



Monitor Escalaties Omgevingsrecht

Resultaat nulmeting fase I

Heike Goudriaan
Nick de Wolf
Luc Verschuren
Sigrid van Hoek

Juni 2020

Projectnummer CBS
Projectnummer BZK/ICTU

305380
VP99322

Inhoudsopgave

Samenvatting	4
1. Inleiding	3
1.1 Aanleiding/achtergrond	3
1.2 Haalbaarheidsonderzoek	4
1.3 Nulmeting	5
2. Databronnen	7
2.1 Registraties van escalaties in het omgevingsrecht	7
2.2 Representativiteit van het E-archief van de Rechtspraak	10
3. Methodiek nulmeting	12
3.1 Voorbewerking onderzoeksdata	12
3.2 Het maken van trainingsets voor machine learning: ‘annoteren’	12
3.3 Kwaliteit van de annotaties en de realisatie van annotatiesets	13
3.4 Modellen maken met behulp van annotatiesets	15
3.5 Implicaties van de gekozen onderzoeksmethodiek	17
4. Resultaten	19
4.1 Geschatte kenmerken gedigitaliseerde rechtbankuitspraken omgevingsrecht naar jaar	20
4.2 Geschatte kenmerken AbRvS-uitspraken omgevingsrecht naar jaar	31
4.3 Nauwkeurigheid van de schattingen	43
5. Conclusies	46
5.1 Opbrengst tot nu toe	46
5.2 Suggesties voor vervolgonderzoek	47
Bijlage 1. Samenstelling Begeleidingsgroep	48
Bijlage 2. Lijst met zoektermen t.b.v. selectie uitspraken omgevingsrecht uit E-archief	49
Bijlage 3. Tabellen bij Hoofdstuk 4	50
Bijlage 4. Rapportage haalbaarheidsonderzoek	52

Samenvatting

De Omgevingswet, die naar verwachting in 2022 in werking treedt, heeft onder andere als doel om de besluitvorming op het gebied van omgevingsrecht en ruimtelijke ordening te versoepelen, zodat bedrijven, burgers en overheden makkelijker en efficiënter zaken kunnen doen op dit gebied. De nieuwe wet beoogt minder en overzichtelijkere regels, een samenhangende benadering van de leefomgeving en snellere besluitvorming. Eén van de verwachtingen vanuit het beleid is dat de invoering van de Omgevingswet zal leiden tot een afname van het aantal conflicten op het gebied van ruimtelijke ordening dat door de rechter moet worden beslecht en in een wijziging van de aard en complexiteit van die conflicten. Met het doel inzicht te krijgen in hoeverre dit wordt bereikt heeft Min BZK/ICTU aan het CBS gevraagd te onderzoeken in hoeverre een zogenoemde nulmeting van de huidige stand van zaken (en in de toekomst mogelijk uitbreiding met een monitoring van de impact van de Omgevingswet) kan worden uitgevoerd door, met behulp van machine learning technieken, gestructureerde statistische informatie te herleiden uit te teksten van rechterlijke uitspraken. In 2019 heeft het CBS een haalbaarheidsonderzoek uitgevoerd, in samenwerking met de Raad voor de rechtspraak (Rvdr). In het eerste halfjaar van 2020 is deze nulmeting uitgevoerd.

Het is mogelijk gebleken om, op basis van informatie in de teksten van rechterlijke uitspraken, op een gestructureerde wijze een selectie van een aantal vooraf gespecificeerde kenmerken te classificeren en, op basis daarvan, statistische informatie te construeren over ontwikkelingen in escalaties in het omgevingsrecht over de tijd. Hierbij is nadrukkelijk geborgd dat onbevoegden geen inzage krijgen in privacygevoelige informatie in de documenten van de rechterlijke uitspraken. In deze eerste nulmeting wordt een eerste kwantitatief beeld gegeven van aard en omvang van deze escalaties in de periode 2007-2019, op basis van door Rvdr en de Raad van State (RvS) beschikbaar gestelde gedigitaliseerde teksten van rechterlijke uitspraken. Ook is er een basissysteem opgezet om in de toekomst deze nulmeting verder uit te kunnen breiden met actuelere data en extra kenmerken. De onderzochte kenmerken in deze nulmeting zijn afgestemd met een begeleidingscommissie en betreffen: geschilsoort, soort eiser/appellant, aantal eisers/appellanten (één of meerdere), (wel/geen) vertegenwoordiging eiser/appellant, soort verweerder, soort kamer (enkelvoudig/meervoudig), uitkomst geschil (gegrond, ongegrond, niet-ontvankelijk) en zaaktype (Eerste Aanleg, Hoger beroep).

De door het CBS geschatte kenmerken van escalaties in het omgevingsrecht over de periode 2007-2019 geven een relatief stabiel beeld in de tijd; de verhoudingen tussen categorieën van kenmerken veranderen slechts marginaal. Hoewel het niet met 100 procent zekerheid uit te sluiten is, zijn er voldoende aanwijzingen om aan te nemen dat deze stabiele resultaten een reëel beeld geven is en dat dit dus geen model-artefact is. De kenmerken voor de rechtbank-uitspraken en voor de uitspraken van de Afdeling bestuursrechtspraak van de Raad van State (de AbRvS-uitspraken) zijn separaat, in afzonderlijke modellen, geschat. Desalniettemin komt een deel van de resultaten redelijk overeen; bijvoorbeeld als je kijkt naar de verhoudingsgetallen van het aantal eisers/appellanten, vertegenwoordiging van eisers/appellanten en ook de uitkomst van het geschil. Er zijn ook zichtbare verschillen, zoals de verhoudingen naar soort eiser/appellant, soort kamer en soort verweerder. Als het gaat om het interpreteren van ontwikkelingen en verschillen moet rekening worden gehouden met een bepaalde mate van onzuiverheid door de toegepaste onderzoeksmethode.

Een belangrijk aspect voor de interpretatie van de resultaten betreft de representativiteit van de beschikbare onderzoeksdata met betrekking tot de rechterlijke uitspraken van rechtbanken. Het E-archief, waaruit deze uitspraken van rechtbanken zijn ontleend, is niet volledig gedigitaliseerd en er is op dit moment nog slechts een beperkt beeld van de omvang en de selectiviteit van deze 'onderdekking'. Voor zover bekend is het uitspraken-registratiesysteem van de RvS wel volledig. Bovendien wordt in dat systeem geregistreerd of een uitspraak omgevingsrecht betreft. Nader onderzoek naar de representativiteit van de geselecteerde uitspraken uit het E-archief wordt sterk aangeraden om de interpretatiewaarde van de uitkomsten van dit deel van het onderzoek te vergroten.

1. Inleiding

1.1 Aanleiding/achtergrond

De Omgevingswet, die naar verwachting in 2022 in werking treedt, heeft onder andere als doel om de besluitvorming op het gebied van omgevingsrecht en ruimtelijke ordening te versoepelen, zodat bedrijven, burgers en overheden makkelijker en efficiënter zaken kunnen doen op dit gebied. De nieuwe wet zorgt voor minder en overzichtelijkere regels, een samenhangende benadering van de leefomgeving en snellere besluitvorming. Zo komt er meer ruimte voor initiatieven van burgers en bedrijven, zijn er minder vergunningen nodig en worden de procedures korter. Eén van de verwachtingen vanuit het beleid is dat de invoering van de Omgevingswet zal leiden tot een afname van het aantal conflicten op het gebied van ruimtelijke ordening dat door de rechter moet worden beslecht en in een wijziging van de aard en complexiteit van die conflicten. Het ministerie van BZK laat onderzoeken in welke mate deze verwachtingen gerealiseerd gaan worden.

Monitor escalaties omgevingsrecht

Bij het bepalen van de aard en omvang van de impact van de invoering van de Omgevingswet op de rechtspraak spelen de volgende vragen een belangrijke rol: Wat is de huidige stand van zaken rondom rechterlijke uitspraken op het gebied van omgevingsrecht? Over hoeveel uitspraken hebben we het, welke uitspraken zijn het en wat weten we over de aard en complexiteit van de onderliggende conflicten? En wat is daadwerkelijk de impact van de Omgevingswet op de omvang en complexiteit van rechtszaken op het gebied van omgevingsrecht? Met het doel inzicht te krijgen in deze vragen heeft Min BZK/ICTU aan het CBS gevraagd te onderzoeken in hoeverre een zogenoemde nulmeting van de huidige stand van zaken (en in de toekomst mogelijk uitbreiding met een monitoring van de impact van de Omgevingswet) kan worden uitgevoerd met behulp van beschikbare rechterlijke uitspraken.

Machine learning

Rechterlijke uitspraken bevatten relevante informatie over de aard en complexiteit van conflicten tussen burgers en bedrijven enerzijds en overheden anderzijds in het omgevingsrecht. Daar het transcript van rechterlijke uitspraken in het bestuursrecht wel wordt vastgelegd, maar bij het archiveren hiervan maar een zeer beperkt aantal kenmerken wordt geclassificeerd (zo wordt bijvoorbeeld überhaupt niet geclassificeerd of sprake is van omgevingsrecht, laat staan over welk overheidsbesluit het conflict tussen eiser/appellant en gedaagde handelde), is onvoldoende *gestructureerde* informatie beschikbaar om dit onderzoek te kunnen baseren op meer traditionele analyses van registerdata. Min BZK/ICTU heeft derhalve het CBS gevraagd om in dit project gebruik te maken van innovatieve big data-technieken, zoals text mining en, waar nodig, machine learning, zodat gestructureerde, analyseerbare informatie onttrokken kan worden aan teksten van rechterlijke uitspraken om inzicht te verkrijgen in ontwikkelingen in de aard en omvang in escalaties in het omgevingsrecht.

Haalbaarheidsonderzoek

Als eerste stap is een haalbaarheidsstudie uitgevoerd, waarin het CBS heeft onderzocht in hoeverre het mogelijk is om, uit de tekstdocumenten van rechterlijke uitspraken op een gestructureerde wijze kenmerken te herleiden die iets zeggen over de aard van de uitspraken, zoals het type eiser, type bestreden besluit (ook wel aangeduid met geschilsoort) en de uitkomst, maar bijvoorbeeld ook over de complexiteit van een zaak. De rapportage van deze haalbaarheidsstudie is opgeleverd op 29 november 2019 en is als bijlage opgenomen bij dit rapport (zie Bijlage 4) Voor het haalbaarheidsonderzoek is gebruik gemaakt van niet-geanonimiseerde gedigitaliseerde rechterlijke uitspraken in het E-archief van de Rechtspraak (Rvdr) in de periode 1 januari 2007 tot en met 31 december 2018. Zoals in Hoofdstuk 2 wordt toegelicht betreft dit een selectie van alle uitspraken door rechtbanken in deze periode. Uitspraken gedaan door de Afdeling bestuursrechtspraak van de Raad van State (AbRvS) zijn in een latere fase, na afronding van het haalbaarheidsonderzoek, toegevoegd.

Leeswijzer

Onderhavig rapport handelt over de eerste uitkomsten van de nulmeting, die volgde op het haalbaarheidsonderzoek. In het vervolg van dit hoofdstuk wordt kort ingegaan op de resultaten van het haalbaarheidsonderzoek en op de scope van deze nulmeting. In Hoofdstuk 2 wordt ingegaan op de gebruikte databronnen en de mogelijkheden en beperkingen van die bronnen en in Hoofdstuk 3 wordt een beschrijving gegeven van de gebruikte methoden en technieken en de voor- en nadelen van deze technieken. Hoofdstuk 4 behandelt de kwaliteit van de uitkomsten van deze eerste nulmeting en hier worden ook de inhoudelijke bevindingen beschreven in de vorm van relatieve verhoudingen van categorieën binnen de onderzochte kenmerken in de periode 2007-2019. In Hoofdstuk 5 worden de conclusies en aanbevelingen voor het vervolg weergegeven.

