

Indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen: een verkennend onderzoek in de gemeente Tilburg

Eindrapport



datum	12 december 2018
auteurs	Yvonne Bolsius MSc dr. Seyit Höcük dr. Patricia Prüfer prof. dr. Emile Kolthoff
versie	1.0



© CentERdata, Tilburg, 2018

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.



Managementsamenvatting

Het doel van dit onderzoek was te bepalen of er indicatoren konden worden gevonden waarmee het voor bestuurders inzichtelijk wordt in welk stadium een bedrijventerrein verkeert in de ontwikkeling of aanwezigheid van (georganiseerde) criminaliteit en ondermijning. Aan de hand van welke indicatoren kan inzichtelijk worden gemaakt waar een bedrijventerrein zich op de 'glijdende schaal' van (georganiseerde) criminaliteit en ondermijning bevindt?

Aanleiding onderzoek en context

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 en landelijk niveau intensivering van de samenwerking tussen betrokken overheden, hun private en maatschappelijke partners en burgers.

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 en Van der Torre, 2015).¹ 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 onderwereld; daartussen bestaan allerlei vormen variërend van bewust weggijken via een graantje meepikken tot aan het verlenen van hand- en spandiensten aan criminelen.

Definities en inkadering

In deze studie is een bedrijventerrein gedefinieerd als een werklocatie van minimaal 1 ha bruto bestemd en geschikt voor gebruik door handel, nijverheid en industrie. Terreinen die buiten deze definitie vallen zijn onder andere zeehaventerreinen, economische zones en kantorenlocaties. Bij het identificeren van de indicatoren en de dataverzameling lag de focus op drugsgelateerde criminaliteit (productie van en handel in hennep en synthetische drugs). Daarnaast is in dit verkennende onderzoek ervoor gekozen om de dataverzameling geografisch te beperken tot 30 bedrijventerreinen in de gemeente Tilburg. Dit betekent dat, hoewel de aanpak en analyses getest konden worden, de resultaten van de analyses door de kleine steekproef van 30 bedrijventerreinen minder betrouwbaar zijn.

Onderzoeksaanpak

Het onderzoek is gestart met het identificeren van potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Dit is gedaan in een vooronderzoek bestaande uit een literatuurstudie, een expertmeeting en interviews. Vervolgens is data met betrekking tot de mogelijke indicatoren verzameld uit diverse bronnen. Om de mate van georganiseerde criminaliteit en ondermijning per bedrijventerrein te inventariseren is een kleine vragenlijst uitgezet waarin respondenten de bedrijventerreinen konden beoordelen². De

¹ Tops, P. & E. van der Torre (2015). Wijkenaanpak en ondermijnende criminaliteit. Den Haag: Boom Lemma.

² Hoe de mate van (georganiseerde) criminaliteit en ondermijning precies is geoperationaliseerd als afhankelijke variabele wordt verder toegelicht bij onderzoeksvraag 3.



data over de potentiële indicatoren en de mate van (georganiseerde) criminaliteit en ondermijning op het bedrijventerrein zijn indien nodig geaggregeerd en daarna gekoppeld, waarna geanalyseerd kon worden welke potentiële indicatoren daadwerkelijk voorspellende waarde hebben voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Om de belangrijkste indicatoren te vinden hebben we gebruik gemaakt van de data science methode Random Forest, waarna we op basis van de 15 belangrijkste indicatoren nog beslissingsbomen hebben gedraaid om de resultaten beter te kunnen duiden. We lichten nu kort de antwoorden op de zes onderzoeksvragen toe.

1. Wat zijn potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?

Ten eerste zijn in het vooronderzoek mogelijke indicatoren geïdentificeerd met betrekking tot de fysieke omgeving (waaronder onderhoud, veroudering en leegstand) en toezicht, handhaving en samenhang op het bedrijventerrein. Andere mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen die in het vooronderzoek naar voren zijn gekomen hebben betrekking op de ligging van het bedrijventerrein, het soort bedrijven (branches) die er aanwezig zijn, het verloop, onroerend goed, meldingen bij de politie en Meld Misdaad Anoniem, meldingen bij de gemeente en overige informatie die bij de gemeente bekend is bijvoorbeeld over incidenten uit het verleden.

2. Welke databronnen zijn beschikbaar en kunnen worden gekoppeld om de voorspellende waarde van potentiële indicatoren te onderzoeken?

Kenmerken van bedrijventerreinen (zoals de aanwezigheid van parkmanagement en de ontsluiting van het terrein) waren beschikbaar in het Integraal Bedrijventerreinen Informatie Systeem (IBIS). Daarnaast hebben we het aantal meldingen bij de politie en Meld Misdaad Anoniem ontvangen. Of een bedrijventerrein een Keurmerk Veilig Ondernemen heeft was online in te zien. De gemeente Tilburg heeft het aantal meldingen bij het Centraal Meldpunt (bijv. over drugsoverlast), het aantal hennepuimingen, het aantal Damocles sluitingen³ en data met betrekking tot de onroerende zaken op de bedrijventerreinen geleverd. Daarnaast zijn data uit de BAG en het bedrijvenregister (waaraan de KvK ten grondslag ligt) ter beschikking gesteld. Tot slot heeft netbeheerder Enexis data geleverd over de aanwezigheid en het weigeren van slimme meters, de capaciteit van elektriciteitsaansluitingen en historische gevallen van fraude. Resultaat was een uitgebreide dataset met mogelijke indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen.

3. Hoe kan de mate van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein worden geoperationaliseerd als afhankelijke variabele die kan worden gebruikt om de voorspellende waarde van potentiële indicatoren te toetsen?

Verschillende manieren om per bedrijventerrein een inschatting van de kans op georganiseerde criminaliteit en ondermijning te verkrijgen zijn tijdens het onderzoek in overweging genomen. Uiteindelijk is ervoor gekozen om de bedrijventerreinen te 'labelen' (dus in te delen in lage, gemiddelde, of hoge kans) door een kleine vragenlijst uit te zetten en respondenten (werkzaam bij bijvoorbeeld de gemeente, politie of parkmanagement) te vragen om een inschatting te maken of de kans op (georganiseerde) criminaliteit en ondermijning op de bedrijventerreinen laag, gemiddeld of hoog is. Vervolgens zijn de antwoorden van de

³ Met het aantal Damocles sluitingen refereren we naar het aantal panden dat gesloten is op basis van de Wet Damocles. Artikel 13b van de Opiumwet (ook wel de Wet Damocles genoemd) biedt burgemeesters de mogelijkheid om bestuursdwang toe te passen en panden te sluiten als daar sprake is van drugshandel.



respondenten gecombineerd tot een eindscore per bedrijventerrein. Hieruit kwam dat acht bedrijventerreinen een lage kans, vijftien terreinen een gemiddelde kans en zeven terreinen een hoge kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning hebben. Beperkingen van deze aanpak zijn dat er maar elf respondenten waren en dat er mogelijk sprake is van subjectiviteit, een gebrek aan kennis en/of eigen belangen, iets dat we hebben proberen te verminderen door zoveel mogelijk respondenten te benaderen van zoveel mogelijk verschillende organisaties en met verschillende rollen en expertisegebieden. Ook zijn de resultaten vergeleken met signalen die zijn voortgekomen uit integrale overheidssamenwerking.

4. Welke potentiële indicatoren hebben daadwerkelijk voorspellende waarde voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?

Om de belangrijkste indicatoren te vinden hebben we gebruik gemaakt van de data science methode Random Forest (RF). In de ranglijst van de 15 meest belangrijke indicatoren staan met name de 'usual suspects' die ook in het vooronderzoek als belangrijk naar voren kwamen. Zo lijkt een combinatie van factoren die betrekking hebben op de fysieke omgeving zoals veroudering en leegstand, de waarde van het onroerend goed, de mate van toezicht (door panden met een publieksfunctie of de aanwezigheid van veel uitzendkrachten) inclusief het aantal meldingen bij de politie en het verloop op een bedrijventerrein belangrijk te zijn. Ook het soort bedrijven lijkt een rol te spelen, bijvoorbeeld de rechtsvorm en grootte. Tot slot kwam de aanwezigheid van veel extra vestigingen, zoals extra KvK inschrijvingen, beheer-bv's en opslagloodsen, naar voren als risico indicator.

Om zicht te krijgen op de richting van de invloed van een bepaalde variabele hebben we aanvullend op de RF beslissingsbomen gedraaid op de top 15 risicofactoren. Deze informatie is namelijk niet direct uit een RF te halen. De beslissingsbomen geven ons tevens meer inzicht in de belangrijkste combinaties van risicofactoren. Deze "giftige cocktail"⁴ op een bedrijventerrein is een mix uit de volgende indicatoren:

1. *Gemiddeld bouwjaar*: hoe lager het gemiddelde bouwjaar op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Dit betekent dat er op bedrijventerreinen met gemiddeld gezien oudere panden een hoger risico is.
2. *Hoog-capaciteit aansluitingen*: hierbij kijken we naar het percentage van de panden op het bedrijventerrein waarbij de capaciteit van de energie-aansluiting groter is dan 35 groepen (3x50, 3x63, 3x80). Het lijkt erop dat als er minder hoog-capaciteit aansluitingen op een bedrijventerrein zijn, de kans op georganiseerde criminaliteit en ondermijning toeneemt. Mogelijk bestaat er een relatie tussen de capaciteit van de elektriciteitsaansluiting en de grootte van een bedrijf of het soort bedrijf.
3. *Aandeel panden laag WOZ*: hoe hoger het percentage van de panden op een bedrijventerrein met een WOZ waarde lager dan €160k, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
4. *Grootteklasse (kleine bedrijven)*: hoe hoger het percentage bedrijven met minder dan vijf werknemers, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
5. *Extra vestiging ja*: hoe meer extra vestigingen aanwezig zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Extra vestigingen zijn vestigingen die niet in het bedrijvenregister worden opgenomen, omdat ze niet

⁴ Met de giftige cocktail wordt de combinatie van indicatoren bedoeld die gezamenlijk lijden tot een hogere kans op georganiseerde criminaliteit en ondermijning.



voldoen aan de voorwaarden (er is niet minstens 1 persoon werkzaam, bijv. extra KvK inschrijvingen, beheer-bv's en opslagloodsen).

6. *Politie drugs en ondermijning*: hoe meer incidenten gerelateerd aan drugs en ondermijning er zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.

Hoewel de resultaten van deze verkennende pilot minder betrouwbaar zijn door de beperkte steekproef van 30 bedrijventerreinen in de gemeente Tilburg, laten we hiermee wel zien dat het technisch gezien mogelijk is om dergelijke analyses uit te voeren en kunnen we voorzichtig afleiden uit de resultaten dat het toepassen van data science methoden kan helpen bij het vinden van indicatoren voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Vervolgonderzoek op grotere schaal is nodig om dit te bevestigen.

5. Hoe kunnen de indicatoren die daadwerkelijk voorspellende waarde hebben worden gebruikt om in de praktijk inzichtelijk te maken welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning?

Inzichten over belangrijke voorspellers zouden gebruikt kunnen worden in toegepaste analyses over welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning. Hiervoor is het echter wel noodzakelijk om eerst op grotere schaal, dus voor meer bedrijventerreinen, onderzoek te verrichten, zodat een betrouwbaar model geschat kan worden. Vervolgens zou ook voor andere bedrijventerreinen, waarvoor nog geen label bekend is, op basis van de indicatoren voorspeld kunnen worden of het terrein een lage, gemiddelde of hoge kans heeft op (georganiseerde) criminaliteit en ondermijning. Daarnaast zou het mogelijk zijn om te kijken hoe veranderingen in de voorspellers de kans op (georganiseerde) criminaliteit en ondermijning veranderen. Hierdoor is het ook mogelijk om verschillende scenario's te draaien, wat kan helpen bij het maken van beleid.

De meerwaarde van data science methoden kan in de toekomst nog meer bewezen worden door analyses uit te voeren op grotere datasets. Echter zullen er ook beperkingen zijn in het gebruik van data science methoden, die met name betrekking hebben op de beschikbaarheid en verkrijgbaarheid van data en het koppelen van datasets (die bijv. verschillende aggregatieniveaus hebben). Daarnaast is het creëren van een geïntegreerde dataset met vele mogelijke indicatoren niet alleen een tijdrovende klus, ook komen er vele juridische aspecten (zoals privacy overweging) bij kijken.

Naast het uitvoeren van onderzoek op grotere schaal en het ontwikkelen van een betrouwbaar model, is een andere mogelijkheid om onderzoeken te doen op kleinere schaal. Gemeentes en regio's die een inventarisatie van hun bedrijventerreinen wensen kunnen ook, net als in dit onderzoek, een aantal experts vragen om de bedrijventerreinen te beoordelen in een vragenlijst en vervolgens zou dit aangevuld kunnen worden met kwalitatief onderzoek om inzichtelijk te maken welke bedrijventerreinen in de gevarenzone zitten.

6. In hoeverre is het haalbaar om in de toekomst dezelfde methodiek als in dit onderzoek toe te passen om indicatoren van (georganiseerde) criminaliteit en/of ondermijning in woonwijken te vinden?

In de toekomst zou dezelfde methodiek als in dit onderzoek ook toegepast kunnen worden om breder te kijken naar indicatoren van (georganiseerde) criminaliteit en ondermijning, bijvoorbeeld in woonwijken. Dit is niet specifiek onderzocht, maar zou wel mogelijk moeten



zijn met enkele aanpassingen, voornamelijk in de gebruikte databronnen. Daarnaast zullen opnieuw verschillende methoden getest moeten worden om te bepalen welke methode het meest geschikt is voor deze nieuwe toepassing. Het zou tevens interessant zijn om de relatie tussen bedrijventerreinen en de omliggende wijken te bekijken. Bedrijventerreinen die dicht bij bewoonde gebieden liggen of daaraan grenzen, hebben een relatie met die woonwijken. Zo vonden Tops en Van der Torre (2015; p. 41)⁵ in een van de door hen onderzochte cases dat *"...op (nabijgelegen) bedrijventerreinen frequent drugsvangsten worden gedaan waar geregeld op de één of andere manier wijkbewoners bij zijn betrokken"*. Volgens hen wordt crimineel geld deels op grote afstand, in andere delen van de wereld, geïnvesteerd, maar investeren criminelen lokaal nog steeds in bedrijven en panden. Door het lectoraat Digitalisering en Veiligheid en het lectoraat Ondermijning van Avans hogeschool wordt op dit moment een onderzoek uitgevoerd naar het in beeld brengen van risicowijken met behulp van data-analyse in de gemeente Tilburg. Het zou interessant zijn om de resultaten van het onderzoek van Avans naast de resultaten van dit onderzoek te leggen om te kijken of de bedrijventerreinen met een hoge kans op criminaliteit en ondermijning ook omringd worden door risicowijken.

⁵ Tops, P. & E. van der Torre (2015). Wijkenaanpak en ondermijnende criminaliteit. Den Haag: Boom Lemma.



Management summary

The objective of this study was to determine whether indicators could be found that illustrate at what stage an industrial area is in the development or presence of (organized) crime and subversion. Put differently: which indicators can predict where an industrial area finds itself on a 'sliding scale' of crime and subversion?

Research reason and context

The integrated approach to organized crime has evolved over the past 20 years. In order to arrive at safe and livable neighborhoods, vital industrial areas and sound business sectors, a broad approach is needed in which criminal law, disruption, enforcement and social policy are combined. This integrated approach requires closer cooperation at a local, regional and national level between the authorities involved, their private and social partners and citizens.

Preventive measures play an increasingly prominent role in the integrated approach to organized crime and subversion, with a great deal of attention being for tackling areas. An area approach and the approach to subversive crime are two sides of the same coin (Tops and Van der Torre, 2015). The accumulation of problems in vulnerable neighborhoods and industrial areas creates a breeding ground for criminal activities. There is a continuum with on the one hand the legitimate society and on the other hand the hard criminal underworld; in between there are many other forms ranging from consciously looking away and getting one's share to lending a hand to criminals.

Definitions and framing

In this study, an industrial area is defined as a work location of at least 1 ha gross that is suitable for use by trade, the manufacturing sector and industry. Areas that fall outside this definition include seaport areas, economic zones and office locations. For the identification of potentially relevant indicators and the data collection, we focused on drug-related crime (production of and trade in hemp and synthetic drugs). In addition, in this pilot study, the data collection was geographically limited to 30 industrial areas in the municipality of Tilburg. This means that, although the approach and analyses could be tested, the results of the analyses are less reliable due to the small sample of 30 industrial areas only.

Research approach

The research started with the identification of potential indicators of (organized) crime and subversion in industrial areas. This was done in a preliminary study consisting of a literature study, an expert meeting and interviews. Subsequently, data relating to the possible indicators were collected from various sources. In order to identify the degree of organized crime and subversion for each industrial area, a small questionnaire was set out in which respondents could assess the industrial areas. Data on the potential indicators and the degree of (organized) crime and subversion in the industrial area were aggregated and then merged if necessary, after which we could analyze which potential indicators actually have predictive value for (organized) crime and subversion in industrial areas. We used the data science method Random Forest to find the most important indicators, after which we ran several decision trees based on the 15 most important indicators to better understand the results. Below, we briefly explain the answers to the six research questions.



1. What are potential indicators of (organized) crime and subversion in industrial areas?

Firstly, in the preliminary study possible indicators have been identified with regard to the physical environment (including maintenance, ageing and empty buildings) and supervision, enforcement and cohesion in the industrial area. Other possible indicators of organized crime and subversion in industrial areas that came forward in the preliminary investigation are related to the location of the industrial area, the type of companies (sectors) that are present, the development, real estate, reports to the police and Meld Misdaad Anoniem⁶, reports to the municipality and other information known to the municipality, for example about incidents in the past.

2. Which data sources are available and can be merged to examine the predictive value of potential indicators?

Characteristics of industrial areas (such as the presence of park management and the access to the area) were available in the Integraal Bedrijventerreinen Informatie Systeem (IBIS).⁷ In addition, we received the number of reports to the police and Meld Misdaad Anoniem. Whether an industrial area has a Keurmerk Veilig Ondernemen⁸ could be found online. The municipality of Tilburg has delivered the number of reports to the Centraal Meldpunt⁹ (e.g. about drug nuisance), the number of cleared buildings where hemp was being produced, the number of Damocles closures¹⁰ and data related to the real estate in the industrial areas. In addition, data from the BAG¹¹ and the company register¹² (based on the Chamber of Commerce) were made available. Finally, grid operator Enexis provided data on the presence and refusal of smart meters, the capacity of electricity connections and historical cases of fraud. The result was a comprehensive dataset with possible indicators of (organized) crime and subversion in industrial areas.

3. How can the extent of (organized) crime and subversion on an industrial area be operationalized as a dependent variable that can be used to test the predictive value of potential indicators?

Different ways of obtaining an estimate of the likelihood of organized crime and subversion per industrial area were taken into consideration during the investigation. Ultimately, the choice was made to 'label' the industrial areas (i.e. to classify them in low, medium or high likelihood) by setting up a small questionnaire and asking respondents (working for example at the municipality, the police or park management) for an estimate whether the likelihood of (organized) crime and subversion in the industrial areas is low, medium or high. Subsequently, the answers of the respondents were combined into a final score for each industrial area. This resulted in eight industrial areas having a low likelihood, fifteen areas with an average likelihood and seven areas with a high likelihood of (organized) crime and subversion. This approach also had limitations: there were only eleven respondents and there may be

⁶ With Meld Misdaad Anoniem, people can report crime anonymously.

⁷ IBIS is a national database with information about industrial areas, see <https://www.ibis-bedrijventerreinen.nl/>

⁸ The 'Keurmerk Veilig Ondernemen' is a quality mark for safe entrepreneurship in industrial areas issued when certain measures are taken to prevent crime and make the area safe.

⁹ Central Reporting Center, where citizens can report crime and nuisance to the municipality.

¹⁰ Damocles closures refer to the buildings that were closed based on the Damocles Law. According to article 13b of the Opium Law (also called the Damocles Law) mayors are authorized to close buildings if drugs are being produced there.

¹¹ The BAG is the registration of addresses and buildings in the Netherlands.

¹² The municipality of Tilburg has a register of the companies that are present in the industrial areas. This register is based on the Chamber of Commerce, but verified, adapted and expanded based on contact they have with the companies.



subjectivity, a lack of knowledge and/or self-interest, something that we have tried to reduce by approaching as many respondents as possible from as many different organizations as possible, and with different roles and areas of expertise. The results were also compared with signals that resulted from integral government cooperation.

4. Which potential indicators actually have predictive value for (organized) crime and subversion in industrial areas?

To find the most important indicators we used the data science method Random Forest (RF). In the ranking of the 15 most important indicators we find the 'usual suspects' that also came forward in the preliminary investigation. For example, a combination of factors relating to the physical environment such as ageing and empty buildings, the value of the real estate, the degree of supervision (caused by buildings with a public function or the presence of many temporary employees), including the number of reports to the police and the development of an industrial area is important. The type of companies also seems to play a role, for example the legal form and size. Finally, the presence of many additional branches in the company register of the municipality, such as additional Chamber of Commerce registrations, came forward as a risk indicator.

In order to gain insights into the direction of the effect of a certain variable, we have also run decision trees with the top 15 risk factors in addition to the RF. This information cannot be retrieved directly from a RF. The decision trees also give us more insight into the most important combinations of risk factors. This "toxic cocktail"¹³ on an industrial area is a mix of the following indicators:

1. *Average year of construction*: the lower the average construction year in an industrial area, the higher the likelihood of organized crime and subversion. This means that there is a higher risk in industrial areas with, on average, older buildings.
2. *High-capacity connections*: here we look at the percentage of buildings in the industrial area where the capacity of the energy connection is more than 35 groups (3x50, 3x63, 3x80). It seems that if there are fewer high-capacity connections in an industrial area, the likelihood of organized crime and subversion increases. There may be a relationship between the capacity of the electricity connection and the size of a company or the type of company.
3. *Share of buildings with a low WOZ value*¹⁴: the higher the percentage of buildings in an industrial area with a WOZ value lower than € 160k, the higher the likelihood of organized crime and subversion.
4. *Company size (small companies)*: the higher the percentage of companies with fewer than five employees, the higher the likelihood of organized crime and subversion.
5. *Additional branches yes*: the more additional branches are present in an industrial area, the higher the likelihood of organized crime and subversion. Additional branches are branches that are not included in the company register because they do not meet the conditions (there is not at least 1 person working, e.g. additional Chamber of Commerce registrations and storage sheds).
6. *Police reports with reference to drugs and subversion*: the more incidents related to drugs and subversion in an industrial area, the higher the likelihood of organized crime and subversion.

¹³ With the toxic cocktail we refer to the combination of indicators that results into a higher likelihood of (organized) crime and subversion in the industrial area.

¹⁴ Value of the building under the Valuation of Immovable Property Act (WOZ).



Although the results of this exploratory study are less reliable due to the limited sample of 30 industrial areas in the municipality of Tilburg, we do show that it is technically possible to carry out such analyses. From these results we can imply that the application of data science methods can help to find indicators for (organized) crime and subversion in industrial areas. Follow-up research on a larger scale is needed to confirm this.

5. How can the indicators that actually have predictive value be used to provide insights in practice about which industrial areas have an increased likelihood of (organized) crime and subversion?

Insights about important predictors could be used in applied analyses on which industrial areas have an increased likelihood of (organized) crime and subversion. However, it is necessary to carry out research on a larger scale first, i.e. for more industrial areas, in order to estimate a reliable model. Subsequently, based on the indicators, it could be predicted for other industrial areas for which no label is known yet whether the area has a low, average or high likelihood of (organized) crime and subversion. In addition, it would be possible to see how changes in the predictors change the likelihood of (organized) crime and subversion. This makes it possible to run different scenarios, which can help in policy making.

The added value of data science methods can be proven even more in the future by carrying out analyses on larger datasets. However, there will also be limitations in the use of data science methods, which in particular relate to the existence and availability of data and the merging of data sets (which, for example, have different aggregation levels). Moreover, creating an integrated dataset with many possible indicators is not only a time-consuming task, but also involves many legal aspects (such as privacy considerations).

Besides conducting research on a larger scale and developing a reliable model, another option is to do research on a smaller scale. Municipalities and regions that want an assessment of their industrial areas can also, like in this study, ask a number of experts to assess the industrial areas using a questionnaire and then this could be supplemented with qualitative research to clarify which industrial areas are in the danger zone.

6. To what extent is it feasible to apply the same methodology as in this study in the future to find indicators of (organized) crime and/or subversion in residential areas?

In the future, the same methodology could also be applied to look more broadly at indicators of (organized) crime and subversion, for example in residential areas. This has not been investigated specifically, but should be possible with some modifications, mainly in the data sources used. Additionally, different methods will have to be tested again to determine which method is most suitable for this new application. It would also be interesting to look at the relationship between industrial areas and the surrounding residential areas. Industrial areas that are close to residential areas have a relationship with these areas. For example, Tops and Van der Torre (2015, p.41) found in one of the cases that they investigated that *"... on (nearby) industrial areas drug seizures are frequently being carried out in which residents are involved in one way or another."* According to them, criminal money is invested partly at a great distance, in other parts of the world, but criminals are still investing locally in companies and premises. The Digitalization and Safety research group and the Subversion research group of Avans University of Applied Sciences are currently conducting research into mapping out risk areas in the municipality of Tilburg using data analysis. It would be interesting to compare the results of Avans' research to the results of this research to see whether the industrial areas



with a high likelihood of (organized) crime and subversion are also surrounded by residential areas with a higher risk.



