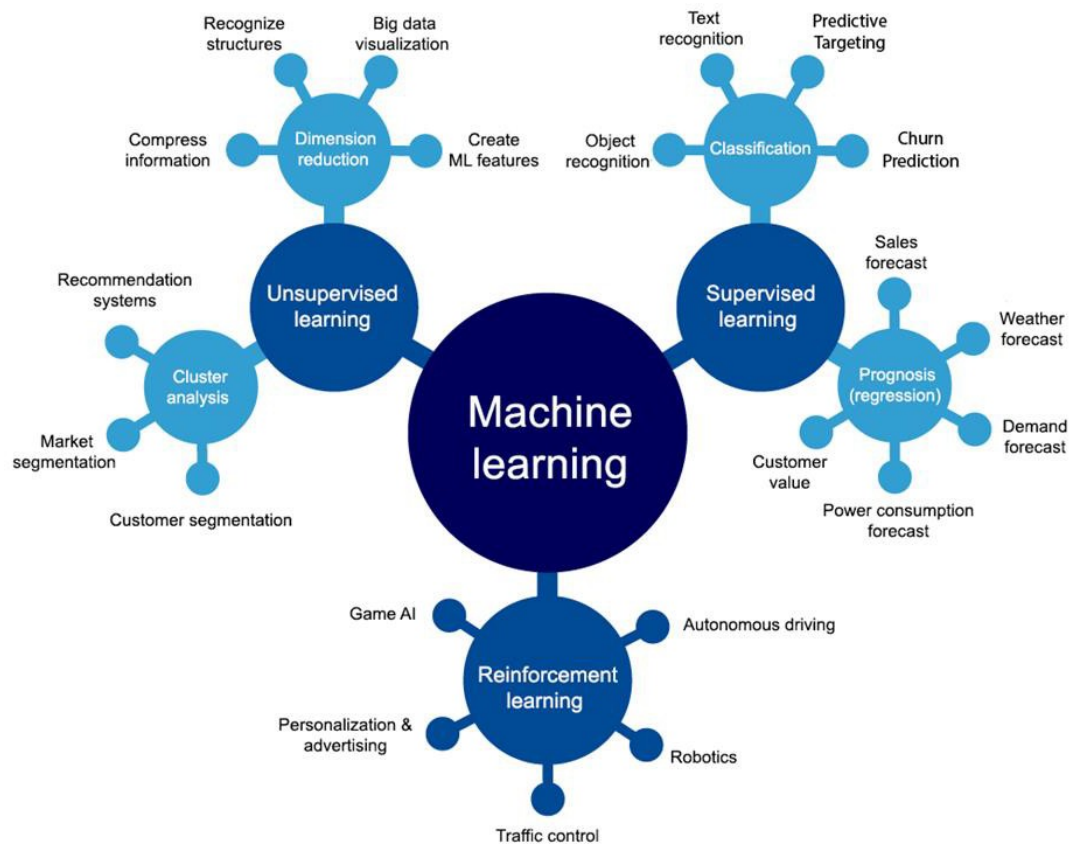
In het vorige hoofdstuk hebben we gezien welke beleidsvragen daaruit voortvloeien. Daar horen ook nieuwe toepassingen van data en informatie bij.

4.1 Data science

Data science omvat de bewerking en het toesnijden van data op de gestelde vragen, de technieken om via algoritmen het proces van machine learning uit te voeren met de data en de toepassing van de gevonden resultaten op het vraagstuk, en de evaluatie daarvan plus de nodige verbeteringen die daaruit voortvloeien. Dit proces is ook internationaal gestandaardiseerd in het zgn. CRISP-DM proces (Cross Industry Standard Process for Data Mining).

Uitleg van het hele systeem van data science voert hier te ver. Van belang is dat data science een reeks van technieken heeft die op verschillende soorten vraagstellingen kunnen worden toegepast. Voor de vraagstellingen in het OV zijn vooral de technieken voor supervised learning en reinforcement learning relevant. Vaak wordt ook trial and error gebruikt waarbij de data door verschillende modellen van machine learning worden gejaagd om te kijken of de uitkomsten een meerwaarde hebben.

Fig. 2: overzicht technieken van Machine learning.