1.2 Haalbaarheidsonderzoek

Gebruikte data

Het haalbaarheidsonderzoek voorafgaand aan de eerste nulmeting is verricht op rechtbankuitspraken in de periode 1 januari 2007 tot en met 31 december 2018. Deze zijn onttrokken uit het E-archief en door de Rvdr aan het CBS ter beschikking gesteld. Data van de RvS zijn later, na afronding van het haalbaarheidsonderzoek, toegevoegd.

Onderzoeksaanpak en uitkomsten haalbaarheidsonderzoek

Op basis van het haalbaarheidsonderzoek is geconstateerd dat het mogelijk is om documenten van rechterlijke uitspraken in rechtbanken in het omgevingsrecht uit het systeem (E-archief) van de Rechtspraak te halen, te verwerken en te ontsluiten ten behoeve van statistisch onderzoek. De uitspraken kunnen worden ontdebeld, gefilterd, samengevoegd en worden opgesplitst in relevante gestructureerde delen teneinde een aantal vooraf bepaalde kenmerken te kunnen classificeren met behulp van technieken voor text mining en machine learning. Ten behoeve van de analyses zijn in het haalbaarheidsonderzoek 360 van de toen ruim 35.000 relevant geachte rechterlijke uitspraken¹ handmatig geannoteerd (gelabeld) met behulp van een daartoe ontwikkelde tool, op de vooraf vastgestelde kenmerken en categorieën, door drie onafhankelijk van elkaar werkende annoteurs. Van een deel van de kenmerken is de haalbaarheid van automatisch annoteren onderzocht en werd vastgesteld dat dit met voldoende kwaliteit bepaald/berekend kan worden op basis van de teksten in de rechterlijke uitspraken; het ging dan in het bijzonder om kenmerken die met behulp van reguliere expressies (regex) kunnen worden gecategoriseerd. Een reguliere expressie is een manier om patronen te beschrijven waardoor een computer softwarematig tekst kan herkennen. Voor het overige, meer complexe deel, bleek meer input van deskundigen nodig te zijn om te komen tot een eerste versie van een geclassificeerd kenmerk.

Representativiteit

Voorts is in het haalbaarheidsonderzoek gebleken dat de voor het onderzoek geleverde data uit het E-archief ver af lag van de onderliggende 'doelpopulatie', namelijk alle omgevingsrecht-gerelateerde rechterlijke uitspraken in de te onderzoeken periode. Ten eerste bleek dat veel uitspraken uit het E-archief waren onttrokken en aangeleverd die bij nadere bestudering geen betrekking hadden op het omgevingsrecht (overdekking). In het haalbaarheidsonderzoek zijn stappen gezet om die overdekking te minimaliseren door nadere filtering van aangeleverde rechtbank-uitspraken. Ten tweede waren uitspraken van de AbRvS nog niet beschikbaar voor het haalbaarheidsonderzoek. Ten derde is vastgesteld dat een deel van de relevante rechtbank-uitspraken ontbrak

¹ Onder deze ruim 35.000 uitspraken bevonden zich meer dan 8.000 (ruim 24%) uitspraken door de voorzieningenrechter. Deze zijn in het haalbaarheidsonderzoek tezamen met de uitspraken ten principale geanalyseerd. Omdat deze uitspraken deels een afwijkende structuur en kenmerken hebben, kan dit de resultaten van het haalbaarheidsonderzoek enigszins (nadelig) hebben beïnvloed, in die zin dat het classificeren van de gevraagde kenmerken minder goed mogelijk leek dan dat dit het geval zal zijn wanneer de uitspraken door de voorzieningenrechter buiten beschouwing zouden zijn gelaten of apart zouden zijn bestudeerd. De uitspraken door de voorzieningenrechter zijn op verzoek van ICTU/BZK in deze eerste uitkomsten van de nulmeting niet meegenomen in de analyses.

doordat deze ofwel met een filter ten onrechte niet uit het E-archief waren onttrokken ofwel doordat ze überhaupt niet zijn gedigitaliseerd in het E-archief (onderdekking). Dat niet alle rechtbank-uitspraken in het E-archief zijn opgenomen was bij de Rvdr op voorhand bekend, maar er was nog geen exact beeld van hoe groot die onderdekking is. Bovendien was nog onbekend in hoeverre hierbij sprake is van selectiviteit. De gevolgen van onderdekking voor statistiek kunnen groot zijn. Vooral als een selecte groep uit de doelpopulatie daardoor niet wordt meegenomen in het onderzoek, waardoor de resultaten een vertekend beeld kunnen opleveren. Zo is het niet ondenkbaar dat er sprake is van selectiviteit in de beschikbare rechtbank-uitspraken, bijvoorbeeld wanneer er sprake is van een toenemend aandeel gedigitaliseerde uitspraken over de tijd of wanneer de kans dat een uitspraak wordt gedigitaliseerd samenhangt met iets als de complexiteit van een zaak. Of wanneer de voor de selectie gebruikte trefwoorden beter aansluiting bij de huidige wetgeving dan die in eerdere jaren. Helaas is onderdekking niet altijd op te lossen (bijv. wanneer een deel van de zaken in het geheel niet is gedigitaliseerd) en ook niet eenvoudig in kaart te brengen. In de rapportage van het haalbaarheidsonderzoek is nader ingegaan op mogelijke oorzaken van de onderdekking en zijn aanbevelingen gedaan voor verder onderzoek hiernaar.

Van haalbaarheidsonderzoek naar nulmeting

Hoewel de eerdergenoemde aandachtspunten over representativiteit relevant zijn voor de interpretatie van resultaten en voor een verdere doorontwikkeling van het onderzoek, werd er voldoende basis gezien voor het maken van een eerste nulmeting over escalaties in het omgevingsrecht; er was een substantiële dataset van documenten van uitspraken uit het E-archief beschikbaar, die bovendien bruikbaar was voor het toepassen van machine learning-technieken. Bovendien was er op dit moment ook het perspectief dat voor de nulmeting ook de uitspraken van de AbRvS toegevoegd zouden kunnen worden aan het onderzoek (zie ook 1.3).

Voor meer informatie over de gebruikte data, de verwerking van deze data, de gebruikte onderzoeksmethoden én de resultaten van het haalbaarheidsonderzoek wordt verwezen naar de haalbaarheidsrapportage, die op 29 november 2019 is opgeleverd (zie Bijlage 4).

1.3 Nulmeting

In deze nulmeting wordt voor een aantal – met de externe begeleidingsgroep (zie Bijlage 1 voor een overzicht van de leden van deze groep) afgestemde – kenmerken, ontwikkelingen geschat voor de gedigitaliseerde rechterlijke uitspraken in het omgevingsrecht in de periode 1 januari 2007 tot en met 31 december 2018 (uitspraken rechtbanken) en 1 juli 2008 tot en met 31 december 2019 (uitspraken AbRvS). Hierbij ligt nadruk op gedigitaliseerd, omdat – zoals in de vorige paragraaf toegelicht – in de haalbaarheidsfase bleek dat een substantieel deel van de rechterlijke uitspraken van rechtbanken niet is gedigitaliseerd (ofwel: niet is opgenomen in het E-archief van de Rechtspraak). De afgestemde kenmerken zijn opgenomen in Tabel 1.1. De kenmerken 5 tot en met 11 worden in deze nulmeting gelabeld en voor zover mogelijk verwerkt tot geschatte verdelingen over categorieën voor de onderzochte periode. Categorieën met een lage frequentie kunnen in deze fase mogelijk (nog) niet met voldoende betrouwbaarheid worden onderscheiden. Wanneer dit het geval is zullen deze worden samengevoegd met een andere categorie. Indien van toepassing is er naast de categorieën in Tabel 1.1 ook een categorie *Overig* gebruikt.

TABEL 1.1. Kenmerken rechterlijke uitspraken

Nr.	Variabele (kolom)	Categorieën
1	ECLI-nummer	Uniek ID per uitspraak, behorende bij de uitspraken van de Rvdr (<i>meta-informatie</i>)
2	Zaaknummer	Uniek ID per uitspraak, behorende bij de uitspraken van de AbRvS (<i>meta-informatie</i>)
3	Datum uitspraak	Datum van de uitspraak (<i>meta-informatie</i>)
4	Rechtbank/AbRvS	Naam Rechtbank of AbRvS (<i>apart aangeleverd/losse bestanden + meta-informatie</i>)
4a	EA/HB	Betreft het een uitspraak in Eerste Aanleg of in Hoger Beroep (<i>later toegevoegde meta-informatie RvS</i>) ^a
5	Soort eiser(s)/appellant(en)	Particulier; niet-particulier ^b ; verschillende soorten eisers
6	Aantal eisers/appellanten	Eén; twee of meer
7	Vertegenwoordiging eiser(s)/appellanten	Wel; geen
8	Geschilsoort	Vergunningsbeschikking ^c ; Handhavingsbeschikking; Planschadebeschikking; Bestemmingsplan (<i>alleen AbRvS</i>); Tracébesluit (<i>alleen AbRvS</i>) ^d
9a	Uitkomst geschil gegrond	Uitkomst geheel of gedeeltelijk gegrond (<i>in HB: uitspraak RB (deels) vernietigd</i>)
9b	Uitkomst geschil ongegrond	Uitspraak geheel of gedeeltelijk ongegrond (<i>in HB: uitspraak RB (deels) bevestigd</i>)
9c	Uitkomst geschil niet-ontvankelijk	Eiser/appellant (deels) niet ontvankelijk verklaard
9d	Uitkomst geschil onbevoegd	Rechter/rechtbank/afdeling (deels) onbevoegd verklaard
10	Soort verweerder	Gemeente; Waterschap; Provincie; Rijk
11	Soort kamer	Enkelvoudige Kamer; meervoudige kamer

^a Uitspraken door de rechtbank betreffen per definitie zaken in eerste aanleg, die bij de AbRvS betreffen zaken in eerste (en enige) aanleg of zaken in hoger beroep.

^b Bij het annoteren worden ook de categorieën overheid, bedrijf en stichting apart onderscheiden. Wanneer in de modelschattingen zou blijken dat (een deel van) deze onderliggende categorieën met voldoende betrouwbaarheid zouden kunnen worden onderscheiden, dan zou dit alsnog apart worden gedaan. Dit blijkt echter in deze fase nog niet het geval.

^c Inclusief vrijstelling, ontheffing, projectbesluit, milieumelding en aanwijzingsbesluit ORAC. Dit kunnen zowel uitspraken door de rechtbanken zijn (dus in EA), als uitspraken door de RvS (in het algemeen – maar hierop zijn uitzonderingen – in HB)

^d Bestemmingsplanzaken en Tracébesluitzaken worden niet behandeld bij een rechtbank, maar gaan in eerste aanleg direct naar de AbRvS.

Positionering nulmeting

Deze nulmeting kan worden gebruikt voor de beantwoording van inhoudelijke onderzoeksvragen zoals die door BZK/ICTU zijn geformuleerd en het is een tussenstap om in de toekomst, ná de invoering van de Omgevingswet, onderzoek te kunnen doen naar samenhang tussen enerzijds de invoering van de nieuwe Omgevingswet en anderzijds ontwikkelingen in aard en omvang van escalaties in het omgevingsrecht. In Hoofdstuk 5 worden ook nog concrete suggesties gedaan voor het verder doorontwikkelen van de nulmeting, alvorens te starten met vervolgmetingen, ná de invoering van de Omgevingswet.

2. Databronnen

Rechterlijke uitspraken in eerste aanleg door rechtbanken zijn onttrokken uit het E-archief van de Rechtspraak. Uitspraken in eerste aanleg en in hoger beroep door de AbRvS zijn onttrokken uit het uitsprakenregistratiesysteem van de RvS. In Paragraaf 2.1 wordt nader ingegaan op beide bronnen, de wijze van onttrekken van de voor dit onderzoek benodigde data en de filtering die na aanlevering op de data heeft plaatsgevonden zodat alleen de voor dit onderzoek relevante uitspraken werden geanalyseerd. Paragraaf 2.2 handelt in het bijzonder over aandachtspunten van het E-archief.

