



Verkeers- veiligheidsmodel

Proactief en risico-
gestuurd werken aan
verkeersveiligheid

Augustus 2019



Inhoud

Voorwoord	3
1 Aanleiding	4
2 Projectomschrijving	5
3 Werking en opbouw model	6
3.1 Wegennetwerk en ongevalsdata	6
3.2 Verkeersgegevens	6
3.3 Wegkenmerken	6
3.4 Omgevingsfactoren	7
3.5 Tijdsgebonden data	8
4 Model in de praktijk	9
4.1 Applicatie (dashboard)	9
4.2 Praktische toepassingen	9
4.3 Prestaties en betrouwbaarheid	9
5 Vervolgstappen	11
5.1 Verbetering ongevalsdata	11
5.2 Extra kenmerken koppelen	11
5.3 Samenwerking en kennisdeling	11
Bijlagen	12
Bijlage 1 - Architectuur	12
Bijlage 2 - Analytics Base Table	12
Bijlage 3 - Datapreparatie	13
Bijlage 4 - Keuze en bouwen voorspelmodel	14

Voorwoord

In de mobiliteitswereld zijn we het er over eens dat een reactieve aanpak voor verkeersveiligheid onvoldoende is. Voorkomen is beter dan genezen. Hedendaagse technieken en de beschikbaarheid van grote hoeveelheden (big) data bieden daarvoor ook steeds meer mogelijkheden. Maar hoe kom je tot een praktische werkwijze, waarmee je het aantal verkeersongevallen daadwerkelijk kunt terugdringen? De gemeente Rotterdam heeft de handschoen opgepakt. Als eerste wegbeheerder in Nederland hebben we in Rotterdam een instrument in handen waarmee we de veiligheid van ons wegennetwerk kunnen bepalen en sturen met behulp van op data gebaseerde risico's: het Verkeersveiligheidsmodel.

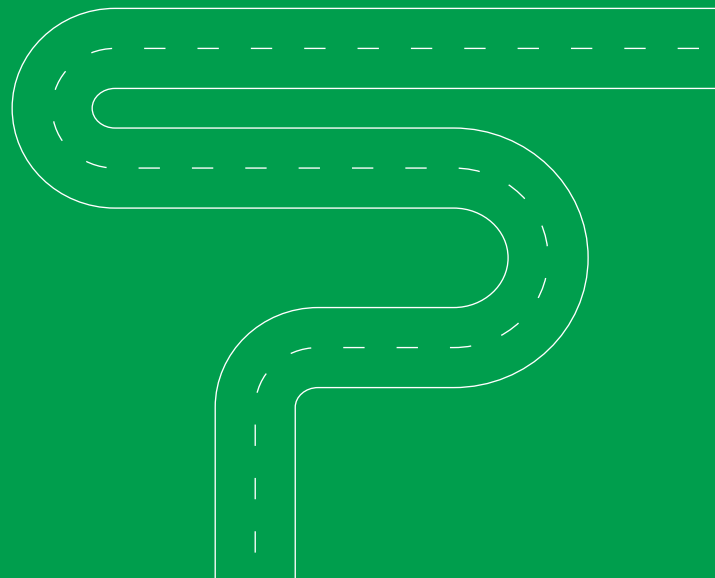
We hebben dit model – een voorspelmodel – de afgelopen anderhalf jaar in eigen beheer en met eigen data en kennis ontwikkeld. Onze specialisten op het gebied van verkeersveiligheid en verkeersdata hebben daarbij samengewerkt met het gemeentelijke Data Science Team. Dank gaat daarbij uit naar het bedrijf Welance.io, dat voor dit project ruime ervaring heeft ingebracht op het gebied van machine learning in relatie tot verkeersveiligheidsprognoses.

De ervaringen en de keuzes die we in dit proces gemaakt hebben, willen we niet voor onszelf houden. We zoeken juist actief de samenwerking met andere gemeenten, wegbeheerders en verkeersprofessionals. Verkeersveiligheid is immers een gemeenschappelijke opgave. We stellen niet alleen kennis over de ontwikkeling en werking van het model beschikbaar, maar vragen andere partijen ook om mee te denken. Hun inzichten bieden immers waardevolle input om het Verkeersveiligheidsmodel verder te verbeteren. Want hoewel een eerste versie van het model inmiddels klaar is voor gebruik, zijn we pas net onderweg.

Gemeente Rotterdam

Afdeling Mobiliteit

Afdeling Onderzoek en Business Intelligence





1 Aanleiding

De komende jaren worden er in Rotterdam duizenden nieuwe woningen bijgebouwd. Ook neemt het toerisme toe, net als het aantal nieuwe voorzieningen in en om de stad.

Deze ontwikkelingen hebben effect op de mobiliteit in de gemeente. Er komen meer mensen op straat, het aantal verkeersdeelnemers groeit en nieuwe verplaatsingsmogelijkheden doen hun intrede, of zijn – zoals de elektrische fiets – inmiddels niet meer uit het straatbeeld weg te denken. Dit alles brengt risico's met zich mee voor de verkeersveiligheid.