4.2 Sturing

De uitkomsten bestaan in een verband met een mate van waarschijnlijkheid tussen een antwoord op de gestelde vraag en een mogelijke veelheid aan factoren die aan dat antwoord bijdragen. Bij voorbeeld het verband tussen het aantal reizigers tussen A en B en de factoren die enerzijds bijdragen aan het gewicht en de aantrekkingskracht van een kern of station, zoals arbeidsplaatsen, inwoners adrestdichtheid en anderzijds de factoren die bijdragen aan de reisweerstand, zoals afstand, overstap en gevoel van (on)veiligheid. Het kwantificeren en relateren van alle factoren vindt plaats met algoritmen en modellen. Het eindresultaat moet zijn een model waarmee het effect van veranderingen (bv. meer inwoners of kortere reistijd) betrouwbaar en gedetailleerd kan worden voorspeld.

Hiermee wordt duidelijk dat naarmate het model beter wordt, er meer gedetailleerd en meer specifiek gestuurd kan worden op factoren die van invloed zijn op het gebruik van het OV. Voor de reiziger zijn waarden als betrouwbaarheid en reistijd centraal. Data science en algoritmen zijn speciaal geschikt om de waarschijnlijkheid van het bieden van betrouwbaar vervoer, het halen van aansluitingen en het analyseren van oorzaken te benaderen. Ook de optimalisering van een netwerk qua capaciteit en lijnvoering kan worden bereikt via algoritmen, waarbij de gewenste waarden als kosten, reistijd en betrouwbaarheid ingesteld kunnen worden.

4.3 Huidige opties en modellen.

De provincie staat voor belangrijke beleidsbeslissingen die regelmatig t.a.v. het provinciale en landelijke OV moeten worden genomen. Die vragen zijn op hoofdlijnen:

1. Wat is het beste netwerk voor de reiziger, de middelen en het beleid.
2. Wat zijn de meest effectieve investeringen in infrastructuur en faciliteiten.
3. Maak betrouwbare prognoses van het effect van beleid en scenario's voor de ontwikkeling van de vervoersvraag.
4. Onderbouw een programma voor de meest effectieve investeringen in de toekomst.

De eerste twee zijn 'statisch': zij worden beantwoord voor de bestaande situatie. De derde en vierde vraagstelling zijn meer voorspellend: zij pogen een antwoord te geven op vragen over te verwachten of denkbare toekomstige maatschappelijke en sociaaleconomische ontwikkelingen en hun invloed op het gebruik van het OV. De vraagstellingen zijn ook te onderscheiden naar enerzijds optimalisering van het netwerk voor de reiziger en de kosten (1 en 3), en anderzijds naar het benaderen van een optimale investeringsagenda (2 en 4).

De huidige modellen gelden als keuzemodellen: de reiziger maakt keuzen op basis van factoren die de reisweerstand groter of kleiner maken. De omvang van de vervoersstroom wordt mede bepaald door de grootte van de kernen of haltes, de mate van verstedelijking, de attractiviteit en de plaats in het netwerk. In feite is hier een zwaartekrachtmodel aan de orde: de aantrekking tussen twee lichamen (locaties), uitgedrukt in aantallen reizigers, gedeeld door een functie van de reisweerstand. Op deze basisformule zijn vele factoren van invloed die de omvang van de reizigersstroom mede bepalen. Met algoritmen kunnen deze factoren uiteengehaald worden en daarmee beter inzichtelijk en beter stuurbaar worden gemaakt.

Voor het OV zijn vanuit de OV chipkaart uiterst gedetailleerde gegevens beschikbaar, veel gedetailleerder dan voor het autoverkeer. Momenteel wordt gepoogd om deze gegevens te incorporeren in de bestaande modellen BBRM en NRM zoals die door de provincie Noord Brabant en door Rijkswaterstaat worden gehanteerd. Binnen de huidige stand van deze modellen is het OV alleen op stedelijk of regionaal niveau te modelleren. Het is de vraag of deze twee totaal verschillende databronnen goed binnen één model gecombineerd kunnen worden.

