



## **X** NS en CQM maken reizigersvoorspellingen

## **U** consistent en inzichtelijk

**NS werkt al jaren met rekenkundige modellen om het aantal reizigers per trein te kunnen voorspellen. Over het algemeen werkten de modellen goed, maar afwijkingen waren soms moeilijk te verklaren en de totstandkoming van de voorspellingen was niet altijd even inzichtelijk. NS wilde het systeem inzichtelijker en consistent in de tijd maken en schakelde hiervoor één partij in: CQM zette NS op het juiste spoor.**

Het voorspellen van het aantal reizigers per trein is voor NS van groot belang voor een goede klanttevredenheid. Het bepaalt de materieelinzet en daarmee de kosten van vervoer en het comfort voor de reiziger: voldoende reizigers moeten kunnen zitten. Gedetailleerde reizigersvoorspellingen maken een optimale balans tussen materieelinzet en vervoerscomfort mogelijk en daarmee ook de balans tussen betaalbaarheid en klanttevredenheid. Maar wat als een voorspelling voor een individuele trein ineens afwijkt? Bijvoorbeeld als het model voorspelt dat er in één specifieke trein in de doorgaans rustige maand

juni juist meer reizigers zitten dan in de drukke maand september? Dan wil NS die vreemde afwijking kunnen verklaren. En dat kon met het eerdere systeem niet. Daarom nam het bedrijf in 2013 een gedurfde stap.

### **Borealis**

“Het oude voorspelsysteem werkte over de hele linie best goed. Maar een organisatie als NS kan het zich niet veroorloven om afwijkingen niet te kunnen verklaren en wil snappen hoe voorspellingen tot stand komen”, zegt consultant Marnix Zoutenbier van CQM. Daarom besloot NS het systeem

te vereenvoudigen. De naam van het project: Borealis. De spannende uitdaging: al na enkele weken de eerste voorspellingen volgens een nieuwe transparante methodiek opleveren.

### **Big data**

“Het was inderdaad erg spannend”, aldus Zoutenbier. “In de eerste plaats omdat we te maken hebben met big data. Voor elke individuele trein op elk specifiek traject in Nederland en op elk moment in het jaar moet een voorspelling van het aantal reizigers worden gedaan.

[lees verder op pagina 2 >](#)



“Het systeem is veel inzichtelijker geworden en consistent in de tijd.”

Eén zogenaamde ‘levering’ bestaat uit maar liefst 125.000 voorspelde combinaties van trein-traject-dag. De input is nog veel groter: bijna 200 miljoen cijfers uit uiteenlopende databronnen. Een steeds belangrijkere bron hierin is de in- en uitcheckgegevens vanuit de OV-chipkaart. Een andere is bijvoorbeeld daadwerkelijke tellingen van het aantal reizigers, die jaarlijks in alle treinen worden gedaan. Ook gebruiken we gegevens uit de dienstregeling. We hebben toen in korte tijd heel intensief samengewerkt met experts van NS om kwaliteit te kunnen leveren. Iedereen maakte daar ook graag tijd voor vrij, omdat we met zijn allen wisten dat we op de goede weg waren. De eerste levering was nog niet optimaal, maar wel voldoende. Het belangrijkste was dat iedereen vertrouwen had dat Borealis de juiste aanpak zou zijn. Juist omdat we nu wél precies weten welke data en modellen we gebruiken en we de oorzaak van afwijkingen dus kunnen begrijpen.”

#### Data science

“Dit is typisch een data science project met big data aspecten”, aldus Zoutenbier. “Elke twee maanden doen we een levering van voorspellingen voor over een half jaar. Dit gebeurt in diverse stappen. Eerst verzamelen wij de data van NS. Die verwerken wij in ons model tot de 125.000 voorspellingen. Samen met de productmanagers van NS vergelijken we in een review de voorspellingen met hun eigen ervaringsgegevens en de meldingen




van reizigers of uit de organisatie. Alle relevante afwijkingen worden geanalyseerd en de voorspellingen worden zo nodig verbeterd. Pas na deze review volgt de feitelijke levering waarmee NS gaat plannen. Daarna volgt nog een analyse van de kwaliteit onder meer op basis van werkelijke tellingen. De verbeterpunten die daaruit komen worden weer naar de volgende ‘levering’ meegenomen.”

