Scriptie

Universiteit van Amsterdam  
  
Bachelor Informatiekunde Scriptie

**Stijn Melchior Willem Hering  
12377422**

**Afstudeerbegeleider: Robin Langerak  
Omvang onderzoek: 18EC**

# **Introductie**

## Inleiding

Volgens cijfers van de World Health Organization (**WHO**) sterven er jaarlijks ongeveer 1,35 miljoen mensen door verkeersongelukken. Nog eens 20 tot 50 miljoen mensen lopen niet-dodelijk letsel op, velen van hen raken gehandicapt als gevolg van dit letsel (WHO, 2020).   
Naast het leed dat mensen ondervinden ten gevolgen van verkeersongelukken veroorzaken verkeersongelukken ook aanzienlijk maatschappelijke kosten. Zowel sociale, medische en immateriële kosten als economische kosten. Het voorkomen van verkeersongelukken met verkeersdoden en verkeersgewonden ten gevolgen zijn hierdoor een volksgezondheidsprioriteit geworden waarvan de sociale en economische implicaties zich tot ver buiten de vervoerssector uitstrekken (Fumagalli et al., 2017).   
Volgens de **WHO** (2020) zijn de totale kosten die gemaakt worden door verkeersongelukken bij het merendeel van de wereld goed voor drie procent van het bruto binnenlands product (bbp). Zo ook in Nederland, volgens de meest recente cijfers van de *Stichting Wetenschappelijk Onderzoek Verkeersveiligheid* (SWOV) uit 2018 worden de totale maatschappelijke kosten van verkeersongevallen geschat op gemiddeld € 17 miljard (€ 15,8 tot € 18,6 miljard); ruim 2% van het bbp. Dit is aanzienlijk hoger dan andere maatschappelijke kosten die door het verkeer in Nederland worden gemaakt. Zo zijn de verkeerskosten met betrekking tot congestie ongeveer 3,3 tot 4,3 miljard en schade aan het milieu goed voor 7 miljard euro (SVOW, 2020a). .

Volgens cijfers van het *Ministerie van Infrastructuur en Waterstaat* (**IenW**)(2019) is meer dan drie kwart van de totale maatschappelijke kosten, ten gevolgen van verkeersongelukken toe te rekenen aan verkeersdoden en verkeersgewonden, het overige deel bestaat uit materiële schade. Gemiddeld bedragen de kosten per verkeersdode circa 2,8 miljoen en per ernstig verkeersgewonde ruim 300.000 . In Nederland bestaan de kosten zowel uit ‘harde’ economische kosten – zoals medische kosten en schade aan voertuigen – als uit immateriële kosten. Bij immateriële kosten gaat het om verlies van levensjaren en kwaliteit van leven. Bijna de helft van de totale kosten zijn immateriële kosten (IenW,2019).

Sinds 1999 was er een dalende trend zichtbaar in het aantal verkeersdoden in Nederland. Van 1999 tot 2010 halveerde het aantal verkeersdoden. De afgelopen tien jaar stagneerde dit cijfer. Al een decennia lang is er geen significante verandering gebracht in het aantal verkeersgewonden en verkeersdoden in Nederland, blijkt uit de cijfers van het **IenW**(2021). Met 19.700 ernstige verkeersgewonden in 2011 en 21.400 in 2019 (Rijkswaterstaat, 2020a) is er in de afgelopen tien jaar zelfs een positieve trend te zien. De doelstelling die waren gezet in Nederland voor het jaar 2020; 500 verkeersdoden en 10.600 ernstige verkeersgewonden (IenW, 2018), zijn zelfs tijdens de Corona pandemie en de restricties die ontstonden door nieuwe regels en noodwetten niet gehaald.   
Ondanks de verplichte verminderde mobiliteit en het noodzakelijke thuiswerken in 2020 en begin 2021, zijn er volgens het**IenW** (2021) 110 méér verkeersdoden geregistreerd dan de doelstelling van maximaal 500. Verder lijkt het onwaarschijnlijk dat de doelstelling omtrent het aantal ernstige verkeersgewonden in 2020 gehaald zal worden als er word gekeken naar het totaal aantal ernstige verkeersgewonden uit 2019. In 2019 ging het om 21.400 ernstige verkeersgewonden, dit is meer dan het dubbelen van wat maximaal word verwacht in 2020. Het verminderen van het aantal verkeersdoden en gewonden is van belang als Nederland zich aan de nationale ambities wilt houden; de halvering van doden en ernstig verkeersgewonden in 2030 (SVOW, 2020b). Verder draagt een vermindering van het aantal verkeersdoden en gewonden bij een verbeterde verkeersveiligheid in Nederland en worden de maatschappelijke lasten gereduceerd.   
Om dit doel te behalen is het essentieel dat er wordt geëxperimenteerd met meerdere innovatieve ideeën om het aantal verkeersincidenten te verminderen. Volgens de SWOV (2020b) zijn er ondanks de maatregelen uit het regeerakkoord van, de investeringsimpuls en de risico gestuurde aanpak vanuit het Strategisch Plan Verkeersveiligheid 2030 meer en effectievere maatregelen nodig (IenW, 2018).

