



Risicotaxatietool criminaliteit en ondermijning op bedrijventerreinen

Eindrapport

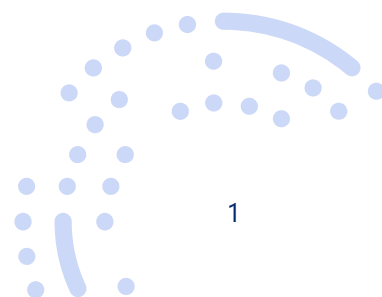
Provincie Noord-Brabant

≡ provincie
Gelderland



avans
hogeschool

Datum 22 april 2022
Auteurs Seyit Höcük (Centerdata)
Thijs Bertram (Centerdata)
Patricia Prüfer (Centerdata)
Jonas Stuurman (Avans)
Jimmy Maan (Avans)
Emile Kolthoff (Avans)





Managementsamenvatting

In opdracht van de provincies Gelderland en Noord-Brabant werd van januari 2021 tot maart 2022 door Centerdata in samenwerking met het lectoraat Ondernijning van Avans Hogeschool een onderzoek uitgevoerd met het doel een risicotaxatietool te ontwikkelen, waarmee voor alle bedrijventerreinen in de beide provincies de relatieve kwetsbaarheid voor ondermijnende activiteiten in beeld wordt gebracht. Dit gebeurde door middel van voorspellende modellen en wordt gevisualiseerd via een interactief dashboard. Het dashboard vormt samen met de handelingsperspectieven op de geïdentificeerde risico-indicatoren het eindproduct van dit project.

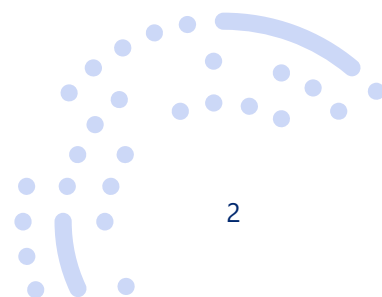
Tot op heden was er geen tool beschikbaar waarmee getaxeerd kon worden hoe kwetsbaar bedrijventerreinen zijn ten aanzien van criminaliteit en ondernijning. Een eerste proeve van een dergelijke techniek werd in 2017 opgeleverd in de vorm van een rapport voor het WODC (Bolsius et al., 2018). De resultaten van die pilot in de gemeente Tilburg waren veelbelovend, maar door een gebrek aan beschikbare data was het toen ontwikkelde model nog lang niet voldragen of operationeel in te zetten. In het hier gepresenteerde onderzoek is voortgebouwd op de pilot uit 2017, met gebruikmaking van meerdere rekentools en een grote hoeveelheid databronnen.

In het onderzoek dienden twee vragen beantwoord te worden:

1. Hoe kan op basis van data vastgesteld worden welke bedrijventerreinen kwetsbaarder zijn dan andere voor criminaliteit en/of ondernijning?
en:
2. Welke handelingsperspectieven hebben provincies en gemeenten om iets aan daadwerkelijk kwetsbare bedrijventerreinen te doen?

De beantwoording van de eerste onderzoeksvraag heeft geleid tot een prioriteringstool in de vorm van een 'dashboard' waarin de relatieve kwetsbaarheid van bedrijventerreinen op een kaart gevisualiseerd wordt. Ook wanneer deze problematiek zich (nog) niet heeft geopenbaard kunnen gemeenten hierop preventief acteren.

In het onderzoek is ervoor gekozen de focus te leggen op ondernijning in het algemeen en drie specifieke ondernijningsvormen: drugscriminaliteit, witwassen, en mensenhandel. De relatieve kwetsbaarheid van terreinen voor twee van de drie ondernijningsvormen, drugs en witwassen, blijkt goed voorspelbaar te zijn binnen het kader van dit onderzoek, evenals ondernijning in het algemeen, welke een combinatie is van de drie ondernijningsvormen. De derde ondernijningsvorm, mensenhandel, doorstaat de testen van voorspellend vermogen en theoretische validatie niet. Er zijn niet genoeg data over mensenhandel en ook te weinig 'gelabelde' doelvariabelen. Naast het bepalen van de relatieve kwetsbaarheid was er ook behoefte aan inzicht in de combinatie van belangrijkste risico-indicatoren om mogelijke aangrijpingspunten voor bestuurders en beleidsmakers inzichtelijk te maken. Het onderzoek maakt aan de hand van combinaties van top-indicatoren duidelijk wanneer een bedrijventerrein relatief kwetsbaar is.





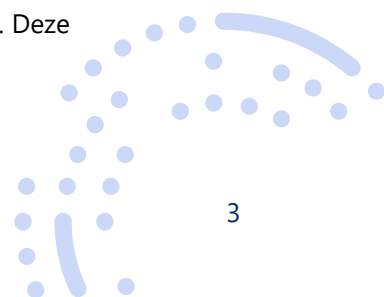
Het project is gestart met het identificeren van potentiële risico-indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Door middel van literatuuronderzoek, expertbijeenkomsten, en interviews is gekeken welke kennis en informatie al beschikbaar was. Er zijn hierbij mogelijke indicatoren geïdentificeerd met betrekking tot de omgevingskenmerken, ondernemingskenmerken, en informatiepositie op een bedrijventerrein.

Na het identificeren van mogelijke risico-indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, zijn data verzameld uit verscheidene bronnen. Kenmerken over bedrijventerreinen waren beschikbaar in het IBIS en het LISA register. Politiedata zijn ontvangen over incidenten, overlast, en handhaving. Data van omgevingsdiensten zijn verkregen via het S@men portaal. Daarnaast hebben de netbeheerders Enexis en Liander data geleverd over de capaciteit van energieaansluitingen. Ook hebben de provincies data uit het KVK-bedrijvenregister, Kadaster, en de BAG ter beschikking gesteld. Verder zijn in het model data van verleende subsidies, ANBI-registraties, Leefbaarometer-scores, en leegstandsmonitors meegenomen. Tot slot zijn er veel buurtgegevens meegenomen die via het CBS beschikbaar zijn gemaakt over gezondheid, leefbaarheid, uitkeringen, voorzieningen, en kerncijfers. Het resultaat was een uitgebreide dataset van 20 onafhankelijke databronnen met honderden mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen.

Om bedrijventerreinen met een verhoogd risico op – of kwetsbaarheid voor - criminaliteit en ondermijning te identificeren, is gebruik gemaakt van geavanceerde data science methoden en zijn drie verschillende machine learning-modellen ingezet: Random Forest, Logistic Regression, en Gradient Boosting. Het Gradient Boosting-model kwam met een voorspellingsscore van 90% voor drugs als beste naar voren tijdens de evaluaties. Dat betekent dat het model 9 van de 10 keer het verwachte risico op ondermijnende criminaliteit in de door experts 'gelabelde' bedrijventerreinen correct wist te voorspellen. Ook is een praktische validatie uitgevoerd waarbij op basis van interviews en observeerbare indicatoren van (kwetsbaarheid voor) ondermijning, de uitkomst van het model zijn getoetst en bevestigd.

Met de combinatie van drie verschillende machine learning-modellen en de methodiek van zowel theoretische als praktische validatie zoals beschreven in het eindrapport, is de verwachting dat het model een voorspellende waarde heeft voor de relatieve kwetsbaarheid voor uitingsvormen van ondermijning van meer dan 80% betrouwbaarheid. Daarmee is in dit onderzoek een nieuwe maatstaf neergelegd voor data gedreven risicomodellen bij veiligheidsonderzoeken. Het betrouwbaarheidspercentage zou nog omhoog gebracht kunnen worden tot circa 90% door een aanvullende praktijktoets uit te voeren. Dit is voorzien als vervolgproject. Hierbij dient de kanttekening geplaatst te worden dat een hoog risico niet automatisch betekent dat er ondermijning plaatsvindt op een bedrijventerrein. Het betreft een *risico*, wat iets zegt over de kwetsbaarheid van een bedrijventerrein. Een hoog risico betekent dat een terrein, vergeleken met andere terreinen, een hoger risico loopt op het ontwikkelen van ondermijnende criminaliteit.

Naast het doen van een betrouwbare voorspelling laat het onderzoek ook zien welke indicatoren bijdragen aan een verhoogd risico op ondermijning, wat weer een aangrijpingspunt voor preventie kan zijn. Per ondermijningsvorm zijn de 20 meest invloedrijke indicatoren geselecteerd. Deze indicatoren bestaan uit een combinatie van factoren die betrekking hebben op



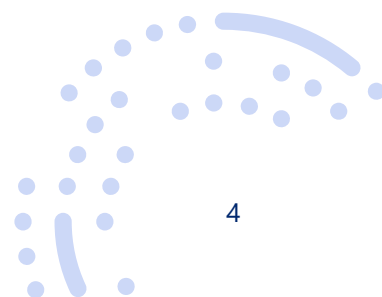


omgeving, onderneming, en informatiepositie. Zo kijkt het model niet alleen naar wat er op een bedrijventerrein gebeurt, maar ook hoeveel er bekend is over een bedrijventerrein, of wat er in de buurt rondom een bedrijventerrein aan de hand is.

Bij de beantwoording van de tweede onderzoeksvraag (het handelingsperspectief van provincies en gemeenten) is uitgegaan van de provinciale taken. De provincie heeft primair een signalerende taak en de inzichten uit het dashboard kunnen worden gebruikt om gemeenten te informeren, de gesignaleerde problematiek te agenderen, en in te zetten bij het aanjagen van lokale en regionale partners met betrekking tot de aanpak van kwetsbare bedrijventerreinen. De uitkomsten uit het dashboard worden door de provincie beschikbaar gesteld aan individuele gemeenten. Als nevenproduct van dit onderzoeksproject is een Handreiking Handelingsperspectieven Risicotaxatietool Bedrijventerreinen opgesteld, die de provincies kunnen gebruiken om met gemeenten in gesprek te gaan over de uitkomsten en waarin een overzicht wordt gegeven van beschikbare instrumenten en informatie over hoe deze in te zetten. Gemeenten kunnen vervolgens allereerst de eigen informatiepositie op een door het dashboard als kwetsbaar aangemerkt bedrijventerrein verdiepen, door middel van het opstellen van een informatiebeeld op een lager aggregatieniveau. Hiervoor zijn uitgewerkte werkprocessen opgesteld in o.a. de Toolbox bedrijventerreinen van Taskforce-RIEC Brabant-Zeeland, die voor alle Nederlandse gemeenten beschikbaar wordt gesteld via het Kennisplatform Ondernijning (KPO). Gemeenten doen dit niet alleen. Er kan worden aangesloten bij bestaande initiatieven zoals in Brabant Taskforce-RIEC Brabant-Zeeland. In Gelderland geldt dit ten aanzien van het RIEC Oost-Nederland en project Theseus. Daarnaast kan aansluiting gezocht worden bij het Platform Veilig Ondernemen (PVO). Ook kunnen andere partners, zoals omgevingsdiensten, herstructureringsmaatschappijen (Gelderland), de Brabantse Ontwikkelmaatschappij (BOM) en het Centrum voor Criminaliteitspreventie en Veiligheid (CCV) worden betrokken bij de aanpak van ondernijning op bedrijventerreinen.

De handreiking staat in principe los van het risicotaxatiemodel. Op basis van de uitkomsten van het risicomodel kunnen ook andere interventies worden gepleegd dan in de handreiking zijn genoemd en de in de handreiking genoemde instrumenten van de toolbox zijn ook in te zetten zonder het risicoanalysemodel te gebruiken, bijvoorbeeld naar aanleiding van klachten of signalen vanuit de politie.

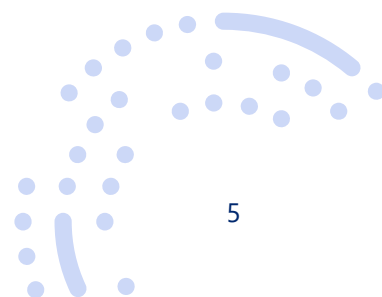
Het opgeleverde dashboard is flexibel van aard en kan verder ontwikkeld en verfijnd worden al naar gelang toekomstige wensen. Zo kunnen bijvoorbeeld nieuwe databronnen en andere vormen van ondernijning aan het model worden toegevoegd. Door het aansluiten van andere provincies of regio's generaliseert het model op een nog breder gebied waardoor inschattingen van de relatieve kwetsbaarheid nog betrouwbaarder worden. Daarnaast kan een uitrol van deze aanpak helpen om tot een landelijke aanpak te komen om georganiseerde criminaliteit en ondernijning tegen te gaan.





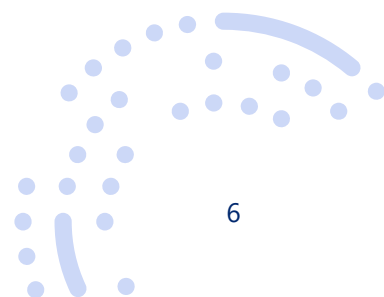
Inhoudsopgave

Managementsamenvatting	2
1 Inleiding	7
1.1 Aanleiding en doel van het rapport	7
1.2 Onderzoeksvragen	7
1.3 Focusgebied risicotaxatietool	8
1.4 Opbouw van het rapport	8
2 Achtergrond	9
2.1 Definiëren van criminaliteit en ondermijning	10
3 Plan van aanpak	12
4 Dataverzameling	13
4.1 Longlist risico-indicatoren	13
4.2 Shortlist risico-indicatoren	14
4.3 Databronnen vergaren	15
5 Doelvariabelen	18
5.1 Vragenlijsten	18
5.2 Labels	19
6 Analyses en resultaten	22
6.1 Voorspellingsscores	22
6.2 Top-indicatoren	24
6.3 Aanvullende theoretische validatie	27
6.4 Belangrijke databronnen	29
6.5 Interpreteren van de uitkomsten	30
7 Praktische validatie	31
7.1 Opzet en organisatie	31
7.2 Werkwijze	31
7.3 Uitkomsten	33
8 Dashboard	35





9	Samenvatting handelingsperspectieven	39
9.1	Informeren	39
9.2	Agenderen	40
9.3	Aanjagen	41
9.4	Handelingsperspectief gemeente	41
10	Slotbeschouwing en toekomstperspectieven	43
10.1	Slotbeschouwing	43
10.2	Toekomstperspectieven	44
	Literatuurlijst	45
	Lijst van Tabellen en Figuren	46
Bijlage A	De lijst van shortlist risico-indicatoren	47
Bijlage B	Korte introductie data science methoden	48
Bijlage C	Voorbeeld van een vragenlijst	50
Bijlage D	Modelleren	52
Bijlage E	Voorbeeld correlatiecheck	59
Bijlage F	Beschrijving classificatiemodellen	60
Bijlage G	Optimale modelparameters	62
Bijlage H	Alle voorspellingsscores	63
Bijlage I	Test op voorspellend vermogen	64
Bijlage J	Uitleg van top-indicatoren	65





1 Inleiding

1.1 Aanleiding en doel van het rapport

In opdracht van de provincies Gelderland en Noord-Brabant, de opdrachtgevers, is begin januari 2021 een gezamenlijk project opgestart met het doel om een risicotaxatietool te bouwen voor bedrijventerreinen in beide provincies. De kwetsbaarheden van bedrijventerreinen zouden door middel van voorspellende technieken in kaart worden gebracht en worden gevisualiseerd via een interactief dashboard. Het dashboard samen met de handelingsperspectieven op de geïdentificeerde risico-indicatoren dienen als het eindproduct van de risicotaxatietool. De datagedreven modellen zijn inmiddels uitgevoerd, de risico-indicatoren van kwetsbare bedrijventerreinen zijn in beeld gebracht, en het dashboard en de handelingsperspectieven, die zelfstandig kunnen worden gebruikt, zijn opgeleverd. Dit eindrapport beschrijft de totstandkoming van het gehele proces.

Het doel van dit rapport is tweeledig. Enerzijds om de ondernomen stappen tot het doel duidelijk in beeld te brengen, van de initiële onderzoeksvragen van de provincies en de daaropvolgende dataverzameling tot aan het modelleren en valideren toe, en anderzijds de bevindingen van de voorspellende analyses te presenteren met voorbeelden vanuit het dashboard. De ingeschatte risico's (kans op ondermijnende criminaliteit) op bedrijventerreinen worden alleen in het dashboard op het gewenste niveau weergegeven vanwege de gevoeligheid van de uitkomsten. De modeluitkomsten zijn zowel theoretisch als praktisch gevalideerd en garanderen een betrouwbaarheid van de uitkomsten. Hierop volgen perspectieven voor de handelingen die kunnen worden ondernomen om preventief in te grijpen op kwetsbare bedrijventerreinen die een negatief uitzicht hebben.

De uitvoering van dit project is een samenwerking tussen onderzoekers en data scientists van Centerdata en onderzoekers en criminologen van Avans Hogeschool.

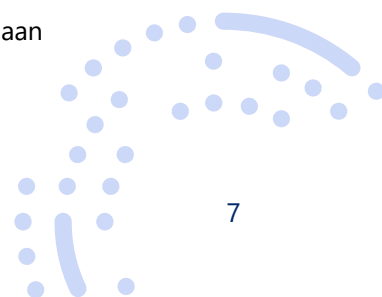
1.2 Onderzoeksvragen

Het project is opgestart met twee onderzoeksvragen van de opdrachtgevers die zijn opgesteld om kwetsbare bedrijventerreinen tijdig aan te kunnen pakken. De onderzoeksvragen zijn hieronder geformuleerd. De eerste hoofdvraag luidt:

1. Hoe kan op basis van data vastgesteld worden welke bedrijventerreinen kwetsbaarder zijn dan andere voor criminaliteit en/of ondermijning?

Een directe vervolgvraag hierop van de opdrachtgevers is welke data hiervoor nodig zijn. Een nog specifiekere verwoording van de eerste hoofdvraag van de opdrachtgevers is of er een tool gecreëerd kan worden waarmee je snel en makkelijk met behulp van beschikbare data een vlag krijgt voor het risico op verschillende vormen van ondermijnende criminaliteit op een specifiek bedrijventerrein. Deze vraag wordt beantwoord door het maken van een 'kaart voor de mate van kwetsbaarheid', waarmee er periodieke monitoring uitgevoerd kan worden. De tweede hoofdvraag luidt:

2. Welke handelingsperspectieven hebben provincies en gemeenten om iets aan daadwerkelijk kwetsbare bedrijventerreinen te doen?





1.3 Focusgebied risicotaxatietool

Mede met behulp van cijfers uit het BVH-bestand¹ van de politie is in de beginfase van het onderzoek en in overleg met de opdrachtgevers definitief bepaald op welke vormen van criminaliteit het onderzoek zich richt. De opdrachtgevers waren initieel geïnteresseerd in het opsporen van twee vormen van (georganiseerde) criminaliteit en ondermijning: drugs en mensenhandel (uitbuiting, illegale arbeid) en indien mogelijk ook een derde vorm: witwassen. Tot op heden is er geen instrument of methodiek beschikbaar waaruit blijkt waar deze gebieden zich op de 'glijdende schaal' bevinden. Wel is bekend dat in de regio het merendeel van ondermijnende criminaliteit zich concentreert op drugscriminaliteit en mensenhandel.

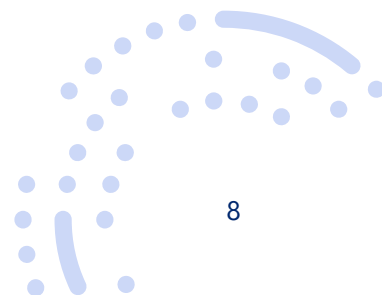
Het vooronderzoek, dataverzamelingstraject, en de modeluitkomsten samen zullen moeten uitwijzen welke ondermijningsvormen uiteindelijk geschikt zijn voor het inzetten van voorspellende modellen.

Het hoofddoel van het uiteindelijke product is om met behulp van preventieve signalen de relatieve kwetsbaarheid van een bedrijventerrein ten aanzien van ondermijnende criminaliteit inzichtelijk te maken voor bestuurders en beleidsmakers.

1.4 Opbouw van het rapport

Dit rapport is als volgt opgebouwd. In hoofdstuk 2 gaan we in op de achtergrond van de problematiek op bedrijventerreinen waarvoor de risicotaxatietool wordt gebouwd. In hoofdstuk 3 laten we een stappenplan zien voor het aanpakken van het project. In hoofdstuk 4 gaan we verder met het vooronderzoek voor het opstellen van een longlist van potentiële risico-indicatoren en vertellen we over de ondernomen stappen voor de dataverzameling. In hoofdstuk 5 gaan we nader in op de methodiek voor het inschatten van de kwetsbaarheid van bedrijventerreinen. Hier worden aan de hand van expertinzichten labels toegekend aan bedrijventerreinen voor de drie ondermijningsvormen (drugs, mensenhandel, witwassen). De labels zijn, naast objectieve data, belangrijk om een goed voorspellend model te kunnen bouwen. In hoofdstuk 6 presenteren we de analyses en de resultaten van het model. De modellen worden in dit gedeelte tevens theoretisch gevalideerd. In hoofdstuk 7 gaan we verder met de praktische validatie van de modeluitkomsten. Dit wordt onafhankelijk van de modellen uitgevoerd middels expedities van studenten en interviews op bedrijventerreinen. In hoofdstuk 8 laten we de werking van het dashboard zien aan de hand van enkele voorbeelden. In hoofdstuk 9 worden bijpassende handelingsperspectieven voorgelegd. De handelingsperspectieven gelden voor zowel de provincies als de gemeenten voor een preventieve aanpak van kwetsbare bedrijventerreinen. Ter afsluiting is er in hoofdstuk 10 een slotbeschouwing.

¹ De Basisvoorziening Handhaving (BVH) is een incidentregistratiesysteem dat wordt gebruikt door de Nederlandse politie.





2 Achtergrond

Witwaspraktijken, hennepsteelt, productie van synthetische drugs, mensenhandel en doorvoer van illegale goederen: op veel bedrijventerreinen is er sprake van verborgen criminaliteit. Zeker verloederde bedrijventerreinen bieden vaak een vruchtbare voedingsbodem voor criminele en ondermijnende activiteiten.

Om ondermijningsproblematiek aan te pakken onderzoeken de provincies de mogelijkheden om via datagedreven oplossingen nieuwe initiatieven te ontwikkelen. Niet alleen in de provincies Gelderland en Noord-Brabant, maar ook op nationaal niveau en zelfs in Europa is criminaliteit en ondermijning een serieus probleem wat hoog op de agenda van de samenleving staat.

De geïntegreerde aanpak van georganiseerde criminaliteit heeft zich de afgelopen 20 jaar doorontwikkeld. Om tot veilige en leefbare wijken, vitale bedrijventerreinen en integere bedrijfssectoren te komen, is een brede aanpak nodig waarin strafrecht, verstoring, handhaving en sociaal beleid gecombineerd worden. Deze geïntegreerde aanpak vergt op lokaal, regionaal, provinciaal en landelijk niveau intensivering van de samenwerking tussen betrokken overheden, hun private en maatschappelijke partners en burgers.

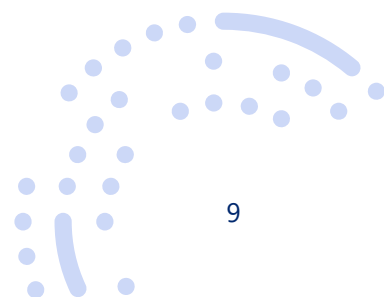
Provincie Gelderland telt volgens IBIS² 490 bedrijventerreinen in 2020. Provincie Noord-Brabant telt 621 bedrijventerreinen. Georganiseerde misdaad en ondermijnende criminaliteit kennen vele verschijningsvormen. In de aanbiedingsbrief bij het Nationaal Dreigingsbeeld georganiseerde criminaliteit 2017-2021, dat door de minister van Veiligheid en Justitie op 1 juni 2017 aan de Tweede Kamer werd aangeboden, werd een 17-tal verschijningsvormen van criminaliteit met het predicaat 'dreiging' gepresenteerd³, zie **Tabel 1**.

Drugs-, mensenhandel, en witwassengerelateerde criminaliteit waren goed voor de meeste dreigingspredicaten. Dit zijn ook de misdrijven die centraal staan in de Monitor Georganiseerde Criminaliteit, waarbij drugsgerelateerde criminaliteit de overhand heeft (Kruisbergen, Van de Bunt & Kleemans, 2012). Hetzelfde geldt voor de cijfers inzake geregistreeerde criminaliteit in de jaarlijkse publicatie Criminaliteit en Rechtshandhaving (Kalidien, 2017).

Preventieve maatregelen spelen een steeds prominentere rol bij de geïntegreerde aanpak van georganiseerde criminaliteit en ondermijning, waarbij veel aandacht is voor de aanpak van gebieden. Gebiedenaanpak en de aanpak van ondermijnende criminaliteit zijn twee kanten van dezelfde medaille (Tops & Van der Torre, 2015). Door cumulatie van problemen in kwetsbare wijken en bedrijventerreinen ontstaat er een voedingsbodem voor criminele activiteiten. Er is sprake van een continuüm met aan de ene kant de normale bovenwereld en aan de andere kant de harde criminele

² IBIS staat voor Integraal Bedrijventerreinen Informatie Systeem en bevat veel beschrijvende data over alle (bijna 3.800) Nederlandse bedrijventerreinen.

³ Brief minister van Veiligheid en Justitie van 1 juni 2017, kenmerk 2082923.





onderwereld; daartussen bestaan allerlei vormen variërend van bewust weggijken via een graantje meepikken tot aan het verlenen van hand- en spandiensten aan criminelen.

Tabel 1: Criminaliteitsvormen met de predicaat 'dreiging' uit het Nationaal Dreigingsbeeld

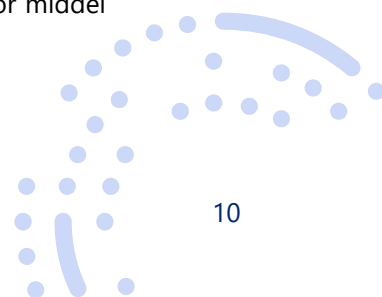
Illegale markten
Handel in en smokkel van cocaïne
Handel in en smokkel van heroïne
Productie van, handel in en smokkel van synthetische drugs
Productie van, handel in en smokkel van cannabis
Seksuele uitbuiting
Arbeidsuitbuiting, criminele uitbuiting en gedwongen dienstverlening
Mensensmokkel
Illegale handel in en smokkel van vuurwapens en explosieven
Productie en verspreiding van kinderpornografie
Fraude en witwassen
Beleggingsfraude
Faillissementsfraude
Verzekeringsfraude
Accijnsfraude
Fiscale fraude
Witwassen
Vermogenscriminaliteit
Woninginbraak
Overvallen

In dit kader zijn reeds verschillende onderzoeken uitgevoerd. Een hiervan is uitgevoerd door het onderzoeksinstituut Centerdata in samenwerking met Avans Hogeschool, en richtte zich expliciet op de verkenning van datagedreven mogelijkheden bij de signalering van ondermijnende criminaliteit op bedrijventerreinen (Bolsius et al. 2018; Prüfer & Kolthoff, 2020).

2.1 Definiëren van criminaliteit en ondermijning

Er bestaan uiteenlopende definities over wat georganiseerde criminaliteit en ondermijning exact inhouden. Zowel bij overheidspartijen die zich bezighouden met de aanpak als in het wetenschappelijk debat lopen definities uiteen en bestaat er nog geen consensus over ondermijning.

Voor het definiëren van georganiseerde criminaliteit sluiten we voor dit onderzoek aan bij die van de zogenaamde Groep Fijnaut in het "Eindrapport Georganiseerde criminaliteit in Nederland" (Tweede Kamer, vergaderjaar 1995–1996, kenmerk 24 072, nr. 16). Deze hanteert een definitie van georganiseerde criminaliteit die zegt 'dat er van zulke criminaliteit sprake is indien groepen die primair zijn gericht op illegaal gewin systematisch misdaden plegen met ernstige gevolgen voor de samenleving, en in staat zijn deze misdaden op betrekkelijk effectieve wijze af te schermen, in het bijzonder door de bereidheid te tonen fysiek geweld te gebruiken of personen door middel van corruptie uit te schakelen.'





De wetenschappelijke literatuur is niet eenduidig over het begrip ondermijning, maar geeft wel een nadere inkadering. Tops & Van der Torre (2015) stellen dat ondermijning criminaliteit is die in empirische zin de volgende aspecten met zich meedraagt:

- Aantasting van de gezagspositie van bestuur en politie
- Aangetast vertrouwen onder frontlijnwerkers
- Sluipende maatschappelijke acceptatie van misdaad(geld)
- Aangetaste marktwerking
- Aantasting van instituten die zich richten op legale kansen

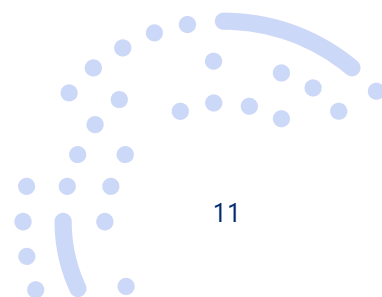
Tegelijkertijd bestaat er gevaar voor inflatie van het onderwerp. Kolthoff & Khonraad (2016) stellen vast dat het begrip ondermijning te pas en te onpas wordt gebruikt voor diverse criminele fenomenen, waarbij het steeds vaker als synoniem wordt gebruikt voor alle georganiseerde misdaad. De auteurs komen met een eerste aanzet tot een nadere duiding van het begrip. Ze stellen dat er sprake is van vier soorten gedragingen die men weliswaar ondermijnd noemt maar die dat lang niet altijd zijn:

- Gedrag dat niet crimineel en niet ondermijnd is
- Gedrag dat wel crimineel is maar niet ondermijnd
- Gedrag dat ondermijnd is maar niet crimineel
- Gedrag dat ondermijnd en crimineel is

In het kader van dit onderzoek zijn we voor wat betreft het element ondermijning voornamelijk geïnteresseerd in gedrag dat past binnen de derde en vierde categorie. Ondermijning wordt dan nader gedefinieerd als:

"Het ontwrichten van de samenhang in wijken en de legale economie, door het op grote schaal systematisch schenden van wettelijke normen, of door intimidatie, omkoping of chantage van burgers, bedrijven, bestuurders of het ambtelijk uitvoerend apparaat."