Om een zo goed mogelijk beeld te krijgen van de kwaliteit van deze data, met name met betrekking representativiteit van de gedigitaliseerde rechtbankuitspraken in het E-archief, is gebruik gemaakt van een aanvullende bron, te weten het financiële archief van de Rechtspraak (InfoRM). Dit komt aan bod in Paragraaf 2.2.

Alle databronnen met privacygevoelige informatie die zijn gebruikt voor dit onderzoek, zijn op een veilige wijze geleverd aan en verwerkt door het CBS. Zie www.cbs.nl voor meer informatie over hoe het CBS omgaat met privacygevoelige data.

2.1 Registraties van escalaties in het omgevingsrecht

Rechtbanken – het E-archief

Rechterlijke uitspraken van de rechtbanken uit de periode 1-1-2007 tot en met 31-12-2018 zijn eind juli 2019 aan het CBS aangeleverd door de Rvdr. De Rvdr heeft hiertoe, met behulp van een vooraf door experts samengestelde lijst van zoektermen (opgenomen in Bijlage 2), rechterlijke uitspraken onttrokken aan het E-archief die betrekking zouden kunnen hebben op het omgevingsrecht. Het gebruik van zoektermen was nodig omdat niet uit de metadata bij een uitspraak kan worden afgeleid of een zaak wel of niet betrekking heeft op het omgevingsrecht. Deze zoektermen moesten resulteren in een zo compleet mogelijke voorselectie van rechterlijke uitspraken die te maken hebben met omgevingsrecht.

Deze werkwijze resulteerde in de aanlevering van een groot aantal XML-bestanden met elk maximaal 1.000 rechterlijke uitspraken. In totaal bevatten deze bestanden bijna 284.000 documenten. Elk document bevat de tekst van een rechterlijke uitspraak en bijbehorende metadata (bijv. datum uitspraak, instantie, uniek identificatienummer). Door de wijze van onttrekken aan het E-archief, namelijk een aparte bevraging voor elk steekwoord, bevatte de geleverde dataset veel documenten dubbel. Bovendien had een aanzienlijk deel van de documenten zoals vooraf werd verwacht geen betrekking op het omgevingsrecht (overdekking). Documenten met enkel de zoekterm 'ontheffing' bleken bijvoorbeeld regelmatig betrekking te hebben op civielrecht en meer specifiek familierecht (ontheffingen uit het ouderlijk gezag).² Omgekeerd heeft de wijze van onttrekken naar verwachting geleid tot het ontbreken van een deel van de uitspraken die wel degelijk betrekking hebben op het omgevingsrecht. Op deze overdekking en onderdekking wordt in de navolgende paragrafen nader ingegaan.

Raad van State – uitsprakenregistratiesysteem

De RvS heeft in februari 2020 uit haar uitsprakenregistratiesysteem een set van in totaal ruim 31.000 documenten geleverd aan het CBS. Deze zijn aangeleverd in de vorm van Word-documenten, waarbij elk document een enkele uitspraak bevatte en geen metadata. De documenten waren uniek te identificeren aan de bestandsnaam met daarin een zaaknummer. Deze AbRvS-uitspraken betreffen de periode 1-7-2008 tot en met 31-12-2019 en het gaat om alle uitspraken van de Ruimtelijke ordeningskamer (R1, R2, R3, R4 en R6), Milieukamer (M1, M2 en M3), Algemene kamer (Unit A1) en H-kamer (Unit H1). Medio maart 2020 is bovendien

² Wanneer voor een vervolg op deze nulmeting een nieuwe extractie uit het E-archief wordt gehaald, is het verstandig de wijze van aanlevering aan te scherpen door vooraf een selectie te maken op alleen bestuursrechtelijke uitspraken (met behulp van beschikbare meta-informatie) en indien mogelijk door toepassing van combinaties van zoektermen in plaats van een losse bevraging per zoekterm.

voor alle documenten aanvullende informatie door de RvS aangeleverd of het een uitspraak in eerste en enige aanleg (EEA; voorts aangeduid met EA) of hoger beroep (HB) betreft. Deze informatie heeft het CBS middels het unieke documentnummer aan de eerder geleverde documenten gekoppeld.

Omdat in het archief van de RvS de uitspraken in het omgevingsrecht als zodanig zijn gelabeld en bovendien wordt verondersteld dat het archief ook volledig (gedigitaliseerd) is, betrof dit reeds een scherpe afbakening van alle omgevingsrecht-uitspraken door de AbRvS. Hier zijn geen indicaties dat sprake is van over- en/of onderdekking.³

Verdere filtering van uitspraken rechtbanken en AbRvS

Het CBS heeft, op basis van de door de Rvdr en de RvS geleverde documenten een nadere filtering uitgevoerd om uiteindelijk tot een minimale overdekking, ofwel een zo scherp mogelijke populatieselectie te komen. Deze filtering vond plaats in afstemming met Min BZK/ICTU en met adviseur prof. dr. Bex, tevens lid van de externe begeleidingsgroep. Voor de rechtbankuitspraken uit het E-archief heeft deze nadere filtering zoals verwacht een groter effect dan voor de AbRvS-uitspraken (zie Tabel 2.1 en 2.2).

TABEL 2.1. Filtering van de geleverde documenten met uitspraken rechtbanken

Stap	Omschrijving	Resterende uitspraken (N)
0.	Initieel geleverde documenten E-archief – Rvdr	283.612
1.	1. Ontdubbelen van dubbel geleverde documenten op basis van ECLI-nummer (<i>metadata</i>) 2. Filteren op instantietype (alleen rechtbank behouden) (<i>metadata</i>) 3. Verwijderen jurisprudentie-stukken 4. Verwijderen bestanden zonder tekst of met tekstveld <150 tekens	64.796
2.	Behoud alleen documenten die betrekking hebben op omgevingsrecht, bestuursrecht algemeen en bestuursprocesrecht (<i>metadata</i>)	46.052
3.	Filter zaken weg op basis van woorden zoals “erfrecht”, “kinderbescherming”, etc. ^a	44.245
4.	Filter lastig geschreven proces-verbalen eruit, evt. in toekomst apart behandelen ^b	42.572
5.	Verwijder anonieme documenten	35.434
6.	Verwijder uitspraken voorzieningenrechter ^c , evt. in toekomst apart behandelen	26.339

^a Dit is uitgevoerd aan de hand van zogenaamde regex-filters (zie ook Paragraaf 3.1 van de Haalbaarheidsrapportage in Bijlage 4), waarbij geautomatiseerd gekeken wordt of bepaalde combinaties van woorden voorkomen in een (specifiek deel van de) tekst. De volledige woordenlijst waarop is gefilterd; “de rechtbank sluit het onderzoek ter zitting en deelt mee dat op”, “bijlage, inhoudende de bewijsmiddelen in de zaak”, “aanvulling bewijsmiddelen”, “samenvatting van de stappen voor het opslaan van een”, “opslaan van uitspraken in het e-archief”, “fout! bladwijzer niet gedefinieerd”, “kinderbescherming”, “sector familie- en jeugdrecht”, “deze uitspraak is vervallen verklaard”, “schuldsanering”, “erfrecht”, “kinderrechter”, “jeugdrecht”, “team handel”, “kantonrecht”, “kanton”, “kopje”, “deze uitspraak is ambtshalve vervallen”, “nakomen koopovereenkomst”, “faillissementswet”, “belastingdienst”, “vreemdelingenzaken”, “hoogheemraadschap”, “het college van dijkgraaf”, “commissie sociale zekerheid”, “rijbewijs”.

^b Dit is eveneens uitgevoerd aan de hand van regex-filters. De woordenlijst waarop is gefilterd: “proces-verbaal van het verhandelde”, “proces-verbaal van de verhandelde”, “proces-verbaal van de zitting”, “proces-verbaal van de behandeling”, “proces-verbaal van de comparitie”, “proces-verbaal van comparitie”, “proces-verbaal van de mondeling”, “proces-verbaal mondelinge uitspraak”, “proces-verbaal van de op de openbare zitting”, “proces-verbaal uitspraak van”, “proces-verbaal ingevolge artikel”, “P R O C E S”, “proces-verbaal van de enkelvoudige kamer”.

^c Ook dit is uitgevoerd aan de hand van regex-filters door in de kop-tekst van de uitspraak te zoeken op woorden als “voorlopige voorziening” of “voorzieningenrecht”.

³ Er is verkend of er mogelijkheden zijn om de aantallen uitspraken in onderhavige nulmeting, en tevens de verhoudingen in de geschatte verdelingen van een aantal van de hierin onderzochte kenmerken, te vergelijken met aantallen en procentuele verdelingen in het Management Informatiesysteem van de RvS. Dit MI-systeem is echter niet gemaakt is om over de specifieke kenmerken in dit onderzoek te rapporteren. Het is ook niet mogelijk om uit het MI-systeem dezelfde bulk aan data te onttrekken. Voor de toekomst is het volgens de onderzoekers echter wel van belang te investeren in een dergelijke vorm van triangulatie – wellicht toch door te putten uit het MI-systeem; of door een andere selectie aan uitspraken uit het uitsprakenregistratiesysteem (die wel ook uit het MI-systeem onttrokken kan worden) middels machine learning te onderzoeken.

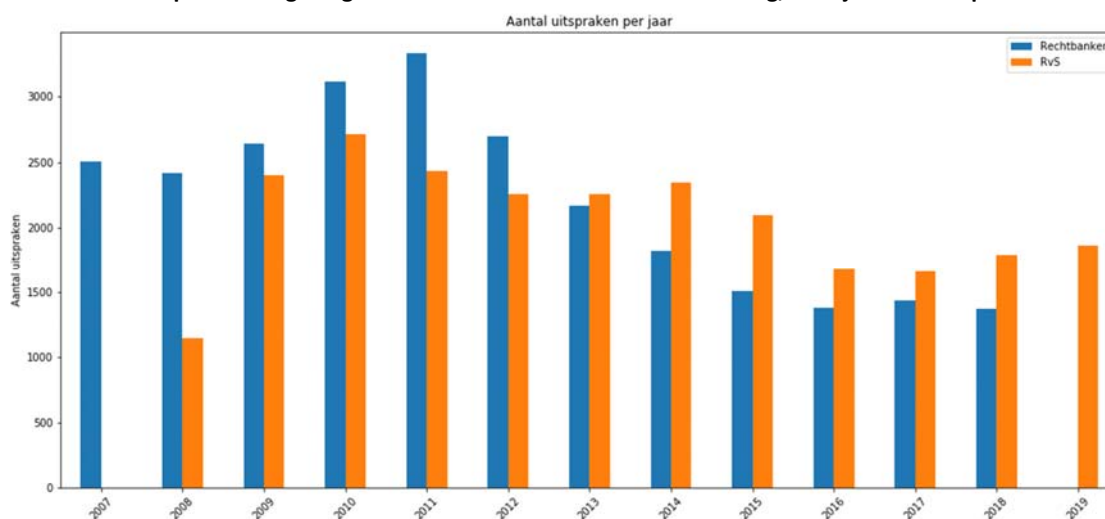
TABEL 2.2. Filtering van de geleverde documenten met uitspraken AbRvS

Stap	Omschrijving	Resterende uitspraken (N)
0.	Initieel geleverde documenten Uitsprakenregistratiesysteem – RvS	31.419
1.	Filter zaken weg op basis van woorden zoals “erfrecht”, “kinderbescherming”, etc. ^a	30.569
2.	Verwijder uitspraken voorzieningenrechter ^a , evt. in toekomst apart behandelen	24.667

^a Zie noten onder Tabel 2.1 voor uitleg over de wijze waarop is gefilterd.

Onderzoeksbestanden

De twee uiteindelijke onderzoeksbestanden bevatten het aantal (gedigitaliseerde) uitspraken in het omgevingsrecht van de rechtbanken en AbRvS na de hierboven beschreven filtering. In Figuur 2.1 zijn deze aantallen uitgesplitst naar jaar van uitspraak. Benadrukt moet worden dat door de beperkte digitalisering van de rechtbankuitspraken deze aantallen niet mogen worden gezien als aantallen rechtbankuitspraken in het omgevingsrecht. Let op: de reeks van uitspraken van de AbRvS loopt niet vanaf 1-1-2008, maar vanaf 1-7-2008, dus dit aantal beslaat slechts de tweede helft van dit jaar.