Inhoudsopgave

Managementsamenvatting	3
Management summary	8
Inhoudsopgave.....	13
1 Inleiding	15
1.1 Aanleiding onderzoek en context.....	15
1.2 Onderzoeksvragen	15
1.3 Definitie bedrijventerrein en verschillende typen	16
1.4 Definitie georganiseerde criminaliteit en ondermijning	18
1.5 Inkadering	21
1.6 Onderzoeksmethodiek	23
1.7 Opbouw rapport.....	27
2 Theoretisch kader	28
2.1 Inleiding	28
2.2 Ligging	30
2.3 Soort bedrijven (branches)	30
2.4 Fysieke omgeving en verloop	31
2.5 Onroerend goed.....	32
2.6 Toezicht, handhaving en samenhang	33
2.7 Aantal meldingen bij politie en Meld Misdaad Anoniem.....	34
2.8 Meldingen bij de gemeente en informatie uit het verleden.....	34
3 Conceptueel model en dataverzameling	36
4 Analyses en resultaten.....	40
4.1 Labelen van de bedrijventerreinen	40
4.2 Datapreparatie	42
4.3 Modelselectie.....	42
4.3.1 Modeltypen	42
4.3.2 Prestatiestatistieken	44
4.3.3 Modelselectie.....	45
4.4 Resultaten	47
5 Slotbeschouwing en mogelijkheden voor toekomstig onderzoek	52
5.1 Slotbeschouwing.....	52
5.2 Mogelijkheden voor toekomstig onderzoek	55
Literatuurlijst	57
A Experts vooronderzoek	59
B Lijst met variabelen.....	60
C Vragenlijst labelen bedrijventerreinen.....	64
C.1 Resultaten vragenlijst labels bedrijventerreinen	65



D	Datapreparatie	67
E	Databewerking	69
E.1	Normalisatie.....	69
E.2	Data balancing/Oversampling	69
E.3	Correlaties	70
F	Modelselectie en -performance	71
F.1	Prestatiestatistieken	71
F.2	Testen	72



1 Inleiding

In dit hoofdstuk wordt de aanleiding voor dit onderzoek toegelicht en wordt de context geschetst (§1.1). Vervolgens worden de onderzoeksvragen besproken (§1.2) en gaan we in op de definitie van bedrijventerreinen en verschillende typen terreinen (§1.3). Daarna geven we een definitie van georganiseerde criminaliteit en ondermijning, gaan we in op verschillende vormen van criminaliteit en lichten we de focus op drugsgerelateerde criminaliteit toe (§1.4). Ook zal de geografische focus (§1.5) en de onderzoeksmethodiek toegelicht worden (§1.6). Tot slot wordt de opbouw van dit rapport vermeld (§1.7).

1.1 Aanleiding onderzoek en context

De afdeling Extern Wetenschappelijke Betrekkingen van het WODC heeft opdracht gegeven om onderzoek te doen naar indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen met als doel te bepalen of er indicatoren kunnen worden gevonden waarmee het voor bestuurders inzichtelijk wordt in welk stadium een bedrijventerrein verkeert in de ontwikkeling of aanwezigheid van criminaliteit en ondermijning. Aan de hand van welke indicatoren kan inzichtelijk worden gemaakt waar een bedrijventerrein zich op de 'glijdende schaal' van georganiseerde criminaliteit en ondermijning¹⁵ bevindt?

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 en landelijk niveau intensivering van de samenwerking tussen betrokken overheden, hun private en maatschappelijke partners en burgers.

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

1.2 Onderzoeksvragen

Dit onderzoek richt zich in eerste instantie specifiek op bedrijventerreinen¹⁶, omdat bekend is dat er bedrijventerreinen zijn waar in meer of mindere mate sprake is van georganiseerde criminaliteit en ondermijning. Tot op heden is er geen instrument of methodiek beschikbaar waaruit blijkt waar deze gebieden zich op de 'glijdende schaal' bevinden. Het doel van deze studie is dan ook te bepalen of er indicatoren kunnen worden ontwikkeld, waardoor het voor

¹⁵ De definitie van georganiseerde criminaliteit en ondermijning zal toegelicht worden in paragraaf 1.4.

¹⁶ De definitie van bedrijventerreinen zal toegelicht worden in paragraaf 1.3.



bestuurders inzichtelijk wordt dat op een bedrijventerrein sprake kan zijn van de ontwikkeling of aanwezigheid van beginnende criminaliteit, criminele netwerken en georganiseerde en ondermijnende criminaliteit. Het gaat hierbij niet om zicht te krijgen op het soort criminaliteit, maar in welke mate een gebied afglijdt naar een situatie van verwevenheid met criminele banden. Om dit zowel conceptueel als kwantitatief te onderzoeken heeft de opdrachtgever de volgende concrete deelvragen geformuleerd:

- Wat zijn potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?
- Welke databronnen zijn beschikbaar en kunnen worden gekoppeld om de voorspellende waarde van potentiële indicatoren te onderzoeken?
- Hoe kan de mate van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein worden geoperationaliseerd als afhankelijke variabele die kan worden gebruikt om de voorspellende waarde van potentiële indicatoren te toetsen?
- Welke potentiële indicatoren hebben daadwerkelijk voorspellende waarde voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?
- Hoe kunnen de indicatoren die daadwerkelijk voorspellende waarde hebben worden gebruikt om in de praktijk inzichtelijk te maken welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning?
- In hoeverre is het haalbaar om in de toekomst dezelfde methodiek als in dit onderzoek toe te passen om indicatoren van (georganiseerde) criminaliteit en/of ondermijning in woonwijken te vinden?

1.3 Definitie bedrijventerrein en verschillende typen

Een eerste belangrijke stap in het onderzoek naar indicatoren van criminaliteit op bedrijventerreinen is het definiëren van 'een bedrijventerrein'. Hiervoor maken we gebruik van het Integraal Bedrijventerreinen Informatie Systeem (IBIS)¹⁷: een landelijke database die gevuld wordt door provincies en gemeenten. In het IBIS wordt er gesproken van zogenaamde 'werklocaties'. Een werklocatie is een terrein van minimaal 1 hectare (ha) bruto, dat vanwege zijn bestemming bestemd en geschikt is voor gebruik door handel, nijverheid, commerciële en niet-commerciële dienstverlening en industrie. Onder deze definitie vallen tevens (delen van) werklocaties die gedeeltelijk bestemd en geschikt zijn voor kantoren.

Terreinen die niet onder de definitie van werklocaties vallen zijn terreinen voor grondstoffenwinning, olie en gaswinning, voor waterwinning, voor agrarische doeleinden zonder verwerkende industrie en logistiek of voor afvalstort. Daarnaast worden kantorenlocaties niet in de IBIS inventarisatie opgenomen.¹⁸ Kantorenlocaties zijn terreinen waarvan het merendeel van de terreinoppervlakte bestemd en geschikt is voor kantoren.

Werklocaties worden door IBIS ingedeeld in de volgende categorieën (zie Figuur 1):

- Bedrijventerrein: een werklocatie van minimaal 1 ha bruto bestemd en geschikt voor gebruik door handel, nijverheid en industrie. Op deze terreinen kan ook enige commerciële en niet-commerciële dienstverlening (zoals kantoorgebouwen, detailhandel) aanwezig zijn, maar deze hebben samen een minderheidsaandeel in de terreinoppervlakte. De volgende terreinen vallen er niet onder: een zeehaventerrein, een economische zone, een kantorenlocatie, een terrein voor

¹⁷ IBIS Bedrijventerreinen Database (2018), zie: <https://www.ibis-bedrijventerreinen.nl/locations/list>.

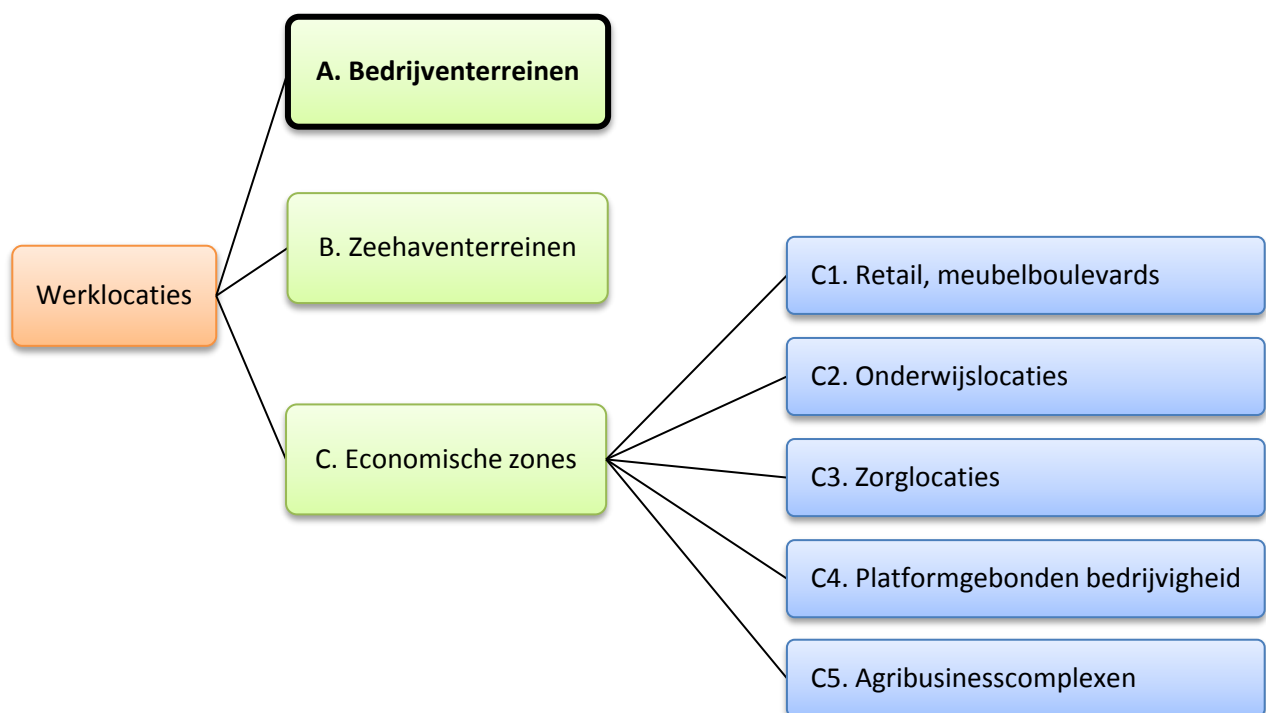
¹⁸ Kantoorlocaties zijn geregistreerd in het BAK-Kantorenbestand.



grondstoffenwinning, een terrein voor olie- en gaswinning, een terrein voor waterwinning, een terrein voor agrarische doeleinden, een terrein voor afvalstort.

- Zeehaventerrein: een werklocatie van minimaal 1 ha bruto waarvan een deel van de kavels een laad- en/of loskade heeft en langs diep vaarwater ligt dat toegankelijk is voor grote zeeschepen.
- Economische zone: een werklocatie van minimaal 1 ha bruto bestemd en geschikt voor commerciële en niet-commerciële dienstverlening. Op deze terreinen kan ook enige handel, nijverheid en industrie aanwezig zijn, maar deze hebben samen een duidelijk minderheidsaandeel in de terreinoppervlakte. Onder economische zones vallen terreinen met verschillende (combinaties van) functies, zie Figuur 1.

In ons onderzoek gebruiken we de IBIS database als startpunt en beperken we ons tot de werklocaties van het type 'bedrijventerrein'.



Figuur 1. Indeling typen bedrijventerreinen (IBIS, 2017)

Samenvatting 'Definitie bedrijventerrein'

We beperken ons in dit onderzoek tot de werklocaties met het type 'bedrijventerrein'. Hiermee laten we de zeehaventerreinen en economische zones (zoals meubelboulevards, zorglocaties etc.) buiten beschouwing. Ook kantorenlocaties (deze zijn niet in de IBIS database opgenomen) en buitengebieden vallen buiten beschouwing in dit onderzoek.



1.4 Definitie georganiseerde 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 wat ondermijning is.

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".¹⁹ Deze hanteert de volgende definitie van georganiseerde criminaliteit: *'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 en 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

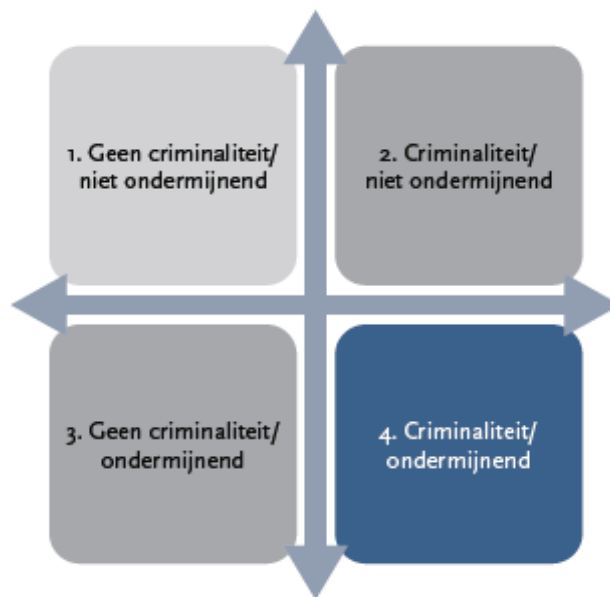
Tegelijkertijd bestaat er gevaar voor inflatie van het onderwerp. Kolthoff en Khonraad (2016) stellen vast dat het begrip ondermijning te pas en te onpas wordt gebruikt voor diverse criminele fenomenen, waarbij het steeds vaker als synoniem wordt gebruikt voor alle vormen van georganiseerde misdaad. 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 ondermijnend noemt maar die dat lang niet altijd zijn:

1. Gedrag dat niet crimineel en niet ondermijnend is
2. Gedrag dat wel crimineel is maar niet ondermijnend
3. Gedrag dat ondermijnend is maar niet crimineel
4. Gedrag dat ondermijnend en crimineel is

Figuur 2 vat dit beknopt samen.

¹⁹ Tweede Kamer, vergaderjaar 1995–1996, 24 072, nr. 16.



Figuur 2. Indeling van gedrag in het kader van ondermijning

In het kader van dit onderzoek zijn we voor wat betreft het element ondermijning 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"*.

Georganiseerde misdaad en ondermijnende criminaliteit kennen vele verschijningsvormen. In de aanbiedingsbrief bij het Nationaal Dreigingsbeeld georganiseerde criminaliteit 2017-2021, dat op 1 juni 2017 door de minister van Veiligheid en Justitie aan de Tweede Kamer werd aangeboden, werd het volgende overzicht gepresenteerd²⁰ (zie Tabel 1).

In het "Nationaal Dreigingsbeeld 2017 Georganiseerde criminaliteit" werd een 17-tal verschijningsvormen van criminaliteit van het predicaat 'dreiging' voorzien. Drugsgelateerde criminaliteit, uitbuiting/mensenhandel en vuurwapengerelateerde criminaliteit waren goed voor 9 dreigingspredicaten en vallen alle onder het kopje illegale markten. Dit zijn ook de misdrijven die centraal staan in de Monitor Georganiseerde Criminaliteit, waarbij drugsgelateerde criminaliteit de overhand heeft (Kruisbergen, Van de Bunt en Kleemans, 2012). Hetzelfde geldt voor de cijfers inzake geregistreeerde criminaliteit in de jaarlijkse uitgave Criminaliteit en Rechtshandhaving (Kalidien, 2017). Daarbij moet wel worden aangetekend dat het dreigingsbeeld ook van een groot aantal 'witte vlekken' spreekt.

²⁰ Brief minister van Veiligheid en Justitie van 1 juni 2017, kenmerk 2082923.



Illegale markten	
Handel in en smokkel van cocaïne	Dreiging
Handel in en smokkel van heroïne	Dreiging
Productie van, handel in en smokkel van synthetische drugs	Dreiging
Productie van, handel in en smokkel van cannabis	Dreiging
Seksuele uitbuiting	Dreiging
Arbeidsuitbuiting, criminele uitbuiting en gedwongen dienstverlening	Dreiging
Mensensmokkel	Dreiging
Orgaanhandel en mensenhandel met het oogmerk van orgaanverwijdering	Geen concrete dreiging
Illegale handel in en smokkel van vuurwapens en explosieven	Dreiging
Productie en verspreiding van kinderpornografie	Dreiging
Productie en verspreiding van vals geld	Geen concrete dreiging
Matchfixing	Geen concrete dreiging
Illegale kansspelen	Geen concrete dreiging
Fraude en witwassen	
Acquisitiefraude	Witte vlek
Hypotheekfraude	Geen concrete dreiging
Beleggingsfraude	Dreiging
Faillissementsfraude	Dreiging
Fraude met betaalmiddelen	Geen concrete dreiging
Fraude met online handel	Geen concrete dreiging
Telecomfraude	Witte vlek
Verzekeringsfraude	Dreiging
Voorschotfraude	Geen concrete dreiging
Accijnsfraude	Dreiging
Fiscale fraude	Dreiging
Witwassen	Dreiging
Vermogenscriminaliteit	
Woninginbraak	Dreiging
Bedrijfsinbraak	Geen concrete dreiging
Winkeldiefstal	Witte vlek
Kraken en geldautomaten	Geen concrete dreiging
Overvallen	Dreiging
Ladingdiefstal	Geen concrete dreiging
Afpersing	Witte vlek
Georganiseerde autocriminaliteit	Voorwaardelijke dreiging
Heling	Witte vlek

Tabel 1. Overzicht kwalificaties uit het Nationaal Dreigingsbeeld

Tot slot betreffen de dagelijkse nieuwsoverzichten van de Taskforce Zeeland-Brabant voor meer dan 80% drugsgerelateerde criminaliteit. In dit onderzoek zullen we bij het vaststellen van de indicatoren en het verzamelen van de data dus de focus leggen op drugsgerelateerde criminaliteit.

Samenvatting 'Georganiseerde criminaliteit en ondermijning'

Een groot gedeelte van de georganiseerde criminaliteit en ondermijning in Nederland betreft drugsgerelateerde criminaliteit. Daarom zal in dit onderzoek bij het vaststellen van de indicatoren en het verzamelen van de data de focus liggen op drugsgerelateerde criminaliteit (productie van en handel in hennep en synthetische drugs).



1.5 Inkadering

Om het onderzoek beheersbaar te houden is er gedurende het onderzoekstraject in overleg met het WODC en de begeleidingscommissie voor gekozen om de dataverzameling geografisch te beperken en een pilot te doen voor de gemeente Tilburg. Hier lagen een aantal beweegredenen aan ten grondslag, maar dit heeft ook een aantal belangrijke consequenties, die we hieronder zullen belichten.

Beweegredenen

Ten eerste bleek uit het vooronderzoek, o.a. uit de interviews, dat het koppelen van verschillende databronnen een struikelblok kan zijn. Zo kan het aggregatieniveau verschillen (bijv. op persoons-, adres-, bedrijfs-, object- of terreinniveau), maar kan ook de naam en identificatie van een bedrijventerrein anders zijn. Op de website www.locatiemonitor.nl²¹, waar een database met kenmerken over bedrijventerreinen te vinden is, wordt aangegeven dat de naam van een locatie gekozen is na raadpleging van diverse bronnen (IBIS, gemeentelijke-/regionale- en provinciale websites en locatie-overzichten van de Kamer van Koophandel), maar dat deze bronnen elkaar helaas veelvuldig tegenspreken op het punt van de naamgeving. Daarnaast komt het regelmatig voor dat de naamgeving zoals die voorkomt op bewegwijzering of plattegronden bij de terreinentree, afwijkt van die in de eerder genoemde bronnen. Dit zou betekenen dat het koppelen van informatie uit verschillende bronnen met informatie over bedrijventerreinen behoorlijk veel handwerk vergt.

Ten tweede wordt de beschikbare hoeveelheid data groter door te focussen op een kleiner regionaal gebied. Zo zitten er in het IBIS van de Provincie Noord-Brabant veel meer kenmerken dan in de landelijke database. Daarnaast werken veel organisaties regionaal. Het was niet haalbaar om voor heel Nederland contact op te nemen met alle gemeenten, netbeheerders van energie etc. Door te focussen op één gemeente, kan informatie van deze bronnen wel meegenomen worden.

Ten slotte was de verwachting dat het verkrijgen van data van externe partijen niet gemakkelijk en erg tijdrovend zou zijn. In interviews werd al aangegeven dat er vaak met een gebruikersovereenkomst gewerkt moet worden om data van externe partijen te krijgen en dat bedrijven en instanties steeds terughoudender worden met het delen van data. Door de ingang van de Algemene verordening gegevensbescherming (AVG) is er extra aandacht gekomen voor privacy en data bescherming. Vanuit onze ervaring met andere projecten weten we ook dat bijvoorbeeld gemeenten steeds terughoudender zijn met het delen van data: vaak moet er eerst een onderzoeksplan ingediend worden en goedgekeurd worden door een privacy-expert etc. Door de focus te leggen op een beperkt geografisch gebied, wordt de hoeveelheid data en de hoeveelheid instanties waarmee we te maken krijgen kleiner, wat de haalbaarheid vergroot.

Keuze voor de gemeente Tilburg

Er waren verschillende redenen om de pilot in de gemeente Tilburg te doen. Ten eerste is de gemeente Tilburg onderdeel van de City Deal 'Zicht op Ondernijning'.²² In de City Deal 'Zicht op Ondernijning' is het de bedoeling dat er meer zicht ontstaat op patronen en structuren van

²¹ <http://www.locatiemonitor.nl/Locatie-identificatie.htm>

²² De City Deal is een samenwerking tussen de gemeenten Amsterdam, Den Haag, Rotterdam, Tilburg en Utrecht, het Centraal Bureau voor de Statistiek (CBS), de Nationale Politie, het OM en de ministeries van Binnenlandse Zaken en Koninkrijksrelaties, Veiligheid en Justitie en Financiën/Belastingdienst.



georganiseerde criminaliteit door beschikbare data te combineren en met behulp van moderne technieken deze informatie te analyseren.²³ De gemeente Tilburg is dus al actief bezig met het gebruik van data (analyse) op het gebied van ondermijning. Daarnaast doet de gemeente Tilburg vaak mee aan onderzoek en zijn er dus al contactpersonen bekend binnen de juiste afdeling. Een bijkomend voordeel is dat de onderzoekers bekend zijn met Tilburg en de bedrijventerreinen. Binnen de gemeente Tilburg is er een verscheidenheid aan bedrijventerreinen, zowel kleine binnenwijkse terreinen als grotere industrieterreinen aan de rand van de stad. Ten slotte wordt er door het lectoraat Digitalisering en Veiligheid en het lectoraat Ondermijning van Avans hogeschool op dit moment een onderzoek uitgevoerd naar het in beeld brengen van risicowijken met behulp van data-analyse in de gemeente Tilburg. De resultaten van dit onderzoek kunnen dus later naast de resultaten van het onderzoek door Avans worden gelegd om een compleet beeld te krijgen.

Consequenties

De keuze voor een pilot in de gemeente Tilburg heeft belangrijke consequenties. Ten eerste wordt de steekproef waarop de analyses worden gedaan sterk verkleind, tot 30 bedrijventerreinen.²⁴ Dit betekent dat, hoewel de aanpak en analyses getest kunnen worden, de resultaten van de analyses minder betrouwbaar zijn. Er kan slechts gekeken worden of de resultaten in dezelfde richting wijzen als bestaande hypothesen, maar er kan geen uitsluitend gegeven worden over het belang en de exacte impact van de indicatoren. Ten tweede zal er vervolgonderzoek nodig zijn om te bepalen of deze aanpak ook voor andere regio's werkt. Data die voor de gemeente Tilburg beschikbaar zijn, zijn niet per definitie ook voor andere regio's beschikbaar. Daarnaast zou er in andere regio's ook sprake kunnen zijn van andere indicatoren. De indicatoren kunnen namelijk ook afhankelijk zijn van de omgeving/context. Tilburg ligt bijvoorbeeld redelijk dicht bij de grens met België en er is veel platteland in de provincie Noord-Brabant. Het model zal dus aangepast en aangevuld moeten worden om het geschikt te maken voor andere regio's.

Samenvatting 'Inkadering'

In dit onderzoek is bij de dataverzameling en analyses ervoor gekozen om een pilot te doen in de gemeente Tilburg. Hier lagen meerdere beweegredenen aan ten grondslag, waaronder het feit dat zowel de dataverzameling als de databewerking (verschillende aggregatieniveaus, verschillende namen voor de terreinen) zeer veel tijd zouden kosten, waardoor het niet haalbaar was om binnen dit onderzoek op grotere schaal data te verzamelen. Een voordeel van onderzoek op kleinere schaal is dat er meer verschillende databronnen beschikbaar zijn (bv. van de gemeente). Het betekent echter wel dat, hoewel de aanpak en analyses getest kunnen worden, de resultaten van de analyses door de kleine steekproef van 30 bedrijventerreinen minder betrouwbaar zijn.

²³ Bron: <https://www.rijksoverheid.nl/actueel/nieuws/2017/06/28/ondermijnende-criminaliteit-beter-in-vizier-met-data-analyse>

²⁴ We nemen namelijk alleen werklocaties met het type 'bedrijventerrein' mee (zie §1.5) en laten bijv. economische zones zoals woonboulevards buiten beschouwing. In het Integraal Bedrijventerreinen Informatie Systeem (IBIS) waren 35 bedrijventerreinen in de gemeente Tilburg te vinden, waarvan er 5 nog in ontwikkeling zijn/nog niet gerealiseerd zijn. Uiteindelijk bestond de steekproef dus uit 30 bedrijventerreinen.



1.6 Onderzoeksmethodiek

Het onderzoek is gestart met een – cruciale – voorbereidende fase. Om te komen tot goede voorspellende indicatoren voor (georganiseerde) criminaliteit en ondermijning is het belangrijk om in eerste instantie te kijken naar bestaande ideeën uit ander onderzoek en uit het veld. Dit hebben we gedaan in een vooronderzoek waarin we door middel van een literatuuronderzoek, een expertmeeting en interviews hebben gekeken welke kennis en informatie al beschikbaar is over (mogelijke indicatoren van) georganiseerde criminaliteit en ondermijning op bedrijventerreinen.

Vooronderzoek

Tijdens het vooronderzoek heeft een expertmeeting plaatsgevonden om bestaande ideeën over potentiële indicatoren in beeld te krijgen. De samenstelling van het expert panel is in overleg met onze interne experts, het WODC en de begeleidingscommissie opgesteld om te komen tot een zo breed mogelijke vertegenwoordiging van kennisgebieden. Er waren twaalf deelnemers aan de expert meeting (de onderzoekers van Avans en CentERdata niet meetellend). Inzichten uit dit expert panel zijn vervolgens samen met onze interne experts geverifieerd en door een literatuurstudie uitgebreid. Ten slotte hebben er nog zes interviews en vele gesprekken plaatsgevonden om de inzichten uit te breiden en databronnen te identificeren die mogelijk voor dit onderzoek gebruikt zouden kunnen worden. In Appendix A is meer informatie te vinden over de bij het vooronderzoek betrokken experts.