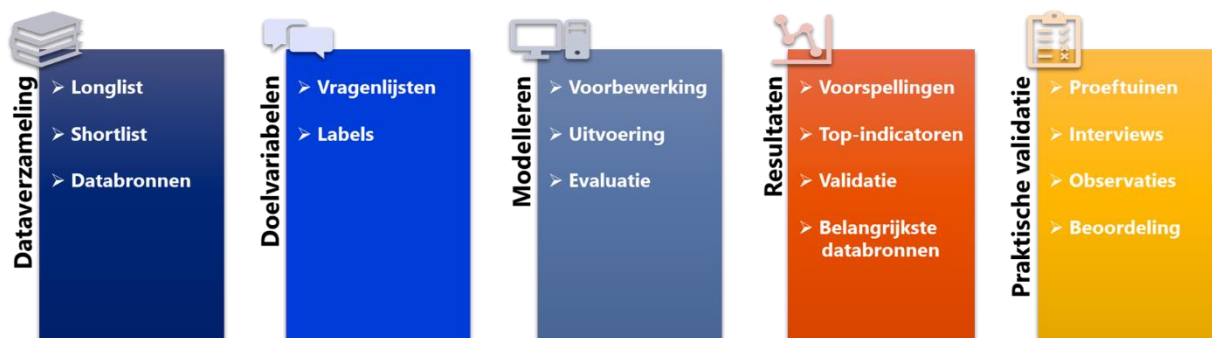
Om een veiligheidsvraagstuk te doorgronden en een aanpak te formuleren zijn inzichten nodig. Bijvoorbeeld over de verschijningsvorm van een fenomeen, de prevalentie in een specifieke geografische omgeving of de wijze waarop er barrières kunnen worden opgeworpen om het fenomeen tegen te gaan. Het voorgestelde onderzoek sluit hierop aan. Een van de belangrijke ontwikkelingen als gevolg van een toename in digitalisering is de wens van gemeenten om datagedreven te werken (Meijer, 2015). Digitalisering maakt het mogelijk te duiden tegen de achtergrond van langere termijn trends in de populatie, op basis van datasets en –stromen waartoe de gemeente toegang heeft of waarvan zij zelf eigenaar is. Digitale data zijn in de huidige praktijk dan ook steeds belangrijker bij het verkrijgen van dit soort inzichten.





3 Plan van aanpak

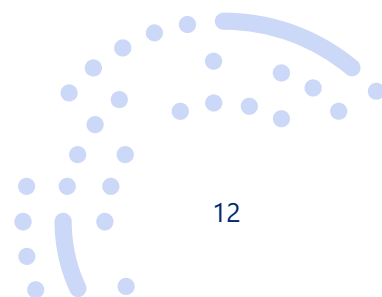
Om de onderzoeksvragen te kunnen beantwoorden via een datagedreven methodiek, is er een plan van aanpak nodig. Hiervoor hebben we een strategie uitgewerkt via een vijfstappenplan, hier weergegeven in **Figuur 1**.



Figuur 1: Plan van aanpak

Dit stappenplan omvat alle benodigde stappen voor het uitvoeren en valideren van de voorspellende modellen. Wat hieruit resulteert zijn handige visuele weergaves, die via een interactief dashboard gepresenteerd worden, en handelingsperspectieven voor de provincies en gemeenten om kwetsbare bedrijventerreinen te kunnen signaleren. Deze twee eindproducten bouwen dus voort op de uitkomsten van het model.

Er is, waar mogelijk, gekozen voor een parallelle aanpak voor het uitvoeren van het stappenplan. Uiteraard zijn hierbij betrokkenheid en inbreng van de opdrachtgevers van cruciaal belang geweest, vooral bij het maken van keuzes tijdens het vooronderzoek, bijvoorbeeld over de relevantie van de risico-indicatoren, en tijdens het dataverzamelingstraject waar er gezamenlijk gekeken is naar de gewenste en beschikbare datasets. Ook bij het uitzetten van de vragenlijsten onder verschillende gemeenten is de inzet van de beleidsmakers van de provincies van wezenlijk belang geweest.





4 Dataverzameling

De dataverzameling is gerealiseerd in twee fasen. In de eerste fase heeft er een vooronderzoek plaatsgevonden waarbij er een longlist is opgesteld van mogelijke risico-indicatoren die van belang zijn voor de voorspellende modellen. Hiervoor zijn er interviews afgenomen met domeinexperts die kennis hebben over criminaliteit en ondermijning op bedrijventerreinen en is eerdere literatuur op dit gebied bestudeerd. Vervolgens is er een (lange) shortlist samengesteld door te kijken naar de haalbaarheid van het bemachtigen van de relevante databronnen. Ook is er gelet op de informatieve waarde van de indicatoren op het uiteindelijk beoogde aggregatieniveau (het niveau van bedrijventerreinen). In de tweede fase is er een gerichte zoektocht gestart naar de relevante databronnen.

4.1 Longlist risico-indicatoren

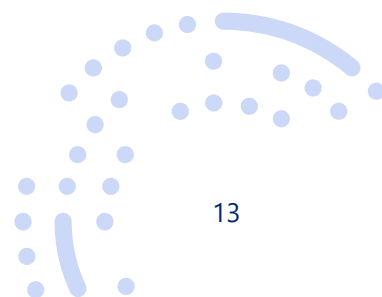
Het eerder aangehaalde onderzoek, uitgevoerd in 2018 in opdracht van het Wetenschappelijk Onderzoek en Documentatiecentrum (WODC), zorgde bij aanvang van het huidige onderzoek voor een basis-indicatorenset.⁴ Het huidige onderzoek verschilt qua omvang en in reikwijdte van het eerder uitgevoerde onderzoek, omdat onder meer specifiek wordt ingezoomd op kwetsbaarheid van bedrijventerreinen aangaande drugscriminaliteit, witwassen, en mensenhandel. Om deze reden heeft een extra vooronderzoek plaatsgevonden. In dit vooronderzoek zijn 21 relevante domeinexperts geïnterviewd. Denk hierbij aan experts ondermijning op bedrijventerreinen, een coördinator synthetische drugs van de politie, een operationeel specialist *human trafficking* van de Regionale Informatie en Expertisecentra (RIEC), en een teamleider opsporing van de Fiscale Inlichtingen en Opsporingsdienst (FIOD). Daarnaast is deskresearch uitgevoerd. Dit richtte zich op het bestuderen van relevante wetenschappelijke publicaties en reeds bestaande documenten die door veiligheidspartners in de praktijk gebruikt worden om op een datagedreven wijze inzicht te krijgen op ondermijning of de genoemde uitingsvormen. Hierbij werden bijvoorbeeld relevante indicatoren verkregen bij programma's zoals Zicht op Ondermijning, en organisaties zoals *Team Follow the Money* van het Functioneel Parket (FP) en verschillende RIEC's.

Met behulp van de beschreven werkwijze is een zogenaamde longlist met indicatoren opgesteld. De focus in deze fase was 'breed' en het ging hierbij vooral om het verzamelen van zoveel mogelijk relevante indicatoren die in een latere fase mogelijk in een voorspellend model opgenomen konden worden. Dit vooronderzoek leidde tot een longlist met in totaal 132 indicatoren. Verdeeld naar de verschillende uitingsvormen komt dit in absolute aantallen neer op verdeling als te zien in **Tabel 2**.

Tabel 2: Longlist risico-indicatoren per ondermijningsvorm

Totaal indicatoren	Drugs	Mensenhandel	Witwassen	Algemeen	Combinatie
132	22	11	35	47	17

⁴ Bolsius et al. 2018; Prüfer & Kolthoff, 2020.





4.2 Shortlist risico-indicatoren

Van 132 unieke risico-indicatoren uit de longlist hebben uiteindelijk 40 de shortlist gehaald. Dit is bereikt in twee iteraties. Eerst is er, samen met de opdrachtgevers, een royale tussenlijst van 70 risico-indicatoren gemaakt na brainstormen over haalbaarheid van indicatoren en databronnen. Er werd hierbij gelet op databronnen die te gevoelig zouden kunnen zijn om te verkrijgen, zoals financiële gegevens van de belastingdienst, banken, of verzekeraars. Ook is er gelet op de beschikbaarheid van data op persoonsniveau die wegens de AVG bemoeilijkt zou worden. Tevens zijn de databronnen die te verspreid beschikbaar zijn, zoals informatie die versnipperd bij de gemeenten ligt, uitgesloten van de shortlist.

Vervolgens is er geprobeerd om zoveel mogelijk van de 70 indicatoren te bemachtigen. Uiteindelijk lukte dit voor 40 indicatoren. De resterende indicatoren konden niet worden verkregen omdat de gecontacteerde dataleveranciers deze om verschillende (geldige) redenen niet konden leveren: er zou te veel handmatig werk nodig zijn om de data te vergaren (bijv. Meld Misdaad Anoniem); de data zouden toch te gevoelig zijn om met anderen te delen (bijv. Inspectieview); en er moesten hoge kosten betaald worden om de data te verkrijgen (bijv. Company.info). Waar mogelijk zijn alternatieve indicatoren die nog wel te maken hadden met de essentie van de longlist indicator gevonden en ingezet.

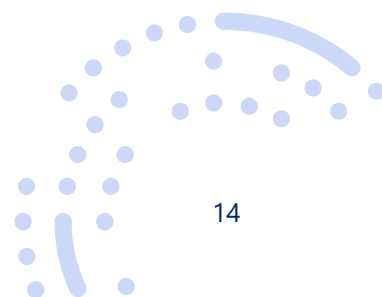
Bij de geshortliste risico-indicatoren ontbreken met name data over financiële gegevens van bedrijven. Ook zijn er geen gegevens op persoonsniveau. De meeste risico-indicatoren die wel zijn bemachtigd zijn voor ondermijningsvorm drugs (16 indicatoren) en de minste voor mensenhandel (5 indicatoren). Het missen van risico-indicatoren kan impact hebben op het voorspellend vermogen van het model. Gezien de hoeveelheid aan verkregen data (risico-indicatoren plus andere relevante variabelen), was de kans op het bouwen van een goed voorspellend model hoog voor ten minste twee van de drie ondermijningsvormen. In **Tabel 3** worden de aantallen en de verhoudingen van de aanvankelijk opgestelde en de uiteindelijk bemachtigde risico-indicatoren per ondermijningsvorm weergegeven.

Naast de drie ondermijningsvormen is er ook een categorie 'algemeen' opgenomen in de tabel. Dit gaat op voor risico-indicatoren die niet specifiek voor een bepaalde ondermijningsvorm gelden, maar die algemeen van toepassing zijn voor ondermijning op bedrijventerreinen. Een voorbeeld hiervan is de aanwezigheid van een parkmanagement.

Tabel 3: Geshortliste risico-indicatoren per ondermijningsvorm

Dataset	Totaal indicatoren	Drugs	Mensenhandel	Witwassen	Algemeen
Longlist	132	40	15	48	46
Shortlist	40	16	5	10	17
Verhouding	30%	40%	33%	21%	37%

De geshortliste risico-indicatoren die meegenomen zijn in de modellen zijn weergegeven in **Bijlage A**.





We beperken ons echter niet alleen tot deze risico-indicatoren voor de voorspellende machine learning-modellen. Een van de risico's bij het uitvoeren van voorspellende modellen is namelijk dat er sprake kan zijn van vooroordelen verankerd in de data of in de selectie van data. De modeluitkomsten zijn immers gebaseerd op de data die erin gestopt worden. Dit betekent dat er een nadelige menselijke voorselectie kan optreden als alleen de risico-indicatoren worden beschouwd. De uitkomsten zijn dan mogelijk niet meer representatief voor de werkelijkheid. In dit onderzoek nemen we daarom veel meer variabelen mee dan alleen de geshortliste risico-indicatoren. Ook de schijnbaar ongerelateerde variabelen zijn meegenomen. Iedere variabele, die natuurlijk wel iets te maken heeft met de doelvariabele, behelst informatie die belangrijk kan zijn voor een goede voorspelling en voegt daardoor iets toe aan het uiteindelijke model. Denk hierbij bijvoorbeeld aan de rechtsvormen van bedrijven, de inschrijvingsredenen bij de Kamer van Koophandel (KVK), de verkoopprijs van grond, of aan de aanwezigheid van een milieu- of geluidszonering op een bedrijventerrein.

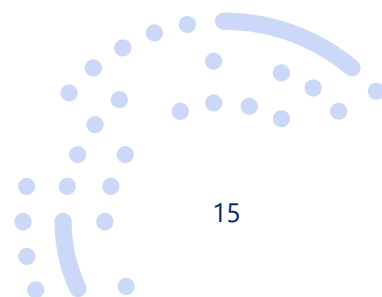
Deze niet-risico-indicator variabelen komen met name uit de databronnen die we al verzamelen om de geshortliste risico-indicatoren te bemachtigen. Een KVK-dataset herbergt al snel vele tientallen variabelen, en veel daarvan bevatten relevante informatie over bedrijventerreinen of vertellen iets over de karakteristieke eigenschappen van bedrijventerreinen. Ook nemen we veel (relevante) data van het Centraal Bureau voor de Statistiek (CBS) mee in de modellen. De online (via StatLine) beschikbare CBS-databestanden zijn makkelijk toegankelijke databronnen die ook nog eens iets vertellen over de regio in en rondom bedrijventerreinen.

4.3 Databronnen vergaren

Het verzamelen van de databronnen is uitgevoerd door eerst meerdere lijsten op te stellen van mogelijke relevante en beschikbare databronnen. Hiervoor zijn brainstormsessies geweest met onder meer de data-analisten van de provincies, omgevingsdiensten, en intern bij Avans en Centerdata. Ook is er aan de hand van de opgestelde longlist een inventarisatie gemaakt van potentiële databronnen. Hiervoor is onder andere de literatuur geraadpleegd.

De opgestelde lijsten zijn samengevoegd tot een uitputtende lijst van mogelijk relevante databronnen waarin we ons hebben verdiept. Na een goed overwogen voorselectie is er contact geweest met verschillende partijen over 44 van deze databronnen. Hiervan zijn uiteindelijk 20 verkregen. Al lijkt het behalen van ongeveer de helft van de databronnen relatief gezien weinig, het gestelde doel van 44 databronnen was al uiterst ambitieus.

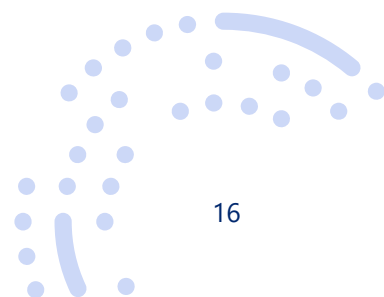
Na een initiële opschoningsslag van variabelen op relevantie hebben we te maken met in totaal maar liefst 663 ruwe variabelen in alle datasets samen. Dit is inclusief de 40 shortlist risico-indicatoren. De ruwe variabelen worden in een later stadium verder bewerkt en opgeschoond, waardoor er een groot aantal nog zal worden uitgesloten van de finale gecombineerde set. Er zal echter ook een aantal nieuwe variabelen worden aangemaakt; dit heet "feature engineering" in data science terminologie. In **Hoofdstuk 6** gaan we hier verder op in, maar uiteindelijk blijven er 310 bruikbare variabelen over voor het model. Meer specifieke bewerkingen die op de variabelen zijn uitgevoerd worden in de opgeleverde codes toegelicht. De complete lijst van 310 variabelen is terug te vinden in de opgeleverde datasets (masterfile) en in het dashboard.





Naast de ruwe variabelen zijn er ook zes beschrijvende variabelen. De beschrijvende variabelen worden met name gebruikt om de verschillende databronnen aan elkaar te koppelen. Deze variabelen zijn: RIN (Rijks Identificerend Nummer), RSIN (Rechtspersonen en Samenwerkingsverbanden Informatienummer), KVK-nummer, gemeentenaam, terreinnaam, en postcode (in dit geval pc6).

De lijst met verkregen databronnen, inclusief het aantal variabelen dat elk databron herbergt, wordt in **Tabel 4** weergegeven. We komen in **Hoofdstuk 6.4** hierop terug door te laten zien welke van deze databronnen als belangrijk naar voren komen aan de hand van de verkregen resultaten van de voorspellende modellen.





Tabel 4: Twintig onafhankelijke databronnen

Nummer	Dataset	Uitleg	Aantal variabelen	Online beschikbaar	Niveau
1	IBIS	Integraal Bedrijventerreinen Informatie Systeem	49	ja	RIN
2	LISA	Landelijk Informatiesysteem van Arbeidsplaatsen	17	nee	RIN
3	BAG	Basisregistratie Adressen en Gebouwen	22	nee	Pand
4	BRK	Basisregistratie Kadaster	9	nee	Perceel
5	KVK	Kamer van Koophandel	90	nee	Bedrijf
6	KVK-regiodata	Met KVK-gegevens van eerdere jaren	73	ja	RIN
7	BVH	Basisvoorziening Handhaving	5	nee	Pc6
8	Netbeheerders Energie	Netbeheerders Enexis en Liander	8	nee	Pc6
9	S@men	Milieu-/overlastmeldingen Omgevingsdiensten	8	nee	Pc6
10	ANBI	Algemeen Nut Beogende Instellingen	3	ja	Bedrijf
11	RVO subsidies	Rijksdienst voor Ondernemend Nederland	7	nee	Pc6
12	Leefbaarometer	Leefbaarheidssituatie en -ontwikkeling	3	ja	Buurt
13	Leegstandsmonitor	Landelijke Monitor Leegstand van min. BZK	6	ja	Buurt
14	Energieleveringen	Energielevering aan woningen en bedrijven (CBS)	10	ja	Pc6
15	Politie misdrijven	Opendata politie via dataportaal	26	ja	Buurt
16	Politie overlastmeldingen	Opendata politie via dataportaal	21	ja	Buurt
17	CBS gezondheid	Via CBS StatLine	84	ja	Buurt
18	CBS uitkeringen	Via CBS StatLine	10	ja	Buurt
19	CBS voorzieningen	Via CBS StatLine	122	ja	Buurt
20	KWB	Kerncijfers Wijken en Buurten (CBS)	90	ja	Buurt





5 Doelvariabelen

Voorspellende machine learning-modellen kunnen op twee manieren worden ingezet om vast te stellen welke bedrijventerreinen kwetsbaarder zijn. De eerste is aan de hand van de clusteringtechniek, een unsupervised ML-techniek, en de tweede is via classificatie, een vorm van supervised machine learning. Meer uitleg over machine learning, supervised machine learning, en andere data science-technieken wordt gegeven in **Bijlage B**.

Om voorspellingen af te geven heeft supervised machine learning echter bijna altijd de voorkeur, omdat de technieken speciaal hierop gericht zijn en tevens ook beter interpreteerbaar zijn. Classificatiemodellen met name leiden tot betrouwbare uitkomsten met hoge nauwkeurigheid. Voor ons doel pakken we de eerste onderzoeksvraag aan door meerdere classificatietechnieken in te zetten.

Een belangrijke vereiste van supervised machine learning is echter dat er *gelabelde* data beschikbaar zijn. Gelabelde data zijn het deel van de data waarvan de uitslag bekend en dus getagd is; bijvoorbeeld bij het voorspellen van fraude beschikt men over de data van bekende fraudegevallen. Het machine learning-model kan hierop worden *getraind*.⁵ In dit geval moeten de labels aangeven wat het risico, oftewel de ingeschatte kans, op ondermijning is op het geselecteerd type van georganiseerde criminaliteit en ondermijning op een bedrijventerrein. Dit hebben we nodig om de kwetsbare bedrijventerreinen te kunnen duiden. Deze risico's vormen onze doelvariabele.

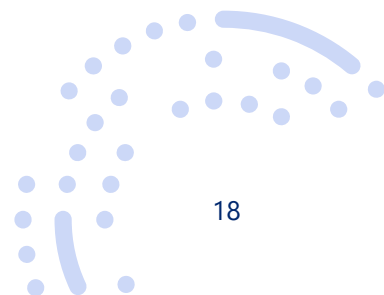
De doelvariabelen worden verkregen door vragenlijsten af te nemen onder experts; mensen die kennis en inzicht hebben van bedrijventerreinen. Het doel is om zoveel mogelijk expertinzichten te vergaren om criminaliteit en ondermijning goed in beeld te krijgen via de collectieve kennis van experts.

5.1 Vragenlijsten

Om de mate van georganiseerde criminaliteit en ondermijning per bedrijventerrein te inventariseren zijn er vragenlijsten in beide provincies uitgezet waarin deskundigen van bedrijventerreinen (experts) hun oordeel konden geven. Denk hierbij aan medewerkers van gemeente, provincie, omgevingsdienst, en politie, en aan parkmanagers, veiligheidsadviseurs, accountmanagers, en mensen die op het gebied van handhaving of beleid werken. Door genoeg 'stemmen' te verzamelen per bedrijventerrein konden we een goed beeld krijgen van het risico op criminaliteit en ondermijning op de bedrijventerreinen.

In totaal zijn 43 gemeenten uit beide provincies geselecteerd om mee te doen aan de steekproef. Bij het selecteren van de gemeenten is er ook gelet op een goede verdeling van de grootte van gemeenten (zowel in oppervlak als in bedrijvigheid) en de geografische locatie van gemeenten (noord vs. zuid en west vs. oost).

⁵ Dit betekent dat het algoritme steeds beter wordt in het vinden van oplossingen, net zoals mensen leren aan de hand van ervaringen.





Er werd gevraagd naar de inschatting en inzicht van deze deskundigen per type ondermijningsvorm (drugs, witwassen, en mensenhandel). De respondenten konden stemmen op een laag, gemiddeld, of hoog risico op ondermijnende criminaliteit. In **Bijlage C** is een voorbeeld te zien van de uitgezette vragenlijsten. De verkregen response op de vragenlijsten is weergegeven in **Tabel 5**.

Tabel 5: De verkregen response van de uitgezette vragenlijsten onder experts

Provincie	Gemeenten	Bedrijventerreinen (BT)	Totaal stemmen	BT's met min. 3 stemmen	BT's met min. 5 stemmen
NB	22	369	2247	155	52
GLD	21	216	1641	112	78
Totaal	43	585	3888	267	130

Verhoudingsgewijs is er iets meer response verkregen vanuit provincie Gelderland. Meer specifiek bedraagt het gemiddeld aantal stemmen voor een bedrijventerrein in Gelderland 7,6 (1641/216), terwijl dit voor bedrijventerreinen in Noord-Brabant, met een gemiddelde van 6,1, (2247/369) meer dan een hele stem lager ligt.⁶

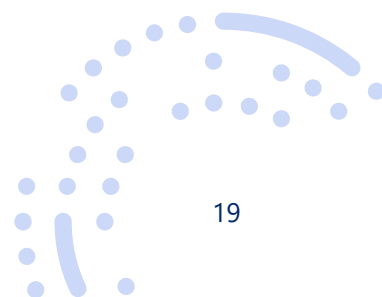
5.2 Labels

Aan de hand van gelabelde bedrijventerreinen kan er een model worden gebouwd dat een voorspelling maakt van de kans op ondermijnende criminaliteit op alle andere (niet gelabelde), huidige of toekomstige, bedrijventerreinen. Hiermee kan de eerste hoofdvraag worden beantwoord. Ook wordt het mogelijk om via het machine learning-model te achterhalen wat de belangrijkste indicatoren zijn die tot de voorspelling hebben geleid. Deze informatie kan worden gebruikt om de tweede hoofdvraag te beantwoorden (zie **Hoofdstuk 1.2** voor beide hoofdvragen).

Het doel is om aan de hand van zoveel mogelijk expertinzichten ondermijning in beeld te krijgen via de collectieve wijsheid van experts. In **Figuur 2** wordt de route geïllustreerd waarmee de geselecteerde bedrijventerreinen gelabeld worden en daarmee dus een beoordeling krijgen op kwetsbaarheid voor de drie typen ondermijnende criminaliteit.

Een belangrijke bevinding is dat na veel onderzoek de beoordeling 'gemiddeld' onbetrouwbare resultaten bleek op te leveren. Er zijn vergaande pogingen gedaan om de stem 'gemiddeld' alsnog te kunnen gebruiken, zoals toekenning via supermajority (minimaal 67% stemmen), en minimum verschil van 1, 3, of 5 stemmen in combinatie met een minimum aantal stemmen voor een valide labeling. Alle pogingen leidden tot slechte modelvoorspellingscores. Ook een verdere inspectie van het voorspellend vermogen van het model liet een negatieve trend zien. Ten overvloede kwamen de toegekende risico's in 48% van de gevallen niet overeen met eerder onderzoek⁴.

⁶ Het gemiddeld aantal stemmen per bedrijventerrein geldt voor alle drie de ondermijningsvormen. Dus de stemmen van de drie ondermijningsvormen zijn bij elkaar opgeteld.



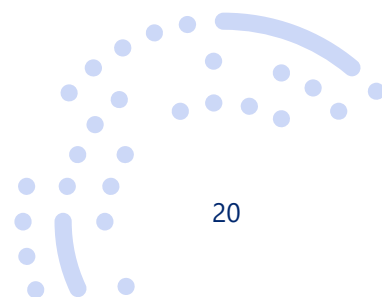


Figuur 2: Route voor het toekennen van risico's aan bedrijventerreinen

De keuze is daarom gemaakt om door te gaan met het toekennen van de kwetsbaarheid van een bedrijventerrein op alleen 'hoog' en 'laag'. De stem 'gemiddeld' is daarvan uitgezonderd. Dit gaf een veel beter beeld met voorspellingscores rond de 80-90%, een positief trend in voorspellend vermogen, en een overeenstemming van 92% met eerder onderzoek. In **Hoofdstuk 6.1** gaan we hier verder op in.

Het uiteindelijke model is iets minder genuanceerd; het voorspelt nu alleen hoge en lage kwetsbaarheid in plaats van hoog, gemiddeld, en laag. Echter, het model is veel beter in het onderscheiden van terreinen met een hoge kwetsbaarheid, met daarbij ook de mogelijkheid om de inschatting van de kans op hoge kwetsbaarheid mee te geven. In theoretische zin maakt het weglaten van de categorie 'gemiddeld' het model veel betrouwbaarder.

We richten ons alleen op de risicobeoordelingen 'hoog' en 'laag' voor bedrijventerreinen. De beoordeling 'gemiddeld' is uitgesloten van de labeling vanwege de lagere betrouwbaarheid.





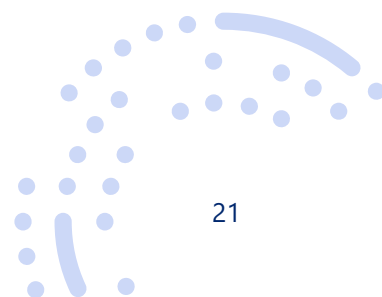
Labels toekennen aan een bedrijventerrein gaat echter gepaard met bepaalde afwegingen. Inzetten op hogere kwaliteit labels door de bedrijventerreinen te kiezen met meer stemmen gaat ten koste van het aantal gelabelde data, wat weer ten nadele is van de het voorspellend vermogen van het uiteindelijke model. Dus zowel de kwaliteit (sterke consensus van experts, aangeduid door de hoeveelheid stemmen per label) als de kwantiteit (hoeveelheid gelabelde bedrijventerreinen) van de labels is belangrijk voor het uiteindelijke model.

We vinden een balans door de bedrijventerreinen te selecteren die een verschil van minimaal drie stemmen hebben tussen 'hoog' en 'laag'.⁷ Zo houden we nog genoeg gelabelde data over van hoge kwaliteit. Het uiteindelijke resultaat is dat voor 173 bedrijventerreinen risico's zijn toegekend met een valide en betrouwbare label van hoog of laag risico. In **Tabel 6** geven we de cijfers van de labeling weer.

Tabel 6: Labels toekennen aan bedrijventerreinen

Provincie	Gemeenten	Bedrijventerreinen	Valide labels 1+ verschil	Valide labels 3+ verschil	Verhouding BT's met label (3+)
NB	22	369	130	85	23 %
GLD	21	216	104	88	41 %
Totaal	43	585	234	173	30 %

⁷ Andere methoden die uitvoerig getoetst zijn onder andere: Marjority voting, supermajority voting (67% stemmen), en 5 verschil (surplus voting). Het eerste van deze drie methoden gaf een te lage kwaliteit label en de laatste twee resulteerden in te weinig labels.





6 Analyses en resultaten

6.1 Voorspellingsscores

Voor de risicotaxatietool schatten we de kans op georganiseerde criminaliteit en ondermijning op alle bedrijventerreinen in de provincies Gelderland en Noord-Brabant op basis van een groot aantal indicatoren. We hebben hiervoor drie verschillende, meest geschikte voorspellende machine learning-modellen ingezet. De opzet, (voor)bewerkingsstappen, en verantwoording van de modellen worden in **Bijlage D–Bijlage G** in detail uitgelegd. We runnen deze machine learning-modellen op alle drie de ondermijningstypen: drugs, witwassen, en mensenhandel. We doen dit ook voor ondermijnende criminaliteit in het algemeen door het samenvoegen van alle drie de ondermijningsvormen.

De voorspellingsscores geven een indicatie van hoe betrouwbaar de modeluitkomsten zijn. Dit is een onderdeel van de theoretische validatie van het model. Aan de hand van de voorspellingsscores kiezen we het beste model voor ons doel. Het best presterende model wordt geselecteerd door de voorspellingsscores van alle ondermijningsvormen te middelen.

Ook worden de modellen getoetst op voorspellend vermogen via verschillende technieken. Een model kan in uitzonderlijke gevallen een relatief hoge voorspellingsscore afgeven, maar toch geen voorspellend vermogen bevatten.⁸ Hiervoor kunnen specifieke tests uitgevoerd worden. Aan de hand daarvan kunnen we aangeven of een ondermijningsvorm uiteindelijk wel of niet voorspelbaar is.

In **Tabel 7** geven we de voorspellingsscores van de drie modellen aan. Meer detail en scores van de andere prestatiepunten worden weergegeven in **Bijlage H**.

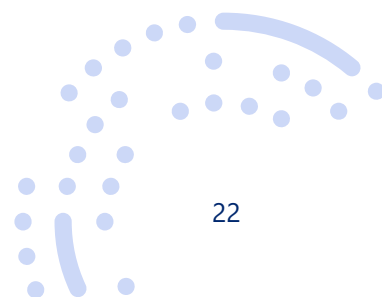
Tabel 7: Voorspellingsscores van de drie machine learning-modellen

Model	F ₁ -score Drugs	F ₁ -score Witwassen	F ₁ -score Mensenhandel	F ₁ -score Algemeen
Random Forest	72 %	75 %	74 %	72 %
Logistic Regression	78 %	78 %	68 %	72 %
Gradient Boosting	90 %	80 %	83 %	76 %

De voorspellingsscores laten duidelijk zien dat voor alle ondermijningsvormen Gradient Boosting beter presteert dan Random Forest of Logistic Regression. Daarom selecteren we Gradient Boosting als ons best presterende model, welke gebruikt wordt voor het doen van inschattingen en voorspellingen.