5. Strategisch ontwerp van een OV-model

5.1 Opbouw rond beleidsvragen

De analyse en bewerking van data om de beoogde resultaten te bereiken is complex en bestaat uit meerdere lagen, processen en bewerkingen. Met het strategisch ontwerp van het model wordt verkend hoe we modelmatig de verbindingen kunnen leggen tussen de data en de wensen van de 'klant', neergelegd in de vier hoofdvragen.

Voor een strategisch ontwerp van een model sluit ik aan bij de vier hoofdvragen uit paragraaf 4.3. Elke vraag heeft een eigen proces om vanuit de data via diverse bewerkingen tot het gewenste resultaat te komen. Ik deel het model aldus in in vier kolommen, een voor elke van de vier vragen en gerelateerd aan Geld, Reisweerstand en Beleidsdoelen (GRB).

1 Beste netwerk op basis GRB	2 Beste investeringen GRB	3 Prognose effecten beleid of scenario's	4 Prognose investeringen bij beleid of scenario
---------------------------------	------------------------------	---	--

De opgave per kolom kan als volgt worden samengevat.

Kolom 1: Beste netwerk op basis GRB

Het invoeren binnen het instrument van de juiste data; dataprocesing waar nodig.
Bepalen van de 'massa' van haltes of kernen binnen het model op basis van relatieve belang van relevante factoren (bv. inwoners, adrestdichtheid, onderwijs, recreatie)
Bepalen van de (factoren van) reisweerstand binnen het (zwaartekracht)model.
Samenstellen van de HB-matrix als model waarmee de herkomsten en bestemmingen van reizigers kunnen worden benaderd en voorspeld op basis van de factoren massa en reisweerstand.

Schaalbaar voor verschillende gebiedsgroottes: stedelijk, stadsregionaal, concessie-landsdelig, landelijk en landelijk + internationaal niveau.

Optimalisering van netwerk naar geld, reisweerstand en beleidsdoelen.

Kolom 2: Beste investeringen GRB

Vertrekpunt is het benaderde optimale netwerk uit kolom 1.

Voorspellen van de reistijden en de robuustheid van het netwerk.

Effecten van investeringen op de reisweerstand voor de reizigers

Stroomschema van de reizigers over de infrastructuur in kaart.

Vergelijkbaar maken van vertraging en de ernst van de vertraging in relatieve en absolute zin binnen het netwerk en voor de reiziger (bepalen van prioriteiten).

Afwegen effect consumptieve uitgave versus investeringen; MKBA-factoren.

Lijst met de meest effectieve investeringen in relatie tot de doelen.

Kolom 3: Prognose effecten beleid of scenario's

Vertrekpunt is het benaderde optimale netwerk uit kolom 1.

Invoer van scenario's, beleidskeuzen en (pakketten) investeringen.

Doorrekenen van die keuzes via het model van de HB-matrix.

Optimaliseren van het ontworpen netwerk naar geld, reisweerstand en beleidsdoelen.

De gevonden waarden (nieuwe fysieke HB-matrix) leggen over geografische en socialegegevenskaarten.

Beeld creëren van nieuwe capaciteit en netwerk in relatie tot reizigersstromen.

Kolom 4: Prognose investeringen bij beleid of scenario

Vertrekpunt is het benaderde optimale netwerk uit kolom 3

Effecten van investeringen op de reisweerstand voor de reizigers

Stroomschema van de reizigers over de infrastructuur in kaart.

Vergelijkbaar maken van vertraging en de ernst van de vertraging in relatieve en absolute zin binnen het netwerk en voor de reiziger (bepalen van prioriteiten).

Afwegen effect consumptieve uitgaven versus investeringen; MKBA-factoren.
Lijst met de meest effectieve toekomstige investeringen in relatie tot de doelen.

5.2 Schematische weergave instrument

In het volgende schema worden de onderdelen van het model weergegeven. Daarbij wordt in kleur aangegeven op welke onderdelen naar mijn inzicht data science een meerwaarde kan betekenen. Ik kom daarbij tot negen toepassingen van modellen van machine learning, die ik in paragraaf 5.3 kort uiteenzet.

Fig. 3: Strategisch ontwerp van het model op hoofdlijnen.

