

## ARTIKELEN

# Met datascience op zoek naar indicatoren van georganiseerde criminaliteit en ondermijning\*

Patricia Prüfer & Emile Kolthoff

## 1 Inleiding: datasciencetoepassingen voor datagedreven beleid

De toenemende digitalisering en datafication waardoor data steeds belangrijker worden in onze maatschappij leiden tot veranderingen in de manier waarop (overheids)instellingen werken en beslissingen genomen worden. Het analyseren van zogeheten big data door middel van datasciencetechnieken kan helpen om nieuwe vormen van criminaliteit te herkennen én meer zicht te krijgen op de daadwerkelijke omvang van al bekende verschijningsvormen van criminaliteit (het zogenoemde *dark number*). Datascience helpt tot nieuwe inzichten te komen door op een innovatieve manier grote hoeveelheden data uit uiteenlopende bronnen te analyseren en interpreteren. De kunst van datascience is het transformeren van beschikbare data in waardevolle acties.

In Nederland zijn al verschillende datasciencetoepassingen voor datagedreven (veiligheids)beleid waar te nemen.<sup>1</sup> In deze bijdrage wordt een onderzoek beschreven waarbij gebruik werd gemaakt van *machine learning*-methoden. *Machine learning* houdt zich bezig met technieken waarmee computers kunnen leren, zonder dat ze er speciaal voor geprogrammeerd zijn.<sup>2</sup> De vraag die centraal stond in het onderzoek was of er indicatoren kunnen worden gevonden waarmee het voor bestuurders inzichtelijk wordt in welk stadium een bedrijventerrein verkeert in de ontwikkeling of aanwezigheid van (georganiseerde) criminaliteit en ondermijning.<sup>3,4</sup> Doel was om verder te exploreren of datascience een bijdrage kan leveren aan de bestrijding van criminaliteit door het ontwikkelen van voorspellende indicatoren.

Het onderzoek is gestart met een – cruciale – voorbereidende fase. Om te komen tot goede voorspellende indicatoren voor (georganiseerde) criminaliteit en ondermijning is het belangrijk om eerst te kijken naar bestaande inzichten over potentiële risicofactoren uit ander onderzoek en uit het veld. Tijdens het vooronder-

\* Dr. Patricia Prüfer is groepsleider Data Science bij CentERdata, Tilburg. Prof. dr. Emile Kolthoff is hoogleraar Criminologie aan de Open Universiteit en lector Ondermijning bij Avans Hogeschool.

1 Zie bijv. de website voor dit project: [www.bigdatagemeenten.nl](http://www.bigdatagemeenten.nl).

2 A. Samuel, 'Some Studies in Machine Learning Using the Game of Checkers', *IBM Journal* 1959, 3, p. 535-554.

3 Y. Bolsius e.a., *Indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen*, Den Haag: WODC 2018.

4 Dit onderzoek werd mede mogelijk gemaakt door een financiering vanuit het WODC.

zoek is daarom deskresearch gedaan en heeft een expertmeeting plaatsgevonden. Inzichten uit dit expertpanel zijn vervolgens geverifieerd en met een literatuurstudie uitgebreid. Ten slotte hebben er nog interviews en gesprekken met andere zowel inhoudelijke als datadeskundigen plaatsgevonden om de inzichten uit te breiden en databronnen te identificeren die mogelijkterwijs voor dit onderzoek gebruikt zouden kunnen worden.

We beginnen deze bijdrage in paragraaf 2 met een korte theoretische beschouwing van de verschijningsvormen van georganiseerde criminaliteit en mogelijke indicatoren die daar op kunnen wijzen. Eerst meer algemeen, vervolgens toegespitst op bedrijventerreinen. Na in paragraaf 3 en 4 respectievelijk in te gaan op het conceptuele model voor ons onderzoek, de dataverzameling en de datapreparatie starten we paragraaf 5 met een korte inleiding in datasciencemethoden, aangezien deze voor veel lezers van *PROCES* relatief onbekend zullen zijn. Paragraaf 5 vervolgt met de toegepaste onderzoeksmethodiek en in paragraaf 6 worden de resultaten gepresenteerd. In paragraaf 7 sluiten we deze bijdrage af met een korte reflectie op het onderzoek.

## 2 Vormen van criminaliteit en gelegenheid

Georganiseerde misdaad en ondermijnende criminaliteit kennen vele verschijningsvormen. In het ‘Nationaal dreigingsbeeld 2017 Georganiseerde criminaliteit’ werden zeventien verschijningsvormen van criminaliteit van het predicaat ‘dreiging’ voorzien.<sup>5</sup> Drugsgelateerde criminaliteit, uitbuiting/mensenhandel en vuurwapengerelateerde criminaliteit vallen allen onder het kopje ‘illegale markten’. Dit zijn ook de misdrijven die centraal staan in de Monitor Georganiseerde Criminaliteit, waarbij drugsgelateerde criminaliteit de overhand heeft.<sup>6</sup> Ook de dagelijkse nieuwsoverzichten van de taskforce Zeeland-Brabant betreffen voor het merendeel drugsgelateerde criminaliteit. Het onderhavige onderzoek heeft zich dan ook vooral gericht op drugsgelateerde criminaliteit.

In de criminologie is de volkswijsheid ‘De gelegenheid maakt de dief’ theoretisch verwoord door onder anderen Felson en Clarke<sup>7</sup> in de gelegenheidstheorie, afgeleid van de eerder ontwikkelde *routine activity theory*.<sup>8</sup> De theorie beoogde een verklaring te bieden voor de steeds stijgende criminaliteitscijfers na de Tweede Wereldoorlog. Cohen en Felson verklaarden deze toename door de enorme groei aan potentiële slachtoffers door de toegenomen welvaart, gepaard gaande met een relatieve afname van toezicht. De *routine activity theory* is een variant van de controletheorieën en gaat uit van de vraag waarom personen ‘geen’ criminaliteit

5 Brief minister van Veiligheid en Justitie van 1 juni 2017, kenmerk 2082923.

6 E. Kruisbergen, H.G. van de Bunt & E. Kleemans, *Georganiseerde criminaliteit in Nederland: vierde rapportage op basis van de Monitor Georganiseerde Criminaliteit*, Den Haag: Boom Lemma uitgevers 2012.

7 M. Felson & R. Clarke, ‘Opportunity makes the thief: Practical theory for crime prevention’, in: B. Webb (red.), *Police Research Series* (paper 98), Londen: Home Office 1998.

8 L. Cohen & M. Felson, ‘Social change and crime rate trends: A routine activity’s approach’, *American Sociological Review* 1979, 44, p. 588-608.

plegen. De *routine activity theory* stelt daarbij dat de aandacht gevestigd moet worden op de situatie in plaats van op de dader. Later ontwikkelden Felson en Clarke hun meer uitgesproken gelegenheidstheorie (*opportunity theory*), die op één enkel principe was gebaseerd: verleidelijke en gemakkelijke gelegenheden om criminaliteit te plegen zetten mensen aan om dit ook te doen. Volgens Felson en Clarke wordt de omvang van criminaliteit hierbij bepaald door drie factoren: aantallen potentiële daders, aantallen aantrekkelijke doelwitten, en de mate van toezicht op en de bescherming van deze doelwitten. Sociale en technologische ontwikkelingen noemden zij nadrukkelijk als nieuwe mogelijkheden voor criminaliteit en wij verstaan daar ook uitdrukkelijk de voortschrijding van digitale technologie onder.