Gradient Boosting is het meest geschikte algoritme om kwetsbare bedrijventerreinen op te sporen door het risico op ondermijnende criminaliteit te voorspellen aan de hand van data.

⁸ Bijvoorbeeld bij zeer ongebalanceerde data of bij hoge mate van "overfitting".





Het blijkt ook dat drugsriminaliteit veel beter te voorspellen is dan de andere ondermijningsvormen. De F_1 -score van 90% overtreft de scores van zowel witwassen als mensenhandel. Dit heeft grotendeels te maken met de beschikbaarheid van data voor drugsriminaliteit. Zoals te zien is in **Tabel 3** zijn er veel meer risico-indicatoren beschikbaar voor drugscriminaliteit dan voor de andere ondermijningsvormen. Ook de niet-risico-indicatoren bevatten meer drugsgerelateerde informatie.

De kwaliteit van de labels is ook van invloed op het verschil in voorspellingsscores. Dit heeft weer te maken met de aanwezigheid en zichtbaarheid van de criminaliteitsvormen. Bij drugscriminaliteit zijn er veel meer bedrijventerreinen gelabeld met een risico 'hoog' terwijl er bij mensenhandel er heel weinig labels zijn met risico 'hoog'. De negatieve invloed van (minder goede) labelkwaliteit is terug te zien in de test op voorspellend vermogen.

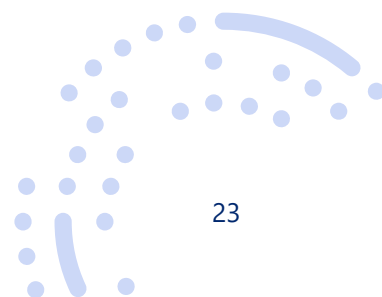
Voorspellend vermogen

We testen het voorspellend vermogen van de modellen op twee manieren. We kijken ten eerste naar de hoogte van de trainingset accuracy, in combinatie met het verschil in de training- en testset accuracies. Deze beoordeling geeft een indicatie of een model aan het overfitten is. Als de trainingset accuracy namelijk te hoog is, of het verschil in de accuracies is te groot, dan kan er sprake zijn van overfitting. Overfitting vindt plaats wanneer een model de details en ruis in de trainingset leert, in die mate dat dit een negatieve invloed heeft op de prestaties van het model op nieuwe data. Dit betekent dat de ruis of willekeurige fluctuaties in de trainingsdata worden opgepikt en als concepten worden geleerd door het model. Dit is ook te zien in de tabellen van **Bijlage H**. Gezien de accuracies lijken onze modellen voor de ondermijningsvorm mensenhandel te overfitten.

Een tweede (betere) manier is door te kijken of de voorspellingsscores veranderen in het geval van een verandering in de grootte van de dataset. We variëren het aantal observaties van de data van 1% tot 100% en trainen de modellen met de veranderende trainingsetgroottes. Als het model voorspellend vermogen heeft, dan zou er een (stijgende) verandering moeten optreden in de voorspellingsscores. Zou de voorspellingsscore constant blijven, dan hebben de data geen impact op het model en dus 'leert' het model niet. We illustreren dit in **Bijlage I**.

Aan de hand van deze twee tests blijkt dat onze modellen geen voorspellend vermogen hebben voor de ondermijningsvorm mensenhandel. Voor de andere twee ondermijningsvormen en ondermijning algemeen is dit wel het geval. Zoals eerder ook gezegd zijn de twee hoofdredenen hiervoor dat er niet genoeg data over mensenhandel beschikbaar zijn waardoor het model niet genoeg input heeft om te leren en dat de inschatting van de experts onvoldoende zijn omdat mensenhandel niet voldoende zichtbaar of aanwezig is. Dat laatste blijkt ook uit het feit dat er te weinig 'hoog' risico labels zijn voor mensenhandel voor de gekozen sample van bedrijventerreinen (maar 6 van de 138 stemmen).

Modellen voor mensenhandel blijken geen voorspellend vermogen te hebben omdat er te weinig risico-indicatoren (data) en te geringe labels (expertmeningen) zijn. Om deze redenen wordt mensenshandel niet meegenomen in de risicotaxatietool.





6.2 Top-indicatoren

Voor ieder machine learning-model en ondermijningsvorm hebben we een lijst met belangrijkste indicatoren die tot de voorspelling leiden. De indicatoren zijn gerangschikt op mate van belangrijkheid en zijn relatief geschaald ten opzichte van elkaar. Uiteindelijk aggregeren we de drie machine learning-modeluitkomsten per ondermijningsvorm om model onafhankelijkheid te garanderen.

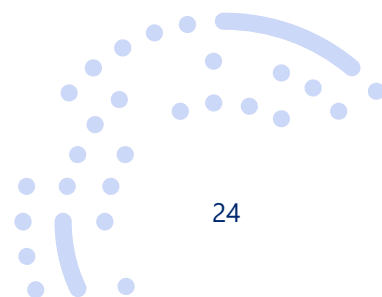
In **Figuur 3** laten we de top-20 indicatoren zien voor de ondermijningsvormen drugs en witwassen, en ondermijning in het algemeen. We zetten de indicatoren op alfabetische volgorde ook uit in **Tabel 8**.

Voor een deel van de 40 top-indicatoren uit de Risicotaxatietool is het mogelijk een criminologische verklaring te geven voor de samenhang van de indicator (of het kenmerk) met een verhoogde kwetsbaarheid voor ondermijning. Deze indicatoren zijn vaak ook beïnvloedbaar (zie ook Handreiking Handelingsperspectieven Risicotaxatietool Bedrijventerreinen). Er is getracht de samenhang van de top-indicatoren met ondermijnende criminaliteit te duiden. Dit is gedaan met behulp van inzichten uit bestaande criminologische theorieën en aan de hand van inzichten uit eerder uitgevoerde onderzoeken naar ondermijnende criminaliteit op bedrijventerreinen. Daarnaast is geput uit de verzamelde data uit interviews die tijdens het vooronderzoek en gedurende de praktische validatie uitgevoerd zijn.

Belangrijk voor de interpretatie van de indicatoren is dat zij niet onafhankelijk zijn van elkaar. De combinatie van indicatoren bepaalt de voorspelde kwetsbaarheid van het bedrijventerrein als geheel, en dus niet van specifieke branches of ondernemingen. Wij noemen deze samenhang van indicatoren een "giftige cocktail". Ook wordt er geen onderscheid gemaakt tussen bedrijventerreinen die (enkel) kwetsbaar zijn, en terreinen waar ondermijning zich reeds manifesteert.

Voor een deel van de top-indicatoren uit het Risicotaxatietool kan geen directe verklaring worden gegeven vanuit bestaande criminologische theorie, eerder uitgevoerde onderzoeken of inzichten uit het door ons uitgevoerde onderzoek. Dit geldt voor in totaal 20 van de top-indicatoren, die volgens het ontwikkelde model wel degelijk een voorspellende waarde hebben voor de verhoogde kwetsbaarheid van een bedrijventerrein ten aanzien van ondermijnende criminaliteit. Aanvullend kwalitatief onderzoek zal nodig zijn om inzicht te krijgen in de achterliggende mechanismen van de samenhang van deze vooraf 'onverwachte' indicatoren van ondermijnende criminaliteit.

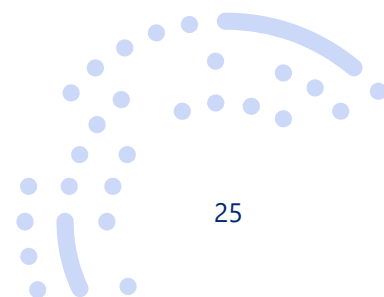
In **Bijlage J** worden de top-indicatoren toegelicht en waar mogelijk van een duiding voorzien.

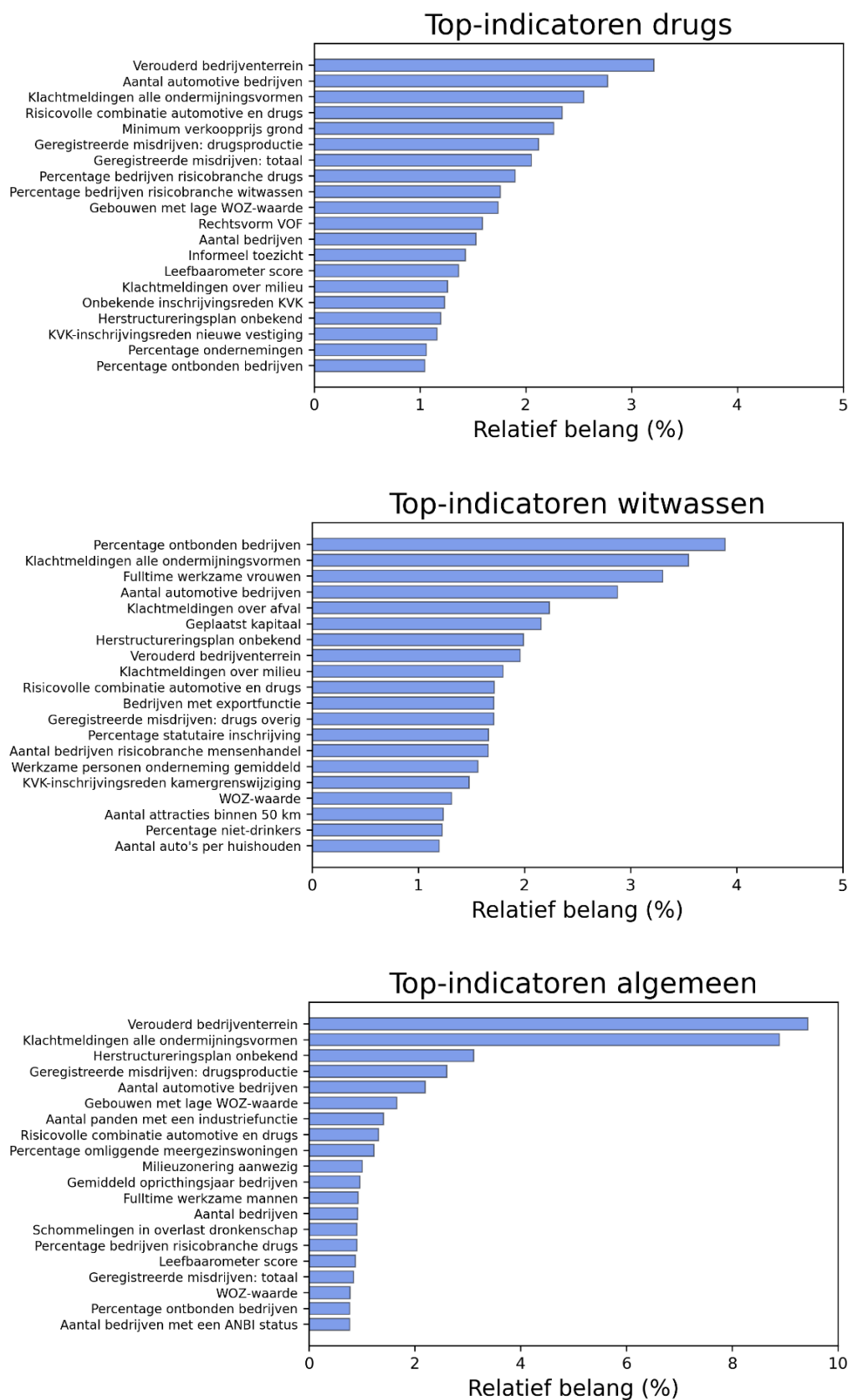




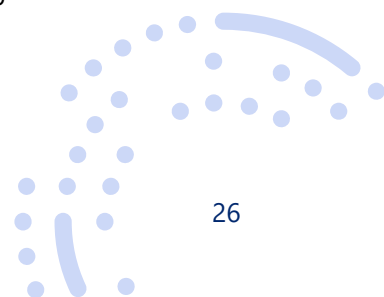
Tabel 8: Lijst met top-indicatoren die voorkomen in de drie top-20 lijsten

Indicator	Algemeen	Drugs	Witwassen
Aantal attracties binnen 50 km			✓
Aantal automotive bedrijven	✓	✓	✓
Aantal auto's per huishouden			✓
Aantal bedrijven	✓	✓	
Aantal bedrijven met een ANBI-status	✓		
Aantal bedrijven risicobranche mensenhandel			✓
Aantal panden met een industriefunctie	✓		
Bedrijven met exportfunctie			✓
Fulltime werkzame mannen	✓		
Fulltime werkzame vrouwen			✓
Gebouwen met een lage WOZ-waarde	✓	✓	
Gemiddeld oprichtingsjaar bedrijven	✓		
Geplaatst kapitaal			✓
Geregistreerde misdrijven: drugs overig			✓
Geregistreerde misdrijven: drugsproductie	✓	✓	
Geregistreerde misdrijven: totaal	✓	✓	
Herstructureringsplan onbekend	✓	✓	✓
Informeel toezicht		✓	
Klachtmeldingen alle ondermijningsvormen	✓	✓	✓
Klachtmeldingen over afval			✓
Klachtmeldingen over milieu		✓	✓
KVK-inschrijvingsreden kamergrenswijziging			✓
KVK-inschrijvingsreden nieuwe vestiging		✓	
Leefbaarometer score	✓	✓	
Milieuzonering aanwezig	✓		
Minimum verkoopprijs grond		✓	
Onbekende inschrijvingsreden KVK		✓	
Percentage bedrijven risicobranche drugs	✓	✓	
Percentage bedrijven risicobranche witwassen		✓	
Percentage niet-drinkers			✓
Percentage omliggende meergezinswoningen	✓		
Percentage ondernemingen		✓	
Percentage ontbonden bedrijven	✓	✓	✓
Percentage statutaire inschrijving			✓
Rechtsvorm VOF		✓	
Risicovolle combinatie automotive en drugs	✓	✓	✓
Schommelingen overlast dronkenschap	✓		
Verouderd bedrijventerrein	✓	✓	✓
Werkzame personen onderneming gemiddeld			✓
WOZ-waarde	✓		✓





Figuur 3: Top-20 indicatoren om 'hoog' risico te voorspellen voor de ondermijningsvormen





Richtingen van de top-indicatoren

Om ook zicht te krijgen op de richting van de invloed van de variabelen op modelvoorspellingen hebben we aanvullend extra modellen (Logistic Regression) gedraaid op de top-indicatoren. Een positief relatie betekent dat meer/hogere waarde leidt tot hoger risico op criminaliteit en ondermijning op bedrijventerreinen. Een negatief relatie betekent dat minder/lagere waarde leidt tot hoger risico.

Als voorbeeld: Meer klachtmeldingen op een bedrijventerrein leiden tot hoger risico op ondermijning. Een lager leefbaarometer score leidt ook tot hoger risico op ondermijning.

In **Figuur 4** geven we de richtingen van de top-indicatoren aan per ondermijningsvorm, inclusief ondermijning in het algemeen.

6.3 Aanvullende theoretische validatie

Naast de voorspellingsscores en de twee toetsen op voorspellend vermogen van de modellen hebben we een aanvullende theoretische validatie gedaan. We passen het best presterende model eerst toe op alle bedrijventerreinen in de twee provincies en krijgen als uitkomst alle risico's terug. Hiermee kunnen we de kwetsbare bedrijventerreinen identificeren. Voor het model zijn dit compleet nieuwe (ongelabelde) data, en dus niet te onderscheiden van toekomstige data.

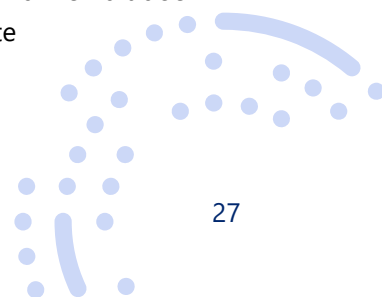
Hierna doen we een gezonde verstand controle door te kijken of de verwachte verhoudingen (hoog versus totaal) die we hebben van de gelabelde data overeenkomen met de verhoudingen van alle voorspelde data. Dit geven we weer in **Tabel 9**.

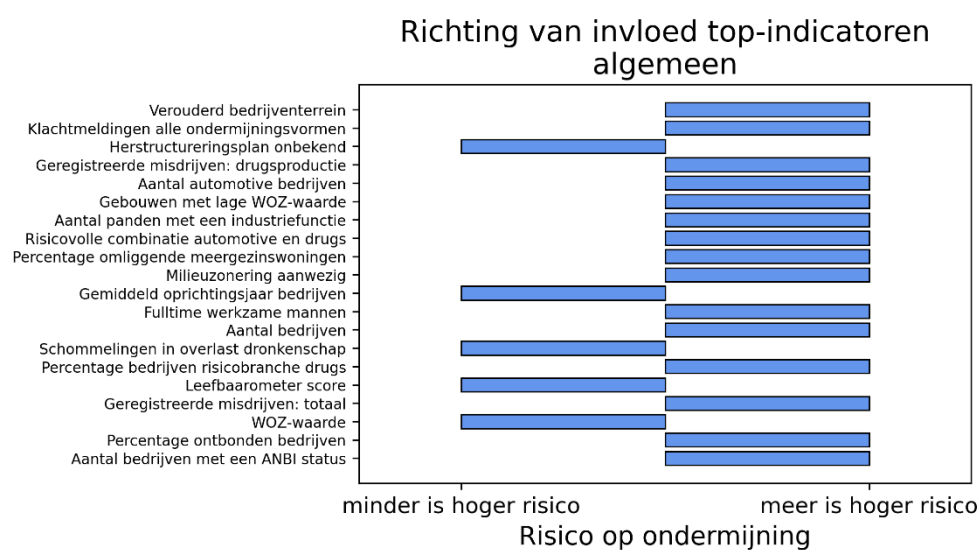
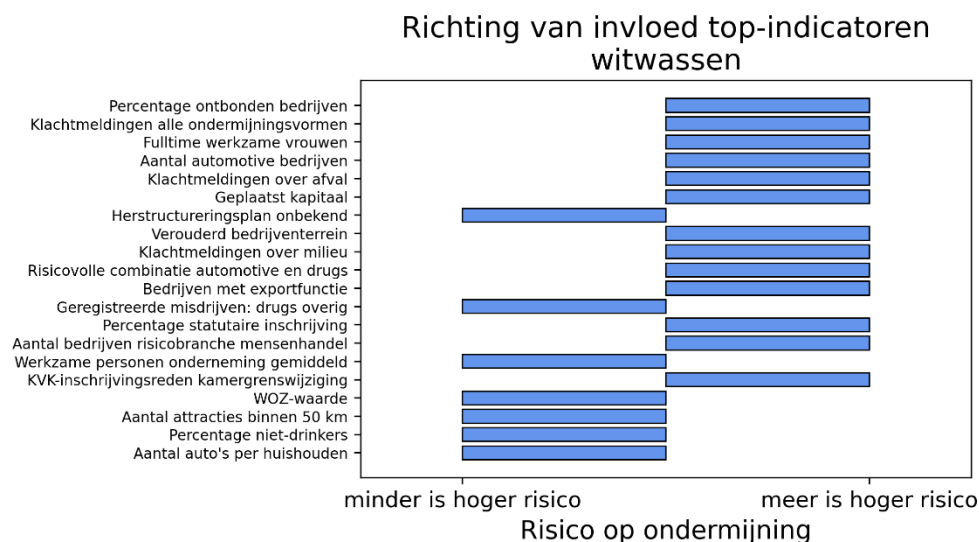
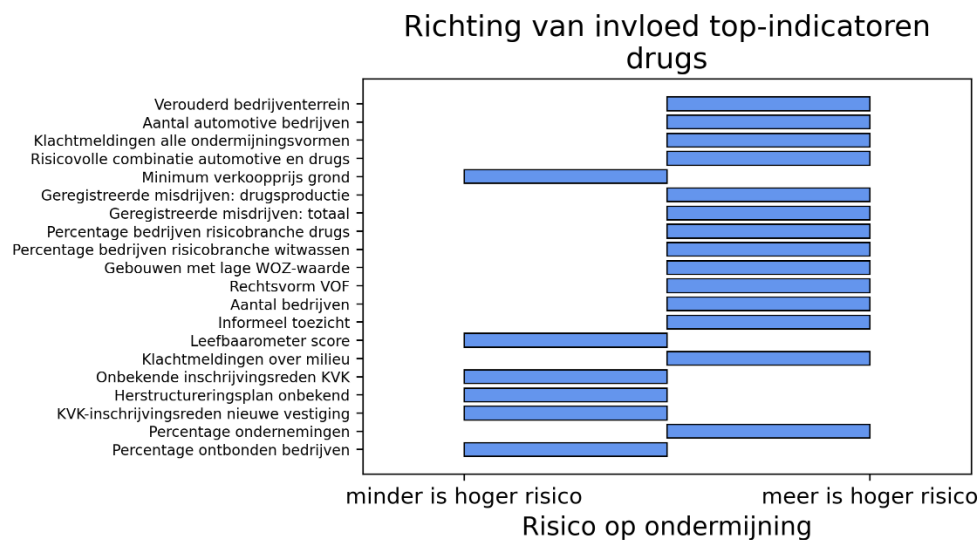
Tabel 9: Extra validatie van het model via verhoudingen van de (voorspelde) klassen

Type ondermijning	Labels			Voorspelling		
	Aantal BT's laag risico	Aantal BT's hoog risico	Verhouding hoog	Aantal BT's laag risico	Aantal BT's hoog risico	Verhouding hoog
Drugs	64	40	0.38	696	313	0.31
Witwassen	59	20	0.25	794	215	0.21
Algemeen	132	41	0.24	834	175	0.17

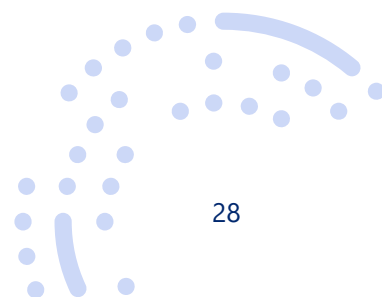
Het is te zien dat de verhoudingen van de voorspelde risico's redelijk goed overeenkomen met de verwachte verhoudingen. Ook via deze methode blijkt dat ondermijningsvorm mensenhandel niet goed voorspeld wordt door het model (dit wordt niet weergegeven in de tabel). De andere ondermijningsvormen worden wel goed voorspeld.

Een opmerking is echter dat de verhoudingen van de voorspellingen consistent iets lager uitvallen dan de verhoudingen van de gelabelde data. Dit kunnen we enigszins kalibreren door de classificatiegrens van ons model bij te stellen. Standaard wordt namelijk bij een voorspelde waarschijnlijkheid van 50% of hoger de klasse 'hoog' risico aangenomen (en anders 'laag'). Door deze grens een beetje lager bij te stellen kan ervoor gezorgd worden dat de verhoudingen beter overeenkomen. We kunnen dit doen in het dashboard door de kleur die samenhangt met de hoogte van de risico iets bij te stellen.





Figuur 4: Richtingen van invloed van de top-indicatoren voor verschillende ondermijningsvormen





6.4 Belangrijke databronnen

Nu we de top-indicatoren hebben, kijken we uit welke bronnen deze indicatoren afkomstig zijn. Zo zien we welke bronnen als de belangrijkste databronnen kunnen worden beschouwd. Gebaseerd op het aantal indicatoren dat in de top-20 lijsten voorkomt, hebben we een tabel van belangrijke, minder belangrijke, en onbelangrijke databronnen opgesteld. De lijsten zijn weergegeven in **Tabel 10**.

Tabel 10: De belangrijke en minder belangrijke databronnen voor het model

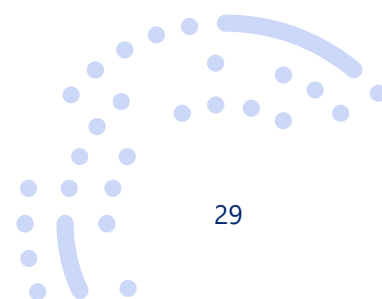
Heel belangrijk	Minder belangrijk	Niet belangrijk
IBIS	Netbeheerders energie	BRK (Kadaster)
LISA	ANBI	KVK-regiodata
BAG	Leefbaarometer	Leegstandsmonitor
KVK	CBS gezondheid	Energieleveringen
BVH	CBS voorzieningen	Politie overlastmeldingen
S@men	Politie misdrijven	CBS uitkeringen
KWB		RVO subsidies

De belangrijkste databronnen hebben twee of meer variabelen die in de top-20 lijsten voorkomen. De minder belangrijke hebben maar één variabele en van de niet belangrijke databronnen zijn er geen variabelen te vinden in de top-20 lijsten.

We merken echter op dat de kwaliteit van de dataset vaak de hoofdoorzaak is waarom een dataset niet of minder belangrijk blijkt te zijn. Kwaliteit van data gaat hier met name om de compleetheid van data. Er zijn bijvoorbeeld veel bedrijventerreinen (ongeveer 35%) die niet opgenomen zijn in de KVK-regiodata. De leegstandsmonitor heeft ook een lage overlap met de lijst van bedrijventerreinen (met missings voor 36% van de bedrijventerreinen). Ook zijn er veel missings in de variabelen van de energieleveringen (variërend van 10% tot 35%). Van de percelengegevens in de BRK-data heeft ongeveer 45% geen postcode of een andere koppelingsvariabele waardoor ze niet meegenomen kunnen worden in het model. Terwijl ongeveer 37% van de RVO- subsidies dataset alleen informatie heeft van de verleende subsidie postcodes op pc4-niveau in plaats van het gewenste pc6-niveau (wegens AVG). Hetzelfde geldt in mindere mate voor de gegevens van netbeheerders energie. De CBS uitkeringen en gezondheid datasets evenals de overlastmeldingen bij de politie bevatten meer indirecte variabelen die met name iets vertellen over de buurten rondom bedrijventerreinen.

Ook al zijn er soms mogelijke aannames te maken om het aantal ontbrekende gegevens te verminderen, bijvoorbeeld door imputatie met mediaan, toch resulteert dit in informatieverlies en is daarom de hoofdreden voor het niet of minder belangrijk zijn van de dataset voor het model.

Een kwalitatief mindere dataset of variabele kan nog altijd een positieve impact hebben op het model. Het voorspellend vermogen van het model wordt desalniettemin versterkt, al is het in mindere mate.





6.5 Interpretieren van de uitkomsten

Enkele van de belangrijke vragen die spelen zijn: Hoe kunnen we de resultaten van voorspellende modellen duiden? Hoe betrouwbaar zijn de uitkomsten en wat zeggen deze ons nu eigenlijk? Op welke topindicatoren moeten we letten? Het voorgaande hoofdstuk gaat in op de technische implicaties van de uitkomsten, maar biedt nog te weinig houvast voor bestuurders en beleidsmakers. Dit hoofdstuk tracht een antwoord te geven op bovenstaande vragen.

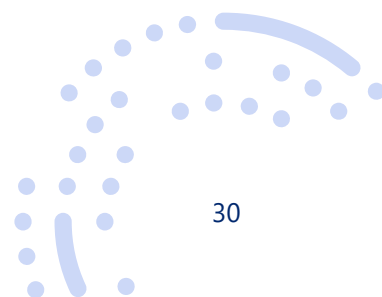
Als eerste kunnen we aan de hand van de voorspellingsscores concluderen dat de modellen voor het inschatten van druggerelateerde ondermijning, witwassen, en ondermijning in het algemeen betrouwbaar zijn. Van de door het model ingeschatte risico's komt respectievelijk 90, 80 en 76 procent overeen met het oordeel van de collectieve kennis van experts. De latente kennis van experts is dan ook goed gevangen door het model, welke kan worden ingezet om een betrouwbare inschatting te maken van het risico op ondermijning voor niet gelabelde bedrijventerreinen.

Daarbij dient de kanttekening geplaatst te worden dat een hoog risico niet automatisch betekent dat er ondermijning plaatsvindt op een bedrijventerrein. Het betreft een *risico*, wat iets zegt over de kwetsbaarheid van een bedrijventerrein. Een hoog risico betekent dat een terrein, vergeleken met andere terreinen, een hoger risico loopt op het ontwikkelen van ondermijnende criminaliteit.

Naast het doen van een betrouwbare voorspelling zijn we ook geïnteresseerd in welke indicatoren bijdragen aan een verhoogd risico op ondermijning. Per ondermijningsvorm zijn de 20 meest invloedrijke indicatoren geselecteerd. Deze indicatoren bestaan uit een combinatie van factoren die betrekking hebben op omgeving, onderneming, en informatiepositie. Zo kijkt het model niet alleen naar wat er op een bedrijventerrein gebeurt, maar ook hoeveel er bekend is over een bedrijventerrein, of wat er in de buurt rondom een bedrijventerrein aan de hand is.

Wanneer je aan de slag wilt met het verkleinen van de kwetsbaarheid van een bedrijventerrein let je eerst op de belangrijkste top-indicatoren. Het is hierbij belangrijk om te weten wat de relatie is van de indicator met kwetsbaarheid, maar ook hoe het bedrijventerrein op deze indicator scoort. Het dashboard kan daarbij helpen om meer inzicht te krijgen. Vooral die indicatoren waarop het bedrijventerrein slecht scoort zijn interessant, omdat dit mogelijke verbeterpunten zijn die de kwetsbaarheid kunnen verkleinen. Denk hierbij bijvoorbeeld aan een verloedering van een bedrijventerrein.

Daarbij is het belangrijk om de invloed van topindicatoren niet los van elkaar te zien. Het model kijkt namelijk naar een combinatie van indicatoren, een giftige mix, bij het doen van een inschatting. Een indicator op zichzelf hoeft niet direct de reden te zijn voor een hoog ingeschat risico op kwetsbaarheid. Het kan bijvoorbeeld zijn dat juist de combinatie van veel automotieve bedrijven en een lage grondprijs voor een verhoogd risico zorgt.





