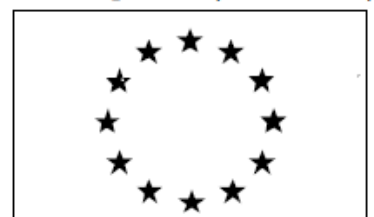


praktikon

Uitbreiding ProVict met ongestructureerde gegevens: Ontwikkeling en validering

Roos Geurts, MSc
Niels Raaijmakers, MSc
Dr. Marc Delsing
Prof. dr. Ron Scholte
Drs. Jacqueline Wientjes

balancing Security and Mobility



EUROPESE UNIE
het Fonds voor interne veiligheid
van de Europese Unie

Uitbreiding ProVict met ongestructureerde gegevens: Ontwikkeling en validering

Praktikon B.V.
Postbus 6909
6503 GK Nijmegen
www.praktikon.nl
info@praktikon.nl

tel. 024-3615480
fax. 024-3611152

© 2021 Praktikon B.V.

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, en evenmin in een retrieval systeem worden opgeslagen zonder de voorafgaande schriftelijke toestemming van Praktikon.

No part of this book/publication may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

Voorwoord

In dit rapport wordt verslag gedaan van de uitbreiding van ProVict met ongestructureerde gegevens. ProVict bepaalt het risico op herhaald slachtofferschap voor personen die als slachtoffer in contact komen met de politie en is uitsluitend gebaseerd op politieregistraties. Het onderzoek is uitgevoerd door Praktikon in opdracht van dhr. Oscar Dros, eenheidschef Oost-Nederland en landelijk portefeuillehouder dienstverlening. Het programma slachtofferrecht valt onder deze portefeuille.

Voor dit onderzoek is een begeleidingscommissie samengesteld ter verhoging van de kwaliteit van het onderzoek en ter bewaking van de onafhankelijkheid en objectiviteit. De begeleidingscommissie stond onder leiding van de directeur Operatiën en bestond uit twee sectorhoofden van politie, een beleidsmedewerker Slachtofferhulp Nederland, een teamchef onderzoek Politieacademie en een hoogleraar forensische orthopedagogiek van de Universiteit van Amsterdam. Tevens is een stuurgroep samengesteld voor de besluitvorming omtrent ProVict. De stuurgroep, onder leiding van Oscar Dros, bestond uit het hoofd Operatiën Oost-Nederland, een hoogleraar van de Radboud Universiteit Nijmegen en een hoogleraar van de Universiteit Tilburg Speciale dank gaat uit naar Jeroen Gijsen en Dick Willems (Team Business Intelligence Amsterdam) voor hun inzet, advies en het generen van de ProVict dataset.

Roos Geurts, MSc & Niels Raaijmakers, MSc
Junior onderzoekers
Praktikon

Dr. Marc Delsing
Senior onderzoeker
Praktikon

Prof. dr. Ron Scholte
Directeur Praktikon
Hoogleraar FSW Radboud Universiteit Nijmegen

Drs. Jacqueline Wientjes
Projectleider ProVict
Programma slachtofferrecht, Politie

November 2021

Managementsamenvatting

In dit rapport wordt verslag gedaan van de uitbreiding van ProVict Fase 1 met ongestructureerde gegevens. ProVict is een verzameling van algoritmen die het risico berekent dat een slachtoffer (van om het even welk delict) binnen twee jaar opnieuw als slachtoffer met de politie in aanraking komt. ProVict is uitsluitend op politiegegevens uit de Basis Voorziening Handhaving (BVH) gebaseerd. ProVict kan worden gebruikt in de dagelijkse politiepraktijk ter voorbereiding op het te voeren slachtoffergesprek in het kader van de verplichte Individuele Beoordeling (IB). De ontwikkeling en validering van ProVict is transparant, reproduceerbaar en overdraagbaar.

ProVict Fase 1 maakt uitsluitend gebruik van gestructureerde politiegegevens om een risico-inschatting van herhaald slachtofferschap te berekenen. Gestructureerde gegevens zijn gegevens met een vaste structuur. Voorbeelden van gestructureerde gegevens in ProVict (ProVict-voorspellers) zijn leeftijd en het aantal registraties van een bepaald delict in een bepaalde rol. In onderhavig rapport wordt gerapporteerd hoe de risico-inschatting van ProVict Fase I kan worden verbeterd door het toevoegen van ongestructureerde gegevens als voorspellers. Ongestructureerde gegevens zijn niet geordend in een vaste structuur omdat het gaat om vrije teksten. Ongestructureerde gegevens uit BVH zijn de gegevens die politiemedewerkers invoeren in formulieren en vrije tekstvelden. Dit kan bijvoorbeeld gaan om verklaringen of additionele toelichtingen.

Er werd onderzoek gedaan naar drie soorten herhaald slachtofferschap: herhaald slachtofferschap van zeer ernstige delicten, ernstige delicten en overige delicten. Voor elk slachtoffer werd per soort herhaald slachtofferschap een risicoscore berekend. Eerst werd onderzocht of ongestructureerde gegevens konden worden gebruikt om een nauwkeurig model voor herhaald slachtofferschap te ontwikkelen. Dit bleek het geval te zijn: ongestructureerde gegevens hadden een voorspellende waarde voor herhaald slachtofferschap. Het risico op herhaald slachtofferschap van zeer ernstige delicten kon het nauwkeurigst worden ingeschat, gevolgd door herhaald slachtofferschap van ernstige delicten. Het risico op herhaald slachtofferschap van overige delicten kon het minst nauwkeurig worden ingeschat.

Na het vaststellen van de voorspellende waarde van ongestructureerde gegevens, werden modellen ontwikkeld die waren gebaseerd op zowel gestructureerde als ongestructureerde gegevens. Ook deze gecombineerde modellen berekenen een score die het risico op herhaald slachtofferschap weergeeft. Van deze modellen is de predictieve

¹ ProVict als samentrekking tussen een afkorting van de Engelse termen 'Protect' en 'Victim'

validiteit vastgesteld. Predictieve validiteit heeft betrekking op de mate waarin de modellen accurate voorspellingen doen wat betreft toekomstig slachtofferschap.

Uit de analyses bleek dat de risico-inschatting uit ProVict Fase 1 (gebaseerd op alleen gestructureerde gegevens) kan worden verbeterd door het toevoegen van ongestructureerde gegevens. De modellen uit ProVict Fase 1 konden (1) herhaald slachtofferschap van zeer ernstige delicten voor alle slachtoffers en (2) herhaald slachtofferschap van ernstige delicten voor 15 tot 23 jarigen het meest nauwkeurig voorspellen. Modellen die gebruik maken van een combinatie van gestructureerde én ongestructureerde gegevens konden (1) herhaald slachtofferschap van zeer ernstige delicten en (2) herhaald slachtofferschap van ernstige delicten voor alle slachtoffers tot 65 jaar het meest nauwkeurig voorspellen. In tegenstelling tot Fase 1 suggereren de resultaten van Fase 2 dat herhaald slachtofferschap van ernstige delicten voor een grotere groep slachtoffers nauwkeurig te voorspellen is.

Voor het praktische gebruik van ProVict is het van belang dat de gebruiker een signaal krijgt wanneer een slachtoffer een verhoogd risico op herhaald slachtofferschap heeft. Hiervoor is het van belang dat er een grens wordt bepaald die aangeeft wanneer dit signaal moet worden getoond. In dit rapport is voor een reeks grenswaarden aangegeven hoe nauwkeurig de voorspellingen zijn. Zo zijn er grenswaarden geïdentificeerd waarbij een meerderheid van de herhaalde slachtoffers terecht wordt ingedeeld in de hoge risicocategorie en tegelijkertijd een meerderheid van de niet-herhaalde slachtoffers terecht in de lage risicocategorie wordt ingedeeld. Door middel van overleg met verschillende partijen en deskundigen dient te worden bepaald welke grenswaarde bij eventuele implementatie zal worden gebruikt.

Dit onderzoek heeft aangetoond dat modellen gebaseerd op een combinatie van gestructureerde en ongestructureerde gegevens een betere risico-inschatting opleveren dan modellen gebaseerd op alleen gestructureerde of alleen ongestructureerde gegevens. ProVict (gebaseerd op zowel gestructureerde als ongestructureerde gegevens) kan op een geautomatiseerde en gevalideerde wijze een risico-inschatting maken van herhaald slachtofferschap van (zeer) ernstige delicten.

Inhoudsopgave

Hoofdstuk 1. Inleiding	1
1.1 Achtergrond	1
1.2 ProVict Fase 1	2
1.2 ProVict Fase 2	2
1.4 Onderzoeksvragen	3
Hoofdstuk 2. Methode	5
2.1 Onderzoeksdesign & data	5
2.2 Variabelen	6
2.3 Data-analyses ongestructureerde data	7
2.4 Combinaties	8
2.5 Risicoscores indelen in risicocategorieën en predictieve validiteit	9
2.6 Robuustheidsanalyses	11
Hoofdstuk 3. Resultaten	12
3.1 Beschrijvende statistieken	12
3.2 Modellen ongestructureerde gegevens	12
3.3 Modellen gestructureerde gegevens.....	13
3.4 Combinatie modellen gestructureerde en ongestructureerde gegevens	13
Hoofdstuk 4. Conclusie	19
4.1 Conclusie.....	19
4.2 Discussie	21
4.3 Aanbevelingen toekomstig onderzoek	23
Literatuurlijst	24
Bijlagen	27
Bijlage 1: Geanalyseerde delicten	27
Bijlage 2: Datapreparatiestrategieën	30
Bijlage 3: Analysemethoden ongestructureerde gegevens	31
Bijlage 4: Combinatiemethoden	33

Hoofdstuk 1. Inleiding

1.1 Achtergrond

Eind 2012 hebben de Europese Raad en het Europees parlement een richtlijn aangenomen die toeziet op het invoeren van minimumnormen voor de rechten, de ondersteuning en de bescherming van slachtoffers van strafbare feiten.² Onderdeel van deze richtlijn is een individuele beoordeling van slachtoffers om beschermingsbehoeften te bepalen (artikel 22 e.v.). Per 16 november 2015 heeft de politie de wettelijke verplichting om voor slachtoffers van strafbare feiten individueel te beoordelen of er sprake is van kwetsbaarheid. Op basis van deze Individuele Beoordeling (IB) dient een keuze te worden gemaakt voor de wijze van bejegening, waarbij al dan niet bepaalde beschermingsmaatregelen kunnen worden geadviseerd en geïnitieerd. Het risico om opnieuw slachtoffer te worden is één van de aspecten van kwetsbaarheid die wordt beoordeeld.

Om politiemedewerkers te ondersteunen bij de inschatting van het risico op herhaald slachtofferschap is ProVict³ ontwikkeld. ProVict is een verzameling van algoritmen die een risicoscore berekent. Deze risicoscore is een inschatting van het risico om opnieuw slachtoffer te worden. Voor deze risicoscores worden grenswaarden vastgesteld die aangeven welke slachtoffers een verhoogd risico hebben op herhaald slachtofferschap. ProVict is transparant, reproduceerbaar, overdraagbaar en op wetenschappelijk verantwoorde wijze ontwikkeld en gevalideerd. In mei 2019 is ProVict Fase 1 afgerond met de oplevering van het rapport 'Ontwikkeling en validering van ProVict' (Geurts et al., 2019). In dat rapport zijn alle details omtrent ProVict Fase 1 terug te vinden.