Een hoopgevend inzicht uit de gelegenheidstheorie is dat er wel iets kan worden gedaan om gelegenheden in te perken. Om hier iets dieper op in te gaan bespreken we nu kort de beweging van *Crime Prevention Through Environmental Design* (CPTED) en Sociaal Veilig Ontwerpen. De theorie van CPTED is gebaseerd op het idee dat criminaliteit mede het gevolg is van de gelegenheid die de fysieke omgeving daartoe biedt.<sup>9</sup> Als dit correct is, moet het ook mogelijk zijn om de fysieke omgeving zodanig te manipuleren dat criminaliteit minder zal optreden, ook op bedrijventerreinen. Een Nederlands instrument, dat gericht betrekking heeft op het sociaal veilig ontwerpen van de buitenruimte, is de checklist Sociaal Veilig Ontwerpen, voortbouwend op het handboek van Van der Voordt en Van Wegen.<sup>10</sup> De volgende punten vatten de belangrijke thema's samen:<sup>11</sup>

- 1 De aanwezigheid van potentiële daders verhoogt het risico.
- 2 De aanwezigheid van sociale ogen (toezicht) verlaagt het risico.
- 3 Zichtbaarheid verlaagt het risico.
- 4 Betrokkenheid bij en verantwoordelijkheid voor de omgeving verlagen het risico.
- 5 Attractiviteit van de omgeving verlaagt het risico.
- 6 Toegankelijkheid en vluchtwegen. Hier moet een balans worden gevonden tussen het zo min mogelijk toegankelijk zijn van het private domein en het aantal vluchtwegen voor potentiële slachtoffers, dat zo groot mogelijk moet zijn.
- 7 Aantrekkelijkheid van een potentieel doelwit verhoogt het risico.

We vervolgen deze inleiding met de bespreking van een aantal mogelijke indicatoren van criminaliteit op bedrijventerreinen, gebaseerd op een combinatie van literatuurstudie en bevindingen uit de expertmeeting (waarover later meer) en interviews die deel uitmaakten van het onderzoek.

9 E. Kube, *Städtebau, Wohnhausarchitektur und Kriminalität: Prevention statt Reaction*, Heidelberg: Kriminalistik-verlag 1982.

10 D. van der Voordt & H. van Wegen, *Sociaal veilig ontwerpen: checklist ten behoeve van het ontwikkelen en toetsen van (plannen voor) de gebouwde omgeving*, Delft: Technische Universiteit Delft, Publicatieburo Bouwkunde 1990.

11 I. Luten (red.), *Handboek Veilig Ontwerp en Beheer*, Bussum: Thoth 2008.

## 2.1 Ligging

De ligging van een bedrijventerrein is tijdens de expertmeeting genoemd als mogelijke indicator van criminaliteit op een bedrijventerrein. Net zoals voor legale bedrijven zijn voor criminele organisaties de ligging van een terrein en de aan- en afvoerroutes een belangrijke factor. Hierbij kan gedacht worden aan de ontsluiting van het terrein via de weg, het spoor of het water. Ook de afstand tot de grens met het buitenland en de afstand tot vliegvelden zouden belangrijke factoren kunnen zijn.

## 2.2 Soort bedrijven (branches)

Naast de ligging van het bedrijventerrein zou ook gekeken kunnen worden naar het soort bedrijven (de branches) die aanwezig zijn op het terrein. Op basis van de literatuur over ondermijnende criminaliteit, drugshandel en heling weten we dat een aantal branches kwetsbaar is voor misdaad, zoals de autobranche, opslag van spullen, transport en de groente- en fruitgroothandel.<sup>12,13</sup> De aanwezigheid van risicobranches werd tijdens de expertmeeting genoemd als een van de belangrijkste indicatoren. Er werd gesteld dat sommige bedrijven per definitie meer risicovol zijn dan andere. Hierbij werden vooral de autobranche, opslagbedrijven en ambulante handel als voorbeelden genoemd. In het algemeen worden branches waarbij weinig vakkennis vereist is en waarvoor geen vergunningsplicht geldt, als meer risicovol genoemd. Omgekeerd kan het ook zo zijn dat bepaalde typen bedrijven juist de kans verkleinen dat een bedrijventerrein afzakt (zoals bedrijven met een publieksfunctie, die zorgen namelijk voor meer sociale controle).

Bij de Wet bevordering integriteitsbeoordelingen door het openbaar bestuur (Wet BIBOB) zijn de transportbranche, woningbouwcorporaties, coffeshops, bordelen, smart- en growshops en de branches bouw, milieu en ICT aangewezen als sectoren die kwetsbaar zijn voor criminaliteit. In het onderzoek ter inventarisatie van de uitbreiding van de Wet BIBOB is die lijst uitgebreid met de speelautomatenbranche, het importeren van vuurwerk, headshops en vastgoedtransacties met de overheid. Risicobranches zijn echter aan verandering onderhevig. Wanneer er streng toezicht wordt gehouden op een bepaalde branche, is het mogelijk dat criminelen uitwijken naar andere branches waar zij minder in de gaten worden gehouden.<sup>14</sup> Uit onderzoek blijkt dat de vraag of door de georganiseerde criminaliteit wordt geïnvesteerd in bedrijven, wordt beïnvloed door vier factoren.<sup>15</sup>

- 12 H. Ferwerda e.a., *Focus op heling – een onderzoek naar het functioneren van de helingmarkt, het beleid tegen en de gevolgen van heling* (Beke-reeks), Den Haag: WODC 2016.
- 13 P. Tops & E. van der Torre, *Wijknaanpak en ondermijnende criminaliteit*, Den Haag: Boom Lemma uitgevers 2015.
- 14 M. te Pest e.a., *Verwevenheid van onder- en bovenwereld bij georganiseerde criminaliteit. Een overzichtsstudie: aard en oorzaken*, Regionaal Informatie en Expertise Centrum (RIEC) Zuid-West Nederland 2012.
- 15 E. Kleemans, M. Brienens & H. van de Bunt, *Georganiseerde criminaliteit in Nederland. Tweede rapportage op basis van de WODC-monitor*, Den Haag: WODC/Boom Juridische uitgevers 2002, p. 133.

- 1 Witwasmogelijkheden: het is eenvoudiger om een legale oorsprong voor criminele winsten voor te wenden als men de beschikking heeft over bedrijven waarin ook legale activiteiten plaatsvinden.
- 2 Logistieke mogelijkheden: wanneer men investeert in bedrijven die nauw aansluiten bij de illegale hoofdactiviteit, hebben zij meer controle over het hele logistieke proces en is men minder afhankelijk van derden.
- 3 Bekendheid met een bepaalde branche: daders investeren liever in een bekende economische activiteit of marktsector waarvan zij weten wat er omgaat en wat de mogelijkheden tot witwassen zijn.
- 4 Etniciteit: de etniciteit van de verdachten is in hoge mate bepalend voor de landen waar misdaadgeld wordt geïnvesteerd.