7 Praktische validatie

7.1 Opzet en organisatie

Nadat een lijst met meetbare en voorspellende indicatoren uit het model is opgesteld, is een praktische validatie van de indicatoren uitgevoerd. Met de uitkomsten wordt in de praktijk gecontroleerd of de voorspellende indicatoren zeggingskracht hebben, ofwel worden herkend in de praktijk situatie. Deze validatie is uitgevoerd door studenten van de minor 'Bestuurlijke aanpak van ondermijning' van de opleiding Bestuurskunde van Avans Hogeschool 's-Hertogenbosch in een tijdsbestek van tien weken, onder begeleiding van onderzoekers van het lectoraat Ondermijning. In totaal zijn 47 studenten verdeeld in tien groepjes van vier of vijf studenten. Deze zijn vervolgens willekeurig toegewezen aan een proeftuingemeente, waar zij vervolgens twee bedrijventerreinen nader hebben onderzocht. Aan de hand van gestructureerde interviews en observaties zijn de volgende opdrachten achtereenvolgens uitgevoerd:

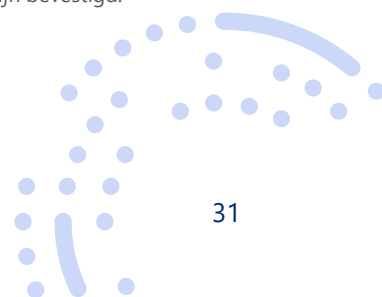
1. Aangeven en onderbouwen welke van de toegewezen bedrijventerreinen respectievelijk het meest en het minst kwetsbaar zijn voor ondermijnende criminaliteit. Dit geldt mede als controle op de eerder beschreven labeling.
2. De validatie van de 15 belangrijkste voorspellende indicatoren uit het model.⁹

7.2 Werkwijze

Selectie proeftuinen

Vanuit de provincie zijn proeftuingemeenten aangewezen waar de praktische validatie van de uitkomsten uit het model uitgevoerd zou gaan worden. Er is hierbij bewust gekozen om gemeenten met verschillende karakteristieken ten aanzien van de omvang, demografische opbouw en ligging te betrekken in het onderzoek. Uiteindelijk zijn per provincie vijf proeftuingemeenten aangewezen voor deelname aan het project. Aan de hand van de resultaten uit de labeling zijn in iedere gemeente twee bedrijventerreinen gekozen als proeftuin. Hierbij ging het steeds om één hoog-risicoterein en één laag-risicoterein. Waar dit niet lukte – vanwege het ontbreken van een als hoog- of laag risico gelabeld terrein – is in plaats hiervan een terrein met een labelscore 'gemiddeld' gekozen. Hiervoor zijn de expertbeoordelingen 'gemiddeld' alleen ter validatie weer teruggehaald. Dit leidde tot de verdeling per provincie weergegeven in twee tabellen, **Tabel 11** en **Tabel 12**.

⁹ Dit waren de top-15 indicatoren voor de uitkomstvariabele 'ondermijning algemeen' omdat hierin beide uitingsvormen samenkwamen. Verfijning van het model na de praktische validatie heeft er toe geleid dat een deel van de indicatoren in de initiële lijst van 15 belangrijkste indicatoren zijn gewisseld ten opzichte van de uiteindelijke top-20. Dit heeft geen invloed gehad op indicatoren die zich leenden voor validatie in de praktijk, en die tijdens de praktische validatie ook zijn bevestigd.





Tabel 11: Proeftuingemeenten Gelderland

Gemeente	Label ondermijning algemeen		
	hoog	gemiddeld	laag
Apeldoorn	✓		✓
Arnhem	✓		✓
Barneveld	✓		✓
Overbetuwe	✓		✓
Wijchen	✓	✓	

Tabel 12: Proeftuingemeenten Noord-Brabant

Gemeente	Label ondermijning algemeen		
	hoog	gemiddeld	laag
's-Hertogenbosch	✓	✓	
Moerdijk	✓		✓
Roosendaal	✓		✓
Valkenswaard		✓	✓
Waalwijk	✓		✓

Gestructureerde interviews

In de tien proeftuingemeenten zijn (gestructureerde) interviews afgenomen met experts. Deze experts waren voornamelijk professionals met kennis van de veiligheidssituatie op de geselecteerde bedrijventerreinen. Het ging hierbij onder andere om Ambtenaren Openbare orde en Veiligheid (AOV'ers) en medewerkers van de afdelingen Economische Zaken van gemeenten. Daarnaast zijn professionals met bij voorkeur operationele functies geïnterviewd, zoals expert-wijkagenten, beveiligingsmedewerkers en parkmanagers. Deze zijn over het algemeen veel aanwezig op het bedrijventerrein en zijn hierdoor relatief gezien goed op de hoogte van de situatie ter plaatse.

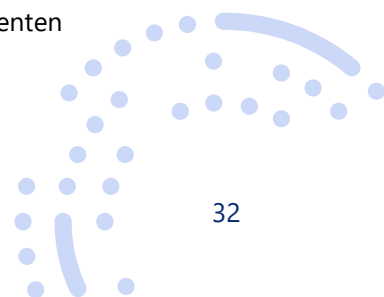
In totaal zijn 48 gestructureerde interviews afgenomen. Hierbij is gebruik gemaakt van vragenlijsten waarin achtereenvolgens de volgende vragen per indicator zijn gesteld.

- Hangt de betreffende indicator samen met een verhoogd risico op ondermijning?
- Zo ja, wat is de richting van dit verband: positief of negatief?
- Waarom hangt dit kenmerk samen met ondermijning op bedrijventerreinen?
- Voor welke uitingsvormen geldt dit in het bijzonder?

Bovenstaande vragen richtten zich expliciet op de twee aangewezen bedrijventerreinen in de proeftuingemeenten. In het laatste deel van het interview is steeds ruimte genomen om door te vragen op mogelijke verklaringen die de domeinexperts gaven voor het door hen gegeven oordeel. Alle interviews zijn getranscribeerd en nagelopen op overeenstemming met de conclusies die studenten hebben getrokken uit bovenstaande vier vragen.

Observaties

Naast de uitgevoerde interviews zijn door de studenten observaties uitgevoerd op de aangewezen bedrijventerreinen. Op ieder bedrijventerrein zijn een of meerdere observatiemomenten geweest. Hierbij is gebruik gemaakt van een webapplicatie, waarin specifieke





indicatoren gerelateerd aan ondermijning zijn opgenomen. Deze indicatoren wijken af van de indicatoren uit systeemdata, die gebruikt zijn in de te ontwikkelen risicotaxatietool. Het ging hierbij met name over indicatoren die met behulp van observaties ter plaatse gemeten kunnen worden en die volgens eerdere onderzoeken geassocieerd kunnen worden met de aanwezigheid van ondermijnende criminaliteit op bedrijventerreinen. Hierbij kan gedacht worden aan indicatoren die iets zeggen over de fysieke kwaliteit van het betreffende terrein, de aanwezigheid van formele of informele sociale controle, de aanwezigheid van risicobranches en de mate van bedrijvigheid op het bedrijventerrein.

7.3 Uitkomsten

Onderscheid hoog- en laagrisico terreinen

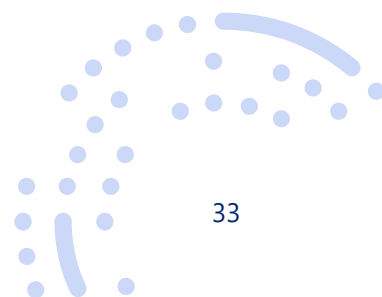
Op basis van de uitgevoerde interviews en observaties is allereerst door de tien groepen studenten beoordeeld in welke mate de door hen bezochte bedrijventerreinen kwetsbaar zijn voor de aanwezigheid van ondermijnende criminaliteit. Zoals hierboven beschreven, waren de studenten niet ingelicht over de uitkomsten uit de labeling. Hen is medegedeeld dat er een verschil in verwachte kwetsbaarheid tussen de terreinen aanwezig was, maar niet hoe deze terreinen voorafgaand aan de praktische validatie door praktijkprofessionals is beoordeeld in de labeling. Het onderscheid tussen hoog- en laag-risicoterreinen werd evenwel zonder problemen in alle 20 terreinen door de studenten herkend. Hierbij kan worden opgemerkt dat door studenten ook nuances bij de kwetsbaarheid van (ons bekende) middelhoog gelabelde terreinen zijn geplaatst, bijvoorbeeld door kwetsbaarheidsverminderende kenmerken van de terreinen mee te nemen in de beoordeling.

Beoordeling indicatoren

De 15 belangrijkste indicatoren zijn op basis van interviews en observaties getoetst aan de praktijk, door middel van verschillende niveaus van herkenning. Dit houdt in dat is onderzocht in hoeverre deze variabelen – waar mogelijk – ook van invloed waren op de onderzochte terreinen, hoe deze variabelen elkaar beïnvloeden (richting) en is gezocht naar een theoretische verklaring¹⁰ voor dit verband. Over het verband met de verschillende uitingvormen doen wij hier geen uitspraken, omdat de ‘top-indicatoren’ zich concentreerden rondom de uitkomstvariabele ‘ondermijning - algemeen’.

- 1) Een deel van de indicatoren was afkomstig uit het vooronderzoek en de theoretische verklaring voor de richting van het verband waren bekend en ‘scoorden’ zoals verwacht hoog: deze werden vrijwel unaniem bevestigd als relevant. De variabelen ‘veroudering van een bedrijventerrein’, ‘aantal automotive bedrijven’ en ‘gemiddelde WOZ waarde’ hadden volgens de respondenten een voorspellende waarde voor de kwetsbaarheid van een terrein, en zij konden deze correlatie ook beargumenteren. De observeerbare kenmerken van deze indicatoren werden ook door de studenten bevestigd naar aanleiding van de eigen observatie. Een afwijkende hoog-scorende variabele was ‘onbekendheid over de aanwezigheid van een herstructureringsplan voor het gebied’. Hoewel een groot deel van de respondenten dit kenmerk als relevant beschouwden, bleek uit de richting van het verband en argumentatie dat deze indicator niet viel te valideren via de methodiek van interviews en observaties.

¹⁰ De theoretische duiding van de top-indicatoren uit het definitieve model is te vinden in **Bijlage J**.



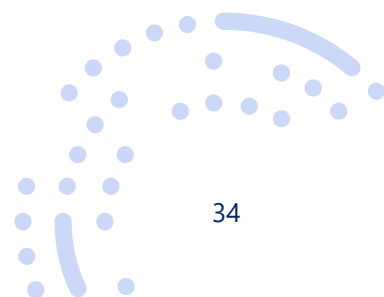


- 2) Een midden-categorie van indicatoren als 'netto oppervlakte van een bedrijventerrein' en 'aantal overlastmeldingen bij de politie' werden eveneens als zeer relevant beoordeeld. Hierbij ontstonden echter in enkele gevallen vraagtekens over de richting van het verband. Zo konden overlastmeldingen volgens sommige respondenten wijzen op de aanwezigheid van sociale controle op een terrein, en zou het daarom een negatief verband hebben met de kwetsbaarheid voor ondermijning. Volgens anderen zijn overlastmeldingen een teken van onregelmatigheden die veelal duiden op achterliggende ondermijningsproblematiek.
- 3) Een ander deel van de indicatoren is niet eenduidig beoordeeld en kon daarom niet bevestigd noch ontkracht worden tijdens de praktische validatie. Ongeveer evenveel respondenten gaven aan dat variabelen als 'afstand tot voorzieningen zoals winkel, museum, zwembad en huisarts' en 'aantal hectare oppervlaktewater binnen een bedrijventerrein' een voorspellende waarde hebben voor de kwetsbaarheid van een terrein, als dat er respondenten waren die zeiden dat deze juist géén voorspellende waarde hebben.
- 4) Voor een deel van de indicatoren bleek validatie in de praktijk door middel van interviews en observaties niet goed mogelijk. Dit waren veelal variabelen die door het model waren 'ontdekt' en niet afkomstig waren uit de longlist aan risico-indicatoren. Specifieke eigenschappen van terreinen, zoals 'aantal parttime werkenden op een bedrijventerrein' en 'het absolute aantal fulltime werkende mannen op een bedrijventerrein' hadden geen verband met ondermijning en/of waren moeilijk van duiding te voorzien. Respondenten gaven hierbij ook veelal aan het niet te weten en hier geen antwoord op te kunnen geven. Het was ook niet mogelijk om voor deze terreinen deze indicatoren op individueel niveau te valideren met systeemdata.

Uit de praktische validatie van de indicatoren kwam naar voren dat indicatoren die zich hier voor leenden (empirisch waarneembaar, of een hoog aantal respondenten met praktijkervaring op dit vlak) ook zijn bevestigd als relevant voor het voorspellen van de kwetsbaarheid van bedrijventerreinen voor (de aanwezigheid van) ondermijnende criminaliteit. Naar mate de indicatoren zich minder goed leenden voor de praktische validatie, ontstond ook meer onduidelijkheid over de aanwezigheid van, de richting van en verklaringen voor verbanden.

De uitkomsten van het model hebben mogelijk implicaties voor theorievorming over relevante variabelen om de kwetsbaarheid van bedrijventerreinen te voorspellen. Interessant aan de variabelen is dat bij een deel hiervan correlaties deels door het model zijn 'ontdekt'; de gelegde verbanden waren onbekend en ontbeerden vooralsnog een theoretische verklaring. Waar mogelijk is een theoretische verklaring geformuleerd over de correlatie tussen de variabelen en de kwetsbaarheid van de bedrijventerreinen. Dit heeft geen nieuwe inzichten opgeleverd. Zoals beschreven werden enkele variabelen niet bevestigd, maar ook niet ontkracht in de praktijk. Het model heeft hiermee mogelijk verbanden blootgelegd die nader onderzoek behoeven.

De praktische validatie (blind uitgevoerd) kwam overeen met de toegekende risico's (labels) bij alle 20 beoordeelde bedrijventerreinen. De praktische validatie kwam grotendeels overeen met de top-indicatoren van het model. De richtingen van invloed waren daarbij vaak intuïtief.





8 Dashboard

Het dashboard is, naast het rapport en de handelingsperspectieven op ondermijning, een van de eindproducten van de risicotaxatietool. Het dashboard maakt de resultaten van de voorspellende analyses op een laagdrempelige en interactieve manier inzichtelijk. Het doel van het dashboard is tweeledig, namelijk:

1. Het risico op ondermijning op bedrijventerreinen op gemeentelijk niveau inzichtelijk maken door risicogebieden in kaart te brengen.
2. Het identificeren van individuele bedrijventerreinen met een verhoogde relatieve kwetsbaarheid. Hierbij worden ook de indicatoren met een grote invloed op het ingeschatte risico gepresenteerd.

Het dashboard is in de eerste plaats een signalerende tool: het biedt een overzicht van de mate van kwetsbaarheid van bedrijventerreinen en geeft een indicatie van de mogelijke oorzaken van verhoogde kwetsbaarheid. Het idee is dat bestuurders en beleidsmakers op basis van preventieve signalen de relatieve kwetsbaarheid van bedrijventerreinen ten aanzien van ondermijnende criminaliteit in de regio kunnen beoordelen.

Om de geïdentificeerde kwetsbaarheid daadwerkelijk te kunnen verminderen is een theoretisch risicomodel alleen niet toereikend. Het kan inzicht bieden in de combinatie van risicofactoren die een belangrijke rol spelen en geeft samen met de opgestelde handelingsperspectieven een startpunt. In de praktijk zijn vervolgens verdere stappen nodig om voor een bepaald terrein in gezamenlijk overleg de specifieke situatie in kaart te brengen.

Hoe gebruik ik het dashboard?

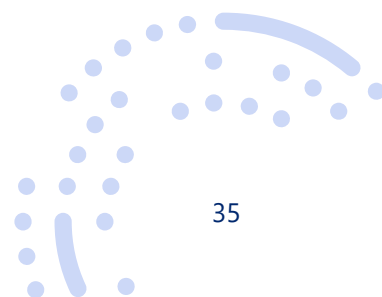
Het dashboard is opgesplitst in twee onderdelen, elk gepresenteerd op een apart tabblad. Het eerste tabblad biedt informatie op gemeenteniveau en kan gebruikt worden om de kwetsbare bedrijventerreinen in de gemeente te identificeren. Vervolgens kunnen kwetsbare bedrijventerreinen in detail bekeken worden aan de hand van het tweede tabblad.

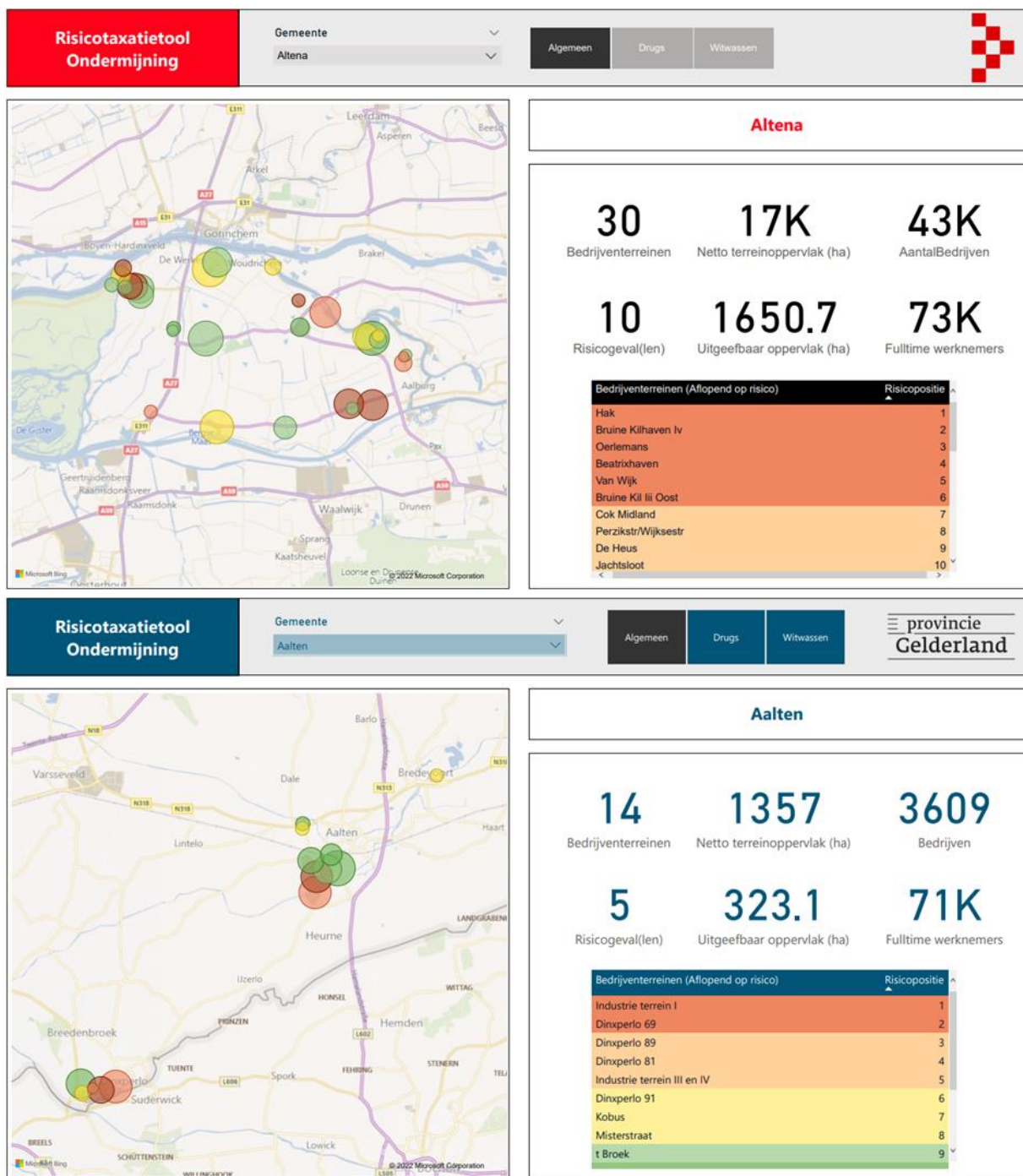
Tabblad 'Gemeente'

Het eerste tabblad bestaat uit een aantal visuele componenten, namelijk:

- Header
- Kaart
- Geaggregeerde informatie
- Ranking van bedrijventerreinen op basis van ingeschatte kwetsbaarheid

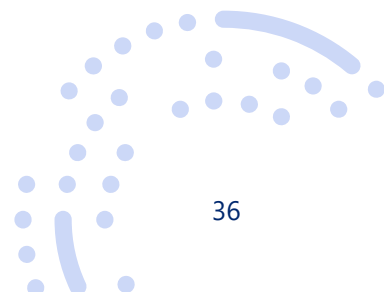
Zie **Figuur 5** voor screenshots van het tabblad 'Gemeente' van de dashboards voor zowel provincie Noord-Brabant als Gelderland.





Figuur 5: Screenshots van het eerste tabblad van het dashboard "Risicotaxatietool" voor provincies Noord-Brabant en Gelderland. Let op: voor deze screenshots zijn de waarheidsgetrouwe waarden vervangen door fictieve waarden. Dit in verband met gevoeligheid

Met behulp van de elementen in de header kan een specifieke gemeente geselecteerd worden, en is het tevens mogelijk om te wisselen tussen de verschillende ondermijningsvormen. Vervolgens worden de risico-inschattingen voor individuele bedrijventerreinen gepresenteerd als cirkels op het kaartelement. De kleur van deze cirkels komt overeen met de ingeschatte risicocategorie, terwijl de grootte een indicatie geeft van het totale oppervlak van een terrein.





Aan de rechterzijde is geaggregeerde informatie te vinden die meer inzicht geeft in de bedrijvigheid in een gemeente. Denk hierbij aan het aantal bedrijventerreinen, de totale grootte van bedrijventerreinen en hoe veel mensen er werkzaam zijn op deze bedrijventerreinen.

Tot slot bevat dit tabblad een ranking van alle bedrijventerreinen, gebaseerd op de inschatting van relatieve risicoscores. Voor elk bedrijventerrein wordt een inschatting gemaakt van de risicoscore, welke relatief is aan alle andere bedrijventerreinen met bekende risico categorieën (door labeling van experts). Aan de hand van deze ranking kun je snel de bedrijventerreinen identificeren die extra aandacht behoeven.

Tabblad "Bedrijventerrein"

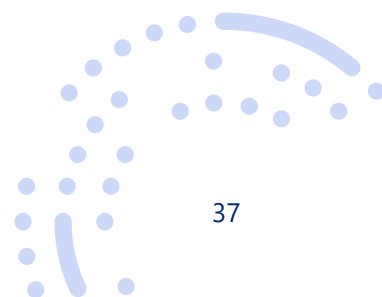
Op het tweede tabblad vind je informatie over specifieke bedrijventerreinen door gebruik te maken van de volgende componenten:

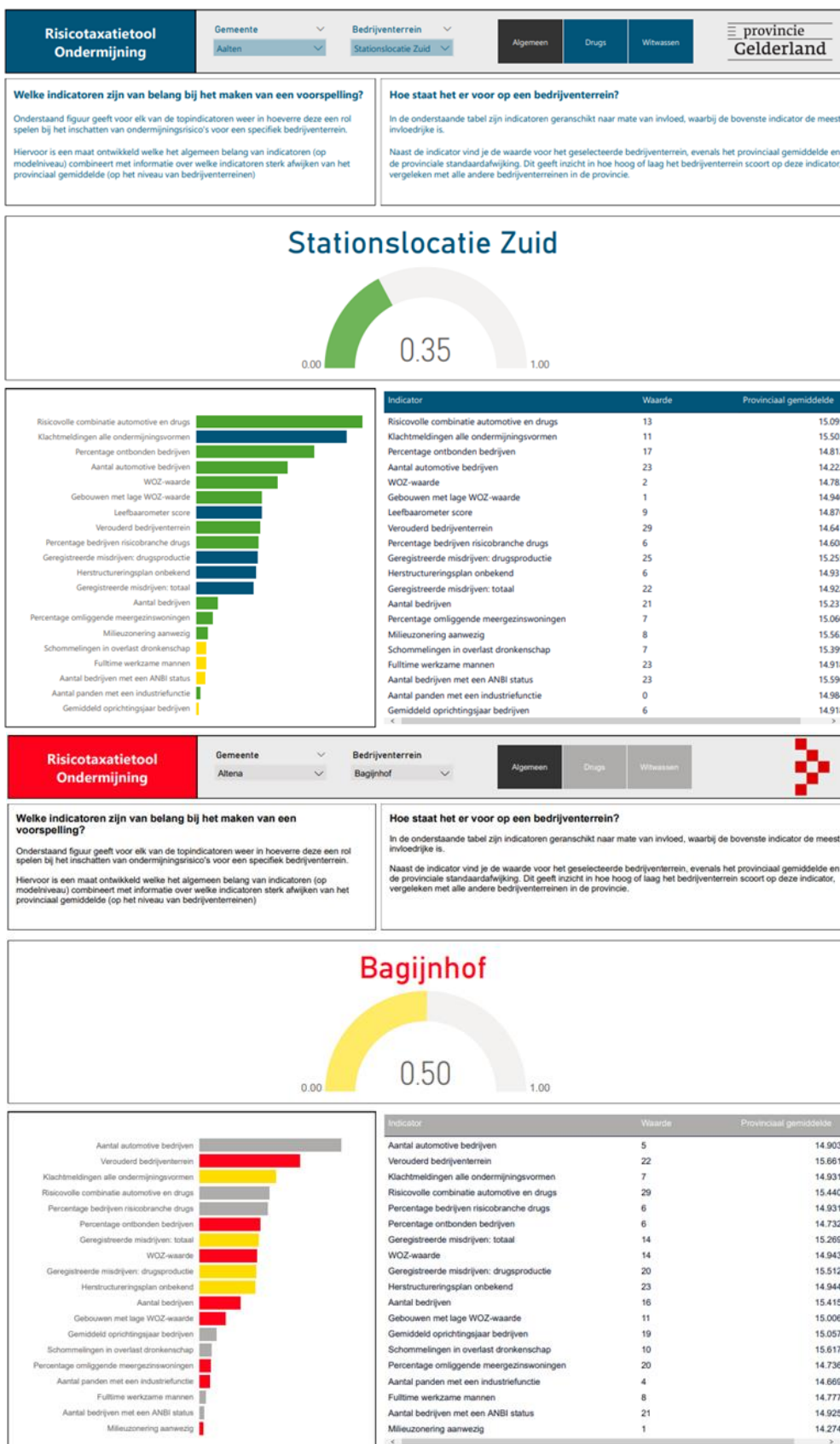
- Header
- Risico-inschatting
- Belangrijke indicatoren en waarden

Zie **Figuur 6** voor screenshots van het tabblad 'Bedrijventerrein' van de dashboards voor zowel provincie Gelderland als Noord-Brabant.

Wanneer een bedrijventerrein geselecteerd is laat de graadmeter de ingeschatte risicoscore zien. Deze risicoscore kan worden geïnterpreteerd als een 'kans' op een verhoogde kwetsbaarheid. Hoe hoger de graadmeter, des te groter de kans dat het bedrijventerrein extra kwetsbaar is voor het geselecteerde type ondermijning. Het staafdiagram toont de meest belangrijke indicatoren en laat voor een bedrijventerrein zien welke indicatoren de grootste invloed hebben gehad op het geschatte risico op kwetsbaarheid. Deze invloed is gebaseerd op een combinatie van het algemene belang van een indicator en de specifieke afwijking van een bedrijventerrein op deze indicator. Let op: Hoewel de indicatoren individueel gepresenteerd zijn, is het altijd een **combinatie** van factoren die zorgt voor de ingeschatte kwetsbaarheid.

Tot slot kun je de tabel aan de rechterzijde gebruiken om te bekijken hoe het geselecteerde bedrijventerrein scoort op de topindicatoren. Hierbij is ook het provinciaal gemiddelde toegevoegd, zodat je weet of het bedrijventerrein hoog of laag scoort op een bepaalde indicator.





Figuur 6: Screenshots van het tweede tabblad van het dashboard "Riscotaxatietool" voor provincies Gelderland en Noord-Brabant. Let op: voor deze screenshots zijn de waarheidsgetrouwe waarden vervangen door fictieve waarden. Dit in verband met gevoeligheid



9 Samenvatting handelingsperspectieven

Om tot een antwoord op de tweede hoofdvraag (Welke handelingsperspectieven hebben provincies en gemeenten om iets aan daadwerkelijk kwetsbare bedrijventerreinen te doen?) te komen hebben wij onderzocht welke handelingsperspectieven het ontwikkelde dashboard voor provincies en gemeenten oplevert voor de aanpak van ondermijnende criminaliteit op bedrijventerreinen¹¹. Uitgangspunt is dat gemeenten zelf aan de slag moeten kunnen met het beeld dat uit de risicotaxatietool naar voren komt. De provincie heeft primair een signalerende taak en de inzichten uit het dashboard worden gebruikt om gemeenten te informeren en te activeren aangaande de aanpak van zogenaamde kwetsbare bedrijventerreinen. Het is nadrukkelijk niet de bedoeling dat de provincie een regierol gaat nemen met betrekking tot de aanpak van ondermijnende criminaliteit. Het provinciale handelingsperspectief omschrijft de taken die logischerwijs voortkomen uit gebruik van het dashboard. Dit zijn:

1. het *informer*en van gemeenten over de resultaten en het wijzen op mogelijkheden om een lokaal informatiebeeld van een specifiek terrein op te stellen;
2. het *agender*en op lokaal en regionaal niveau van de in kaart gebrachte ondermijningsproblematiek en;
3. het *aanjagen* van de aanpak door het maken van concrete afspraken met gemeenten en hen in contact brengen met relevante partners in de veiligheidszorg en bestaande initiatieven.

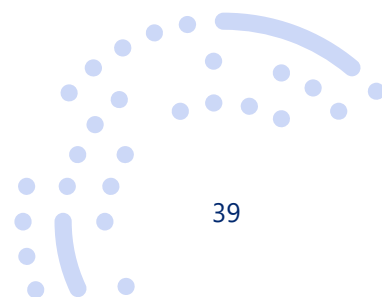
Gemeentelijke handelingsperspectieven focussen zich op het verbeteren van de eigen informatiepositie en het initiëren van concrete interventies.

9.1 Informeren

De provincies willen gemeenten helpen meer zicht te krijgen op kwetsbare bedrijventerreinen. De ontstane veiligheidsbeelden uit het dashboard kunnen op initiatief van de provincie, of op aanvraag van een individuele gemeente, met gemeenten gedeeld worden. De provincie kan gemeenteambtenaren met een taakaccent veiligheid op bedrijventerreinen – bijvoorbeeld een ambtenaar Economische Zaken of een adviseur Openbare orde en Veiligheid – in een gesprek informeren over de uitkomsten uit het dashboard. Op (sub-)regionaal niveau is het mogelijk om individuele beelden te delen om deze te vergelijken, bijvoorbeeld op indicatorniveau, om regionale samenwerking in de aanpak te stimuleren.

