

Managementsamenvatting

Verkeersveiligheidsmodel Provincie Utrecht

Arjan Knol, Martijn van der Boor, Martijn Koole (welance.io)
24-01-2020

Introductie

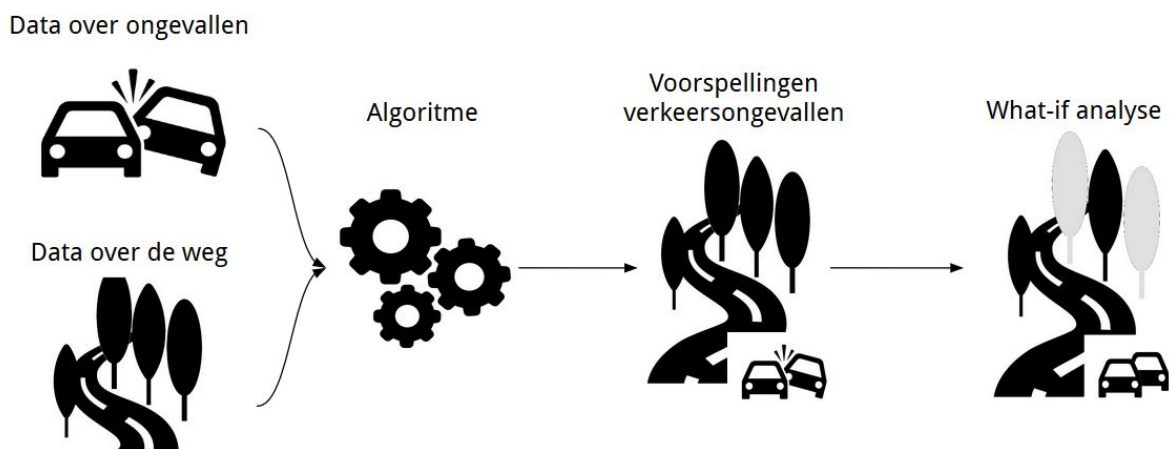
De stad, mens, mobiliteit en techniek veranderen. Een reactieve aanpak voor verkeersveiligheid is niet altijd meer genoeg. Dit is terug te zien in de groei van het aantal ongevallen. Er is behoefte aan proactief inzicht in waar zich mogelijk verkeersonveilige situaties voordoen om verder te kijken dan de traditionele blackspotlocaties. Hedendaagse machine learning technieken en de beschikbaarheid van grote hoeveelheden data maken voorspel-instrumenten mogelijk om bij te dragen aan een risicogestuurde en proactieve aanpak voor verkeersongevallen.

[Welance.io](https://welance.io) heeft in samenwerking met de provincie Utrecht een **proof of concept verkeersveiligheidsmodel** ontwikkeld. Het voorspellend verkeersveiligheidsmodel geeft op basis van een brede dataset voor elk wegvak/hectopunt en kruispunt een risicoscore op verkeersongevallen van zeer laag (1, groen) tot en met zeer hoog (5, zwart). Ook geeft het model inzicht in welke factoren op een specifiek hectopunt of kruispunt van belang zijn ter advisering aan verkeerskundigen en voor what-if analyse.

Deze managementsamenvatting geeft een beknopt overzicht van de **aanpak** van dit project, de **data** die gekoppeld is, het **voorspelmodel**, de **waarde van factoren** en de **dashboard applicatie**. Tot slot volgt een overzicht van gebruikte **hardware en software**.

Aanpak

Voor het verkeersveiligheidsmodel zijn data over historische ongevallen gekoppeld aan gegevens over de weg, over het verkeer en over de omgeving binnen de invloedsfeer van de (ongevals)locatie. Met behulp van een zelflerend algoritme zijn modellen ontwikkeld die per wegvak/hectopunt of per kruispunt een risicoscore berekenen. Met een zogenaamde what-if analyse kan vervolgens worden bepaald wat de invloed is op de risicoscore als input-parameters worden gewijzigd (bijvoorbeeld verlagen maximumsnelheid). De aanpak is schematisch weergegeven in Figuur 1.



Figuur 1: schematische weergave van verkeersveiligheidsmodel

Data

Er zijn voorspelmodellen gebouwd voor wegvakken/hectopunten en voor kruispunten. Voor beide modellen zijn specifieke data gekoppeld.