Vanwege veranderende gelegenheidsstructuren en het verplaatsingseffect is een uitsluitende lijst met branches niet te geven. Wel kunnen de eigenschappen van een branche worden genoemd die voor criminelen een gelegenheid bieden tot het succesvol witwassen van crimineel geld. Dit zijn volgens Te Pest e.a.:<sup>16</sup>

- 1 branches waarin de waarde van 'producten' moeilijk objectief vast te stellen is;
- 2 branches waarin grote sommen (contant) geld omgaan;
- 3 branches waar de omzet gemakkelijk kan worden gemanipuleerd;
- 4 branches waar weinig tot geen toetredingseisen gelden.

Een aantal branches loopt door organisatiekenmerken, maar ook door sociale factoren op individueel niveau (zoals het lage opleidingsniveau of een zorgelijke financiële positie van in dienst genomen personeel), extra risico om al dan niet bewust betrokken te raken bij criminele activiteiten. Risicobranches die in verband met bedrijventerreinen in de hiervoor genoemde onderzoeken steeds weer worden genoemd, zijn de handel in kunst, antiek, dieren, sieraden, afval, voertuigen, prostitutie, de reinigingsbranche, belwinkels, avondkappers en massagesalons, de speelautomatenbranche en gokhallen, de transportsector, distributieknooppunten, de uitzendbranche, de land- en tuinbouwsector, en growshoptoeleveranciers en dekmantelbedrijven die speciaal zijn opgericht of aangekocht om criminele activiteiten af te schermen.

### 2.3 Fysieke omgeving en verloop

Een ander belangrijk kenmerk van een bedrijventerrein waar naar kan worden gekeken is de fysieke omgeving. In lijn met de in de inleiding besproken gelegenheidstheorie kwam de fysieke omgeving ook tijdens de expertmeeting naar voren als een belangrijke indicator voor criminaliteit op bedrijventerreinen. Ten eerste werden een rommelig terrein, afvallozing en onderhoud genoemd, analoog aan de *broken windows theory*.<sup>17</sup> Deze theorie stelt dat omgevingen die reeds vervuild zijn meer vuil aantrekken, en verloederde terreinen waar zich veel criminelen bevinden zullen dan ook nieuwe criminelen aantrekken. Ten tweede kwam leegstand

16 M. te Pest e.a. 2012, p. 60.

17 G. Kelling & J.Q. Wilson, 'Broken windows: The police and neighborhood safety', *The Atlantic Monthly* 1982, p. 29-38.

zowel in de expertmeeting als in interviews naar voren als een belangrijke indicator. Als derde indicator met betrekking tot de fysieke omgeving werd veroudering genoemd. Volgens Beekmans<sup>18</sup> lopen de gevolgen van veroudering erg uiteen: van fysieke aftakeling tot sociaal onwenselijke verschijnselen zoals een toename van criminaliteit en de bijbehorende onveiligheidsgevoelens. Beekmans inventariseerde een aantal criteria voor veroudering van bedrijventerreinen, ontleend aan cijfers van het Planbureau voor de Leefomgeving. Ook hierin komen achterstallig onderhoud en leegstand naar voren als belangrijkste criteria.

Het verloop van en binnen bedrijven zou iets over de kwaliteit van een bedrijventerrein kunnen zeggen. Verloop zou het gevolg kunnen zijn van criminelen die hun misdaden willen verhullen. Hierbij kan aan de ene kant gedacht worden aan het verloop van bedrijven op een bedrijventerrein: wisselen de bedrijven die op een terrein zitten snel of zitten dezelfde bedrijven al jaren op het terrein? Daarnaast kan verloop betrekking hebben op een specifiek bedrijf: wisselt een bedrijf vaak van eigenaar?

#### 2.4 Onroerend goed

Wanneer het met een buurt beter gaat, is dat terug te zien aan een stijging van de waarde van het vastgoed.<sup>19</sup> Omdat bedrijventerreinen geen bewoners met sociaal-economische kenmerken kennen, zijn veel van de gebruikte indicatoren bij het proces van gentrification<sup>20</sup> niet bruikbaar voor onderzoek naar veroudering van bedrijventerreinen. Dat geldt echter niet voor de waarde van vastgoed. Veroudering en verloederen doen de waarde van onroerend goed op bedrijventerreinen afnemen. Daar staat tegenover dat bepaalde indicatoren die een bedrijventerrein aantrekkelijk maken, zoals bereikbaarheid en specialisatie (bijvoorbeeld logistiek), los van verouderingsprocessen de waarde van onroerend goed hoog houden. De waarde van onroerend goed op het bedrijventerrein zou dus een indicator kunnen zijn voor criminaliteit en ondermijning. In de interviews met inhoudsdeskundigen kwam naar voren dat een (significante) afwijking tussen marktwaarde en verkoopwaarde voor vastgoed een indicator zou kunnen zijn. Hierbij werd wel opgemerkt dat de WOZ-waarde bij niet-woningen vaak niet overeenkomt met de echte marktwaarde.

#### 2.5 Toezicht, handhaving en samenhang

Een indicator die genoemd werd tijdens de expertmeeting is de hoeveelheid toezicht en handhaving op het terrein en de mate van samenhang. Een belangrijk voorbeeld zijn de verloederde bedrijventerreinen, waar de overheid in de loop der jaren de teugels heeft laten vieren door langdurig weg te kijken. De problemen beginnen vaak klein, maar doordat de overheid weinig interesse toont of de regels niet handhaaft worden de problemen steeds groter en groeien dergelijke ter-

18 J. Beekmans, 'Verouderingsprocessen op bedrijventerreinen', in: E. van Krabben, C.-J. Pen & F. de Feijter (red.), *De Markt voor Bedrijventerreinen: Uitkomsten van Onderzoek en Beleid*, Den Haag: Platform31 2015, p. 58.

19 R. Walks & R. Maaranen, 'Gentrification, social mix, and social polarization: Testing the linkages in large Canadian cities', *Urban Geography* 2008, 4, p. 293-326, m.n. p. 296.

20 E. Kolthoff, *Basisboek criminologie*, Den Haag: Boom criminologie 2016, p. 153.

reinen uit tot ware vrijplaatsen. Hoe kan de overheid in deze kwetsbare gebieden haar gezag herstellen? Dat kan volgens Lam, Van der Wal en Kop mede door strikte repressie en handhaving. Door het stellen van duidelijke grenzen kunnen maatschappelijke normen en waarden in ere worden hersteld. Handhaving, ook op kleine afwijkingen, is hierbij noodzakelijk.<sup>21</sup>

Of de ondernemers verenigd zijn in een ondernemers- of eigenarenvereniging zou een indicator kunnen zijn. Dit zou namelijk kunnen betekenen dat er sprake is van meer samenhang, wat ertoe zou kunnen leiden dat sociale normen meer gelden en gehandhaafd worden. Als mogelijke verdieping zou gekeken kunnen worden welke ondernemers wel en welke geen lid zijn van de ondernemersvereniging. Op dezelfde manier zou gekeken kunnen worden naar lidmaatschap van brancheverenigingen.