De uitkomsten uit het dashboard worden door de provincie in een pdf-bestand beschikbaar gesteld aan individuele gemeenten. De provincie vervult op deze wijze haar signalerende en informerende taak. De inzichten uit het dashboard vormen de aanleiding voor een gesprek tussen relevante ambtenaren van de provincie en eerder genoemde taakaccenthouders en domeinexperts van de gemeente.

¹¹ De volledige Handreiking Handelingsperspectieven Risicotaxatietool wordt samen met dit eindrapport opgeleverd aan de provincies.





In dit gesprek moet ook aandacht zijn voor de diagnose ten aanzien van individuele 'hoog-risico' bedrijventerreinen, waaraan de gemeente haar prioriteit zou kunnen geven bij de aanpak van ondermijnende criminaliteit.

Omdat het dashboard relatief weerbare gebieden onderscheidt van relatief kwetsbare gebieden, is het zaak om voorzichtigheid te betrachten met het delen van de gegenereerde beelden. Het risico bestaat dat negatieve aandacht voor een terrein bestaande positieve ontwikkelingen verstoren. Gemeenten krijgen daarom enkel in samenspraak met de provincie inzage in hun 'eigen' terreinen en medewerkers van de provincie helpen de gemeentefunctionarissen met de interpretatie van de resultaten uit het dashboard, door de uitkomsten van context te voorzien. Het gesprek wordt vervolgens als input gebruikt voor de mogelijke oplossingsrichtingen waarin een gemeente zich kan gaan bewegen.

Deelgebieden

Uit het dashboard komen indicatoren of kenmerken naar voren die op zichzelf, of in combinatie met andere kenmerken, een voorspellende waarde hebben voor de kwetsbaarheid van een bedrijventerrein. Voor elke set van kenmerken geldt dat zij met het deelgebied waartoe zij behoren gekoppeld kunnen worden aan specifieke handelingsperspectieven voor gemeenten. De risicotaxatietool bedrijventerreinen is in deze context vooral een *prioriteringstool*. Een deel van deze indicatoren of kenmerken heeft een beïnvloedbaar karakter, welke onder te verdelen zijn in een drietal deelgebieden:

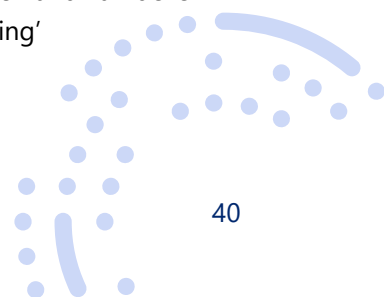
A. De *informatiepositie op een bedrijventerrein* betreft een categorie indicatoren bestaande uit data uit officiële registraties, zelfrapportages en (klacht)meldingen. Denk hierbij aan geregistreerde misdrijven, score op de Leefbaarometer en klachtmeldingen over ondermijningsvormen. Hoewel deze indicatoren niet direct een handelingsperspectief voor de provincie opleveren, is het belangrijk dat de provincie de gemeente wijst op het belang van het versterken van de eigen informatiepositie. Uit de praktische validatie van de indicatoren blijkt dat de inzichten uit het dashboard slechts een algemeen beeld schetsen over de daadwerkelijke situatie op een individueel bedrijventerrein.

B. De *omgevingskenmerken*, die iets zeggen over de ligging, de fysieke kenmerken of de fysieke kwaliteit van het bedrijventerrein of het daarop aanwezige vastgoed. Denk hierbij aan de gemiddelde WOZ-waarde, de minimum verkoopprijs van de grond of veroudering van het bedrijventerrein.

C. De *ondernemingskenmerken*, die samenhangen met de aanwezigheid van risicobranches en bepaalde criminogene of kwetsbaarheid-verhogende kenmerken van ondernemingen op het terrein. Denk hierbij bijvoorbeeld aan een oververtegenwoordiging van de autobranche op een terrein, een hoog aandeel van kwetsbare rechtsvormen zoals een vennootschap onder firma (VOF) of de aanwezigheid van relatief veel bedrijven met ANBI-status.

9.2 Agenderen

De provincie communiceert over de algemene inzichten uit het dashboard en agendeert de aanpak van de ondermijningsproblematiek op (sub-)regionaal niveau en bij gemeenten. Aan de hand van deze inzichten kan de provincie sturen op de programmering van het thema 'ondermijning' en de planning van de aanpak daarvan op bedrijventerreinen. Door dit thema als





'speerpunt' te benoemen kan extra aandacht gegenereerd worden voor het stimuleren van de aanpak op regionaal en lokaal niveau. De provincies zijn momenteel samen met de regio's en de gemeenten bezig om kwalitatief gestuurde afspraken te maken ten aanzien van de aanpak van bedrijventerreinen. In deze overleggen kan de voorgestelde aanpak, gestuurd door het dashboard, geagendeerd worden. De huidige regionale gesprekstafels en het overleg werklocaties bieden daarbij een mogelijkheid tot afstemming en gezamenlijke besluitvorming ten aanzien van de aanpak van ondermijnende criminaliteit. Gezien de aandachtspunten die naar voren komen uit het dashboard, dienen de afspraken zich te richten op de drie beïnvloedbare kenmerken: het vergroten van de informatiepositie, het verbeteren van de fysieke kwaliteit en het stimuleren van een gezond ondernemersklimaat op de hoog-risico terreinen. Hierbij kan gedacht worden aan afspraken aangaande het vestigingsbeleid ten aanzien van bedrijventerreinen en afspraken over de fysieke inrichting en onderhoud van reeds bestaande en toekomstig te ontwikkelen bedrijventerreinen. Hierbij hebben individuele gemeenten ook een duidelijke rol, bijvoorbeeld vanuit de afdelingen Openbare Orde en Veiligheid en/of Ruimtelijke Ordening en Economie. In die zin dient de agendering van de aanpak van ondermijning op bedrijventerreinen ook plaats te vinden op het niveau van de Commissaris van de Koning en burgemeesters, zodat iedereen op zijn eigen verantwoordelijkheden wordt aangesproken.

9.3 Aanjagen

Nadat de inzichten op bovenstaande wijze gedeeld zijn op de verschillende niveaus, is het raadzaam om tot concrete afspraken te komen met oog op te behalen resultaten. In deze laatste fase worden afspraken gemaakt met onder andere de regio's en met gemeenten en wordt gezamenlijk tot concrete actie overgegaan. Deze acties kunnen zowel een preventief als repressief karakter hebben. De uitkomsten uit het dashboard zijn hierbij richtinggevend.

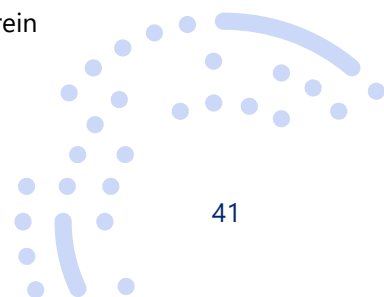
Allereerst moeten voor bedrijventerreinen waarop te weinig zicht is, op strategisch niveau concrete afspraken worden gemaakt over te leveren inspanningen om de belangrijkste gelegenheidsstructuren voor ondermijnende criminaliteit in kaart te brengen.

Indien (kwetsbaarheid voor) ondermijningsproblematiek op een bedrijventerrein als verrassing komt voor een gemeente, is mogelijk sprake van een blinde vlek en dient op lokaal niveau de informatiepositie te worden versterkt. Hiervoor kan gebruik worden gemaakt van diverse tools zoals de *Handreiking informatiebeeld bedrijventerreinen voor gemeenten* uit de *Toolbox Bedrijventerreinen*.

9.4 Handelingsperspectief gemeente

Het lectoraat ondermijning van Avans heeft in opdracht van Taskforce-RIEC Brabant-Zeeland een *Toolbox bedrijventerreinen* ontwikkeld, die concrete instrumenten biedt aan de hand waarvan gemeenten, o.a., 1. kwetsbare terreinen beter in kaart kunnen brengen, 2. concrete handelingsperspectieven op gemeentelijk niveau aangereikt krijgen en 3. van producten worden voorzien om de bewustwording van ondernemers te vergroten. De *Toolbox bedrijventerreinen* wordt voor gemeenten in heel Nederland beschikbaar gesteld via het Kennisplatform Ondermijning (KPO).

Wanneer inzicht is verkregen in de aard van bestaande (kwetsbaarheid van het terrein voor) ondermijnende activiteiten, is het van belang op basis van de eerder



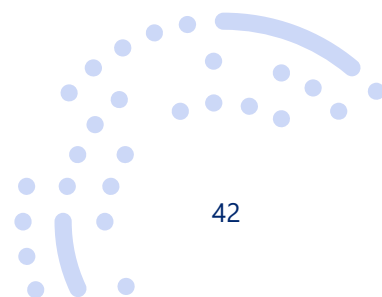


genoemde deelgebieden de beïnvloedbare kenmerken van het terrein aan te pakken. Voor een terrein dat relatief gezien slecht scoort op *omgevingskenmerken*, dienen handelingsperspectieven zich te richten op het verbeteren van de fysieke staat van een bedrijventerrein. Een ingrijpende mogelijkheid is (het bieden van financiële ondersteuning voor) herstructurering van kwetsbare terreinen. Een toegankelijker mogelijkheid als het aankomt op een groot aantal terreinen, is bijvoorbeeld het opstarten van een Bedrijven Investeringszone (BIZ), waarbij ondernemers meebetalen aan verbetering van de fysieke kwaliteit van het bedrijventerrein, zoals verbeterde groenvoorzieningen, verlichting of camerabewaking. De gemeente zal in dit geval in gesprek moeten gaan met ondernemers en hen moeten bewegen om hieraan deel te nemen.

Wanneer kwetsbaarheid van het terrein voortkomt uit *ondernemingskenmerken*, dient hier op te worden ingezet. De provincie kan gemeenten bijvoorbeeld ondersteunen bij het vestigingsbeleid op een bedrijventerrein, zodat de gewenste (type) bedrijven zich op de juiste plaats vestigen. Zo kan overaanbod van bepaalde bedrijven of branches tegengegaan worden en toegewerkt worden naar een toekomstbestendig bedrijventerrein. Dit thema is reeds onderwerp van gesprek bij de regiotafels van werklocaties, wat het mogelijk maakt om de uitkomsten uit het dashboard te gebruiken als input. Daarnaast kan de provincie gemeenten adviseren om te investeren in de organisatiegraad van ondernemers op de bedrijventerreinen die relatief slecht scoren op ondernemingskenmerken. Hierbij kan gedacht worden aan het starten van een Keurmerk Veilig Ondernemen traject (KVO) of indien die er nog geen is, het beschikbaar stellen van een subsidie voor de oprichting van een ondernemersvereniging of een andere vorm van publiek-private samenwerking.

Gemeenten doen dit niet alleen. Hiervoor is voor beide provincies het gremium aan meest relevante initiatieven en partners in de veiligheidszorg in kaart gebracht. Voor wat betreft het versterken van de informatiepositie kunnen de provincies, de regio's en gemeenten in Brabant samen optrekken met de Taskforce-RIEC Brabant-Zeeland. In Gelderland geldt dit ten aanzien van het RIEC Oost Nederland en bij het Platform Veilig Ondernemen (PVO). Ook kunnen andere veiligheidspartners, zoals omgevingsdiensten, herstructureringsmaatschappijen (Gelderland), de Brabantse Ontwikkelmaatschappij (BOM) en het Centrum voor Criminaliteitspreventie en Veiligheid (CCV) worden betrokken bij de aanpak van ondermijning op bedrijventerreinen. Een intensieve samenwerking van deze veiligheidspartners en begeleiding van de eerste 'voorbeeld-gemeenten' kan een positieve impuls geven aan de bekendheid onder gemeenten van de mogelijkheden die er zijn om een bedrijventerrein aan te pakken.

Met verschillende van de genoemde actoren is inmiddels afgesproken om kort na de lancering van de risicotaxatietool en de *Toolbox bedrijventerreinen* een kennisbijeenkomst te organiseren om geïnteresseerde partijen te informeren over de mogelijkheden hiervan. Het lijkt voor de hand te liggen om dit per provincie apart te organiseren, omdat ook de organisatiestructuur per provincie verschilt. De kennisbijeenkomst genereert aan de ene kant aandacht voor de ontwikkelde tools en biedt daarnaast de mogelijkheid om het praktische gebruik van de verschillende tools aan een groter publiek te communiceren.





10 Slotbeschouwing en toekomstperspectieven

10.1 Slotbeschouwing

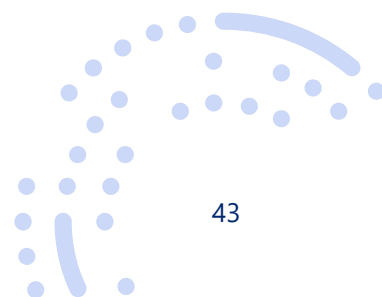
Tot op heden was er geen tool beschikbaar waarmee getaxeerd kon worden hoe kwetsbaar bedrijventerreinen zijn ten aanzien van criminaliteit en ondermijning. Het primaire doel van deze risicotaxatietool was daarom ook om via voorspellende modellen in te schatten welke bedrijventerreinen relatief gezien kwetsbaarder zijn voor ondermijnende criminaliteit. Voor deze tool zijn drie ondermijningsvormen gekozen: drugs, witwassen, en mensenhandel. Naast het bepalen van de relatieve kwetsbaarheid was er ook behoefte aan inzicht in de combinatie van belangrijkste risico-indicatoren om mogelijke aangrijpingspunten voor bestuurders en beleidsmakers inzichtelijk te maken. De doelen zijn nu gehaald.

De relatieve kwetsbaarheid van terreinen voor twee van de drie ondermijningsvormen, drugs en witwassen, blijkt goed voorspelbaar te zijn binnen het kader van dit onderzoek, evenals ondermijning in het algemeen, welke een combinatie is van de drie ondermijningsvormen. De derde ondermijningsvorm, mensenhandel, doorstaat de testen van voorspellend vermogen en theoretische validatie niet. Er zijn niet genoeg data over mensenhandel en ook te weinig 'gelabelde' doelvariabelen.

Het onderzoek maakt aan de hand van combinaties van top-indicatoren duidelijk wanneer een bedrijventerrein kwetsbaar is. Deze combinaties zijn vervolgens praktisch gevalideerd. Vaak gaat het om usual suspects die bevestigd worden door zowel bedrijventerreinexperts als via observaties. Hiermee bekrachtigt de praktische validatie de uitkomsten en de betrouwbaarheid van het model. Enkele indicatoren konden niet worden bevestigd. Dit waren veelal variabelen die door het model waren 'ontdekt' en niet afkomstig waren uit de longlist aan risico-indicatoren. Daarnaast waren er een paar indicatoren die niet konden worden beoordeeld op voorspellende waarde, zoals afstand tot voorzieningen (bijv. winkels, huisartsen, musea), waardoor ze noch bevestigd noch ontkracht konden worden tijdens de praktische validatie. Deze moeilijker beoordeelbare of onbevestigde indicatoren vergen veelal kennis over de omgeving op grotere schaal en de globale situatie op alle bedrijventerreinen, wat moeilijker in te schatten is door individuen, maar geen probleem zijn voor datagedreven modellen.

Het onderzoek biedt daarnaast inzicht in de richting van de invloed van een bepaalde variabele op het risico en geeft een duiding van de meest belangrijke indicatoren.

Samenvattend biedt het onderzoek voldoende handelingsperspectieven voor zowel provincies als gemeenten om iets aan daadwerkelijk kwetsbare bedrijventerreinen te doen. Het uitgangspunt is hierbij dat gemeenten zelf aan de slag moeten kunnen met het beeld dat uit de risicotaxatietool naar voren komt. De provincie heeft primair een signalerende taak en de inzichten uit het dashboard worden gebruikt om gemeenten te *informer*en, de gesignaleerde problematiek te *agender*en, en het *aanjagen* van lokale en regionale partners aangaande de aanpak van zogenaamde kwetsbare bedrijventerreinen.





10.2 Toekomstperspectieven

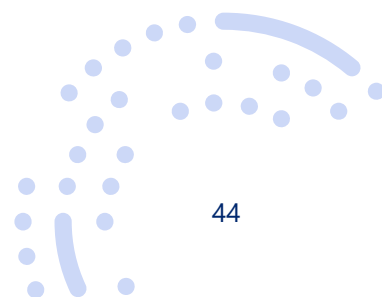
Zoals uit dit project blijkt, is het goed mogelijk om het risico op criminaliteit en ondermijning op bedrijventerreinen betrouwbaar te voorspellen voor verschillende vormen van criminaliteit. Voor de opgeleverde risicotaxatietool zijn hiervoor verschillende geavanceerde AI-technieken ingezet. De inzichten uit de risicotaxatietool zijn goed te visualiseren door middel van een dashboard. De uitkomsten zijn goed te duiden en te vertalen naar handelingsperspectieven.

Het opgeleverde dashboard is flexibel en kan verder ontwikkeld en verfijnd worden al naar gelang de wensen in de toekomst. Uiteraard is het mogelijk om in de loop van de tijd eventuele updates door te voeren in het dashboard door bijvoorbeeld nieuwe data toe te voegen of door de functionaliteit aan te passen of uit te breiden.

Ook kan het model voor alle ondermijningsvormen uitgebreid en verbeterd worden door meer data en labels toe te voegen. Dit is met name ook van belang voor de ondermijningsvorm mensenhandel. Een van de redenen waarom deze niet voorspelbaar bleek tijdens het hele traject is dat er niet genoeg data beschikbaar waren over mensenhandel. Van de 15 risico-indicatoren die oorspronkelijke voor mensenhandel op de longlist stonden zijn er uiteindelijk maar 5 verkregen. Pogingen kunnen ondernomen worden om de resterende risico-indicatoren te bemachtigen. Hiervoor kunnen nieuwe en op het doel gefocuste dataverzamelingsstappen uitgevoerd worden, waarmee er getracht kan worden om het model ook geschikt en praktisch toepasbaar te maken voor mensenhandel.

Daarnaast kunnen meer ondermijningsvormen toegevoegd worden aan het model. Voor sommige daarvan zijn mogelijk al veel relevante gegevens beschikbaar in de huidige data, daarnaast kunnen ook nieuwe databronnen aangeboord worden. De haalbaarheid van de methodiek is bewezen, waarmee de kans op succes groot is om nieuwe ondermijningsvormen toe te voegen. Dit kan helpen bij het uitbreiden van beleid op bedrijventerreinen.

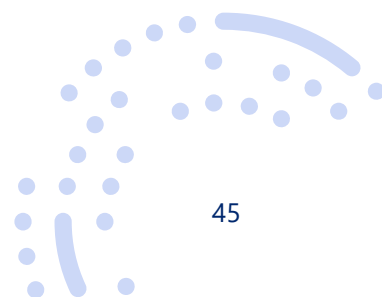
Tot slot kunnen ook andere provincies of regio's aansluiten. Door het vergroten van de schaal van de risicotaxatietool, generaliseert het model op een breder gebied waardoor inschattingen van de relatieve kwetsbaarheid nog betrouwbaarder zijn voor zowel nieuwe als toekomstige bedrijventerreinen. Daarnaast kan een uitrol van deze aanpak helpen om tot een landelijke aanpak te komen om georganiseerde criminaliteit en ondermijning tegen te gaan.





Literatuurlijst

- Bolsius, Y., Höcük, S., Prüfer, P. & Kolthoff, E. (2018). Indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Den Haag: Wetenschappelijk Onderzoek en Documentatiecentrum (WODC), <https://repository.wodc.nl/handle/20.500.12832/2353>.
- Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. (2002). Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, pp. 321–357.
- Kalidien, S. (2017). Criminaliteit en rechtshandhaving 2016. Ontwikkelingen en samenhangen. Den Haag: Wetenschappelijk Onderzoek en Documentatiecentrum (WODC).
- Kolthoff, E. & Khonraad, S. (2016). Ondermijnende aspecten van georganiseerde criminaliteit en de rol van de bovenwereld. *Tijdschrift voor Criminologie*, 58(2), pp. 76-90.
- Kruisbergen, E., Van de Bunt, H.G. & Kleemans, E. (2012). Georganiseerde criminaliteit in Nederland: Vierde rapportage op basis van de Monitor Georganiseerde Criminaliteit. Den Haag: Boom Lemma.
- Meijer, A. (2015). Bestuur in de datapolis. Slimme stad, blijde burger? Oratie Universiteit Utrecht. Den Haag: Boom Bestuurskunde.
- Prüfer, P. & Kolthoff, E. (2020). Met data science op zoek naar indicatoren van georganiseerde criminaliteit en ondermijning. *PROCES 2020*, p. 85-101.
- Slof, D. (2014). Voorspellen van webwinkel aankopen met een Random Forest. Erasmus Universiteit Rotterdam, Econometrie en Operationele Research.
- Tops, P. & Van der Torre, E. (2015). Wijkenaanpak en ondermijnende criminaliteit. Den Haag: Boom Lemma.





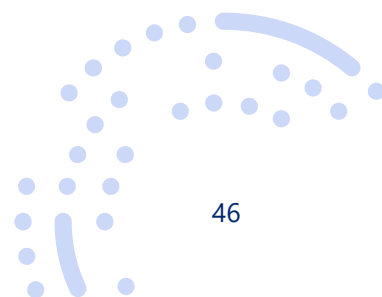
Lijst van Tabellen en Figuren

Lijst van tabellen in het rapport

Tabel 1: Criminaliteitsvormen met de predicaat 'dreiging' uit het Nationaal Dreigingsbeeld	10
Tabel 2: Longlist risico-indicatoren per ondermijningsvorm.....	13
Tabel 3: Geshortliste risico-indicatoren per ondermijningsvorm	14
Tabel 4: Twintig onafhankelijke databronnen	17
Tabel 5: De verkregen response van de uitgezette vragenlijsten onder experts	19
Tabel 6: Labels toekennen aan bedrijventerreinen.....	21
Tabel 7: Voorspellingsscores van de drie machine learning-modellen.....	22
Tabel 8: Lijst met top-indicatoren die voorkomen in de drie top-20 lijsten	25
Tabel 9: Extra validatie van het model via verhoudingen van de (voorspelde) klassen	27
Tabel 10: De belangrijke en minder belangrijke databronnen voor het model	29
Tabel 11: Proeftuingemeenten Gelderland	32
Tabel 12: Proeftuingemeenten Noord-Brabant.....	32
Tabel 13: Niveautransformatiestappen tot het niveau van bedrijventerreinen (RIN).....	53
Tabel 14: Het aantal variabelen dat verwijderd wordt bij het opschonen van data	54

Lijst van figuren in het rapport

Figuur 1: Plan van aanpak	12
Figuur 2: Route voor het toekennen van risico's aan bedrijventerreinen	20
Figuur 3: Top-20 indicatoren om 'hoog' risico te voorspellen voor de ondermijningsvormen	26
Figuur 4: Richtingen van invloed van de top-indicatoren voor verschillende ondermijningsvormen.....	28
Figuur 5: Screenshots van het eerste tabblad van het dashboard "Risicotaxatietool" voor provincies Noord-Brabant en Gelderland. Let op: voor deze screenshots zijn de waarheidsgetrouwe waarden vervangen door fictieve waarden. Dit in verband met gevoeligheid	36
Figuur 6: Screenshots van het tweede tabblad van het dashboard "Risicotaxatietool" voor provincies Gelderland en Noord-Brabant. Let op: voor deze screenshots zijn de waarheidsgetrouwe waarden vervangen door fictieve waarden. Dit in verband met gevoeligheid	38
Figuur 7: De datavoorbewerkingsstappen.....	52
Figuur 8: Het uitvoeren van een voorspellend model is een cyclisch proces	55
Figuur 9: Schematische weergave van Random Forest	60
Figuur 10: Schematische weergave van Gradient Boosting.....	61
Figuur 11: Schematische weergave van Logistic Regression	61





Bijlage A De lijst van shortlist risico-indicatoren

Dit zijn de 40 geshortliste risico-indicatoren die uiteindelijk meegenomen zijn in de modellen.

Aandeel panden met lage WOZ-waarde	Aansluiting op binnenvaart	Aantal aanvragen van subsidies of toeslagen	Aantal 'extra vestigingen' op BT	Aanwezigheid 'problematische types'
Aanwezigheid van meerdere ondernemingen op één adres	Aanwezigheid van stichtingen met een ANBI-status	Bedrijf in staat van faillissement	Bedrijfsactiviteit – automotieve / garagebedrijven	Bedrijven die in financieel zwaar weer zitten
Bedrijven op BT hebben adreswijzigingen in KVK	Combinatie risicobranche algemeen	Eenmanszaken	Fluctuatie groei binnen branche	Grootteklasse (kleine bedrijven)
Het gemiddelde bouwjaar (op een BT)	Hoog verloop (van ondernemingen op BT)	Hoog-capaciteit aansluitingen (op BT)	Kwetsbare branches - drugscriminaliteit	Kwetsbare branches - mensenhandel
Kwetsbare branches - witwassen	Leegstand - gasverbruik	Leegstand - stroomverbruik	Leegstand op BT	Ligging in een 'buitengebied'
Ligging van bedrijventerrein - minder informeel toezicht	Ligging van bedrijventerrein - vervoersmogelijkheden	Nieuwe bedrijven	Nieuwe importeurs/exporteurs	Overlastmeldingen
Papieren BV	Parkmanagement/beheer	Publieke samenkomstfuncties	Risicovolle combinatie 'automotive' en 'groente/fruit'	Start bedrijf sinds medio 2020 (recent opgericht)
Toename van een bepaalde branche in korte tijd	Verloederd bedrijventerrein	Vertrokken bedrijven	Weigeren slimme meter	Woon-werkfunctie BT

Indirect verkregen variabelen zijn in grijs tint weergegeven. Indirect kan bijvoorbeeld zijn dat de indicator op een geaggregeerde wijze is verkregen; zoals de WOZ-waarde die per buurt is in plaats van per gebouw. Of dat alleen een deel of de essentie van de indicator is verkregen, zoals bijvoorbeeld dat alleen het aantal aanvragen voor subsidies is verkregen maar geen toeslagen (dus deels) en bij 'Papieren BV' de indicatie of een bedrijf economisch actief is als representatief is genomen (dus essentie).





Bijlage B Korte introductie data science methoden

Data science is een discipline die erop gericht is om grote hoeveelheden data te analyseren. Het gaat vaak om het interpreteren van data die vaak complex en ongestructureerd is, en vaak hoogfrequent gegenereerd worden. De kunst van data science is het transformeren van deze data in waardevolle acties. Dit vakgebied biedt meer mogelijkheden met data dan klassieke statistiek.

Met data science is het mogelijk om onzichtbare patronen in de data te herkennen en op basis hiervan voorspellingen te doen. Data science onderscheidt zich van conventionele methoden door de holistische benadering. Op die manier kunnen er aanbevelingen worden gedaan voor belangrijke beslissingen die genomen moeten worden.

Big Data

Het verzamelen, het opslaan, het delen en het verwerken van data wordt steeds goedkoper en gemakkelijker. Hierdoor groeit de hoeveelheid aan data continu. Databestanden worden daardoor steeds groter en rijker. Ook is het tegenwoordig steeds beter mogelijk om allerlei data te koppelen. Zo wordt bijvoorbeeld de data die het CBS tot haar beschikking heeft, vaak gekoppeld met andere datasets (in een beveiligde omgeving), zoals vragenlijsten, politiedata en gemeentelijke data met als resultaat grote en vaak complexe datasets.

We spreken van big data als aan een aantal kenmerken voldaan wordt. Deze kenmerken zijn met name de hoeveelheid aan data, de snelheid waarmee de data binnenkomen of opgevraagd worden en de diversiteit van de data. Data science is een vakgebied dat zich met uitstekend leent voor het analyseren van dergelijke data.

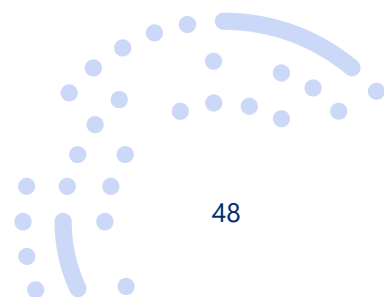
Big data vragen om bijzondere manieren van betrouwbare en veilige dataopslag, verwerking en toegankelijkheid. Analyses worden vaak gedaan door gebruik te maken van meerdere computers en meerdere rekenunits, zogenaamd 'high-performance computing' (HPC) zoals Hadoop clusters en Spark-Streaming, en van virtuele omgevingen. Daardoor is het mogelijk om met voldoende rekenkracht een veelvoud aan verschillende tools, technieken en datamanagementstructuren bijeen te brengen in data science.

Machine Learning

Machine Learning is een geavanceerd onderzoeksveld binnen data science dat zich bezighoudt met technieken waarmee computersystemen kunnen leren, zonder dat ze er speciaal voor geprogrammeerd zijn. Met de technieken van machine Learning kunnen uit grote databestanden verbanden in de data worden geïdentificeerd. Doorgaans wordt machine learning onderverdeeld in drie brede categorieën: Supervised machine Learning, Unsupervised machine Learning en Reinforcement Learning.

Supervised machine Learning

Supervised machine Learning is de benaming van een set van geavanceerde algoritmen die gebruik maken van de informatie van bekende uitslagen, de zogenaamde *labels*, om de voorspellingen te kunnen maximaliseren. Gelabelde data zijn het deel van de data





waarvan de uitslag bekend en dus getagd is; bijvoorbeeld bij het voorspellen van uithuisplaatsingen beschikt men over de data van kinderen die al uit huis zijn geplaatst. Het machine Learning-model kan hierop worden *getraind*.¹² Hiermee worden ook inzichten verkregen in de features (de variabelen die gebruikt worden voor een voorspelling) door iteratieve multidimensionale dataverkenning. Door middel van validatie bij elke iteratiestap, en bij voorkeur een teststap aan het eind, worden de algoritmen getraind om te worden geoptimaliseerd tegen "onbekende" positieve en negatieve patronen uit toekomstige data. Dat is de kracht van supervised machine Learning.

Unsupervised machine Learning

Machine learning kan ook unsupervised worden uitgevoerd en worden gebruikt om baseline gedragsprofielen voor verschillende entiteiten te leren en vast te stellen om die vervolgens te gebruiken om zinvolle anomalieën te vinden. Als er geen labels (geen bekende uitkomsten) beschikbaar of moeilijk te genereren zijn, wordt de voorkeur gegeven aan unsupervised learning. Een vorm van unsupervised machine Learning kan bijvoorbeeld ingezet worden om informatie uit teksten te extraheren die op het eerste gezicht lastig uit de data te halen is.