#### Sprong voorwaarts

Borealis-projectleider Raoul Klein Kranenburg van NS vindt de nieuwe voorspelmethode een belangrijke verbetering voor de reiziger en voor NS: “Ten eerste omdat de gegevens vanuit de OV-chipkaart nu, uiteraard anoniem, optimaal door het systeem worden benut. Sinds in- en uitchecken met de OV-chipkaart de standaard is geworden, is dit de belangrijkste databron. Het is dus goed dat we daar afgelopen jaren

het model met Borealis op hebben voorbereid om zo de voorspellingen steeds beter te maken. In het model gebruiken we nu ook de meldingen van reizigers gegeven via de NS-app en bij Klantenservice. Dit helpt om te drukke treinen in beeld te krijgen. Met Borealis hebben we daarnaast ook de lege treinen in beeld. Maar het belangrijkste is dat het hele systeem nu veel inzichtelijker is geworden en bovendien consistent in de tijd.” Het is de bedoeling dat NS Borealis nog dit jaar volledig in eigen huis gaat uitvoeren, onder begeleiding van CQM. “Het blijft een continu proces om Borealis op nieuwe ontwikkelingen aan te passen. Daarvoor kunnen we altijd terugvallen op CQM. Maar het nieuwe model kan met veranderingen omgaan én we hebben dit cruciale proces voor NS in eigen hand om de voorspellingen continu te blijven verbeteren”, aldus Klein Kranenburg.

 Marnix Zoutenbier T (040) 750 2325

## Evenement

**4-5 juni 2015 – Nacht van Eindhoven.** De nachtelijke strijd tussen universiteiten op het gebied van toegepaste statistiek en operations research werd dit jaar voor de negende keer georganiseerd. Acht deelnemende teams van drie master studenten streden om eeuwige roem én een geldprijs. Naast opdrachten als het slim inkopen van bier en de bevoorrading van een snoepfabriek, kregen de studenten dit jaar de kans om te ervaren hoe het is om een adviesbureau te runnen. Zelfs keuzes als locatie, werknemers en uit te voeren projecten blijken dus statistisch onderbouwd (en geoptimaliseerd) te kunnen worden...

De winnaars dit jaar: Emma Wagner, Pim Otte en Jasper Velthoen van Studievereniging Christiaan Huygens (TU Delft).



# Data science, hoe werkt het in de praktijk?

**Data science is de term voor het winnen en analyseren van kennis uit data door middel van technieken en theorieën uit vakgebieden zoals wiskunde, statistiek en informatietechnologie. De input voor data science kan big data zijn. Dit is de term voor datasets die zo groot of complex zijn dat ze niet door reguliere database management systemen kunnen worden verwerkt.**

CQM heeft inmiddels 35 jaar projectervaring met data science. Dit is zelfs 20 jaar langer dan de term bestaat. In dit artikel komen vier gerealiseerde projecten aan bod, met daarbij verschillende aspecten van **data science** en **BIG DATA** waarmee CQM te maken kreeg. Hoe gingen we ermee om? Of met andere woorden: data science, hoe werkt het in de praktijk?



In een nieuw te bouwen universiteits-ziekenhuis in Canada zullen alle goederen zoals voedsel, medicijnen, linnengoed en afval door middel van AGV's (automated guided vehicles) op hun bestemming gebracht worden. Frog AGV Systems levert de systemen en de software om het aan te sturen. Het gaat hierbij om veel bewegingen en beslismomenten voor de routing van de AGV's (**VOLUME**). Extra uitdagend is dat de AGV's ook goederen tussen verschillende verdiepingen in verschillende ziekenhuistorens moeten vervoeren. CQM werd gevraagd om een algoritme te ontwikkelen voor de meest efficiënte liftaansturing in de drie ziekenhuistorens (**Prescriptive**). Doorstroming is belangrijk om slim met de beschikbare liftcapaciteit om te gaan.

Daarnaast moeten AGV's met hoge prioriteit (bijvoorbeeld warm eten) voorrang krijgen. Door middel van een simulatie gaf CQM inzicht in de duur van de diverse processtappen en werd zichtbaar welke tijden door een slim algoritme kunnen worden beïnvloed (**Descriptive**). Dit algoritme moet bovendien uiterlijk binnen 0,5 seconde een optimaal antwoord geven (**VELOCITY**).