Een andere mogelijke indicator is de aanwezigheid van parkmanagement op het bedrijventerrein. Parkmanagement is een manier om de kwaliteit van bedrijventerreinen in stand te houden of te verbeteren. Gebruikelijke parkmanagementactiviteiten zijn: beheer en onderhoud van de openbare ruimte, bewegwijzering, collectieve beveiliging, gezamenlijke afvalinzameling, en verwerking en gezamenlijke inkoop van onderzoek en advies. Op nieuwe bedrijventerreinen is deelname vaak verplicht, op bestaande is het meestal een zaak van de ondernemersvereniging.

Naast parkmanagement kan ook gekeken worden naar de aanwezigheid van een Keurmerk Veilig Ondernemen voor bedrijventerreinen (KVO-B).

## 2.6 Aantal meldingen en mutaties bij politie, gemeente en Meld Misdaad Anoniem

Als er veel meldingen of incidenten met betrekking tot een bedrijventerrein geregistreerd staan bij de politie, kan dit een signaal zijn dat daar criminaliteit voorkomt. Aan de andere kant zou de afwezigheid van meldingen en aangiftes kunnen betekenen dat mensen bang zijn of de moed hebben verloren dat er iets aan de problemen wordt gedaan. Als het lage aantal meldingen het gevolg is van een lage meldingsbereidheid doordat mensen bedreigd worden door criminelen en bang zijn voor represailles, zou juist de afwezigheid van aangiftes/meldingen als indicator voor georganiseerde criminaliteit kunnen gelden. Er kan dus geen eenduidige conclusie getrokken worden met betrekking tot de relatie tussen het aantal meldingen bij de politie en de kans op criminaliteit. Daarnaast is er bij de gemeente veel informatie bekend over situaties die zich in het verleden hebben voorgedaan op het bedrijventerrein, zoals het aantal hennepuimingen of het aantal panden dat is gesloten (Damocles-sluitingen).<sup>22</sup> Naast de meldingen bij de politie is ook Meld Misdaad Anoniem (M.) een interessante databron. M. is een onafhankelijk meldpunt waar men anoniem informatie kan geven over criminaliteit en misdaad. Naast het absolute aantal meldingen bij M. zou ook het aantal meldingen in ver-

21 J. Lam, R. van der Wal & N. Kop, *Sluipend gif. Een onderzoek naar ondermijnende criminaliteit*, Den Haag: Boom criminologie 2018, p. 214.

22 Met het aantal Damocles-sluitingen refereren we naar het aantal panden dat gesloten is op basis van de Wet Damocles. Art. 13b van de Opiumwet (ook wel de Wet Damocles genoemd) biedt burgemeesters de mogelijkheid om bestuursdwang toe te passen en panden te sluiten als daar sprake is van drugshandel.

houding tot de meldingen bij de politie interessant kunnen zijn (zijn er bijvoorbeeld weinig meldingen bij de politie, maar wordt er wel veel anoniem gemeld).

### 3 Conceptueel model en dataverzameling

Op basis van het vooronderzoek is een conceptueel model opgesteld met mogelijk nuttige indicatoren om georganiseerde criminaliteit en ondermijning te voorspellen. Bij de dataverzameling vanuit verschillende bronnen waren we afhankelijk van de beschikbaarheid en de verkrijgbaarheid van de data. Kenmerken van bedrijventerreinen (zoals de aanwezigheid van parkmanagement en de ontsluiting van het terrein) waren beschikbaar in het Integraal Bedrijventerreinen Informatie Systeem (IBIS). Daarnaast hebben we het aantal meldingen bij de politie en bij M. ontvangen. Of een bedrijventerrein een Keurmerk Veilig Ondernemen heeft, was online in te zien. De gemeente Tilburg heeft het aantal meldingen bij het Centraal Meldpunt (bijvoorbeeld over drugsoverlast), het aantal hennepuimingen, het aantal Damocles-sluitingen en data met betrekking tot de onroerende zaken op de bedrijventerreinen geleverd. Daarnaast zijn data uit de Basisregistratie adressen en gebouwen (BAG) en het bedrijvenregister (waaraan de registratie bij de Kamer van Koophandel (KvK) ten grondslag ligt) ter beschikking gesteld. Tot slot heeft netbeheerder Enexis data geleverd over de aanwezigheid en het weigeren van slimme meters, de capaciteit van elektriciteitsaansluitingen, en historische gevallen van fraude. Dit resulteerde in een uitgebreide dataset met mogelijke indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen.

In figuur 1 is een globale weergave van het conceptueel model te vinden. Dit model vormt de basis voor de analyses die in de volgende paragraaf worden beschreven.