Verkeersveiligheid draait niet alleen om het minimaliseren van het aantal verkeersslachtoffers. Verkeersveiligheid speelt ook een rol in de afweging van mensen om wel of niet de straat op te gaan. Zo kan verkeersveiligheid ook bijdragen aan de transitie naar gezonde en efficiënte mobiliteit. Als mensen zich veilig voelen in het verkeer, zullen ze zich vaker verplaatsen om te sporten, te winkelen, zelfstandig naar school te gaan en vrienden te bezoeken. Aandacht voor verkeersveiligheid kan levens redden, leidt tot minder files, verbetert het welzijn, en is daarmee van groot maatschappelijk belang.

De traditionele aanpak bij het verbeteren van de verkeersveiligheid kent een reactief karakter. Na een ongeval vindt een analyse plaats en wordt onderzocht hoe dit ongeval kon gebeuren en welke oplossingen er nodig zijn om herhaling te voorkomen. Centraal in deze gebruikelijke aanpak staan

de zogeheten *blackspots*: locaties waar zich (geregistreerde) letselongevallen hebben voorgedaan. Er is een groeiende behoefte om verder te kijken dan die blackspots en van een reactieve aanpak tot een proactieve en risicogestuurde aanpak te komen.

Bij het verbeteren van de verkeersveiligheid is het belangrijk om onderscheid te maken tussen beïnvloeding van gedrag (van verkeersdeelnemers) en aanpassingen in de infrastructuur. Beide aspecten zijn belangrijk en in de aanpak kunnen ze elkaar versterken. In het Verkeersveiligheidsmodel ligt de nadruk op aanpassingen in de infrastructuur. Een belangrijke reden voor deze keuze is de beschikbaarheid van grote hoeveelheden objectieve data op dat vlak. Naast gegevens over de infrastructuur, gaat het hierbij ook om gegevens over het verkeer en de buitenruimte. Hedendaagse (zelflerende) analysetechnieken maken het mogelijk om die data effectief om te zetten naar inzichten.

Datagedreven werken en innovatie passen goed bij het beleid van de gemeente Rotterdam. Alleen al op het gebied van mobiliteit zijn in en rondom de stad meer dan vijftig vernieuwende initiatieven te vinden. Van zelfrijdende *shuttles* bij Rotterdam The Hague Airport tot elektrische deelvoertuigen in de binnenstad. Als het gaat om verkeersveiligheid wil Rotterdam landelijk koploper zijn op het gebied van proactief en risicogestuurd werken. Een optimaal gebruik van data en algoritmes staat bij de gemeente hoog op de agenda. Het gemeentelijk programma Datagedreven Werken en het Data Science Team van de afdeling Onderzoek en Business Intelligence (OBI) vormen hier illustraties van.

2 Projectomschrijving

In het Verkeersveiligheidsmodel koppelen we data over historische ongevallen aan gegevens over de weg, over het verkeer en over de omgeving binnen de invloedssfeer van de (ongevals)locatie. Door gebruik te maken van een zelflerend algoritme ontstaat een model met een voorspellend karakter.

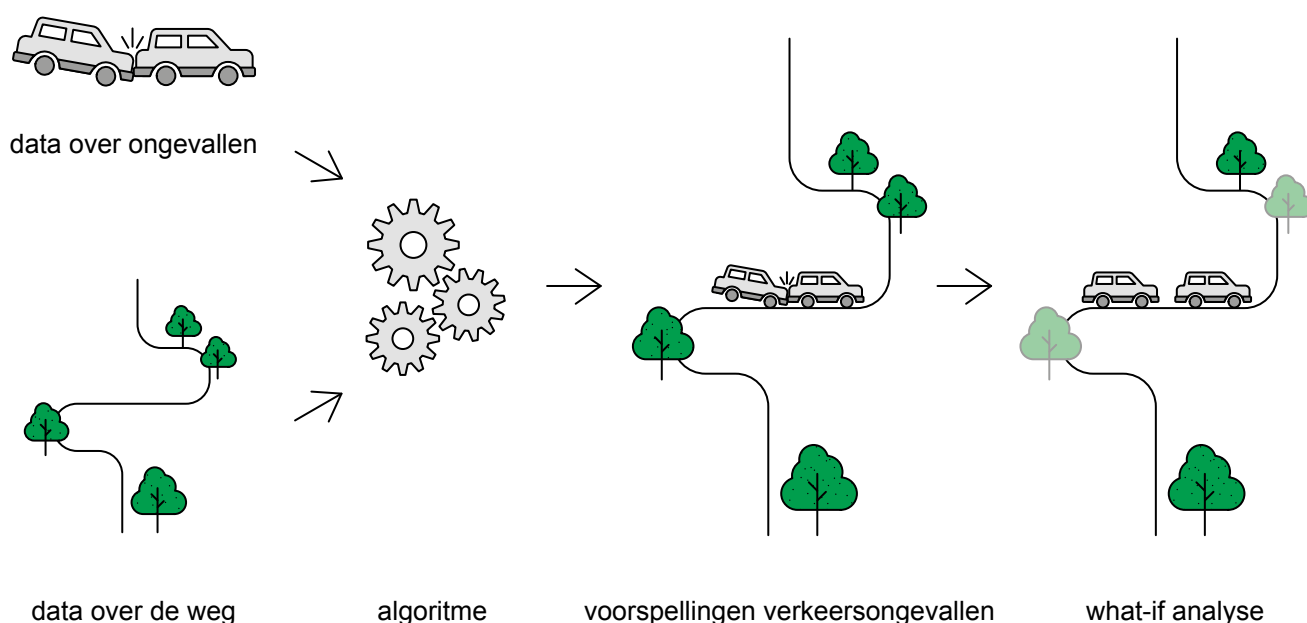
Bij de ontwikkeling van het Verkeersveiligheidsmodel hebben de verkeerskundigen van de gemeente Rotterdam intensief samengewerkt met gemeentelijke data scientists. De beoogde gebruikers (verkeerskundigen) hebben op basis van praktijkervaringen en kennis de mogelijke risico-indicatoren aan de voorkant in beeld gebracht. Ook hebben zij de resultaten van het model beoordeeld en de applicatie getest, waarmee geborgd is dat het Verkeersveiligheidsmodel goed aansluit bij de praktijk.