Er is het afgelopen decennia veel onderzoek gedaan naar het gebruik van data wetenschap (*data science*), machinaal leren (*machine learning*) en kunstmatige intelligentie (*AI*) in het verkeer. Dit onderzoek wordt voornamelijk gedaan op het gebied van het voorspellen van de *verkeerssituatie*, *het voorspellen van verkeersstromen* en *het* *voorspellen van de reisbehoefte*, en veel van deze onderzoeken hebben gezorgd voor nieuwe ontwikkelingen en innovaties omtrent het verkeer en de verkeersveiligheid (Yuan & Li, 2021). In Nederland zijn dit soort onderzoeken zeer beperkt, terwijl de voordelen van deze onderzoeken voor de verkeersveiligheid veelvuldig zijn bewezen. De Nederlands staat heeft wel stappen genomen om onderzoek omtrent AI en Machine Learning te stimuleren. Door het creëren van het OPENDATA portaal via *Overheid.nl* kunnen burgers en professionals zoeken in regelgeving, bekendmakingen en andere datacollecties van de overheid (Dataregister van de Nederlandse Overheid | Data overheid, z.d.).

Het doel van dit onderzoek is bij te dragen aan het verbeteren van de verkeersveiligheid in Nederland door technieken en methoden uit het domein van de statistiek, data wetenschap (*data science*) en machinaal leren (*machine learning*) toe te passen op het *Bestand Geregistreerde Ongevallen in Nederland* (**BRON**). Het BRON is beschikbaar is gesteld als *OPENDATA* (Rijkswaterstaat, 2020) door het Ministerie van Economische Zaken en Klimaat (EZK) via de Digitale Delta (Ministerie van Economische Zaken en Klimaat, 2019) en wordt vanaf 2010 bijgehouden door de Nederlandse politie en/of weginspecteurs van Rijkswaterstaat. Het BRON bevat veel informatie over verkeersongevallen zoals; tijd en dag van incident, locatie, staat van het wegdek (nat, droog, etc), informatie over het weer, informatie over bestuurders/inzittende et cetera.   
In dit onderzoek word met behulp van het BRON antwoord worden geven op de volgende onderzoekvraag:

*Is het mogelijk om de verkeersveiligheid in Nederland te verbeteren door het voorspellen van verkeersongelukken aan de hand van historische data?*

*Benoem ik hier ook de sub-vragen?*

Door middel van kwantitatief onderzoek, via een data-analyse en het maken van een voorspellingsmodel, word kennis en inzicht verkregen over factoren die invloed hebben op verkeersincidenten.Data omtrent verkeersongelukken is verzameld via *Overheid.nl*. De data analyse is uitgevoerd via *JupyterLab* een web gebaseerde interactieve ontwikkelomgeving voor *Jupyter notebooks*, code, en data. HIER KUNNEN LATER NOG DE MODELLEN EN METHODEN WORDEN INGEZET In deze environment is een vervolgens ook een voorspellend model gebouwd. Uit de data-analyse is duidelijk geworden welke factoren en omstandigheden het meeste invloed hebben op de kansen op verkeersongelukken. Deze factoren worden gebruikt bij het maken van voorspellingen door het voorspellend model. Door het maken van voorspellingen is duidelijk geworden waar, wanneer en met welke factoren de kans op verkeersongelukken het grootst is. Hierdoor wordt er bijgedragen aan een verminderen van het aantal verkeersdoden-en gewonden en een vermindering in maatschappelijke kosten.

In het volgende segment word er toegelicht welke data wetenschap en machinaal leren toepassingen er al zijn onderzocht en ontwikkeld met betrekking tot het voorspellen van om verkeer en transport onderwerpen . In het segment dat volgt wordt de relevantie en het nut van het voorspellen van verkeersongelukken toegelicht. Vervolgens word de voorgestelde methodologie toegelicht en word de data kort beschreven. Vervolgens worden de resultaten van de data-analyse en het voorspellend model geëvalueerd en beoordeeld. Het laatste segment behandeld de conclusie van dit onderzoek.