Conceptueel model en dataverzameling

Op basis van het vooronderzoek is een conceptueel model opgesteld. Nadat was vastgesteld welke indicatoren nuttige input vormen voor de uit te voeren analyses, moest er gekeken worden voor welke variabelen er gegevens beschikbaar waren. Vervolgens is de dataverzameling gestart. In het IBIS), dat informatie over alle bedrijventerreinen in Nederland bevat, waren 35 bedrijventerreinen in de gemeente Tilburg te vinden. Nader onderzoek wees uit dat vijf van deze terreinen nog in ontwikkeling zijn/nog niet gerealiseerd zijn. Deze terreinen zijn niet meegenomen in dit onderzoek, waardoor de uiteindelijke steekproef bestond uit 30 bedrijventerreinen. Data over deze 30 bedrijventerreinen zijn zoveel mogelijk verzameld op bedrijventerreinniveau. Het doel van het onderzoek was immers om inzichten te verkrijgen over georganiseerde criminaliteit en ondermijning op bedrijventerreinen. Data over bedrijven en adressen zijn geaggregeerd per bedrijventerrein en data op persoonsniveau zijn niet meegenomen. De meest recente data zijn opgevraagd en vaak zijn data van meerdere jaren opgevraagd (met als richtlijn de jaren 2014 t/m 2018, indien beschikbaar), zodat deze later nog geaggregeerd kon worden als de data te summier bleek te zijn.

Naast data met betrekking tot de indicatoren, moest er ook bepaald worden welke bedrijventerreinen een lage, een gemiddelde of hoge kans hebben op de aanwezigheid van criminaliteit en ondermijning, zodat het model vervolgens kon bepalen welke indicatoren hier bepalend voor zijn. Deze 'labeling' van de bedrijventerreinen (in lage, gemiddelde, of hoge kans) heeft plaatsgevonden door aan meerdere mensen met verschillende achtergronden te vragen deze categorisatie te maken in een kleine vragenlijst en vervolgens alle resultaten te combineren tot een eindindeling. De vragenlijst is ingevuld door elf van de 21 uitgenodigde personen en de respondenten waren werkzaam bij de gemeente Tilburg (in verschillende rollen), de politie, de Veiligheidsregio Midden-Brabant en een parkmanagement organisatie. Dit wordt verder toegelicht in hoofdstuk 4.



Data koppeling

Na de dataverzameling moesten de verschillende bestanden gekoppeld worden op het niveau van bedrijventerrein. Waar dit niet mogelijk was is een alternatief gezocht, bijvoorbeeld door op basis van de postcodes te achterhalen op welk bedrijventerrein de data betrekking hadden.

Data analyse

Het doel van de analyses was om te zoeken naar verbanden tussen (georganiseerde) criminaliteit en ondermijning enerzijds en alle andere beschikbare variabelen anderzijds. In andere woorden, we willen de voorspellende waarde van potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein onderzoeken. Hierbij is er gebruik gemaakt van data science methoden, zie Box 1 en 2.



Box 1: Inleiding data science en voorbeelden van toepassingen voor datagedreven beleid²⁵

Inleiding Data Science

Data science is een relatief nieuwe discipline die leidt tot nieuwe innovaties. Met data science wordt bedoeld: het analyseren en interpreteren van grote hoeveelheden data die, vaak complex en ongestructureerd, tegenwoordig dagelijks gegenereerd worden. De kunst van data science is het transformeren van data in waardevolle acties. Dit vakgebied beslaat meer dan het conventionele gebruik van data en statistiek. Wat data science onderscheidt van statistiek is de holistische benadering. Er zijn steeds meer data 'in het wild' te vinden en data scientists verzamelen deze data, gieten de data in een bruikbare vorm en presenteren het verhaal dat de data vertellen. Conventioneel statistisch onderzoek is over het algemeen veel meer gericht op het verleden. Er worden hypothesen opgesteld en met behulp van data worden deze getoetst om aan te tonen dat bepaalde processen en verbanden in het verleden zo hebben plaatsgevonden. Data science daarentegen is er veel meer op gericht om ook voorspellingen te doen en op die manier aanbevelingen te doen voor belangrijke beslissingen die genomen moeten worden.

Voorbeelden van data science toepassingen voor datagedreven beleid

In Nederland is er al sprake van diverse data science toepassingen voor datagedreven beleid. De nationale politie is bezig met het Criminaliteit Anticipatie Systeem (CAS) waarin criminaliteit wordt voorspeld door een algoritme op basis van een groot aantal kenmerken. Na succesvolle pilots in Amsterdam en Den Haag wordt CAS nu uitgerold naar de eenheden in alle andere regio's. Ook op gemeenteniveau zijn diverse data science aanpakken gaande. De gemeente Alkmaar is gestart met een pilotstudie waarbij data verkregen met behulp van sensoren een rol kunnen spelen in de verbetering van de leefomgeving. De gemeente Utrecht is momenteel bezig met meerdere projecten, bijvoorbeeld een pilot in het kader van inbraakpreventie en een "Living Lab Air" waarin sensordata uit diverse bronnen gecombineerd worden. Daarnaast loopt in Amsterdam een project waarbij de intensiteit van potentiële overlast wordt voorspeld op basis van portofoon- en digitale bonddata, omgevingskenmerken en weerdata. Deze informatie kan vervolgens worden ingezet voor een efficiëntere handhaving. Dit zijn slechts enkele voorbeelden en er bestaan talloze andere voorbeelden van (big) data gebruik door Nederlandse gemeenten.

De toenemende digitalisering en datafication waardoor data steeds belangrijker worden in onze maatschappij leidt niet alleen tot veranderingen in de manier hoe (overheids-) instellingen werken en beslissingen genomen worden, maar ook tot een verandering van de aard van de criminaliteit, denk aan cyber crime, en een verschuiving van criminaliteit die waargenomen en geregistreerd wordt naar criminaliteit die (nog) niet gemeten en geregistreerd wordt. Juist het analyseren van zogenaamde big data door middel van data science technieken kan helpen om nieuwe vormen van criminaliteit te herkennen én het dark number van de criminaliteit (beter) op te sporen.

²⁵ Verwijzingen voor voorbeelden in box 1: voor CAS, zie www.politie.nl; voor pilot Alkmaar zie https://www.atlasleefomgeving.nl/documents/10156/5184970/Ronde+2+Internet+of+things+Sensorpilot-Alkmaar_2.pdf/bd2feb78-f0c2-4743-8a1f-4b2795aa8ec0; voor Utrecht zie <https://www.stadszaken.nl/smart/data/1004/misdaad-voorspellen-misdaad-met-algoritmes> en <https://www.uu.nl/nieuws/nieuw-innovatielab-leefomgeving-meet-luchtkwaliteit-op-utrecht-science-park> en voor Amsterdam zie www.bigdatagemeenten.nl. Op deze laatste pagina zijn ook nog andere voorbeelden van (big) data gebruik door Nederlandse gemeenten te vinden.



Box 2: Big data en Data Science

Big data en Data Science

We spreken van big data als er sprake is van meerdere kenmerken, met name de hoeveelheid data, de snelheid waarmee de data binnenkomen en/of opgevraagd worden, en de diversiteit van de data. De data is ongestructureerd en kan niet in een traditionele database opgeslagen worden. Andere mogelijke factoren zijn de variatie in de data (verschillende bronnen kunnen elkaar tegenspreken), de kwaliteit van de data (de ene bron is minder betrouwbaar dan de andere) en de complexiteit van de data (soms is het lastig om ongestructureerde data van verschillende bronnen met elkaar te combineren).

Een veel gebruikte bron voor big data is het zogenaamde 'web data scraping', waarbij informatie van het internet wordt gehaald. Ook andere databronnen die betrekking hebben op natuurlijke taal kunnen worden gebruikt voor data science, zoals open antwoorden, tekstbestanden, aantekeningen van klantcontacten, rapporten of e-mails. Internet-logbestanden en de metadata van zoekmachines kunnen eveneens interessante informatie bieden over trends in de tijd. Het minen, clusteren en analyseren van deze ongestructureerde gegevensbronnen vereist het gebruik van analytische technieken voor natuurlijke taal, bijvoorbeeld sentimentanalyse, die subjectieve informatie uit taal kan extraheren. Dit wordt ook Natural Language Processing (NLP) genoemd.

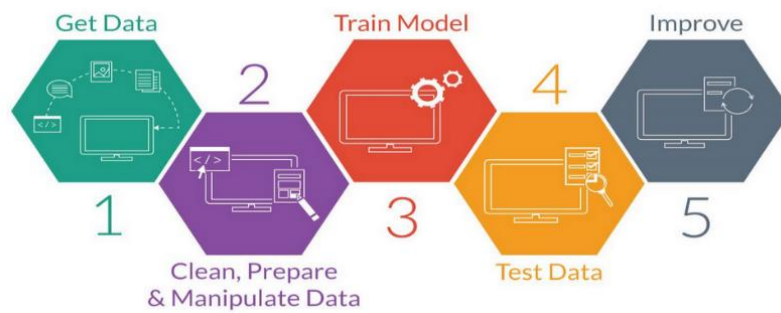
Data science methoden

Door machine learning classificatietechnieken en clustertechnieken toe te passen op (zeer) grote databestanden kunnen verbanden worden geïdentificeerd.²⁶ Machine learning is een geavanceerd onderzoeksveld binnen data science dat zich bezighoudt met technieken waarmee computers kunnen leren, zonder dat ze er speciaal voor geprogrammeerd zijn. Machine learning wordt over het algemeen onderverdeeld in twee categorieën: supervised en unsupervised machine learning.

Bij **unsupervised learning** methoden kan geen gebruik gemaakt worden van bestaande informatie, de zogenaamde 'labels', en is het doel om de onderliggende structuur of verdeling in de data te modelleren. Het wordt meestal gebruikt voor verkennende gegevensanalyse om verborgen patronen of clustering in gegevens te detecteren. De basis van deze technieken is het gebruik van het herkennen van patronen voor complexe en multispectrale data. Door software autonoom te maken of door iteratieve terugkoppeling te gebruiken om samenhang in de data te ontdekken, worden patronen gevonden en anomalieën ontdekt.

Een ander type data science methode is **supervised learning**, een set van geavanceerde algoritmen die wel gebruik maken van bekende informatie/'labels'. Deze labels kunnen categorieën zijn, waarmee het een classificatieprobleem is, of getallen, wat neerkomt op een regressieprobleem. Er is een doelvariabele en het algoritme leert om de output (onbekende labels of toekomstige data) te voorspellen op basis van de input data (en bestaande labels).

²⁶ Normaal gesproken worden data science technieken gebruikt bij grote datasets met veel observaties en variabelen. In dit onderzoek waren we door omstandigheden genooddaakt om een data science methode toe te passen op een kleine dataset met slechts 30 observaties, maar wel bijna 100 variabelen. We laten hiermee zien dat het technisch gezien mogelijk is om dergelijke analyses uit te voeren en we kunnen voorzichtig afleiden uit de resultaten dat het toepassen van data science methoden kan helpen bij het vinden van indicatoren voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Vervolgonderzoek op grotere schaal is nodig om dit te bevestigen.



Figuur 4. Kernstappen supervised machine learning

In dit onderzoek is er gebruik gemaakt van de analysetechniek Random Forest, een 'tree based supervised learning' techniek. Het idee van een Random Forest is dat een groot aantal beslissingsbomen gecombineerd wordt om tot een uiteindelijke voorspelling te komen.

Samenvatting 'Onderzoeksmethodiek'

In een vooronderzoek hebben we door middel van een literatuuronderzoek, een expertmeeting en interviews gekeken welke kennis en informatie al beschikbaar was. Op basis hiervan is een conceptueel model opgesteld. Vervolgens zijn data over de indicatoren verzameld, gekoppeld en geaggregeerd per bedrijventerrein. Daarnaast zijn de bedrijventerreinen gelabeld (ingedeeld in drie categorieën: een lage, gemiddelde, of hoge kans op georganiseerde criminaliteit en ondermijning) door aan meerdere mensen met verschillende achtergronden en vanuit verschillende organisaties te vragen deze categorisatie te maken in een kleine vragenlijst en vervolgens alle resultaten te combineren tot een eindindeling. Toen per bedrijventerrein de data over de indicatoren en het kans op georganiseerde criminaliteit en ondermijning vastgesteld was, zijn de data geanalyseerd door gebruik te maken van de supervised learning techniek Random Forest.

1.7 Opbouw rapport

Dit rapport is als volgt opgebouwd. In hoofdstuk 2 schetsen we het theoretische kader, waarna we het conceptuele model en de databronnen toelichten in hoofdstuk 3. In hoofdstuk 4 gaan we vervolgens in op de analyses en resultaten. Hoofdstuk 5 bevat de slotbeschouwing en mogelijkheden voor toekomstig onderzoek.



2 Theoretisch kader

In dit hoofdstuk wordt het theoretische kader geschetst. Eerst lichten we in de inleiding verschillende theorieën vanuit de criminologie toe (§2.1), waarna we mogelijke indicatoren van (georganiseerde) criminaliteit op bedrijventerreinen schetsen. Hierbij gaan we als eerst in op de ligging van het bedrijventerrein (§2.2). Vervolgens komen het soort bedrijven (branches) die er aanwezig zijn aan bod (§2.3). Daarna bespreken we de fysieke omgeving (waaronder onderhoud, veroudering en leegstand) en het verloop (§2.4) en de indicatoren met betrekking tot onroerend goed (§2.5). Dan weiden we uit over toezicht, handhaving en samenhang op het bedrijventerrein (§2.6). Tot slot gaan we in op meldingen bij de politie en Meld Misdaad Anoniem (§2.7) en meldingen bij de gemeente en overige informatie die bij de gemeente bekend is (§2.8).

2.1 Inleiding

In de criminologie is de volkswijsheid 'De gelegenheid maakt de dief' theoretisch verwoord door onder anderen Felson en Clarke (1998) in de gelegenheidstheorie. De gelegenheidstheorie is echter lastig te onderscheiden van de eerder ontwikkelde routine activity theory (Cohen en Felson, 1979). De theorie werd ontwikkeld om een verklaring te bieden voor de steeds stijgende criminaliteitscijfers na de Tweede Wereldoorlog. Cohen en Felson verklaarden deze toename door de enorme groei aan potentiële slachtoffers door de toegenomen welvaart, gepaard gaande met een relatieve afname van toezicht. De routine activity theory is een variant van de controletheorieën, die uitgaat van de vraag waarom personen 'geen' criminaliteit plegen. De routine activity theory stelt daarbij dat de aandacht gevestigd moet worden op de situatie in plaats van op de dader. In die zin vormt deze theorie de inspiratie voor de gelegenheidstheorie. Later ontwikkelden Felson en Clarke (1998) hun meer uitgesproken gelegenheidstheorie (opportunity theory), die op één enkel principe was gebaseerd: verleidelijke en gemakkelijke gelegenheden om criminaliteit te plegen zetten mensen aan om dit ook te doen. Volgens Felson en Clarke (1998) wordt de omvang van criminaliteit hierbij bepaald door drie factoren: aantallen potentiële daders, aantallen aantrekkelijke doelwitten, en de mate van toezicht op en de bescherming van deze doelwitten.

Felson en Clarke (1998) vatten hun gelegenheidstheorie samen in tien uitgangspunten:

1. Gelegenheid speelt een rol in het veroorzaken van alle soorten criminaliteit.
2. Gelegenheden voor criminaliteit zijn zeer specifiek.
3. Gelegenheden voor criminaliteit zijn gebonden aan tijd en ruimte.
4. Gelegenheden voor criminaliteit zijn afhankelijk van alledaagse activiteiten.
5. Een specifiek delict creëert gelegenheden voor andere delicten.
6. Specifieke goederen bieden meer verleidelijke gelegenheden voor criminaliteit.
7. Sociale en technologische ontwikkelingen bieden nieuwe mogelijkheden voor criminaliteit.
8. De gelegenheid voor het plegen van criminaliteit kan worden verminderd.
9. Het verminderen van de gelegenheid voor het plegen van criminaliteit leidt in het algemeen niet tot verplaatsingseffecten.
10. Gerichte gelegenhedsvermindering kan leiden tot een nog grotere criminaliteitsafname.



Onder de sociale en technologische ontwikkelingen zoals genoemd onder punt 7 dienen we ook uitdrukkelijk de voortschrijding van digitale technologie te verstaan. Een van de belangrijke ontwikkelingen als gevolg van een toename in digitalisering is de wens van gemeenten om datagestuurd te werken (Meijer, 2015). Digitalisering kan duiding geven aan 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. Maar er is ook een keerzijde. De bevindingen die we in een onderzoek naar integriteitschendingen in verband met georganiseerde criminaliteit onder rechtshandhavers aantreffen (Nelen en Kolthoff, 2017), lijken onverkort te gelden voor de criminelen zelf. Ze communiceren via digitale weg, kunnen hun activiteiten versluieren, manifesteren zich op sociale media en schermen hun activiteiten af met behulp van hoogwaardige technologie.

Een meer hoopgevend inzicht uit de gelegenheidstheorie is dat er dus wel iets aan kan worden gedaan en dat hier ook indicatoren uit kunnen worden gedestilleerd, waar men deze inzichten niet hanteert. Om hier iets dieper op in te gaan bespreken we hier kort de beweging van *Crime Prevention Through Environmental Design* (CPTED) en Sociaal Veilig Ontwerpen.

De theorie van CPTED is gebaseerd op het idee dat criminaliteit mede het gevolg is van de gelegenheid die de fysieke omgeving daartoe biedt (zie ook Kube, 1982). Als dit correct is, moet het ook mogelijk zijn om de fysieke omgeving zodanig te manipuleren dat criminaliteit minder zal optreden. Dit uitgangspunt vertoont overeenkomsten met het voorbeeld uit de geschiedenis van de gezondheidszorg, waarbij verbeteringen in de fysieke omgeving diep hebben ingegrepen in de gezondheidssituatie en de levensverwachting in westerse landen gedurende de afgelopen eeuwen. De enorme vooruitgang in de gezondheidssituatie kan slechts gedeeltelijk worden toegeschreven aan de ontwikkeling van de medische praktijk. Hij werd vooral veroorzaakt door ingenieurs. Door de inspanning van architecten, planners, stedenbouwkundigen en technici kregen grote groepen mensen de beschikking over schoon drinkwater, riolering en afvalverwerking, met als gevolg een enorme verbetering van hun gezondheid (Kolthoff, 2016).

De term CPTED werd geïntroduceerd door Jeffery (1971), die betoogde dat preventie van criminaliteit vooral gericht moet zijn op het verkleinen van de mogelijkheden voor criminaliteit. In Engeland wordt voor CPTED het begrip *Designing Out Crime* gehanteerd en in Nederland spreekt men vaak van Sociaal Veilig Ontwerpen. Een Nederlands instrument, dat gericht betrekking heeft op het sociaal veilig ontwerpen van de buitenruimte, is de checklist Sociaal Veilig Ontwerpen, voortbouwend op het handboek van Van der Voordt en Van Wegen (1990). De volgende punten vatten de belangrijke thema's samen (Luten, 2008):

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



De ontwikkeling van barrière modellen²⁷ door het Centrum voor Criminaliteitspreventie en Veiligheid (CCV) sluit naadloos aan op die ontwikkeling. In de vele barrière modellen die al zijn ontwikkeld (helaas nog niet specifiek voor bedrijventerreinen) zie je als rode draad steeds de drieslag 'leegstand, infrastructuur en gebrek aan toezicht' terugkomen.²⁸

Nu we verschillende theorieën vanuit de criminologie hebben besproken, gaan we dieper in op de mogelijke indicatoren van criminaliteit op bedrijventerreinen die in het vooronderzoek naar voren zijn gekomen. De bron bij onderstaande paragrafen betreft een combinatie van literatuurstudie en bevindingen uit de expert meeting en interviews.

2.2 Ligging

De ligging van een bedrijventerrein is tijdens de expertmeeting genoemd als mogelijke indicator van criminaliteit op een bedrijventerrein. Net zoals voor legale bedrijven, is voor criminele organisaties de ligging van een terrein en de transportmogelijkheden hier vandaan een facilitator. Hierbij kan gedacht worden aan de ontsluiting van het terrein via de weg, het spoor of het water. Ook de afstand tot de grens met het buitenland en de afstand tot vliegvelden zouden belangrijke factoren kunnen zijn. Deze laatste zijn echter minder relevant voor dit onderzoek, aangezien alle bedrijventerreinen in deze pilot zich in de gemeente Tilburg bevinden. Wat wel een rol zou kunnen spelen is of het een binnenwijken bedrijventerrein betreft of niet.

2.3 Soort bedrijven (branches)

Naast de ligging van het bedrijventerrein zou ook gekeken kunnen worden naar het soort bedrijven (de branches) dat aanwezig is op het terrein. Op basis van de literatuur over ondermijnende criminaliteit, drugshandel en heling weten we dat een aantal branches kwetsbaar is voor misdaad, zoals de autobranche, opslag van spullen, transport en de groente- en fruitgroothandel.²⁹ De aanwezigheid van risicobranches wordt tijdens de expertmeeting genoemd als een van de belangrijkste indicatoren. Er wordt gesteld dat sommige bedrijven per definitie meer risicovol zijn dan andere. Hierbij worden vooral de autobranche, opslagbedrijven en ambulante handel als voorbeelden genoemd. In het algemeen worden branches waarbij weinig vakkennis vereist is en waarvoor er geen vergunningsplicht geldt meer risicovol genoemd. Omgekeerd kan het ook zo zijn dat bepaalde type bedrijven juist de kans verkleinen dat een bedrijventerrein afzakt (zoals bedrijven met een publieksfunctie, die zorgen namelijk voor meer sociale controle).

Bij de Wet BIBOB zijn de transportbranche, woningbouwcorporaties, coffeeshops, bordelen en smart- en growshops, en de branches bouw, milieu en ICT aangewezen als sectoren die kwetsbaar zijn voor criminaliteit. En in het onderzoek ter inventarisatie van de uitbreiding van de Wet BIBOB is die lijst uitgebreid met de speelautomatenbranche, het importeren van vuurwerk, headshops, en vastgoedtransacties met de overheid. Risicobranches zijn echter aan verandering onderhevig. Wanneer er streng toezicht wordt gehouden op een bepaalde

²⁷ Een barrière model brengt in kaart welke stappen een crimineel moet zetten om een delict te kunnen plegen en welke partijen en gelegenheden hierbij nodig zijn. Hierdoor wordt er inzicht gegeven in welke barrières er opgeworpen kunnen worden om het werk van criminelen te verstoren (bron: <https://hetccv.nl>).

²⁸ <https://www.barrieremodellen.nl/>.

²⁹ O.a. Ferwerda et al. (2016) over heling en Tops en Van der Torre (2015) over o.a. transportsector, autobranche en witwassen.



branche, is het mogelijk dat criminelen uitwijken naar andere branches waar zij minder in de gaten worden gehouden (Te Pest et al., 2012). Uit onderzoek blijkt dat de bedrijven waarin door de georganiseerde criminaliteit wordt geïnvesteerd, beïnvloed worden door vier factoren (Kleemans et al., 2002; p. 133):

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

Vanwege veranderende gelegenheidsstructuren en het verplaatsingseffect, is een uitsluitende lijst met branches niet te geven. Wel kunnen de eigenschappen van een branche worden genoemd die voor criminelen een gelegenheid bieden tot het succesvol witwassen van crimineel geld. Dit zijn volgens Te Pest et al. (2012; p. 60):

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

Een aantal branches loopt door organisatiekenmerken, maar ook door sociale factoren op individueel niveau (zoals het lage opleidingsniveau of een zorgelijke financiële positie van in dienst genomen personeel), extra risico om al dan niet bewust betrokken te raken bij criminele activiteiten. Zo noemen Te Pest et al. (2012; p. 73) energiemaatschappijen, de transportsector, distributieknooppunten en de financiële dienstverlening als risicobranches. Risicobranches die in verband met bedrijventerreinen in bovengenoemde onderzoeken steeds weer worden genoemd zijn de handel in kunst, antiek, dieren, sieraden, afval, voertuigen, prostitutie, de reinigingsbranche, belwinkels, avondkappers en massagesalons, de speelautomatenbranche en gokhallen, de transportsector, distributieknooppunten, de uitzendbranche, de land- en tuinbouwsector en growshop toeleveranciers en dekmantelbedrijven die zijn opgericht om criminele activiteiten af te schermen.

2.4 Fysieke omgeving en verloop

Een ander belangrijk kenmerk van een bedrijventerrein waar naar gekeken kan worden is de fysieke omgeving: is er sprake van veroudering, hoe wordt het terrein onderhouden en is er sprake van leegstand. In lijn met de in de inleiding besproken gelegenheidstheorie, de beweging van CPTED en Sociaal Veilig Ontwerpen, en de ontwikkeling van barrière modellen door het CCV, kwam de fysieke omgeving ook tijdens de expertmeeting naar voren als een belangrijke indicator voor criminaliteit op bedrijventerreinen. Ten eerste werden een rommelig terrein, afvallozing en onderhoud genoemd, in lijn met de *broken windows theory* (Kelling en Wilson, 1982). Deze theorie stelt dat omgevingen die reeds vervuild zijn, meer vuil aantrekken en verloederde terreinen waar zich veel criminelen bevinden zullen dan ook nieuwe criminele



aantrekken (Kelling en Wilson, 1982). Ten tweede kwam leegstand zowel in de expertmeeting als in interviews naar voren als een belangrijke indicator. Als derde indicator met betrekking tot de fysieke omgeving werd veroudering genoemd. Volgens Beekmans (2015; p. 58) lopen de gevolgen van veroudering erg uiteen: van fysieke aftakeling tot sociaal onwenselijke verschijnselen zoals een toename van criminaliteit en de bijbehorende onveiligheidsgevoelens. Beekmans (2015) inventariseerde een aantal criteria voor veroudering van bedrijventerreinen dat is samengevat in Tabel 2, en dat is ontleend aan cijfers van het Planbureau voor de Leefomgeving (PBL). Hierin komen ook onderhoud en leegstand weer naar voren als belangrijke criteria.