Via een overzichtelijke gebruikersapplicatie krijgen verkeersprofessionals hiermee de volgende mogelijkheden:

- Proactief signaleren van potentieel onveilige locaties
- Prioriteren van onveilige kruispunten en wegvakken
- Analyseren van de belangrijkste Safety Performance Indicators (SPI's); risico's voor verkeersonveiligheid die voortkomen uit kenmerken van de weg, het verkeer of de omgeving
- Berekenen van het effect van verkeerskundige ingrepen op de verkeersveiligheidsscore ('what-if-analyse')

In dit project zijn achtereenvolgens de volgende stappen doorlopen:

- Bepalen mogelijke risico-indicatoren (op basis van ervaring verkeerskundigen en literatuur)
- Verzamelen databronnen
- Koppeling databronnen onderling
- Keuze model/algoritme
- Optimaliseren algoritme
- Testen voorspelmodel
- Bouwen van applicatie
- Testen van applicatie bij gebruiker



Figuur 1 Schematische weergave Verkeersveiligheidsmodel

3 Werking en opbouw model

In het Verkeersveiligheidsmodel zijn circa 1500 variabelen meegenomen. Het is gebaseerd op machine learning: in de grote hoeveelheden beschikbare data gaat een zelflerend algoritme zelf op zoek naar patronen. De werking van dit algoritme en de (technische) keuzes die hierbij gemaakt zijn staan beschreven in de bijlagen. Dit hoofdstuk beschrijft met welke typen data het Verkeersveiligheidsmodel werkt. De eerste versie van het model is gebaseerd op data uit de jaren 2014-2018. De gegevens worden ieder jaar geüpdatet.

3.1 Wegennetwerk en ongevalsdata

De volgende twee bestanden vormen het uitgangspunt voor het voorspelmodel:

- Nationaal Wegenbestand (NWB), een openbare database met het wegennet van alle wegbeheerders in Nederland.
- Bestand geRegistreerde Ongevallen Nederland (BRON), alle ongevallen die door de politie zijn vastgelegd in processen-verbaal of registratiesets.

Door BRON en NWB aan elkaar te koppelen weten we per wegvak en kruispunt in Rotterdam hoeveel ongevallen er door de jaren heen hebben plaatsgevonden. We maken onderscheid in kruispunten en wegvakken vanwege hun significant verschillende eigenschappen. Op een wegvak speelt de snelheid van het verkeer bijvoorbeeld een andere rol dan op kruispunten.

De geregistreerde ongevallen in de jaren 2014-2018 (circa 5.000 ongevallen op kruispunten en circa 5.000 ongevallen op wegvakken) vormen samen de zogenaamde 1-set. Deze dataset is vergeleken met de set locaties waar geen ongevallen hebben plaatsgevonden, de 0-set. Het voorspelmodel zoekt naar verschillen tussen de 1-set en de 0-set en berekent op basis daarvan risico-scores. De volgende stap was om te onderzoeken wat tussen die locaties de verschillen zijn in de (aanwezigheid en type van) zogeheten risico-indicatoren. Risico-indicatoren zijn variabelen – zoals snelheid en verkeersintensiteit – die van belang zijn voor het doen van voorspellingen over de kans op en de impact van een ongeval.

3.2 Verkeersgegevens

Verkeersintensiteit en verkeerssamenstelling zijn de risico-indicatoren die als eerste in het voorspelmodel zijn opgenomen. Hoe druk het op de weg is en het type verkeersdeelnemers beïnvloeden immers de kans op ongevallen. Om deze variabelen in het voorspelmodel mee te kunnen nemen, is gebruik gemaakt van gegevens (uit basisjaar 2016) van het verkeersmodel van de Metropoolregio Rotterdam Den Haag (V-MRDH 2.0).

Voor motorvoertuigen maakt het verkeersmodel V-MRDH 2.0 onderscheid in verkeersintensiteit per werkdag en per etmaal en hanteert het een onderverdeling naar licht, middelzwaar en zwaar verkeer. In dit verkeersmodel staan ook fietsgegevens, afkomstig uit het fietsnetwerk van de fietsersbond. Daardoor worden ook fietsroutes en -infrastructuur meegenomen. Ten slotte houdt het model V-MRDH 2.0 rekening met de aanwezigheid van OV-verbindingen (frequentie van tram-, metro- en buslijnen).