FIGUUR 2.1. Uitspraken omgevingsrecht rechtbanken en AbRvS na filtering, naar jaar van uitspraak

In onderstaande tabel is het aantal in het E-archief gedigitaliseerde en als omgevingsrecht geselecteerde rechtbank-uitspraken weergegeven per rechtbank per jaar van uitspraak. Ook hier gaat het om het aantal uitspraken ná nadere filtering zoals beschreven in Tabel 2.1.

TABEL 2.3. Gedigitaliseerde uitspraken omgevingsrecht rechtbanken, naar rechtbank^a en jaar van uitspraak

Rechtbank	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Oost-Brabant	5	172	348	565	644	485	250	199	175	157	145	163
Overijssel	166	129	155	342	160	185	74	136	120	100	62	55
Noord-Nld	188	162	163	220	224	242	167	116	150	105	104	85
Limburg	348	268	273	235	266	229	236	177	127	115	115	117
Zeeland/W-Br.	292	274	439	303	349	244	262	237	190	185	170	182
Amsterdam	275	270	212	187	258	112	101	75	63	63	99	74
Den Haag	400	394	317	356	402	286	247	216	150	173	156	145
Midden-Nld	191	161	108	189	274	189	168	108	114	96	99	105
Noord-Holland	272	216	254	262	234	190	162	124	98	81	99	112
Rotterdam	151	177	185	226	223	210	234	247	122	127	161	124
Gelderland	212	192	183	224	292	327	139	183	193	174	216	206
Oost-Nld	0	0	0	0	0	0	120	0	0	0	0	0

^a Situatie per 1 april 2013 met de huidige 11 rechtbanken, plus de tijdelijke rechtbank Oost-Nederland. Tot 1 januari 2013 waren er 19 rechtbanken. De Rechtbank Oost-Nederland was van 1 januari 2013 tot en met 31 maart 2013 een van de rechtbanken in Nederland. De rechtbank ontstond uit een samenvoeging van de rechtbanken Arnhem, Almelo, Zutphen en het Overijsselse deel van de rechtbank Zwolle-Lelystad. Op 1 april 2013 werd deze rechtbank weer gesplitst in de rechtbank Overijssel en de rechtbank Gelderland, maar deze splitsing naar de huidige 11 rechtbanken was met de beschikbare gegevens niet te maken.

De uitspraken uit het E-archief van de Rechtspraak en die uit het uitsprakenregistratiesysteem van de RvS die na filtering overbleven hebben een relatief vaste structuur en consistente terminologie. Dat dit voor de data uit het E-archief het geval is bleek reeds in het haalbaarheidsonderzoek. Wel verschilt de structuur en gebruikte terminologie enigszins tussen de beide bronnen. Omdat bovendien de periode waarover documenten zijn aangeleverd in beide bronnen verschillend was, is ervoor gekozen de twee bronnen in het verdere onderzoek separaat te behandelen. Ook de onderzoeksresultaten en geschatte kenmerken worden separaat berekend en in deze rapportage telkens apart beschouwd.

2.2 Representativiteit van het E-archief van de Rechtspraak

Na het haalbaarheidsonderzoek is de filtering van de aangeleverde documenten verder aangescherpt, om de overdekking nog verder te verkleinen. In het haalbaarheidsonderzoek werden mogelijke oorzaken van onderdekking in (de aangeleverde rechtbankuitspraken uit) het E-archief aangestipt. Deze konden in de haalbaarheidsfase nog niet daadwerkelijk worden onderzocht wegens het ontbreken aan nadere informatie. In de onderzoeksfase van de eerste nulmeting is hier wél nader onderzoek naar verricht:

Onvolledigheid en representativiteit E-archief

In een poging zicht te krijgen op de representativiteit van de beschikbare dataset uit het E-archief heeft de Rvdr in nauwe samenwerking met het CBS een representativiteitsanalyse uitgevoerd. Hiertoe is door de Rvdr een koppeling gemaakt tussen het E-archief en het Datawarehouse van de Rechtspraak (InfoRM). Dit kon worden gedaan voor uitspraken in de jaren 2014-2018. Omdat InfoRM tevens een financieel archief is, wordt verondersteld dat dit een (nagenoeg) volledige gedigitaliseerde registratie bevat van alle omgevingsrecht-zaken. De Rvdr heeft, zo goed als mogelijk, een 'micro-koppeling' gemaakt tussen beide registraties, dat wil zeggen een koppeling op zaaksniveau. Een dergelijke exercitie is (in ieder geval in theorie) de beste manier om goed inzicht te krijgen in de mismatch tussen beide registraties. Hierbij waren echter enkele complicerende factoren:

- Het E-archief is op het niveau van rechterlijke uitspraken en InfoRM op het niveau van rechtszaken; dat zijn verschillende entiteiten
- Beide registraties bevatten weliswaar een 'zaaknummer', maar de notatie van dit zaaknummer is niet uniform, hetgeen een koppeling bemoeilijkt
- Net als in het E-archief is er in InfoRM geen vast label voor omgevingsrecht; er zijn alleen regelgevingscodes om in te schatten of een zaak wel of geen omgevingsrecht betreft

- InfoRM bevat data vanaf 2014; deze exercitie levert dus alleen inzichten op voor de periode 2014-2018 en niet voor de periode 2007-2013.

Al met al heeft de Rvdr ondanks deze complexiteit een goede micro-koppeling kunnen maken, op zaaksniveau, tussen de selectie uit het E-archief en InfoRM (95% van de uitspraken konden worden gekoppeld aan een zaak in InfoRM), waardoor de representativiteitsanalyse de volgende inzichten heeft opgeleverd:

- Uit een selectie van zaken met 'soort regelgeving = WABO' uit InfoRM is 30 procent van deze zaken terug te vinden in de query uit het E-archief. Onder de aanname dat 'soort regelgeving=WABO' uitsluitend omgevingsrecht betreft, betekent dit dat voor dit deel van omgevingsrecht dus ongeveer 70 procent ofwel niet is gedigitaliseerd ofwel onbedoeld niet is meegekomen in de query uit het E-archief. Rond dit dekkingspercentage van 30 (100-70) procent zit vermoedelijk nog een ruime marge⁴, als men conclusies zou willen trekken voor alle omgevingsrecht-uitspraken in de periode 2007-2018:
 - De vergelijking kon niet worden uitgevoerd voor de jaren 2007-2013
 - WABO is slechts een deel van totaal omgevingsrecht
 - Het is niet 100 procent zeker dat alle zaken in InfoRM zuiver zijn geclassificeerd; dat is een aanname.
- De onderdekking in de geleverde uitspraken uit het E-archief, voor het deel WABO (zie hierboven), lijkt regionaal te verschillen. Onder voorbehoud van verschillende aannames, ontstaat het beeld op basis van de analyses dat bijvoorbeeld dat de regio's Gelderland en Oost-Brabant een relatief hoge dekking hebben (resp. 41 en 40%) en de regio's Amsterdam en Overijssel een relatief lage dekking (resp. 17 en 22 %). In sommige regio's lijkt de dekking te zijn gestegen sinds 2014 en in andere regio's juist gedaald.
- Tot slot heeft deze exercitie, in samenwerking met de Rvdr, ook bijgedragen aan meer inzicht in resterende ruis (dus juist de overdekking) in de query uit het E-archief. Op basis van deze inzichten heeft het CBS de filtering, zoals in Paragraaf 2.1 beschreven, nog verder aan kunnen scherpen.

Conclusie representativiteit

Al met al is er op dit moment een eerste beeld van de representativiteit van het E-archief. Voor de periode 2014-2018 lijkt er voor het deel WABO een substantieel deel (ongeveer 70%) te missen doordat uitspraken niet gedigitaliseerd zijn, óf doordat zij ten onrechte niet zijn meegekomen in de zoektermen-query uit het E-archief. Een uitdaging voor de toekomst van de Monitor Escalaties Omgevingsrecht is om hier beter zicht op te krijgen. Ook is het van belang om in de doorontwikkeling van de nulmeting te onderzoeken in hoeverre er relevante zoektermen zijn vergeten (waardoor relevante uitspraken nu ten onrechte niet zijn geselecteerd). De in deze rapportage getoonde aantallen en kenmerken hebben derhalve betrekking op de *geselecteerde en gedigitaliseerde* uitspraken van rechtbanken in het omgevingsrecht in de onderzochte periode en kunnen niet zonder meer worden ggeneraliseerd naar alle uitspraken in het omgevingsrecht.

⁴ Eerder heeft de Rvdr zelf een analyse gedaan van de dekkingsgraad van het E-archief in verschillende rechtsgebieden. Uit die analyse, over het jaar 2014, bleek dat de dekkingsgraad van bestuursrecht, met 75 procent, ten opzichte van strafrecht en civiel recht (resp. 13 en 40%) juist relatief hoog is. Die conclusie gaat over het totaal van bestuursrechtzaken (dus niet alleen omgevingsrecht).

3. Methodiek nulmeting

In dit hoofdstuk wordt achtereenvolgens ingegaan op de voorbereiding van de twee onderzoeksbestanden met rechtbankuitspraken en AbRvS-uitspraken (Paragraaf 3.1), het maken van trainingsets voor het trainen van de modellen (3.2 en 3.3) en de in dit onderzoek gebruikte modellen zelf (3.4). Omdat deze zaken deels ook al staan toegelicht in de Haalbaarheidsrapportage in Bijlage 4, zal hier waar mogelijk naar worden verwezen.

3.1 Voorbereiding onderzoeksdata

De tekst van de uitspraken is onderworpen aan enkele voorbereidingsstappen. Allereerst zijn deze ‘opgeschoond’ zoals beschreven in Bijlage 4 in Paragraaf 1.4 onder de kop ‘Preprocessing’ van de Haalbaarheidsrapportage.

Splitsing uitspraakteksten

Omdat de voor het onderzoek relevante kenmerken in het algemeen telkens te vinden zijn in een specifiek deel van de uitspraaktekst, zijn de uitspraken in de twee gefilterde onderzoeksbestanden gesplitst in verschillende tekstblokken. Door de uitspraken vooraf op deze wijze te splitsen krijgt het algoritme minder ruis binnen en het vergemakkelijkt bovendien het handmatig annoteren (zie Paragraaf 3.2). De uitspraken bevatten in het algemeen de elementen zoals weergegeven in Tabel 3.1.

In 96 tot 99 procent van de uitspraken komen de tekstelementen bovendien in de volgorde zoals in Tabel 3.1 weergegeven voor in de uitspraak. In enkele uitspraken is de volgorde anders of missen er duidelijke splitsingen tussen de onderdelen. Zo lopen Procesverloop en Overwegingen wel eens samen en ook de splitsing tussen Overwegingen en Conclusie is niet altijd goed te maken.

TABEL 3.1. Tekstelementen in uitspraken rechtbanken en AbRvS

Naam	Omschrijving
Kop	Bevat de uitspraakdatum, instantie, zaaknummers en soort kamer in rechtbankzaken.
- Eisers	Staat in de Kop. Bevat informatie over de eisers/appellanten.
- Verweerders	Staat in de Kop. Bevat de informatie over de verweerders. Eventuele derden staan hier vaak ook voor het eerst vermeld.
Procesverloop	Bevat informatie over het bestreden besluit (geschilsoort) en over de stappen die tot dit punt zijn ondernomen in het geding. Vermeldt ook welke partijen (met eventuele vertegenwoordiging) aanwezig waren op de zitting. Eventuele Derden worden hier ook (opnieuw) genoemd.
Overwegingen	Bevat informatie over de overwegingen van de rechter(s).
Conclusie	Bevat per eis de uitkomst van het geschil, ofwel het besluit van de rechter, soms met een korte uitleg. Ook proceskostenveroordeling wordt hier vaak kenbaar gemaakt. In AbRvS-uitspraken staat hier de informatie over enkelvoudige/meervoudige kamer.

Zie voor meer informatie over hoe deze splitsing is gemaakt Paragraaf 3.1 van de Haalbaarheidsrapportage (Bijlage 4).