In onderhavig rapport staat de beschrijving van ProVict Fase 2 centraal. ProVict Fase 1 behandelde de predictieve validiteit van risico-inschattingen op basis van gestructureerde gegevens. ProVict Fase 2 bouwt voort op dat onderzoek en bekijkt in hoeverre de inschatting van herhaald slachtofferschap kan worden verbeterd door het toevoegen van ongestructureerde gegevens als voorspellers. Allereerst wordt de voorspellende waarde van uitsluitend ongestructureerde gegevens onderzocht. Vervolgens wordt de toegevoegde waarde daarvan ten opzichte van het gebruik van alleen gestructureerde gegevens bepaald. De kwaliteit van de risico-inschattingen wordt beoordeeld door de predictieve validiteit vast te stellen. Predictieve validiteit heeft betrekking op de mate waarin de modellen accurate inschattingen maken wat betreft herhaald slachtofferschap. Daarbij wordt onder andere

² Richtlijn 2012/29/EU van het Europees Parlement en de Raad van 25 oktober 2012 tot vaststelling van minimumnormen voor de rechten, de ondersteuning en de bescherming van slachtoffers van strafbare feiten, en ter vervanging van Kaderbesluit 2001/220/JBZ.

³ ProVict als samentrekking tussen afkortingen van de Engelse termen 'Protect' en 'Victim'.

nagegaan in hoeverre slachtoffers die op basis van ProVict als risicovol worden beoordeeld ook daadwerkelijk opnieuw slachtoffer worden.

1.2 ProVict Fase 1

ProVict is ontwikkeld met als doel om op geautomatiseerde wijze een inschatting te maken van het risico op herhaald slachtofferschap voor slachtoffers die met de politie in aanraking komen. ProVict is uitsluitend gebaseerd op politieregistraties. Voor ProVict Fase 1 betrof dit uitsluitend gestructureerde gegevens. Gestructureerde gegevens zijn gegevens met een vaste structuur. Voorbeelden van gestructureerde gegevens in ProVict (ProVict-voorspellers) zijn leeftijd en het aantal registraties van een bepaald delict in een bepaalde rol. De ProVict-voorspellers zijn uitsluitend afkomstig uit de Basis Voorziening Handhaving (BVH).

De resultaten van het valideringsonderzoek van ProVict Fase 1 toonden aan dat met behulp van ProVict een accurate inschatting kan worden gemaakt van het risico dat een slachtoffer binnen twee jaar opnieuw slachtoffer wordt. De meest accurate inschatting kon worden gemaakt voor herhaald slachtofferschap van zeer ernstige delicten en van ernstige delicten (voor 15- tot 23-jarigen). Voor zeer ernstige delicten en voor ernstige delicten (bij 15- tot 23-jarigen) werd de meerderheid van de slachtoffers die opnieuw slachtoffer werden door ProVict correct als risicovol beoordeeld. Toch werd nog steeds een substantieel deel van de herhaalde slachtoffers door het instrument als niet-risicovol ingeschat. Ook werd een deel van de niet-herhaalde slachtoffers onterecht als risicovol aangemerkt. Dit suggereert dat er, ondanks de goede resultaten, ruimte is voor verbetering van de risico-inschatting van herhaald slachtofferschap door ProVict.

1.2 ProVict Fase 2

Naast gestructureerde gegevens bevat BVH ook ongestructureerde gegevens. Ongestructureerde gegevens uit BVH zijn de gegevens die politiemedewerkers invoeren in formulieren en vrije tekstvelden. Dit kan bijvoorbeeld gaan om verklaringen of additionele toelichtingen in registraties. Ongestructureerde gegevens zijn niet geordend in een vaste structuur omdat het gaat om vrije teksten. Het extraheren en verzamelen van zinvolle informatie uit ongestructureerde gegevens is daardoor een arbeidsintensief proces, wanneer dit handmatig wordt uitgevoerd (Grimmer & Stewart, 2013). Recente ontwikkelingen op het gebied van text mining en machine learning hebben het echter mogelijk gemaakt om ongestructureerde gegevens beter te analyseren (Hassani et al., 2016). Ongestructureerde gegevens bevatten mogelijk gedetailleerde informatie over een incident of slachtoffer die niet terug te vinden is in de gestructureerde gegevens. Gedetailleerde informatie in de ongestructureerde gegevens kan potentieel relevant zijn voor de inschatting van het risico op

herhaald slachtofferschap. Het gebruik van ongestructureerde gegevens zou daarmee bij kunnen dragen aan een verbetering van de risico-inschatting van ProVict.

Wetenschappelijke fundering

Net als bij ProVict Fase 1 is het van groot belang dat ProVict Fase 2 op een wetenschappelijk verantwoorde wijze wordt ontwikkeld en getest. Eerder onderzoek naar risico-inschattingen heeft de waarde aangetoond van (a) ongestructureerde politiegegevens, (b) het combineren van verschillende typen gegevens en (c) het gebruik van machine learning technieken (Adamou et al., 2018; Amrit et al., 2017; Berk & Sorenson, 2020; Chen et al., 2004; Haleem et al., 2019; Härkänen et al., 2020; Hassani et al., 2016; Hatef et al., 2020; Hegelich, 2016; Jonnagaddala et al., 2015; Karystianis et al., 2018, 2019; Laar et al., 2020; Poelman et al., 2010).

Meer dan 90% van alle informatie in het politiesysteem is als ongestructureerde gegevens opgeslagen (Poelman et al., 2010). Voorbeelden van ongestructureerde politiegegevens zijn verklaringen van slachtoffers, getuigen of verdachten en additionele informatie die politiemedewerkers registreren. Ongestructureerde gegevens worden al geruime tijd beschouwd als zeer waardevol voor risicotaxatie (Chen et al., 2004; Hassani et al., 2016). Desondanks zijn er maar weinig instrumenten ontwikkeld op basis van ongestructureerde politiegegevens (Haleem et al., 2019; Karystianis et al., 2018). Daarnaast waren studies die gebruik hebben gemaakt van ongestructureerde gegevens niet zozeer gericht op risicotaxatie van personen, maar op identificatie van incidenten gerelateerd aan bijvoorbeeld huiselijk geweld (Poelman et al., 2010) en geestelijke gezondheidsproblemen (Haleem et al., 2019; Karystianis et al., 2018), of op de identificatie van verschillende soorten huiselijk geweld en letsel van slachtoffers (Karystianis et al., 2019). Tot op heden ontbreekt een instrument dat voor individuen, op basis van ongestructureerde politiegegevens, een inschatting maakt van het risico op herhaald slachtofferschap.

Studies in diverse onderzoeksgebieden hebben aangetoond dat gestructureerde en ongestructureerde gegevens samen de beste risico-inschatting opleveren (Adamou et al., 2018; Amrit et al., 2017; Härkänen et al., 2020; Hatef et al., 2020; Jonnagaddala et al., 2015; Laar et al., 2020). Het is de vraag of dit ook het geval is voor de risico-inschatting van herhaald slachtofferschap op basis van gestructureerde én ongestructureerde politiegegevens.

1.4 Onderzoeksvragen

In dit rapport worden twee onderzoeksvragen beantwoord. Ten eerste wordt onderzocht wat de voorspellende waarde is van ongestructureerde gegevens voor herhaald slachtofferschap. Ten tweede wordt de voorspellende waarde onderzocht van de combinatie van twee typen

gegevens: risicoscores gebaseerd op gestructureerde én risicoscores gebaseerd op ongestructureerde gegevens.

Dit rapport beantwoordt hiermee de volgende twee onderzoeksvragen.

- In hoeverre kan het risico op herhaald slachtofferschap worden ingeschat op basis van ongestructureerde gegevens uit BVH?
- In hoeverre kunnen ongestructureerde gegevens de risico-inschatting van ProVict Fase 1 (gestructureerde gegevens) verbeteren?

In de volgende hoofdstukken worden achtereenvolgens de methode, resultaten en conclusies beschreven.

Hoofdstuk 2. Methode

2.1 Onderzoeksdesign & data

Voor de beantwoording van de onderzoeksvragen is een dataset verzameld die bestond uit een random steekproef ($n = 150.000$) van alle personen die tussen 1 juli 2016 en 30 juni 2017 (selectiejaar) als slachtoffer met de politie in aanraking zijn gekomen. Het laatste slachtofferincident in het selectiejaar werd het indexincident genoemd. Informatie uit BVH tot 4 jaar voor het indexincident (historische periode) werd gebruikt om het risico op herhaald slachtofferschap binnen 2 jaar na het indexincident (toetsingsperiode) in te schatten (zie Figuur 1). Het indexincident was daardoor geen onderdeel van de historische of toetsingsperiode en speelde geen rol in de risico-inschatting. ProVict dient namelijk ter voorbereiding op het slachtoffergesprek, het indexincident is dan nog niet geregistreerd in BVH. Het onderzoeksdesign is analoog aan ProVict Fase I, maar er is een vernieuwde dataset verzameld met de meest actuele registraties uit BVH op dat moment. Voor specifiekere informatie over het onderzoeksdesign kan het rapport 'Ontwikkeling en validering van ProVict' (Geurts et al., 2019) worden geraadpleegd.

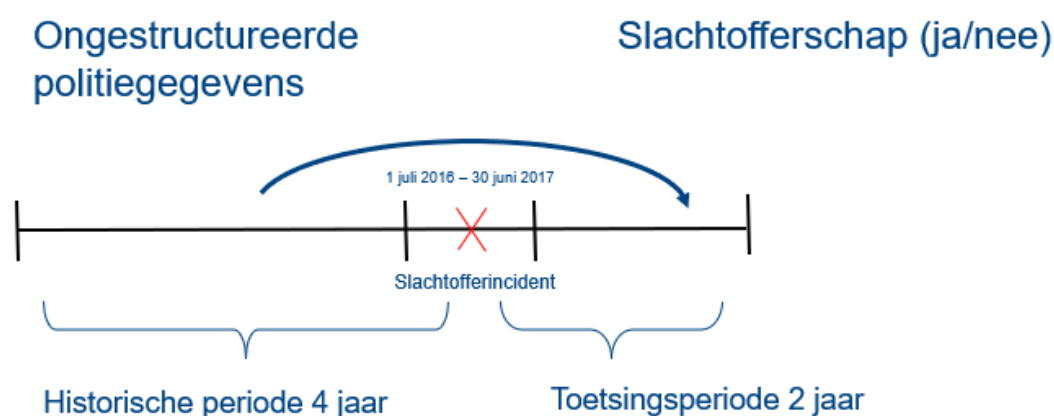
Zowel de gestructureerde als ongestructureerde gegevens waren afkomstig uit BVH. BVH wordt landelijk gebruikt door medewerkers van de politie bij het registreren van hun werkzaamheden en bevindingen: politiecontacten met slachtoffers, verdachten en betrokkenen worden in BVH vastgelegd door politiemedewerkers. Voor de gestructureerde gegevens werd van alle slachtoffers in de steekproef informatie verzameld met betrekking tot de ProVict-voorspellers in de historische periode. De ongestructureerde gegevens werden verkregen door voor elk slachtoffer in de steekproef alle registraties waaraan het slachtoffer was gekoppeld en daarin opgemaakte documenten uit de historische periode mee te nemen. Alleen de ongestructureerde gegevens in de tekstvelden *Toelichting*, *Verklaring* en *Bevinding* werden meegenomen.

Na verwijdering van personen die binnen twee jaar na het indexincident waren overleden, geen (geldige) geboortedatum hadden of van wie het geslacht onbekend was, bleven er nog 145.824 personen over in de dataset. Deze dataset werd random gesplitst in een ontwikkelset ($n = 72.922$) en een valideringsset ($n = 72.902$). De ontwikkelset werd gebruikt ten behoeve van de analyses voor de ontwikkeling van het op ongestructureerde gegevens gebaseerde model. Dit model werd vervolgens gevalideerd in de valideringsset. Tevens werd de valideringsset gebruikt om de validiteit te onderzoeken van de combinatie van op gestructureerde en ongestructureerde gegevens gebaseerde risicoscores.