3.3 Wegkenmerken

Een volgende risico-indicator voor verkeersveiligheid is snelheid. In het Verkeersveiligheidsmodel is de maximaal toegestane snelheid opgenomen. Mede op basis van deze snelheidsregimes worden ook de functies van de wegen meegenomen (bijvoorbeeld ontsluitingsweg, verblijfsgebied, industriestraat, etc). De wens is om ook informatie over daadwerkelijk gereden snelheden (floating car data) te verwerken in het model (zie paragraaf 5.2).

Naast de snelheid is ook de inrichting van de weg van belang voor het voorspellen van de kans op ongevallen. Of er sprake is van een fietsstrook op de rijbaan of een fietspad dat vrij ligt, zegt bijvoorbeeld iets over de mate waarin verkeer veilig kan doorstromen. Parkeervakken en bushaltes zorgen voor snelheidsverschillen op de weg. En bochten in een weg hebben invloed op de snelheid van het verkeer en de complexiteit van de rijtaak.

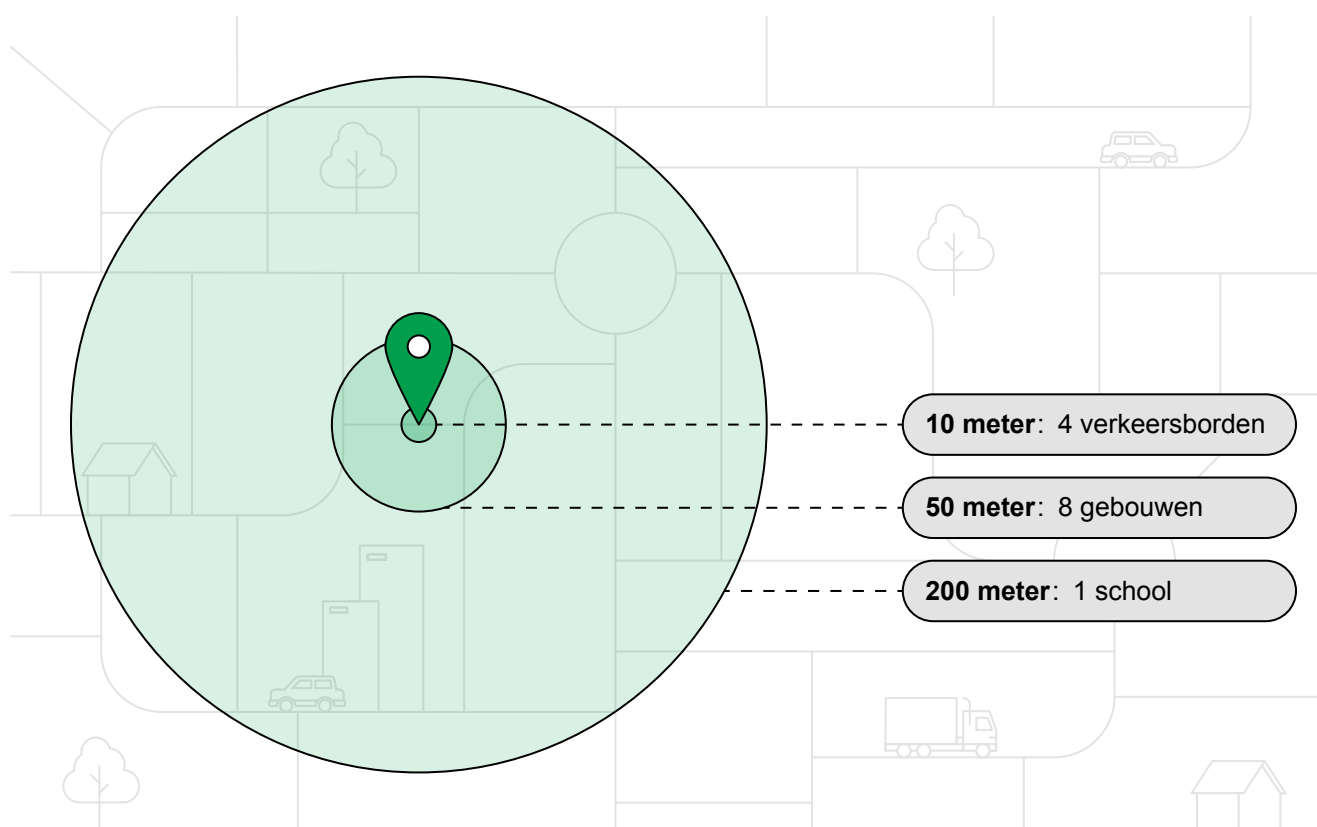
Voorbeelden van meegenomen wegkenmerken:

- Aantal rijstroken (per rijrichting)
- Type wegverharding
- Laad- en loslocaties
- Voetgangersoversteekplaats
- Tram wel of niet op de rijbaan

Ook de gegevens over de inrichting van de weg zijn afkomstig uit het verkeersmodel van de Metropoolregio Rotterdam Den Haag.

Hoe vormt het Verkeersveiligheidsmodel een beeld van de omgeving?

Per locatie (wegvak en kruispunt) trekken we voor elke omgevingsfactor (bijvoorbeeld gebouwen, scholen, verkeersborden) een cirkel en tellen we het aantal objecten binnen die cirkel bij elkaar op. Hoe groot we de cirkel maken, verschilt per omgevingsfactor. Voor verkeersborden is de straal bijvoorbeeld 10 meter, om te voorkomen dat ook verkeersborden van andere straten worden meegenomen. Voor scholen hanteren we een straal van 200 meter; dat is voldoende om een beeld te krijgen van het type buurt.