Ongevallen

Voor data over verkeersongevallen is BRON de primaire bron. Voor de periode 2014-2018 zijn er 2112 exact gekoppelde ongevallen geregistreerd op wegen in beheer van de provincie Utrecht, waarvan 1226 op een wegvak en 886 op een kruispunt. Verder zijn er twee 'subjectieve' bronnen gebruikt: meldingen van RTV Utrecht en ongevalsclusters bepaald door Arcadis. De laatste twee bronnen zijn geen input voor het model, maar gebruikt voor de evaluatie van de uitkomsten van het model.

Tabel 1: datasets verkeersongevallen

Bron	Status	Gebied	Omschrijving
BRON	Gekoppeld	Landelijk	Bestand geRegistreerde Ongevallen in Nederland
Analyse MOVE RTV Utrecht	Gekoppeld (voor validatie)	Provincie Utrecht	Subjectieve data veiligheid provinciale weg.
Ongevalsclusters Arcadis	Gekoppeld (voor validatie)	Provincie Utrecht	Analyse van onveilige wegvakken provinciale wegen
Spoedeisendhulp cijfers	Nog niet beschikbaar	Provincie Utrecht	Cijfers aanrijden ambulance
Handhavingsdata	Nog niet beschikbaar	Provincie Utrecht	
Wild aanrijdingen	Nog niet beschikbaar	Provincie Utrecht	
RAVU	Nog niet beschikbaar	Provincie Utrecht	Data ziekenhuizen van ongevallen (vragenlijsten)

Verder is nog gesproken over gebruik van data over ongevallen uit registraties van de spoedeisende hulp, RAVU (ambulance), aanrijdingen met wild en data van handhaving van de provincie. Omdat deze datasets in deze fase nog niet beschikbaar waren, zijn deze nog niet meegenomen.

Wegvakken/hectopunten

Tabel 2 toont de bronnen die zijn gebruikt voor het voorspelmodel voor wegvakken/hectopunten.

Tabel 2: datasets wegvakken

Bron	Status	Gebied	Omschrijving
NWB	Gekoppeld	Landelijk	Ligging en definitie van alle wegvakken, hectometers en kruispunten.
WKD	Gekoppeld	Landelijk	Maximum snelheden
Openstreetmaps	Gekoppeld	Landelijk	Openstreetmaps geselecteerde wegeigenschappen via Overpass API
Seasonality	Gekoppeld	Landelijk	Seizoensafhankelijkheden (vakanties, spits...)
KNMI	Gekoppeld	Landelijk	Weer
BGT	Gekoppeld	Landelijk	Basisregistratie Grootchalige Topografie (objecten infrastructuur, wegkenmerken, etc)
CBS	Gekoppeld	Landelijk	Cijfers over o.a. bevolkingssamenstelling op buurt- en gemeenteniveau
Onderhoudsstaat van de weg	Nog niet beschikbaar	Provincie Utrecht	Onderhoudsstaat van de weg.
Verkeersintensiteiten	Gekoppeld	Provincie Utrecht	Verkeersintensiteiten
Gereden snelheid (NDW)	Gekoppeld	Provincie Utrecht	S85 gekoppeld vanuit NDW
Duurzaamveilig dataset	Gedeeltelijk gekoppeld	Provincie Utrecht	Onderliggende data Duurzaamveilig onderzoek
Data verkeersmodel	Gekoppeld	Provincie Utrecht	Parameters uit verkeersmodel
Fietspaden (Fietzersbond)	Gekoppeld	Provincie Utrecht	Data van fietzersbond, ligging en eigenschappen fietspaden

Kruispunten

De datasets die zijn gebruikt voor het voorspelmodel voor kruispunten overlappen deels met wegvakken/hectopunten, behalve dat een aantal van de voorgaande datasets minder relevant zijn voor kruispunten of minder goed te koppelen zijn dan voor wegvakken/hectopunten.