Criteria	Aantal keer genoemd
Achterstallig onderhoud en beheer	53
Leegstand/verouderd vastgoed	35
Verkeersproblemen	25
Aard van de gevestigde bedrijvigheid	15
Leeftijd	10
Gebrek aan economische dynamiek	10
Beeldkwaliteit	8
Extensief ruimtegebruik	8
Kwaliteit ondergrondse infrastructuur	4
Overige	5

Tabel 2. Criteria voor veroudering van bedrijventerreinen volgens 64 gemeenten (Beekmans, 2015, p. 59)

Ook het verloop zou iets over het bedrijventerrein kunnen zeggen. Het verloop werd als belangrijkste indicator genoemd tijdens de expertmeeting. Verloop zou het gevolg kunnen zijn van criminelen die hun misdaden willen verhullen. Hierbij kan aan de ene kant gedacht worden aan het verloop van bedrijven op een bedrijventerreinen: wisselen de bedrijven die op een terrein zitten snel of zitten dezelfde bedrijven al jaren op het terrein? Daarnaast kan verloop betrekking hebben op een bedrijf: wisselt een bedrijf vaak van eigenaar?

2.5 Onroerend goed

Wanneer het met een buurt beter gaat is dat terug te zien aan een stijging van de waarde van het vastgoed (zie bijvoorbeeld Walks en Maaranen, 2008; p. 296). Omdat bedrijventerreinen geen bewoners met sociaaleconomische kenmerken kennen, zijn veel van de gebruikte indicatoren bij het proces van gentrification³⁰ (Kolthoff, 2016; p. 153) niet bruikbaar voor onderzoek naar veroudering van bedrijventerreinen. Dat geldt echter niet voor de waarde van vastgoed. Veroudering en verloedering doen de waarde van onroerend goed op bedrijventerreinen afnemen. Daar staat tegenover dat bepaalde indicatoren die een bedrijventerrein aantrekkelijk maken, zoals bereikbaarheid en specialisatie (bv. logistiek) los van verouderingsprocessen de waarde van onroerend goed hoog houden. De waarde van onroerend goed op het bedrijventerrein zou dus een indicator kunnen zijn voor criminaliteit en ondermijning. In de interviews kwam naar voren dat een (significante) afwijking tussen marktwaarde en verkoopwaarde voor vastgoed een indicator zou kunnen zijn.

³⁰ Gentrification is een proces waarbij stadswijken een opwaardering doormaken op sociaal, cultureel en economisch gebied. Dit gaat veelal gepaard met een stijging van de waarde van het onroerend goed in de wijk en een stijging van het gemiddeld inkomensniveau van de wijkbewoners (door het aantrekken van kapitaalkrachtige nieuwe bewoners/gebruikers).



Veronderstellend dat de marktwaarde gebaseerd is op een objectieve economische inschatting van de daadwerkelijke waarde van vastgoed, zou een grote afwijking van de verkoopwaarde er op kunnen wijzen dat er sprake is van een subjectieve inschatting van de waarde die niet (alleen) gebaseerd is op een economische inschatting, maar eventueel op andere zaken zoals omkoping, bevoordeling van zakenpartners, of andere criminele activiteiten. Hierbij werd wel opgemerkt dat de WOZ waarde bij niet-woningen vaak niet overeenkomt met de echte marktwaarde.

2.6 Toezicht, handhaving en samenhang

Een indicator die genoemd werd tijdens de expertmeeting is de hoeveelheid toezicht en handhaving op het terrein en de mate van samenhang. Een belangrijk voorbeeld zijn de verloederde bedrijventerreinen, waar de overheid in de loop der jaren de teugels heeft laten vieren door langdurig weg te kijken. De problemen beginnen vaak klein, maar doordat de overheid weinig interesse toont of de regels niet handhaaft, worden de problemen steeds groter en groeien dergelijke terreinen uit tot ware vrijplaatsen. Hoe kan de overheid in deze kwetsbare gebieden haar gezag herstellen? Dat kan volgens Lam et al. (2018) mede door strikte repressie en handhaving. Door het stellen van duidelijke grenzen kunnen maatschappelijke normen en waarden in ere worden hersteld. Handhaving, ook op kleine afwijkingen, is hierbij noodzakelijk (Lam et al., 2018; p. 214).

Op de bedrijventerreinen die door Lam et al. (2018) worden beschreven zijn wietplantages en xtc-laboratoria aangetroffen. Bij politie en justitie ontstaat het idee om naast het opsporen van die criminaliteit in te zetten op publiek-private samenwerking. Het idee is om via voorlichting de ondernemers bewust te maken van de schadelijke gevolgen van deze criminaliteit, waarbij vooral wordt ingezet op het thema ondermijning. De aanwezigheid van criminele activiteiten op het bedrijventerrein geeft het een slechte naam en kan uiteindelijk schadelijke gevolgen hebben voor de bedrijvigheid in het algemeen (Lam et al., 2018; p. 156).

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

Een andere mogelijke indicator is de aanwezigheid van parkmanagement op het bedrijventerrein. Parkmanagement is een manier om de kwaliteit van bedrijventerreinen in stand te houden of te verbeteren. Gebruikelijke parkmanagement activiteiten zijn: beheer en onderhoud van de openbare ruimte, bewegwijzering, collectieve beveiliging, gezamenlijke afvalinzameling, en verwerking en gezamenlijke inkoop van onderzoek en advies. Parkmanagement kan op verschillende manieren georganiseerd worden: door commerciële organisaties die parkmanagement aanbieden of door ondernemersverenigingen die het management van hun terrein zelf verzorgen. Op nieuwe bedrijventerreinen is deelname vaak verplicht, op bestaande is het meestal een zaak van de ondernemersvereniging.

Naast parkmanagement, kan ook gekeken worden naar de aanwezigheid van een Keurmerk Veilig Ondernemen voor bedrijventerreinen (KVO-B). MKB-Nederland is in samenwerking met



het toenmalige ministerie van Veiligheid en Justitie in 2005 gestart met dit keurmerk. In het KVO-B maken ondernemers samen met de gemeente, brandweer, politie en eventuele andere partijen afspraken om overlast en criminaliteit aan te pakken en onderhoud en beheer te regelen. Voorbeelden van maatregelen die worden genomen zijn:

- het organiseren van bijeenkomsten voor en door ondernemers over *'schoon, heel en veilig'*
- beter georganiseerd toezicht door politie en beveiliging
- betere verlichting en het plaatsen van beveiligingscamera's
- investeringen in de openbare ruimte.³¹

2.7 Aantal meldingen bij politie en Meld Misdaad Anoniem

Het aantal meldingen over criminaliteit en overlast bij de politie kan als een mogelijke indicator worden gezien. Als er veel meldingen of incidenten geregistreerd staan bij de politie is dit een signaal dat er criminaliteit plaatsvindt. Aan de andere kant zou de afwezigheid van meldingen en aangiftes kunnen betekenen dat mensen bang zijn, of de moed hebben verloren dat er iets aan de problemen wordt gedaan. Als het lage aantal meldingen het gevolg is van een lage meldingsbereidheid doordat mensen bedreigd worden door criminelen en bang zijn voor represailles, zou juist de afwezigheid van aangiftes/meldingen als indicator voor georganiseerde criminaliteit kunnen gelden. Er kan dus geen eenduidige conclusie getrokken worden met betrekking tot de relatie tussen het aantal meldingen bij de politie en de kans op criminaliteit.³²

Naast de meldingen bij de politie, is ook Meld Misdaad Anoniem (M.) een interessante databron. M.³³ is een onafhankelijk meldpunt waar men anoniem informatie kan geven over criminaliteit en misdaad. Naast het absolute aantal meldingen bij M. zou ook het aantal meldingen in verhouding tot de meldingen bij de politie interessant kunnen zijn (zijn er bijvoorbeeld weinig meldingen bij de politie, maar wordt er wel veel anoniem gemeld).

2.8 Meldingen bij de gemeente en informatie uit het verleden

Het aantal meldingen bij het Centraal Meldpunt (CMP) van de gemeente, bijvoorbeeld over de fysieke omgeving op het bedrijventerrein (zie ook §2.4), zoals meldingen van zwerfvuil of illegale stortingen, maar ook over bijvoorbeeld hennepkwekerijen, zou een indicator kunnen zijn. Daarnaast is er bij de gemeente veel informatie bekend over situaties die zich in het verleden hebben voorgedaan op het bedrijventerrein, zoals het aantal hennepstuiven of het aantal panden dat is gesloten (Damocles sluitingen).

³¹ MKB-Nederland, zie <https://www.mkb.nl/kvob>, geraadpleegd op 10 september 2018.

³² Alhoewel dit een vaak genoemd fenomeen is waarover diverse krantenartikelen zijn geschreven, zijn er geen goede bronnen in de literatuur te vinden. Om hier goed onderbouwde uitspraken over te kunnen doen zou er sprake moeten zijn van aanvullend onderzoek.

³³ <https://www.meldmisdaadanoniem.nl/>.



Samenvatting 'Theoretisch kader'

In de criminologie is de volkswijsheid 'De gelegenheid maakt de dief' theoretisch verwoord door Felson en Clarke (1998) in hun gelegenheidstheorie, die op één enkel principe is gebaseerd: verleidelijke en gemakkelijke gelegenheden om criminaliteit te plegen zetten mensen aan om dit ook te doen. Als criminaliteit mede het gevolg is van de gelegenheid die de fysieke omgeving daartoe biedt, moet het ook mogelijk zijn om de fysieke omgeving zodanig te manipuleren dat criminaliteit minder zal optreden. Dit is de kern van de beweging van 'Crime Prevention Through Environmental Design' (CPTED), wat in Engeland ook wel 'Designing Out Crime' en in Nederland 'Sociaal Veilig Ontwerpen' genoemd wordt. Gerelateerd aan bovenstaande, zijn in dit onderzoek een aantal indicatoren meegenomen die betrekking hebben op de fysieke omgeving (waaronder onderhoud, veroudering en leegstand) en toezicht, handhaving en samenhang op het bedrijventerrein. Andere mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen die in het vooronderzoek naar voren zijn gekomen hebben betrekking op de ligging van het bedrijventerrein, het soort bedrijven (branches) die er aanwezig zijn, het verloop, onroerend goed, meldingen bij de politie en Meld Misdaad Anoniem, meldingen bij de gemeente en overige informatie die bij de gemeente bekend is bijvoorbeeld over incidenten uit het verleden. In het volgende hoofdstuk zullen de beschikbare databronnen met betrekking tot bovengenoemde indicatoren besproken worden en wordt het conceptueel model gepresenteerd.



3 Conceptueel model en dataverzameling

Na het vooronderzoek zijn de belangrijkste indicatoren vanuit de verschillende bronnen verzameld. Hierbij waren we afhankelijk van de beschikbaarheid en de verkrijgbaarheid van de data. We zullen in dit hoofdstuk slechts de databronnen die voor dit onderzoek verzameld konden worden beschrijven en gaan hierbij ook in op het soort informatie dat we uit deze bronnen kunnen halen. De volledige lijst met variabelen met een toelichting is bijgevoegd als Appendix B. Data zijn zoveel mogelijk op bedrijventerreinniveau verzameld en data op andere aggregatieniveaus (bijvoorbeeld bedrijf, adres of postcode) zijn later geaggregeerd tot bedrijventerreinniveau (dit lichten we verder toe in §4.3). De meest recente data is opgevraagd en vaak zijn data van meerdere jaren opgevraagd (met als richtlijn de jaren 2014 t/m 2018, indien beschikbaar), zodat deze later nog geaggregeerd kon worden als de data te summier bleek te zijn.

Integraal Bedrijventerreinen Informatie Systeem (IBIS)

Het IBIS bevat informatie over alle bedrijventerreinen in Nederland. Gemeenten vullen per bedrijventerrein een vragenlijst in, waarna de vragenlijsten door de provincies verzameld en gecontroleerd worden. In het IBIS staat een aantal kenmerken per bedrijventerrein, en daarnaast wordt de locatie van het terrein vastgelegd op basis van een geometriebestand. De kenmerken die we bekijken zijn onder andere oppervlakte (bruto, netto, nog uitgeefbaar), structurele leegstand (en percentage), schaalniveau, opzet, kwaliteit terrein, functiemenging met wonen, minimale en maximale milieucategorie, bereikbaarheid en parkeerdruk, ontsluiting van het terrein (per weg, spoor, water en openbaar vervoer), veroudering en parkmanagement.

Politie

Uit het incidentregistratiesysteem dat wordt gebruikt door de politie zijn data ter beschikking gesteld voor dit onderzoek, te weten: het aantal incidenten per incidentsoort per jaar (van 2014 t/m gedeelte 2018). Omdat het om kleine aantallen gaat zijn de data voor alle jaren geaggregeerd in drie categorieën:

- Drugsgelateerde incidenten, bijvoorbeeld het bezit van, handel in of vervaardigen van soft- of harddrugs.
- Incidentsoorten die gerelateerd zouden kunnen zijn aan georganiseerde misdaad, zoals bedreiging, bezit vuurwapens, chantage/afpersing, doodslag/moord en witwassen.
- Overige incidentsoorten, zoals diefstal.

Meld Misdaad Anoniem

Meld Misdaad Anoniem (M.) is het onafhankelijke meldpunt waar anoniem informatie gegeven kan worden over criminaliteit en misdaad. Dit meldpunt wordt aangeboden door NL Confidential. Op langere termijn worden alleen meta-data van M. bewaard, niet de oorspronkelijke data. Deze bestaan uit datum, locatie (op postcode-6 niveau) en criminaliteitsklasse. Wederom is de data geaggregeerd voor alle jaren en in categorieën:

- Incidentsoorten die gerelateerd zouden kunnen zijn aan georganiseerde misdaad, zoals crimineel onverklaarbaar vermogen, energiediefstal, harddrugs, hennepkwekerij, laboratorium XTC/GHB en softdrugs.
- Overige incidentsoorten, zoals diefstal.



Keurmerk Veilig Ondernemen Bedrijventerreinen (KVO-B)

Een lijst van bedrijventerreinen die een KVO-B certificering hebben is te vinden op de website <https://www.hetkeurmerkveiligondernemen.nl/over-het-kvo/wie-heeft-een-kvo/>. Om de informatie op de website te verifiëren, hebben we tevens contact gehad met de landelijk projectleider KVO-B, die ons de meest recente gegevens heeft aangeleverd. Zo wisten we dat een bepaald bedrijventerrein tijdens dit onderzoek haar certificaat heeft behaald, en dat een van de certificaten geldt voor twee IBIS-bedrijventerreinen. In de analyses hebben we de variabele 'Keurmerk Veilig Ondernemen Bedrijventerreinen ja/nee' meegenomen.

Gemeente

Vanuit de gemeente Tilburg hebben we verschillende data ontvangen. Ten eerste, data over meldingen die burgers doen bij de gemeente via het Centraal Meldpunt. Het aantal meldingen is geaggregeerd over de periode 2016 t/m 24-6-2018 en verdeeld in twee categorieën: drugsgelateerd (bv. hennepafval, hennepkwekerij en drugsoverlast) en overig. Ten tweede hebben we inzicht gekregen in het aantal hennepruiming (data zijn geaggregeerd voor de jaren 2016 en 2017) en Damocles sluitingen³⁴ per bedrijventerrein (data zijn geaggregeerd voor de periode 2016 t/m 24-6-2018). Verder hebben we data gekregen met betrekking tot de onroerende zaken, bijvoorbeeld over het gebruik van gebouwen (bv. opslag/distributie) en de WOZ waarde.

Ten slotte zijn er nog data uit overige bronnen ter beschikking gesteld voor dit onderzoek:

- Basisregistraties Adressen en Gebouwen (BAG): Dit bevat informatie over onder andere de oppervlakte en het bouwjaar van gebouwen en het gebruiksdoel (bv. winkelfunctie, bijeenkomstfunctie). Het BAG bestand bevat ook de naam van de bedrijventerreinen, wat van groot belang was voor dit onderzoek. Hierdoor kon data op adres of postcode niveau namelijk herleid worden tot een bedrijventerrein en vervolgens dus geaggregeerd worden tot bedrijventerreinniveau.
- Bedrijvenregister Tilburg: Dit wordt beheerd door de gemeente Tilburg. Het register is bedoeld om de werkgelegenheid naar sector per gemeente/buurt/wijk/bedrijventerrein e.d. in kaart te brengen, over minimaal de afgelopen tien jaar. Hiervoor wordt jaarlijks een werkgelegenheidsenquête gehouden en krijgt men maandelijks een mutatiebestand van de Kamer van Koophandel (KvK). Niet alle mutaties van de KvK worden doorgevoerd in het Bedrijvenregister. Niet alle inschrijvingen worden bijvoorbeeld opgenomen in het register, want er moet wel daadwerkelijk iemand werkzaam zijn. Pensioen-bv's en beheer-bv's e.d. worden dus niet opgenomen. Ook loodsen, waarvan men aangeeft dat er alleen opslag is en waarin geen personen werkzaam zijn, worden niet opgenomen. Extra inschrijvingen van hetzelfde bedrijf op hetzelfde adres worden ook niet opgenomen in het register om dubbeltellingen van banen en vestigingen te voorkomen. Het register bevat per vestiging informatie over het aantal werkzame personen (ingedeeld in grootteklassen) en onderverdelingen naar fulltime/parttime en geslacht. Tevens bevat het register data uit het register van de KvK, zoals de rechtsvorm en de SBI-code (sector).

³⁴ Met het aantal Damocles sluitingen refereren we naar het aantal panden dat gesloten is op basis van de Wet Damocles. Artikel 13b van de Opiumwet (ook wel de Wet Damocles genoemd) biedt burgemeesters de mogelijkheid om bestuursdwang toe te passen en panden te sluiten als daar sprake is van drugshandel.



Netbeheerder energie (Enexis)

Enexis Netbeheer transporteert elektriciteit en gas naar huizen en bedrijven, onder andere in de gemeente Tilburg. Enexis heeft per bedrijventerrein geanonimiseerde/geaggregeerde data³⁵ ter beschikking gesteld over de capaciteit van elektriciteitsaansluitingen, de aanwezigheid van slimme meters³⁶, het aantal keren dat een slimme meter geweigerd is, de (contractuele) leegstand en de som van het aantal dagen dat er geen contract aanwezig was, het aantal historische gevallen van fraude en het aantal spanningsklachten.

In Figuur 3 is een globale weergave van het conceptueel model te vinden. Dit model vormt de basis voor de analyses die in het volgende hoofdstuk worden beschreven. De complete lijst met variabelen die zijn meegenomen in de analyses is te vinden in Appendix B.

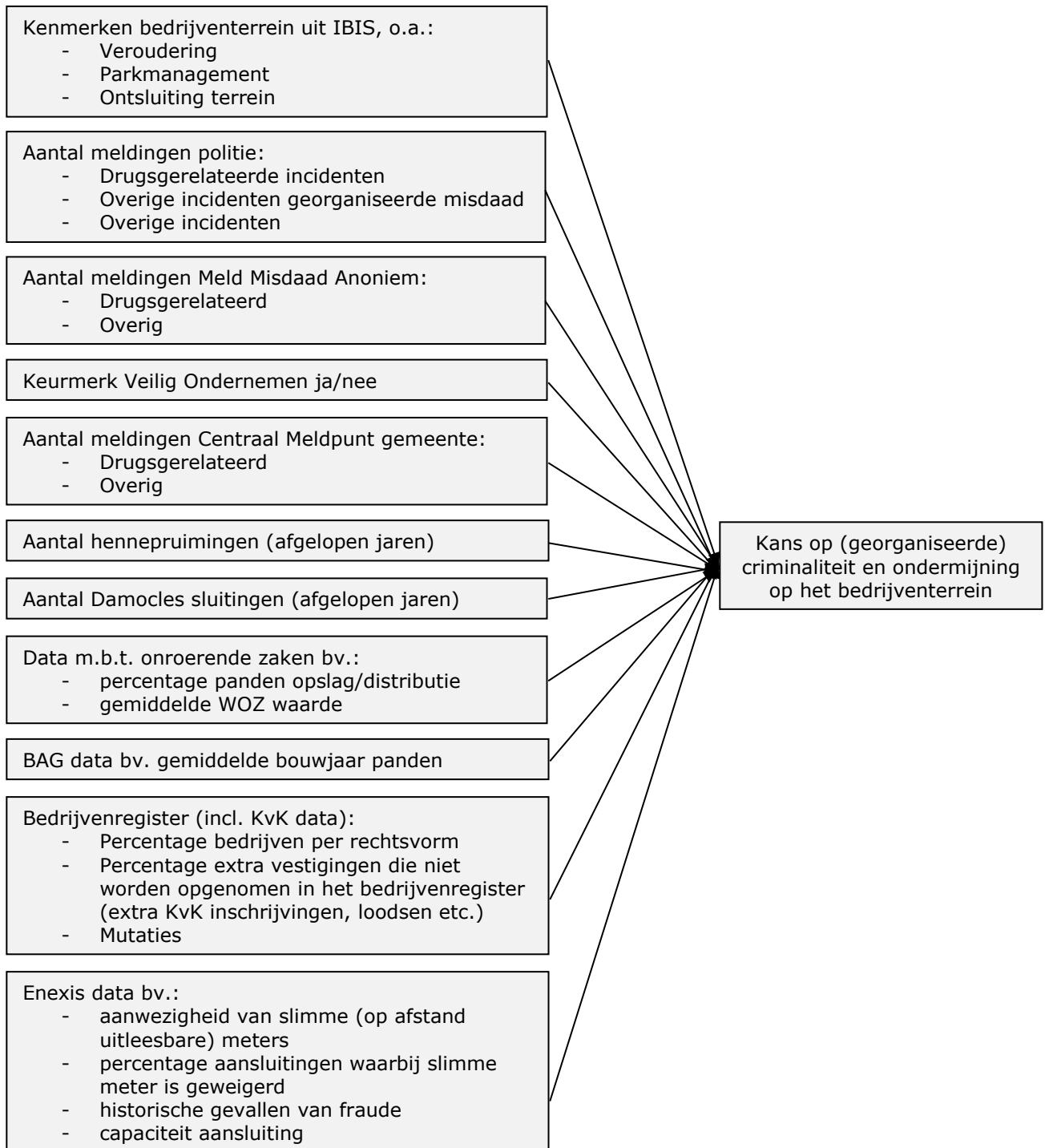
Samenvatting 'Dataverzameling'

Na het identificeren van mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, is data verzameld uit verscheidene bronnen. Kenmerken over bedrijventerreinen (zoals de aanwezigheid van parkmanagement en de ontsluiting van het terrein) waren beschikbaar in het IBIS. Daarnaast hebben we het aantal meldingen bij de politie en Meld Misdaad Anoniem ontvangen. Of een bedrijventerrein een Keurmerk Veilig Ondernemen heeft was online in te zien. De gemeente Tilburg heeft het aantal meldingen bij het Centraal Meldpunt (bijv. over drugsoverlast), het aantal hennepuimingen, het aantal Damocles sluitingen en data met betrekking tot de onroerende zaken op de bedrijventerreinen geleverd. Daarnaast hebben ze data uit de BAG en het bedrijvenregister (waaraan de KvK ten grondslag ligt) ter beschikking gesteld. Tot slot heeft netbeheerder Enexis data geleverd over de aanwezigheid en het weigeren van slimme meters, de capaciteit van elektriciteitsaansluitingen en historische gevallen van fraude. Op basis van het bovenstaande zullen we nu het conceptueel model presenteren.

In Figuur 3 is een globale weergave van het conceptueel model te vinden. Dit model vormt de basis voor de analyses die in het volgende hoofdstuk worden beschreven. De complete lijst met variabelen die zijn meegenomen in de analyses is te vinden in Appendix B.

³⁵ Dat betekent dat de deze gegevens niet te herleiden zijn tot een natuurlijk persoon.

³⁶ Een slimme meter is het nieuwste type gas- en elektriciteitsmeter. Deze meter kan meterstanden op afstand doorsturen en is geschikt voor het registreren van dubbeltarief. Ook kan de meter bijhouden hoeveel energie wordt teruggeleverd aan het energienet. Een slimme meter kan uw energieverbruik digitaal meten en opslaan. Klanten kunnen zelf gegevens uit het geheugen van hun slimme meter halen en Enexis kan dat ook. (Bron: www.enexis.nl)



Figuur 3. Conceptueel model



4 Analyses en resultaten

In dit hoofdstuk worden de analyses en resultaten besproken. Eerst gaan we in op het labelen van de bedrijventerreinen (§4.1). In §4.2 bespreken we vervolgens de datapreparatie. Tot slot lichten we de gebruikte analysemethode Random Forest (§4.3) toe met de bijbehorende uitkomsten (§4.4).