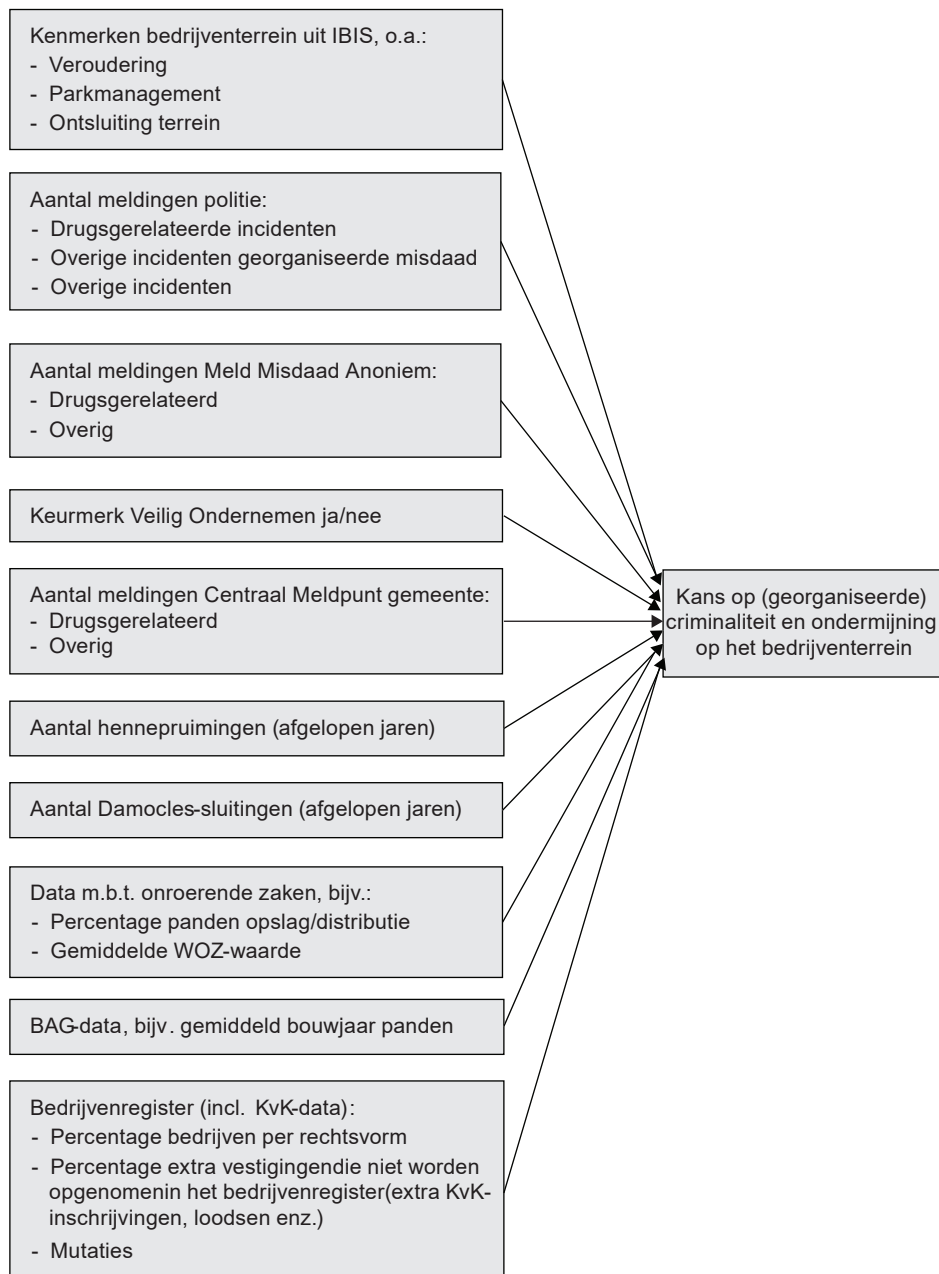
### 4 Datapreparatie en -koppeling

De datapreparatie is een zeer belangrijk onderdeel om tot betrouwbare resultaten te komen. Na de dataverzameling moesten de verschillende bestanden gekoppeld worden op het niveau van bedrijventerrein. Waar dit niet mogelijk was, is een alternatief gezocht, bijvoorbeeld door op basis van de postcodes te achterhalen op welk bedrijventerrein de data betrekking hadden. De meest recente data zijn opgevraagd en vaak zijn data van meerdere jaren opgevraagd (met als richtlijn de jaren 2014 tot en met 2018, indien beschikbaar), zodat deze later nog geaggregeerd konden worden als de data te summier bleken te zijn.

De uiteindelijke steekproef bestond uit dertig bedrijventerreinen in de gemeente Tilburg. Voor deze bedrijventerreinen is een grote hoeveelheid data verzameld met betrekking tot mogelijke risicofactoren. Een aantal databronnen is aangeleverd op het niveau van bedrijventerrein. Dat wil zeggen: er was één observatie per bedrijventerrein. Er waren echter ook databronnen die op een ander aggregatieniveau zijn aangeleverd (bijvoorbeeld postcode 6-niveau) en die dus door de onder-



Met datascience op zoek naar indicatoren van georganiseerde criminaliteit en ondermijning

**Figuur 1** *Conceptueel model*

zoekers geaggregeerd moesten worden tot bedrijventerreinniveau. Daarnaast moesten de verschillende databronnen aan elkaar gekoppeld worden.

De meeste variabelen die aantallen van een bedrijventerrein weergeven, bijvoorbeeld aantal werkend personeel, aantal vestigingen, oppervlakte nog uitgeefbaar of aantal Centraal Meldpunt-meldingen zijn genormaliseerd met de netto oppervlakte van elk bedrijventerrein. Daarnaast zijn de variabelen met een MinMaxScaler herschaald. De keuze voor een MinMaxScaler in plaats van andere schalingsmethoden (zoals *mean normalization*) is empirisch bepaald. We hebben hier niet te maken met uitschieters in de data, dus MinMaxScaler was in dit geval een goede methode.

De grootste problemen bij dit onderzoek waren het kleine aantal beschikbare observaties (30 bedrijventerreinen) en de ongebalanceerdheid van de gelabelde data (8 laag, 15 gemiddeld, 7 hoog). Om een gebalanceerde dataset te verkrijgen kunnen oversamplingtechnieken worden toegepast. Wij gebruikten de Synthetic Minority Oversampling Technique (SMOTE)<sup>23</sup> om onze data te oversamplen. In ons geval oversamplen we de data van 30 naar 45 (naar 15 van elke klasse).

## 5 Onderzoeksmethodiek

### 5.1 Inleiding datasciencemethoden

Datascience helpt tot nieuwe inzichten te komen door op een innovatieve manier grote hoeveelheden data uit uiteenlopende bronnen te analyseren en interpreteren. De kunst van datascience is het transformeren van data in waardevolle acties waarbij er sprake is van een holistische benadering: datascientists verzamelen data die steeds meer ‘in het wild’ te vinden zijn, gieten deze in een bruikbare vorm en presenteren zo het verhaal dat de data vertellen. Datascience is gericht op het doen van voorspellingen om zo aanbevelingen te doen voor beslissingen die genomen moeten worden.<sup>24</sup> Deze grote hoeveelheden data, zogenoemde ‘big data’, zijn meestal complex en vaak ongestructureerd en worden tegenwoordig continu gegenereerd.<sup>25</sup> Een veelgebruikte bron voor big data is het zogeheten ‘web data scraping’, waarbij informatie van het internet wordt gehaald. Ook databronnen die betrekking hebben op natuurlijke taal kunnen worden gebruikt voor datascience, zoals open antwoorden, tekstbestanden, aantekeningen van klantcontacten, rapporten of e-mails. Internetlogbestanden en de metadata van zoekmachines kunnen eveneens interessante informatie bieden over trends in de tijd. Ongeveer 80% van big data komt tegenwoordig uit dit soort bronnen.<sup>26</sup>

Uiteraard kunnen ook andere databronnen geanalyseerd worden op basis van datasciencetechnieken. Onze wereld is vol met apparatuur en toepassingen die in hoog tempo en in grote hoeveelheden data genereren, opslaan en verzenden. Daardoor zijn enorme hoeveelheden data beschikbaar voor allerlei (realtime)

- 23 N. Chawla e.a., ‘Smote: synthetic minority over-sampling technique’, *Journal of Artificial Intelligence Research* 2002, 16, p. 321-357.
- 24 S. Mullainathan & J. Spiess, ‘Machine Learning: An Applied Econometric Approach’, *Journal of Economic Perspectives* 2017, 2, p. 87-106.
- 25 V. Mayer-Schönberger & K. Cukier, ‘The Rise of Big Data’, *Foreign Affairs* 2013, mei/juni.
- 26 D. Cogburn & M. Hine, ‘Introduction to Text Mining in Big Data Analytics’, *Proceedings of the 50th Hawaii International Conference on System Sciences* 2017.

Met datascience op zoek naar indicatoren van georganiseerde criminaliteit en ondermijning

analyses. Denk bijvoorbeeld aan openbare wifi, socialemediadata, (zelfrijdende) auto's en vrachtwagens vol met sensoren, intelligente thuis- en kantoorapparatuur of beveiligingssystemen, digitale camerabeelden, sensoren in de openbare ruimte, smartphone-apps en andere wearables.