Tabel 3: datasets kruispunten

Bron	Status	Gebied	Omschrijving
NWB	Gekoppeld	Landelijk	Ligging kruispunten, aantal vertakkingen
Openstreetmaps	Gekoppeld	Landelijk	Openstreetmaps geselecteerde eigenschappen via Overpass API
Seasonality	Gekoppeld	Landelijk	Seizoensafhankelijkheden (vakanties, spits...)
KNMI	Gekoppeld	Landelijk	Weer
BGT	Gekoppeld	Landelijk	Basisregistratie Grootchalige Topografie (objecten infrastructuur, wegkenmerken, etc)
CBS	Gekoppeld	Landelijk	Cijfers over o.a. bevolkingssamenstelling op buurt- en gemeenteniveau
Roodlichtnegatie (VRI)	Gekoppeld	Provincie Utrecht	Percentage door rood licht rijden, intensiteiten per richting
Gereden snelheid (FCD)	Gekoppeld	Provincie Utrecht	S85 vanuit NDW
Data verkeersmodel	Gekoppeld	Provincie Utrecht	Type/soort kruispunt

Voorspelmodellen voor wegvakken/hectopunten en voor kruispunten

Voor het maken van de voorspellingen is gebruik gemaakt van machine learning algoritmes. In de dataset met kenmerken van de weg/het kruispunt gekoppeld aan ongevaldata wordt door het algoritme gezocht naar patronen. De data is gesplitst in train- en testsets. De trainsets zijn gebruikt om modellen te trainen. Tijdens deze leer of train periode worden relaties tussen de kenmerken en de ongevaldata gevonden. Om te kunnen beoordelen hoe goed het model presteert, zijn de prestaties van het model vervolgens geverifieerd op de testsets (die het model niet eerder gezien heeft).

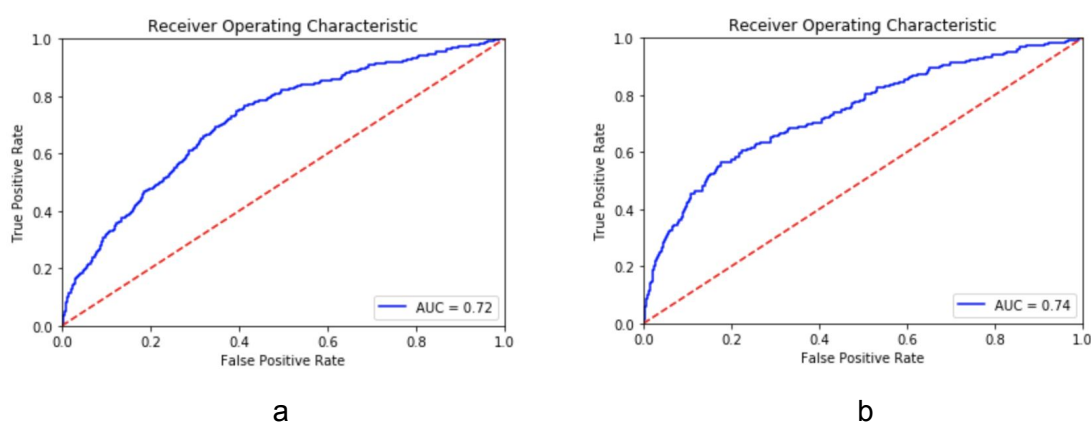
Er is gebruik gemaakt van een XGBoost algoritme¹. Voor het trainen van de modellen is data van 2014 tot en met 2017 gebruikt. Voor het testen van de modellen is data uit 2018 gebruikt.

¹ <https://xgboost.readthedocs.io/en/latest/>

Tabel 4: aantal rijen in dataset (train en test bij elkaar), 0-set zijn wegvakken/kruispunten zonder ongevallen in een bepaald jaar, 1-set zijn wegvakken/kruispunten waarop in een jaar 1 of meer ongevallen zijn gebeurd

Model	Set	Aantal
Wegvak	0	17874
Wegvak	1	1226
Kruispunt	0	6184
Kruispunt	1	886

De performance van het model is bepaald met behulp van de ROC-curve², waarmee kan worden bepaald hoe de prestatie van het model varieert door veranderen van de classificatie-drempelwaarde. De algehele performance van het model kan worden beschreven door de oppervlakte onder deze curve (AUC), waarbij een AUC van 1 overeenkomt met een perfect model, en AUC van 0.5 met compleet random (model presteert even goed als wanneer je random zou gokken). Figuur 2 toont de ROC-curves voor het wegvakken en kruispunten model.