Personen zonder registraties in de historische periode werden niet meegenomen in de ontwikkeling en validering van de modellen op basis van ongestructureerde gegevens. Deze

personen hebben geen gegevens in de tekstvelden die meegenomen worden voor de ongestructureerde gegevens. In de modellen op basis van de gestructureerde gegevens zijn deze personen wél meegenomen, omdat in deze modellen, naast politieregistraties, ook gebruik werd gemaakt van informatie over de leeftijd en het geslacht van het slachtoffer. Personen zonder registraties in de historische periode werden ook meegenomen in de analyses ten behoeve van de modellen die gestructureerde en ongestructureerde gegevens combineren, omdat in de praktijk behoefte is aan een risico-inschatting voor álle slachtoffers die bij de politie komen en niet alleen voor slachtoffers met eerdere politiecontacten.

Figuur 1: Onderzoeksdesign ProVict Fase 2



2.2 Variabelen

2.2.1 Afhankelijke variabelen

Herhaald slachtofferschap werd gedefinieerd als het opnieuw slachtoffer worden van een delict in de toetsingsperiode, ongeacht het type delict van het indexincident. Herhaald slachtofferschap werd geoperationaliseerd op basis van gestructureerde gegevens. Drie soorten herhaald slachtofferschap werden onderscheiden als afhankelijke variabelen: slachtofferschap van zeer ernstige delicten, ernstige delicten, en overige delicten. Bovenstaande is analoog aan ProVict Fase 1.

In Bijlage 1 is een overzicht van de geanalyseerde zeer ernstige, ernstige en overige delicten weergegeven. Zeer ernstige delicten zijn delicten die in het kader van de ontwikkeling van de individuele beoordeling (IB) zijn aangeduid als delicten met een hoge

impact⁴. Slachtoffers van dit type delicten worden verondersteld een verhoogd risico te hebben op herhaald slachtofferschap, secundaire victimisatie, intimidatie of vergelding. In het rapport van ProVict Fase 1 werden deze delicten ‘IB-delicten’ genoemd. Stalking, verkrachting en zware mishandeling zijn voorbeelden van zeer ernstige delicten. Het al dan niet slachtoffer worden van een zeer ernstig delict in de toetsingsperiode was de eerste afhankelijke variabele.

Ernstige delicten zijn delicten die niet behoren tot de lijst met IB-delicten, maar toch door slachtoffers als ingrijpend ervaren kunnen worden. In het rapport van ProVict Fase 1 werden deze delicten ‘overige ingrijpende delicten’ genoemd. Chantage, bedreiging en diefstallen met geweld zijn voorbeelden van ernstige delicten. Het al dan niet slachtoffer worden van een ernstig delict in de toetsingsperiode was de tweede afhankelijke variabele.

Tot slot werden delicten die niet in de andere twee categorieën vallen ingedeeld in de categorie ‘overige delicten’. Deze variabele omvat een groot aantal verschillende soorten delicten. Enkele voorbeelden van overige delicten zijn fietsendiefstal, heling en eenvoudige mishandeling. Het al dan niet slachtoffer worden van een overig delict was de derde afhankelijke variabele.

2.3 Data-analyses ongestructureerde data

2.3.1 Datapreparatie ongestructureerde data

De ongestructureerde gegevens werden verzameld door voor elk slachtoffer in de dataset alle registraties en daarin opgemaakte documenten⁵ uit de historische periode mee te nemen, ongeacht de rol die het slachtoffer bij deze registraties had. Met andere woorden, voor elk slachtoffer werd, binnen de eerder genoemde tekstvelden, gezocht naar alle beschikbare vrije teksten in de historische periode van registraties waar het slachtoffer aan gekoppeld was. Deze werden vervolgens samengevoegd in één overkoepelend document. Stopwoorden (veel voorkomende woorden die niet informatief zijn, zoals ‘de’, ‘het’ of ‘is’) werden verwijderd uit dit document. De aldus opgeschoonde teksten dienden als input voor de analyses. Voordat de vrije teksten konden worden geanalyseerd dienden de samengevoegde documenten te worden omgezet in getallen. Hiervoor zijn twee datapreparatiestrategieën gebruikt: count-vector en TF/IDF. Deze twee strategieën zijn beschreven in Bijlage 2.

⁴ Betreft 28 delict gerelateerde Maatschappelijke klasse en vijf projectcodes die i.s.m. het Openbaar Ministerie en partners in de strafrecht keten zijn vastgesteld.

⁵ Aan deze documenten kunnen ook andere personen verbonden zijn. Hierdoor kunnen de woorden in de vrije teksten kunnen, naast het slachtoffer, ook betrekking hebben op andere personen die aan het incident verbonden zijn (bijvoorbeeld verdachten of andere betrokkenen) of op het incident/de omstandigheden zelf.

2.3.2 Ontwikkeling model ongestructureerde data

Nadat de ongestructureerde data waren gekwantificeerd, werden twee machine learning methoden gebruikt voor het bepalen van de risicoscore voor herhaald slachtofferschap, namelijk decision tree en random forest. De twee analysemethoden, alsmede de voor- en nadelen van deze methoden, zijn beschreven in Bijlage 3.

2.3.3 Validering model ongestructureerde data

De twee datapreparatiestrategieën (count-vector en TF/IDF) en de twee analysemethoden (decision tree en random forests) resulteerden in vier modellen per afhankelijke variabele. In totaal werden er dus twaalf modellen ontwikkeld. Deze modellen werden ontwikkeld op de ontwikkelset en vervolgens toegepast op de valideringsset. Dit hield in dat voor elk slachtoffer in de valideringsset risicoscores werden berekend op basis van de modellen die waren ontwikkeld in de ontwikkelset.

Vervolgens werd de nauwkeurigheid van de modellen bepaald in termen van de AUC-waarden (Area Under the Curve). Deze AUC-waarden werden verkregen door middel van ROC-analyses (ROC: Receiver Operating Characteristic). De AUC-waarde is een veelgebruikte en aanbevolen indicator om de nauwkeurigheid van modellen te toetsen (Messing & Thaller, 2013; van der Put et al., 2019). In dit onderzoek staan AUC-waarden voor de kans dat een willekeurig herhaald slachtoffer een hogere risicoscore heeft dan een willekeurig niet-herhaald slachtoffer (Singh, 2013). Voor het beoordelen van de gevonden AUC-waarden baseerden we ons op de algemeen gehanteerde vuistregel dat een model acceptabel is als het een AUC-waarde heeft tussen 0,70 en 0,80, dat een model excellent is met een AUC-waarde tussen de 0,80 en 0,90 en dat een model voortreffelijk is bij een AUC-waarde boven de 0,90 (Hosmer & Lemeshow, 2000). Ten behoeve van de vervolganalyses werd op basis van de AUC-waarden per afhankelijke variabele het beste model gekozen.

2.4 Combinaties

Nadat de voorspellende waarde van modellen die gebaseerd zijn op ongestructureerde data was vastgesteld, werden er risicoscores berekend die op zowel gestructureerde als ongestructureerde data gebaseerd waren. Dit werd gedaan door de modellen op basis van gestructureerde gegevens (ProVict Fase I) te combineren met de modellen gebaseerd op ongestructureerde gegevens. Hieronder wordt eerst besproken hoe de op gestructureerde data gebaseerde risicoscore tot stand kwam en vervolgens hoe beide typen risicoscores werden gecombineerd.

2.4.1 Risicoscores gestructureerde data

De modellen die zijn ontwikkeld voor ProVict Fase I zijn voor het huidige onderzoek toegepast op de valideringsset. Dit betekent dat voor elk individu in de valideringsset risicoscores zijn berekend op basis van de modellen uit ProVict Fase 1. Voor de inhoud van deze modellen verwijzen we naar het rapport van ProVict Fase 1. De AUC-waarden van deze modellen zijn berekend om te onderzoeken in hoeverre het ontwikkelde en gevalideerde ProVict model uit Fase 1 ook op een andere (recentere) dataset valide risico-inschattingen oplevert. Daarmee werd ook een indicatie gegeven van de robuustheid van het model uit ProVict Fase 1.

2.4.2 Combinatiemethoden

Verschillende combinatiemethoden werden uitgevoerd om na te gaan in hoeverre een combinatie van op gestructureerde (ProVict Fase 1) en ongestructureerde gegevens gebaseerde risicoscores betere inschattingen oplevert dan beide typen risicoscores afzonderlijk. In dit onderzoek werden vijf verschillende combinatiemethoden onderzocht. Het doel van alle combinatiemethoden was om de twee risicoscores te combineren tot één risicoscore. De vijf combinatiemethoden waren *hoogste leidend*, *laagste leidend*, *gemiddelde* en twee *regressie modellen*. De beschrijving van deze methoden is te vinden in Bijlage 4.

De vijf combinatiemethoden werden vergeleken op basis van de AUC-waarden. Hierdoor kon per afhankelijke variabele worden bepaald wat de beste combinatiemethode was. Personen zonder politiecontacten in de historische periode ($n = 14.683$) hebben geen risicoscore gebaseerd op ongestructureerde gegevens. Voor deze personen was de score uit het model op basis van gestructureerde gegevens automatisch de nieuwe risicoscore, omdat het in de praktijk van belang is dat voor alle slachtoffers een score kan worden berekend, ook als er geen eerdere politiecontacten beschikbaar zijn.

2.5 Risicoscores indelen in risicocategorieën en predictieve validiteit

Per afhankelijke variabele werd de predictieve validiteit van de combinatiemethode met de hoogste AUC-waarde nader bestudeerd. Hiervoor werd een aantal afkappunten vastgesteld. In de praktijk is het nodig om met een afkappunt te werken. Dit afkappunt is een grenswaarde op basis waarvan twee risicocategorieën worden gecreëerd. Een politiemedewerker wil in het gebruik immers in één oogopslag kunnen zien of er sprake is van een verhoogd risico. Slachtoffers met een risicoscore hoger dan of gelijk aan de grenswaarde hebben een verhoogd risico op herhaald slachtofferschap. Slachtoffers met een risicoscore lager dan de grenswaarde hebben een laag risico op herhaald slachtofferschap.

Bij de keuze van het afkappunt spelen verschillende afwegingen een rol. Om deze afwegingen te faciliteren is per afhankelijke variabele van de beste combinatiemethode de predictieve validiteit berekend bij een reeks afkappunten. Hiervoor zijn de risicoscores ingedeeld in risicocategorieën. Het betrof 10 categorieën (oftewel risicodecielen) die elk 10

procent van de slachtoffers in de ontwikkelset vertegenwoordigden. De ondergrens van ieder deciel vertegenwoordigt hierbij een mogelijk afkappunt. Risicoscores groter of gelijk aan dit afkappunt vallen in de hoge risicocategorie, risicoscores onder dit afkappunt vallen in de lage risicocategorie. Uiteindelijk resulteerde deze exercitie in tien mogelijke afkappunten.