3.2 Het maken van trainingsets voor machine learning: ‘annoteren’

Het trainen van algoritmes

Om gebruik te kunnen maken van de voor dit onderzoek benodigde technieken voor text mining en machine learning, is het noodzakelijk om bij een deel van de data de ‘uitkomsten’ (hier: categorieën van kenmerken) reeds te kennen. Hierop kan vervolgens een wiskundig model (algoritme) worden losgelaten om vervolgens

voorspellingen te doen voor data waarvan de uitkomsten nog niet bekend zijn. Het algoritme leert welke kenmerken van een dataset leiden tot een bepaalde uitkomst. Dat lerend vermogen heet het trainen van een algoritme. Dit proces herhaalt zich een aantal keer, omdat men diverse algoritmes laat rekenen en op die manier laat welk model de beste resultaten behaalt. Wanneer er een succesvol getraind model is, kan hiermee voor nieuwe data de uitkomsten worden voorspeld.

Het maken van geannoteerde datasets

In dit onderzoek zijn apart geannoteerde datasets gemaakt voor de rechtbankuitspraken ($n=478$) en voor de AbRvS-uitspraken ($n=446$). Dit is gedaan door bij twee subsets van de uitspraken handmatig de te onderzoeken kenmerken te classificeren, ofwel te ‘annoteren’. De te annoteren subset van rechtbankuitspraken is voor een deel samengesteld en geannoteerd in de haalbaarheidsfase en deze is aangevuld voor de nulmeting. In de haalbaarheidsfase zijn ook nog niet alle in de nulmeting overeengekomen kenmerken geannoteerd. Daarnaast zijn in de haalbaarheidsfase de voorzieningenrechterzaken meegenomen. Omdat die in deze eerste nulmeting buiten beschouwing zijn gehouden, was een deel van de oorspronkelijk annotaties onbruikbaar. Uiteindelijk konden 210 geannoteerde rechtbankuitspraken uit de haalbaarheidsfase in de nulmeting worden gebruikt en zijn er 268 aanvullende rechtbankuitspraken geannoteerd.

De te annoteren uitspraken zijn aselekt geselecteerd uit de onderzoeksbestanden, maar hierbij is wel gestratificeerd naar de drie rechtsgebieden die in de totale dataset voorkwamen (zie Tabel 2.1 stap 2 voor deze drie rechtsgebieden). Bij de te annoteren subset van AbRvS-uitspraken is enkel gestratificeerd naar het jaar van de uitspraak (in de AbRvS-uitspraken is geen metadata beschikbaar over het rechtsgebied), zodat de dataset uit elk beschikbaar jaar evenveel uitspraken bevatte.

Het resultaat bestaat uit twee geannoteerde datasets. Het maken van dergelijke datasets is een arbeidsintensief proces. Zo worden alle voor de trainingsets geselecteerde uitspraken onafhankelijk van elkaar door drie personen geannoteerd, om het aantal menselijke beoordelingsfouten te verkleinen. Deze drie personen, de *annotateurs*, dienen voldoende inhoudelijke kennis van het omgevingsrecht en van het juridisch jargon in rechterlijke uitspraken te hebben om in staat te zijn dit te doen.⁵

Annotatietool en handleiding

Het annoteerwerk is verricht in een speciaal hiervoor gemaakte tool (zie Figuur 3.1 in de Haalbaarheidsrapportage in Bijlage 4) en er is gebruik gemaakt van een handleiding met nadere toelichting bij de te classificeren kenmerken (in welk deel van de tekst is het kenmerk i.h.a. te vinden en waaraan zijn de categorieën te herkennen, welke combinaties kunnen wel of juist niet voorkomen, etc.). Voor de handleiding is input geleverd door enkele inhoudelijk deskundigen uit de Begeleidingsgroep.

3.3 Kwaliteit van de annotaties en de realisatie van annotatiesets

Mate van overeenstemming tussen de annotateurs

Nadat alle voor de trainingsets geselecteerde uitspraken onafhankelijk van elkaar door drie personen zijn geannoteerd is de mate van overeenstemming/consensus tussen de annotateurs berekend. De resultaten hiervan staan in Tabel 3.2 en 3.3. Voor de rechtbankuitspraken kon een deel van de in de haalbaarheidsfase uitgevoerde annotaties worden hergebruikt. In Tabel 3.2 staan echter alleen de resultaten van de nieuw uitgevoerde

⁵ Vooraf is in verschillende sessies door enkele experts uit de begeleidingsgroep en van Min BZK input aan de onderzoekers van het CBS geleverd over welke mogelijke categorieën bij de te onderzoeken kenmerken te onderscheiden zijn en hoe deze categorieën te herkennen zouden moeten zijn. Daar waar onzekerheid was over de kwaliteit van de annotaties, is bovendien achteraf door experts meegekeken bij uitspraken waar de drie annotateurs het niet met elkaar eens waren – deze experts traden dan op als extra annotateur en kregen daarbij een dubbel gewicht. Omdat een hoge kwaliteit van de trainingsets ook de uiteindelijke uitkomsten van het onderzoek ten goede zal komen, wordt het aanbevolen voor een vervolg op deze eerste nulmeting bij het uitbreiden van de annotatiebestanden (met nieuwe zaken en aanvullende kenmerken) de annotateurs vooraf te laten trainen door inhoudelijk experts.

annotaties weergegeven, omdat in de haalbaarheidsfase bij verschillende kenmerken in meer categorieën werd geïnclassificeerd dan voor deze eerste nulmeting is gedaan. De mate van overeenstemming is daardoor niet goed vergelijkbaar. Zie eventueel Paragraaf 3.2 van de Haalbaarheidsrapportage in Bijlage 4 voor informatie over de mate van overeenstemming tussen de annoteurs bij de haalbaarheidsfase.

Zoals uit de twee tabellen hieronder valt op te maken, bleken niet alle kenmerken even eenduidig te categoriseren voor de annoteurs. De overeenstemming bleek zowel bij de rechtbankuitspraken als de AbRvS-uitspraken het hoogst bij het Aantal eisers/appellanten en het Soort kamer. Bij deze kenmerken waren alle drie de annoteurs het in ruim 90 tot ruim 95 procent van de gevallen met elkaar eens. Dit heeft mogelijk deels te maken met het feit dat deze beide kenmerken slechts drie categorieën kenden: *één* versus *meerdere* eisers/appellanten en *enkelvoudige* versus *meervoudige* kamer, en daarnaast een categorie *onzeker*. De kans dat alle drie de annoteurs een andere categorie kiezen is deze is bij deze kenmerken dus op voorhand al kleiner dan bij bijvoorbeeld Geschilsoort (bij de AbRvS-uitspraken waren hier 5 mogelijke categorieën). De kenmerken die het lastigst bleken te categoriseren voor de annoteurs, althans, waar de annoteurs het minste overeenstemming over hadden, waren de vraag of de eisers zich *wel* of *niet* lieten Vertegenwoordigen (in beide sets 78% volledige consensus) en de vraag naar het Soort geschil (81% volledige consensus bij de rechtbankuitspraken en 80% bij de uitspraken door de AbRvS). Het annoteren bij de Vertegenwoordiging van de eiser(s)/appellant(en) was lastiger omdat deze informatie niet altijd op dezelfde plek in de tekst van de uitspraak staat. Soms staat dit direct in de Kop, onder de eisers, soms pas in het Procesverloop bij de informatie over de zitting. Bij Geschilsoort was de lagere consensus vooral het gevolg van de complexiteit van het gedeelte van de tekst waar dit kenmerk te vinden was. Hoewel de informatie in het algemeen in het tekstdeel met het Procesverloop stond, was het hier niet altijd even eenvoudig uit te filteren. De annoteurs hebben hier relatief veel aantekeningen bij gemaakt waarin ze hun twijfel uitten over de gekozen categorie en bij twijfel werd de waarde overig/onzeker gekozen.

TABEL 3.2. Mate van overeenstemming annoteurs rechtbankuitspraken (n=268)

Kenmerk	Mate van overeenstemming (%)		
	3/3 eens	2/3 eens	oneens
Is het omgevingsrecht en is het geen voorlopige voorziening?	91,8	7,5	0,7
Geschilsoort	81,0	16,4	2,6
Aantal eisers	94,4	5,6	0,0
Soort eiser(s)	90,3	8,6	1,1
Vertegenwoordiging eiser(s)	77,6	20,9	1,5
Soort verweerder	91,4	7,1	1,5
Soort kamer	95,1	3,7	1,1
Uitkomst geschil ^a	86,9	11,2	1,9

^a Hierbij konden alle combinaties van gegrond, ongegrond, niet-ontvankelijk en onbevoegd voorkomen.

TABEL 3.3. Mate van overeenstemming annoteurs AbRvS-uitspraken (n=446)

Kenmerk	Mate van overeenstemming (%)		
	3/3 eens	2/3 eens	oneens
Is het omgevingsrecht en is het geen voorlopige voorziening?	89,7	10,3	0,0
Geschilsoort	79,8	16,8	3,4
Aantal appellanten	90,4	9,6	0,0
Soort appellant(en)	86,1	12,6	1,3
Vertegenwoordiging appellant(en)	78,5	19,7	1,8
Soort verweerder	87,9	11,0	1,1
Soort kamer	93,9	6,1	0,0
Uitkomst geschil ^a	82,3	15,2	2,5

^a Hierbij konden alle combinaties van gegrond, ongegrond, niet-ontvankelijk en onbevoegd voorkomen.

Omdat in de geautomatiseerde filtering van uitspraken mogelijk nog documenten zijn overgebleven die toch niet onder het omgevingsrecht blijken te vallen óf die uitspraken van de voorzieningenrechter bleken te bevatten, is ook dat in het annotatieproces meegenomen als kenmerk om te scoren. In beide annotatiesets bleek bij zo'n 8 tot 10 procent van de documenten niet alle drie de annoteurs het over deze kwestie eens te zijn. Uitspraken die volgens twee of meer annoteurs niet tot het omgevingsrecht behoorden, danwel die volgens hen uitspraken door de voorzieningenrechter betroffen, zijn niet in de geannoteerde datasets opgenomen. Nadien zijn bovendien deze uitspraken handmatig bekeken om de filtering op het totale bestand verder aan te scherpen.

Realisatie van geannoteerde datasets

Wanneer minimaal twee van de drie annoteurs het eens waren over de waarde van een kenmerk, dan is deze waarde direct opgenomen in de trainingset (het geannoteerde databestand). Daar waar de annoteurs alle drie een andere categorie hadden gescoord, is de uitspraak door een vierde persoon met meer domeinkennis gecontroleerd om een uiteindelijke waarde toe te kennen. Het totaal aantal geannoteerde rechtbankuitspraken en AbRvS-uitspraken is respectievelijk 478 en 446. Van deze uitspraken zijn alle te voorspellen kenmerken gelabeld.

3.4 Modellen maken met behulp van annotatiesets

In deze paragraaf wordt de methode beschreven om de verschillende kenmerken te classificeren, op basis van de geannoteerde uitspraken (zoals beschreven in de vorige paragraaf).⁶ Deze paragraaf is relatief technisch van aard en vooral bedoeld om ingewijden in machine learning een transparant overzicht te geven van de stappen die zijn doorlopen en de keuzes die daarin zijn gemaakt.

Verwerking uitspraken

Voor de classificatie met behulp van machine learning zijn de tekstelementen van de rechterlijke uitspraken vertaald naar vectoren met getallen. Conform de haalbaarheidsfase zijn de teksten eerst voorbereid.⁷ Hierbij is gebruikt gemaakt van Bag-of-Words in combinatie met tf-idf (afkorting van 'Term Frequency – Inversed Document Frequency').⁸ Deze methodes worden toegelicht in Paragraaf 1.4 van de Haalbaarheidsrapportage

⁶ De auteurs zijn dank verschuldigd aan dr. Arnout van Delden, senior methodoloog bij het CBS, voor zijn inbreng bij dit onderdeel van het onderzoek.