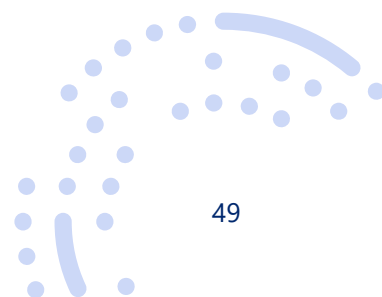
Transparantie

Een vaak voorkomend vereiste is dat een gebruikte machine learning-algoritme transparant en uitlegbaar moet zijn. Overheidsinstanties hebben bijvoorbeeld een verplichting richting de burgers, waardoor alle gemaakte beslissingen communiceerbaar, uitlegbaar en transparant moeten zijn. Een 'black box' algoritme, zoals een neuraal netwerk of Deep Learning, is vaak daardoor geen optie.

Door te gaan voor een AI-model dat zich verleent voor transparency, explainability, traceability, privacy en data governance voldoet een uitvoerder ook aan veel van de kernonderdelen van de Europese High-Level Expert Group on Artificial Intelligence.¹³

¹² Dit betekent dat het algoritme steeds beter wordt in het vinden van oplossingen, net zoals mensen leren aan de hand van ervaringen.

¹³ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1>





Bijlage C Voorbeeld van een vragenlijst

Een voorbeeld van een vragenlijst opgestuurd naar experts van bedrijventerreinen ziet er als volgt uit:

Over welke gemeenten in **Noord-Brabant** kunt u iets vertellen over criminaliteit en ondermijning op bedrijventerreinen?

- | | | |
|--|---------------------------------------|--|
| <input type="checkbox"/> Altena | <input type="checkbox"/> Halderberge | <input type="checkbox"/> s-Hertogenbosch |
| <input type="checkbox"/> Bernheze | <input type="checkbox"/> Helmond | <input type="checkbox"/> Steenbergen |
| <input checked="" type="checkbox"/> Best | <input type="checkbox"/> Heusden | <input type="checkbox"/> Tilburg |
| <input type="checkbox"/> Bladel | <input type="checkbox"/> Meijerijstad | <input type="checkbox"/> Valkenswaard |
| <input type="checkbox"/> Breda | <input type="checkbox"/> Moerdijk | <input type="checkbox"/> Veldhoven |
| <input type="checkbox"/> Eindhoven | <input type="checkbox"/> Oss | <input type="checkbox"/> Waalwijk |
| <input type="checkbox"/> Geldrop-Mierlo | <input type="checkbox"/> Roosendaal | <input type="checkbox"/> Geen van deze gemeenten |
| <input type="checkbox"/> Goirle | <input type="checkbox"/> Rucphen | |

Vorige

Verder

Er zijn drie mogelijke stemmen, 'laag', 'gemiddeld', en 'hoog' risico. Als de afnemer geen kennis heeft over een bedrijventerrein, kan de afnemer ook 'weet niet' aanvinken.

Gemeente Best

[Klik hier voor een overzicht op de kaart van de bedrijventerreinen.](#)

Hieronder worden bedrijventerreinen in de **gemeente Best** getoond.

Kunt u een inschatting maken van de kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning op de volgende bedrijventerreinen?

Geef per bedrijventerrein aan of u deze kans laag, gemiddeld, of hoog inschat als het gaat om:

- drugscriminaliteit
- witwassen
- mensenhandel

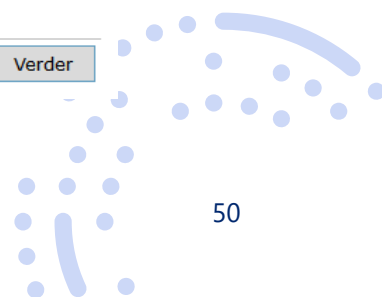
Wanneer u een bedrijventerrein niet kent, kies dan voor de optie 'Weet niet'.

Wanneer u een overzicht op de kaart wilt zien van waar de bedrijventerreinen zich precies bevinden, kunt u op de link bovenin klikken.

	Risico op drugscriminaliteit				Risico op witwassen				Risico op mensenhandel			
	Laag	Gemiddeld	Hoog	Weet niet	Laag	Gemiddeld	Hoog	Weet niet	Laag	Gemiddeld	Hoog	Weet niet
Breeven	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
T Zand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Heide	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ploegstraat - Park Batadorp	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

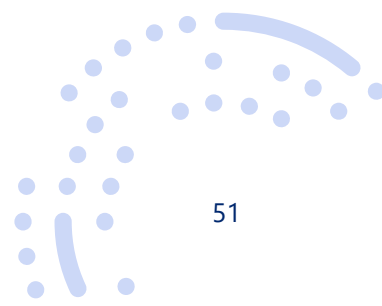
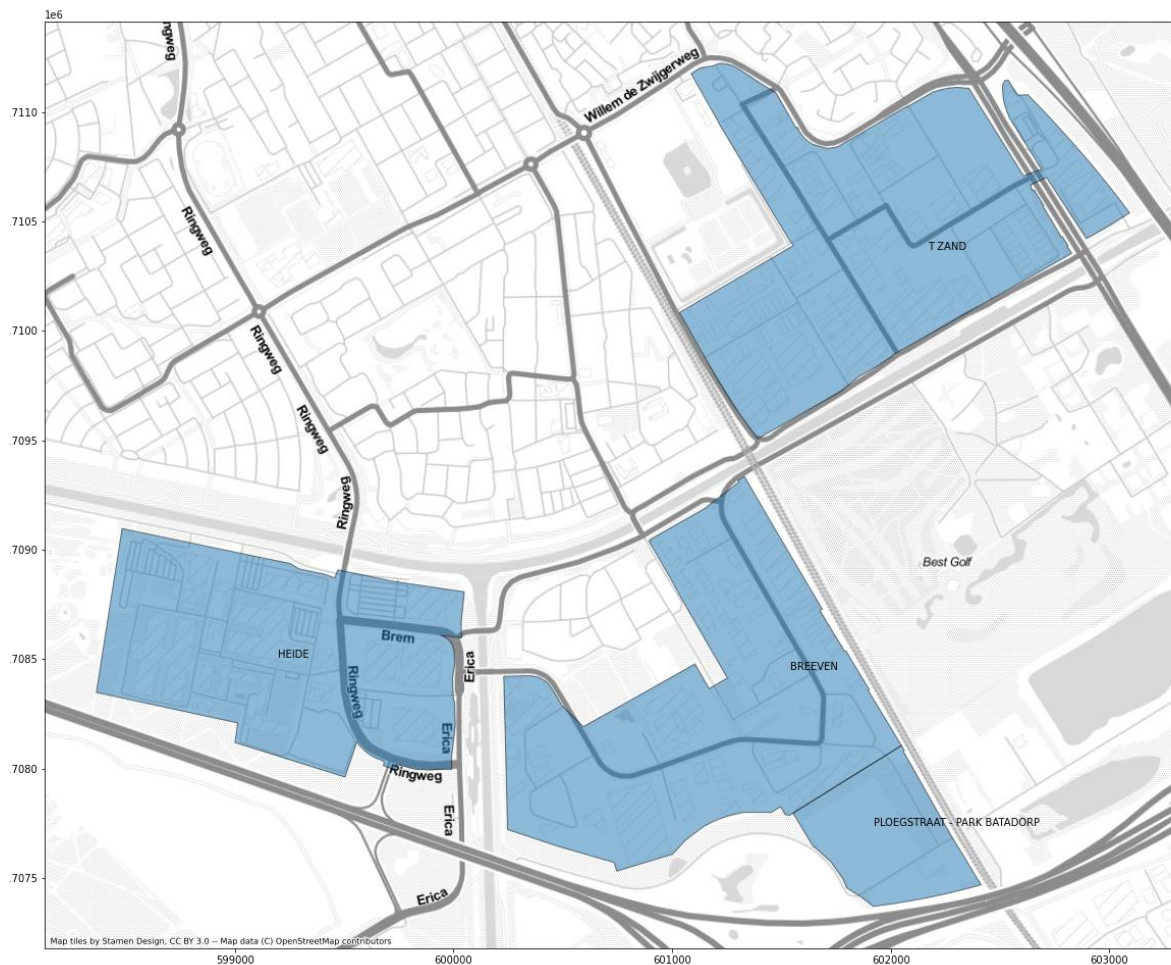
Vorige

Verder





Per vragenlijst kunnen meerdere gemeenten geselecteerd worden. Bij elke gemeente is een plattegrond met locaties van bedrijventerreinen toegevoegd.



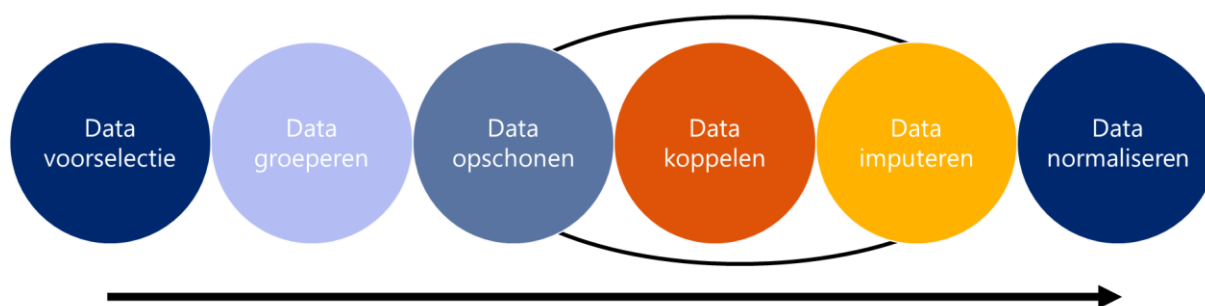


Bijlage D Modelleren

Het modelleren van de data om tot een antwoord op de eerste hoofdvraag te komen omvat meerdere fasen. De data moeten eerst gekoppeld en voorbereikt worden voordat het modelleren kan starten. Hierna wordt het voor ons doel meest geschikte model geselecteerd en uitgevoerd. Soms kunnen meerdere modellen ingezet worden om model onafhankelijkheid te creëren. Als laatste worden de modeluitkomsten getoetst en geëvalueerd. Ieder fase bestaat uit meerdere (sub)onderdelen.

Datavoorbewerking

De datavoorbewerking is een van de belangrijkste, maar ook meest tijdrovende klussen. Het omvat meerdere onderdelen om de ruwe data tot een bruikbare en waardevolle dataset te transformeren. In **Figuur 7** laten we de onderdelen zien die nodig zijn tijdens de datavoorbewerking.

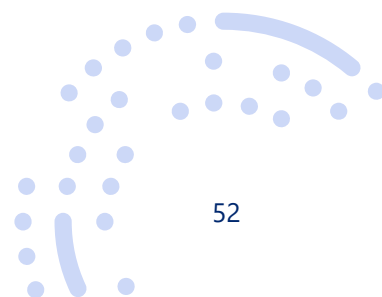


Figuur 7: De datavoorbewerkingsstappen

De datavoorbewerking begint met ruwe data (helemaal links in de figuur) en geeft als uitkomst bewerkte data die gemodelleerd kunnen worden (helemaal rechts in de figuur). Het opschonen van data overlapt meerdere onderdelen (te zien aan de zwarte curves in de figuur). Wat in dit plaatje mist is feature engineering; een stap waarbij nieuwe variabelen worden aangemaakt uit al bestaande variabelen. Feature engineering wordt zowel uitgevoerd tijdens de data voorselectie als na imputatie.

Data voorselectie

De data voorselectie gebeurt op individuele datasets (zie **Tabel 4**). Ongerelateerde variabelen zoals 'gemeente', 'terreinnaam', 'datum', 'id-nummer', of 'provincie' worden bij voorbaat al verwijderd. Ook te uitgebreide gegevens, zoals bij de RVO subsidies dataset de 'subsidie thema' (bijv. geothermie), of informatie die bekend is bij alleen één van de provincies, worden niet meegenomen in de finale versie van de bewerkte dataset. Daarnaast worden sommige variabelen samengevoegd tot een enkele variabele die de risico-indicator vastlegt. De variabele laagcapaciteitsaansluiting is bijvoorbeeld een samenvoeging van de capaciteiten 1x6 tot en met 3x35. Andere variabelen worden daarentegen weer opgesplitst tot meerdere variabelen, zoals bijvoorbeeld de categorische variabelen waarbij ieder categorie een afzonderlijke onafhankelijke variabele wordt. Verder worden alle gegevens die gaan over locaties die buiten bedrijventerreinen vallen gefilterd uit de finale dataset. We hebben hiervoor een lange lijst met postcodes (pc6) die vallen op bedrijventerreinen in de twee provincies.





Data groeperen

We willen alle data op het niveau van bedrijventerreinen (oftewel RIN-nummer). Is dat niet mogelijk, dan vindt er een niveautransformatie plaats. Een transformatie kan bijvoorbeeld zijn een aggregatie van data van een hoger niveau of een segregatie van data van een lager niveau. In **Tabel 13** laten we de transformatiestappen zien die nodig zijn om alle datasets op hetzelfde niveau te krijgen.

Tabel 13: Niveautransformatiestappen tot het niveau van bedrijventerreinen (RIN)

Niveau van de dataset	Segregatie naar niveau pc6	RSIN koppelen aan bedrijf	Aggregatie naar niveau pc6	Aggregatie naar niveau RIN
Pc6				✓
Pand / Perceel			✓	✓
Bedrijf			✓	✓
RSIN (KVK-nummer)		✓	✓	✓
Buurt	✓			✓

Het segregeren van een buurt naar het niveau van pc6 gaat gepaard met benaderingen. De belangrijkste benadering is dat de waarde van een buurt gelijkmatig verdeeld wordt onder alle postcodes binnen die buurt. Het aggregeren van pc6 naar bedrijventerrein gaat gepaard met imperfecte grensbepaling van een bedrijventerrein. Sommige postcodes overstijgen namelijk de grenzen van een bedrijventerrein. Deze postcodes worden alsnog in zijn geheel meegenomen.

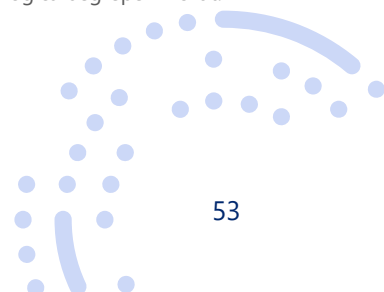
Data opschonen

Data opschonen betekent het verwijderen van onbruikbare variabelen en het wegwerken van fouten en missings.¹⁴ Voor de finale versie van de bewerkte dataset verwijderen we de volgende variabelen: beschrijvende en koppelvariabelen zoals het RIN-nummer of het KVK-nummer, redundante gegevens zoals het aantal voorzieningen binnen 1, 3, 5, 10, 20 km waarbij een enkel afstand volstaat, gevoelige variabelen zoals etniciteit, heel schaars gevulde variabelen van meer dan 33% missings, variabelen die weinig variantie tonen zoals bijvoorbeeld alles ingevuld met een 'ja', sterk gecorreleerde variabelen met een Pearson's $|r|$ van >0.90 , en ongerelateerde variabelen zoals het aantal werknemers in 2014. In Error! Reference source not found. geven we een voorbeeld van een correlatiecheck.

Ook corrigeren we incidentele fouten in de data zoals een bouwjaar in de toekomst (van meer dan 1 jaar) en werken we user missings¹⁵ weg door de betekenis van de user missings te achterhalen en ze te vervangen door reële gegevens. We laten in **Tabel 14** zien hoeveel variabelen we uit ieder categorie verwijderen. Uiteindelijk houden we 310 variabelen over.

¹⁴ Een missing is gedefinieerd als een ontbrekende waarde binnen een observatie voor een bepaalde variabele. Missings zijn vaak niet-gevulde of leeggelaten datapunten.

¹⁵ Een user missing is wanneer een missende datapunt wordt vervangen door een code, vaak 888 of 999, wat uitleg geeft over de reden van het missen van het datapunt. Soms kunnen user missings alsnog worden opgevuld als de logica begrepen wordt.





Tabel 14: Het aantal variabelen dat verwijderd wordt bij het opschonen van data

Beschrijvend	Ongelateerd	Redundant	Gevoelig	Missings	Variantie	Correlatie
5	23	166	4	98	1	99

Data koppelen

Alle 20 datasets worden vervolgens aan elkaar gekoppeld op het niveau van RIN. We koppelen geen data van een lager aggregatieniveau dan RIN, wat bijvoorbeeld een wijk of gemeente kan zijn. Bij het aggregeren van variabelen houden we rekening met het datatype van de variabele. De variabeletypen kunnen absoluut, percentage, of gemiddelde zijn. Absolute waarden worden opgesomd en relatieve waarden (percentages en gemiddelden) worden gemiddeld. In het geval waarbij beide typen van belang kunnen zijn voor het model, nemen we beide mee als twee aparte variabelen. Dat betekent dat we zowel de absolute vorm als een percentage van een variabele meenemen. Een voorbeeld hiervan is het aantal bedrijven op een bedrijventerrein dat ooit een faillissement heeft aangevraagd. Naast het absolute aantal nemen we ook het percentage van bedrijven op een bedrijventerrein mee.

Data imputeren

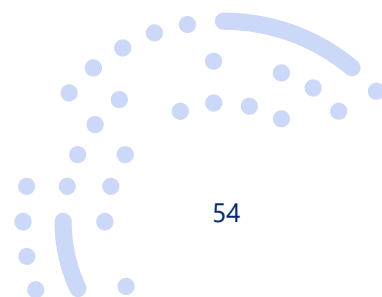
Imputeren is het proces van (realistische) waarden toekennen aan missende datapunten. Tijdens de voorbereidingsstappen worden de lege datapunten zoveel mogelijk geïmputeerd door bekende of verwachte waarden. We kunnen in sommige gevallen aannemen dat de missings '0' moeten zijn. Als er bijvoorbeeld missings zijn bij misdrijven of overlastmeldingen op een postcode in de data, betekent dat dat daar géén (dus 0) misdrijven en overlastmeldingen zijn geweest. Alle resterende missings worden geïmputeerd via de mediaan van de variabele, als de variabele niet in zijn geheel uitgesloten is uit de finale set doordat er te veel missings zijn (zie tekst over data opschonen).

Het aantal overgebleven missings in de opgeschoonde data is verhoudingsgewijs 6.2%. Dit aantal daalt na het invullen van de variabelen met bekende of verwachte waarden met 1.8%. Het resterende aantal missings dat nog geïmputeerd moet worden (via mediaan) is 4.4%.

Data normaliseren / standaardiseren

Om variabelen beter met elkaar te kunnen vergelijken tussen verschillende bedrijventerreinen moet er een normalisatie plaatsvinden. Zonder normalisatie geven absolute waarden een vertekend beeld. Denk bijvoorbeeld aan het aantal inbraken op een groot bedrijventerrein ten opzichte van een klein bedrijventerrein. Normalisatie kan gedaan worden door de absolute getallen te delen door het aantal bedrijven, werknemers, of oppervlakte (hectare) van een bedrijventerrein. Vanwege eenduidigheid en deels wegens incomplete en onzekere gegevens over de aantallen van bedrijven en werknemers in de data, normaliseren we alle absolute waarden door het netto oppervlak van een bedrijventerrein.

Relatieve waarden, zoals percentages of gemiddelden, of binaire variabelen, zoals 'parkmanager aanwezig' (ja/nee), worden niet genormaliseerd. Het aantal variabelen die een normalisatie behoeven komen neer op 100 variabelen (van de 310).

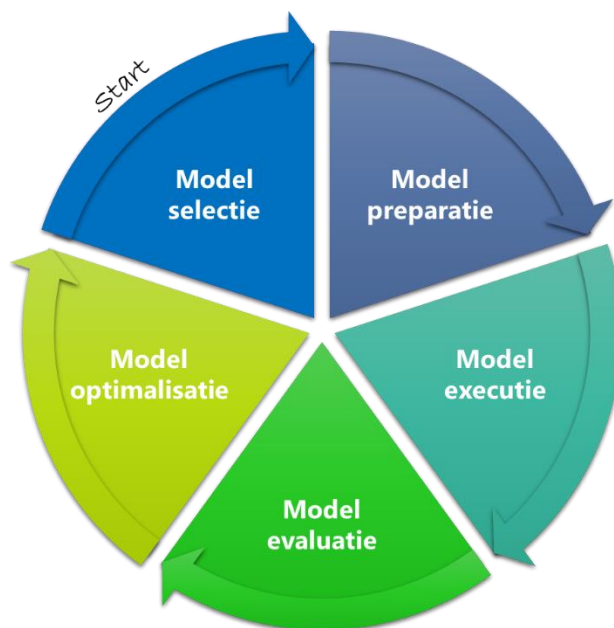




Modellen uitvoeren

Het uitvoeren van een voorspellend machine learning-model is een cyclisch proces. Het begint met het selecteren van het meest geschikte model voor het onderzoek (of beschikbare data) en eindigt met het optimaliseren van dit model. Bij het optimaliseren kan het resultaat zijn dat er een beter of geschikter model geselecteerd wordt of dat alleen de modelparameters aangepast worden. Soms is het zelfs nodig om een stap verder terug te doen en de datavoorbewerkingsstappen opnieuw uit te voeren.

Figuur 8 illustreert het cyclische proces.



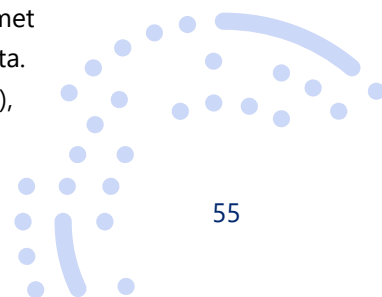
Figuur 8: Het uitvoeren van een voorspellend model is een cyclisch proces

Model selectie

Er bestaan veel machine learning-modellen, maar we hebben hier te maken met classificatietechnieken voor supervised machine learning. Classificatietechnieken zijn algoritmes gebouwd voor het voorspellen van een categorale variabele op basis van een aantal indicatoren. Hier gaat het dus om het lage of hoge risico op ondermijnende criminaliteit op het bedrijventerrein.

Verschillende selectiecriteria zijn van belang voor het kiezen van het best passende model. Voor ons doel hebben we een transparant en verklaarbaar model nodig. Overheidsinstanties hebben verplichtingen tegenover de burgers, waardoor alle gemaakte beslissingen communiceerbaar, uitlegbaar, en transparant moeten zijn. Een 'black box' algoritme, zoals een neuraal netwerk of deep learning, is daardoor vaak geen optie. Daarnaast moet het model krachtig genoeg zijn in voorspellende waarde, omdat de resultaten beleidsbepalend kunnen zijn. Er wordt veel waarde gehecht aan de modeluitkomsten. De betrouwbaarheid van de modeluitkomsten wordt daarom ook getoetst aan de hand van een praktische validatie (zie **Hoofdstuk 7**).

Als laatste moet het model geschikt zijn voor het doel en voor de beschikbare data voor dit werk. Sommige algoritmen vergen veel data, terwijl andere juist beter overweg kunnen met weinig data. Nog andere modellen kunnen goed overweg met ongebalanceerd data. Voor ons doel zijn er drie machine learning-modellen geschikt: Random Forest (RF),





Gradient Boosting (GB), en Logistic Regression (LR). Het beste model zullen we kiezen we aan de hand van de beste voorspellingsscores.

Voor de indicatoren passen we echter een ander manier toe. In plaats van één van de drie modellen te kiezen, passen we alle drie de modellen toe en nemen we het gemiddelde resultaat als de finale uitkomst. Deze methodiek wordt ook wel stacking genoemd en maakt de modeluitkomsten betrouwbaarder en robuuster door de betere generaliseerbaarheid en modelonafhankelijkheid. In **Bijlage F** beschrijven we de drie technieken in meer detail.

Model preparatie

Er zijn verschillende keuzes gemaakt bij het ontwerpen van het model. We leggen hier de keuzes uit. Alle keuzes zijn grondig getest op de impact die ze hadden. Waar de keuzes grotere impact hadden op de uitkomsten, zijn er meerdere modellen gedraaid en zijn de gemiddelde uitkomsten aangenomen.

We hebben de data opgesplitst in een training- en een testset. We hebben hiervoor verschillende train-testratio's gebruikt. Deze zijn respectievelijk 80/20, 75/25, en 70/30 (%). De opsplitsingen zijn gestratificeerd op de klassen van de labels (de expertmeningen 'hoog' en 'laag' risico) zodat de klassenverhoudingen behouden blijven. Verder passen we datanormalisatie toe via de MinMaxScaler.¹⁶

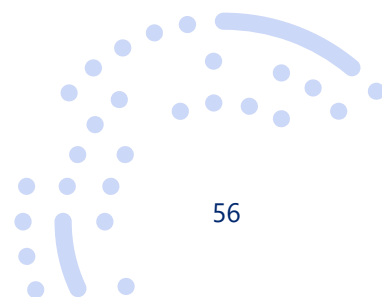
Vanwege de klassenonevenwichtigheid komt het in sommige gevallen voor dat er te weinig datapunten zijn van een bepaalde klasse. Dit is vooral het geval voor de klasse 'hoog' bij de ondermijningsvormen mensenhandel en witwassen. De toegepaste oplossing hiervoor is een synthetische overbemonstering. We hebben hiervoor de techniek SMOTE¹⁷ (Chawla et al., 2002) ingezet. Hiermee genereren we nieuwe synthetische data voor de trainingset om de modellen nog nauwkeuriger te maken. Verschillende gradaties van overbemonstering zijn getest en uitgevoerd.

De missende datapunten in de data worden aangevuld via de mediaan van de desbetreffende variabele. Het gaat hier om ongeveer 4% van de datapunten. Andere imputatietechnieken zijn ook uitgeprobeerd, zoals het imputeren via het gemiddelde, het meest frequente, of de geavanceerde k-NN techniek. De conclusie is echter dat het uiteindelijk weinig uitmaakt voor de uitkomsten.

Als laatste zijn de parameters van het model van belang. Ieder model heeft zijn eigen set van optimalisatieparameters. Deze worden ook hyperparameters genoemd. De keuze hierin is vaak om de standaardparameters aan te nemen, en dit is met name zo bij een *data-centric* aanpak. Dat betekent dat het standaardmodel als goed genoeg wordt beschouwd en de prestatieverbeteringen met name via de data-aanpassingen gaan. Toch hebben we een uitgebreide grid-search ingezet om de beste hyperparameters te vinden en de gemiddelde uitkomst van de twee beste hyperparametercombinaties als eindresultaat genomen. De gevonden beste hyperparameters zijn weergegeven in **Bijlage G**.

¹⁶ De MinMaxScaler schaalt alle waarden van de variabelen naar een waarde tussen 0 en 1.

¹⁷ Synthetic Minority Over-sampling Technique.





Model evaluatie

De prestaties van voorspellende machine learning-modellen worden aan de hand van verschillende evaluatiemetrieken beoordeeld. De in dit onderzoek in beschouwing genomen evaluatiemetrieken zijn:

- **Accuracy**, voor zowel de train- als de testset.
- **Precision**, ook wel bekend als Positive Predictive Value (PPV).
- **Recall**, ook wel bekend als Sensitivity of True Positive Rate (TPR).
- **F₁-score**, het harmonische gemiddelde van precision en recall.

We focussen voor onze toepassing op de F₁-score. De F₁-score is met name heel erg geschikt voor ongebalanceerde data, wat bij ons het geval is. De doelvariabelen bevatten namelijk (veel) meer 'laag' geïdentificeerde risicoklassen dan 'hoog'. De F₁-score is zowel een optimalisatie voor de correct voorspelde meerderheidsklasse ('laag') als een minimalisatie van alle verkeerd voorspelde klassen (zowel 'hoog' als 'laag'). Aangezien de evaluatiestatistieken op klasniveau worden gegeven, nemen we de strikte macro-gemiddelde klassenscore aan als de uiteindelijke modelprestatie.

De evaluatiemetrieken worden ook vaak aangeduid in correct of fout voorspelde klassen. In binaire gevallen wordt er dan vaak gesproken over een 'Positive' en een 'Negative'. Een True Positive is bijvoorbeeld een correct voorspelde positieve klasse. Hier zou het dan gaan om een correct voorspelde 'laag' risico bedrijventerrein. De vier mogelijkheden zijn dan: True Positive, True Negative, False Positive, en False Negative. Deze korten we af als TP, TN, FP, en FN.

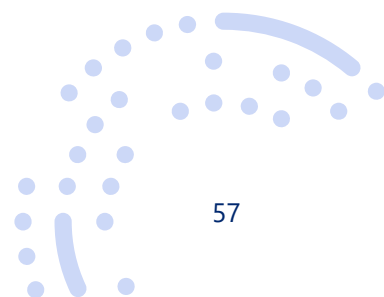
In de lijst hieronder worden de evaluatiemetrieken in de hierboven aangegeven termen geformuleerd, waarbij ieder term staat voor het aantal keer dat die term voorkomt in de voorspelling.

- **Accuracy** = $(TP + TN) / (TP + TN + FP + FN)$
- **Precision** = $TP / (TP + FP) = PPV$
- **Recall** = $TP / (TP + FN) = TPR$
- **F₁-score** = $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Modeloptimalisatie

Na de modevaluatie worden modellen geoptimaliseerd en begint de hele cyclus opnieuw. Na de aanpassingen aan het model worden de modellen opnieuw getraind en wederom geëvalueerd. De modeloptimalisatietechnieken bestaan onder andere uit een hyperparameteroptimalisatie. Hiervoor wordt er een grid search ingezet voor een zoektocht naar de beste hyperparameters, zoals bijvoorbeeld de learning rate, het aantal iteraties, of de loss-function.

Veel van de andere optimalisaties gaan nog een stap verder terug door te kijken naar mogelijke verbeteringen in de data, zoals beter data opschonen, meer feature engineering, het gebruik van een andere imputatietechniek, en zelfs dataverrijking door middel van reële of synthetische data. De optimalisaties worden toegepast om uiteindelijk het best presterend voorspellend model te vinden.