De predictieve validiteit bij deze afkappunten werd bepaald aan de hand van de volgende, in de literatuur aanbevolen, indicatoren: *sensitiviteit*, *specificiteit*, *Positive Predictive Value (PPV)*, *Negative Predictive Value (NPV)*, *Number Needed to Focus (NNF)* en *Number Safely Discharged (NSD)* (Singh, 2013). Deze indicatoren zijn schematisch weergegeven in Tabel 1. Bij sensitiviteit gaat het om hoeveel procent van de herhaalde slachtoffers in de hoge risicocategorie zit. Bij specificiteit gaat het om het percentage niet-herhaalde slachtoffers dat in de lage risicocategorie is ingedeeld. De PPV heeft betrekking op het percentage herhaalde slachtoffers binnen de hoge risicocategorie: het percentage echt-positieven. Echt-positieven zijn slachtoffers die in de hoge risicocategorie zijn ingedeeld en ook daadwerkelijk herhaald slachtoffer zijn in de toetsingsperiode. De NPV heeft betrekking op het percentage niet-herhaalde slachtoffers binnen de lage risicocategorie: het percentage echt-negatieven. Echt-negatieven zijn personen die in de lage risicocategorie werden ingedeeld en daadwerkelijk geen slachtofferschap ervaarden in de toetsingsperiode. Op basis van de PPV en de NPV werden ook de NNF en NND berekend. De NNF heeft betrekking op het aantal slachtoffers dat men in de hoge risicocategorie dient te ‘selecteren’ zodat daar minimaal één herhaald slachtoffer bij zit. De NSD heeft betrekking op het aantal slachtoffers dat men in de lage risicocategorie kan ‘negeren’ zonder dat daar één herhaald slachtoffer bij zit. De PPV, NPV, NNF en NSD zijn sterk afhankelijk van de verdeling van de afhankelijke variabele (in dit geval herhaald slachtofferschap) in de steekproef (Helmus & Babchishin, 2017). Hierdoor is het mogelijk dat bij een zeer nauwkeurig model, deze waarden desondanks ongunstig uitvallen als gevolg van het feit dat het aantal herhaalde slachtoffers veel lager is dan het aantal personen zonder herhaald slachtofferschap.

Tabel 1: Definities indicatoren predictieve validiteit

Indicator	Toelichting
Sensitiviteit	Het percentage herhaalde slachtoffers dat in de hoge risicocategorie is ingedeeld
Specificiteit	Het percentage niet-herhaalde slachtoffers dat in de lage risicocategorie is ingedeeld
PPV	Het percentage herhaalde slachtoffers in de hoge risicocategorie
NPV	Het percentage niet-herhaalde slachtoffers in de lage risicocategorie
NNF	Het aantal slachtoffers in de hoge risicocategorie voordat er één herhaald slachtoffer in de hoge risicocategorie aanwezig is
NSD	Het aantal niet-herhaalde slachtoffers in de lage risicocategorie voordat er één herhaald slachtoffer in deze categorie aanwezig is

Noot: PPV = positive predictive value, NPV = negative predictive value, NNF = number needed to focus on, NSD = number safely discharged.

2.6 Robuustheidsanalyses

Tenslotte is het voor de praktijk belangrijk om te weten in hoeverre de ontwikkelde modellen nauwkeurig zijn voor verschillende groepen. Vandaar dat ook de robuustheid van de modellen getest is. De robuustheid van de modellen werd onderzocht door de AUC-waarden te berekenen voor verschillende subgroepen (gevormd op basis van leeftijd of geslacht).

Hoofdstuk 3. Resultaten

In dit hoofdstuk worden allereerst de resultaten gepresenteerd van de modellen die uitsluitend zijn gebaseerd op respectievelijk ongestructureerde en gestructureerde gegevens. Vervolgens worden de resultaten gepresenteerd van de combinaties van risicoscores afkomstig van de twee soorten modellen. Tot slot worden de resultaten van de robuustheidsanalyses besproken.

3.1 Beschrijvende statistieken

De leeftijd van slachtoffers in de steekproef varieerde van 0 tot 100 jaar ($M = 41,63$; $SD = 17,28$). Van alle slachtoffers in de steekproef was 56% man en 44% vrouw. In de totale dataset vond herhaald slachtofferschap van een zeer ernstig delict plaats bij 0,7% van de slachtoffers. Voor ernstige delicten en overige delicten was dit percentage respectievelijk 2,2% en 10,9% in de hele dataset.

3.2 Modellen ongestructureerde gegevens

Tabel 2 geeft een overzicht van de AUC-waarden van de verschillende modellen op basis van de ongestructureerde gegevens. De analysemethode random forest leverde voor alle drie de afhankelijke variabelen hogere AUC-waarden op dan de decision tree methode. Vervolgens is gekeken welke datapreparatiestrategie (count-vector of TF/IDF) binnen de random forest methode de hoogste AUC-waarden opleverde. Voor zeer ernstige delicten en ernstige delicten was dit de count-vector methode (zeer ernstige delicten = 0,796, ernstige delicten = 0,760). Het model voor overige delicten behaalde de hoogste AUC-waarde met de TF/IDF methode (0,631), al was het verschil met de count-vector methode minimaal. Ook voor overige delicten werd het random forest count-vector model gekozen als beste model, omwille van de consistentie met de twee andere afhankelijke variabelen en omdat de verschillen minimaal waren.

Voor zeer ernstige delicten en ernstige delicten was de AUC-waarde boven de 0,70, hetgeen duidt op een acceptabele risico-inschatting. De AUC-waarde van het model voor overige delicten was lager dan 0,70. De inschatting voor herhaald slachtofferschap van overige delicten was dus het minst nauwkeurig en valt onder de grens die door Hosmer en Lemeshow (2000) is vastgesteld als acceptabel.

Tabel 2: AUC-waarden valideringsset modellen ongestructureerde gegevens

Analyse methode	Datapreparatie strategie	Zeern ernstige delicten	Ernstige delicten	Overige delicten
Decision tree	Count-vector	0,735	0,702	0,605
	TF/IDF	0,737	0,712	0,611
Random forest	Count-vector	0,796	0,760	0,630
	TF/IDF	0,794	0,758	0,631

Noot: Valideringsset, $n = 72.902$

3.3 Modellen gestructureerde gegevens

De ontwikkelde en gevalideerde modellen op basis van gestructureerde gegevens (zoals beschreven in ‘Ontwikkeling en validering van ProVict’) zijn toegepast op de valideringsset van ProVict Fase 2. In Tabel 3 zijn de bijbehorende AUC-waarden weergegeven. Deze AUC-waarden zijn vergelijkbaar met de AUC-waarden zoals gepresenteerd in het rapport van ProVict Fase 1. Dit betekent dat de modellen uit ProVict Fase 1 vergelijkbaar presteren op een nieuwe, recentere dataset. Voor zeer ernstige delicten lag de AUC-waarde ruimschoots boven 0,70, hetgeen duidt op een acceptabele risico-inschatting. Voor ernstige delicten lag de AUC-waarde slechts een fractie onder 0,70, waardoor ook hier nog gesproken kan worden van een acceptabele risico-inschatting (Hosmer & Lemeshow, 2000). Overige delicten konden minder nauwkeurig worden voorspeld gezien de AUC-waarde van 0,627.

Tabel 3: AUC-waarden valideringsset modellen gestructureerde gegevens (ProVict Fase 1)

	AUC-waarden
Zeern ernstige delicten	0,757
Ernstige delicten	0,693
Overige delicten	0,627

Noot: Valideringsset, $n = 72.902$

3.4 Combinatie modellen gestructureerde en ongestructureerde gegevens

De risicoscores op basis van gestructureerde gegevens en ongestructureerde gegevens zijn gecombineerd op vijf verschillende manieren (zie Bijlage 4). Voor de gestructureerde gegevens werden de risicoscores berekend op basis van de modellen die ontwikkeld zijn voor ProVict Fase 1. Voor de ongestructureerde gegevens gebeurde dit op basis van de modellen die uitgaan van de count-vector datapreparatiestrategie en de random forest analysemethode.

Tabel 4 laat voor alle combinatiemethoden de AUC-waarden zien van de risicoscores gebaseerd op gestructureerde en ongestructureerde gegevens. De methode waarbij het

gemiddelde wordt berekend van beide risicoscores leverde de hoogste AUC-waarden op voor zeer ernstige delicten (0,815) en ernstige delicten (0,752). Voor overige delicten zijn de verschillen tussen de combinatiemethoden minimaal.

Met betrekking tot zeer ernstige en ernstige delicten kan op basis hiervan worden geconcludeerd dat de methode waarbij het gemiddelde van de twee scores wordt berekend tot het beste model leidt. Omwille van de consistentie wordt ook voor overige delicten geconcludeerd dat de combinatiemethode die gebaseerd is op het gemiddelde tot het beste model leidt. De AUC-waarde van dit combinatiemodel voor zeer ernstige delicten is groter dan 0,80, hetgeen duidt op een uitstekende risico-inschatting. Het combinatiemodel voor ernstige delicten heeft een AUC-waarde boven de 0,70 en kan daarmee worden beschouwd als een acceptabel model. Het model voor overige delicten is het minst nauwkeurige model met een AUC-waarde lager dan 0,70.

Tabel 4: AUC-waarden valideringsset combinaties count-vector random forest modellen en ProVict-modellen

Combinatiemethoden	Ze er n st i g e d e l i c t e n	E r n s t i g e d e l i c t e n	O v e r i g e d e l i c t e n
Hoogste leidend	0,809	0,743	0,644
Laagste leidend	0,786	0,733	0,641
Gemiddelde	0,815	0,752	0,653
Logistische regressie beide scores	0,811	0,740	0,653
Logistische regressie beide scores + interactie term	0,808	0,748	0,653

Noot: Valideringsset, n = 72.902

3.4.1 Predictieve validiteit combinatiemodel

Tabellen 5, 6 en 7 laten voor de beste combinatiemodellen de predictieve validiteitsindicatoren zien bij verschillende afkappunten. Tabel 5 laat deze indicatoren zien voor het model voor zeer ernstige delicten. Zoals eerder aangegeven is de keuze voor het afkappunt afhankelijk van verschillende afwegingen. Zo is deze keuze afhankelijk van de afweging tussen enerzijds de voordelen van het terecht toekennen van een verhoogd risico aan een herhaald slachtoffer, en anderzijds de nadelen van het ten onrechte toekennen van een verhoogd risico aan een niet-herhaald slachtoffer, of het ten onrechte toekennen van een laag risico aan een herhaald slachtoffer. Als sensitiviteit en specificiteit beide optimaal dienen te zijn kan het afkappunt bij de top 30% worden gelegd. In dat geval wordt 78,6% van de herhaalde slachtoffers in de hoge categorie ingedeeld en 70,3% van de niet-herhaalde slachtoffers in de lage categorie. Een ruime meerderheid van beide groepen (herhaalde en

niet-herhaalde slachtoffers) wordt dan in de juiste categorie ingedeeld. De NNF is bij dit afkappunt 50, wat wil zeggen dat er zich onder elke 50 slachtoffers in de hoge risicocategorie één herhaald slachtoffer (echt-positieve) bevindt. De NSD is 427, wat wil zeggen dat onder elke 428 slachtoffers in de lage risicocategorie zich één herhaald slachtoffer (vals-negatieve) bevindt. Wanneer naar de PPV gekeken wordt blijkt dat bij dit afkappunt slechts een klein deel van de hoge risicocategorie bestaat uit herhaalde slachtoffers (2,0% in de valideringsset). Zoals eerder aangegeven heeft dit te maken met de verdeling van herhaald slachtofferschap voor zeer ernstige delicten in de toetsingsperiode. Slechts een zeer klein percentage (0,8% in de valideringsset) is in de toetsingsperiode slachtoffer van een dergelijk delict.