⁷ Anders dan in de haalbaarheidsfase zijn getallen vervangen door 'XXX', omdat het feit dat er getallen in de teksten staan informatie bevat. Daarnaast zijn komma's bij sommige kenmerken niet verwijderd, aangezien het aantal komma's informatie kan bevatten over bijv. het aantal eisers.

⁸ Andere opties voor input-data van de modellen zijn andere geclassificeerde kenmerken en het extraheren van karakteristieken door middel van regex. Aanvullende analyses lieten zien dat de kwaliteit van de modellen niet beter werd wanneer andere classificaties meegenomen werden. Om de modellen robuust te houden tegen woordkeuzes heeft regex niet de voorkeur.

(zie Bijlage 4) en bevatten twee belangrijke zogenaamde ‘hyperparameters’: het aantal opeenvolgende woorden dat gezamenlijk wordt geanalyseerd (Ngram) en het maximale aantal combinaties van woorden.

Modelkeuze

Per onderzocht kenmerk is een apart classificatiemodel geschat voor de rechtbank-uitspraken en de AbRvS-uitspraken. Hiervoor is gekozen vanwege de verschillen tussen categorieën van kenmerken van rechtbankuitspraken en uitspraken door de AbRvS en de verschillen in tekstelementen waar de kenmerken beschreven worden. Soms komt het ook voor dat de relevante informatie in meerdere (opeenvolgende) tekstelementen kan staan, zoals bij ‘Soort kamer’.⁹

Op basis van de annotaties en verwerkte uitspraken zijn twee types machine-learning-modellen getraind: Random Forest (RF) en Support-Vector Machines (SVM) met lineaire kernel. Bij RF bepalen twee hyperparameters de grootte van het model: het aantal beslissingsbomen en de diepte van de beslissingsbomen. Bij SVM’s zijn de veelgebruikte loss functies ‘hinge loss’ en ‘squared hinge loss’ vergeleken.¹⁰ Per kenmerk is allereerst het type model bepaald met behulp van 5-fold cross validation. Bij 5-fold cross validation wordt de trainingset gesplitst in vijf even grote delen. Voor elk van de vijf delen worden steeds vijf minus één delen (80%) gebruikt om het model te trainen en één deel (20%) om het model te valideren. De nauwkeurigheid is vervolgens het gemiddelde over de vijf modellen. Wanneer het type model is bepaald, worden de hyperparameters afgesteld op basis van de accuracy. De accuracy is de proportie correcte categoriseringen in een kenmerk. Ofwel: de som van het aantal modelschattingen dat overeenkomt met de werkelijkheid gedeeld door het totaal aantal modelschattingen.

Wanneer het model de eigenaardigheden van de trainingset (de geannoteerde datasets) te nauw volgt, geeft het model geen goede beschrijving van de echte ontwikkelingen, maar eerder een product van de eigenaardigheden in de geannoteerde dataset. Dit fenomeen wordt ‘overfitting’ genoemd. Het model classificeert dan heel nauwkeurig de trainingset, maar doet dit niet betrouwbaar voor de overige data. Om dit te voorkomen is ervoor gekozen om de modellen niet te flexibel te laten zijn door restrictieve waardes van hyperparameters te gebruiken.

Nauwkeurigheid

Om informatie te verkrijgen over de nauwkeurigheid van de modelschattingen, zijn op basis van de confusion matrices (waarin per kenmerk of per categorie de voorspelde/geschatte frequenties zijn uitgezet tegen de werkelijke frequenties en waarmee het aantal correct voorspelde classificaties en het aantal foutief voorspelde classificaties in beeld wordt gebracht), naast de hierboven genoemde accuracy, ook de balanced accuracy en de F1-score per kenmerk berekend. De balanced accuracy houdt rekening met het aantal annotaties per categorie. Het verschil tussen de accuracy en balanced accuracy geeft derhalve informatie over de disbalans tussen de categorieën. Omdat het van belang is dat alle categorieën goed geclassificeerd worden, is deze metriek relevant. De F1-score is een functie van de precision¹¹ en de recall¹² en deze maat houdt ook rekening met disbalans tussen de categorieën.

⁹ De relevante informatie over het soort kamer kan naast de Kop ook voorkomen in het Procesverloop of (bij RvS-uitspraken) in de Conclusie. Om ruis te voorkomen zijn alleen de zin of zinnen waarin ‘kamer’, ‘enkelvoud’ of ‘meervoud’ in de voorbereekte procesverloop-tekst voorkomt meegenomen.

¹⁰ De volgende vijf hyperparameters zijn gebruikt: aantal woordcombinaties (met waardes 20, 50 of 100), ngram (met waardes 1, 2, of 3), aantal beslissingsbomen (met waardes 50 of 100), diepte beslissingsbomen (met waardes 2 of 3), loss functie (SVM hinge loss of squared hinge loss).

¹¹ Het aantal keren dat een categorie correct als aanwezig wordt voorspeld (de zogenaamde true positives), gedeeld door het totaal aantal keren dat een categorie als aanwezig wordt voorspeld (de som van de true positives en de false positives).

¹² Het aantal keren dat een categorie correct als aanwezig wordt voorspeld (true positives), gedeeld door het werkelijk aantal keren dat een categorie voorkomt (de som van de true positives en de false negatives).

Tabel 3.4 geeft de (balanced) accuracy en *F1*-score per kenmerk van rechtbankuitspraken en AbRvS-uitspraken aan. Voor alle drie de maatstaven geldt dat deze waardes aannemen tussen 0 en 1, waarbij 1 een perfecte score zou zijn. Uit Tabel 3.4 blijkt dat Geschilsoort een lastig te schatten kenmerk is, terwijl Soort kamer juist goed te schatten is.

TABEL 3.4. Nauwkeurigheid van de modellen

	Rechtbankuitspraken			Uitspraken AbRvS		
	Accuracy	Balanced accuracy	<i>F1</i> -score	Accuracy	Balanced accuracy	<i>F1</i> -score
Geschilsoort	0,79	0,86	0,76	0,83	0,87	0,81
Aantal eisers/appellanten	0,97	0,98	0,98	0,98	0,99	0,99
Soort eiser(s)/appellant(en)	0,86	0,86	0,79	0,95	0,96	0,96
Vertegenwoordiging eiser(s)/appellanten	0,91	0,93	0,87	0,89	0,89	0,90
Soort verweerders	0,97	0,96	0,90	0,94	0,93	0,90
Soort kamer	1,00	0,96	0,98	1,00	1,00	1,00
Uitkomst geschil gegrond	0,96	0,98	0,97	0,98	0,99	0,99
Uitkomst geschil ongegrond	0,96	0,97	0,96	0,91	0,94	0,93
Uitkomst geschil niet-ontvankelijk	0,98	0,99	0,98	0,96	0,99	0,99

3.5 Implicaties van de gekozen onderzoeksmethodiek

Splitsing uitspraakteksten

Doordat de uitspraken voorafgaand van het trainen van de modellen zijn gesplitst in tekstelementen, bevatte de input-data van het model weinig ruis. Wanneer een kenmerk in een uitspraaktekst echter toevallig in een ander tekstelement wordt beschreven, kan het model deze uitspraak niet goed classificeren. Als de (volgorde van) tekstelementen in rechterlijke uitspraken in de toekomst wijzigen, heeft dit gevolgen voor de kwaliteit van de classificaties. Het is dus relevant om dit aspect in de toekomst goed te monitoren.

Restcategorieën

Voor de kenmerken Soort kamer, Geschilsoort en Soort verweerder is een restcategorie *Onzeker/overig* toegevoegd. Bij het Soort kamer bestaat deze categorie volledig uit uitspraken waarin geen informatie over het Soort kamer staat - de categorieën *enkelvoudig* en *meervoudig* zijn hier uitputtend. Voor Geschilsoort zijn de categorieën niet volledig uitputtend en bestaat categorie *Onzeker/overig* uit uitspraken die niet in de andere categorieën vallen (o.a. enkele *bestemmingsplanbesluiten* bij rechtbankuitspraken) óf waarbij (eenduidige) informatie over de aard van het bestreden besluit (geschilsoort) ontbrak. Ook is hier de categorie *Tracébesluit* (bij AbRvS-uitspraken) toegevoegd, omdat deze in de trainingset te weinig voor bleek te komen om dit betrouwbaar te kunnen schatten. Bij Soort verweerder bevat *Onzeker/overig* zowel uitspraken die de annoteurs reeds als onzeker markeerden, als de categorieën *particulier*, *bedrijf*, *waterschappen*, *stichting*, *vereniging* en *verschillende* (bij meerdere verweerders) die allen te weinig uitspraken bevatten. Als er voor een vervolgonderzoek meer geannoteerd zou worden, dan zouden deze categorieën mogelijk alsnog nader geclassificeerd kunnen worden. In Paragraaf 3.6 van de Haalbaarheidsrapportage (zie Bijlage 4) werd deze beperking reeds beschreven.

Onderlinge samenhang van kenmerken en classificaties

Het aantal mogelijke combinaties van kenmerken is groot; de onderlinge samenhang van classificaties in deze eerste nulmeting is nog niet grondig onderzocht. Een geclassificeerd kenmerk meenemen als input-data voor een model van een ander kenmerk zou mogelijk kunnen leiden tot een beter resultaat van de modelschatting (modelfit). Bij de uitkomst van het geschil, dat in drie separate kenmerken is gesplitst (Gegrond, Ongegrond en

Niet-ontvankelijk), is wel geprobeerd de onderlinge samenhang mee te nemen in de classificaties, maar dit leidde niet tot een betere nauwkeurigheid dan wanneer geen rekening werd gehouden met deze onderlinge samenhang. Tot slot kan door de onderlinge samenhang mee te nemen een betere schatting van de totale combinatie van kenmerken per uitspraak worden gemaakt. In Paragraaf 3.5 van de Haalbaarheidsrapportage (zie Bijlage 4) wordt uitgelegd waarom dit belangrijk is.¹³

Onzekerheidsmarges en mogelijke vertekening over de tijd

Het doel van de modellen is om zo goed mogelijke individuele uitspraken classificeren. Uit de confusion matrices blijkt echter dat het aantal uitspraken in een bepaalde categorie soms over- of onderschat wordt. Wanneer de fouten bij het voorspellen niet tegen elkaar uitmiddelen bij het aggregeren, ontstaat een vertekening in de geschatte kenmerken over de tijd. De schattingen hebben bovendien altijd een mate van onzekerheid, waar in Hoofdstuk 4 verder op in wordt gegaan. Het aantal annotaties is relatief klein. Hoewel er geen harde vuistregel is met betrekking tot de minimale grootte van trainingsets, geldt wel dat hoe groter de beschikbare trainingset, hoe nauwkeuriger de mate van vertekening bepaald kan worden en hoe beter hier vervolgens voor gecorrigeerd kan worden. Wanneer sprake is van veel mogelijke categorieën, die bovendien niet in dezelfde mate voorkomen in de rechterlijke uitspraken, zijn meer annotaties nodig voor nauwkeurige en betrouwbare schattingen, dan wanneer de te onderzoeken kenmerken slechts uit twee mogelijke categorieën bestaan die bovendien ieder voorkomen in ongeveer de helft van de uitspraken. Naast dat meer annotaties leidt tot meer nauwkeurigheid in het bepalen van de vertekening, zal het ook de classificatiemodellen ten goede komen. Om complexere kenmerken, zoals Geschilsoort, te leren onderscheiden heeft het model voldoende annotaties nodig.

Analysemogelijkheden op verrijkt microbestand

De classificaties van kenmerken die in deze eerste nulmeting zijn meegenomen worden opgeslagen in een verrijkt microbestand, waarin ieder kenmerk voor iedere uitspraak is geclassificeerd. Op dit bestand zijn eerste, beperkte analyses mogelijk. Multivariate analyses zijn in deze fase nog een stap te ver. In de eerste plaats omdat de classificering van elk kenmerk geschat is, en dus een bepaalde onzekerheid bevat. In de tweede plaats omdat, zoals genoemd, bij het schatten geen rekening is gehouden met de onderlinge samenhang tussen kenmerken. Analyses op de verrijkte micro-data naar die samenhang zou gemakkelijk tot foutieve conclusies kunnen leiden. Een logische vervolgstap zou zijn om voor de bestaande kenmerken en eventuele nieuwe kenmerken, met input van juridisch-inhoudelijk deskundigen (over logische samenhang) de modellen verder te verfijnen.