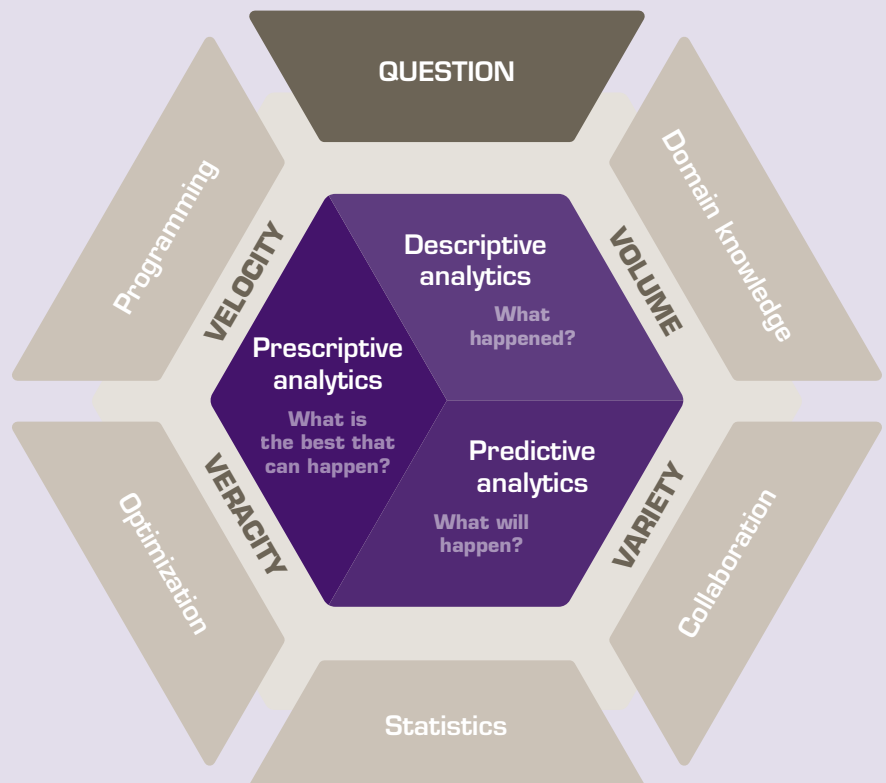


AgroEnergy is dé energiespecialist voor de agrarische sector in Nederland. Het bedrijf helpt tuinders om optimale resultaten in hun energiehandel te bereiken. Vorig jaar introduceerde

[lees verder op pagina 4 >](#)

## CQM, al 35 jaar data science specialist

Samen met u waarde creëren voor uw organisatie. Dat is wat wij al meer dan 35 jaar in meer dan 3500 projecten en voor meer dan 350 klanten doen. Lang voordat de termen data science en big data bestonden werkte CQM al met grote en gevarieerde hoeveelheden data. Hoewel vooral de term big data soms als 'antwoord der antwoorden' wordt gezien, weten wij door onze jarenlange ervaring dat het verzamelen van data nooit een doel op zich is. CQM begint altijd bij de vraag en niet bij het antwoord. Daarom vormen data nooit het startpunt, maar kunnen ze wel de weg naar de oplossing zijn. Elk vraagstuk vereist een aanpak op maat, waarbij complexe algoritmieken alleen wordt gebruikt als het echt nodig is. Maar als het nodig is, hebben wij de kennis en ervaring van statistiek, optimalisatie en software engineering om op de juiste manier met small, medium en big data als onderdeel van data science om te gaan.





AgroEnergy BiedOptimaal, een innovatieve plug-in die het biedproces van tuinders nog makkelijker maakt: tegen welke prijs koop ik welke hoeveelheid gas en elektra in? BiedOptimaal bepaalt APX-biedingen waarbij de warmtekostprijs voor de tuinder wordt geminimaliseerd. CQM hielp AgroEnergy bij de ontwikkeling van BiedOptimaal. Het project werd volgens een data science-aanpak uitgevoerd. BiedOptimaal draait dagelijks, vlak voordat de APX-biedingen moeten worden ingediend. Er worden realtime veel verschillende databronnen gebruikt (**VARIETY**). Steeds worden de meest recente gegevens zoals weersvoorspellingen en net gesynchroniseerde buffervulling gebruikt om snel tot een biedadvies te komen. De voorspellingen voor warmtebehoefte en energieprijzen (**Predictive**) komen samen in een optimalisatiemodel, dat de optimale bieding voor de tuinder bepaalt (**Prescriptive**). BiedOptimaal is sinds oktober 2014 beschikbaar voor tuinders en wordt in de loop van 2015 ook geschikt voor belichtende tuinders. AgroEnergy wil een soortgelijke dienst ook voor andere sectoren aanbieden, zoals de gebouwde omgeving.