Feature importance

Het vinden van de top-indicatoren gebeurt aan de hand van ingebouwde technieken die ook bekend staan onder de naam feature importance. De gebruikte machine learning-modellen hebben kant-en-klare methoden om de belangrijkste indicatoren te rangschikken aan de hand van verschillende metrieken. Random Forest gebruikt bijvoorbeeld de mean impurity decrease¹⁸. Het staat ook bekend als het Gini-belang. Dit is een ingebouwde optie van het scikit-learn¹⁹ model en de specifieke functie hiervoor heet "*feature_importances_*".²⁰ Bij Logistic Regression²¹ is het belang van de indicatoren veel directer; er wordt simpelweg gekeken naar de absolute waarde van de feature coëfficiënten van het model. XGBoost²² heeft ook een ingebouwde optie om de feature importance weer te geven. Net als bij Random Forest gebruikt het de functie "*feature_importances_*", waarbij er verschillende opties zijn voor het bepalen van het belang van de indicatoren via de parameter "*importance_type*".²³ Het standaardtype "*weight*" is aangenomen.

¹⁸ https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html

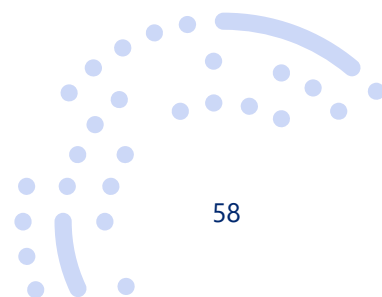
¹⁹ <https://scikit-learn.org/stable/>

²⁰ <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

²¹ https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

²² <https://xgboost.readthedocs.io/en/stable/>

²³ https://xgboost.readthedocs.io/en/stable/python/python_api.html

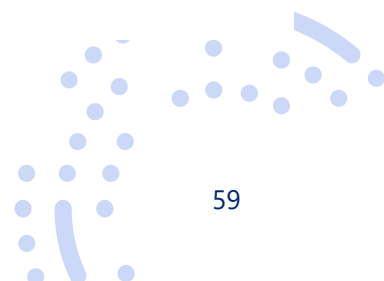
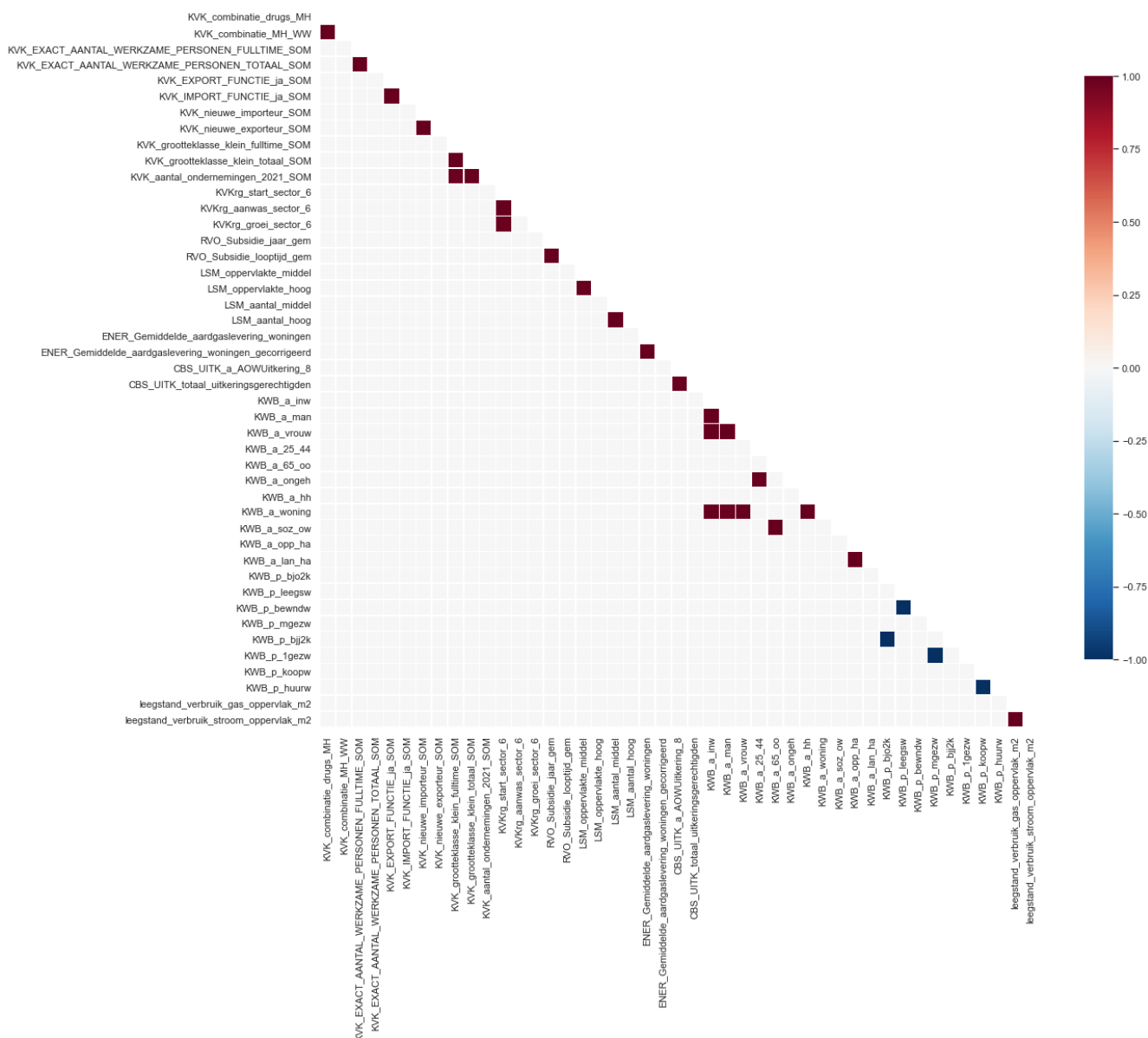




Bijlage E Voorbeeld correlatiecheck

Hier laten we een voorbeeld van een correlatiecheck zien. We gebruiken in dit voorbeeld een correlatiecoëfficiënt van Pearson's $|r| \geq 0.99$. Voor het onderdeel correlatie-cleaning tijdens de datavoorbewerkingsstappen wordt echter een Pearson's waarde van 0.90 aangehouden in plaats van de 0.99 waarde die ter illustratie is weergegeven hier.

Bij een sterke (anti)correlatie wordt de variabele in de verticale as verwijderd. De naamgeving in de grafische weergave heeft in dit geval nog de originele naamgeving vanuit de datasets. Hierdoor kan het mogelijk minder begrijpelijk zijn wat de variabele precies inhoudt. Voor meer informatie kan de documentaties van de originele databronnen geraadpleegd worden. De databronnen worden aangeduid door de eerste paar (hoofd)letters of woord bij de namen van de variabelen.

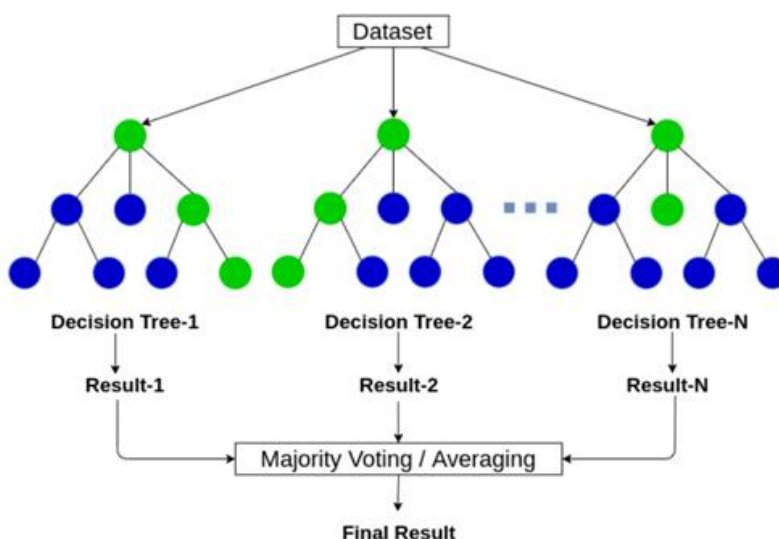




Bijlage F Beschrijving classificatiemodellen

Random Forest

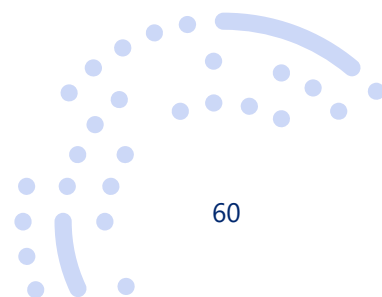
Random Forest is een 'tree-based supervised learning' techniek, waarbij een groot aantal beslissingsbomen (Decision Trees) gecombineerd worden om tot de uiteindelijke voorspelling te komen. Hierdoor is deze methode veel stabielere dan een enkele beslissingsboom. Als de doelvariabele categoraal is, wordt de uiteindelijke uitkomst bepaald door middel van 'majority voting'. Met andere woorden: de uitkomst van de meeste bomen geldt als einduitkomst. De verzameling bomen wordt random genoemd, omdat elke boom getraind wordt op een willekeurige selectie van variabelen en observaties. Als meerdere modellen gecombineerd worden in een groot model spreken we van een ensemble model. Het combineren van vele losse beslissingsbomen tot een ensemble model resulteert in een hogere precisie en ook in stabielere voorspellingen. Een Random Forest geeft over het algemeen veel betere voorspellingen dan een enkel beslissingsboom (Slof, 2014). Een illustratie van Random Forest is te zien in **Figuur 9**.

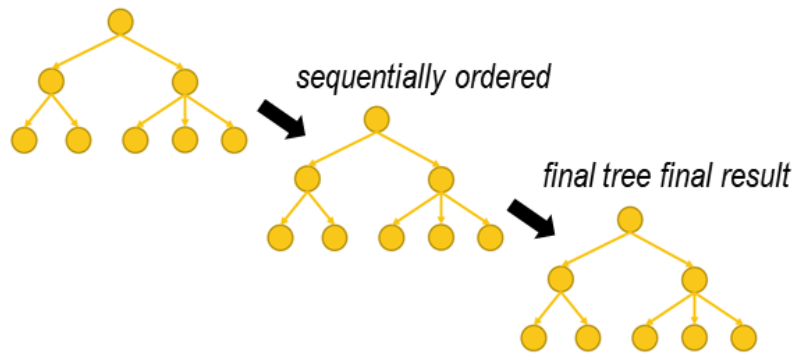


Figuur 9: Schematische weergave van Random Forest

Gradient Boosting

Gradient Boosting is ook een 'tree-based supervised learning' techniek. Bij gradient boosting wordt er gebruik gemaakt van boosting technieken om de voorspellende waarde van classificatie en/of regressie te optimaliseren. Gradient boosted tree, doorgaans ook gewoon gradient boosting genoemd, is net als Random Forest een ensemble methode om beslissingsbomen te verbeteren. In een stapsgewijs proces (in plaats van parallel) worden er meerdere modellen gecreëerd, waarbij elke nieuw model leert van de fouten van het voorgaande model en het generaliseert hen door optimalisatie van een willekeurige differentieerbare verliesfunctie toe te staan. Zie **Figuur 10** voor een illustratie van Gradient Boosting.



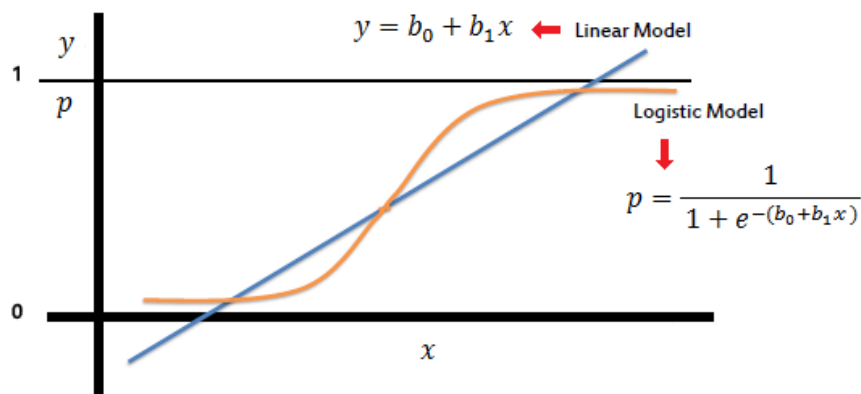


Figuur 10: Schematische weergave van Gradient Boosting

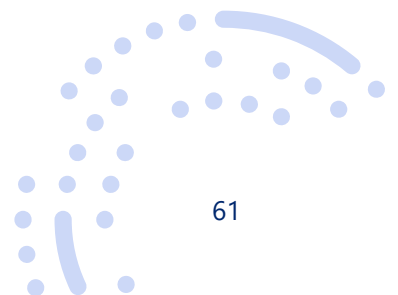
Logistic Regression

Logistic Regression, of logistische regressie, is een classificatietechniek om de (binaire) uitkomst van een afhankelijke variabele te voorspellen met behulp van een reeks onafhankelijke variabelen. Het algoritme gebruikt een lineaire combinatie van meerdere verklarende variabelen net zoals lineaire regressie dat doet (bijv. $Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n$). De lineaire combinatie van variabelen wordt anders dan bij lineaire regressie gebruikt als argument voor de sigmoid-functie. Hiermee wordt er een waarschijnlijkheid afgegeven. De sigmoïde functie kan worden gegeven als $= 1/(1+e^{(-Y)})$.

De corresponderende uitvoer van de sigmoid-functie is een getal tussen 0 en 1. Een invoer die een waarschijnlijkheid $\geq 0,5$ retourneert, wordt bestempeld als klasse 1, terwijl een waarschijnlijkheid $< 0,5$ wordt bestempeld als klasse 0. In tegenstelling tot lineaire regressie, waarbij de afhankelijke variabele continu is, is bij logistische regressie de afhankelijke variabele een categorische (discrete) variabele. Zie **Figuur 11** voor een illustratie van Logistic Regression.



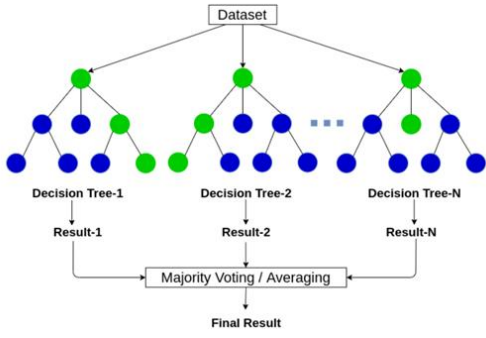
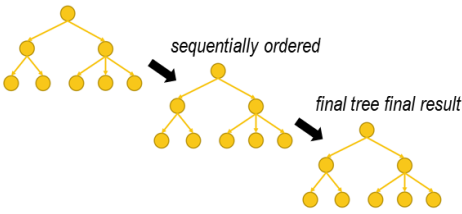
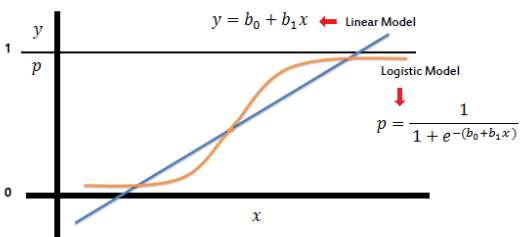
Figuur 11: Schematische weergave van Logistic Regression

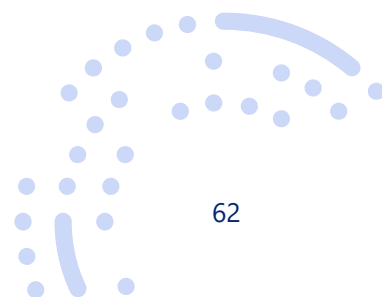




Bijlage G Optimale modelparameters

We presenteren hier de geoptimaliseerde hyperparameters van de drie machine learning-modellen. Deze zijn gevonden aan de hand van een grid-search-techniek. We laten alleen de belangrijkste hyperparameters zien. De resterende hyperparameters zijn van minder belang waarbij standaardwaarden van de algoritmen zijn aangenomen. Zie hiervoor de webpagina van scikit-learn: <https://scikit-learn.org/stable/>.

	<p>Random Forest</p> <ul style="list-style-type: none">• N estimators = 100• Max. depth = 3• Min. samples leaf = 5• Min. samples split = 7• Split criterion = 'gini'• Weighting = 'Balanced'
	<p>Gradient Boosting</p> <ul style="list-style-type: none">• N estimators = 200• Max. depth = 2• Eval metric = 'error'• Objective function = 'binary:logistic'• Learning rate = 0.05• Colsample (all) = 0.3
	<p>Logistic Regression</p> <ul style="list-style-type: none">• N iterations = 500• Regularization type = 'Ridge' (= 'l1')• Regularization par. = 1.0• Solver = 'lbfgs'• Tolerance = 0.05• Weighting = 'Balanced'





Bijlage H Alle voorspellingscores

Hier geven we in meer detail de scores van de prestatie metriekeken voor alle ondermijningsvormen weer. Dit is inclusief mensenhandel die uiteindelijk niet opgenomen is in het model.

Ondermijningsvorm: Drugs

Model	Train accuracy	Test accuracy	Precision (macro)	Recall (macro)	F ₁ -score (macro)
Random Forest	94 %	91 %	74 %	72 %	72 %
Logistic Regression	77 %	78 %	77 %	79 %	78 %
Gradient Boosting	95 %	91 %	91 %	89 %	90 %

Ondermijningsvorm: Witwassen

Model	Train accuracy	Test accuracy	Precision (macro)	Recall (macro)	F ₁ -score (macro)
Random Forest	93 %	83 %	80 %	72 %	75 %
Logistic Regression	90 %	83 %	78 %	78 %	78 %
Gradient Boosting	99 %	83 %	82 %	80 %	80 %

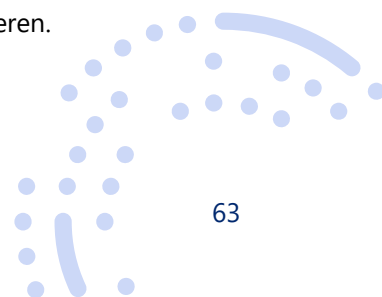
Ondermijningsvorm: Mensenhandel*

Model	Train accuracy	Test accuracy	Precision (macro)	Recall (macro)	F ₁ -score (macro)
Random Forest	99 %	95 %	74 %	74 %	74* %
Logistic Regression	99 %	93 %	65 %	73 %	68* %
Gradient Boosting	99 %	98 %	99 %	75 %	83* %

Ondermijningsvorm: Algemeen

Model	Train accuracy	Test accuracy	Precision (macro)	Recall (macro)	F ₁ -score (macro)
Random Forest	92 %	79 %	71 %	75 %	72 %
Logistic Regression	99 %	79 %	71 %	75 %	72 %
Gradient Boosting	96 %	81 %	74 %	79 %	76 %

* De voorspellingscores voor mensenhandel zijn heel erg gevoelig voor parameterkeuzes zoals bijvoorbeeld de gekozen random seed. Hierdoor kunnen de scores heel erg fluctueren.

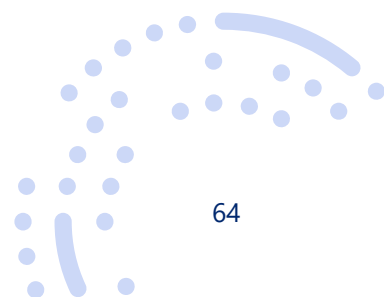
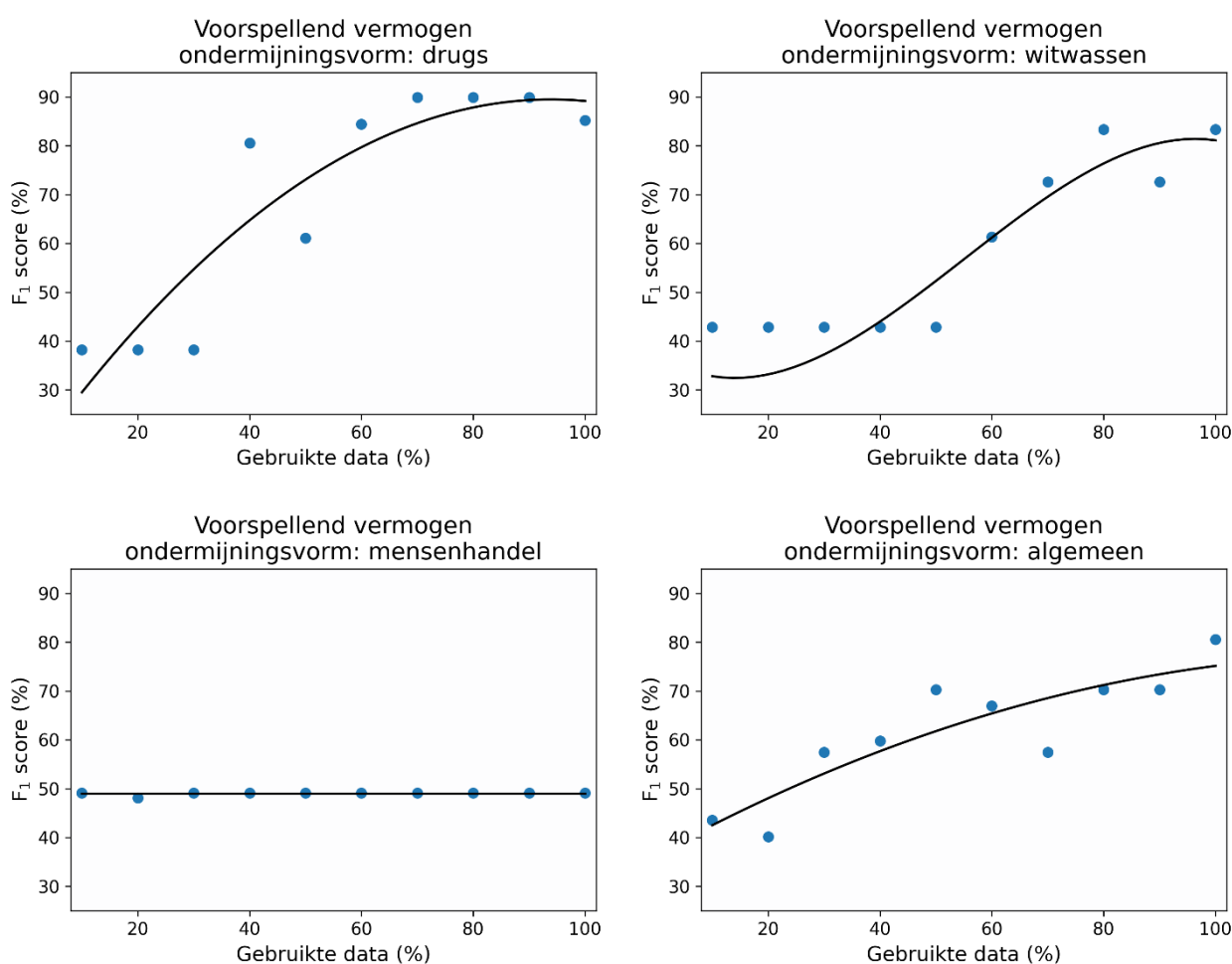




Bijlage I Test op voorspellend vermogen

Hier geven we de impact van datasetgrootte (in fractie van observaties) op de voorspellingsscores weer per ondermijningsvorm. We presenteren hier alleen de grafieken van voorspellend vermogen voor het Gradient Boosting-model. Voor enige robuustheid zijn hier de gemiddelde scores over meerdere parameterkeuzes geplot. Deze kunnen wat afwijken van de eerder gepresenteerde 'beste' scores. De ononderbroken zwarte lijnen vertegenwoordigen de beste (least squares) curve fits volgens een polynoomfunctie op de blauwe datapunten.

Te zien is dat het model voor de ondermijningsvorm mensenhandel niet 'leert' van meer (van de huidige) data; de voorspellingsscores blijven nagenoeg onveranderd. Meer relevante data en meer labels van de minderheidsklasse zou naar verwachting het model voor mensenhandel wel helpen.





Bijlage J Uitleg van top-indicatoren

Voor een deel van onderstaande (40) top-indicatoren uit de Risicotaxatietool is het mogelijk een criminologische verklaring te geven voor de samenhang van de indicator (of het kenmerk) met een verhoogde kwetsbaarheid voor ondermijning. Deze indicatoren zijn vaak ook beïnvloedbaar (zie ook Handreiking Handelingsperspectieven Risicotaxatietool Bedrijventerreinen). Er is getracht de samenhang van de top-indicatoren met ondermijnende criminaliteit te duiden. Dit is gedaan met behulp van inzichten uit bestaande criminologische theorieën en aan de hand van inzichten uit eerder uitgevoerde onderzoeken naar ondermijnende criminaliteit op bedrijventerreinen. Daarnaast is geput uit de verzamelde data uit interviews die tijdens het vooronderzoek en gedurende de praktische validatie uitgevoerd zijn.

Belangrijk voor de interpretatie van de indicatoren is dat zij niet onafhankelijk zijn van elkaar. De combinatie van indicatoren bepaalt de voorspelde kwetsbaarheid van het bedrijventerrein als geheel, en dus niet van specifieke branches of ondernemingen. Wij noemen deze samenhang van indicatoren een "giftige cocktail". Ook wordt er geen onderscheid gemaakt tussen bedrijventerreinen die (enkel) kwetsbaar zijn, en terreinen waar ondermijning zich reeds manifesteert.

Voor een deel van de top-indicatoren uit het Risicotaxatietool kan geen directe verklaring worden gegeven vanuit bestaande criminologische theorie, eerder uitgevoerde onderzoeken of inzichten uit het door ons uitgevoerde onderzoek. Dit geldt voor in totaal 20 van de top-indicatoren, die volgens het ontwikkelde model wel degelijk een voorspellende waarde hebben voor de verhoogde kwetsbaarheid van een bedrijventerrein ten aanzien van ondermijnende criminaliteit. Aanvullend kwalitatief onderzoek zal nodig zijn om inzicht te krijgen in de achterliggende mechanismen van de samenhang van deze vooraf 'onverwachte' indicatoren van ondermijnende criminaliteit. De richting van het verband tussen deze indicatoren en de verschillende ondermijningsvormen kan u vinden in **Figuur 4**.

Hieronder worden de top-indicatoren toegelicht en waar mogelijk van een duiding voorzien.

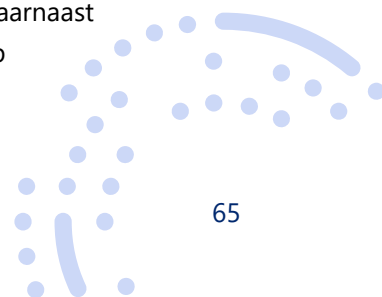
Lijst van top-indicatoren met uitleg en duiding

Verouderd bedrijventerrein (IBIS)

Er is sprake van veroudering als de kwaliteit van het terrein niet langer voldoet aan de vestigingseisen van bedrijven. Er worden verschillende typen veroudering onderscheiden: technische, economische, maatschappelijke, en ruimtelijke. De oorzaak wordt aangegeven via andere variabelen (*hoofdoorzaak veroudering A, B, C, en D*).

Eenheid: Absoluut

Duiding: Veroudering van een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen, drugsgereleerde criminaliteit en witwassen op bedrijventerreinen. Veroudering van een bedrijventerrein hangt samen met o.a. achterstallig onderhoud en beheer, leegstand en verouderd vastgoed. Dit maakt een bedrijventerrein minder aantrekkelijk voor nieuwe ondernemingen die hogere eisen stellen aan de kwaliteit van het vastgoed en de (technische) infrastructuur. Daarnaast kan veroudering leiden tot verloederingsomgeving. Een verloederde omgeving kan duiden op





afwezigheid van formeel en sociale controle. Dit kan een aanzuigende werking hebben voor criminelen omdat zij gebaat zijn bij weinig toezicht.

Aantal automotive bedrijven (KVK)

Het aantal automotive bedrijven op een bedrijventerrein. Een bedrijf wordt aangemerkt als automotive als de SBI-code van de hoofd- of nevenactiviteit voorkomt in de lijst met automotive SBI-codes. De SBI-codes worden door KVK toegekend.

Eenheid: Per oppervlak (ha)

Duiding: Een hoog aantal autobedrijven duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen, drugsgerelateerde criminaliteit en witwassen op bedrijventerreinen. De autobranche is een kwetsbare branche voor (onbedoelde) facilitering of het plegen van ondermijnende criminaliteit. De branche voorziet in diverse diensten en producten die benodigd zijn voor het criminele bedrijfsproces zoals het leveren van vervoer middels huurauto's, het inbouwen van verborgen ruimten en heling van onderdelen. De autobranche biedt hiernaast mogelijkheden tot witwassen en btw-fraude via de import en export van voertuigen. Het gaat hierbij dus om de sector in de breedte.

Klachtmeldingen alle ondermijningsvormen (S@men)

Aantal klachtmeldingen gerelateerd aan de drie typen ondermijning (drugs, witwassen, en mensenhandel) die vallen op postcodes binnen bedrijventerreinen. De klachten over leefomgeving komen binnen bij ondernemingsdiensten via het platform S@men. Dit is een samenwerkingsverband tussen provincie Gelderland, provincie Noord-Brabant, de Gelderse, Brabantse en Zeeuwse Omgevingsdiensten en een aantal Gelderse gemeenten. De lijst van mogelijke klachten zijn: illegale bouwsels, onveiligheid, chemische lucht / teer / asfalt, water en chemische verontreiniging, bodemverontreiniging, zoet-zuur.

Eenheid: Per oppervlak (ha)

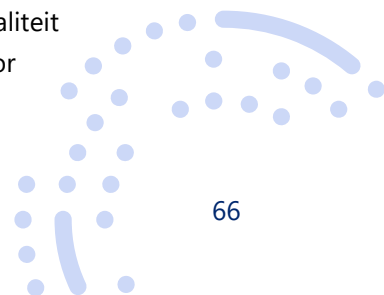
Duiding: Een hoog aantal klachtmeldingen op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen, drugsgerelateerde criminaliteit en witwassen op bedrijventerreinen. Klachtmeldingen duiden op onregelmatigheden op het bedrijventerrein. Deze signalen kunnen veroorzaakt worden door onderliggende ondermijningsproblematiek zoals overlast door illegale huisvesting, onveiligheidsgevoelens door intimidatie en ander algemeen regel-overtredend gedrag door in het gebied aanwezige personen.

Risicovolle combinatie automotive en drugs (KVK)

Percentage bedrijven op een bedrijventerrein die de risicovolle combinatie hebben tussen automotive sector en risicobranche voor drugs. Deze combinatie is een opvallende combinatie. Hiervoor worden er naar de SBI-codes van zowel de hoofd- als de nevenactiviteiten van bedrijven gekeken. Een lijst met SBI-codes die behoren tot de risicobranche is opgesteld door criminologen. De SBI-codes worden door KVK toegekend.