Tabel 5: Predictieve validiteitsindicatoren per afkappunt voor zeer ernstige delicten; combinatiemethode gebaseerd op gemiddelde van de twee scores

Afkappunt	Sensitiviteit	Specificiteit	PPV	NPV	NNF	NSD
Top 100%	100,0%	0,0%	0,8%	0,0%	131	0
Top 90%	98,4%	10,0%	0,8%	99,9%	120	803
Top 80%	96,9%	19,8%	0,9%	99,9%	109	842
Top 70%	95,0%	30,2%	1,0%	99,9%	97	780
Top 60%	92,4%	40,2%	1,2%	99,9%	85	693
Top 50%	89,6%	50,3%	1,4%	99,8%	73	627
Top 40%	85,4%	60,2%	1,6%	99,8%	62	537
Top 30%	78,6%	70,3%	2,0%	99,8%	50	427
Top 20%	69,1%	80,4%	2,6%	99,7%	38	338
Top 10%	49,6%	90,3%	3,8%	99,6%	26	233

Noot: Valideringsset, n = 72.902

In Tabel 6 zijn de predictieve validiteitsindicatoren per afkappunt gepresenteerd voor het model voor ernstige delicten. In Tabel 4 was al te zien dat de AUC-waarde van het model voor ernstige delicten lager was dan voor zeer ernstige delicten. Dit is ook terug te zien in de waarden van de sensitiviteit en specificiteit. Beide waarden zijn optimaal wanneer het afkappunt bij de top 30% wordt gelegd. Van de herhaalde slachtoffers wordt 67,3% dan in de hoge categorie ingedeeld, terwijl 70,8% van de niet-herhaalde slachtoffers in de lage categorie wordt ingedeeld. Met andere woorden, ook voor dit model wordt bij dit afkappunt de meerderheid van de herhaalde slachtoffers en niet-herhaalde slachtoffers in de juiste categorie ingedeeld. De NNF bij dit afkappunt is 21, wat inhoudt dat er zich onder elke 21 slachtoffers in de hoge risicocategorie één herhaald slachtoffer bevindt (echt-positieve). De

NSD is 99 wat inhoudt dat zich onder elke 100 slachtoffers in de lage risicocategorie één herhaald slachtoffer (vals-negatieve) bevindt. Aangezien herhaald slachtofferschap in de toetsingsperiode voor ernstige delicten vaker voorkomt (2,1% in de valideringsset) dan voor zeer ernstige delicten is de PPV hier iets hoger. Echter is bij het gehanteerde afkappunt nog steeds slechts een klein deel (4,8% in de valideringsset) van de slachtoffers in de hoge risicocategorie uiteindelijk herhaald slachtoffer.

Tabel 6: Predictieve validiteitsindicatoren per afkappunt voor ernstige delicten; combinatiemethode gebaseerd op gemiddelde van de twee scores

Afkappunt	Sensitiviteit	Specificiteit	PPV	NPV	NNF	NSD
Top 100%	100,0%	0,0%	2,1%	0,0%	47	0
Top 90%	97,8%	10,1%	2,3%	99,5%	43	206
Top 80%	93,6%	20,3%	2,5%	99,3%	40	145
Top 70%	91,2%	30,3%	2,8%	99,4%	36	157
Top 60%	87,1%	40,6%	3,1%	99,3%	32	144
Top 50%	81,7%	50,7%	3,5%	99,2%	29	126
Top 40%	76,3%	60,8%	4,1%	99,2%	24	117
Top 30%	67,3%	70,8%	4,8%	99,0%	21	99
Top 20%	56,4%	80,8%	6,1%	98,8%	17	84
Top 10%	38,0%	90,6%	8,2%	98,5%	12	67

Noot: Valideringsset, n = 72.902

In Tabel 7 zijn de predictieve validiteitsindicatoren van het model voor overige delicten weergegeven. Dit model heeft de laagste AUC-waarde vergeleken met de modellen voor zeer ernstige en ernstige delicten. Dit is terug te zien in de sensitiviteit en de specificiteit. Wanneer slachtoffers met de top 40% risicoscores worden ingedeeld in de hoge risicocategorie zijn sensitiviteit en specificiteit optimaal. Een meerderheid van de herhaalde slachtoffers (60,3 %) wordt dan terecht in de hoge categorie ingedeeld, terwijl een meerderheid van de niet-herhaalde slachtoffer (62,4%) terecht in de lage categorie wordt ingedeeld. Ondanks dat een meerderheid van de herhaalde en niet-herhaalde slachtoffers nog steeds correct wordt ingedeeld, zijn de percentages bij dit model lager dan voor de modellen voor zeer ernstige delicten en ernstige delicten. Bij het afkappunt van de top 40% geeft de NNF aan dat per 6 slachtoffers in de hoge risicocategorie, één persoon herhaald slachtoffer is (echt-positieve). De NSD van 13 geeft aan dat er zich onder elke 14 slachtoffers in de lage risicocategorie één herhaald slachtoffer bevindt dat onterecht in die categorie is ingedeeld (vals-negatieve). De

PPV voor dit model is bij het gehanteerde afkappunt een stuk hoger (16,4%) dan bij de eerdere modellen. Hierbij bestaat echter een groot deel van de hoge risicocategorie nog steeds uit niet-herhaalde slachtoffers.

Tabel 7: Predictieve validiteitsindicatoren per afkappunt voor overig delicten; combinatiemethode gebaseerd op gemiddelde van de twee scores

Afkappunt	Sensitiviteit	Specificiteit	PPV	NPV	NNF	NSD
Top 100%	100,0%	0,0%	10,9%	0,0%	9	0
Top 90%	95,6%	10,6%	11,5%	95,2%	9	20
Top 80%	89,9%	21,1%	12,2%	94,5%	8	17
Top 70%	83,5%	31,6%	13,0%	94,0%	8	16
Top 60%	76,7%	42,0%	13,9%	93,7%	7	15
Top 50%	68,9%	52,3%	15,0%	93,3%	7	14
Top 40%	60,3%	62,4%	16,4%	92,8%	6	13
Top 30%	49,3%	72,3%	17,8%	92,1%	6	12
Top 20%	37,4%	82,1%	20,3%	91,5%	5	11
Top 10%	22,4%	91,5%	24,4%	90,6%	4	10

Noot: Valideringsset, $n = 72.902$

3.4.2 Robuustheidsanalyses

Tabel 8 laat zien hoe de combinatiemodellen voor de drie afhankelijke variabelen presteren voor verschillende subgroepen. Het model voor zeer ernstige delicten blijkt voor mannen en vrouwen nagenoeg even nauwkeurig (AUC-waarden respectievelijk 0,820 en 0,812). Wat betreft leeftijd blijkt het model voor 0 tot 12-jarigen en 65+ minder nauwkeurige inschattingen op te leveren dan voor andere leeftijdsgroepen. Voor het model voor ernstige delicten geldt dit alleen voor de groep 65+. Voor mannen, vrouwen en de leeftijdsgroepen onder 65 jaar levert het model nauwkeurige risico-inschattingen op: alle AUC-waarden liggen boven de 0,70. Het model voor overige delicten bleek alleen voor de groep 0 tot 12 jaar boven de 0,70 uit te komen. Hier kunnen echter geen conclusies aan worden verbonden, aangezien de AUC-waarde hier is gebaseerd op een zeer kleine groep herhaalde slachtoffers.

Tabel 8: AUC-waarden valideringsset voor subgroepen; combinatiemethode gebaseerd op gemiddelde van de twee scores

	Zeer ernstige delicten	Ernstige delicten	Overige delicten
Mannen	0,820	0,750	0,647
Vrouwen	0,812	0,760	0,656
0 tot 12 jaar	0,645	0,733	0,750
12 tot 15 jaar	0,724	0,770	0,643
15 tot 18 jaar	0,803	0,806	0,660
18 tot 23 jaar	0,744	0,780	0,634
23 tot 65 jaar	0,808	0,761	0,643
65+ jaar	0,624	0,591	0,594

Noot: Valideringsset, $n = 72.902$

Hoofdstuk 4. Conclusie

4.1 Conclusie

In onderhavig rapport wordt de ontwikkeling en validering beschreven van ProVict Fase 2. ProVict is een verzameling algoritmen die het risico berekent dat een slachtoffer (van om het even welk delict) binnen twee jaar opnieuw als slachtoffer met de politie in aanraking komt. Voor ProVict Fase 1 werden modellen ontwikkeld die uitsluitend op gestructureerde politiegegevens waren gebaseerd. In ProVict Fase 2 werden deze modellen uitgebreid met ongestructureerde gegevens. Twee onderzoeksvragen stonden centraal:

1. In hoeverre kan het risico op herhaald slachtofferschap worden ingeschat op basis van ongestructureerde gegevens uit BVH?
2. In hoeverre kunnen ongestructureerde gegevens de risico-inschatting van ProVict Fase 1 (gestructureerde gegevens) verbeteren?

De eerste onderzoeksvraag werd beantwoord door voor drie afhankelijke variabelen (zeer ernstige delicten, ernstige delicten en overige delicten) modellen te ontwikkelen op basis van ongestructureerde gegevens. Per afhankelijke variabele werden vier modellen ontwikkeld die varieerden wat betreft de datapreparatiestrategie en analysemethode. De ontwikkelde modellen werden gevalideerd in een onafhankelijke valideringsset. De resultaten toonden aan dat de datapreparatiestrategie count-vector in combinatie met de analysemethode random forest tot de hoogste AUC-waarden leidden. Een bijkomend voordeel van het feit dat de count-vector strategie tot het beste model leidde is dat de resultaten hiervan makkelijker te interpreteren zijn dan die van de TD-IDF strategie. Dit is voor de politiepraktijk belangrijk in het kader van transparantie. De AUC-waarden gaven aan dat de modellen voor zeer ernstige en ernstige delicten de meest nauwkeurige inschattingen van het risico op herhaald slachtofferschap opleveren. De resultaten lieten zien dat de modellen voor overige delicten de laagste AUC-waarden hadden en dus het minst presteerden wat betreft nauwkeurigheid.

Om de tweede onderzoeksvraag te beantwoorden werden op vijf verschillende manieren de risicoscores gebaseerd op gestructureerde gegevens (ProVict Fase 1) en de risicoscores gebaseerd op ongestructureerde gegevens gecombineerd. Voor elk van de drie afhankelijke variabelen leverde de methode waarbij een gemiddelde van de twee risicoscores werd berekend de beste modellen op. Voor deze methode waren de AUC-waarden namelijk het hoogst. Voor zeer ernstige delicten was de AUC-waarde groter dan 0,80, hetgeen duidt op een uitstekend model wat betreft algemene nauwkeurigheid. Voor ernstige delicten was de AUC-waarde groter dan 0,70, hetgeen duidt op een acceptabel model. Het model voor overige delicten was minder nauwkeurig met een AUC-waarde onder de 0,70. De AUC-waarden

van de combinatiemodellen waren hoger dan de AUC-waarden van ProVict Fase 1 en hoger dan de modellen gebaseerd op uitsluitend ongestructureerde gegevens. Hieruit blijkt dat het risico op herhaald slachtofferschap beter kan worden ingeschat met een combinatie van gestructureerde en ongestructureerde gegevens dan met gestructureerde of ongestructureerde gegevens alleen.

De AUC-waarden per subgroep suggereren dat de modellen voor zeer ernstige delicten en ernstige delicten nauwkeurig waren voor zowel mannen als vrouwen. Met betrekking tot verschillende leeftijdsgroepen suggereren de resultaten dat het model voor zeer ernstige delicten nauwkeurig was voor slachtoffers van 12 tot 65 jaar. Er dient rekening te worden gehouden dat het model voor zeer ernstige delicten voor de groepen 0 tot 12 jaar en 65+ minder nauwkeurige inschattingen opleverde. Voor ernstige delicten gold dit alleen voor de groep 65+, voor de andere leeftijdsgroepen werden nauwkeurige voorspellingen gevonden.