In 2008 schakelde Corus (tegenwoordig Tata) CQM al in voor een data science opdracht. In dit geval om klantklachten over de staalkwaliteit te helpen reduceren. Een grote diversiteit aan databronnen (**VARIETY**) binnen Corus vormden de input. Door een sterke nadruk te leggen op het doel ontstond de mogelijkheid om relevante data heel duidelijk te onderscheiden. Om de potentie ervan te kunnen benutten moesten grote hoeveelheden data (**VOLUME**) aan elkaar worden gekoppeld. Die data bestonden echter in verschillende formaten en waren dus niet eenduidig te verbinden. Daarnaast bevatten de data moeilijk te detecteren vervuilingen (**VERACITY**). Ook waren er praktische uitdagingen bij het volgen van staalrollen door het productieproces. Staalrollen worden meermaals op- en afgerold waardoor

de materiaalvolgorde elke keer wordt gespiegeld. Daarnaast worden staalrollen steeds dunner gewalst, waardoor de lengte toeneemt en de plaats van oppervlakte-defecten verschuift en wordt uitgesmeerd. Corus gebruikte de data voor snelle kwaliteitsmonitoring (**Descriptive**) en startte verbeterprojecten voor de meest voorkomende klantklachten (**Predictive**).

## ProRail

Samen met ProRail heeft CQM de Infra-Monitor applicatie ontwikkeld. Binnen deze applicatie kan de infrastructuur van het spoor en de dienstregeling in combinatie met andere databronnen (**VARIETY**) worden gevisualiseerd. Met deze detailinformatie kunnen gebruikers door middel van interactieve data analyses verschillende vraagstukken beantwoorden (**Descriptive**). Een voorbeeld hiervan is de RisicoRegister Rijnwegen analyse. Met deze analyse is het mogelijk om systematisch die punten in de infra te identificeren waar het risico op een botsing tussen twee treinen hoger is dan gemiddeld. De analyse kan ook het effect van maatregelen als flankdekkingswissels, ATBVV en doorschietlengte inzichtelijk maken (**Predictive**). Door de resultaten daarna te visualiseren in een schematische layout van het spoor, krijgen de experts direct inzicht in de maatregelen en bijbehorende consequenties. ProRail gebruikte de RisicoRegister Rijnwegen analyse al voor verschillende veiligheidsstudies.

Roel Wijgers T (040) 750 2313

## Terugblik

### Workshop big data in het OV

Met de huidige beschikbaarheid van veel verschillende databronnen ontstaan nieuwe kansen. Voor het herkennen en benutten van die kansen is creativiteit nodig. Op uitnodiging van OV Next Step – een groep enthousiastelingen van 35-45 jaar, werkzaam in de gehele breedte van het OV – begeleidde CQM op 18 mei 2015 een creatieve brainstorm rond het thema big data in het Openbaar Vervoer. De vraag: hoe kan het OV verbeterd worden met de inzet van databronnen? Zo'n 30 mensen hebben actief en gestimuleerd door creatieve denktechnieken maar liefst tweehonderd verrassende ideeën bedacht. Van het differentiëren van prijzen van OV naar vraag, aanbod en kwaliteitsniveau tot het gebruiken van de door klanten aangevraagde reisadviezen om aanbod van OV-reizen aan te scherpen.



## Alles zelf doen is optellen. Samenwerken is vermenigvuldigen.



CQM gelooft in de wisselwerking tussen verschillende specialismen om samen optimale oplossingen te kunnen bieden. Op het gebied van data science zijn wij druk bezig een ecosysteem op te bouwen. Zo werken we nauw samen met kennisinstellingen als DSC/e (Data Science Center Eindhoven) en technologiepartners als KPMG, IBM en SAS.



Quant wordt gratis toegestuurd aan klanten en andere relaties van CQM. Aanmelden voor een abonnement kan via [info@cqm.nl](mailto:info@cqm.nl). Het volgende nummer zal in oktober 2015 verschijnen. Meer informatie: +31 40 750 2323, [www.cqm.nl](http://www.cqm.nl)

Openbaar maken en/of vervoelvuldigen van teksten/beelden is alleen toegestaan na uitdrukkelijke toestemming van CQM.