4.1 Labelen van de bedrijventerreinen

Twee manieren om per bedrijventerrein een inschatting van de kans op georganiseerde criminaliteit en ondermijning te verkrijgen zijn tijdens het onderzoek in overweging genomen. Het aantal meldingen bij politie en/of Meld Misdaad Anoniem (M.) zou mogelijkwijs een indicatie kunnen geven. Echter, zoals in paragraaf 2.7 besproken is, kan het ook zo zijn dat er juist een lage meldingsbereidheid is doordat er georganiseerde criminaliteit en ondermijning plaatsvindt. Vanuit de theorie bestaat er geen eenduidige aanwijzing voor de relatie tussen politiemeldingen en (georganiseerde) criminaliteit en ondermijning (positief of negatief). Onder andere hierdoor was het gebruiken van het aantal geregistreerde incidenten bij politie als absolute maatstaf voor de mate van georganiseerde criminaliteit en ondermijning lastig. Daarom is ervoor gekozen om een kleine vragenlijst af te nemen om de bedrijventerreinen te categoriseren. Hierin werd aan respondenten gevraagd om voor elk van de bedrijventerreinen in de gemeente Tilburg een inschatting te maken of de kans op (georganiseerde) criminaliteit en ondermijning laag, gemiddeld of hoog is. Daarnaast was er een 'weet niet' optie, zodat respondenten die een bedrijventerrein niet kenden de vragenlijst toch konden invullen. De vragenlijst is opgenomen als Appendix C. De vragenlijst is verspreid binnen een diverse groep respondenten met verschillende achtergronden/functies. Uiteindelijk hebben elf van de 21 uitgenodigde respondenten de vragenlijst ingevuld (zie Tabel 3).³⁷

Respondent	Organisatie	Rol / expertisegebied
Respondent 1	Gemeente Tilburg	beleid
Respondent 2	Gemeente Tilburg	informatie- en kenniscentrum
Respondent 3	Gemeente Tilburg	accountmanager bedrijventerreinen
Respondent 4	Gemeente Tilburg	accountmanager bedrijventerreinen
Respondent 5	parkmanagement organisatie	parkmanager
Respondent 6	parkmanagement organisatie	parkmanager
Respondent 7	Gemeente Tilburg	handhaving
Respondent 8	Gemeente Tilburg	omgevingsmanager/wijkregisseur
Respondent 9	politie	coördinator
Respondent 10	Gemeente Tilburg	ondermijning
Respondent 11	Veiligheidsregio Midden West-Brabant	adviseur

Tabel 3. Respondenten vragenlijst labelen bedrijventerreinen

Beperkingen van deze aanpak zijn dat er maar elf respondenten waren en dat er mogelijkwijs sprake is van subjectiviteit, een gebrek aan kennis en/of eigen belangen. We hebben geprobeerd om dit te verminderen door zoveel mogelijk respondenten te benaderen van zoveel mogelijk verschillende organisaties en met verschillende rollen en

³⁷ In Tabel 3 zijn de organisaties en een globale omschrijving van de functies/rollen/expertisegebieden van de respondenten toegevoegd om een beeld te geven van de respondenten die de vragenlijst hebben ingevuld, maar tegelijkertijd de privacy van de respondenten te waarborgen.



expertisegerieden. Echter bestaat er, net als in ander kwalitatief inductief onderzoek, wel het risico op contaminatie doordat de respondenten wellicht de indicatoren in gedachten houden en laten meewegen bij hun beoordeling van de bedrijventerreinen. In onderhavig onderzoek lijkt dit niet tot problemen te leiden, aangezien de resultaten uit de vragenlijst overeenkomen met signalen die zijn voortgekomen uit integrale overheidssamenwerking (zie hieronder).

Resultaten labelen bedrijventerreinen

De afzonderlijke scores van de respondenten zijn gecombineerd tot een eindscore per bedrijventerrein door gebruik te maken van 'majority voting'. De categorie die door de meeste respondenten gekozen werd is dus gebruikt als totaalscore.³⁸ In Appendix C.1 zijn de volledige resultaten van de vragenlijst opgenomen en is toegelicht hoe is omgegaan met twijfelgevallen. Volgens de respondenten hebben acht bedrijventerreinen een lage kans, vijftien terreinen een gemiddelde kans en zeven terreinen een hoge kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning.

Om extra te controleren of de resultaten uit de vragenlijst een goede indicatie te geven van de mate van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, hebben we een vergelijking gemaakt met signalen die zijn voorgekomen uit integrale overheidssamenwerking.³⁹ Hierbij is gebruik gemaakt van signalen die in 2013 zijn vastgesteld. Voor vijf van de zeven terreinen die als hoge kans naar boven kwamen in de vragenlijst waren er ook signalen vanuit de integrale overheidsaanpak. Bedrijventerrein 15⁴⁰, waar wel signalen voor waren, wordt in onze vragenlijst als gemiddeld gecategoriseerd, al zijn er ook drie respondenten die dit terrein als hoog hebben gecategoriseerd (tegenover vijf respondenten die het terrein als gemiddeld beoordeelden). Navraag bij de gemeente wees uit dit bedrijventerrein zich in de afgelopen jaren over het algemeen goed heeft ontwikkeld, wat een mogelijke verklaring zou kunnen zijn dat het terrein nu vaker als 'gemiddeld' beoordeeld wordt in plaats van als 'hoog'. Tot slot waren er signalen vanuit de integrale overheidssamenwerking over bedrijventerrein 4. Echter is ook hier in de afgelopen vijf jaren redelijk veel veranderd, wat een mogelijke verklaring zou kunnen zijn voor het verschil tussen de signalen uit de integrale overheidssamenwerking uit 2013 en dit onderzoek in 2018. In het algemeen zijn de resultaten uit de vragenlijst in lijn met de signalen die in 2013 naar voren zijn gekomen in de integrale overheidssamenwerking.

³⁸ Bij majority voting wordt aan ieders beoordeling een gelijk gewicht toegekend. Een beperking hierbij is dat de ene expert meer kennis kan hebben dan andere experts. Aangezien het niet mogelijk was om op een objectieve en betrouwbare wijze verschillende gewichten toe te kennen aan de respondenten, is er gewerkt met gelijke gewichten.

³⁹ Op verzoek is de bron geanonimiseerd.

⁴⁰ De bedrijventerreinen worden in dit rapport niet bij naam genoemd, omdat uit dit onderzoek geen betrouwbare conclusies getrokken kunnen worden over de specifieke terreinen. Daarnaast was het doel van dit onderzoek om de belangrijke risicofactoren te verkennen en niet om de risico's op de specifieke bedrijventerreinen in de gemeente Tilburg vast te stellen en openbaar te maken.



Samenvatting 'Labelen van de bedrijventerreinen'

Twee manieren om per bedrijventerrein een inschatting van de kans op georganiseerde criminaliteit en ondermijning te verkrijgen zijn tijdens het onderzoek in overweging genomen. Uiteindelijk is er in dit onderzoek voor gekozen om de bedrijventerreinen te 'labelen' (in lage, gemiddelde, of hoge kans) door aan meerdere mensen met verschillende achtergronden en vanuit verschillende organisaties te vragen deze categorisatie te maken in een kleine vragenlijst en vervolgens alle resultaten te combineren tot een eindindeling. Hieruit kwam dat acht bedrijventerreinen een lage kans, vijftien terreinen een gemiddelde kans en zeven terreinen een hoge kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning hebben. De resultaten uit de vragenlijst waren grotendeels in lijn met signalen die zijn voortgekomen uit integrale overheidssamenwerking.

4.2 Datapreparatie

Voor de 30 bedrijventerreinen in de gemeente Tilburg is een grote hoeveelheid data verzameld met betrekking tot de mogelijke indicatoren. Een aantal databronnen is aangeleverd op het niveau van bedrijventerrein; dat wil zeggen dat er 1 observatie per bedrijventerrein was. Echter waren er ook databronnen die op een ander aggregatieniveau zijn aangeleverd (bijvoorbeeld postcode-6 niveau) en die dus door de onderzoekers geaggregeerd moesten worden tot bedrijventerreinniveau. Daarnaast moesten de verschillende databronnen aan elkaar gekoppeld worden. In Appendix D en E wordt de data pre-processing en koppeling uitgebreid beschreven.

4.3 Modelselectie

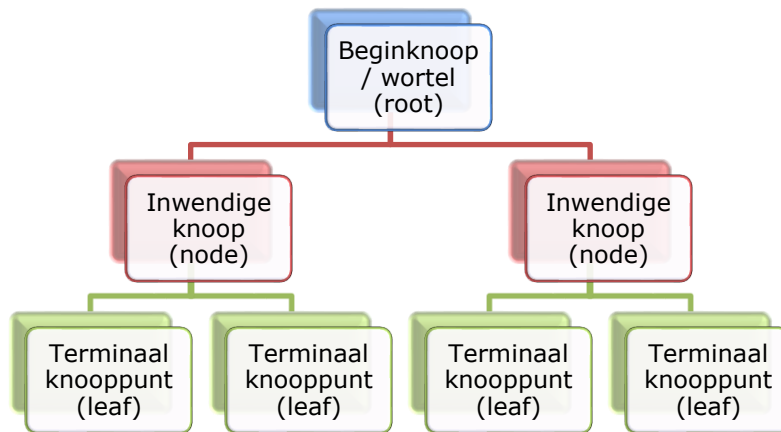
In deze paragraaf bespreken we de modelselectie. Eerst zullen we verschillende modeltypen bespreken die mogelijk gebruikt zouden kunnen worden (§4.3.1). Vervolgens gaan we in op de prestatiestatistieken die we kunnen gebruiken om verschillende modellen te beoordelen en met elkaar te vergelijken (§4.3.2). Tot slot wordt het gekozen model toegelicht (§4.3.3).

4.3.1 Modeltypen

Er bestaan verschillende classificatie technieken voor het voorspellen van een categorale variabele (in dit onderzoek: een lage/gemiddelde/hoge kans op (georganiseerde) criminaliteit en ondermijning op het bedrijventerrein) op basis van een aantal indicatoren. Hieronder zullen we de verschillende technieken beschrijven.

Beslissingsbomen, de bouwstenen van een Random Forest

Een eenvoudig interpreteerbare en veelgebruikte classificatiemethode is de beslissingsboom, in het Engels ook wel *decision tree* genoemd. Bij het maken van een beslissingsboom beginnen we bij de wortel (zie Figuur 4). De training observaties worden zo heterogeen mogelijk opgesplitst in twee subgroepen. Bij elk knooppunt wordt bekeken welke variabele de data het beste op kan splitsen in twee nieuwe knooppunten. Op deze manier wordt de data steeds verder opgesplitst, net zolang totdat aan een stopcriterium is voldaan (bijvoorbeeld minder dan n training observaties per knooppunt). Op basis van de waarden voor de variabelen, valt elke observatie uiteindelijk in één klasse (dus in één leaf).

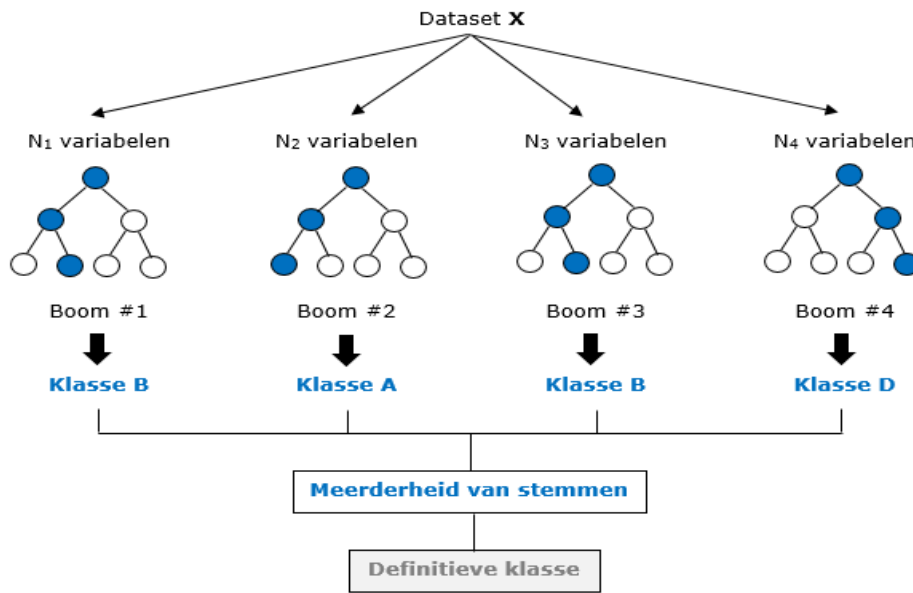


Figuur 4. Schematische weergave beslissingsboom

Een voordeel van een beslissingsboom is dat het begrijpelijk is, dat de resultaten goed te interpreteren zijn en dat de resultaten grafisch kunnen worden weergegeven. Het nadeel van beslissingsbomen is echter dat ze instabiel kunnen zijn en dat een relatief kleine verandering in de data kan resulteren in een andere boom. Met andere woorden: een beslissingsboom kan knopen bevatten die gecreëerd zijn door specifieke gevallen in de trainingsdataset, waardoor het model zich slecht laat generaliseren naar andere data. Het model heeft dan een grote 'generalization error', een fenomeen dat ook wel 'overfitting van de data' genoemd wordt. Een manier om dit te voorkomen is het gebruik van ensemble methoden, zoals bijvoorbeeld een Random Forest. Deze methode zullen we nu beschrijven.

Random Forest

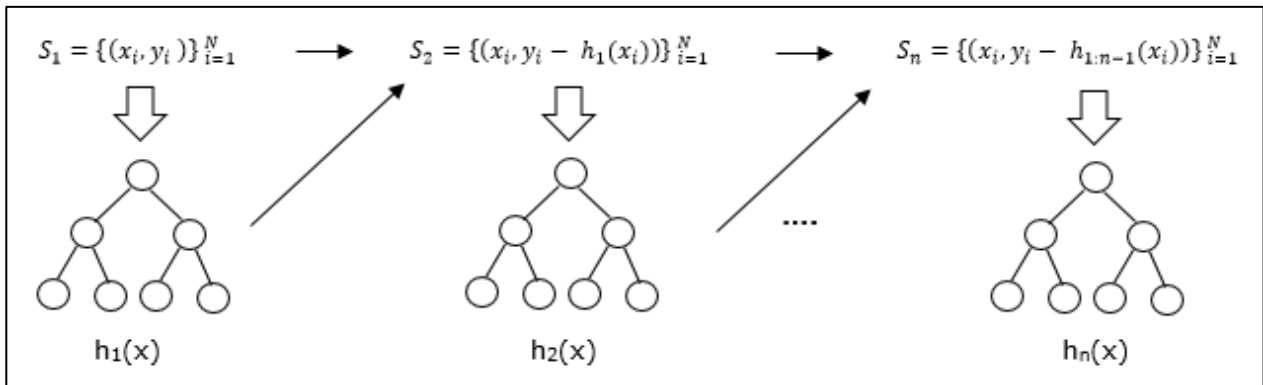
Random forest (RF) is een 'tree-based supervised learning' techniek, waarbij een groot aantal beslissingsbomen gecombineerd wordt om tot de uiteindelijke voorspelling te komen. Hierdoor is deze methode veel minder instabiel dan een enkele beslissingsboom. Als de doelvariabele die we willen voorspellen categoriaal is, wordt de uiteindelijke uitkomst bepaald door middel van 'majority voting'. Met andere woorden: de uitkomst van de meeste bomen geldt als einduitkomst (zie Figuur 5). 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 RF geeft over het algemeen veel betere voorspellingen dan een beslissingsboom (Slof, 2014).



Figuur 5. Schematische weergave Random Forest algoritme

Gradient Boosting

Bij gradient boosting wordt er gebruik gemaakt van boosting technieken om de voorspellende waarde van classificatie en regressie te optimaliseren. Gradient boosted tree, doorgaans ook gewoon gradient boosting genoemd, is een ensemble methode om beslissingsbomen te verbeteren. In een stapsgewijs proces worden er meerdere modellen gecreëerd, waarbij elke nieuw model leert van het voorgaande model en het generaliseert hen door optimalisatie van een willekeurige differentieerbare verliesfunctie toe te staan (zie Figuur 6).



Figuur 6. Gradient Boosting

4.3.2 Prestatiestatistieken

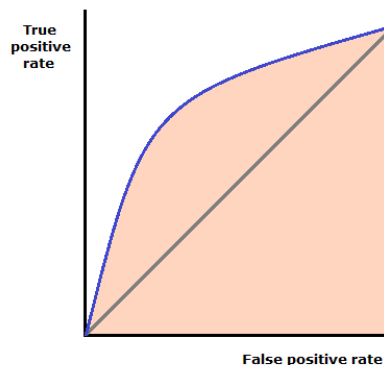
Om de prestatie van een model te beoordelen en verschillende modellen te vergelijken, kan gebruik gemaakt worden van verschillende statistieken (zie ook Appendix F):

- Accuracy (nauwkeurigheid): het deel van de observaties dat correct voorspeld wordt.
- Sensitivity (gevoeligheid): het deel van de observaties dat voor de desbetreffende klasse correct was voorspeld. Dit kan per klasse of gemiddeld over alle klassen.
- Precision (precisie): het deel van de voorspellingen dat voor de desbetreffende klasse correct geassocieerd is. Dit kan per klasse of gemiddeld over alle klassen.



- F1-score. De F1-score meet de prestaties van een model als het harmonische gemiddelde tussen precision en sensitivity (Provost & Fawcett, 2013). Het resultaat is een waarde tussen nul en één, waarbij een waarde van één perfecte precisie en gevoeligheid impliceert.

Naast bovenstaande prestatie maatstaven kunnen we kijken naar de ROC curve (zie Figuur 7 voor een voorbeeld). Dit is een grafische weergave van de modelprestatie die gemaakt wordt door de true positive rate en de false positive rate te plotten (Provost en Fawcett, 2013). Hierin staat de diagonale grijze lijn voor de verwachting bij willekeurige classificatie. Hoe groter de afstand tussen de curve en de diagonaal richting de linker boven hoek, oftewel hoe verder de curve van de diagonaal af staat, hoe beter de modelprestatie. Bij de best presterende modellen zal de curve dus door het gebied linksboven in de grafiek gaan. Een veelgebruikte maatstaf om de modelprestatie te beoordelen is de "Area Under the Curve" (AUC), het gebied onder de ROC curve (in Figuur 7 het oranje gekleurde gebied). De AUC ligt altijd tussen de nul en één (Schwarz et al., 2017), waar een AUC van 1 een perfecte voorspelling betreft.



Figuur 7. ROC curve en Area Under the Curve (AUC)

4.3.3 Modelselectie

Zoals in paragraaf 4.3.1. besproken is, bestaan er verschillende classificatie technieken voor het voorspellen van een categorale variabele. In dit onderzoek proberen we de kans op (georganiseerde) criminaliteit en ondermijning op het bedrijventerrein (laag, gemiddeld of hoog) te voorspellen op basis van een aantal indicatoren. We lichten hieronder toe welke overwegingen een rol hebben gespeeld bij de modelselectie. Allereerst verschillen de methoden in interpreteerbaarheid en voorspellingskracht. Sommige modeltypen zijn over het algemeen beter in het voorspellen van de correcte uitkomst, maar zijn meer een 'black box' als het gaat over hoe het model tot deze classificatie is gekomen (bijv. gradient boosting). Andere modeltypen zijn minder goed in het voorspellen, maar geven wel meer inzicht in de invloed van de verschillende indicatoren bij het bepalen van de uitkomst (bijv. een beslisboom). Daarom komt het soms voor dat in een onderzoek meerdere modeltypen gecombineerd worden. Meestal worden er in data science onderzoeken meerdere algoritmen getest en worden de prestatie statistieken (zie §4.3.2) met elkaar vergeleken. Als er een goed beeld is van de voorspellingskracht van de verschillende modellen, kan dit worden meegewogen in de uiteindelijke modelselectie. Een belangrijk punt bij het selecteren van een model is de prestatie van het model bij nieuwe data, dus data die niet gebruikt zijn om het model te trainen (zie Appendix F.2 voor een uitgebreide toelichting). In Tabel 4 geven we de prestatie statistieken van de geteste modellen weer.



Model	Accuracy (k-fold=10)	Sensitivity (micro)	Precision (micro)	F1 score (micro)	AUC (micro)
Decision Tree	0.53	0.67	0.67	0.67	0.75
Random Forest	0.67	0.89	0.89	0.89	0.92
Gradient Boosting	0.65	0.67	0.67	0.67	0.80

Tabel 4. Prestatiestatistieken van drie getrainde classificatiemodellen

Kijkend naar de prestatie-statistieken, scoort de methode Random Forest het beste. Deze uitkomst wordt ook bevestigd door de ROC-curves van alle classificatiemodellen (zie Appendix F.2). Daarnaast zijn er nog een aantal andere sterke punten van deze classificatiemethode. De techniek kan namelijk goed omweg met grote hoeveelheden variabelen.⁴¹ Een ander voordeel is dat de methode op een natuurlijk manier omgaat met een combinatie van numerieke en categorale data. De methode is bovendien ongevoelig voor de aanwezigheid van ruis, ontbrekende waarden en outliers (Slof, 2014). Het is een erg geschikte methode als er in de dataset sprake is van multicollineariteit en interacties tussen predictoren. In tegenstelling tot lineaire modellen, kan een RF de niet-lineaire relaties in de data goed omschrijven, doordat vele beslissingsbomen worden gecombineerd. Ten slotte kan RF goed overweg met ongebalanceerde data, waarbij gebeurtenissen onevenredig verdeeld zijn. Op basis van deze overwegingen, en gebaseerd op meerdere tests voor een geschikte modelselectie en modelprestatie is in dit onderzoek gekozen voor RF.

De accuracy van het RF model met de 15 meest belangrijke indicatoren is 67%. Dit betekent dat in twee derde van de gevallen de risicoklassen (laag – gemiddeld – hoog) juist worden voorspeld. Bij een derde ligt de voorspelling een klasse ernaast (hoog versus gemiddeld of gemiddeld versus laag). In geen van de gevallen zitten de misclassificaties er totaal naast (hoog versus laag). De algemene voorspellingskracht van het geschatte Random Forest model is met een Area Under the Curve (AUC) van 0.92 bijzonder goed. Dit betekent dat we met de opgenomen indicatoren/potentiële risicofactoren de expert opinion classificatie (lage, gemiddelde of hoge kans op georganiseerde criminaliteit en ondermijning) goed kunnen voorspellen, ook voor verschillende drempelinstellingen (zie Appendix F).

⁴¹ In dit onderzoek hebben we (na data cleaning, preparatie en variabele selectie op basis van o.a. correlaties) 59 variabelen meegenomen in het RF algoritme.



Samenvatting 'Modelselectie'

Er bestaan verschillende classificatie technieken voor het voorspellen van een categorale variabele (in dit onderzoek: de kans op (georganiseerde) criminaliteit en ondermijning op het bedrijventerrein). Een eenvoudig interpreteerbare en veelgebruikte classificatiemethode is de beslissingsboom, waarbij telkens wordt gekeken welke variabele de training observaties het beste zo heterogeen mogelijk kan opsplitsen in twee subgroepen, net zo lang totdat aan een bepaald stopcriterium is voldaan. Random forest (RF) is een 'tree based supervised learning' techniek, waarbij een groot aantal beslissingsbomen gecombineerd worden om tot de uiteindelijke voorspelling te komen. Elke boom wordt getraind op een willekeurige selectie van variabelen en observaties en de uitkomst van de meeste bomen geldt als einduitkomst ('majority voting'). Na het overwegen van verschillende modeltypen, presteerde de random forest het beste volgens de verschillende prestatiestatistieken die in deze paragraaf zijn toegelicht. Daarnaast heeft RF een aantal andere voordelen. Niet-lineaire relaties in de data kunnen bijvoorbeeld goed omschreven worden en RF kan goed omgaan met ongebalanceerde data. Uiteindelijk is in dit onderzoek de keuze gemaakt om RF te gebruiken.

4.4 Resultaten

In eerste instantie is Random Forest toegepast op alle variabelen (features) in de gekoppelde dataset met alle potentiële indicatoren. Zoals toegelicht in §4.1 hebben we gebruik gemaakt van bekende informatie (labels) over de risicoclassificatie van alle bedrijventerreinen in de gemeente Tilburg. Met behulp van deze labels is de doelvariabele, de kans op (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein, geoperationaliseerd en kunnen de voorspellende waarden van potentiële indicatoren in kaart gebracht worden. De RF levert een ranglijst van de belangrijkheid van alle variabelen op. Hieruit zijn de top 15 meest belangrijke indicatoren geselecteerd om het RF algoritme nog een keer hierop te trainen. Dit maakt de uitslag van het model robuuster, omdat de coëfficiënten ten opzichte van elkaar geoptimaliseerd worden.⁴²

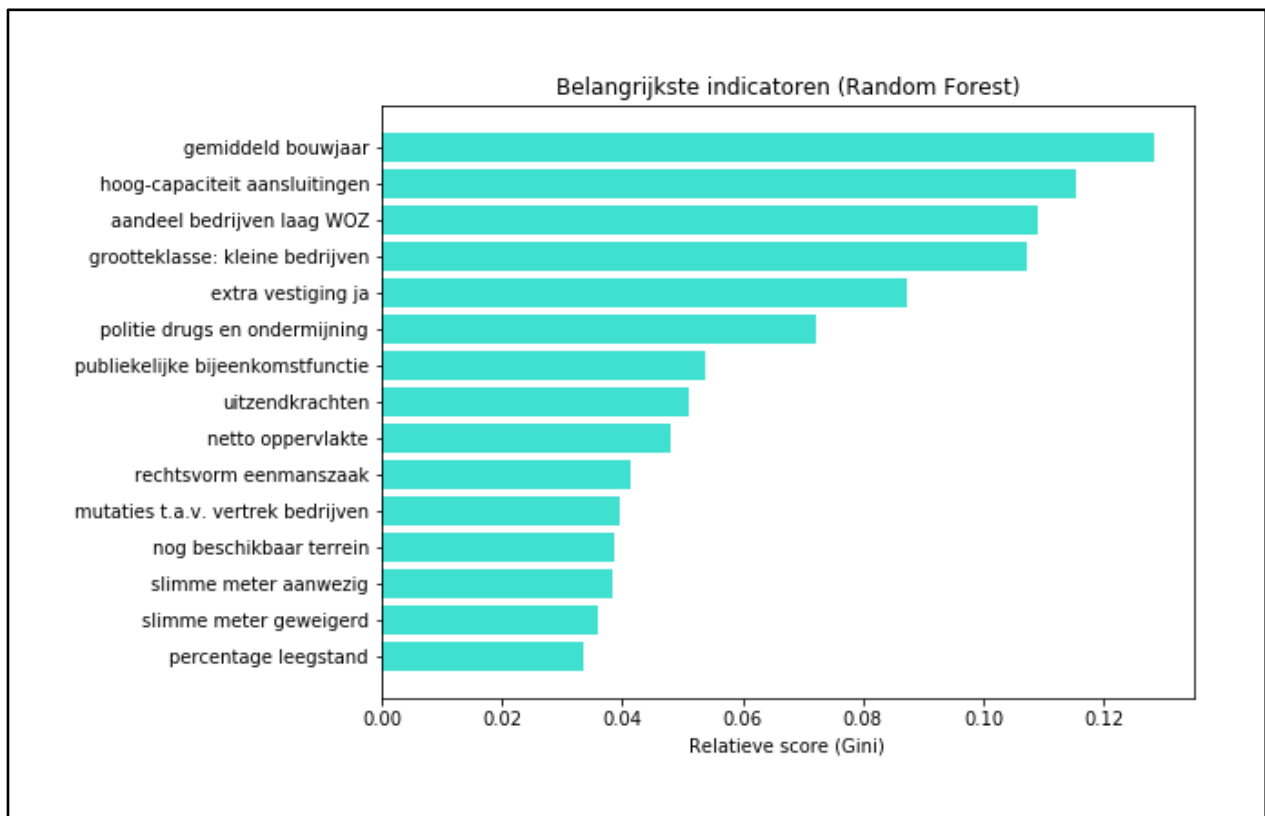
In Figuur 8 zijn de top 15 belangrijkste indicatoren te zien. De lijst is aflopend gesorteerd op de Gini-coëfficiënt, die vaak gebruikt wordt om de variabelen aan te duiden die het belangrijkste zijn en het grootste verschil maken voor de voorspelling.⁴³ Anders gezegd, dit zijn de variabelen die de bedrijventerreinen het beste kunnen scheiden en een indicator met een hoge Gini-coëfficiënt scheidt in dit geval dus de bedrijventerreinen met een hoge kans op georganiseerde criminaliteit en ondermijning van die met een gemiddelde of lage kans etc. Deze ranglijst geeft dus inzicht in de meest belangrijke voorspellers van georganiseerde criminaliteit en ondermijning op bedrijventerreinen en we zien in de top 15 dat als we de supervised machine learning methode Random Forest toepassen de 'usual suspects' naar boven komen.