Figuur 2 Werkwijze omgevingsfactoren

3.4 Omgevingsfactoren

De omgeving van een kruispunt of wegvak kan op verschillende manieren invloed hebben op de verkeersveiligheid. Verkeersdeelnemers kunnen bijvoorbeeld afgeleid worden door de bebouwing rondom een rijbaan. Daarom neemt het voorspelmodel ook de aanwezigheid en kenmerken (dichtheid en hoogte) van bebouwing mee. Ook de gebiedsfunctie kan impact hebben. Een schoolzone in een woonwijk kent bijvoorbeeld een andere verkeersdynamiek dan een vrijliggend wegvak dat een woonwijk ontsluit. En verplaatsingen door een drukke winkelstraat zijn anders dan op een bedrijventerrein. Van scholen en bedrijven neemt het voorspelmodel ook de omvang mee (respectievelijk aantal leerlingen/studenten en aantal medewerkers).

Ook objecten op en langs de rijbaan kunnen risico's met zich meebrengen. De aanwezigheid van bomen bijvoorbeeld kan zorgen voor slechter zicht en dat kan er toe leiden dat bestuurders onverwachts moeten uitwijken.

Voorbeelden van meegenomen omgevingsfactoren:

- Verkeersborden
- Laadpalen
- Straatlantaarns
- Winkelgebieden
- Schoolomgeving

3.5 Tijdsgebonden data

Ook het moment op de dag of in het jaar kan impact hebben op de verkeersveiligheid. In het weekend zijn de risico's anders dan doordeweeks, en in de spits anders dan midden in de nacht. Op sommige wegen wordt er bijvoorbeeld harder gereden wanneer het rustig is op de weg. Het model neemt ook mee of het wel of geen vakantieperiode is.

Ten slotte zijn ook weersomstandigheden op verschillende manieren van invloed op het verkeer en de verkeersveiligheid. Bij regen, hagel of sneeuw kan het zicht bijvoorbeeld beperkt zijn, net als bij een laagstaande zon of 'donker' weer. En bij harde wind kunnen verkeersdeelnemers minder controle over hun voertuig ervaren. Voor tijdsgebonden data is onder meer gebruik gemaakt van informatie van het KNMI.

Voorbeelden van meegenomen tijdsgebonden data:

- Dag/nacht
- Weersgesteldheid
- Weeralarm



4 Model in de praktijk

Om het model – dat op basis van de veelheid aan data voorspellingen doet – in de praktijk toe te kunnen passen, is een applicatie ontwikkeld. Deze applicatie maakt verschillende toepassingen mogelijk. Hierbij is het belangrijk dat de uitkomsten van het Verkeersveiligheidsmodel op een goede manier worden geïnterpreteerd.

4.1 Applicatie (dashboard)

De webapplicatie van het Verkeersveiligheidsmodel visualiseert voorspellingen op een dashboard en kent onder meer de volgende mogelijkheden:

- Gevaarlijke wegvakken en kruispunten zijn in een oogopslag zichtbaar op een kaart met kleurcoderingen.
- Per wegvak en kruispunt is de volgende informatie zichtbaar:
 - risicoscore voor ongevallen (op een schaal van 1 tot 5)
 - aantal ongevallen in het verleden
 - top 10 belangrijkste risico-indicatoren voor ongevallen
- Door middel van een zoekfunctie kan binnen de data snel worden ingezoomd op een bepaalde wijk, buurt of straat.
- Een geïntegreerde *streetview*-weergave geeft verkeerskundigen een directe mogelijkheid de weg ook zelf te beoordelen.
- Voor een geselecteerd wegvak of kruispunt kan via een what-if-analyse het effect van verkeerskundige ingrepen direct worden doorgerekend.

De applicatie is beveiligd en alleen toegankelijk met een inlognaam en wachtwoord. Deze worden ter beschikking gesteld aan verkeerskundigen.

4.2 Praktische toepassingen

Het Verkeersveiligheidsmodel zorgt ervoor dat de gemeente Rotterdam verkeersonveilige locaties eerder en beter kan signaleren. Wanneer een locatie op basis van de historie en data als 'potentieel onveilig' wordt aangemerkt (objectief), nemen we ook mee of mensen die plek als onveilig ervaren (subjectief). Met behulp van het Verkeersveiligheidsmodel kunnen we objectieve en subjectieve informatie beter combineren.

Het Verkeersveiligheidsmodel geeft ook aan welke factoren bijdragen aan een hoge of lage risicoscore. Is dat bijvoorbeeld de intensiteit van het autoverkeer, de aanwezigheid van parkeerplaatsen langs de rijbaan, of een combinatie

van beide? Dit soort inzichten helpen de verkeerskundigen met het vinden van een effectieve aanpak. Wat gebeurt er wanneer we de snelheid ergens verlagen van 50 km/u naar 30 km/u? En wat als we de hoeveelheid rijstroken terugbrengen van twee per richting naar één per richting? De what-if-analyse laat direct het effect van de verschillende ingrepen op de verkeersveiligheidsscore zien.