In eerste instantie werd de AUC-waarde gebruikt om de algemene nauwkeurigheid van een model te testen. De AUC-waarde geeft echter slechts een globaal beeld van de predictieve validiteit van een model. De evaluatie van een model hangt daarom niet alleen af van de algemene indicator. Ook specifiekere indicatoren dienen in ogenschouw te worden genomen. Om een beter beeld te krijgen van de predictieve validiteit van de combinatiemodellen met de hoogste AUC-waarden werden predictieve validiteitsindicatoren berekend voor meerdere afkappunten. In de praktijk is een afkappunt noodzakelijk, aangezien een politiemedewerker in één oogopslag wil kunnen zien of er sprake is van een verhoogd risico op herhaald slachtofferschap. Uit de resultaten voor de predictieve validiteitsindicatoren bleek dat bij een afkappunt waarbij zowel sensitiviteit als specificiteit werden gemaximaliseerd, een meerderheid van de herhaalde slachtoffers terecht in de hoge risicocategorie werd ingedeeld (echt-positieven). Tegelijkertijd werd een meerderheid van de niet-herhaalde slachtoffers terecht in de lage risicocategorie ingedeeld (echt-negatieven). Dit was het geval voor alle drie de afhankelijke variabelen. Analooq aan de resultaten op basis van de AUC-waarden bleek het model voor zeer ernstige delicten het meest nauwkeurig in termen van sensitiviteit en specificiteit, gevolgd door het model voor ernstige delicten. Het model voor overige delicten bleek het minst nauwkeurig. Daarnaast bleek voor alle drie de afhankelijke variabelen dat de hoge risicocategorie slechts voor een klein deel bestond uit herhaalde slachtoffers. Dit heeft slechts deels met de nauwkeurigheid van het model te maken en voornamelijk met de verdeling van de afhankelijke variabele: het aantal herhaalde slachtoffers in de dataset is namelijk zeer gering. Voor het gebruik van een model in de praktijk is het echter relevant inzicht te hebben in het aantal vals-positieven en vals-negatieven bij een bepaald afkappunt.

De uiteindelijke keuze om een model in de praktijk in te zetten hangt af van een zorgvuldige afweging van verschillende indicatoren. De AUC-waarde geeft hierbij een algemeen beeld, maar indicatoren als de specificiteit en de sensitiviteit geven een beeld van de praktische gevolgen van een model bij een bepaald afkappunt. Zowel de algemene indicator als de specifieke indicatoren kunnen bijdragen aan de keuze voor een model. De AUC-waarde is, met andere woorden, niet leidend in de keuze om een model wel of niet in de praktijk in te zetten. De keuze hangt dan ook af van het belang dat de politiepraktijk hecht aan bijvoorbeeld de sensitiviteit en specificiteit van de modellen bij een bepaald afkappunt. De uiteindelijke keuze voor een geschikt afkappunt hangt hiermee samen. Vanuit de politiepraktijk kan het bijvoorbeeld wenselijk zijn om waarde te hechten aan óf de sensitiviteit óf de specificiteit. Ook is het mogelijk dat er andere aspecten (zoals capaciteit) meespelen in de keuze voor een afkappunt. Door middel van overleg met verschillende partijen en deskundigen dient te worden bepaald in welke mate belang wordt gehecht aan respectievelijk sensitiviteit en specificiteit.

Samengevat kunnen de modellen uit ProVict Fase 1 worden verbeterd door ongestructureerde gegevens hieraan toe te voegen. De modellen uit ProVict Fase 1 konden (1) herhaald slachtofferschap van zeer ernstige delicten voor alle slachtoffers en (2) herhaald slachtofferschap van ernstige delicten voor 15- tot 23-jarigen het meest nauwkeurig voorspellen. Modellen die gebruik maken van een combinatie van gestructureerde én ongestructureerde gegevens konden (1) herhaald slachtofferschap van zeer ernstige delicten en (2) herhaald slachtofferschap van ernstige delicten voor alle slachtoffers tot 65 jaar het meest nauwkeurig voorspellen. Het risico op herhaald slachtofferschap kan beter worden ingeschat met een combinatie van gestructureerde en ongestructureerde gegevens dan met deze gegevens afzonderlijk.

4.2 Discussie

Doel van dit rapport is de politie te voorzien van relevante inzichten op het gebied van risicotaxatie van herhaald slachtofferschap. De resultaten laten zien dat het analyseren van politiegegevens kan leiden tot nauwkeurige inschattingen van het risico op herhaald slachtofferschap. De meest nauwkeurige inschattingen worden gerealiseerd door gebruik te maken van zowel gestructureerde als ongestructureerde politiegegevens. Deze inschattingen kunnen politiemedewerkers ondersteunen in hun dienstverlening richting slachtoffers. Bovenstaande bevindingen dienen echter van enkele kanttekeningen te worden voorzien.

Ten eerste hebben we bij de ontwikkeling en validering van ProVict Fase 1 en 2 uitsluitend gebruik gemaakt van officiële politiegegevens. Enerzijds kent dit een aantal voordelen: in tegenstelling tot het gebruik van bijvoorbeeld vragenlijstgegevens speelt sociale

wenselijkheid of selectieve herinnering geen rol bij dergelijke gegevens. Ook maakt het gebruik van uitsluitend officiële politiegegevens het beter mogelijk een dataset te genereren die van dezelfde personen over een lange periode informatie bevat. Anderzijds kleven er ook enkele nadelen aan het uitsluitende gebruik van politiegegevens. Zo bestaat het risico dat de werkelijke mate waarin herhaald slachtofferschap voorkomt, onderschat wordt als gevolg van een lage aangiftebereidheid. Ervaringen bij eerdere aangiftes zijn van invloed op de keuze om aangifte te doen bij nieuw slachtofferschap. Wanneer slachtoffers het gevoel hebben gehad niet rechtvaardig of correct behandeld te zijn bij hun aangifte is de kans kleiner dat zij bij toekomstig slachtofferschap contact opnemen met de politie en dus opnieuw aangifte doen (Bolger & Bolger, 2019; Koster, 2017).

Ten tweede zijn ongestructureerde gegevens geanalyseerd die afkomstig zijn uit registraties waaraan het slachtoffer verbonden was, in om het even welke rol. Aan deze registraties kunnen ook andere personen verbonden zijn. De woorden in de vrije teksten kunnen dus, naast het slachtoffer, ook betrekking hebben op andere personen die aan het incident verbonden zijn (bijvoorbeeld verdachten of andere betrokkenen) of op het incident/de omstandigheden zelf. Ondanks dat de woorden wellicht niet uitsluitend op het slachtoffer van toepassing zijn, zijn ze wel gebruikt in een registratie waarbij het slachtoffer betrokken was. Die woorden kunnen namelijk wel degelijk informatie bevatten die voor de risico-inschatting relevant is. Daarnaast was het doel van dit onderzoek om de voorspellende waarde van ongestructureerde gegevens vast te stellen en eventueel ProVict Fase 1 uit te breiden met ongestructureerde gegevens. Het doel was niet te onderzoeken welke specifieke ongestructureerde gegevens (woorden) voorspellend waren voor herhaald slachtofferschap.

Ten derde werd in dit rapport voor de ontwikkeling van modellen op basis van ongestructureerde gegevens gebruik gemaakt van twee machine learning methoden: decision trees en random forests. Zoals beschreven in Bijlage 3 hebben beide methoden hun eigen voor- en nadelen. Decision trees zijn bijvoorbeeld beter te interpreteren, maar minder robuust. Random forests zijn meestal robuuster dan decision trees, maar zijn minder goed te interpreteren. De resultaten toonden aan dat modellen die waren ontwikkeld met de random forest-methode de hoogste AUC-waarden hadden. Aangezien de random forest modellen dus de beste predictieve validiteit (in termen van de AUC-waarde) lieten zien zijn ze gebruikt voor de combinatiemodellen. Hierdoor is er gebruik gemaakt van een model dat wel nauwkeurig is, maar minder goed te interpreteren. Met het oog op de transparantie is het daarom belangrijk dat er aandacht wordt besteed aan de uitlegbaarheid van deze modellen in de praktijk.

4.3 Aanbevelingen toekomstig onderzoek

Naar aanleiding van het huidige onderzoek worden enkele suggesties voor toekomstig onderzoek geformuleerd. Ten eerste zou vervolgonderzoek zich kunnen richten op verdere verfijning van de risicotaxatie op basis van ongestructureerde gegevens. Zoals aangegeven is het op dit moment mogelijk dat woorden uit de vrije teksten geen betrekking hebben op het slachtoffer, maar op andere personen die betrokken zijn bij een registratie of op algemene omstandigheden. Het is mogelijk dat dit zorgt voor een minder nauwkeurige risico-inschatting. Middels extra analyses zouden alleen woorden die betrekking hebben op het slachtoffer kunnen worden meegenomen in de analyses. Toekomstig onderzoek kan uitwijzen of dit tot nauwkeurigere risico-inschattingen leidt.

Ten tweede kan er in toekomstig onderzoek meer aandacht worden besteed aan de interpretatie van de voorspellers van het model dat gebaseerd is op ongestructureerde gegevens. Toekomstig onderzoek zou de inhoud en betekenis van deze woorden kunnen onderzoeken. Deze woorden kunnen leiden tot meer kennis met betrekking tot mogelijke risico- en beschermende factoren voor herhaald slachtofferschap en eventuele interventies.

Ten derde kan vervolgonderzoek zich richten op herhaald slachtofferschap van overige delicten. Herhaald slachtofferschap van overige delicten was in zowel ProVict Fase 1 en Fase 2 het minst nauwkeurig te voorspellen. Dit kan mogelijk verklaard worden door het feit dat de categorie overige delicten erg heterogeen was: de categorie bestond uit vele verschillende soorten slachtofferschap. Toekomstig onderzoek kan verkennen of de overige delicten uit verschillende subcategorieën bestaand, die wellicht wel nauwkeurig voorspeld kunnen worden.

Tot slot wordt aanbevolen de validiteit van ProVict periodiek te onderzoeken. Criminaliteit, en daarmee ook slachtofferschap, is onderhevig aan veranderingen. Dit geldt ook voor de registratiesystemen (in dit geval BVH) van de politie. Op basis van onderhoud kunnen de modellen zo nodig worden aangepast. Het is van belang dergelijk onderhoud in de politieorganisatie te beleggen. Aangezien dit rapport aantoont dat een combinatie van gestructureerde en ongestructureerde gegevens de meest nauwkeurige inschattingen van herhaald slachtofferschap oplevert, is het van belang te bepalen hoe en waar het onderhoud van een dergelijk combinatiemodel kan plaatsvinden.