Producten	1 Beste netwerk op basis GRB	2 Beste investeringen GRB	3 Prognose effecten beleid of scenario's	4 Prognose investeringen bij beleid of scenario
Resultaat	Optimaliseren netwerk	Optimaliseren investeringen	Optimaal netwerk bij scenario	Optimale investeringen bij scenario
Beleidsdoelen	[4] Afweging tussen Geld, Reisweerstand, Beleidsdoelen (GRB)	[8] Consumptieve uitgaven versus investeringen; opbrengst/subsidie	[4] Afweging tussen Geld, Reisweerstand, Beleidsdoelen (GRB)	[8] Consumptieve uitgaven versus investeringen; opbrengst/subsidie
Tussenproduct	[3] HB-matrix: de kans dat er X reizigers van A naar B rijden op tijdstip Y	[7] Kwaliteit x reizigers Stroomschema reizigers en reisweerstand over de infra	[9] Prognose HB-matrix voor scenario	[7] Kwaliteit x reizigers Stroomschema reizigers en reisweerstand over de infra bij scenario
Tussenproduct	[2] (technisch) massa HB, [1] verplaatsingsweerstand	[6] Reistijden en [5] Robuustheid Input: optimale netwerk en HB-matrix	Input: scenario-info, HB-matrix, nieuwe infra, woningbouw.	[6] Reistijden en [5] Robuustheid Input: optimale netwerk en HB-matrix bij scenario
DATA: Onderhoud, opslag, processing, actualisering.	Data top 16; parameters	Data top 16; parameters	Mobiliteitsbeleid Maatschappij en cultuur RO en SE ontwikkeling Scenario's Strategie Investeringen	Mobiliteitsbeleid Maatschappij en cultuur RO en SE ontwikkeling Scenario's Strategie Investeringen

5.3 Vereisten aan de negen modellen van machine learning.

1. **De verplaatsingsweerstand.** Verplaatsingsweerstand bestaat uit diverse 'harde' en 'zachte' elementen: de reistijd, de overstap(tijd), de ervaren reissnelheid, het comfort van het voertuig, de mogelijkheden tot werken of amusement, de ervaren (on)veiligheid, reisinformatie, betaalgemak, kosten van de reis en de concurrentie/kwaliteitsverschil met andere modaliteiten. Via een model kunnen deze factoren in een onderlinge verhouding worden geplaatst. Tot nu toe gebeurt dat via enquêtes. Het isoleren van de goede data is een belangrijke opgave.

Het hiervoor geschikte model of modellen: Random forest voor het bepalen van de relatieve betekenis van de relevante factoren in relatie tot de aantallen reizigers.

2. **De (technische) massa van herkomst en bestemming.** Aan elke stad, station, halte, wijk kan een massa worden toegekend binnen de zwaartekrachtformule waarmee de reizigersstromen kunnen worden benaderd en voorspeld. Deze massa bestaat uit een reeks van factoren, zoals inwonersaantal, bedrijven en

onderwijsinstellingen, recreatieve bestemmingen, adressendichtheid enz. De onderlinge verhouding van die factoren en de relatie met de reizigersaantallen (herkomst en bestemming) kan via een model worden gelegd.

Het hiervoor geschikte model of modellen: Random forest voor het bepalen van de relatieve betekenis van de relevante factoren in relatie tot de aantallen reizigers.

3. **De samenstelling van de HB-matrix volgens een waarschijnlijkheidsmodel.**

Voor het OV bestaan data per reis en reiziger, van halte naar halte en het tijdstip waarop gereisd wordt. Voor autoverkeer en OV-reizen via de huidige metingen is die matrix veel grover en kunnen uitspraken gedaan worden op de schaal van een stad, regio of grotere postcodegebieden. Het samenstellen van een gedetailleerde HB-matrix voor het OV is een complexe bewerking, die binnen enkele jaren landelijk gerealiseerd zou moeten worden. Het beheer daarvan ligt bij Translink systems en de belangen van privacy, bedrijfsvertrouwelijkheid en mededinging geven tamelijk strakke kaders aan de beschikbaarheid en bruikbaarheid van die data. Een alternatief kan worden bereikt door het combineren van algoritmen en nieuwe databronnen (de HB-matrix OV) met bestaand modellen op basis van verplaatsingsweerstand en zwaartekracht. Daarmee kan de nauwkeurigheid en bruikbaarheid van modellen voor het OV m.i. sterk worden verbeterd. De chipkaartdata kunnen dan periodiek worden gebruikt om het model te kalibreren.