¹³ Om een indruk te krijgen van de onderlinge samenhang is in april 2020 een aantal omgevingsrecht-experts benaderd hun hypotheses hieromtrent te beschrijven. Het zou waardevol zijn de kennis die hierover in het veld bestaat te gebruiken voor toekomstige modelleringen.

4. Resultaten

In dit hoofdstuk worden de resultaten getoond van de geschatte ontwikkelingen naar jaar van de gelabelde kenmerken van escalaties in het omgevingsrecht in deze nulmeting. De cijfers worden getoond als relatieve percentages; in figuren is per indicator per jaar weergegeven wat het geschatte aandeel (in %) is van elke categorie in het totaal aantal uitspraken in dat jaar. Deze figuren geven derhalve geen inzicht in ontwikkelingen in het totaal aantal uitspraken. In Paragraaf 4.1 worden de geschatte kenmerken getoond van de *gedigitaliseerde* rechtbankuitspraken en in Paragraaf 4.2 van de AbRvS-uitspraken in de onderzochte periodes. De geschatte percentages waarop de figuren in beide paragrafen zijn gebaseerd zijn opgenomen in tabelvorm in Bijlage 3. In Bijlage 3 zijn bovendien de geschatte absolute aantallen weergegeven, waarmee inzicht kan worden verkregen in het verloop van het aantal uitspraken en voorkomen van de verschillende categorieën van kenmerken over de onderzochte periode.

Benadrukt moet worden dat alle figuren zijn gebaseerd op geschatte percentages. Hierbij is dus sprake van een bepaalde mate van onzekerheid. Op de nauwkeurigheid van de huidige modelschattingen wordt in Paragraaf 4.3 ingegaan.

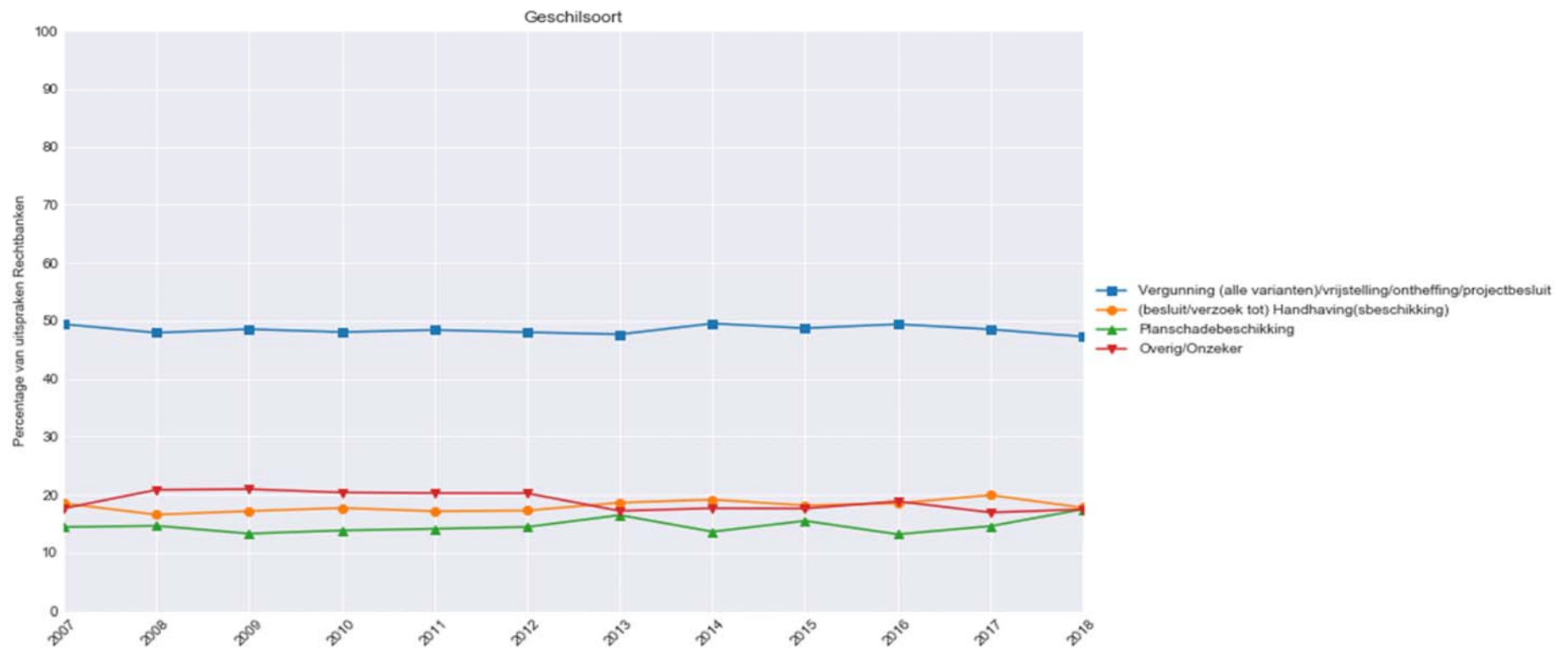
4.1 Geschatte kenmerken gedigitaliseerde rechtbankuitspraken omgevingsrecht naar jaar

Geschilsoort

Figuur 4.1 toont voor alle gedigitaliseerde rechtbankuitspraken in het omgevingsrecht in de periode 1-1-2007 tot en met 31-12-2018 de schatting van het geclassificeerde kenmerk 'geschilsoort'. Het gaat hier om de aard van het onderliggend besluit dat door de eiser wordt bestreden. De hierbij onderscheiden categorieën zijn *vergunningbeschikkingen* (alle varianten en inclusief vrijstellingen, ontheffingen, projectbesluiten, milieumeldingen en aanwijzingsbesluiten ORAC), *handhavingsbeschikkingen*, *planschadebeschikkingen* en een categorie *overig/onzeker* waaronder o.a. conflicten met betrekking tot bestemmingsplannen vielen. Informatie over dit kenmerk is in het algemeen te vinden in het Procesverloop.

Geschat wordt dat bijna de helft van de gedigitaliseerde rechtbankuitspraken betrekking heeft op conflicten over vergunningbeschikkingen en dat dit voor zowel handhavingsbeschikkingen als planschadebeschikkingen tussen de 10 en 20 procent is. Het aandeel geclassificeerde uitspraken in de restcategorie 'overig/onzeker' is, met zo'n 20 procent, opvallend hoog, ook nog nadat er met hulp van een aanscherping van de annotaties in deze categorie heeft plaatsgevonden.

FIGUUR 4.1. Geschilsoort: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

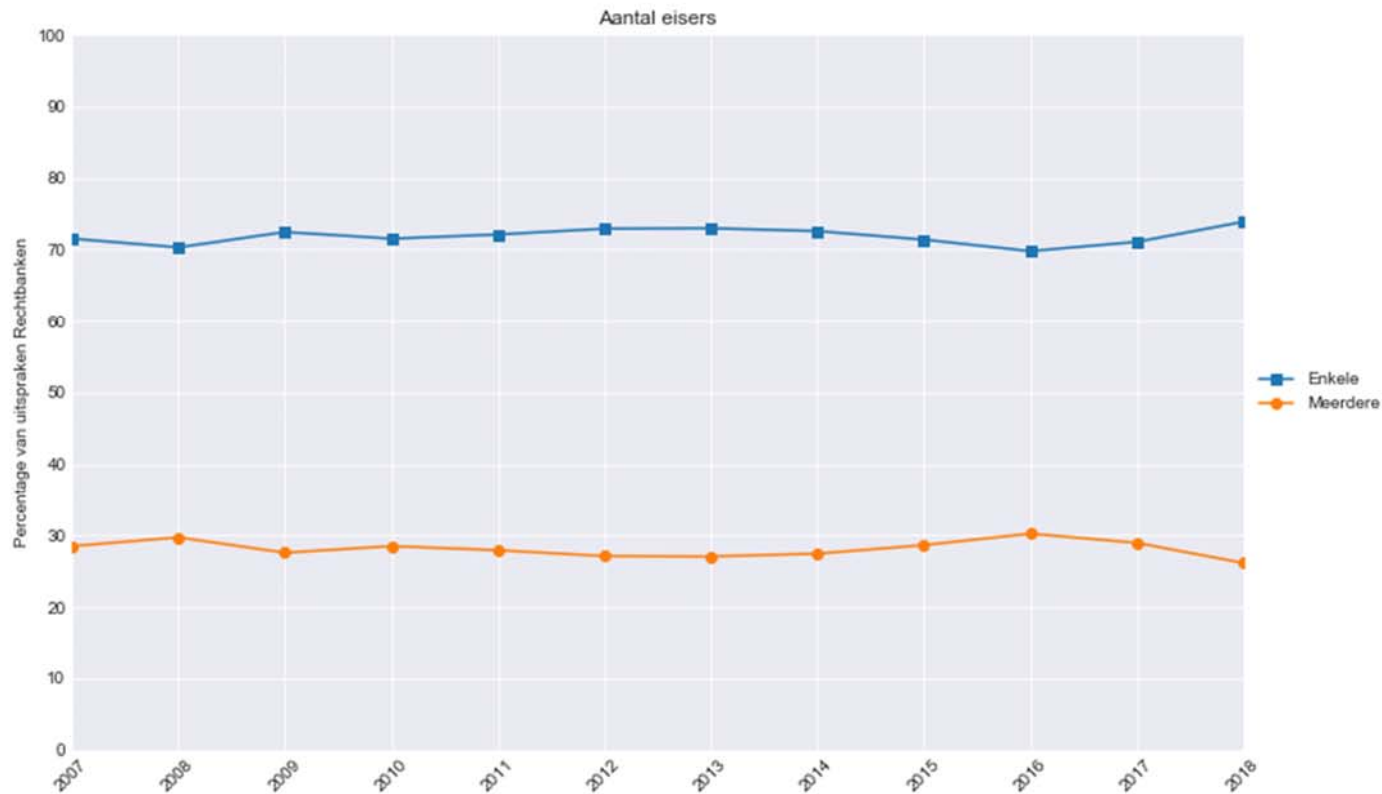


Aantal eisers

Figuur 4.2 toont de schatting van het geclassificeerde kenmerk 'aantal eisers'. De hierbij onderscheiden categorieën zijn *één* of *meerdere*. Informatie over dit kenmerk is te vinden in de Kop van de uitspraken.

Geschat wordt dat bij ruim 7 op de 10 gedigitaliseerde rechtbankuitspraken sprake is van een enkele eiser. Dit aandeel lijkt in de meest recente jaren licht toe te nemen.