⁴² In eerste instantie is geprobeerd om de branches mee te nemen in de analyses. Echter was de SBI code te specifiek om te gebruiken in de analyse van de kleine dataset met 30 terreinen, en bij aggregatie naar een hoger niveau (namelijk SBI sector) kwam het niet naar voren als belangrijke indicator.

⁴³ Deze meet de ongelijkheid tussen waarden van een frequentieverdeling (in dit onderzoek dus de classificatie van laag, gemiddeld en hoog). Een Gini-coëfficiënt van nul geeft de perfecte gelijkheid weer, waarbij alle waarden hetzelfde zijn (bijvoorbeeld waar allemaal gemiddeld geclassificeerd zijn). Een Gini-coëfficiënt van 1 (of 100%) drukt de maximale ongelijkheid tussen waarden uit.



Uit onze resultaten blijkt dat de sterkste risicofactor voor de bedrijventerreinen in Tilburg, de variabele die het beste onderscheid maakt tussen bedrijventerreinen met een hoge en lage kans op georganiseerde criminaliteit en ondermijning, het gemiddelde bouwjaar van de panden op een bedrijventerrein is. De top 15 omvat naast het gemiddelde bouwjaar verder vijf andere risicofactoren die beduidend sterker naar voren komen ten opzichte van de andere indicatoren (Gini-coëfficiënt groter dan het gemiddelde van de top 15).



Figuur 8. Belangrijkste indicatoren (Random Forest)

Om zicht te krijgen op de richting van de invloed van een bepaalde variabele hebben we aanvullend op de RF beslissingsbomen gedraaid op de top 15 risicofactoren (voor één zo'n voorbeeld zie Figuur 9). Deze informatie is namelijk niet direct uit een RF te halen. De beslissingsbomen geven ons tevens meer inzicht in de belangrijkste combinaties van risicofactoren. Deze "giftige cocktail"⁴⁴ op een bedrijventerrein is een mix uit de volgende indicatoren:

1. **Gemiddeld bouwjaar:** hoe lager het gemiddelde bouwjaar op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Dit betekent dat er op bedrijventerreinen met gemiddeld gezien oudere panden een hoger risico is.
2. **Hoog-capaciteit aansluitingen:** hierbij kijken we naar het percentage van de panden op het bedrijventerrein waarbij de capaciteit van de energie-aansluiting groter is dan 35 groepen (3x50, 3x63, 3x80). De capaciteit van de energie-aansluiting bepaalt hoeveel gas of stroom er door de aansluiting kan, oftewel de zwaarte van de aansluiting. Het lijkt erop dat als er minder hoog-capaciteit aansluitingen op een

⁴⁴ Met de giftige cocktail wordt de combinatie van indicatoren bedoeld die gezamenlijk lijden tot een hogere kans op georganiseerde criminaliteit en ondermijning.



bedrijventerrein zijn, de kans op georganiseerde criminaliteit en ondermijning toeneemt. Mogelijk bestaat er een relatie tussen de capaciteit van de elektriciteitsaansluiting en de grootte van een bedrijf of het soort bedrijf.

3. **Aandeel panden laag WOZ:** hoe hoger het percentage van de panden op een bedrijventerrein met een WOZ waarde lager dan €160k, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
4. **Grootteklasse (kleine bedrijven):** hoe hoger het percentage bedrijven met minder dan vijf werknemers, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
5. **Extra vestiging ja:** hoe meer extra vestigingen aanwezig zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Extra vestigingen zijn vestigingen die niet in het bedrijvenregister worden opgenomen, omdat ze niet voldoen aan de voorwaarden (er is niet minstens 1 persoon werkzaam, bijv. extra KvK inschrijvingen, beheer-bv's en opslagloodsen).
6. **Politie drugs en ondermijning:** hoe meer incidenten gerelateerd aan drugs en ondermijning er zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.

Andere risicofactoren die in combinatie met de top risicofactoren als belangrijk naar voren komen zijn:

- **publiekelijke bijeenkomstfuncties:** de kans op georganiseerde criminaliteit en ondermijning is hoger als er minder panden zijn met een publieke bijeenkomstfunctie. Dit zou te maken kunnen hebben met het feit dat als er veel panden met publiekelijke bijeenkomstfuncties zijn (winkels en/of sport en onderwijs gerelateerd), er meer 'ogen' zijn, waardoor er meer sociale controle op het bedrijventerrein is en afscherming van criminele activiteiten lastiger is.
- **gemiddeld aantal uitzendkrachten:** hoe lager het gemiddelde aantal uitzendkrachten, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Het gemiddeld aantal uitzendkrachten zou samen kunnen hangen met de grootte van een bedrijf en het soort bedrijf.
- **netto oppervlakte van een bedrijventerrein:** hoe groter de netto oppervlakte, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Een mogelijke verklaring hiervoor zou kunnen zijn dat de sociale cohesie achteruit gaat en controles op grote terreinen moeilijker zijn.
- **eenmanszaken:** de kans op georganiseerde criminaliteit en ondermijning wordt groter naarmate er meer eenmanszaken aanwezig zijn. Dit is een welbekende indicator (zie bv. Nieuwenhuis en Pannekoek, 2012) die nu ook bevestigd wordt door machine learning.
- **vertrekkende bedrijven:** hoe meer bedrijven er vertrokken zijn, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Hierbij kijken we naar het aantal bedrijven dat tussen 2016 en 2018 verhuisd is naar een andere locatie.
- **nog beschikbaar terrein:** hierover kon geen eenduidige conclusie getrokken worden op basis van verdere dataexploratie.
- **aanwezigheid van een slimme meter:** hierover kon geen eenduidige conclusie getrokken worden op basis van verdere dataexploratie.
- **het weigeren van een slimme meter:** hoe vaker een slimme meter geweigerd is, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Het zou zo kunnen zijn dat in sommige gevallen het weigeren van een slimme, op afstand uitleesbare, meter gerelateerd is aan het willen verbergen van bepaalde activiteiten.

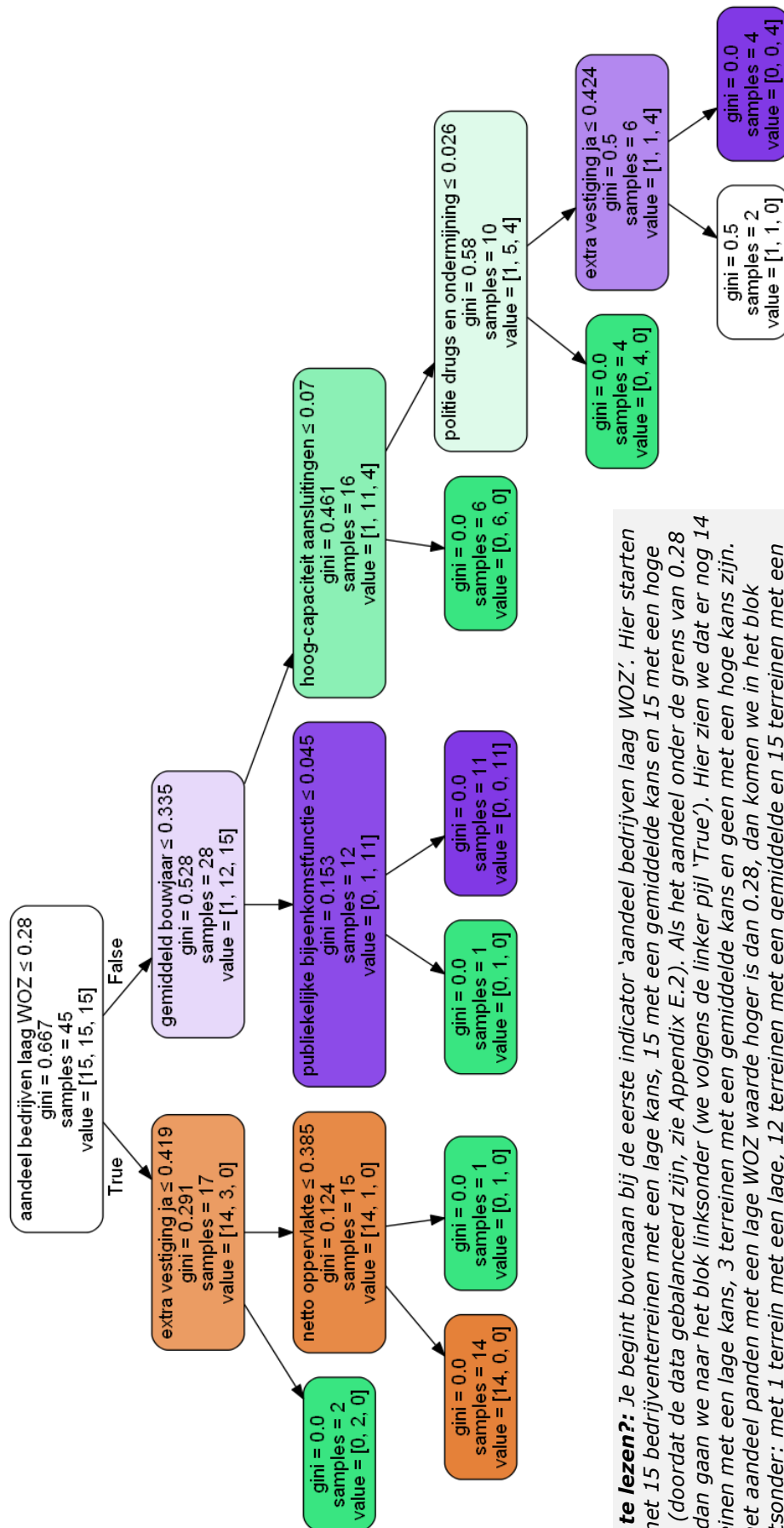


- **percentage leegstand:** hoe minder leegstand, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Dit is enigszins verrassend, de verwachting zou zijn dat meer leegstand juist leidt tot een hoger risico.

Naast het kijken naar de belangrijkste indicatoren, geeft het kijken naar de indicatoren die niet als belangrijk naar voren komen ook extra inzichten. Op basis van onze analyses lijkt het er bijvoorbeeld op dat het aantal hennepuimingen in het verleden geen onderscheidende factoren zijn bij het categoriseren van bedrijventerreinen op basis van de kans op georganiseerde criminaliteit en ondermijning. Ook de aanwezigheid van parkmanagement en het bezit van het Keurmerk Veilig Ondernemen komen niet naar voren als onderscheidende factoren, wat mogelijk verklaard zou kunnen worden door de beperkte spreiding in deze variabelen. Daarnaast zijn de drugsgerelateerde meldingen bij het Centraal Meldpunt van de gemeente geen indicator, maar de meldingen bij de politie wel. De ontsluiting van een bedrijventerrein kwam tevens niet naar voren als indicator. Hierbij moeten we wel weer de kanttekening plaatsen dat de resultaten minder betrouwbaar zijn door de zeer kleine steekproefomvang van dertig bedrijventerreinen en dat we dus niet met zekerheid kunnen stellen dat bovengenoemde variabelen geen rol spelen in het voorspellen van georganiseerde criminaliteit en ondermijning op bedrijventerreinen. Het uitvoeren van vervolgonderzoek op grotere schaal is dan ook van groot belang om de resultaten uit dit onderzoek te valideren (zie ook §5.2).

Samenvatting 'Resultaten'

Om de belangrijkste indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen te vinden is eerst een Random Forest algoritme gedraaid, waarna we op basis van de 15 belangrijkste indicatoren nog een beslissingsboom hebben gedraaid om de resultaten beter te kunnen duiden. Uit de analyses blijkt dat een aantal indicatoren voorspellende waarde hebben voor de mate van aanwezigheid van (georganiseerde) criminaliteit en ondermijning. In de ranglijst van de 15 meest belangrijke indicatoren staan met name de 'usual suspects' die ook in het vooronderzoek als belangrijk naar voren kwamen. Zo lijkt een combinatie van factoren die betrekking hebben op de fysieke omgeving zoals veroudering en leegstand, de waarde van het onroerend goed, de mate van toezicht (bv. door panden met een publieksfunctie), inclusief het aantal meldingen bij politie en het verloop op een bedrijventerrein belangrijk te zijn. Ook het soort bedrijven lijkt een rol te spelen, bijvoorbeeld de rechtsvorm en grootte. Tot slot kwam de aanwezigheid van veel extra vestigingen, zoals extra KvK inschrijvingen, beheer-bv's en opslagloodsen, naar voren als risico indicator. Samenvattend kan dus geconcludeerd worden dat in deze verkennende data science analyse op basis van een supervised machine learning model bekende inzichten grotendeels bevestigd lijken te worden, evenals de hoofdlijnen van de in hoofdstuk 2 geschetste criminologische theorieën. Echter zijn de resultaten van dit onderzoek minder betrouwbaar door de zeer kleine steekproefomvang van dertig bedrijventerreinen, dus vervolgonderzoek op grotere schaal is nodig om met zekerheid te kunnen zeggen welke indicatoren wel of niet bepalend zijn bij het voorspellen van de kans op georganiseerde criminaliteit en ondermijning op bedrijventerreinen.



Hoe te lezen?: Je begint bovenaan bij de eerste indicator 'aandeel bedrijven laag WOZ'. Hier starten we met 15 bedrijventerreinen met een lage kans, 15 met een gemiddelde kans en 15 met een hoge kans (doordat de data gebalanceerd zijn, zie Appendix E.2). Als het aandeel onder de grens van 0.28 ligt, dan gaan we naar het blok links onder (we volgen de linker pijl 'True'). Hier zien we dat er nog 14 terreinen met een lage kans, 3 terreinen met een gemiddelde kans en geen met een hoge kans zijn. Als het aandeel panden met een lage WOZ waarde hoger is dan 0.28, dan komen we in het blok rechts onder: met 1 terrein met een lage, 12 terreinen met een gemiddelde en 15 terreinen met een hoge kans. Vervolgens wordt weer gekeken op basis van welke variabele de data verder opgesplitst kunnen worden. N.B.: de grenswaarden zijn soms lastig te interpreteren door de databewerkingen. Echter kijken we voor een indicatie van de relatie tussen de variabele en de kans op georganiseerde criminaliteit en ondermijning vooral naar de richting (hoger of lager), gezien het feit dat de exacte resultaten niet betrouwbaar zijn door de kleine steekproef van 30 bedrijventerreinen.

Figuur 9. Een beslissingsboom uitgevoerd op resultaten van de RF ter illustratie van de evaluatie van de meeste belangrijke indicatoren



5 Slotbeschouwing en mogelijkheden voor toekomstig onderzoek

Dit laatste hoofdstuk bevat de slotbeschouwing (§5.1), waarna er aanbevelingen voor toekomstig onderzoek worden gegeven (§5.2).

5.1 Slotbeschouwing

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

De eerste onderzoeksvraag was: *'Wat zijn potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?'*. Het onderzoek is dan ook gestart met het identificeren van potentiële indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Dit hebben we gedaan in een vooronderzoek waarin we door middel van een literatuuronderzoek, een expertmeeting en interviews hebben gekeken welke kennis en informatie al beschikbaar was. Ten eerste zijn mogelijke indicatoren geïdentificeerd met betrekking tot de fysieke omgeving (waaronder onderhoud, veroudering en leegstand) en toezicht, handhaving en samenhang op het bedrijventerrein. Andere mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen die in het vooronderzoek naar voren zijn gekomen hebben betrekking op de ligging van het bedrijventerrein, het soort bedrijven (branches) die er aanwezig zijn, het verloop, onroerend goed, meldingen bij de politie en Meld Misdaad Anoniem, meldingen bij de gemeente en overige informatie die bij de gemeente bekend is bijvoorbeeld over incidenten uit het verleden.

Na het identificeren van mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen, zijn data verzameld uit verscheidene bronnen. De tweede onderzoeksvraag was namelijk: *'Welke databronnen zijn beschikbaar en kunnen worden gekoppeld om de voorspellende waarde van potentiële indicatoren te onderzoeken?'*. In dit verkennende onderzoek is ervoor gekozen om de dataverzameling geografisch te beperken tot 30 bedrijventerreinen in de gemeente Tilburg. Kenmerken over bedrijventerreinen (zoals de aanwezigheid van parkmanagement en de ontsluiting van het terrein) waren beschikbaar in het IBIS. Daarnaast hebben we het aantal meldingen bij de politie en Meld Misdaad Anoniem ontvangen. Of een bedrijventerrein een Keurmerk Veilig Ondernemen heeft was online in te zien. De gemeente Tilburg heeft het aantal meldingen bij het Centraal Meldpunt (bijv. over drugsoverlast), het aantal hennepuimingen, het aantal Damocles sluitingen en data met betrekking tot de onroerende zaken op de bedrijventerreinen geleverd. Daarnaast hebben ze data uit de BAG en het bedrijvenregister (waaraan de KvK ten grondslag ligt) ter beschikking gesteld. Tot slot heeft netbeheerder Enexis data geleverd over de aanwezigheid en het weigeren van slimme meters, de capaciteit van elektriciteitsaansluitingen en historische



gevallen van fraude. Resultaat was een uitgebreide dataset met mogelijke indicatoren van georganiseerde criminaliteit en ondermijning op bedrijventerreinen.

De volgende stap was het operationaliseren van de mate van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein als afhankelijke variabele. De derde onderzoeksvraag was namelijk: *'Hoe kan de mate van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein worden geoperationaliseerd als afhankelijke variabele die kan worden gebruikt om de voorspellende waarde van potentiële indicatoren te toetsen?'*. Twee manieren om per bedrijventerrein een inschatting van de kans op georganiseerde criminaliteit en ondermijning te verkrijgen zijn tijdens het onderzoek in overweging genomen. Uiteindelijk is er in dit onderzoek voor gekozen om de bedrijventerreinen te 'labelen' (in lage, gemiddelde, of hoge kans) door aan meerdere mensen met verschillende achtergronden en vanuit verschillende organisaties te vragen deze categorisatie te maken in een kleine vragenlijst en vervolgens alle resultaten te combineren tot een eindindeling. Hieruit kwam dat acht bedrijventerreinen een lage kans, vijftien terreinen een gemiddelde kans en zeven terreinen een hoge kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning hebben. Beperkingen van deze aanpak zijn dat er maar elf respondenten zijn en dat er mogelijk sprake is van subjectiviteit, een gebrek aan kennis en/of eigen belangen. We hebben geprobeerd dit te verminderen door zoveel mogelijk respondenten te benaderen van zoveel mogelijk verschillende organisaties en met verschillende rollen en expertisegebieden. Ook waren de resultaten uit de vragenlijst grotendeels in lijn met signalen die naar voren zijn gekomen in de integrale overheidssamenwerking.

De vierde onderzoeksvraag was: *'Welke potentiële indicatoren hebben daadwerkelijk voorspellende waarde voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen?'*. Om hier achter te komen moesten we eerst een analysemethode kiezen. Om de belangrijkste indicatoren te vinden hebben we gebruik gemaakt van de data science methode Random Forest (RF). Zoals in §4.4 is te zien staan in de ranglijst van de 15 meest belangrijke indicatoren met name de 'usual suspects' die ook in het vooronderzoek als belangrijk naar voren kwamen. Zo lijkt een combinatie van factoren die betrekking hebben op de **fysieke omgeving** (§2.4) zoals veroudering en leegstand, de waarde van het **onroerend goed** (§2.5), de mate van **toezicht** (§2.6) (door panden met een publieksfunctie of de aanwezigheid van veel uitzendkrachten) inclusief het **aantal meldingen bij de politie** (§2.7) en het **verloop** op een bedrijventerrein (§2.4) belangrijk te zijn. Ook het **soort bedrijven** (§2.3) lijkt een rol te spelen, bijvoorbeeld de rechtsvorm en grootte. Tot slot kwam de aanwezigheid van veel extra vestigingen, zoals extra KvK inschrijvingen, beheer-bv's en opslagloodsen, naar voren als risico indicator. Samenvattend kan dus geconcludeerd worden dat in deze verkennende data science analyse op basis van een supervised machine learning model bekende inzichten grotendeels bevestigd lijken te worden, evenals de hoofdlijnen van de in hoofdstuk 2 geschetste criminologische theorieën.

Om zicht te krijgen op de richting van de invloed van een bepaalde variabele hebben we aanvullend op de RF beslissingsbomen gedraaid op de top 15 risicofactoren. Deze informatie is namelijk niet direct uit een RF te halen. De beslissingsbomen geven ons tevens meer inzicht in de belangrijkste combinaties van risicofactoren. Deze "giftige cocktail"⁴⁵ op een bedrijventerrein is een mix uit de volgende indicatoren:

⁴⁵ Met de giftige cocktail wordt de combinatie van indicatoren bedoeld die gezamenlijk lijden tot een hogere kans op georganiseerde criminaliteit en ondermijning.



1. **Gemiddeld bouwjaar:** hoe lager het gemiddelde bouwjaar op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Dit betekent dat er op bedrijventerreinen met gemiddeld gezien oudere panden een hoger risico is.
2. **Hoog-capaciteit aansluitingen:** hierbij kijken we naar het percentage van de panden op het bedrijventerrein waarbij de capaciteit van de energie-aansluiting groter is dan 35 groepen (3x50, 3x63, 3x80). Het lijkt erop dat als er minder hoog-capaciteit aansluitingen op een bedrijventerrein zijn, de kans op georganiseerde criminaliteit en ondermijning toeneemt. Mogelijk bestaat er een relatie tussen de capaciteit van de elektriciteitsaansluiting en de grootte van een bedrijf of het soort bedrijf.
3. **Aandeel panden laag WOZ:** hoe hoger het percentage van de panden op een bedrijventerrein met een WOZ waarde lager dan €160k, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
4. **Grootteklasse (kleine bedrijven):** hoe hoger het percentage bedrijven met minder dan vijf werknemers, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.
5. **Extra vestiging ja:** hoe meer extra vestigingen aanwezig zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning. Extra vestigingen zijn vestigingen die niet in het bedrijvenregister worden opgenomen, omdat ze niet voldoen aan de voorwaarden (er is niet minstens 1 persoon werkzaam, bijv. extra KvK inschrijvingen, beheer-bv's en opslagloodsen).
6. **Politie drugs en ondermijning:** hoe meer incidenten gerelateerd aan drugs en ondermijning er zijn op een bedrijventerrein, hoe hoger de kans op georganiseerde criminaliteit en ondermijning.

Hoewel de resultaten van deze verkennende pilot minder betrouwbaar zijn door de beperkte steekproef van 30 bedrijventerreinen in de gemeente Tilburg, laten we hiermee wel zien dat het technisch gezien mogelijk is om dergelijke analyses uit te voeren en kunnen we voorzichtig afleiden uit de resultaten dat het toepassen van data science methoden kan helpen bij het vinden van indicatoren voor (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Vervolgonderzoek op grotere schaal is nodig om dit te bevestigen.

De vijfde onderzoeksvraag luidde: *'Hoe kunnen de indicatoren die daadwerkelijk voorspellende waarde hebben worden gebruikt om in de praktijk inzichtelijk te maken welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning?'.* De indicatoren die een belangrijke voorspellende waarde hebben zouden in de toekomst gebruikt kunnen worden om in de praktijk inzichtelijk te maken welke bedrijventerreinen een verhoogde kans hebben op (georganiseerde) criminaliteit en ondermijning. Hiervoor is het echter wel noodzakelijk om eerst op grotere schaal, dus voor meer bedrijventerreinen, onderzoek te verrichten, zodat een betrouwbaar model geschat kan worden. Vervolgens zou ook voor andere bedrijventerreinen, waarvoor nog geen label bekend is, op basis van de indicatoren voorspeld kunnen worden of het terrein een lage, gemiddelde of hoge kans heeft op (georganiseerde) criminaliteit en ondermijning. Daarnaast zou het mogelijk zijn om te kijken hoe veranderingen in de waardes van de voorspellers de inschatting van de kans op (georganiseerde) criminaliteit en ondermijning veranderen. Hierdoor is het ook mogelijk om verschillende scenario's te draaien, wat kan helpen bij het maken van beleid.

De meerwaarde van data science methoden kan in de toekomst dus nog meer bewezen worden door analyses uit te voeren op grotere datasets. Echter zullen er ook beperkingen zijn in het gebruik van data science methoden, die met name betrekking hebben op de beschikbaarheid



en verkrijgbaarheid van data en het koppelen van datasets (die bijv. verschillende aggregatieniveaus hebben). Daarnaast is het creëren van een geïntegreerde dataset met vele mogelijke indicatoren niet alleen een tijdrovende klus, ook komen er vele juridische aspecten (zoals privacy overweging) bij kijken.

Naast het uitvoeren van onderzoek op grotere schaal en het ontwikkelen van een betrouwbaar model, is een andere mogelijkheid om onderzoeken te doen op kleinere schaal. Gemeentes en regio's die een inventarisatie van hun bedrijventerreinen wensen kunnen ook, net als in dit onderzoek, een aantal experts vragen om de bedrijventerreinen te beoordelen in een vragenlijst en vervolgens zou dit aangevuld kunnen worden met kwalitatief onderzoek om inzichtelijk te maken welke bedrijventerreinen in de gevarezone zitten.