Het hiervoor geschikte model of modellen zijn: Supervised regression model t.b.v. voorspellen hoeveel reizigers er op tijdvak Y reizen van A naar B, voor het hele netwerk.

4. **De relatieve en instelbare afweging tussen geld, reisweerstand en reistijd (GRB).** Dit is de optimaliseringsfunctie, waarmee bij een gewenst budget en de gegeven reisweerstand de beste reistijden voor de reiziger kunnen worden gegenereerd via een afgeleid geoptimaliseerd netwerk.

Het hiervoor geschikte model of modellen: Hiervoor zijn verschillende modellen denkbaar. Supervised regression, waarbij steeds de best mogelijke uitkomst wordt gezocht voor de toevoeging van eenheden (reizen).; Unsupervised, gericht op segmentation van reizigersgroepen tegen laagste kosten of reinforcement learning waarbij beste oplossingen over de infrastructuur worden geleerd.

5. **De robuustheid van de reistijden per traject.** Afgeleid wordt de mate van zekerheid dat de reiziger op tijd, t vroeg of te laat aan zal komen, of zijn aansluiting haalt of mist.

Het hiervoor geschikte model of modellen: Supervised regression model waardoor de gevoeligheid van reistijden wordt berekend en voorspeld aan de hand van gegevens over infra, voertuigen en gevoeligheden voor storing en vertraging.

6. **De te verwachten reistijden per traject.** Er zijn nauwkeurige gegevens over de spreiding van de feitelijke reistijden t.o.v. de dienstregeling. Door deze te relateren aan de infrastructuur, kunnen de verbanden tussen wijzigingen in de infra en de daarbij behorende reistijden worden benaderd. Dit kan via een model (de waarschijnlijke verdeling/spreiding van reistijden op een traject) of via een directe berekening en meting (basistijd met een afwijking).

Het hiervoor geschikte model of modellen: Supervised regression model waardoor reistijden worden berekend en voorspeld aan de hand van gegevens over infra en voertuigen.

7. **De kwaliteit voor de reiziger,** uitgedrukt in een stroomschema van de omvang van reisweerstand en reizigers. Hier wordt afgeleid wat de waarde voor de reiziger en het netwerk is van een verbetering, die invloed heeft op de factoren van de

reisweerstand. Bij voorbeeld de effecten van een kortere of langere busbaan. Het is mogelijk om voor alle belangrijke baten uit de MKBA (Maatschappelijke kosten-batenanalyse) een inzicht te krijgen op basis van deze aanpak.

Het hiervoor geschikte model of modellen: Supervised learning via classification of regression, wat het beste werkt.

8. **De verandering in de optimalisering van het netwerk** door de investeringen (verschuiving consumptief naar investeringen vv. en rendement vs. kostendekking). Hiermee wordt benaderd of en hoe een verschuiving van middelen van exploitatie (consumptie) naar infrastructuur (investeringen) een positief effect heeft voor de reiziger en de functie van het OV binnen de mobiliteit.

Het hiervoor geschikte model of modellen: Supervised learning via classification of regression, wat het beste werkt.

9. **De prognose op basis van scenario of input** kan worden berekend op basis van enerzijds scenario-input en anderzijds het model van het netwerk en de HB-matrix.

Qua model lijkt hiervoor een regression aan de orde.

6. De bestaande beleidsomgeving

De (complexe) beleidsomgeving is hieronder vereenvoudigd weergegeven.

Fig 4: Relevante actoren in het OV.