Het model kan in sommige gevallen ook bijdragen aan het creëren van draagvlak. Bij een herinrichting van een straat of een kruispunt is het bijvoorbeeld mogelijk op voorhand het effect op de verkeersveiligheid inzichtelijk te maken. Daarmee kunnen we aan de hand van een extra analyse uitleggen waarom we bepaalde keuzes maken.

Tot slot helpt het voorspelmodel de gemeente bij het prioriteren van de aanpak van onveilige locaties. Daar waar volgens het model de meeste risico's zijn, geven we voorrang in onze aanpak. Hierbij focussen we in eerste instantie op locaties met risicoscores (4) en (5).

4.3 Prestaties en betrouwbaarheid

Het Verkeersveiligheidsmodel is getest door een analyse uit te voeren op een zogeheten testset (data die niet is meegenomen in de 'training' van het model (zie bijlage voor toelichting)). Uit die test bleek dat het model goed presteert. Voor de wegvakken waarvan het voorspelmodel een hoge risicoscore geeft, gebeuren in 81% van de gevallen ook één of meerdere ongevallen op de testset. Voor de kruispunten lag dit percentage op 96%.

Omdat het om een volledig nieuw model gaat, zet Rotterdam zich de komende tijd in voor verdere (wetenschappelijke) validatie. Hiervoor is onder meer een klankbordgroep opgericht met experts van CROW (kennisplatform voor infrastructuur, openbare ruimte, verkeer en vervoer), Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV) en Rijkswaterstaat.

Dat het model niet feilloos is, is goed te verklaren. Zoals eerder aangegeven, hangt verkeersveiligheid samen met infrastructuur én gedrag. Wanneer een automobilist bijvoorbeeld op een kaarsrechte weg wordt afgeleid door zijn telefoon, wordt dit gedrag niet meegenomen in het model. Dit geeft aan hoe belangrijk het is dat de resultaten op de werkvloer op een goede manier geïnterpreteerd moeten worden. Het model helpt bij het vinden van verbanden tussen oorzaken van verkeersonveiligheid. Het is daarmee een aanvulling op bestaande (gebieds)kennis van verkeerskundigen, informatie over (locatie en toedracht van) ongevallen en klachten over onveilige locaties.

5 Vervolgstappen

De gemeente Rotterdam lanceert in 2019 een nieuw beleidsplan voor verkeersveiligheid. Vanaf dat moment wordt het Verkeersveiligheidsmodel door verkeerskundigen officieel in gebruik genomen. Ook daarna blijven we het model continu verbeteren.

5.1 Verbetering ongevalsdata

De afgelopen jaren is de politie op een andere manier ongevallen gaan registreren. Sommige kenmerken worden steeds beter geregistreerd, bijvoorbeeld het vervoersmiddel waarmee de verkeersgewonde zich ten tijde van het ongeval verplaatste. Andere kenmerken worden juist minder goed geregistreerd. Dat geldt bijvoorbeeld voor de oorzaak van het ongeval en de eigenschappen (vervoersmiddel) van een eventuele veroorzakende partij.

Wanneer er nu een ongeval met een fietser heeft plaatsgevonden, kan het model die informatie gebruiken om de veiligheid voor fietsers op die locatie in te schatten. Het model weet echter niet altijd of die fietser zelf op een paaltje is gebotst of door een auto met hoge snelheid is aangereden. Als het model ook die informatie mee kan nemen, kunnen ook de risico's tussen verkeersdeelnemers onderling worden ingeschat. Daarmee is het mogelijk de resultaten van het voorspelmodel verder te verbeteren.

5.2 Extra kenmerken koppelen

De koppeling van extra datasets zal het Verkeersveiligheidsmodel verder uitbreiden en verbeteren. Zo voegen we de werkelijk gereden gemiddelde snelheid per wegvak (floating car data) aan het model toe; hiermee nemen we 'automatisch' de aanwezigheid mee van bijvoorbeeld verkeersdrempels, die niet als objecten in het voorspelmodel zijn opgenomen. Andere gegevens waar het voorspelmodel nog mee kan worden uitgebreid zijn informatie over wegdelen, zoals wegvakken en zijstraten.

5.3 Samenwerking en kennisdeling

Als initiatiefnemer van het eerste gemeentelijk Verkeersveiligheidsmodel zoeken we actief samenwerking met andere partijen. We kiezen er niet alleen voor om kennis over de ontwikkeling en werking van het model beschikbaar te stellen, maar delen ook onze ervaringen. Bijvoorbeeld met het interpreteren van de resultaten en het creëren van draagvlak, zowel binnen de organisatie als landelijk (politiek). Zo kunnen we andere wegbeheerders op weg helpen met een eigen proactieve en risicogestuurde aanpak voor verkeersveiligheid. Daarnaast biedt het combineren van data van verschillende gemeenten de mogelijkheid om regionale voorspelmodellen te maken, waar verschillende wegbeheerders gezamenlijk baat bij hebben.

De inzichten uit het voorspelmodel zijn niet alleen interessant voor wegbeheerders. Ook andere partijen kunnen er hun voordeel mee doen. De verkeerspolitie kan de data bijvoorbeeld gebruiken om controles uit te voeren op locaties met de grootste risico's op ongevallen. Daarnaast is het mogelijk de inzichten uit het voorspelmodel te gebruiken als input voor (wetenschappelijk) onderzoek naar oorzaken en oplossingen op het gebied van verkeersveiligheid.