Literatuurlijst

- Adamou, M., Lagani, V., Antoniou, G., Charonyktakis, P., Greasidou, E., & Tsamardinos, I. (2018). Mining free-text medical notes for suicide risk assessment. *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, 1-8.
<https://doi.org/10.1145/3200947.3201020>
- Amrit, C., Paauw, T., Aly, R., & Lavric, M. (2017). Identifying child abuse through text mining and machine learning. *Expert Systems with Applications*, 88, 402-418.
<https://doi.org/10.1016/j.eswa.2017.06.035>
- Berk, R. A., & Sorenson, S. B. (2020). Algorithmic approach to forecasting rare violent events: An illustration based in intimate partner violence perpetration. *Criminology and Public Policy*, 19, 213-233. <https://doi.org/10.1111/1745-9133.12476>
- Bolger, M. A., & Bolger, P. C. (2019). Predicting fear of crime: results from a community survey of a small city. *American Journal of Criminal Justice*, 44(2), 334-351.
<https://doi.org/10.1007/s12103-018-9450-x>
- Brand, J. E., Xu, J., Koch, B., & Geraldo, P. (2021). Uncovering sociological effect heterogeneity using tree-based machine Learning. *Sociological Methodology*, 1-35.
<https://doi.org/10.1177/0081175021993503>
- Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer*, 37(4).
<http://hdl.handle.net/10722/45461>
- Geurts, R., Raaijmakers, N., Delsing, M., Scholte, R., & Wientjes, J. (2019). *Ontwikkeling en validering van ProVict*.
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267-297.
<https://doi.org/10.1093/pan/mps028>
- Haleem, M. S., Han, L., Harding, P. J., & Ellison, M. (2019). An automated text mining approach for classifying mental-ill health incidents from police incident logs for data-driven intelligence. *IEEE International Conference on Systems, Man and Cybernetics*, 2279-2284. <https://doi.org/10.1109/SMC.2019.8914240>
- Härkänen, M., Vehviläinen-Julkunen, K., Murrells, T., Paananen, J., Franklin, B. D., & Rafferty, A. M. (2020). The contribution of staffing to medication administration errors: A text mining analysis of incident report data. *Journal of Nursing Scholarship*, 52(1), 113-123. <https://doi.org/10.1111/jnu.12531>
- Hassani, H., Huang, X., Silva, S. E., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(3), 141-154. <https://doi.org/10.1002/sam>

- Hatef, E., Ma, X., Rouhizadeh, M., Singh, G., Weiner, J., & Kharrazi, H. (2020). Assessing the impact of social needs and social determinants of health on health care utilization: Using patient- and community-level data. *Population Health Management*, 1-9. <https://doi.org/10.1089/pop.2020.0043>
- Hegelich, S. (2016). Decision trees and random forests: Machine learning techniques to classify rare events. *European Policy Analysis*, 2(1), 98-120. <https://doi.org/10.18278/epa.2.1.7>
- Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior*, 44(1), 8-25. <https://doi.org/10.1177/0093854816678898>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Sons. <https://doi.org/10.1002/0471722146>
- Jonnagaddala, J., Liaw, S. T., Ray, P., Kumar, M., Chang, N. W., & Dai, H. J. (2015). Coronary artery disease risk assessment from unstructured electronic health records using text mining. *Journal of Biomedical Informatics*, 58, S203-S210. <https://doi.org/10.1016/j.jbi.2015.08.003>
- Karystianis, G., Adily, A., Schofield, P., Knight, L., Galdon, C., Greenberg, D., Jorm, L., Nenadic, G., & Butler, T. (2018). Automatic extraction of mental health disorders from domestic violence police narratives: Text mining study. *Journal of Medical Internet Research*, 20(9), 1-16. <https://doi.org/10.2196/11548>
- Karystianis, G., Adily, A., Schofield, P. W., Greenberg, D., Jorm, L., Nenadic, G., & Butler, T. (2019). Automated analysis of domestic violence police reports to explore abuse types and victim injuries: Text mining study. *Journal of Medical Internet Research*, 21(3), 1-12. <https://doi.org/10.2196/13067>
- Koster, N. S. N. (2017). Victims' perceptions of the police response as a predictor of victim cooperation in the Netherlands: a prospective analysis. *Psychology, Crime and Law*, 23(3), 201-220. <https://doi.org/10.1080/1068316X.2016.1239098>
- Laar, S. A. Van, Gombert-handoko, K. B., Guchelaar, H., & Zwaveling, J. (2020). An electronic health record text mining tool to collect real-world drug treatment outcomes: A validation study in patients with metastatic renal cell carcinoma. *Clinical Pharmacology & Therapeutics*, 108(3), 644-652. <https://doi.org/10.1002/cpt.1966>
- Lehmann, R. J. B., Hanson, R. K., Babchishin, K. M., Gallasch-Nemitz, F., Biedermann, J., & Dahle, K. P. (2013). Interpreting multiple risk scales for sex offenders: Evidence for averaging. *Psychological Assessment*, 25(3), 1019-1024. <https://doi.org/10.1037/a0033098>
- Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner

- violence risk assessment instruments. *Journal of Interpersonal Violence*, 28(7), 1537-1558. <https://doi.org/10.1177/0886260512468250>
- Olver, M. E., & Jung, S. (2017). Incremental prediction of intimate partner violence: An examination of three risk measures. *Law and Human Behavior*, 41(5), 440-453. <https://doi.org/10.1037/lhb0000251>
- Poelman, J., Elzinga, P., Viane, S., & Dedene, G. (2010). Curbing domestic violence: Instantiating C-K theory with formal concept analysis and emergent self-organizing maps. *Intelligent Systems in Accounting, Finance and Management*, 17, 161-176. <https://doi.org/10.1002/isaf>
- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, 17(2), 156-167. <https://doi.org/10.1037/1040-3590.17.2.156>
- Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, 31, 8-22. <https://doi.org/10.1002/bsl>
- van der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, 47, 100-116. <https://doi.org/10.1016/j.avb.2019.03.008>

Bijlagen

Bijlage 1: Geanalyseerde delicten

Zeer ernstige delicten

B70	STRAATROOF
B72	OVERVAL IN WONING
B73	OVERVAL OP OVERIGE OBJECTEN
B74	OVERVAL OP GELD- EN WAARDETRANSPORT
F50	DISCRIMINATIE
F520	OPENBARE SCHENNIS DER EERBAARHEID
F521	VERKRACHTING
F522	AANRANDING
F523	OVERIGE ZEDENMISDRIJVEN
F525	PORNOGRAFIE
F526	INCEST / AFHANKELIJKHEID / WILSONBEKWAME
F527	SEKSUEEL MISBRUIK KINDEREN (GEEN INCEST)
F5291	KINDERPORNOGRAFIE
F5295	SEXTING
F5296	GROOMING
F533	STALKING
F540	DOODSLAG / MOORD
F542	OVERIGE MISDRIJVEN TEGEN HET LEVEN
F551	ZWARE MISHANDELING
F561	MENSENHANDEL SEKSUELE UITBUITING
F562	MENSENHANDEL ARBEIDSUITBUITING
F565	MENSENHANDEL OVERIGE VORMEN VAN UITBUITING

Ernstige delicten

A20	GEKWAL. DIEFSTAL IN / UIT WONING
A30	DIEFSTAL IN / UIT WONING (NIET GEKWAL.)
A82	CHANTAGE / AFPERSING
B10	DIEFSTAL MET GEWELD UIT / VANAF PERSONENAUTO
B12	DIEFSTAL MET GEWELD UIT / VANAF ANDER VERVOERMID
B20	GEKWAL. DIEFSTAL MET GEWELD IN / UIT WONING
B21	GEKWAL. DIEFSTAL MET GEWELD IN / UIT BOX / GARAGE / SCHUUR
B22	GEKWAL. DIEFSTAL MET GEWELD IN / UIT WINKEL
B23	GEKWAL. DIEFSTAL MET GEWELD IN / UIT BEDRIJF / KANTOOR
B24	GEKWAL. DIEFSTAL MET GEWELD IN / UIT SPORTCOMPLEX
B25	GEKWAL. DIEFSTAL MET GEWELD IN / UIT HOTEL / PENSION
B27	GEKWAL. DIEFSTAL MET GEWELD IN / UIT ANDERE GEBOUWEN
B30	DIEFSTAL MET GEWELD IN / UIT WONING (NIET GEKWAL.)
B32	DIEFSTAL MET GEWELD IN / UIT BEDRIJF / KANTOOR (NIET GEKWAL.)
B34	DIEFSTAL MET GEWELD IN / UIT BOX / GARAGE / SCHUUR (NIET GEKWAL.)
B50	WINKELDIEFSTAL MET GEWELD
B60	DIEFSTAL MET GEWELD PERSONENAUTO
B61	DIEFSTAL MET GEWELD MOTOR
B62	DIEFSTAL MET GEWELD FIETS
B63	DIEFSTAL MET GEWELD BROMFIETS / SNORFIETS
B95	OVERIGE DIEFSTALLEN MET GEWELD

F12 OPENLIJKE GEWELDPLEGING TEGEN PERSONEN
F530 BEDREIGING
F531 OVERIGE MISDRIJVEN TEGEN DE PERSOONLIJKE VRIJHEID
F532 GIJZELING/ONTVOERING

Overige delicten

A10 DIEFSTAL UIT/VANAF PERSONENAUTO
A11 DIEFSTAL UIT/VANAF VAARTUIG
A12 DIEFSTAL UIT/VANAF ANDERE VERVOERMIDDELEN
A21 GEKWAL. DIEFSTAL IN/UIT BOX/GARAGE/SCHUUR
A22 GEKWAL. DIEFSTAL IN/UIT WINKEL
A23 GEKWAL. DIEFSTAL IN/UIT BEDRIJF/KANTOOR
A24 GEKWAL. DIEFSTAL IN/UIT SPORTCOMPLEX
A25 GEKWAL. DIEFSTAL IN/UIT HOTEL/PENSION
A26 GEKWAL. DIEFSTAL IN/UIT SCHOOL
A27 GEKWAL. DIEFSTAL IN/UIT ANDERE GEBOUWEN
A31 DIEFSTAL IN/UIT SCHOOL (NIET GEKWAL.)
A32 DIEFSTAL IN/UIT BEDRIJF/KANTOOR (NIET GEKWAL.)
A33 DIEFSTAL IN/UIT HOTEL/PENSION (NIET GEKWAL.)
A34 DIEFSTAL IN/UIT BOX/GARAGE/SCHUUR/ERF (NIET GEKWAL.)
A35 DIEFSTAL IN/UIT SPORTCOMPLEX (NIET GEKWAL.)
A36 DIEFSTAL IN/UIT ANDERE GEBOUWEN (NIET GEKWAL.)
A40 ZAKKENROLLERIJ/TASSENROLLERIJ
A50 WINKELDIEFSTAL
A60 DIEFSTAL DIER
A70 DIEFSTAL PERSONENAUTO
A71 DIEFSTAL MOTOR
A72 DIEFSTAL FIETS
A73 DIEFSTAL BROMFIETS/SNORFIETS
A74 DIEFSTAL ANDER VERVOERMIDDEL
A75 DIEFSTAL VAARTUIG
A76 DIEFSTAL VRACHTAUTO/BESTELAUTO
A80 VERDUISTERING (EVT. IN DIENSTBETREKKING)
A81 HELING
A90 OVERIGE (EENVOUDIGE) DIEFSTAL
A95 OVERIGE GEKWAL. DIEFSTAL
C10 VERNIELING VAN/AAN AUTO
C20 VERNIELING VAN/AAN OPENBAAR VERVOER/ABRI
C30 VERNIELING VAN/AAN OPENBAAR GEBOUW
C40 VERNIELING OVERIGE OBJECTEN
C50 VANDALISME/BALDADIGHEID
D10 VERKEERSONGEVAL MET UITSLUITEND MATERIELE SCHADE
D11 VERKEERSONGEVAL MET LETSEL
D12 VERKEERSONGEVAL MET DODELIJKE AFLOOP
D13 VERLATEN PLAATS NA VERKEERSONGEVAL
D50 JOYRIDING
D51 VALS KENTEKEN / VALSE KENTEKENPLATEN
D52 OVERIG VERKEERSMISDRIJF
D70 AGRRESSIEF/ONVEILIG RIJGEDRAG
F000 OVERTREDING OVERIG (WETBOEK VAN STRAFRECHT)
F001 OVERTREDING OVERIG
F0010 OVERTREDING APV
F010 OPENBARE DRONKENSCHAP