Maar ook vanuit traditionele databronnen kunnen door middel van moderne datasciencebenaderingen nieuwe inzichten gecreëerd worden. Zo kunnen ook (grote hoeveelheden) administratieve gegevens op deze manieren worden geanalyseerd, of data uit eerder afgenomen enquêtes, zoals de Veiligheidsmonitor of slachtofferenquêtes. Vooral door een combinatie van big data en *machine learning* met administratieve gegevens, surveydata en 'traditioneel' empirisch onderzoek kan er veel meerwaarde ontstaan.<sup>27</sup>

Door steeds geavanceerde *machine learning*-classificatietechnieken en -cluster technieken toe te passen kunnen verbanden worden geïdentificeerd. *Machine learning* wordt over het algemeen onderverdeeld in twee categorieën: *unsupervised* en *supervised machine learning*.<sup>28,29</sup> Bij *unsupervised machine learning*-methoden kan geen gebruik worden gemaakt van bestaande informatie, de zogenoemde 'labels', en is het doel om de onderliggende structuur of verdeling in de data te modelleren voor verkennende gegevensanalyse om verborgen patronen of clustering in gegevens te detecteren. De andere categorie is *supervised machine learning*, waarin wél gebruik wordt gemaakt van bekende informatie/labels. Er is een doelvariabele, en het algoritme leert om de output (onbekende labels of toekomstige data) te voorspellen op basis van de input (data en bestaande labels).<sup>30</sup> Een overzicht van recente technieken en toepassingen op uiteenlopende vraagstukken, ook op het gebied van veiligheid en stabiliteit van instituties en organisaties, wordt gegeven in de onderzoek agenda van Prüfer & Prüfer.<sup>31</sup>

## 5.2 Classificatietechniek Random Forest

In dit onderzoek is gebruikgemaakt van *supervised machine learning*-methoden. Er bestaan verschillende classificatietechnieken voor het voorspellen van een categorale variabele (in dit onderzoek: de kans op (georganiseerde) criminaliteit en ondermijning op het bedrijventerrein) op basis van een aantal indicatoren. Hiervoor is de *supervised machine learning*-techniek *Random Forest* (RF) gebruikt.<sup>32</sup> Een groot aantal beslissingsbomen ('decision trees') wordt gecombineerd om tot de uiteindelijke voorspelling te komen, waarbij de uiteindelijke uitkomst bepaald wordt door middel van 'majority voting'. Met andere woorden: de uitkomst van de meeste bomen geldt als einduitkomst. De verzameling bomen wordt random

27 J. Prüfer & P. Prüfer, 'Data Science for Entrepreneurship Research: Studying Demand Dynamics for Entrepreneurial Skills in the Netherlands', *Small Business Economics* 2019, p. 1-22, doi: 10.1007/s11187-019-00208-y.

28 C. Bishop, *Pattern Recognition and Machine Learning*, New York: Springer 2011.

29 K. Murphy, *Machine Learning: A Probabilistic Perspective*, Cambridge: MIT Press 2012.

30 T. Hastie, R. Tibshirani & J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer 2019.

31 J. Prüfer & P. Prüfer, 'Data science for institutional and organizational economics', in: *A Research Agenda for New Institutional Economics*, Edward Elgar Publishers 2018, p. 248-259.

32 L. Breiman, 'Statistical Modeling: The Two Cultures', *Statistical Science* 2001, 3, p. 199-231.

genoemd omdat elke boom getraind wordt op een willekeurige selectie van variabelen en observaties.

Deze methode is veel stabielier dan een enkele beslissingsboom, wat betekent dat een relatief kleine verandering in de data niet kan resulteren in compleet andere uitkomsten. Een RF leidt daarom over het algemeen tot betere voorspellingen dan een beslissingsboom. Daarnaast kan de techniek goed overweg met grote hoeveelheden observaties en/of variabelen en kunnen zowel numerieke als categorische variabelen worden gebruikt. De methode is bovendien ongevoelig voor de aanwezigheid van ruis, ontbrekende waardes en outliers.<sup>33</sup> Het is een erg geschikte methode als er in de dataset sprake is van multicollineariteit en interacties tussen predictoren (voorspellers), waardoor een RF zelfs niet-lineaire relaties in de data goed kan omschrijven. Ten slotte kan RF goed overweg met ongebalanceerde data, waarbij gebeurtenissen onevenredig verdeeld zijn, bijvoorbeeld als er veel meer observaties zijn waarin er geen sprake is van een crimineel incident dan observaties met criminele gevallen. Op basis van deze overwegingen, en gebaseerd op meerdere tests voor een geschikte modelselectie en modelprestatie, is in dit onderzoek gekozen voor deze classificatietechniek.<sup>34</sup>

### 5.3 Labels van de bedrijventerreinen

Voor *supervised machine learning* is het noodzakelijk om de mate van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein te operationaliseren als afhankelijke variabele. Voor elk bedrijventerrein hebben we daarom een score nodig, zodat we ons model kunnen trainen aan de hand van deze zogenoemde 'labels'. In dit onderzoek zijn twee manieren in overweging genomen. Het aantal meldingen bij politie en/of bij M. zou mogelijk een indicatie kunnen geven. Echter, het kan ook zo zijn dat er juist een lage meldingsbereidheid is doordat er georganiseerde criminaliteit en ondermijning plaatsvindt. Vanuit de theorie bestaat er geen eenduidige aanwijzing voor de relatie tussen politiemeldingen en (georganiseerde) criminaliteit en ondermijning (positief of negatief). Daarom was het gebruiken van het aantal geregistreerde incidenten bij de politie als absolute maatstaf voor de mate van georganiseerde criminaliteit en ondermijning lastig en is ervoor gekozen om een korte vragenlijst af te nemen om de bedrijventerreinen te categoriseren. Hierin werd aan respondenten gevraagd om voor elk van de bedrijventerreinen in de gemeente Tilburg een inschatting te maken of de kans op (georganiseerde) criminaliteit en ondermijning laag, gemiddeld of hoog is.

33 D. Slof, *Voorspellen van webwinkel aankopen met een Random Forest*, Rotterdam: Erasmus Universiteit 2014.

34 De modelprestatie, getoetst door middel van kruisvalidatie, geeft een nauwkeurigheid (*accuracy*) aan van 67%. Dit betekent dat in twee derde van de gevallen de risicoklassen (hoog – gemiddeld – laag) juist worden voorspeld. Bij een derde ligt de voorspelling een klasse ernaast (hoog versus gemiddeld of gemiddeld versus laag). In geen van de gevallen zitten de misclassificaties er totaal naast (hoog versus laag). De algemene voorspellingskracht van het model is met een Area Under Curve (AUC) van 0,92 (waar AUC = 1,0 een perfecte voorspelling betreft) bijzonder goed. Dit betekent dat we met de opgenomen indicatoren/potentiële risicofactoren de *expert opinion*-classificatie (lage, gemiddelde en hoge kans op misdaad en ondermijning) goed kunnen voorspellen.

De afzonderlijke scores van de respondenten zijn gecombineerd tot een eindscore per bedrijventerrein door gebruik te maken van ‘majority voting’. De categorie die door de meeste respondenten gekozen werd, is dus gebruikt als totaalscore.<sup>35</sup> Volgens de respondenten hebben acht bedrijventerreinen een lage kans, vijftien terreinen een gemiddelde kans en zeven terreinen een hoge kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning. Om extra te controleren of de resultaten uit de vragenlijst een goede indicatie geven van de mate van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, hebben we een vergelijking gemaakt met signalen die in 2013 zijn voortgekomen uit een integrale overheidssamenwerking.<sup>36</sup> Hieruit blijkt dat in het algemeen de resultaten uit de vragenlijst in lijn zijn met deze signalen.