Ten slotte maakt samenwerking het mogelijk om bestaande functies in het model en (daarmee) de kwaliteit van de resultaten te verbeteren. Hoe meer gemeenten meedoen, hoe meer informatie we verzamelen over bijvoorbeeld verkeersonveilige locaties op 50 km/u wegen. Meer informatie betekent meer mogelijkheden om verbanden te leggen tussen historische ongevallen en de kans op toekomstige ongevallen. Kortom, hoe meer we samenwerken, hoe groter onze gezamenlijke bijdrage aan de landelijke opgave op het gebied van verkeersveiligheid.

Meer informatie?

Geïnteresseerden in het Verkeersveiligheidsmodel kunnen contact opnemen met de afdeling Mobiliteit van de gemeente Rotterdam via verkeersveiligheid@rotterdam.nl

Bijlagen

Bijlage 1 - Architectuur

In het project hebben we gebruikgemaakt van twee open source servers.

De interne open source server van de gemeente Rotterdam is gebruikt voor het koppelen van organisatiespecifieke data en voor het bouwen van voorspelmodellen.

Hardware:

Intel Quadcore i5
64GB RAM
500GB HDD

Software:

Linux Ubuntu 18.04
R 3.5.1 + packages
Rstudio Server
Java 11.0.3
H2o.ai 3.24.0.2

De externe open source server ongevalrisico.nl is gebruikt voor het koppelen van landelijk beschikbare data en voor de dashboardapplicatie.

Hardware:

Intel Quadcore Xeon
E3-1225
16GB RAM
1TB SSD

Software:

Linux Ubuntu 18.04
R 3.5.1 + Rshiny + packages
Java 11.0.3
Shinyproxy.io
Docker 18.09
PostgreSQL 10.7
Nginx 1.14
H2o.ai 3.24.0.2

Het geheel is geprogrammeerd in de open source programmeertaal R.

Bijlage 2 - Analytics Base Table

Alle gekoppelde en geprepareerde data die gebruikt zijn voor het bouwen van het model staan in de Analytics Base Table. Deze tabel (data frame) bevat van 2014 tot en met 2018 voor elk jaar alle wegvakken en kruispunten in de gemeente Rotterdam volgens het Nationaal Wegen Bestand (op peildatum 1 januari).

Gekoppelde data

Per opgeknipt wegvak/kruispunt bevat de tabel kolommen met bijvoorbeeld:

- Klasse wegvak of kruispunt
- Aantal verkeersongevallen (de target kolom van het voorspelmodel 0 of 1 - elk verkeersongeval staat op een eigen rij met een datum en tijdstip)
- Datum en tijdstip verkeersongeval, afgerond op hele uren
- Algemene eigenschappen van het wegvak/kruispunt, zoals straat, plaats en coördinaten
- Verkeersmodel met o.a. intensiteiten, wegtype en maximumsnelheden
- Kenmerken/objecten binnen 10m en 100m uit Open Streetmap
- Omgevingsobjecten kenmerken/objecten binnen diverse radii van 5m tot 200m (bijvoorbeeld aantal gebouwen, aantal verkeersborden, nabijheid tramlijnen, etc.)
- Tijdsgebonden data zoals weersomstandigheden op tijdstip van het ongeluk/niet-ongeluk, weekday, maand, vakantie en spits
- Wegbeleving meldingen onveilige wegen van burgers (enkel gebruikt voor validatie van het voorspelmodel)



Bijlage 3 - Datapreparatie

Undersampling 0-set

De Analytics Base Table bestaat uit wegvakken en kruispunten waarop verkeersongevallen hebben plaatsgevonden (de 1- set) en wegvakken en kruispunten waarop geen verkeersongevallen hebben plaatsgevonden met een willekeurig tijdstip (de 0-set). Om te voorkomen dat het voorspelmodel te snel 0 voorspelt, is ervoor gekozen de 0-set te undersamplen:

- Elk wegvak/kruispunt uit de 0-set bevat één willekeurig tijdstip in een jaar (één rij) en niet meerdere willekeurige tijdstippen.
- Wegvakken en kruispunten waarop over de jaren heen een of meer ongevallen zijn gebeurd, maar in een specifiek jaar niet, zijn verwijderd uit de 0-set.
- Elke wegvaklijn is opgeknipt in punten om de 10 meter. Zo kan bijvoorbeeld het aantal gebouwen binnen 50 meter worden bepaald. Sommige wegvakken bevatten hierdoor meer dan 200 punten (rijen) in de 0-set. Om te voorkomen dat het voorspelmodel te snel 0 voorspelt, is het aantal punten op wegvakken waarop geen ongevallen zijn gebeurd gedeeld door 10.