F10 OVERIGE DELICTEN OPENBARE ORDE
F11 OPENLIJKE GEWELDPLEGING TEGEN GOEDEREN
F13 BRANDSTICHTING
F15 HUISVREDEBREUK
F16 LOKAALVREDEBREUK
F17 WEDERSPANNIGHEID (VERZET)
F18 NIET VOLDOEN AAN BEVEL/VORDERING
F19 OVERIGE MISDRIJVEN TEGEN HET OPENBAAR GEZAG
F30 VALSE IDENTITEIT OPGEVEN
F40 BEZIT HARDDRUGS (LIJST I)
F41 BEZIT SOFTDRUGS (LIJST II)
F42 HANDEL E.D. HARDDRUGS (LIJST I)
F45 VERVAARDIGEN SOFTDRUGS (LIJST II)
F47 OVERIGE DRUGSDELICTEN
F51 BELEDIGING
F550 EENVOUDIGE MISHANDELING
F611 VALS GELD UITGEVEN
F614 FRAUDE MET BETAALPRODUCTEN
F616 IE-FRAUDE/NAMAAKGOEDEREN
F617 IDENTITEITSFRAUDE
F620 OVERIGE HORIZONTALE FRAUDE
F621 UITKERINGSFRAUDE
F622 VERZEKERINGSFRAUDE OF ASSURANTIEFRAUDE
F624 VALSE AANGIFTE
F625 FAILLISSEMENTSFRAUDE
F631 KREDIET-, HYPOTHEEK- EN DEPOTFRAUDE
F632 ACQUISITIEFRAUDE
F633 VASTGOEDFRAUDE
F634 FRAUDE MET KILOMETERTELLERS
F635 FRAUDE IN DE ZORG
F636 FRAUDE MET ONLINE HANDEL
F637 VOORSCHOTFRAUDE
F638 TELECOMFRAUDE
F639 BELEGGINGSFRAUDE
F649 OVERIGE VERTICALE FRAUDE
F70 BEZIT VUURWAPENS
F72 BEZIT OVERIGE WAPENS
F90 CYBERCRIME
F93 MISDRIJVEN ANDERS
F95 OVERTREDING HUISVERBOD
L41 SCHEEPSONGEVAL BPR
M011 OP/IN BODEM BRENGEN AFVALSTOFFEN
M051 GEZONDHEID EN WELZIJN DIEREN EN DIERENVERVOER / WET DIEREN
M103 VISSERIJWET
M134 BEZITTEN/VERVAARDIGEN/VOORHANDEN HEBBEN/AFLEVEREN VUURWERK
M22 GELUIDSHINDER OVERIG

Bijlage 2: Datapreparatiestrategieën

Count-vector

Bij de count-vector procedure wordt gekeken naar hoe vaak elk woord in een tekst voorkomt, hetgeen leidt tot een overzicht van tellingen. Elk element in dit overzicht vertegenwoordigt dus een woordfrequentie. Om het aantal woorden dat in aanmerking komt voor de analyses te beperken zijn alleen de 10.000 meest voorkomende woorden vertegenwoordigd in het overzicht. Uit deze 10.000 woorden is vervolgens een nieuwe selectie gemaakt om de efficiëntie van het proces te verbeteren. Bij deze selectie zijn de 2.000 woorden met de sterkste samenhang met de afhankelijke variabele geselecteerd als features voor de modelbouw. Deze woorden zijn geselecteerd met een chi-kwadraat selectiealgoritme.

TF/IDF

De TF/IDF (term frequency / inverse document frequency) procedure is verwant aan de count-vector procedure. Echter worden in de TF/IDF procedure de woordfrequenties gewogen met hun “inverse document frequency”. Normaliter is dit een maat voor hoe vaak een woord voorkomt in verschillende documenten. Alle verzamelde tekstvelden van een persoon zijn samengevoegd in een overkoepeld document. In dit onderzoek heeft ieder persoon één overkoepelend document, vandaar dat de IDF hier een maat is voor hoe vaak een term voorkomt bij verschillende personen. Aangezien naar het inverse gekeken wordt zullen woorden die veelvuldig voorkomen bij veel personen uiteindelijk een lager gewicht krijgen dan zeldzame woorden. Om de efficiëntie van het proces te verbeteren zijn ook hier de 2.000 woorden met de sterkste samenhang met de afhankelijke variabele geselecteerd als features voor de modelbouw. Deze termen zijn eveneens geselecteerd met het chi-kwadraat selectiealgoritme.

Bijlage 3: Analysemethoden ongestructureerde gegevens

Bij de start van het onderzoek werden drie analysemethoden gebruikt om de ongestructureerde gegevens te analyseren: logistische regressie, decision tree en random forest modellen. Dit zijn drie populaire analysemethoden die vaak gebruikt worden bij de ontwikkeling en validering van risicotaxatie instrumenten. De resultaten van de logistische regressiemodellen zijn uiteindelijk niet uitgewerkt en gepresenteerd, omdat de logistische regressiemodellen namelijk tot minder nauwkeurige modellen leidden dan de overige twee analysemethoden. Vandaar dat de resultaten voor de modellen op basis van ongestructureerde gegevens alleen gebaseerd zijn op decision trees en random forests. Deze twee machine learning technieken hebben de laatste jaren aan populariteit gewonnen op het gebied van risicotaxatie en zijn geschikt voor het analyseren van relatief zeldzame gebeurtenissen als herhaald slachtofferschap (Berk & Sorenson, 2020; Hegelich, 2016). Daarnaast zijn ze in staat om subpopulaties en interacties tussen variabelen te identificeren (Brand et al., 2021).

Decision tree

Een decision tree, of beslisboom, splitst op basis van kenmerken die het sterkst samenhangen met de afhankelijke variabele (in dit geval wel/geen slachtofferschap in de toetsingsperiode) de dataset herhaaldelijk in verschillende subgroepen. Dit proces verloopt in een aantal stappen. In de eerste stap wordt de dataset gesplitst op basis van het kenmerk dat het sterkst samenhangt met de afhankelijke variabele. Dit resulteert in twee of meerdere subgroepen (ook wel takken genoemd). Vervolgens wordt elke subgroep opnieuw gesplitst op basis van het kenmerk dat voor die subgroep het sterkst samenhangt met de afhankelijke variabele. Dit proces gaat door totdat in de subgroepen geen variabelen meer significant samenhangen met de afhankelijke variabele of totdat de groepen een minimale omvang hebben bereikt of totdat het maximaal aantal vertakkingen dat van tevoren is ingesteld bereikt is. Zo ontstaat een aantal risicogroepen van slachtoffers met dezelfde soort risicofactoren en eenzelfde risico op herhaald slachtofferschap.

Een belangrijk voordeel voor het gebruik van beslisbomen in de politiepraktijk is de interpreteerbaarheid. Beslisbomen kunnen snel inzichtelijk maken welke invloed voorspellers hebben op een afhankelijke variabele. Een nadeel is het gebrek aan robuustheid. Kleine veranderingen in de data kunnen leiden tot significante veranderingen in het model en dus de inschattingen. Hierdoor kan de nauwkeurigheid van een decision tree model sterk verschillen tussen datasets. Om dit punt (deels) te ondervangen werd het ontwikkelde decision tree model ook op een andere dataset getest.

Random Forest

Om de beperking op het gebied van de robuustheid van een decision tree model te limiteren, kan een random forest model worden gebruikt. Een random forest is een verzameling van decision trees. Met deze methode wordt een bepaald aantal decision trees ontwikkeld. Elke boom wordt ontwikkeld op basis van een random steekproef uit de ontwikkelset. Bij het uitvoeren van deze methode kunnen criteria worden gesteld aan het aantal bomen dat ontwikkeld wordt en de diepte (aantal vertakkingen) van deze bomen. Het proces stopt dan ook wanneer aan deze criteria voldaan is. In dit onderzoek werden 25 bomen ontwikkeld met een diepte van tien vertakkingen. Nadat deze subbomen zijn ontwikkeld, worden ze samengevoegd tot een model op basis waarvan een risicoscore wordt berekend. Het voordeel van een random forest is dat het resulteert in een robuust model, aangezien het meerdere beslisbomen met elkaar combineert. Een nadeel van deze methode is dat het uiteindelijke model minder goed te interpreteren is dan een decision tree. Het is namelijk lastig om visueel uiteen te zetten hoe voorspellers met de afhankelijke variabele samenhangen.

Bijlage 4: Combinatiemethoden

Eerdere wetenschappelijke studies presenteren verschillende methoden om risicoscores te combineren en geven aan dat het nog onduidelijk is welke combinatiemethode de voorkeur heeft (Lehmann et al., 2013; Olver & Jung, 2017; Seto, 2005). Daarom zijn in dit onderzoek verschillende combinatiemethoden onderzocht. De eerste twee combinatiemethoden waren respectievelijk *hoogste leidend* en *laagste leidend*. In het eerste geval werd per slachtoffer de hoogste risicoscore gezien als de nieuwe risicoscore, en in het tweede geval de laagste. De derde combinatiemethode hield in dat de nieuwe risicoscore werd berekend op basis van het *gemiddelde* van de twee risicoscores. Tot slot zijn twee combinatiemethoden toegepast die gebruik maken van *logistische regressieanalyses*.

De combinatie van de twee risicoscores

Ten eerste werden de twee risicoscores als onafhankelijke variabelen opgenomen in een logistische regressieanalyse om het risico op herhaald slachtofferschap in te schatten. Deze logistische regressieanalyse leverde gewichten op waarmee een nieuwe risicoscore werd berekend. Ten tweede werden de twee risicoscores én de interactieterm van deze twee scores opgenomen als onafhankelijke variabelen in een logistische regressieanalyse. Een interactieterm houdt rekening met de mogelijkheid dat de ene risicoscore het effect van de andere risicoscore kan versterken of afzwakken. Uit deze logistische regressieanalyse volgden ook weer gewichten die gebruikt werden om een nieuwe (combinatie)risicoscore te berekenen.

balancing Security and Mobility



EUROPESE UNIE

het Fonds voor interne veiligheid
van de Europese Unie

Colofon

Praktikon BV is een zelfstandige en onafhankelijke organisatie voor onderzoek en ontwikkeling in de jeugdzorg en onderwijsgerelateerde zorg. Praktikon wil instellingen voor jeugdzorg en scholen voor speciaal onderwijs helpen hun werk bij de aanpak van opgroei- en opvoedingsproblemen zo goed mogelijk te doen. Praktikon hoopt hiermee een bijdrage te leveren aan het verbeteren van de kwaliteit van het werk en aan het verantwoorden hiervan. Vanuit deze doelstellingen wordt samengewerkt met de Radboud Universiteit Nijmegen en de Stichting Entréa, een organisatie voor jeugdzorg, onderwijs en onderzoek in Gelderland.

Over de auteurs

Roos Geurts, MSc en Niels Raaijmakers, MSc zijn als junior onderzoekers verbonden aan Praktikon

Dr. Marc Delsing is als senior onderzoeker verbonden aan Praktikon

Prof. dr. Ron Scholte is directeur van Praktikon en als hoogleraar verbonden aan de Faculteit Sociale Wetenschappen van de Radboud Universiteit Nijmegen

Projectleiding

Jacqueline Wientjes is projectleider van ProVict

Dit onderzoek is uitgevoerd in opdracht van de Politie.