De laatste onderzoeksvraag was: *'In hoeverre is het haalbaar om in de toekomst dezelfde methodiek als in dit onderzoek toe te passen om indicatoren van (georganiseerde) criminaliteit en/of ondermijning in woonwijken te vinden?'*. In de toekomst zou dezelfde methodiek als in dit onderzoek ook toegepast kunnen worden om breder te kijken naar indicatoren van (georganiseerde) criminaliteit en ondermijning, bijvoorbeeld in woonwijken. Dit is niet specifiek onderzocht, maar zou wel mogelijk moeten zijn met enkele aanpassingen, voornamelijk in de gebruikte databronnen. Daarnaast zullen opnieuw verschillende methoden getest moeten worden om te bepalen welke methode het meest geschikt is voor deze nieuwe toepassing. Het zou tevens interessant zijn om de relatie tussen bedrijventerreinen en de omliggende wijken te bekijken. Bedrijventerreinen die dicht bij bewoonde gebieden liggen of daaraan grenzen, hebben een relatie met die woonwijken. Zo vonden Tops en Van der Torre (2015; p. 41) in een van de door hen onderzochte cases dat *"...op (nabijgelegen) bedrijventerreinen frequent drugsvangsten worden gedaan waar geregeld op de één of andere manier wijkbewoners bij zijn betrokken"*. Volgens Tops en Van der Torre (2015; p. 58) wordt crimineel geld deels op grote afstand, in andere delen van de wereld, geïnvesteerd, maar investeren criminelen lokaal nog steeds in bedrijven en panden. Door het lectoraat Digitalisering en Veiligheid en het lectoraat Ondermijning van Avans hogeschool wordt op dit moment een onderzoek uitgevoerd naar het in beeld brengen van risicowijken met behulp van data-analyse in de gemeente Tilburg. Het zou interessant zijn om de resultaten van het onderzoek van Avans naast de resultaten van dit onderzoek te leggen om te kijken of de bedrijventerreinen met een hoge kans op criminaliteit en ondermijning ook omringd worden door risicowijken.

5.2 Mogelijkheden voor toekomstig onderzoek

Gedurende dit onderzoek zijn veel ideeën ontstaan voor toekomstig onderzoek, die we hieronder zullen toelichten. Een eerste mogelijkheid is schaalvergroting. Na deze verkennende pilot voor de gemeente Tilburg is een logische vervolgstap om de schaal te vergroten en onderzoek te doen voor een groter gebied, bijvoorbeeld voor heel Nederland. Ook zou het interessant kunnen zijn om andere soorten bedrijventerreinen mee te nemen. Daarnaast zouden er indicatoren toegevoegd kunnen worden, zoals de afstand tot de grens met het buitenland en de afstand tot vliegvelden. Er moet rekening gehouden worden met het feit dat niet alle databronnen/variabelen die gebruikt zijn in dit onderzoek ook beschikbaar zijn op grotere schaal. Zoals eerder vermeld, biedt onderzoek op grotere schaal veel meer mogelijkheden. Zo kan er een betrouwbaar model geschat worden, waardoor het draaien van



verschillende scenario's/simulaties een mogelijkheid wordt, wat kan helpen bij het maken van beleid.

Een tweede mogelijkheid voor toekomstig onderzoek is het toevoegen van additionele databronnen die voor dit onderzoek niet verkrijgbaar waren of om een andere reden niet meegenomen konden worden in dit onderzoek. Een inventarisatie, waarbij de bedrijventerreinen in de gemeente Tilburg geanalyseerd worden op basis van andere databronnen (bijvoorbeeld data van de Belastingdienst) en de resultaten vervolgens naast de resultaten van dit onderzoek gelegd kunnen worden, zou tevens erg interessant kunnen zijn. Daarnaast zijn de analyses binnen dit onderzoek op bedrijventerreinniveau gedaan, maar er zouden ook nog analyses gedaan kunnen worden op bedrijf- of adresniveau en/of zelfs op persoonsniveau.



Literatuurlijst

- Beekmans, J.J.H. (2015). Verouderingsprocessen op bedrijventerreinen. In: E. van Krabben, C.-J. Pen & F. de Feijter (red.), *De Markt voor Bedrijventerreinen: Uitkomsten van Onderzoek en Beleid*, pp. 57-69. Den Haag: Platform31.
- Chawla, N.V., K.W. Bowyer, L.O. Hall & W.P. Kegelmeyer (2002). Smote: synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, 16, pp. 321–357.
- Clarke, R. (2009). Situational crime prevention: Theoretical background and current practice. In M.D. Krohn, A. Lizotte & G. Penly Hall (red.), *Handbook on Crime and Deviance*. Handbooks of sociology and social research (pp. 259-276). New York: Springer.
- Cohen, L. & M. Felson. (1979). Social change and crime rate trends: A routine activity's approach. *American Sociological Review*, 44, pp. 588-608.
- Emke, E., D. Vughs, A. Kolkman & P. de Voogt (2018). Wastewater-based epidemiology generated forensic information: Amphetamine synthesis waste and its impact on a small sewage treatment plant. *Forensic Science International*, 286, e1-e7.
- Felson, M. (1998). *Crime and Everyday Life* (2e editie). Thousand Oaks, CA: Pine Forge Press/Sage.
- Felson, M. & R. Clarke. (1998). Opportunity makes the thief: Practical theory for crime prevention. In B. Webb (red.), *Police Research Series, paper 98*. London: Home Office.
- Ferwerda, H., T. van Ham, L. Scholten & D. Jager (2016), *Focus op heling – Een onderzoek naar het functioneren van de helingmarkt, het beleid tegen en de gevolgen van heling*. Den Haag: WODC/Beke reeks.
- James, G., D. Witten, T. Hastie & R. Tibshirani (2013). *An Introduction to Statistical Learning* (112). New York: Springer.
- Jeffery, C.R. (1971). *Crime Prevention Through Environmental Design*. Beverly Hills, CA: Sage.
- Kalidien, S. (2017). Criminaliteit en rechtshandhaving 2016. Ontwikkelingen en samenhangen. Den Haag: WODC.
- Keizer, K., S. Lindenberg & L. Steg (2008). The spreading of disorder. *Science*. 332, p. 1691.
- Kelling, G.L. & Wilson, J.Q. (1982). Broken windows: The police and neighborhood safety. *The Atlantic Monthly*, 29-38.
- Kleemans, E.R., M.E.I. Brien en & H.G. van de Bunt (2002). *Georganiseerde criminaliteit in Nederland. Tweede rapportage op basis van de WODC-monitor*. Den Haag: WODC/Boom Juridische Uitgevers.
- Kolthoff, E. (2016). *Basisboek criminologie*, 2e druk, Den Haag: Boom Criminologie.
- Kolthoff E. & S. Khonraad (2016). Ondermijnende aspecten van georganiseerde criminaliteit en de rol van de bovenwereld. *Tijdschrift voor Criminologie*, 58(2), pp. 76-90.



- Kruisbergen, E., H.G. van de Bunt & E. Kleemans (2012). *Georganiseerde criminaliteit in Nederland: Vierde rapportage op basis van de Monitor Georganiseerde Criminaliteit*. Den Haag: Boom Lemma.
- Kube, E. (1982). *Städtebau, Wohnhausarchitektur und Kriminalität: Prevention statt Reaction*. Heidelberg: Kriminalistik-verlag.
- Lam, J., R. van der Wal & N. Kop. (2018). *Sluipend gif. Een onderzoek naar ondermijnende criminaliteit*. Den Haag: Boom Criminologie.
- Luten, I. (red.) (2008). *Handboek Veilig Ontwerp en Beheer*. Bussum: Thoth.
- Meijer, A. (2015). *Bestuur in de datapolis. Slimme stad, blijde burger?* Oratie Universiteit Utrecht. Den Haag: Boom Bestuurskunde.
- Nelen, H. & E. Kolthoff (2017). *Schaduwen over de rechtshandhaving. Georganiseerde criminaliteit en integriteitsschendingen van functionarissen in de rechtshandhaving*. Den Haag: Boom Criminologie.
- Nieuwenhuis, F. & M. Pannekoek (2012). *Werkwijzen van georganiseerde criminaliteit*. Verslag van een onderzoek voor het National dreigingsbeeld 2012. Dienst IPOL, Zoetermeer.
- Pest, M. te, S. Nieuwdorp, B. Smeets & E. van Wijnen (2012). *Verwevenheid van onder- en bovenwereld bij georganiseerde criminaliteit*. Een overzichtsstudie: aard en oorzaken. Regionaal Informatie en Expertise Centrum (RIEC) Zuid-West Nederland.
- Provost, F. & T. Fawcett (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*, O'Reilly Media, Inc.
- Schwartz, I. M., P. York, E. Nowakowski-Sims, & A. Ramos-Hernandez (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The broward county experience, *Children and Youth Services Review*, 81, pp. 309–320.
- Slof, D. (2014). *Voorspellen van webwinkel aankopen met een Random Forest*. Erasmus Universiteit Rotterdam, Econometrie en Operationele Research.
- Tops, P. & E. van der Torre (2015). *Wijkenaanpak en ondermijnende criminaliteit*. Den Haag: Boom Lemma.
- Voordt, D.J.M. van der & H.B.R. van Wegen (1990). *Sociaal veilig ontwerpen: Checklist ten behoeve van het ontwikkelen en toetsen van (plannen voor) de gebouwde omgeving*. Delft: Technische Universiteit Delft, Publikatieburo Bouwkunde.
- Walks, R. A. & R. Maaranen (2008). Gentrification, social mix, and social polarization: Testing the linkages in large Canadian cities. *Urban Geography*, 29(4), pp. 293-326.
- Wiele, D. van der (2016). Normloosheid op het bedrijventerrein. *Secondant*, 2 mei 2016. Geraadpleegd via: <https://www.ccv-secondant.nl/platform/article/normloosheid-op-het-bedrijventerrein/> op 10 april 2018.



A Experts vooronderzoek

De onderzoekers hebben tijdens het vooronderzoek dankbaar gebruik gemaakt van de expertise en de medewerking van personen werkzaam bij onderstaande organisaties/initiatieven:

Expert meeting:

- Belastingdienst – externe samenwerking & fraude
- Bureau voor Toegepast Veiligheidsonderzoek
- Centrum voor Criminaliteitspreventie en veiligheid (CCV)
- City Deal Zicht op Ondernijning
- DCMR Milieudienst Rijnmond - handhaving
- Experts in Media en Maatschappij (EMMA) – onderzoeker/adviseur
- Gemeente Tilburg – onderzoeker veiligheid
- Regionaal Informatie en Expertise Centrum (RIEC) Zeeland - West-Brabant
- Partners + Pröpper - onderzoeker/adviseur
- Programma Georganiseerde Ondernijnende Criminaliteit
- Solaris Parkmanagement B.V. - parkmanager
- Tilburg University - Hoogleraar Tilburg Law School

Interviews:

- Belastingdienst – externe samenwerking & fraude
- DCMR Milieudienst Rijnmond - handhaving
- Gemeente Rotterdam – bestrijding criminaliteit bedrijventerrein
- Gemeente Tilburg – onderzoeker veiligheid
- Regionaal Informatie en Expertise Centrum (RIEC) Zeeland - West-Brabant
- VNO-NCW / MKB Nederland

Tevens bedanken we alle personen en organisaties die ons te woord hebben gestaan, ons hebben ontvangen of langs zijn gekomen op ons kantoor en/of data ter beschikking hebben gesteld voor dit onderzoek.



B Lijst met variabelen

Variable	Bron	Toelichting
ondernemers of eigenarenvereniging - ja/nee	Gemeente Tilburg	Wel of geen aanwezigheid van een ondernemers- of eigenaren vereniging op het bedrijventerrein
KVO certificaat basis samenwerken - ja/nee	Gemeente Tilburg	Wel of geen beschikbaarheid van een KVO certificaat op het bedrijventerrein
parkmanagement volgens ibis - ja/nee	Gemeente Tilburg	Wel of geen parkmanagement aanwezig op bedrijventerrein
damocles sluitingen aantal totaal 2014tm24062018	Gemeente Tilburg	Het aantal sluitingen volgens het handhavingsbeleid Damocles
bedrijvenregister aantal vestigingen april2018	Gemeente Tilburg	Totaal aantal vestigingen in april 2018.
mutaties t.a.v. oprichting bedrijven	Gemeente Tilburg	Verandering in het bedrijvenregister tussen 2016 en 2018 door oprichting van bedrijven (nieuw opgericht bedrijf vestigt zich op het terrein)
mutaties t.a.v. vestiging bedrijven	Gemeente Tilburg	Verandering in het bedrijvenregister tussen 2016 en 2018 door vestiging van bedrijven (bestaand bedrijf vestigt zich op het terrein)
mutaties t.a.v. opheffing bedrijven	Gemeente Tilburg	Verandering in het bedrijvenregister tussen 2016 en 2018 door opheffing van bedrijven (bedrijf stopt te bestaan)
mutaties t.a.v. vertrek bedrijven	Gemeente Tilburg	Verandering in het bedrijvenregister tussen 2016 en 2018 door vertrek van bedrijven (bedrijf blijft bestaan, maar vertrekt naar een andere locatie)
hennepuimingen 2016plus2017	Gemeente Tilburg	Het aantal hennepuimingen in 2016 en 2017
bedrijvenregister vestigingen 2018min2016	Gemeente Tilburg	Het netto groei van het aantal vestigingen tussen 2016 en 2018
bedrijvenregister aantal vestigingen april2018 genormaliseerd	Gemeente Tilburg	Het aantal vestigingen genormaliseerd door het netto oppervlak van het bedrijventerrein
total nr adressen met melding 2016tm2018	CMP meldingen	Het totaal aantal adressen dat een melding heeft gekregen
cmp alle drugs gerelateerd	CMP meldingen	Alle drugs gerelateerde meldingen
netto oppervlakte	IBIS bedrijventerrein Tilburg	Netto oppervlakte (in ha) van de werklocatie. De som van alle uitgeefbare of in erfpacht uitgeefbare kavels in hectares, zowel niet-uitgegeven als uitgegeven kavels
nog beschikbaar terrein	IBIS bedrijventerrein Tilburg	Nog beschikbaar aantal hectaren dat uitgegeven kan worden
percentageleegstand	IBIS bedrijventerrein Tilburg	Een inschatting van het aandeel (in %) van de panden dat structureel (langer dan 3 jaar) leeg staat
minimalemilieucategorie	IBIS bedrijventerrein Tilburg	De minimaal toegestane milieucategorie volgens de indeling in 'Bedrijven en Milieuzonering' (editie 2009)
maximalemilieucategorie	IBIS bedrijventerrein Tilburg	De maximaal toegestane milieucategorie volgens de indeling in 'Bedrijven en Milieuzonering' (editie 2009)



netto verouderd oppervlak	IBIS bedrijventerrein Tilburg	Het netto oppervlakte (in ha) dat verouderd is. Er is sprake van veroudering als de kwaliteit van het terrein niet langer voldoet aan de vestigingseisen van bedrijven
bruto-netto	IBIS bedrijventerrein Tilburg	Bruto minus netto oppervlakte (in ha)
verouderd bruto-netto	IBIS bedrijventerrein Tilburg	Verouderde bruto minus netto (in ha)
structureleleegstand ja	IBIS bedrijventerrein Tilburg	Structurele leegstand. Hier aangeven of er op het terrein bedrijfspanden aanwezig zijn die al langer dan drie jaar leeg staan
schaalniveau r	IBIS bedrijventerrein Tilburg	Hier aangeven of het terrein bedoeld is voor de opvang van bovenlokale of regionale bedrijven (r) of dat het vooral bedoeld is voor de opvang van lokale bedrijven
vol nee	IBIS bedrijventerrein Tilburg	Categorische variabele: bedrijventerrein is nog niet helemaal uitgegeven
mutaties ja	IBIS bedrijventerrein Tilburg	Aanduiding van een wijzigingen in de oppervlakte van een bedrijventerreinen
waterbeheer nee	IBIS bedrijventerrein Tilburg	Geen waterbeheer aanwezig. Bij waterbeheer gaat het om het beperken van wateroverlast bij hoosbuien, door te zorgen voor meer opvangcapaciteit en betere infiltratie van water
breeam nee	IBIS bedrijventerrein Tilburg	Er is geen BREEAM-NL gebiedsontwikkeling certificaat
bereikbaarheidextern goed	IBIS bedrijventerrein Tilburg	Er zijn geen files in de omgeving en de afstand tot de snelweg is minder dan 10 kilometer
parkeerdrukopenbaar nee	IBIS bedrijventerrein Tilburg	Er is altijd voldoende parkeergelegenheid op of vlak bij het eigen terrein van bedrijven. Bezoekers/werknemers kunnen op de locatie parkeren
opzet g	IBIS bedrijventerrein Tilburg	Het terrein is grootschalig van opzet. Dit wil zeggen dat er ook kavels groter dan 5000 m2 zijn
opzet gk	IBIS bedrijventerrein Tilburg	Het terrein is gedeeltelijk grootschalig en gedeeltelijk kleinschalig
opzet k	IBIS bedrijventerrein Tilburg	Het terrein is kleinschalig van opzet. Dit wil zeggen dat de kavels kleiner dan 5000 m2 zijn
kwaliteitsterrein f	IBIS bedrijventerrein Tilburg	De kwaliteit van het terrein is functioneel. Er zijn zeer brede wegen, ingericht op vrachtwagens, geen trottoir/fietspad. Er zijn opstelplaatsen voor vrachtauto's in openbare ruimte en personenauto's kunnen op eigen terrein. Er is 10-15% groen op het terrein, in smalle stroken
functiemenging gedeeltelijk	IBIS bedrijventerrein Tilburg	Functiemenging met wonen is gedeeltelijk toegestaan op het terrein
functiemenging ja	IBIS bedrijventerrein Tilburg	Er is functiemenging met wonen toegestaan op het terrein
ontsluitingweg l	IBIS bedrijventerrein Tilburg	De inrit van het terrein komt uit op een lokale weg
ontsluitingweg p	IBIS bedrijventerrein Tilburg	De inrit van het terrein komt uit op een provinciale weg



ontsluitingweg r	IBIS bedrijventerrein Tilburg	De inrit van het terrein grenst direct aan een afslag van een rijksweg
ontsluitingspoor a	IBIS bedrijventerrein Tilburg	De werklocatie wordt niet ontsloten door een spoor
ontsluitingwater e	IBIS bedrijventerrein Tilburg	De werklocatie wordt ontsloten door een spoor met aftakking, publiek gebruik
ontsluitingov n	IBIS bedrijventerrein Tilburg	De werklocatie wordt niet ontsloten via het OV
glasvezel nee	IBIS bedrijventerrein Tilburg	Er is geen supersnelle internetverbinding beschikbaar
herstructureringsplan n	IBIS bedrijventerrein Tilburg	Er is geen herstructureringsplan
verouderd gedeeltelijk	IBIS bedrijventerrein Tilburg	Een gedeelte van het terrein is verouderd. Op deze plekken voldoet het terrein niet aan de vestigingseisen van bedrijven
verouderd ja	IBIS bedrijventerrein Tilburg	Het bedrijventerrein is verouderd
energie nee	IBIS bedrijventerrein Tilburg	Er wordt geen duurzame energie opgewekt op het bedrijventerrein
politie drugs gerelateerd	Politie	Aantal incidenten die te maken hebben met drugs
politie ondermijning gerelateerd	Politie	Aantal incidenten die gerelateerd zijn aan ondermijning
politie rest	Politie	Aantal incidenten die niet onder drugs of ondermijning vallen
regulier bedrijventerrein	BAG	Het gebruiksdoel van het pand
nevenadres	BAG	Adressoort wel of geen nevenadres (t.o.v. hoofadres)
gemiddeld bouwjaar	BAG	Het gemiddelde bouwjaar van alle panden op het bedrijventerrein
publiekelijke bijeenkomstfunctie	BAG	Het aandeel (in %) van terrein met publieke gebruiksdoelen (winkel, sport of onderwijs)
median bouwjaar	BAG	Het mediaan bouwjaar van alle panden op het bedrijventerrein
oudste bouwjaar	BAG	Het oudste bouwjaar van alle panden van het bedrijventerrein. Het jaar waarin het eerste pand gebouwd werd
nieuwste bouwjaar	BAG	Het nieuwste bouwjaar van alle panden van het bedrijventerrein. Het jaar waarin het laatste pand gebouwd werd
percentage pand als opslag	Gemeente OZ	Het aandeel (in %) van het bedrijventerrein dat als opslag gebruikt wordt
woz waarde median	Gemeente OZ	Het mediaan van de WOZ-waarde van alle panden op het bedrijventerrein
aandeel panden laag WOZ	Gemeente OZ	Het aandeel (in %) van de panden met een lage WOZ-waarde (< 160k €).
aandeel panden gemiddeld WOZ	Gemeente OZ	Het aandeel (in %) van de panden met een gemiddelde WOZ-waarde (160k-400k €).



aandeel panden hoog WOZ	Gemeente OZ	Het aandeel (in %) van de panden met een hoge WOZ-waarde (> 400k €).
opgeheven sinds 2016	Bedrijvenregister	Aantal bedrijven opgeheven sinds 2016
vaste pers	Bedrijvenregister	Aantal personen vast in dienst, fulltime en parttime
uitzendkrachten	Bedrijvenregister	Gemiddeld aantal uitzendkrachten per bedrijventerrein (fulltime en parttime)
totaal pers	Bedrijvenregister	Totale aantal personen in dienst
extra vestiging ja	Bedrijvenregister	Extra vestigingen zijn vestigingen die niet in het bedrijvenregister worden opgenomen, omdat ze niet voldoen aan de voorwaarden (er is niet minstens 1 persoon werkzaam, bijv. extra KvK inschrijvingen, beheer-bv's en opslagloodsen).
reden einde verandering eigenaar	Bedrijvenregister	Percentage van de bedrijven die niet meer bestaat vanwege de verandering van eigenaar
reden einde faillissement	Bedrijvenregister	Percentage van de bedrijven dat niet meer bestaat vanwege faillissement
rechtsvorm eenmanszaak	Bedrijvenregister	Percentage van bedrijven met als rechtsvorm een eenmanszaak
rechtsvorm bv	Bedrijvenregister	Percentage van bedrijven met als rechtsvorm een bv
rechtsvorm stichting	Bedrijvenregister	Percentage van bedrijven met als rechtsvorm een stichting
grootteklasse kleine bedrijven	Bedrijvenregister	Het aandeel (in %) van bedrijven met minder dan 5 werkzame personen (minimaal 1)
grootteklasse grote bedrijven	Bedrijvenregister	Het aandeel (in %) van bedrijven met 5 of meer werkzame personen
SBI sector	Bedrijvenregister	Het aandeel (in %) bedrijven met SBI sector X (waarbij er meerdere sectoren zijn en er voor elke sector een aparte variabele is gemaakt).
aantal aansluitingen	Enexis	Aantal energie-aansluitingen op het bedrijventerrein v.a. 2015
fraudegevallen	Enexis	Aantal bekende fraudegevallen op het bedrijventerrein v.a. 2015
capaciteit onbekend	Enexis	Het aandeel (in %) van panden van het bedrijventerrein waar de capaciteit van de energie-aansluiting onbekend is
slimme meter aanwezig	Enexis	Het aandeel (in %) van panden van het bedrijventerrein met een slimme meter
slimme meter geweigerd	Enexis	Het aandeel (in %) van panden van het bedrijventerrein waar het automatisch uitlezen van de slimme meter is geweigerd
Enexis dagen leegstand	Enexis	Aantal dagen van leegstand vanaf 2015 volgens Enexis.
Enexis laag capaciteit	Enexis	Het aandeel (in %) van panden van het bedrijventerrein met capaciteit van energie-aansluiting kleiner dan 35 groepen
hoog-capaciteit aansluitingen	Enexis	Het aandeel (in %) panden van het bedrijventerrein met capaciteit van energie-aansluiting groter dan 35 groepen (3x50, 3x63, 3x80)



C Vragenlijst labelen bedrijventerreinen

In opdracht van het WODC voert onderzoeksinstituut CentERdata een onderzoek uit naar indicatoren van (georganiseerde) criminaliteit en ondermijning op bedrijventerreinen. Binnen dit onderzoek is de **gemeente Tilburg uitgekozen voor een pilot**.

Om voor dit onderzoek een beter beeld te krijgen van de mate van aanwezigheid van georganiseerde misdaad en (drugsgelateerde) criminaliteit op de bedrijventerreinen in de gemeente Tilburg hebben we **uw hulp nodig**. Zou u per bedrijventerrein een inschatting hiervan kunnen maken? **Het kost u slechts 2 minuten om dit te doen**. Hiermee helpt u ons ontzettend in het onderzoek. Uw antwoorden zullen gebruikt worden om aanbevelingen voor beleid te kunnen formuleren.

Probeer eerlijk te zijn bij het inschatten van de kans op aanwezigheid van (georganiseerde) criminaliteit en ondermijning op een bedrijventerrein. De gegevens zullen vertrouwelijk worden behandeld en individuele antwoorden zullen niet worden gerapporteerd. Er bestaan geen goede of foute antwoorden. In dit onderzoek gaat het om uw perceptie.

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

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.

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 rechts bovenin klikken.

v1_1 <<Naam bedrijventerrein>>

v1_2 <<Naam bedrijventerrein>>

...

...

Categories:

1. Laag
2. Gemiddeld
3. Hoog
4. Weet niet

Hebt u nog aanvullende opmerkingen (bv. over specifieke bedrijventerreinen)?

1. Ja
2. Nee

U kunt uw opmerking hieronder kwijt.



C.1 Resultaten vragenlijst labels bedrijventerreinen

In onderstaande frequentietabel is aangegeven hoe vaak het label laag/gemiddeld/hoog per bedrijventerrein gekozen is. De categorie die het meest gekozen is, is geel gemarkeerd. Door middel van majority voting is de eindscore bepaald. Voor vier bedrijventerreinen bestond er twijfel. Hierover zijn de volgende beslissingen genomen:

- Bedrijventerrein 6: Twijfel over 'laag' of 'gemiddeld'. Aangezien er geen stemmen zijn voor 'hoog', en de antwoorden dus 'skewed to the left' zijn, is ervoor gekozen om dit terrein in de categorie 'laag' te plaatsen. Dit maakt de data tevens meer gebalanceerd voor de analyses.
- Bedrijventerrein 16: Zowel 'laag' als 'gemiddeld' hebben drie stemmen gekregen, maar er is ook een keer voor 'hoog' gekozen. Daarom is dit terrein in de categorie gemiddeld geplaatst.
- Bedrijventerrein 17: 'laag', 'gemiddeld' en 'hoog' hebben elk een stem gekregen en er is veel 'weet niet' geantwoord. Dit terrein is in de categorie 'gemiddeld' geplaatst.
- Bedrijventerrein 26: Twijfel over 'gemiddeld' of 'hoog'. Aangezien er geen stemmen zijn voor 'laag', en de antwoorden dus 'skewed to the right' zijn, is ervoor gekozen om dit terrein in de categorie 'hoog' te plaatsen. Dit maakt de data tevens meer gebalanceerd voor de analyses.