De verhouding tussen het aantal ongevallen (1-set) en geen ongevallen (0-set) op wegvakken/kruispunten is:

Wegvlakken:

Aantal ongevallen:	5558
Gemiddelde kans op ongeval in dataset:	0.77%
Gemiddelde kans op ongeval in trainset:	4.24% (undersampling aantal punten)
Gemiddelde kans op ongeval in testset:	1.05%

Kruispunten:

Aantal ongevallen:	5303
Gemiddelde kans op ongeval in dataset:	13.21%
Gemiddelde kans op ongeval in trainset:	12.24%
Gemiddelde kans op ongeval in testset:	15.99%

Train- en testset

Het definitieve model is getraind op de jaren 2014-2017 uit de Analytics Base Table. De betrouwbaarheid van het model is getest met data uit het jaar 2018 (deze data heeft het model tijdens het trainen niet gezien). Om het model robuust te maken, is het ook getraind op 80% willekeurige rijen uit de Analytics Base Table en vervolgens getest op de overige 20% die het model tijdens het trainen niet gezien heeft. De resultaten hiervan kwamen overeen met de betrouwbaarheidsresultaten van het definitieve voorspelmodel.

Voor verdere (wetenschappelijke) validatie van het voorspelmodel is een klankbordgroep opgericht met experts van het CROW, SWOV en Rijkswaterstaat.

Bijlage 4 - Keuze en bouwen voorspelmodel

Getoetste algoritmes

Voor het voorspelmodel hebben we diverse supervised (machine learning) algoritmes uitgeprobeerd:

- Lineaire regressie
- Logistische regressie
- Decision tree
- Random forest
- Neural network
- Gradient boosting machine

Binnen het gradient boosting machine algoritme hebben we diverse verdelingen (distributions) uitgeprobeerd: bernoulli (1 of 0), multinomial, poisson en gaussian.

Voor poisson en gaussian is de Analytics Base Table met wegvakken geaggregeerd naar een regressietabel per wegvak per jaar (zonder tijdsgebonden data). Zo is voor elk wegvak/kruispunt het totale aantal ongelukken per jaar bepaald. Kolommen, zoals die met het aantal gebouwen in een straal van 50 meter, zijn geaggregeerd met het minimum en maximum aantal en met een standaard-afwijking om de spreiding over het wegvak en verschillen in de omgeving per jaar goed weer te geven.

Hyperparameter tuning

Met behulp van een grid search en een validatieset hebben we afzonderlijke instellingen van het voorspelmodel geoptimaliseerd (bijvoorbeeld learning rate en diepte per boom voor gradient boosting machines).

Classificatie

Omdat we gebruik maken van seasonality-variabelen op bepaalde tijdstippen, zoals weersomstandigheden en spits, is elk verkeersongeval één rij in de Analytics Base Table met een bijbehorend (punt op een) wegvak/kruispunt. Uit de diverse getoetste algoritmes is gebleken dat een classificatie gradient boosting machine (bernoulli-verdeling) het best presteert op de testset. De data is daarom niet geaggregeerd voor het uiteindelijke model.

Betrouwbaarheid testset: precisie en bereik

Bij de afstelling van het gradient boosting algoritme hebben we gekeken naar precisie en bereik.

- Precisie: hoeveel 1-en in de testset (locaties waarop ongevallen zijn gebeurd) heeft het model daadwerkelijk goed voorspeld?
- Bereik: hoeveel 1-en in de testset zijn door het model herkend?

We hebben het algoritme zodanig afgesteld dat er balans tussen precisie en bereik is. Het model voorspelt voldoende locaties correct (voldoende bereik) en van de locaties waar het model een hoog risico voorspelt zijn we er vrij zeker van dat hier ook ongevallen gebeuren (voldoende precisie).

De betrouwbaarheidscijfers op de testset zijn:

Wegvakken:

	0	1
0	178196	1482
1	98	418

accuracy: 99.12
precision: 81.01
recall: 22
AUC: 0.89

Kruispunten:

	0	1
0	8675	433
1	47	1227

accuracy: 95.38
precision: 96.31
recall: 73.92
AUC: 0.96

Van voorspellingen naar risicoscores

Het voorspelmodel geeft per (punt op een) wegvak/kruispunt een voorspelling tussen 0 en 1. Voor de wegvakken zijn de voorspellingen van afzonderlijke punten geaggregeerd naar gemiddelde voorspellingen per wegvak. Vervolgens zijn de voorspellingen tussen 0 en 1 omgezet naar een risicoscore schaal 1 t/m 5:

- zeer laag: onderste 25e percentiel (groen)
- laag: tussen 25e en 60e percentiel (groen)
- gemiddeld: tussen 60e en 90e percentiel (oranje)
- hoog: tussen 90e en 99.5e percentiel (rood)
- zeer hoog: bovenste 99.5e percentiel (top 0.5%) (zwart)

Voor de omgevingsfactoren is een bereik aangegeven voor het gehele wegvak, op basis van de posities waar de minste en waar de meeste omgevingsfactoren voorkomen. Wanneer bijvoorbeeld een wegvak op één locatie 5 gebouwen heeft in een straal van 50m, en op een andere locatie 10 gebouwen, dan is het bereik voor het gehele wegvak 5-10.

Inzicht in factoren per rij

Het voorspelmodel maakt gebruik van Shapley Values: een methode binnen het vakgebied Interpretable Machine Learning waarmee per rij inzicht kan worden gegeven in de factoren die meespelen in de voorspelling. Het Shapley Values algoritme maakt voorspellingen met en zonder een factor en kijkt op basis hiervan in hoeverre een factor bijdraagt aan een specifieke voorspelling.





Gemeente Rotterdam