Eenheid: Percentage (%)

Duiding: Een hoog percentage bedrijven met een risicovolle combinatie van bedrijfsactiviteiten op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen, drugsgerelateerde criminaliteit en witwassen op bedrijventerreinen. De autobranche is een kwetsbare branche voor





(onbedoelde) facilitering of het plegen van ondermijnende criminaliteit. Het is opvallend wanneer bedrijven in deze branche tevens bedrijfsactiviteiten uitvoeren uit andere branches die kwetsbaar zijn, zoals bijvoorbeeld de handel in sieraden of activiteiten in de beveiligingsbranche.

Minimum verkoopprijs grond (IBIS)

De minimum verkoopprijs van grond (kavels) per m² op een bedrijventerrein (in € exclusief BTW). Als er voor bepaalde kavels toeslagen of kortingen gelden, dan kan dat verwerkt zijn in de minimumprijs.

Eenheid: Absoluut

Duiding: Een lage minimum verkoopprijs van grond op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor drugsgerelateerde criminaliteit op bedrijventerreinen. Een lage minimum verkoopprijs van grond hangt samen met een lage marktwaarde, wat een signaal kan zijn dat het gebied niet populair is om in te investeren voor het bouwen van vastgoed. Wanneer investeringen in het gebied uitblijven kan een negatieve spiraal in werking treden waardoor het gebied steeds onaantrekkelijker maakt voor nieuwe bedrijven en investeerders. Dit kan leiden tot verloedering.

Geregistreerde misdrijven: drugsproductie (BVH)

Aantal drugsproductie gerelateerde misdrijven in de BVH-data op een bedrijventerrein. In het BVH-systeem registreert de politie meldingen en aangiften om inzicht te krijgen in de geregistreerde omvang van crimineel gebruik van bedrijfsruimten. De registratie in BVH vindt plaats aan de hand van een maatschappelijke code. Onder de drugsproductie categorie vallen: vervaardigen harddrugs (F44) en vervaardigen softdrugs (code F45).

Eenheid: Per oppervlak (ha)

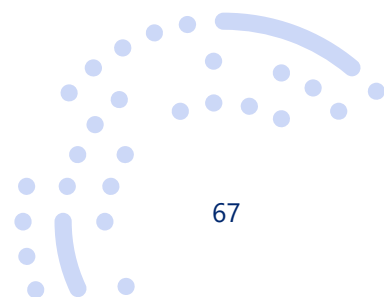
Duiding: Een hoog aantal geregistreerde misdrijven in relatie tot drugsproductie op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen en drugsgerelateerde criminaliteit op bedrijventerreinen. Geregistreerde misdrijven op het vlak van drugsproductie zijn vaak 'het topje van de ijsberg', welke kunnen duiden op al aanwezige drugsproductie in het gebied. Hiernaast kan het terrein specifieke kenmerken hebben die het aantrekkelijk maakt voor het vestigen van een drugsproductielocatie. Uit onderzoek blijkt dat drugsproductielocaties en drugsopslag vaak worden aangetroffen in panden waar dit al eens het geval is geweest.

Geregistreerde misdrijven: totaal (BVH)

De som van het aantal misdrijven in de BVH-data die vallen onder de categorieën voor drugs en mensenhandel op een bedrijventerrein. In het BVH-systeem registreert de politie meldingen en aangiften om inzicht te krijgen in de geregistreerde omvang van crimineel gebruik van bedrijfsruimten. De registratie in BVH vindt plaats aan de hand van een maatschappelijke code. Onder de categorieën voor drugs en mensenhandel vallen: de codes F42 t/m F47, F5294, en F561 t/m F565.

Eenheid: Per oppervlak (ha)

Duiding: Een hoog aantal geregistreerde misdrijven op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen en drugsgerelateerde criminaliteit op bedrijventerreinen. Geregistreerde misdrijven zijn vaak 'het topje van de ijsberg', welke kunnen duiden op al aanwezige ondermijningsproblematiek in het gebied. Hiernaast kan het terrein specifieke kenmerken hebben die het aantrekkelijk maakt voor criminelen.





Percentage bedrijven risicobranche drugs (KVK)

Het gemiddelde aantal van bedrijven op een bedrijventerrein die behoren tot de risicobranche voor drugs. Hiervoor worden er naar de SBI-codes van zowel de hoofd- als de nevenactiviteiten van bedrijven gekeken. Een lijst met SBI-codes die behoren tot de risicobranche is opgesteld door criminologen. De SBI-codes worden door KVK toegekend.

Eenheid: Percentage (%)

Duiding: Een hoog percentage bedrijven behorende tot een risicobranche voor drugscriminaliteit op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen en drugsgelateerde criminaliteit op bedrijventerreinen. Een concentratie van branches die verhoogd kwetsbaar zijn voor (onbedoelde) facilitering of het plegen van ondermijnende criminaliteit, verhoogt de kwetsbaarheid van het terrein als geheel.

Percentage bedrijven risicobranche witwassen (KVK)

Het gemiddelde aantal van bedrijven op een bedrijventerrein die behoren tot de risicobranche voor witwassen. Hiervoor worden er naar de SBI-codes van zowel de hoofd- als de nevenactiviteiten van bedrijven gekeken. Een lijst met SBI-codes die behoren tot de risicobranche is opgesteld door criminologen. De SBI-codes worden door KVK toegekend.

Eenheid: Percentage (%)

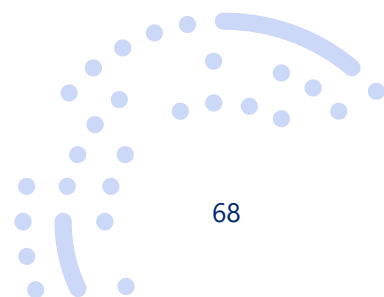
Duiding: Een hoog percentage bedrijven behorende tot een risicobranche voor witwassen op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor drugsgelateerde criminaliteit op bedrijventerreinen. Een concentratie van branches die verhoogd kwetsbaar zijn voor (onbedoelde) facilitering of het plegen van ondermijnende criminaliteit, verhoogt de kwetsbaarheid van het terrein als geheel.

Gebouwen met lage WOZ-waarde (KWB)

Percentage van postcodegebieden op een bedrijventerrein met een lage gemiddelde WOZ-waarde. Een lage WOZ-waarde wordt toegekend als de gemiddelde WOZ-waarde valt in de onderste 25 percentiel van alle WOZ-waardes in en rondom bedrijventerreinen van de provincies Noord-Brabant en Gelderland. In 2021 komt dit overeen met een WOZ-grenswaarde van 226k. Alles kleiner dan 226k wordt aangemerkt als een lage WOZ-waarde.

Eenheid: Percentage (%)

Duiding: Een hoog percentage bouwen met een lage WOZ-waarde op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen en drugsgelateerde criminaliteit op bedrijventerreinen. Een lage WOZ-waarde hangt samen met een lage marktwaarde, wat een signaal kan zijn dat het gebied niet populair is voor ondernemers om zich te vestigen doordat dat de panden bijvoorbeeld erg klein of verouderd zijn. Ondermijningsproblematiek wordt veelal gesignaleerd in concentraties van kleine, verouderde panden op een bedrijventerrein. Hierdoor kan een negatieve spiraal in werking treden waardoor het bedrijventerrein steeds onaantrekkelijker wordt voor nieuwe bedrijven en investeerders.





Rechtsvorm VOF (KVK)

Percentage bedrijven op een bedrijventerrein met de rechtsvorm vennootschap onder firma. De rechtsvorm geeft aan waaronder een vestiging, vereniging, of stichting aan het economisch en/of maatschappelijk verkeer deelneemt.

Eenheid: Percentage (%)

Duiding: Onbekend

Aantal bedrijven (LISA)

Totaal aantal bedrijven op een bedrijventerrein zoals die geregistreerd staat in het landelijk informatiesysteem (LISA) van 2019.

Eenheid: Absoluut

Duiding: Een hoge dichtheid van bedrijven op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor drugsgerelateerde criminaliteit op bedrijventerreinen. Een relatief hoge dichtheid van ondernemingen zou kunnen duiden op een concentratie van kleine ondernemingen op locaties als opgesplitste loodsen, garageboxen en bedrijfsverzamelgebouwen. De anonimiteit die dergelijke locaties met zich meebrengen bieden een gelegenheidsstructuur voor het ontplooiën van criminele activiteiten.

Informeel toezicht (BAG & CBS)

Deze indicator is opgebouwd uit meerdere variabelen. Het geeft het aantal panden weer met een publieke samenkomstfunctie op een bedrijventerrein (BAG) plus het aantal publieke voorzieningen binnen een straal van 1 km rondom een bedrijventerrein (CBS). De samenkomstfuncties zijn: winkelfunctie, bijeenkomstfunctie, sportfunctie, onderwijsfunctie, gezondheidszorgfunctie, en logiesfunctie. De publieke voorzieningen zijn: school, kinderdagverblijf, buitenschoolse opvang, huisartsenpraktijk, fysiotherapeut, grote supermarkt, overige dagelijkse levensmiddelen, café, cafetaria, en restaurant.

Eenheid: Absoluut

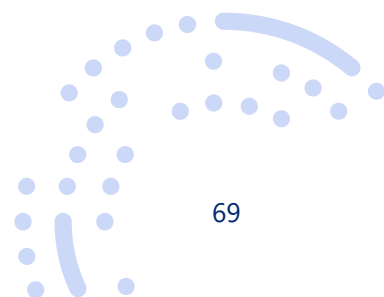
Duiding: Onbekend

Leefbaarometer score (Leefbaarometer)

Een modelscore die laat zien wat de leefbaarheid van een buurt is. De gemiddelde score van alle buurten in en rondom een bedrijventerrein wordt hier genomen. Leefbaarheid is gedefinieerd als de mate waarin de leefomgeving aansluit bij de voorwaarden en behoeften die er door de mens aan worden gesteld. Het geeft de situatie in de wijk weer, maar ook ontwikkelingen en achtergronden van de buurt. Om leefbaarheid in beeld te brengen wordt gebruik gemaakt van 100 objectieve indicatoren, onderverdeeld in 5 dimensies. De leefbaarometerscores van 2018 zijn hierbij gebruikt.

Eenheid: Gemiddeld (buurten)

Duiding: Een hoge score op de Leefbaarometer op een bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verlaagde kwetsbaarheid voor ondermijning algemeen en drugsgerelateerde criminaliteit op bedrijventerreinen. Een hoge leefbaarometer score zegt iets over de kwaliteit van (de omgeving van) het bedrijventerrein. Dat de omgeving van het terrein positief wordt ervaren door bewoners en ondernemers duidt op een gedeeld verantwoordelijkheidsgevoel waarin verloedering en ondermijning moeilijk voet aan de grond kunnen krijgen.





Klachtmeldingen over milieu (S@men)

Aantal klachtmeldingen gerelateerd aan milieu die vallen op postcodes binnen bedrijventerreinen. De klachten over leefomgeving komen binnen bij ondernemingsdiensten via het platform S@men. S@men is een samenwerkingsverband tussen provincie Gelderland, provincie Noord-Brabant, de Gelderse, Brabantse en Zeeuwse Omgevingsdiensten en een aantal Gelderse gemeenten. De lijst van mogelijke klachten zijn: water en chemische verontreiniging, bodemverontreiniging, aantasting beschermde dier- of plantensoorten, rook, stof, uitlaatgassen / smog, fakkelen, ongedierte / plagen, illegale houtkap.

Eenheid: Per oppervlak (ha)

Duiding: Onbekend

Onbekende inschrijvingsredenen KVK (KVK)

Percentage van bedrijven op een bedrijventerrein waarbij de reden van inschrijving bij het KVK niet bekend is. Dit kan zijn omdat ze bijvoorbeeld niet zijn ingevuld of omdat het niet om een onderneming gaat. Sinds 1986 wordt deze rubriek bij nieuwe inschrijvingen voor ondernemingen altijd ingevuld. Ongeveer een op acht van de KVK-inschrijvingsredenen zijn onbekend.

Eenheid: Percentage (%)

Duiding: Onbekend

Herstructureringsplan onbekend (IBIS)

Verouderde bedrijventerreinen krijgen soms een herstructureringsplan van de gemeente. Is er een herstructureringsplan aanwezig voor het bedrijventerrein? Mogelijkheden zijn: ja, nee, of onbekend. Van de bedrijventerreinen hebben 79% een onbekend herstructureringsplan (13 % ja en 8 % nee). Er kan ook aangegeven worden wat voor soort herstructureringsplan er is. Dat wordt middels een ander variabele opgevangen.

Eenheid: Absoluut

Duiding: Onbekend

KVK-inschrijvingsredenen nieuwe vestiging (KVK)

Percentage van bedrijven op een bedrijventerrein waarbij de reden van inschrijving bij het KVK gaat om een nieuwe vestiging (oprichting algemeen). Sinds 1986 wordt deze rubriek bij nieuwe inschrijvingen voor ondernemingen altijd ingevuld. Dit is de vaakst voorkomende inschrijvingsredenen (42.9%).

Eenheid: Percentage (%)

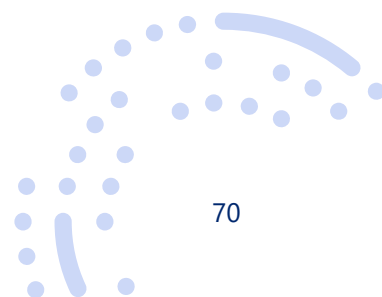
Duiding: Onbekend

Percentage ondernemingen (KVK)

Percentage van bedrijven op een bedrijventerrein die vallen onder soort organisatie onderneming. Van alle KVK-inschrijvingen gaat het in 89,9% van de gevallen om een onderneming. De andere mogelijkheden op bedrijventerreinen in de provincies zijn: statutaire inschrijving (0,5%), stichting (5,4%), vereniging (4,0%), kerkgenootschap (0,04%), en publiekrechtelijk rechtspersoon (0,2%). Er komen geen privaatrechtelijk rechtspersonen voor.

Eenheid: Percentage (%)

Duiding: Onbekend





Percentage ontbonden bedrijven (KVK)

Het percentage van bedrijven op een bedrijventerrein die om welk reden dan ook ontbonden zijn. De redenen of soort ontbinding van de inschrijving kunnen zijn: besluit van de algemene vergadering / stichtingsbestuur, rechterlijke uitspraak, statutaire bepaling, staat van insolventie, opheffing faillissement wegens de toestand van de boedel, geheel ontbreken van leden, verstrijken van de duur, uittreden van een of meer vennoten, faillissement van een of meer vennoten, insolventie of opheffing faillissement wegens toestand boedel, ontbinding door de Kamers van Koophandel. Een derde van alle bedrijventerreinen hebben een of meer bedrijven die ontbonden zijn. Het gemiddelde ontbindingspercentage van bedrijventerreinen is 0,6%.

Eenheid: Percentage (%)

Duiding: Een relatief hoog percentage ontbonden bedrijven op het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen, drugsgesrelateerde criminaliteit en witwassen op bedrijventerreinen. Wanneer op een bedrijventerrein ondernemingen ophouden te bestaan kan dat betekenen dat het terrein minder vitaal wordt. Deze economische achteruitgang werkt verdere achteruitgang in de hand, wat het terrein minder aantrekkelijk maakt voor investeerders en bonafide ondernemers. Ook kan er sprake zijn van een verhullingstactiek, doordat het ontbinden (en opnieuw starten) van ondernemingen overheidsdiensten belemmert om een informatiebeeld op te bouwen op specifieke ondernemingen.

Aantal panden met een industriefunctie (BAG)

Het BAG bevat gemeentelijke basisgegevens van alle adressen en gebouwen en registreert daarbij de gebruiksdoelen en functies van panden. Hier gaat het om het aantal panden op een bedrijventerrein met functie industrie. Onder industriefunctie vallen de volgende gebruiksdoelen: opslag, bedrijf, niet voor publiek toegankelijk atelier, fabriek, opslag/distributie, agrarisch object/kas, werkplaats, en montagehal.

Eenheid: Per oppervlak (ha)

Duiding: Onbekend

Percentage omliggende meergezinswoningen (KWB)

Het aantal meergezinswoningen in en rondom bedrijventerreinen vermeld als percentage van de totale woningvoorraad. Wat is een meergezinswoning? Elke woning die samen met andere woonruimten c.q. bedrijfsruimten een geheel pand vormt. Hieronder vallen flats, galerij-, portiek-, beneden- en bovenwoningen, appartementen en woningen boven bedrijfsruimten, voor zover deze zijn voorzien van een buiten de bedrijfsruimte gelegen toegangsdeur.

Eenheid: Percentage (%)

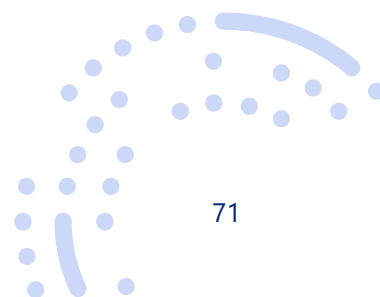
Duiding: Onbekend

Fulltime werkzame mannen (LISA)

Het totaal aan fulltime werkende mannen op een bedrijventerrein. Dit aantal is gemeten en opgenomen in het landelijk informatiesysteem (LISA) in 2019.

Eenheid: Absoluut

Duiding: Onbekend





Milieuozonering aanwezig (IBIS)

De milieuozonering zorgt ervoor dat nieuwe bedrijven een juiste plek krijgen in de nabijheid van woningen. Indien de gemeente in het kader van de vaststelling of herziening van een bestemmingsplan besluit om op een werklocatie zogenaamde grote lawaaimakers toe te laten, moet een **geluidzone** rond de werklocatie worden vastgesteld. Het gaat hierbij om bedrijven die zeer veel geluid produceren en als zodanig in het Inrichtingen en Vergunningenbesluit (IVB) bij de Wet milieubeheer zijn aangewezen. Deze zogenaamde art. 2.4 IVB-inrichtingen waren vroeger bekend als A-inrichtingen in het kader van de Wet geluidhinder. Indien deze bedrijven op een werklocatie voorkomen of in het bestemmingsplan niet uitdrukkelijk zijn uitgesloten, moet in het bestemmingsplan een zonering voor geluid worden opgenomen.

Eenheid: Absoluut

Duiding: Onbekend

Gemiddeld oprichtingsjaar bedrijven (KVK)

Het gemiddelde jaar van oprichting van bedrijven op een bedrijventerrein waarop de rechtspersoon, VOF of CV is opgericht. Het oprichtingsjaar is niet altijd gevuld bij een NV of BV. Onbekende gevallen worden niet meegenomen in de middeling.

Eenheid: Gemiddeld (bedrijven)

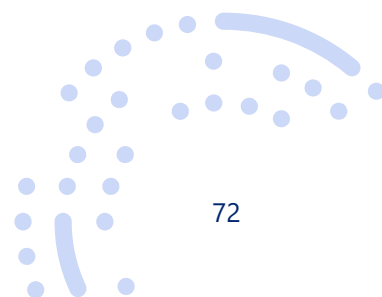
Duiding: Een gemiddeld oud oprichtingsjaar van bedrijven duidt met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen op bedrijventerreinen. De gemiddelde leeftijd van bedrijven kan samenhangen met de leeftijd van het terrein als geheel, wat duidt op veroudering. Veroudering van een bedrijventerrein hangt samen met o.a. achterstallig onderhoud en beheer, leegstand en verouderd vastgoed. Dit maakt een bedrijventerrein minder aantrekkelijk voor nieuwe ondernemingen die hogere eisen stellen aan de kwaliteit van het vastgoed en de (technische) infrastructuur. Daarnaast leidt veroudering vaak tot verloedering. Een verloederde omgeving kan worden gepercipieerd aan een gebrek aan formeel en informeel toezicht, uit wat een aanzuigende werking heeft voor criminelen omdat zij gebaat zijn bij weinig toezicht.

WOZ-waarde (KWB)

De gemiddelde waarde onroerende zaken van woonobjecten in en rondom bedrijventerreinen gebaseerd op de Wet Waardering Onroerende Zaken. Voor de bepaling van de gemiddelde woningwaarde is alleen gebruik gemaakt van die WOZ-objecten omschreven als woningen dienend tot hoofdverblijf en woningen met praktijkruimte met een waarde groter dan nul euro. De gemiddelde woningwaarde wordt bepaald met de waardepeildatum van voorgaand jaar (hier 1-1-2019).

Eenheid: Gemiddeld (woningen)

Duiding: Een hoge WOZ-waarde van woonobjecten in en rondom het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verlaagde kwetsbaarheid voor ondermijning algemeen en witwassen op bedrijventerreinen. Naar verwachting duidt deze indicator op een hoge kwaliteit van vastgoed in het gebied, wat de aantrekkelijkheid voor criminele activiteiten verlaagt.





Schommelingen in overlast dronkenschap (Politie)

Dit laat de schommelingen zien in geregistreerde overlastmeldingen van dronkenschap in en rondom bedrijventerreinen. De schommelingen zijn gemeten over een periode van 5 jaar (hier 2015 tot 2020).

Eenheid: Per oppervlak (ha)

Duiding: Onbekend

Fulltime werkzame vrouwen (LISA)

Het totaal aan fulltime werkende vrouwen op een bedrijventerrein. Dit aantal is gemeten en opgenomen in het landelijk informatiesysteem (LISA) in 2019.

Eenheid: Absoluut

Duiding: Onbekend

Aantal bedrijven met een ANBI-status (ANBI)

Totaal aantal bedrijven of instellingen op een bedrijventerrein die geregistreerd staan als algemeen nut beogende instellingen in 2021.

Eenheid: Absoluut

Duiding: Een hoog aantal bedrijven met ANBI-status op het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor ondermijning algemeen op bedrijventerreinen. Instellingen met deze status zijn kwetsbaar voor misbruik doordat zij belastingvoordelen genieten. Ook blijkt uit onderzoek dat c.q. 50% van deze instellen de verplichting om inzage te geven in de eigen financiële gangen niet nakomt.

Klachtmeldingen over afval (S@men)

Aantal klachtmeldingen gerelateerd aan afval die vallen op postcodes binnen bedrijventerreinen. De klachten over leefomgeving komen binnen bij ondernemingsdiensten via het platform S@men.

S@men is een samenwerkingsverband tussen provincie Gelderland, provincie Noord-Brabant, de Gelderse, Brabantse en Zeeuwse Omgevingsdiensten en een aantal Gelderse gemeenten. De lijst van mogelijke klachten zijn: afval en verontreiniging.

Eenheid: Per oppervlak (ha)

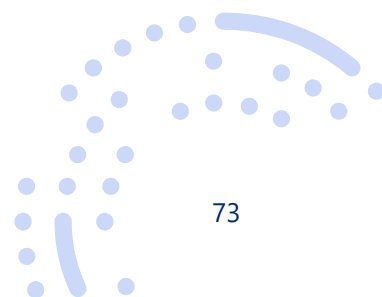
Duiding: Onbekend

Bedrijven met exportfunctie (KVK)

Percentage ondernemingen op een bedrijventerrein met de distributiefunctie export. Dit geeft dus aan of de vestiging exporteert. Ondernemers zijn niet verplicht deze informatie op te geven. 13.5% van alle bedrijven in de provincies hebben een exportfunctie. 91.4% van alle bedrijventerreinen hebben minimaal een bedrijf met exportfunctie.

Eenheid: Percentage (%)

Duiding: Een hoog percentage bedrijven met exportfunctie op het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor witwassen op bedrijventerreinen. Het exporteren van goederen is een bedrijfsactiviteit die kwetsbaar is voor witwassen vanwege beperkte gegevensdeling met andere landen en verschillen in belastingconstructies.





Geplaatst kapitaal (KVK)

Dit geeft het gemiddelde bedrag van geplaatst kapitaal op een bedrijventerrein weer waarvoor aandelen van naamloze vennootschappen zijn geplaatst. Het bedrag staat weergegeven in € 1000. Minimaal 20% van het maatschappelijk kapitaal moet geplaatst zijn. Minimaal geplaatst kapitaal voor NV is € 45.000,-. Bij 63% van alle bedrijven in de provincies is er een kapitaal vermeld van meer dan 0 euro.

Eenheid: Gemiddeld (bedrijven)

Duiding: Onbekend

Aantal attracties binnen 50 km (CBS)

Het aantal attracties binnen een straal van 50 km rondom een bedrijventerrein gemeten over de weg voor alle inwoners in dat gebied. Hierbij is het gemiddelde aantal van alle inwoners genomen.

Attractie: Pretpark, dierentuin en binnenspeeltuin.

Eenheid: Absoluut

Duiding: Onbekend

Geregistreerde misdrijven: drugs overig (BVH)

Aantal drugs gerelateerde misdrijven in de BVH-data die vallen onder de overige drugs categorie op een bedrijventerrein. In het BVH-systeem registreert de politie meldingen en aangiften om inzicht te krijgen in de geregistreerde omvang van crimineel gebruik van bedrijfsruimten. De registratie in BVH vindt plaats aan de hand van een maatschappelijke code. Onder de overige drugs categorie vallen: aantreffen drugs, geen verdachte (code F46) en overige drugsdelicten (code F47).

Eenheid: Per oppervlak (ha)

Duiding: Onbekend

Percentage statutaire inschrijving (KVK)

Percentage van bedrijven op een bedrijventerrein die vallen onder soort organisatie statutaire inschrijving. Van alle KVK-inschrijvingen gaat het in 0,5 % van de gevallen om een statutaire inschrijving. De andere mogelijkheden op bedrijventerreinen in de provincies zijn: onderneming (89,9%), stichting (5,4%), vereniging (4,0%), kerkgenootschap (0,04%), en publiekrechtelijk rechtspersoon (0,2%). Er komen geen privaatrechtelijk rechtspersonen voor.

Eenheid: Percentage (%)

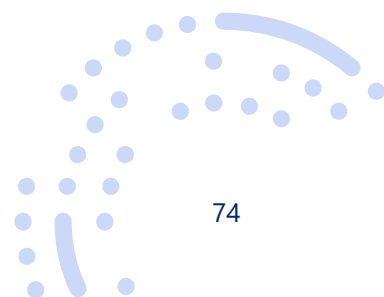
Duiding: Onbekend

KVK-inschrijvingsreden kamergrenswijziging (KVK)

Percentage van bedrijven op een bedrijventerrein waarbij de reden van inschrijving bij het KVK gaat om een Kamergrenswijziging. Sinds 1986 wordt deze rubriek bij nieuwe inschrijvingen voor ondernemingen altijd ingevuld. Deze inschrijvingsredenen bij het KVK komt weinig voor bij ondernemingen (1.2%).

Eenheid: Percentage (%)

Duiding: Onbekend





Aantal bedrijven risicobranche mensenhandel (KVK)

Het totaal aantal bedrijven op een bedrijventerrein die behoren tot de risicobranche voor mensenhandel. Hiervoor worden er naar de SBI-codes van zowel de hoofd- als de nevenactiviteiten van bedrijven gekeken. Een lijst met SBI-codes die behoren tot de risicobranche is opgesteld door criminologen. De SBI-codes worden door KVK toegekend.

Eenheid: Per oppervlak (ha)

Duiding: Een hoog aantal bedrijven behorende tot een risicobranche voor mensenhandel op het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verhoogde kwetsbaarheid voor witwassen op bedrijventerreinen. Een concentratie van branches die verhoogd kwetsbaar zijn voor (onbedoelde) facilitering of het plegen van ondermijnende criminaliteit, verhoogt de kwetsbaarheid van het terrein als geheel.

Werkzame personen onderneming gemiddeld (KVK)

Het gemiddeld aantal werkzame personen (parttime en fulltime) van de totale onderneming van alle ondernemingen op een bedrijventerrein. Een persoon is parttime werkzaam als het aantal gewerkte uren per week minder is dan vijftien.

Eenheid: Gemiddeld (totale ondernemingen)

Duiding: Een hoog gemiddeld aantal werkzame personen duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verlaagde kwetsbaarheid voor witwassen op bedrijventerreinen. Kleine ondernemingen zijn kwetsbaar om misbruikt te worden voor het witwassen van geld omdat hier minder (juridisch vereiste) controlemechanismen gelden.

Percentage niet-drinkers (CBS)

Het percentage personen in en rondom bedrijventerreinen van 19 jaar en ouder dat geen of maximaal een glas alcohol per dag drinkt. De percentages zijn gebaseerd op een modelmatige schatting door het CBS. De percentages worden door het CBS gegeven op buurniveau. Voor bedrijventerreinen worden de percentages van buurten in en rondom bedrijventerreinen gemiddeld.

Eenheid: Percentage (%)

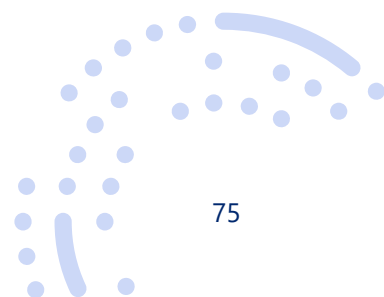
Duiding: Onbekend

Aantal auto's per huishouden (KWB)

Het gemiddeld aantal personenauto's per (particulier) huishouden in of rondom bedrijventerreinen. De personenauto's worden regionaal ingedeeld met behulp van de kentekenregistratie. Personenauto's die geregistreerd staan op het adres van het lease- of verhuurbedrijf vertekenen daarom de autodichtheid per huishouden. Het aantal personenauto's per huishouden is vermeld bij een waarde van maximaal 2,5 personenauto's per huishouden.

Eenheid: Gemiddeld (huishoudens)

Duiding: Een gemiddeld hoog aantal auto's per huishouden rondom of op het bedrijventerrein duidt in samenhang met de andere risico-indicatoren in de risicotaxatietool op een verlaagde kwetsbaarheid voor witwassen op bedrijventerreinen. Deze indicator kan er op duiden dat de personen die op en rondom het bedrijventerrein wonen financieel sterk staan. Een vitaal gebied is minder aantrekkelijk voor criminelen omdat zij moeilijker gebruik kunnen maken van de financiële kwetsbaarheid van bijvoorbeeld ondernemers en pandeigenaren.





Centerdata

Postbus 90153
5000 LE Tilburg

+31 (0)13 206 3500
info@centerdata.nl

Contactpersoon

Seyit Höcük
+31 (0)13 206 3531
seyit.hocuk@centerdata.nl

centerdata.nl