IBISWERKLOCATIE	Laag	Gemiddeld	Hoog	Weet niet	Conclusie	Beslissing
Bedrijventerrein 1	4	3	1	3	Laag	Laag
Bedrijventerrein 2	2	1	0	8	Laag	Laag
Bedrijventerrein 3	4	1	0	6	Laag	Laag
Bedrijventerrein 4	4	3	1	3	Laag	Laag
Bedrijventerrein 5	6	0	1	4	Laag	Laag
Bedrijventerrein 6	2	2	0	7	Twijfel	Laag
Bedrijventerrein 7	4	1	1	5	Laag	Laag
Bedrijventerrein 8	6	0	1	4	Laag	Laag
Bedrijventerrein 9	0	3	1	7	Gemiddeld	Gemiddeld
Bedrijventerrein 10	0	3	1	7	Gemiddeld	Gemiddeld
Bedrijventerrein 11	0	3	1	7	Gemiddeld	Gemiddeld
Bedrijventerrein 12	1	5	3	2	Gemiddeld	Gemiddeld
Bedrijventerrein 13	0	3	1	7	Gemiddeld	Gemiddeld
Bedrijventerrein 14	1	3	2	5	Gemiddeld	Gemiddeld
Bedrijventerrein 15	0	5	3	3	Gemiddeld	Gemiddeld
Bedrijventerrein 16	3	3	1	4	Twijfel	Gemiddeld
Bedrijventerrein 17	1	1	1	8	Twijfel	Gemiddeld
Bedrijventerrein 18	1	6	0	4	Gemiddeld	Gemiddeld
Bedrijventerrein 19	1	2	1	7	Gemiddeld	Gemiddeld
Bedrijventerrein 20	1	6	1	3	Gemiddeld	Gemiddeld
Bedrijventerrein 21	3	4	1	3	Gemiddeld	Gemiddeld
Bedrijventerrein 22	2	5	1	3	Gemiddeld	Gemiddeld
Bedrijventerrein 23	0	3	0	8	Gemiddeld	Gemiddeld
Bedrijventerrein 24	0	1	6	4	Hoog	Hoog
Bedrijventerrein 25	1	2	4	4	Hoog	Hoog



Bedrijventerrein 26	0	2	2	7	Twijfel	Hoog
Bedrijventerrein 27	1	2	3	5	Hoog	Hoog
Bedrijventerrein 28	0	3	6	2	Hoog	Hoog
Bedrijventerrein 29	1	2	4	4	Hoog	Hoog
Bedrijventerrein 30	2	0	3	6	Hoog	Hoog



D Datapreparatie

Een aantal databronnen is aangeleverd op het niveau van bedrijventerrein dat wil zeggen er was 1 observatie per bedrijventerrein. Er waren echter ook databronnen die op een ander aggregatieniveau zijn aangeleverd (bijvoorbeeld postcode-6 niveau) en die dus door de onderzoekers geaggregeerd moesten worden tot bedrijventerreinniveau. Hieronder wordt beschreven hoe hiermee is omgegaan.

Politie

Het feit dat niet alle incidenten geregistreerd worden met een straatnaam en huisnummer was een uitdaging bij het verwerken van de data van de politie. Het komt regelmatig voor dat alleen de straatnaam geregistreerd wordt (als een incident op straat plaatsvindt). Dit maakt de koppeling met een bedrijventerrein lastiger. Sommige straten vallen namelijk gedeeltelijk binnen een bedrijventerrein en gedeeltelijk erbuiten. Het is dus niet met zekerheid vast te stellen of een incident heeft plaatsgevonden op het bedrijventerrein of daarbuiten. Dit is met name problematisch voor grote straten, zoals de Ringbaan Zuid in Tilburg. Deze meldingen zijn aangeduid als twijfelgevallen in de data. Niet alleen de meldingen die niet bevestigd konden worden op precieze locatie, maar ook onduidelijkheden over het type misdaad zijn gemarkeerd als twijfelgeval. Het percentage aan twijfelgevallen bedraagt 54% (2685 van de 5877). Alle twijfelgevallen zijn uit de data gehaald en dus niet gebruikt.

Omdat politie data relatief gezien schaars zijn, hebben we alle data van de jaren 2014 tot en met 2018 samengevoegd. Vervolgens hebben we de data opgesplitst in twee onderdelen: (1), drugs en ondermijning gerelateerde misdaad (de focus van dit onderzoek) en (2) alle andere soorten misdaad (zoals bv. diefstal of ruzie). De data is vervolgens geaggregeerd op bedrijventerreinniveau waardoor er gekoppeld kan worden met alle andere data op hetzelfde aggregatieniveau. De resulterende twee datasets betreft het aantal plegingen van misdaad per bedrijventerrein sinds 2014. Om niet afhankelijk van de grootte van het bedrijventerrein te zijn (immers grote bedrijventerreinen hebben een grotere kans op misdaad), hebben we het aantal misdaad plegingen gedeeld door de netto oppervlakte van het bedrijventerrein.

Meld Misdaad Anoniem

Ook hier was het een uitdaging om de (postcode-6 niveau) data te aggregeren tot bedrijventerreinniveau. Sommige postcodes kwamen namelijk op meerdere terreinen voor (terreinen die naast elkaar liggen). Per postcodegebied hebben we gekeken naar het aantal panden per bedrijventerrein. Dit doen we op basis van het BAG ID, een uniek referentienummer voor een object in de BAG. Als een postcode bij meerdere terreinen hoort, wijzen we de data toe aan het terrein waarop de meeste panden staan met deze postcode. Hiermee creëren we unieke postcodes die aan één bedrijventerrein gebonden zijn. Dit heeft de voorkeur boven een splitsing naar rato, omdat je dan te maken krijgt met incidenten die onsplitsbaar zijn (bv. er is één incident op een postcodegebied).

Ook hier zijn alle data van de jaren 2014 tot en met 2018 geaggregeerd en alle drugs en ondermijning gerelateerde misdaad meldingen samengevoegd. De uiteindelijke data is alsnog relatief dun verspreid, maar nu wel bruikbaar. We houden drie variabelen over; meldingen gerelateerd aan drugs en ondermijning, alle overige meldingen, en de som van deze twee (het totaal aantal meldingen). Alle postcodes worden uiteindelijk gegroepeerd op



bedrijventerreinniveau en gekoppeld met overige data. Ook hier worden de anonieme meldingen genormaliseerd met de netto oppervlakte van het bedrijventerrein.

BAG

De BAG data is een grote dataset op niveau van gebouwen (dus postcode 6 + huisnummer) met bijna 4000 BAGID's. Er zijn enkele bruikbare variabelen te halen uit de BAG data, zoals het gebruiksdoel van het pand (binnenwijken of regulier bedrijventerrein), adressoort (hoofd of nevenadres), pand status (in gebruik of niet) en het bouwjaar. We hebben de BAG omschrijving van het bedrijventerrein gekoppeld aan de IBIS werklocatie door handmatig de naamgevingen te controleren. Behalve enkele kleine verschillen in de naam, die we corrigeerden, was er één extra locatie gemarkeerd als bedrijventerrein in de BAG data, het AaBe-complex. Dit is niet het type bedrijventerrein van dit onderzoek, maar in feite een winkelcentrum. Deze locatie hebben we daarom uit de data verwijderd.

Van het bouwjaar per pand hebben we het gemiddelde, jongste en oudste bouwjaar van een bedrijventerrein gecreëerd als bruikbare variabelen. Alle andere variabelen zijn per pand gecategoriseerd door middel van one-hot encoding. Deze zijn: wel of geen nevenadres, wel of geen regulier bedrijventerrein, en wel of geen winkel-, sport- of onderwijsbijeenkomst (dat wil zeggen publiekelijke gebruiksdoelen). Bij het aggregeren zijn de categorie aantallen gesommeerd per bedrijventerrein. Vervolgens zijn deze aantallen getransformeerd naar percentages per bedrijventerrein.

KvK

De data van de Kamer van Koophandel is op niveau van individuele bedrijven. De KvK data is in de eerste fase gecombineerd met het Tilburgs Bedrijvenregister. Van de KvK data worden uiteindelijk geen variabelen in de directe zin gebruikt, maar de variabelen die overgenomen of afgeleid zijn uit de KvK dataset worden wel meegenomen. Deze zitten voornamelijk in de Tilburgs Bedrijventerreinen dataset. De meegenomen variabelen zijn: rechtsvorm van bedrijf, SBI sector code, grootteklasse bedrijf, en datum einde bedrijf. Bedrijven worden op basis van de IBIS werklocatie geaggregeerd naar het niveau van bedrijventerreinen.

Alleen bedrijven die nog bestaan op 1 januari 2016 worden meegenomen in de analyses (alle data met einddatum vóór 2016 zijn verwijderd). Van de rechtsvorm nemen we alleen de bedrijven die een eenmanszaak, besloten vennootschap (BV) of een stichting zijn in beschouwing. Alle overige rechtsvormen (42) zijn buiten beschouwing gelaten. Bedrijven die na 1 januari 2016 een einde hebben, worden geclassificeerd als 'opgeheven bedrijven' en tevens opgesplitst in twee subcategorieën met als reden 1) faillissementen en 2) verandering van eigenaar. Van de SBI sector nemen we alle 19 sectoren in beschouwing door middel van one-hot encoding. De gecodeerde variabelen worden opgeteld per bedrijventerrein en gedeeld door het totaal (van de bedrijventerrein) om een procentueel inzicht te krijgen.

Enexis

Enexis data is gekoppeld op postcode-6 niveau en geaggregeerd tot bedrijventerreinen niveau door de bestaande koppeling tussen postcode 6 en bedrijventerreinen van IBIS. De Enexis data is verzameld vanaf 1 januari 2015 en is bijgehouden tot oktober 2018.

Alle capaciteiten van energie-aansluiting groter dan 35 groepen hebben we beschouwd als hoog-capaciteit aansluitingen (3x50, 3x63 en 3x80) en al het andere als laag-capaciteit aansluitingen.



E Databewerking

E.1 Normalisatie

De meeste variabelen die aantallen weergeven van een bedrijventerrein, bv. aantal werkend personeel, aantal vestigingen, oppervlakte nog uitteefbaar of aantal Centraal Meldpunt (CMP) meldingen zijn genormaliseerd met de netto oppervlakte van elk bedrijventerrein. Dit is om de aantallen onafhankelijk te maken van de grootte van het bedrijventerrein. We zijn tevens consistent gebleven met het normalisatiefactor (de netto oppervlakte). Variabelen die als fractie van een bedrijventerrein gecalculerd zijn, bijvoorbeeld gebouwen onderverdeeld in laag, gemiddeld of hoog WOZ waarde, faillissementen sinds 2016, panden gebruikt als opslag of eenmanszaken zijn natuurlijkerwijs niet genormaliseerd, maar worden procentueel weergegeven.

Daarnaast zijn de variabelen met een MinMaxScaler herschaald. De keuze voor een MinMaxScaler in plaats van andere schalingsmethoden (zoals mean normalization) is empirisch bepaald. We hebben hier niet te maken met uitschieters in de data, dus MinMaxScaler is in dit geval een goede methode.

E.2 Data balancing/Oversampling

De grootste problemen bij dit onderzoek zijn het kleine aantal beschikbare observaties (30 bedrijventerreinen) en de ongebalanceerdheid van de gelabelde data (8 laag, 15 gemiddeld, 7 hoog). Om een gebalanceerde dataset te verkrijgen kunnen oversampling technieken worden toegepast. Wij gebruiken hiervoor de Synthetic Minority Oversampling Technique (SMOTE, Chawla et al., 2002) om onze data te oversamplen. In ons geval oversamplen we de data van 30 naar 45 (naar 15 van elke klasse).

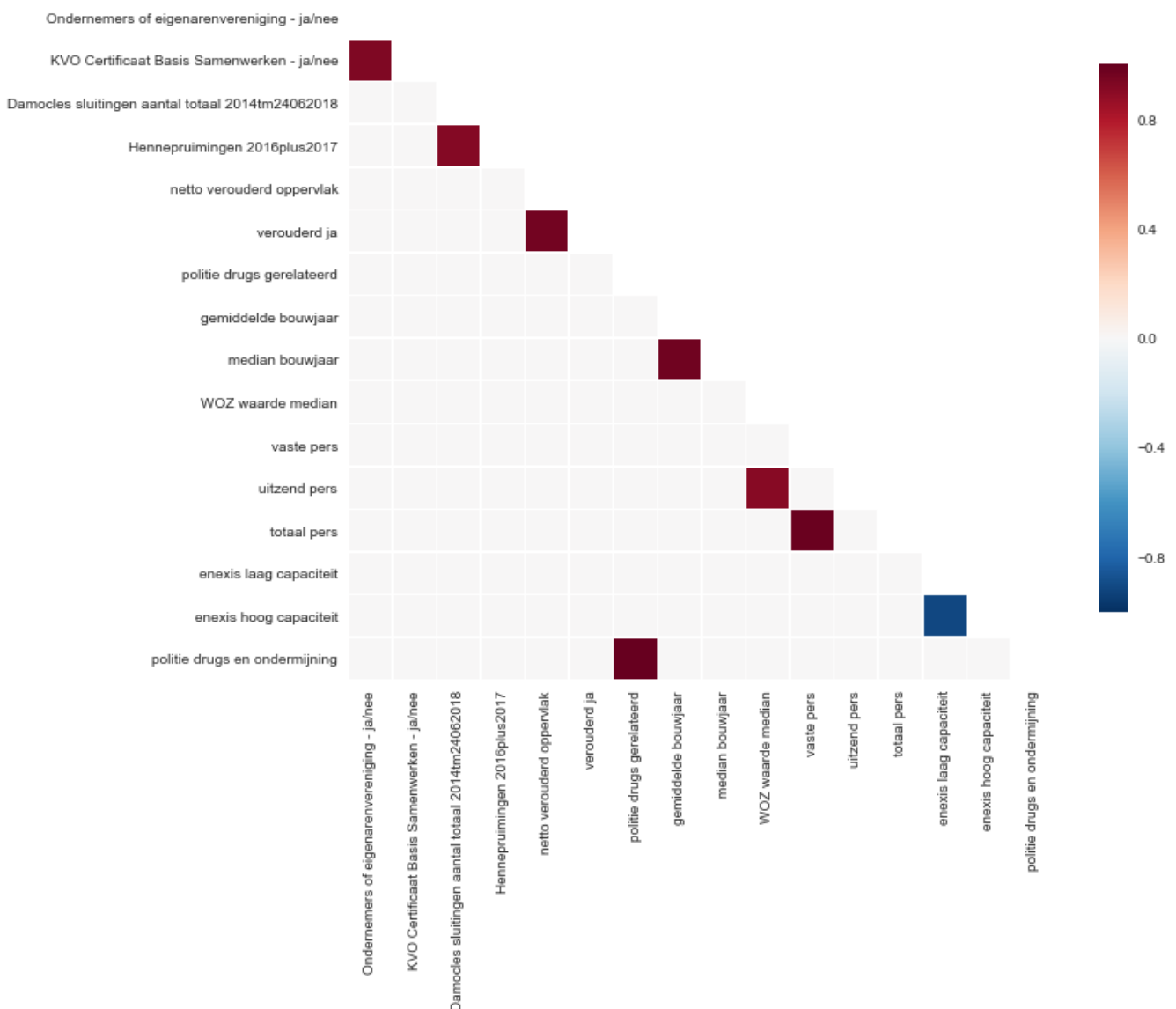
SMOTE

Wanneer de classificatiecategorieën van een dataset niet gelijk zijn gedistribueerd, is een dataset ongebalanceerd. In machine-learning komen ongebalanceerde datasets vaak voor bij voorspellingen van "zeldzame" voorvallen, zoals fraude of ziekten. De prestaties van een voorspellend model, indien gebaseerd op een ongebalanceerde dataset, wordt sterk gereduceerd. SMOTE is een over-sampling benadering waarbij de minderheidsklasse wordt verhoogd door het creëren van "synthetische" punten in de dataset. De minderheidsklasse wordt bemonsterd door middel van het nemen van een steekproef onder elk van de minderheidsklassen en het introduceren van synthetische datapunten langs de lijnsegmenten die aansluiten bij één of alle van de k-nearest neighbours van de minderheidsklassen. De k-nearest neighbours zijn willekeurig gekozen, afhankelijk van het benodigde hoeveelheid oversampling.



E.3 Correlaties

Tijdens het verkennende vooronderzoek (Exploratory Data Analysis, EDA) merkten we al snel dat er sterk gecorreleerde variabelen bestaan. In de meeste gevallen zijn deze eenvoudig te verklaren, bv. WOZ waarde laag en WOZ waarde hoog. We laten de variabelen vallen die een sterke (anti-)correlatie met elkaar vertonen en eenduidig zijn. De enige correlatie die niet simpelweg verklaarbaar is, maar wel een sterke correlatie waarde laat zien, is tussen de variabelen “uitzend personeel” en “WOZ waarde median”. We behouden beide variabelen in de data. Hieronder wordt een plot weergegeven van variabelen die een Pearson correlatie van groter dan 0.9 vertonen, zie Figuur E.1.



Figuur E.1. Correlatiematrix van hoog gecorreleerde variabelen



F Modelselectie en -performance

F.1 Prestatiestatistieken

Om de prestatie van een model te beoordelen en verschillende modellen te vergelijken, kan gebruik gemaakt worden van verschillende statistieken. In een confusion matrix wordt voor een binaire uitkomstvariabele (d.w.z. een uitkomstvariabele met twee categorieën) de voorspelde klasse vergeleken met de werkelijke klasse (zie Figuur F.1.). In dit onderzoek is er echter sprake van een uitkomstvariabele met drie categorieën in plaats van twee, namelijk: een lage, gemiddelde of hoge kans op georganiseerde criminaliteit en ondermijning. Bij de aanwezigheid van meerdere klassen kan er per klasse gekeken worden naar de prestatie. Het is gebruikelijk om vervolgens te kijken naar het gewogen gemiddelde. Het middelen kan op twee manieren gedaan worden: het micro- of macrogemiddelde.⁴⁶ Micro- en macrogemiddelden (voor welke maatstaf dan ook) zullen enigszins verschillende dingen berekenen en hun interpretatie verschilt. Het microgemiddelde wordt doorgaans iets vaker gebruikt, omdat het dichterbij de prestaties van het model voor de grotere klasse ligt. Wij houden daarom in dit onderzoek het microgemiddelde aan.

		Voorspelde klasse	
		0	1
Geobserveerde klasse	0	True Negative (TN)	False Positive (FP)
	1	False Negative (FN)	True Positive (TP)

Figuur F.1. Confusion matrix

Uit de confusion matrix kunnen andere prestatie maatstaven worden afgeleid, namelijk:

- Accuracy (nauwkeurigheid): het deel van de observaties dat correct voorspeld wordt.

$$Accuracy = \frac{\text{aantal correct voorspelde observaties}}{\text{totaal aantal observaties}} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Sensitivity (gevoeligheid): het deel van de observaties dat voor de desbetreffende klasse correct was voorspeld. Dit kan per klasse of gemiddeld over alle klassen. Voor klasse 1 bij een dichotome variabele zou de formule bijvoorbeeld als volgt zijn:

$$Sensitivity = \frac{\text{aantal 'true positives'}}{\text{aantal werkelijke positives}} = \frac{TP}{TP + FN}$$

- Precision (precisie): het deel van de voorspellingen dat voor de desbetreffende klasse correct geassocieerd is. Dit kan per klasse of gemiddeld over alle klassen. Voor klasse 1 bij een dichotome variabele zou de formule bijvoorbeeld als volgt zijn:

$$Precision = \frac{\text{aantal 'true positives'}}{\text{aantal voorspelde positives}} = \frac{TP}{TP + FP}$$

⁴⁶ Een voorbeeld van een micro- vs. macrogemiddelde voor de maatstaf 'precision' is als volgt:
 Macrogemiddelde precision = $0.5 \times (\text{Precision1} + \text{Precision2})$
 Microgemiddelde precision = $TP1 + TP2 / (TP1 + TP2 + FP1 + FP2)$



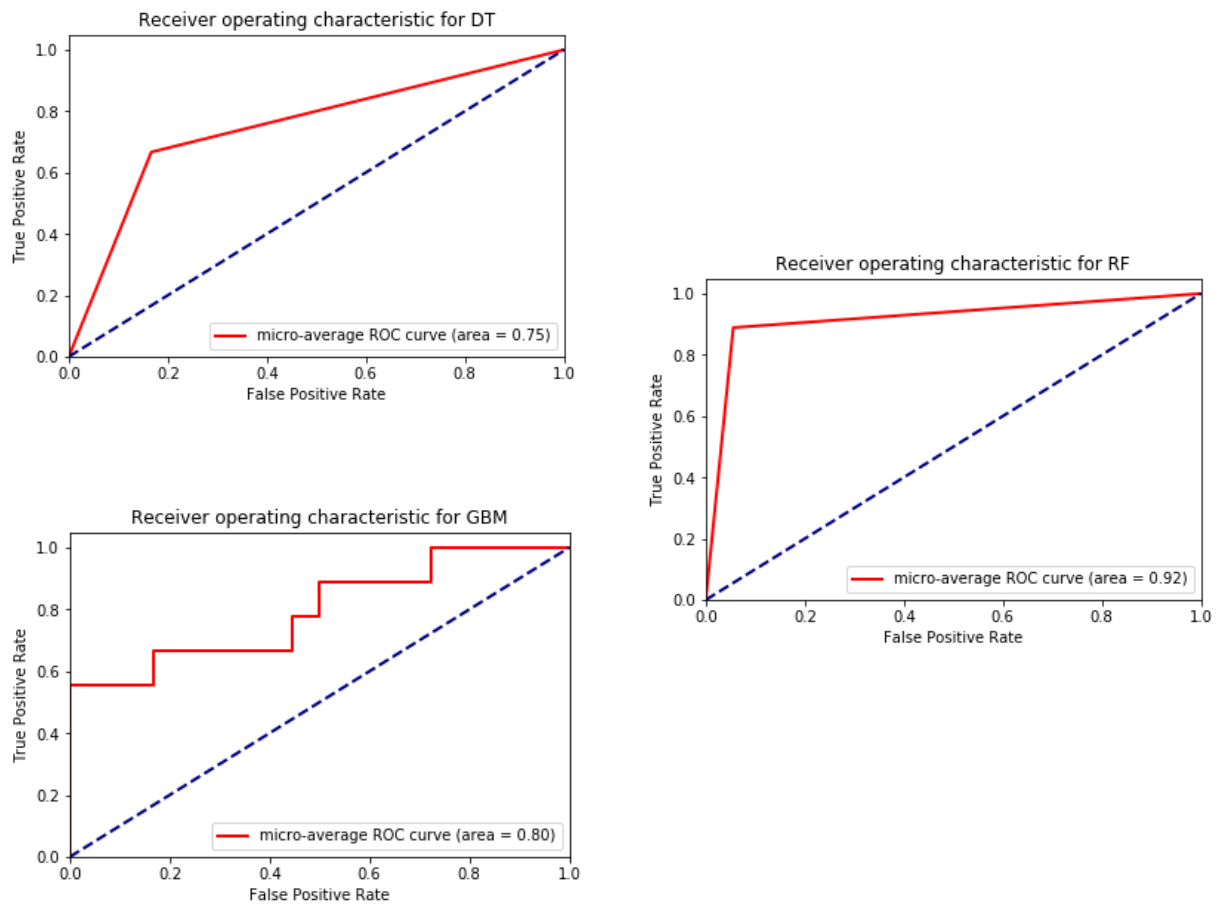
- F1-score. De F1-score meet de prestaties van een model als het harmonische gemiddelde tussen precision en sensitivity (Provost & Fawcett, 2013). Het resultaat is een waarde tussen nul en één, waarbij een waarde van één perfecte precisie en gevoeligheid zou impliceren. De F1-score wordt berekend door de volgende formule:

$$F1 - score = 2 * \frac{precision * sensitivity}{precision + sensitivity}$$

F.2 Testen

Een belangrijk punt bij het selecteren van een model is de prestatie van het model bij nieuwe data, dus data die niet gebruikt is om het model te trainen. Een model kan goed scoren op de getrainde data, maar slecht op alle andere data. Dit heet overfitting. Om dit te voorkomen wordt er doorgaans een (random) deel van de data achter de hand gehouden als een test-set, bijvoorbeeld 20%. Daarnaast kan er nog een deel van de data als validatieset achter de hand worden gehouden om een model te optimaliseren, bijvoorbeeld ook 20%. Het nadeel hiervan is echter dat er te veel achter de hand kan worden gehouden en in geval van kleinere datasets (weinig observaties) de modeltraining er onder lijdt. In zulke gevallen wordt er vaak voor gekozen om de validatie weg te laten of deze tegelijk te doen met de test-set. Wij hebben daarom in onze resultaten gekozen voor een evaluatie met een cross-validatie (k-fold=10) en de prestatiestatistieken berekend aan de hand van de test-set⁴⁷. Deze aanpak is constant gehouden bij alle geteste methoden, zodat we de methoden goed konden vergelijken. In Figuur F.2. zijn de ROC curves van de geteste methoden weergegeven.

⁴⁷ Door de prestatiestatistieken enkel over de test-set te berekenen, dus voor de data waarop het model getraind is, kunnen de statistieken hoger uitvallen dan het geval zou zijn als er ook nieuwe data meegenomen zou worden.



Figuur F.2. ROC-curves van drie classificatiemodellen. Van boven naar onder: Beslissingsboom (DT), Random forest (RF) en Gradient boosting machine (GBM). De gestippelde lijn geeft een willekeurige kansverdeling weer.