	Modaliteiten en partners O: hoofdfactor x: betrokken	Trein Infra	Trein exploitatie	BTM Infra	BTM Exploitatie	Regiotaxi c.a.	Contractueel groepsvervoer	Collectief Flex	Collectief Fiets	Collectief Auto, Taxi	Te Voet
1	Rijk systeemverantwoordelijk	O	O	x	x	x		x	x	O	
2	Rijk Railbeheerder (ProRail)	O	x	x							
3	Rijk concessieverlener	x	O						x	x	
4	Rijk wegbeheerder (RWS)			O							
5	Provincie/VR concessieverl.		O		O	O		O			
6	Provincie als opdrachtgever	x	x	O		x			x		
7	Provincie wegbeheerder			O							
8	Gemeente als opdrachtgever					O	O	x			
9	Gemeente wegbeheerder	x		O		O		x	x	x	x
10	Vervoerder	x	O	x	O	O	O	O	O	O	
11	Beheerder HUB / Station	O		O		x		x	x	x	x
12	Reizigers(organisaties)		x		x	x		x			
13	Dienstverleners (MaaS e.a.)		x		x	x		x	x	x	x

Deze complexiteit maak voortgang en vernieuwing weerbarstig. De wijze waarop dit kan is een aparte en lastige vraag. Daarvoor zijn mij de volgende uitgangspunten aangereikt.

- Werk samen met collega-overheden en met de landelijke kennisplatforms.
- Werk samen op nationaal niveau en maak toetreden per regio/concessie mogelijk.
- Betrek kennisinstituten en HBO/universitaire vakgroepen.
- Denk multidisciplinair
- Vorm in overleg een kopgroep die wil experimenteren en uitproberen.

- Begin met het verzamelen van en bekend worden met de data.
- Maak van de ontwikkeling van de HB-matrix als model een experiment naar de opzet van een Hackathon.
- Sluit zoveel mogelijk aan bij de bestaande modellering van verkeer en OV.

7. Stappenplan

Bij het schrijven van dit paper staat mij het volgende stappenplan voor ogen. Mogelijk heeft mijn inzicht zich op het moment van presentatie tijdens het colloquium verrijkt. Uiteraard deel ik dat dan met de deelnemers.

Binnen deze opdracht richt ik mij op een statisch model, gericht op verbeteringen over langere periodes. Het is gericht op beleidsanalyse en advisering van optimale keuzes van lijnvoering en investeringen in voorzieningen, gegeven het budget en de te verwachten ruimtelijke en sociaaleconomische ontwikkelingen. In wezen steeds een van tevoren beschreven uitgangssituatie.

1. Waarderen en leren kennen wat al kan: leren werken met en waarderen van de nieuwe data uit de OV-chipkaart c.a. binnen de bestaande verkeers- en rekenmodellen. Eerste optie hiervoor is een onderzoek ten behoeve van de voorbereiding van de komende aanbesteding in Zuidoost Brabant. Trans Link Systems (TLS) is de organisatie waar alle info uit de OV-chipkaart bij elkaar komt. Zij leveren onder strikte voorwaarden data over het OV op basis van herkomsten en bestemmingen.
2. Ontwikkelen van een HB-matrix als model, als *'digital twin'* van de werkelijke reizigersstromen. Er zijn veel factoren van invloed binnen een model dat de werkelijkheid van de reizigersstromen in het OV moet kunnen benaderen. Globaal gaat het om twee soorten factoren.
 - 1) Factoren die het gewicht van een kern of halte bepalen, zoals aantal inwoners, arbeidsplaatsen, onderwijs, winkels, adresdichtheid enz.
 - 2) Factoren die de reisweerstand bepalen, zoals afstand, overstappen, historische verbanden en grenzen, rijsnelheid, gevoel van veiligheid, verblijfswaarde in het voertuig en op haltes en stations enz.

Dit kan via een experiment in de vorm van een hackathon, waarbij de initiators de data en faciliteiten verzorgen voor de deelnemende teams.

De vraag is naar een flexibele HB-matrix met een grote 'dieptescherpte', ook voor kleine stromen.
3. Met de HB-matrix als model kunnen we naar twee kanten verder ontwikkelen:
 - 1) De optimalisering van netwerken, zowel qua dienstverlening als qua infrastructuur, met als resultaat een betere verhouding tussen kosten, opbrengsten en dienstverlening.
 - 2) Betrouwbare en gedetailleerde prognoses en scenario's opstellen voor ruimtelijke, maatschappelijke en economische ontwikkelingen en voor investeringen in infrastructuur.
4. Bouw een integraal doorstroommodel voor weg en rail en voeg daar ook de data aan toe die nodig zijn voor de batenkant van een MKBA. Maak e.e.a. schaalbaar naar de beleidsbehoefte van Rijk, concessieverleners en investeerders in infrastructuur, d.w.z. maar landelijk niveau (incl. grensoverschrijdend), landsdelig, regionaal/per concessie en per gemeente of stedelijke regio.