# **Theoretisch kader**

Het gebruik van data-wetenschap, machinaal leren en kunstmatige intelligentie bij het analyseren en voorspellen van verkeersgerelateerde kwesties is de afgelopen jaren enorm toegenomen. Deze ontwikkeling zijn versneld door het gebruik van big-data en het Internet of Things (IoT) in het verkeer. Het verzamelen van verkeersdata gebeurt tegenwoordig voornamelijk automatisch door middel van *slimme apparaten* die in een netwerk in verbinding staan met elkaar, ook wel het Internet of Things (IoT) genoemd. Deze netwerken van *slimme apparaten* worden beheerd door (branche)organisatie of overheidsinstantie die een belang hebben bij het monitoren en managen van de wegen en het verkeer (Barceló, Kuwahara & Miska, 2010). Verkeersdata wordt gebruikt voor verschillende doeleinden gebruikt. Zo wordt verkeersdata gebruikt voor het vermindering van de congestie, detectie van verkeersopstoppingen, optimalisatie van de bestaande infrastructuur door een beter gebruik van het huidige wegennet, verbeterde informatiediensten(e.g. verkeersinformatie), dynamische routegeleiding en het voorspellen van verschillende verkeersonderwerp (Leduc, 2008).   
In dit onderzoek zullen we ons richten op de factoren analyse en voorspellings-toepassingen op verkeersdata. Het maken van verkeersvoorspellingen vormen een grote uitdaging in de wetenschappelijke gemeenschap. Hierdoor lopen de methoden en toepassingen die zijn onderzocht enorm uit een. De reden hiervoor is te herleiden naar de vorm van de verkeersdata. Verkeersdata komt voornamelijk voor in de ruimtelijk en temporeel (spatio-temporal) vorm. Hierdoor veranderd verkeersdata voortdurend met de tijd en de ruimte en vertoont het complexe en dynamische ruimtelijk-temporele afhankelijkheden. Verder wordt ruimtelijke en temporaire verkeersdata ook beïnvloed door een heleboel externe factoren, zoals weersomstandigheden, gebeurtenissen of wegkenmerken. Hierdoor vertonen verkeerstekens een sterke dynamische correlatie in zowel ruimtelijke als temporele dimensies (Yin et all., 2020).

In dit deel zullen een aantal van meest voorkomende toepassingen worden besproken en worden afgewogen en zullen de gekozen modellen in dit onderzoek verder toegelicht worden.

### Deep Learning

# **Referentie**

Barceló, J., Kuwahara, M., & Miska, M. (2010). Traffic data collection and its standardization. *In Traffic Data Collection and its Standardization (pp. 1-10)*. Springer, New York, NY.

*Dataregister van de Nederlandse Overheid | Data overheid.* (z.d.). Overheid.nl. Geraadpleegd op 1 juni 2021, van https://data.overheid.nl/

Fumagalli, E., Bose, D., Marquez, P., Rocco, L., Mirelman, A., Suhrcke, M., & Irvin, A. (2017). The high toll of traffic injuries: unacceptable and preventable. World Bank.

Leduc, G. (2008). Road traffic data: Collection methods and applications. *Working Papers on Energy, Transport and Climate Change, 1*(55), 1-55.

Ministerie van Infrastructuur en Waterstaat. (2018, december). Het strategisch Plan Verkeersveiligheid 2030: Veilig van deur tot deur. Rijksoverheid. <https://www.rijksoverheid.nl/documenten/rapporten/2018/12/05/bijlage-1-het-strategisch-plan-verkeersveiligheid-2030-veilig-van-deur-tot-deur>

Ministerie van Infrastructuur en Waterstaat. (2021, 20 april). Actuele verkeersongevallencijfers. Rijkswaterstaat. <https://www.rijkswaterstaat.nl/wegen/wegbeheer/onderzoek/verkeersveiligheid-en-ongevallencijfers/actuele-verkeersongevallencijfers>

Ministerie van Infrastructuur en Waterstaat (2019). Kennisinstituut voor Mobiliteitsbeleid KiM, 2019, 203 p., KiM-19-A12 - ISBN 978-90-8902-216-5

Ministerie van Economische Zaken en Klimaat. (2019, 31 mei)., Dutch Digital Delta., *Informatie- en communicatietechnologie (ICT)*., Rijksoverheid.nl.,  
<https://www.rijksoverheid.nl/onderwerpen/ict/ict-en-economie/dutch-digital-delta>

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV). (2020a). Kosten van verkeersongevallen. SWOV-factsheet, maart 2020, SWOV Den Haag.

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid (SWOV). (2020). De Staat van de Verkeersveiligheid 2020 Doelstellingen 2020 worden niet gehaald. Geraadpleegd van: <https://www.swov.nl/publicatie/de-staat-van-de-verkeersveiligheid-2020>.

World Health Organization. (2020, 7 februari). Road traffic injuries. World Health Organization (WHO). <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>

Wijnen, W., Weijermars, W., Schoeters, A., van den Berghe, W., Bauer, R., Carnis, L., ... & Martensen, H. (2019). An analysis of official road crash cost estimates in European countries. Safety science, 113, 318-327.

Rijkswaterstaat. (2020, 6 november). *Verkeersongevallen - Bestand geRegistreerde Ongevallen Nederland | data overheid* [Dataset]. Overheid.nl. <https://data.overheid.nl/dataset/9841-verkeersongevallen---bestand-geregistreerde-ongevallen-nederland>

Yuan, H., & Li, G. (2021). A Survey of Traffic Prediction: from Spatio-Temporal Data to Intelligent Transportation. Data Science and Engineering, 6(1), 63-85.

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2020). A comprehensive survey on traffic prediction. *arXiv preprint arXiv:2004.08555*.