## 6 Resultaten en discussie

In eerste instantie is RF toegepast op alle variabelen (features) in de gekoppelde dataset met alle potentiële indicatoren. Door gebruik te maken van bekende informatie (labels) over de risicoclassificatie van alle Tilburgse bedrijventerreinen is de doelvariabele, de kans op (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein, geoperationaliseerd en werden de voorspellende waarden van potentiële indicatoren in kaart gebracht.

De RF levert een ranglijst van alle variabelen op; deze top 15 van belangrijkste indicatoren (risicofactoren) is te zien in figuur 2.<sup>37</sup> De lijst is aflopend gesorteerd op de Gini-coëfficiënt, die vaak gebruikt wordt om de variabelen aan te duiden die het belangrijkste zijn en het grootste verschil maken voor de voorspelling.<sup>38</sup> Anders gezegd, dat zijn de variabelen die het beste kunnen scheiden. En een indicator met een hoge waarde van de Gini-coëfficiënt scheidt in dit geval dus de bedrijventerreinen die in goede staat verkeren optimaal van terreinen in een slechte staat. Deze ranglijst geeft dus inzicht in de meest belangrijke voorspellers van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, en zoals uit de top 15 blijkt zijn dit net name de ‘usual suspects’ die ook door middel van *supervised machine learning* naar voren komen.

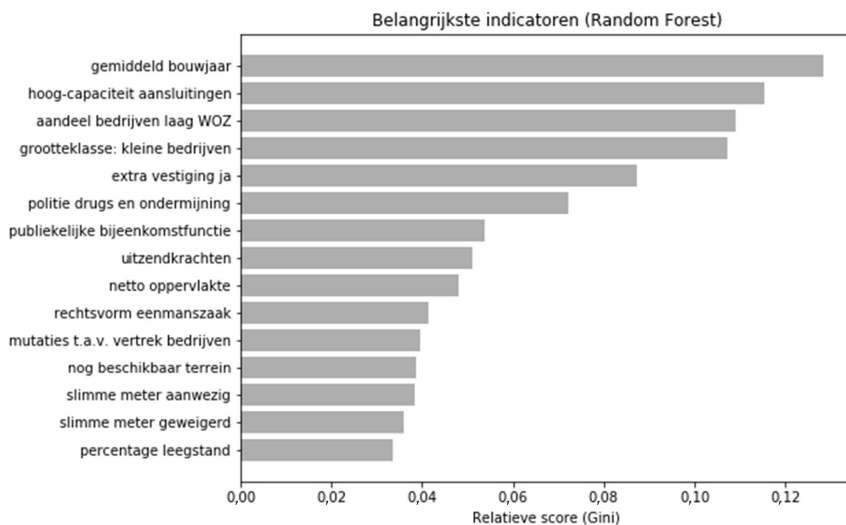
Uit onze resultaten blijkt dat de sterkste risicofactor voor de bedrijventerreinen in Tilburg, de variabele die het beste onderscheid maakt tussen goed en slecht, het gemiddeld bouwjaar van de panden op een bedrijventerrein is. De top 15

35 Bij *majority voting* wordt aan ieders beoordeling een gelijk gewicht toegekend. Een beperking hierbij is dat de ene expert meer kennis kan hebben dan andere experts. Aangezien het niet mogelijk was om op een objectieve en betrouwbare wijze verschillende gewichten toe te kennen aan de respondenten, is er gewerkt met gelijke gewichten.

36 Op verzoek is de bron geanonimiseerd.

37 Uit de lijst van alle mogelijk relevante variabelen is de top 15 van belangrijkste indicatoren geselecteerd om de RF-algoritme nog een keer te trainen. Dit maakt de uitslag van het model robuuster, omdat de coëfficiënten ten opzichte van elkaar geoptimaliseerd worden.

38 Deze meet de ongelijkheid tussen waarden van een frequentieverdeling (in dit onderzoek dus de classificatie van laag, gemiddeld en hoog). Een Gini-coëfficiënt van nul geeft de perfecte gelijkheid weer, waarbij alle waarden hetzelfde zijn (bijv. waar ze allemaal gemiddeld geïnclassificeerd zijn). Een Gini-coëfficiënt van 1 (of 100%) drukt de maximale ongelijkheid tussen waarden uit.

**Figuur 2** Belangrijkste indicatoren (Random Forest-algoritme)

omvat naast het gemiddelde bouwjaar verder vijf andere risicofactoren die beduidend sterker naar voren komen ten opzichte van de andere indicatoren (Gini-coëfficiënt groter dan het gemiddelde van de top 15).

Om zicht te krijgen op de richting van de invloed van een bepaalde variabele zijn aanvullend op de RF een beslissingsboom en een logistische regressie gedraaid op de top 15 risicofactoren.<sup>39</sup> Deze informatie is niet direct uit een RF te halen, maar wel informatie over de meest belangrijke combinatie van risicofactoren. Deze 'giftige cocktail' op een bedrijventerrein is een mix uit de volgende indicatoren:

- 1 *Gemiddeld bouwjaar*: hoe lager het gemiddelde bouwjaar op een bedrijventerrein, des te slechter het scoort.
- 2 *Aandeel panden laag WOZ*: hoe groter het percentage van panden op een bedrijventerrein met een WOZ-waarde lager dan € 160.000, des te slechter het scoort.
- 3 *Hoog-capaciteitaansluitingen*: hoe minder het aantal hoog-capaciteitaansluitingen op een bedrijventerrein (ten opzichte van laag-capaciteit), des te slechter het scoort.
- 4 *Grootteklasse: kleine bedrijven*: hoe groter het percentage bedrijven met minder dan vijf werknemers, des te slechter het scoort.
- 5 *Extra vestiging ja*: hoe meer extra vestigingen van bedrijven aanwezig zijn op een bedrijventerrein, des te slechter het scoort.

39 De uitkomsten van deze *decision tree* zijn op verzoek verkrijgbaar bij de auteurs of te vinden in het onderzoeksrapport op de website van het WODC: [www.wodc.nl/onderzoeksdatabase/2836-de-ontwikkeling-van-\(voorspellende\)-criminogene-indicatoren-voor-wijken-bedrijvengebieden-en-bedrijfstacken.aspx](http://www.wodc.nl/onderzoeksdatabase/2836-de-ontwikkeling-van-(voorspellende)-criminogene-indicatoren-voor-wijken-bedrijvengebieden-en-bedrijfstacken.aspx).

## 6 *Politie drugs en ondermijning*: hoe meer incidenten gerelateerd aan drugs en ondermijning er zijn op een bedrijventerrein, des te slechter het scoort.

Het is niet verrassend dat bij de politie bekende incidenten (gerelateerd aan drugs en ondermijning) een risicofactor zijn voor slecht scorende terreinen. Het verleden is een belangrijke indicatie, maar geen zekerheid voor het heden of de toekomst. Daarentegen is de hoeveelheid als extra vestiging op een bedrijventerrein geregistreerde panden wel een interessante – misschien zelfs verrassende – risicofactor. Extra vestigingen zijn vestigingen waar niet gewerkt wordt, maar waarvan het adres toch van belang is. Extra vestigingen zijn anderzijds ook de panden waarvan alleen een adres bekend is en aanvullende informatie ontbreekt.