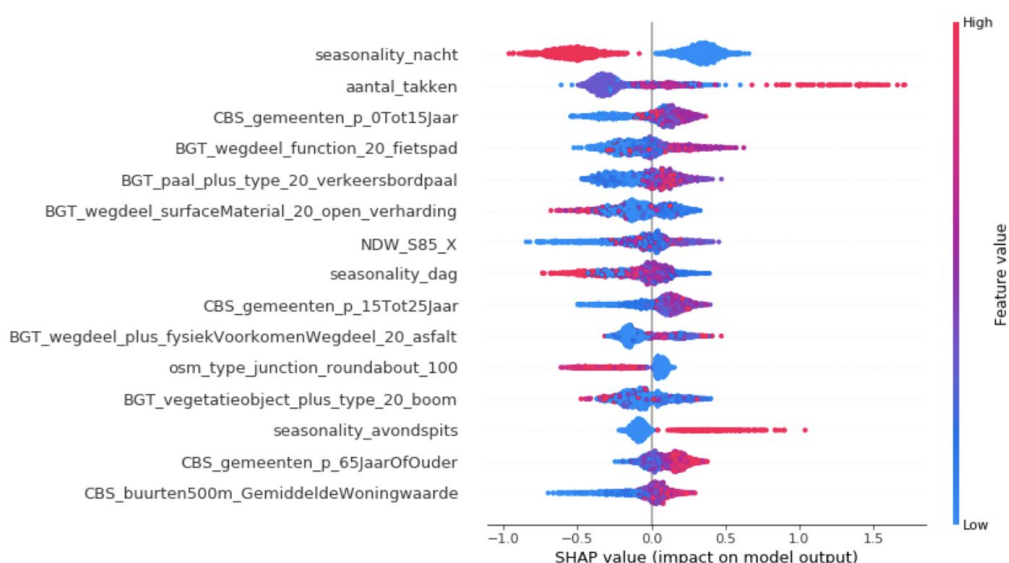
Figuur 2: ROC-curves voor wegvakken model (a) en kruispunten model (b), berekend voor de test-set.

Hieruit kan worden geconcludeerd dat de modellen redelijk presteren (AUC 0.72-0.74), maar dat verbeteringen nog mogelijk zijn. Uit de hellingsgraad van de blauwe lijn vlakbij de oorsprong kan worden bepaald hoe goed het model presteert voor de hoogste risicoscores. Met name bij het kruispunten model loopt de lijn in het begin steil omhoog, wat betekent dat de kruispunten met de hoogste risicoscores ook vrijwel allemaal kruispunten zijn waarop in de testset (2018) ongevallen zijn gebeurd.

² https://en.wikipedia.org/wiki/Receiver_operating_characteristic

Waarde van factoren

Met behulp van SHAP (SHapley Additive exPlanations)³ is de impact van de verschillende input parameters op de output van het model bepaald. Hiermee kan per wegvak/hectopunt en kruispunt worden bepaald welke factoren in die specifieke situatie de grootste invloed hebben (verwerkt in de dashboard applicatie). Figuur 3 toont een visualisatie van alle SHAP values voor het kruispunten model, waaruit o.a. kan worden afgeleid dat kruispunten met een hoog aantal takken (aansluitende wegen) over het algemeen een hoger ongevalrisico kennen, en kruispunten met rotondes in de buurt (osm_type_junction_roundabout_100) over het algemeen een lager ongevalrisico.



Figuur 3: SHAP values voor belangrijkste 15 factoren van het kruispunten model, berekend voor testset (2018)