Literatuur en bronnen

- Brands T.; Spruijt J.; Drift S. van der (2019) Data in gebruik: bouw landelijk OV-netwerk. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 21 en 22 november 2019, Leuven.
- CBS (2018) Onderweg in Nederland (ODiN) 2018, Onderzoeksbeschrijving
- CBS (2020) Onderweg in Nederland (ODiN) Plausibiliteitsrapportage 2019.
- DOVA 2021-22 Vergaderstukken MIPOV, Datamanagement.
- Giro. Over de HASTUS OV-software. Website <https://www.giro.ca/en-ca/our-solutions/hastus-software/>
- Heynickx, M.; Egeraat, M. van; Kiel, J.; (2013) De BrabantBrede ModelAanpak (BBMA) – nieuwe toekomstinzichten in de praktijk gebracht. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 2013, 21 en 22 november, Rotterdam.
- JADS Leergang digitalisering voor de overheid. Diverse auteurs en docenten; 2021-22
- Kiel, J.; Grol, R. van; Heynickx, M.; (2015) De provinciebrede modelaanpak. Aanpakken of loslaten? Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 19 en 20 november 2015, Antwerpen.
- Koopal, R.J.; Heynickx, M.; (2017) De BrabantBrede ModelAanpak, lessons learned? Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 23 en 24 november 2017, Gent
- Loop, H van der; Peter Bakker, P.; Savelberg, F.; Kouwenhoven, M.; Helder, E.; (2018) Verklaring van de ontwikkeling van het ov-gebruik in Nederland over 2005-2016. Ministerie van Infrastructuur en Waterstaat/KIM Den Haag.
- Ministerie IenM (2017) Nationale Markt en Capaciteitsanalyse 2017. Hoofdrapport
- Oort, N. van; Ebben, M.; Kant, P.; (2015) Innovatieve toepassingen van OV chipkaartdata: buiten de lijntjes van datagedreven OV onderzoek. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 19 en 20 november 2015, Antwerpen.
- Peter Bakker, P.; Warffemius, P.; (2017) De bereikbaarheidsindicator uitgewerkt voor openbaar vervoer: BBI-ov - Het beleidskader, uitwerking en toepassing. Kennisinstituut voor Mobiliteitsbeleid (KiM)
- Planting T.J. (2016) Ontwerpmethoden van Dienstregelingen; Verkennend onderzoek naar de ontwerpmethodiek per fase in het planningsproces bij NS & ProRail. Masterthesis 2016 TUDelft.
- Provinciaal bestuur van Noord-Brabant. (2019) Addendum bij Bestuursakkoord 2020-2023. Stuknr. 4580481.
- Provinciaal bestuur van Noord-Brabant. (2019) Bestuursakkoord 2020-2023 'Samen, Slagvaardig en Slim: Ons Brabant'
- Provinciaal bestuur van Noord-Brabant. (2021) Begroting 2022 en bijlagenbundel Provincie Noord-Brabant (PAS 55-21)
- Provinciaal bestuur van Noord-Brabant. (Geen jaarvermelding) Waardengedreven digitaal transformeren, Datavisie Provincie Noord-Brabant 2020-2025. Stuknr 4606251.
- Romph, E. de ; Bruno Kochan, B.; Clerx, W.; (2018) Groei van de steden prikkelt modelbouwers! Een Activity Based Model voor Rotterdam, het ABMR. Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk 22 en 23 november 2018, Amersfoort
- Warffemius, P.; (2015) Effecten van veranderingen in reistijd en daaraan gerelateerde kwaliteitsaspecten in het openbaar vervoer. Kennisinstituut voor Mobiliteitsbeleid
- Wikipedia: Gravitatiewet van Newton.
https://nl.wikipedia.org/wiki/Gravitatiewet_van_Newton