## 7 Ten slotte

Tot op heden was er geen instrument of methodiek beschikbaar waaruit blijkt waar bedrijventerreinen zich bevinden op een ‘glijdende schaal’ van criminaliteit en ondermijning. Het doel van deze studie was dan ook om te bepalen of er indicatoren konden worden ontwikkeld waardoor het voor bestuurders inzichtelijk wordt dat op een bedrijventerrein sprake kan zijn van de ontwikkeling of aanwezigheid van (georganiseerde) criminaliteit en ondermijning. In deze verkennende pilot hebben we hiervoor de eerste stappen gezet.

De indicatoren die een belangrijke voorspellende waarde hebben, zouden in de toekomst gebruikt kunnen worden om in de praktijk inzichtelijk te maken welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning. Hiervoor is het echter wel noodzakelijk om eerst op grotere schaal, dus voor meer bedrijventerreinen, onderzoek te verrichten, zodat een betrouwbaar model geschat kan worden. Vervolgens zou ook voor andere bedrijventerreinen, waarvoor nog geen label bekend is, op basis van de indicatoren voorspeld kunnen worden of het terrein een lage, gemiddelde of hoge kans op (georganiseerde) criminaliteit en ondermijning heeft. Daarnaast kunnen de inzichten uit dit onderzoek gebruikt worden om de risicoverhogende indicatoren in de praktijk aan te pakken.

In de toekomst zou dezelfde methodiek als in dit onderzoek ook toegepast kunnen worden om breder te kijken naar indicatoren van (georganiseerde) criminaliteit en ondermijning, bijvoorbeeld in woonwijken. Er zullen dan wel wat aanpassingen gedaan moeten worden, voornamelijk in de gebruikte databronnen, en er zullen opnieuw verschillende methoden getest moeten worden om te bepalen welke methode het meest geschikt is voor deze nieuwe toepassing.

Naast nieuwe methoden en technieken zoals in deze bijdrage beschreven, blijft aanvullend klassiek handwerk van levensbelang. Een sprekend voorbeeld van de onzichtbaarheid (en afschermingsmethoden) van criminaliteit op bedrijventerreinen vinden we in de uitgave van de Politieacademie, getiteld *Sluipend gif*. Een van de bedrijven in de beschreven casus is een groothandel in tuinbenodigdheden. Meer specifiek levert het bedrijf verlichting, ventilatie en aanverwante artikelen aan particulieren en bedrijven. Het is een bedrijf van IKEA-achtige proporties. Op

het eerste gezicht lijkt het een keurige onderneming, maar volgens betrokkenen is het in werkelijkheid een toeleverancier van benodigdheden voor growshops. Het keurige voorkomen van het bedrijf zorgt ervoor dat het lang onder de radar van de overheid kon blijven. Wanneer de Belastingdienst in 2008 een routinecontrole in het bedrijf uitvoert, komt het andere gezicht naar buiten. Eenmaal binnen bekruipt de medewerkers van de Belastingdienst een gevoel van onveiligheid.<sup>40</sup> Dit voorbeeld geeft het belang van observaties ter plekke aan.

Ondernemers op bedrijventerreinen, in horecagelegenheden en winkels weten vaak precies waar de zaken niet kloppen. De uitdaging is om deze partijen te benutten als bron van informatie en tegelijkertijd hen weerbaar te maken tegen de effecten van georganiseerde ondermijnende criminaliteit. In een onderzoek naar criminaliteit op bedrijventerreinen werd een groot aantal ondernemers geïnterviewd die deze vorm van criminaliteit 'goed georganiseerd' noemden.<sup>41</sup> Criminelen zouden er wel voor zorgen dat niemand last van hen had. 'Ze rijden echt niet met getrokken pistolen langs', is een breed gedragen opvatting op de bedrijventerreinen. Toch is angst een belangrijke reden dat zij wantoestanden niet snel melden bij de politie. 'Want dan weten ze me te vinden.' Witwassen duidt op andere vormen van criminaliteit. De verschillende handhavingspartners kunnen een gezamenlijke strategie bepalen voor een krachtigere aanpak. Dat is volgens dit onderzoek een voorwaarde voor het zichtbaar maken van die problematiek, zodat ondernemers zien dat er niet meer wordt gedoogd. Wil de overheid meer meldingen, dan moeten de ondernemers erop kunnen vertrouwen dat hun meldingen worden opgepakt.

Een discrepantie tussen de werkelijke situatie op een bedrijventerrein en wat er in registraties is opgenomen ('de papieren werkelijkheid'), werd tijdens de expertmeeting ook genoemd als mogelijke indicator. Daarom is het belangrijk om de informatie op papier te valideren, aan te vullen en te verrijken met wat er in de praktijk geobserveerd wordt. Het zou bijvoorbeeld verdacht kunnen zijn als er erg weinig bedrijvigheid is op het terrein of bij een op klanten gericht bedrijf, terwijl er wel hoge omzetten opgegeven worden bij de Belastingdienst. Ook een groot aantal inschrijvingen bij de KvK op een adres, terwijl er in werkelijkheid maar één of zelfs geen bedrijf is, wordt genoemd als mogelijke indicator.

Daarnaast kwam uit de interviews naar voren dat je veel dingen niet in de data kunt zien: zachte relaties, dat er een laag stof op auto's ligt, dat er geen bedrijvigheid is, hoe mensen gekleed gaan (trainingspakken) of dat mensen in te dure auto's rijden. Ook tijdens de expertmeeting werden verschillende indicatoren genoemd die gebaseerd zijn op observaties ter plekke, bijvoorbeeld het feit dat er op ongebruikelijke tijden (voor het type bedrijven dat er zit) veel activiteit is op het terrein. De voertuigen op het terrein kunnen volgens de experts ook verdeningen oproepen. Het feit dat er veel huurbusjes rondrijden, dat dure auto's het terrein op en af rijden voor veelal korte bezoeken en dat er veel auto's/busjes met

40 Lam, Van der Wal & Kop 2018, p. 128.

41 D van der Wiele, 'Normloosheid op het bedrijventerrein', *Secondant* 2016, 2 mei, geraadpleegd op 10 april 2018 via [www.ccv-secondant.nl/platform/article/normloosheid-op-het-bedrijven-terrein/](http://www.ccv-secondant.nl/platform/article/normloosheid-op-het-bedrijven-terrein/).



Met datascience op zoek naar indicatoren van georganiseerde criminaliteit en ondermijning

buitenlandse kentekens rondrijden (waarbij er ook op gelet kan worden uit welk land de voertuigen komen) wordt als verdacht beschouwd. Ook een sterke afwijkende geur die aanwezig is in het gebied kan een belangrijke indicator zijn die niet is vastgelegd in databestanden.

In vervolgonderzoek zou de huidige onderzoeksmethode nog uitgebreid kunnen worden met een aantal additionele – ook niet-openbare – databronnen. Binnen dit onderzoek zijn alle data verzameld of geaggregeerd tot bedrijventerreinniveau. Er zijn echter nog vele mogelijkheden voor vervolgonderzoek als ook data van individuele bedrijven en personen in de analyses worden betrokken.