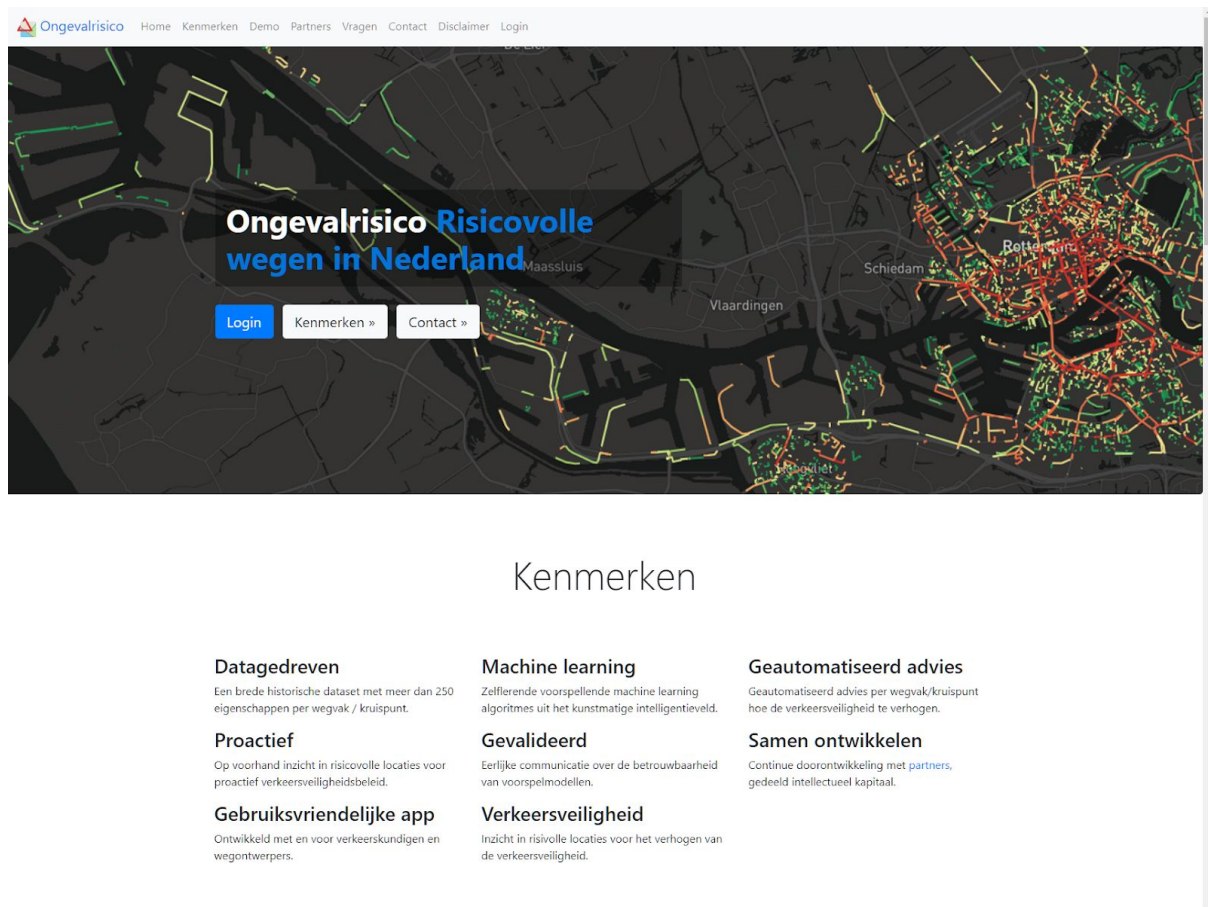
Dashboard applicatie

Gekoppelde data en risicoscores van de voorspelmodellen zijn beschikbaar gesteld via een dashboard applicatie: <https://www.ongevalrisico.nl>. De dashboard applicatie is ontworpen met en voor verkeerskundigen. De dashboard applicatie biedt onder andere de volgende functionaliteiten:

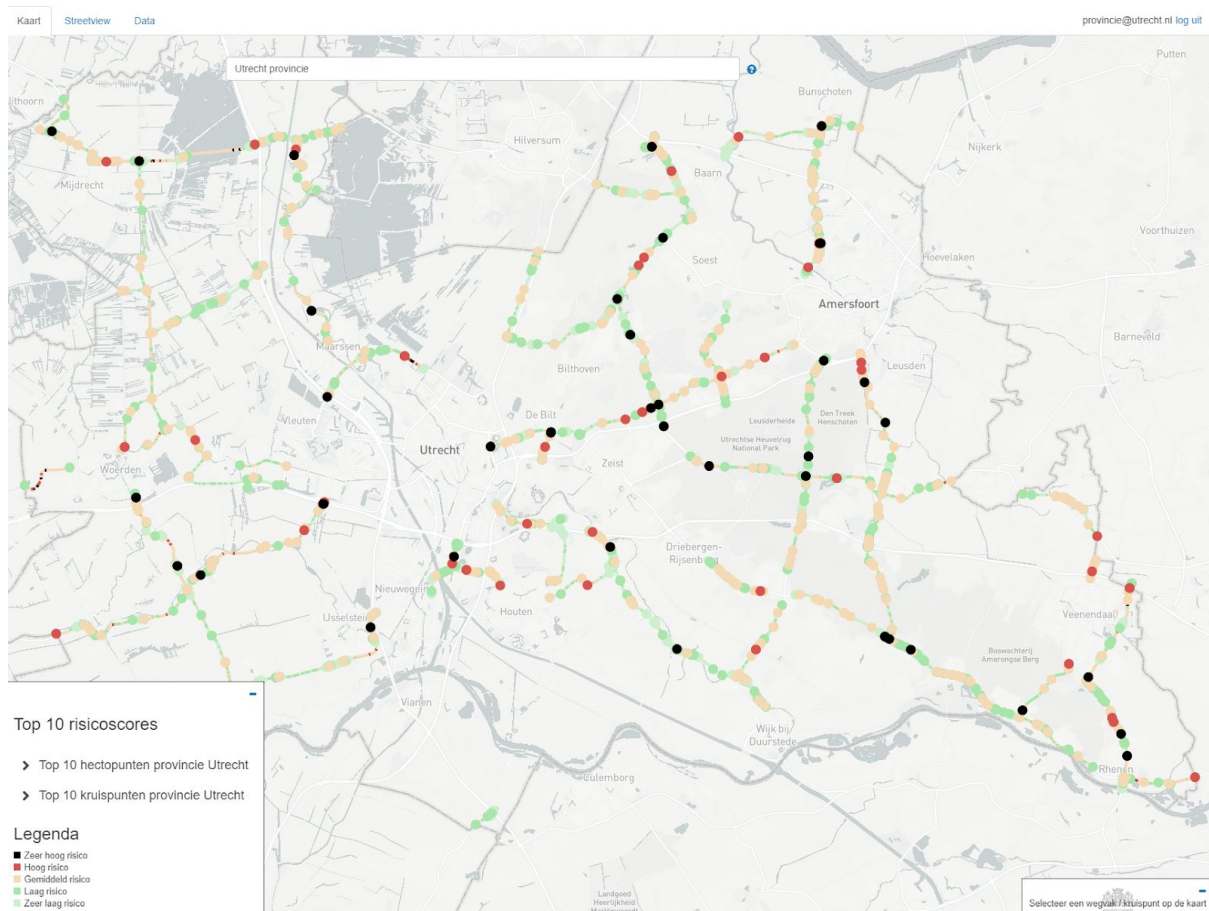
- Landing page met uitleg en achtergrondinformatie over het verkeersveiligheidsmodel
- Authenticatie: het dashboard is te openen na inloggen met een gebruikersnaam en wachtwoord voor de provincie Utrecht
- Kaartweergave met:
 - Wegvakken/hectopunten en kruispunten in de provincie Utrecht
 - Risicoscores per wegvak/hectopunt en kruispunt (van groen naar oranje naar rood naar zwart)

³ <https://github.com/slundberg/shap#citations>

- Een 'intelligent' zoekveld om bijvoorbeeld te zoeken op N-wegen rondom de gemeente Utrecht of een specifieke N-weg
- Een data pane met essentiële informatie over wegvak/hectopunt/kruispunt als deze geselecteerd wordt op de kaart (BPS code hectopunt of id en straatnamen kruispunt, risicoscore, aantal ongevallen afgelopen jaren, maximumsnelheid / werkelijk gereden snelheid V85, verkeersintensiteit, top 10 factoren, etc.)
- Een data pane met top 10 risicolocaties voor de hele provincie Utrecht en voor een geselecteerd gebied op de kaart
- Een legenda van de risicoscores kleurenschaal
- Streetviewweergave met:
 - Streetviewbeelden van een geselecteerd wegvak/hectopunt/kruispunt op de kaart
 - Een tabel met gekoppelde data inclusief een zoekfunctie
- Dataweergave waar CSV exports gedownload kunnen worden van de hele dataset en datasets van gebieden op de kaart.



Figuur 4: landing page ongevalrisico.nl met achtergrondinformatie



Figuur 5: kaartweergave dashboard app app.ongevalrisico.nl

ICT architectuur

Het verkeersveiligheidsmodel en de dashboard applicatie zijn ontwikkeld op servers/laptops/desktops van welance.io tevens gebruikmakend van cloud servers van Google. Voornaamste programmeertalen/gebruikte software is:

- R 3.x met bijbehorende packages en Rstudio Server
- Python 3.x met bijbehorende packages waaronder scikit-learn en Jupyter Notebooks
- Google Cloud Storage
- Google Bigquery
- PostgreSQL en PostGIS
- Authenticatie via Auth0
- Docker containers (schaalbaar voor meerdere gebruikers)