FIGUUR 4.2. Aantal eisers: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

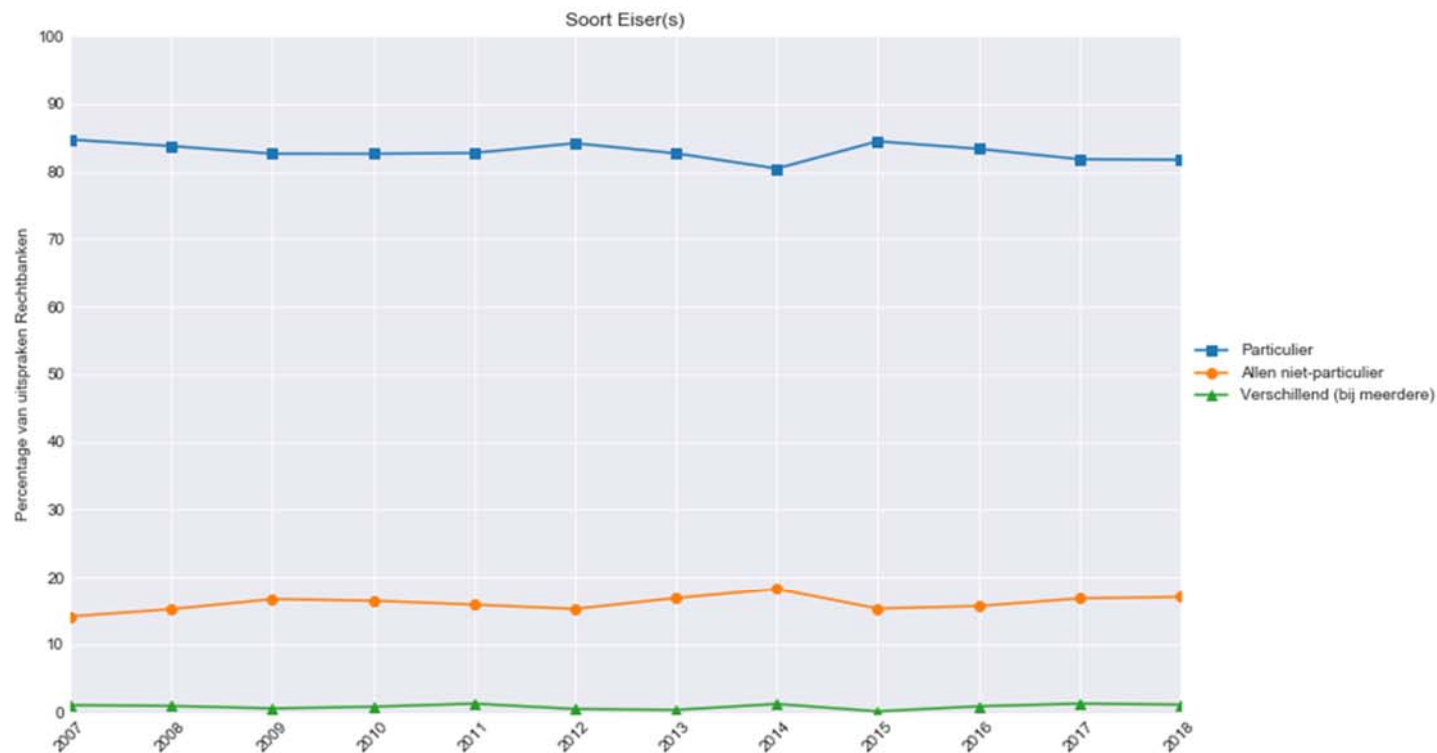


Soort eiser(s)

Figuur 4.3 toont de schatting van het geclassificeerde kenmerk 'soort eiser(s)'. De hierbij onderscheiden categorieën zijn *(allen) particulier*, *(allen) niet particulier* of *verschillend* (bij zowel particuliere en niet-particuliere eisers; gemengd dus). Informatie over dit kenmerk is te vinden in de Kop van de uitspraken. Alleen wanneer enkel eigennamen werden genoemd bij de eiser(s), werd dit kenmerk door de annoteurs beschouwd als *particulier*. Wanneer bij een eigenaam ook een bedrijfsnaam of bijvoorbeeld een stichting werd genoemd, waar de betreffende persoon bijvoorbeeld directeur van was, dan werd dit gezien als *niet particulier*. Dit kenmerk meet dus niet het onderscheid tussen natuurlijke personen en rechtspersonen.

Geschat wordt dat bij ruim 4 op de 5 gedigitaliseerde rechtbankuitspraken sprake is van (een) particuliere eiser(s).

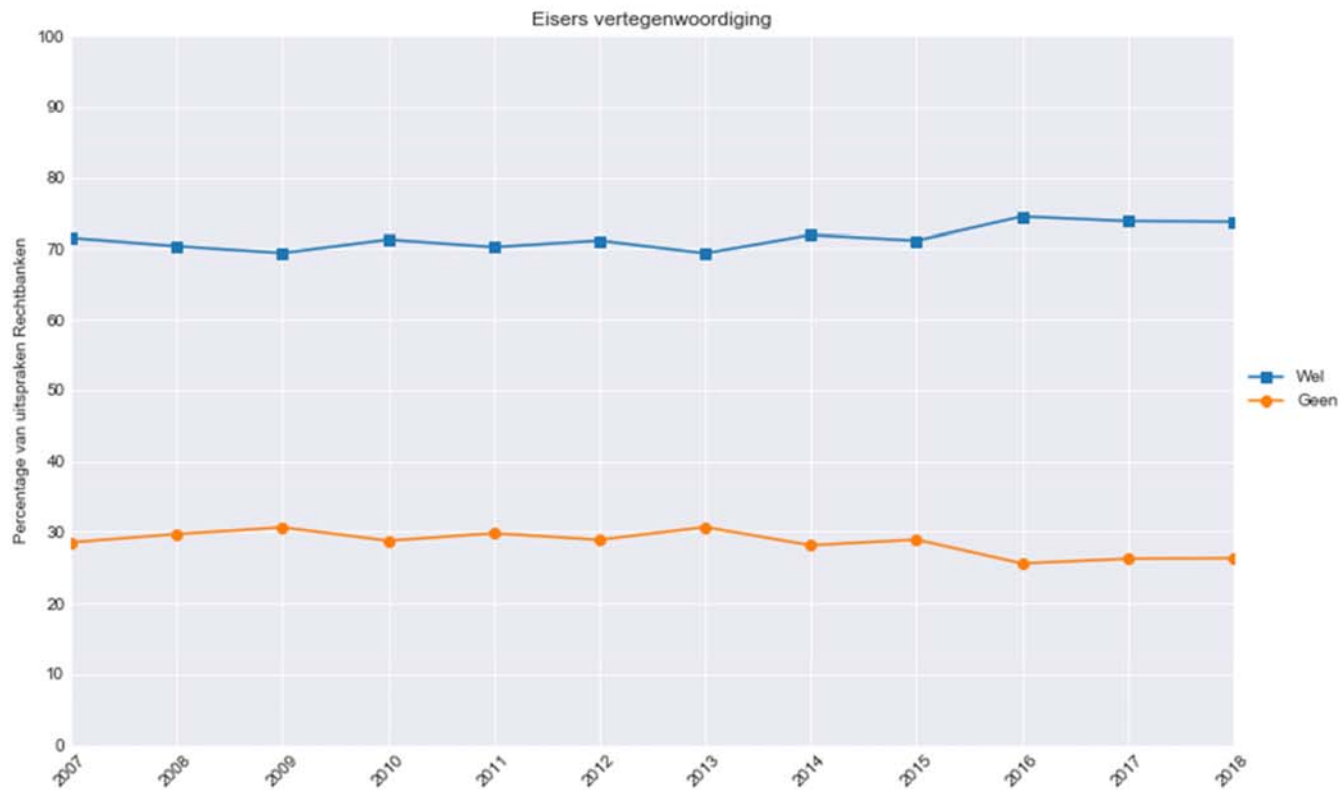
FIGUUR 4.3. Soort eiser(s): geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018



Vertegenwoordiging eiser(s)

Figuur 4.4 toont de schatting van het geclassificeerde kenmerk 'vertegenwoordiging eiser(s)'. De hierbij onderscheiden categorieën zijn *wel* of *geen*. Een vertegenwoordiger kan een advocaat zijn, maar dit hoeft niet het geval te zijn. Ook wanneer sprake was van een gemachtigd familielid, werd dit door de annoteurs gescoord als *wel* vertegenwoordiging. Informatie over dit kenmerk is soms te vinden in de Kop van de uitspraken onder de eiser, soms is het te vinden in het Procesverloop. Geschat wordt dat bij ruim 7 op de 10 gedigitaliseerde rechtbankuitspraken de eiser zich door iemand heeft laten vertegenwoordigen. Dit aandeel schommelt in de meest recente jaren rond de 75 procent.

FIGUUR 4.4. Vertegenwoordiging eiser(s): geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

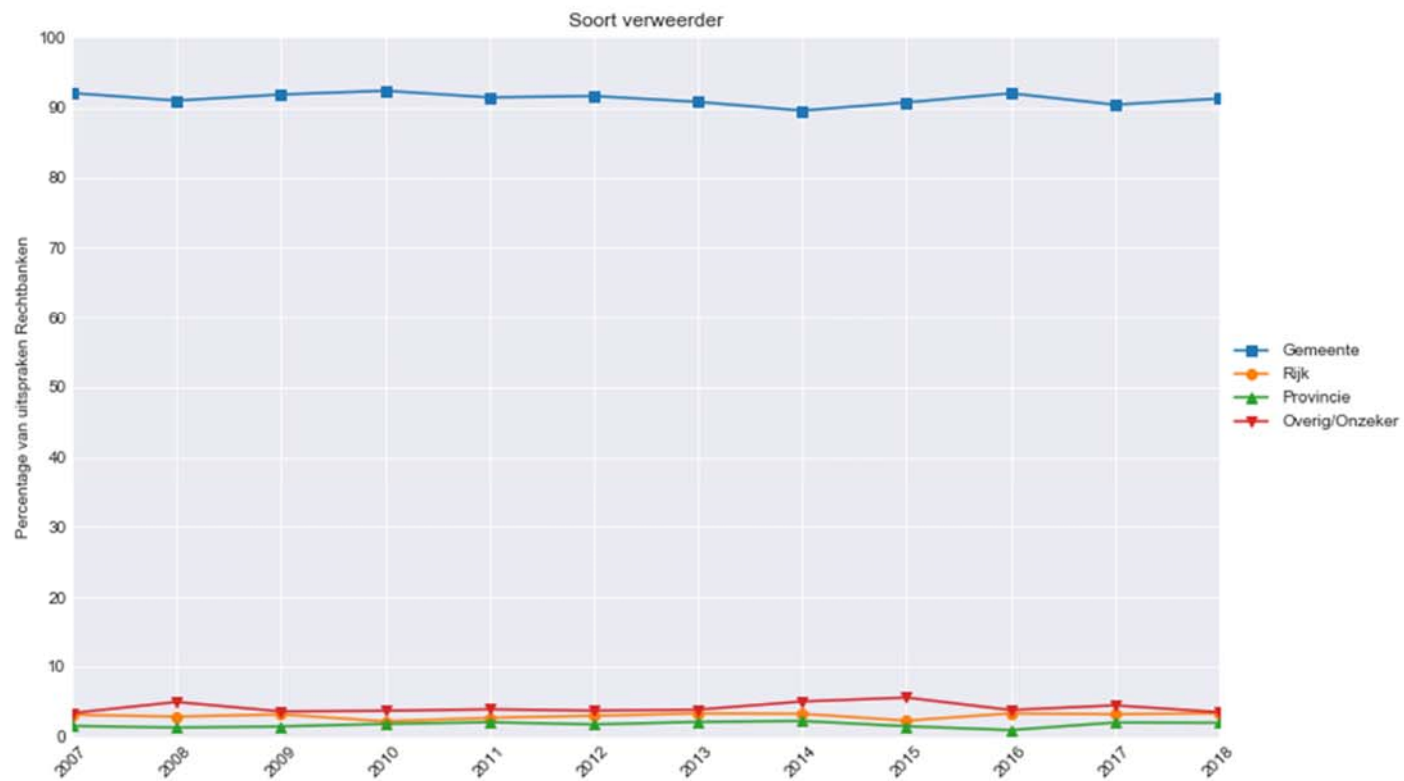


Soort verweerder

Figuur 4.5 toont de schatting van het geclassificeerde kenmerk 'soort verweerder'. De hierbij onderscheiden categorieën zijn *gemeente*, *provincies*, *rijk* en *overig/onzeke*. Het was de bedoeling hier ook *waterschappen* als aparte categorie te onderscheiden, maar deze bleek te weinig voor te komen (slechts 3x in het geannoteerde bestand) om goed te kunnen schatten. Deze categorie is derhalve geschaard onder de restcategorie *overig/onzeke*. Informatie over dit kenmerk is in het algemeen te vinden in de Kop van de uitspraak.

Geschat wordt dat bij ruim 9 van de 10 gedigitaliseerde rechtbankuitspraken de verweerder een gemeente is (meestal betreft dit het college van B&W).

FIGUUR 4.5. Soort verweerder: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

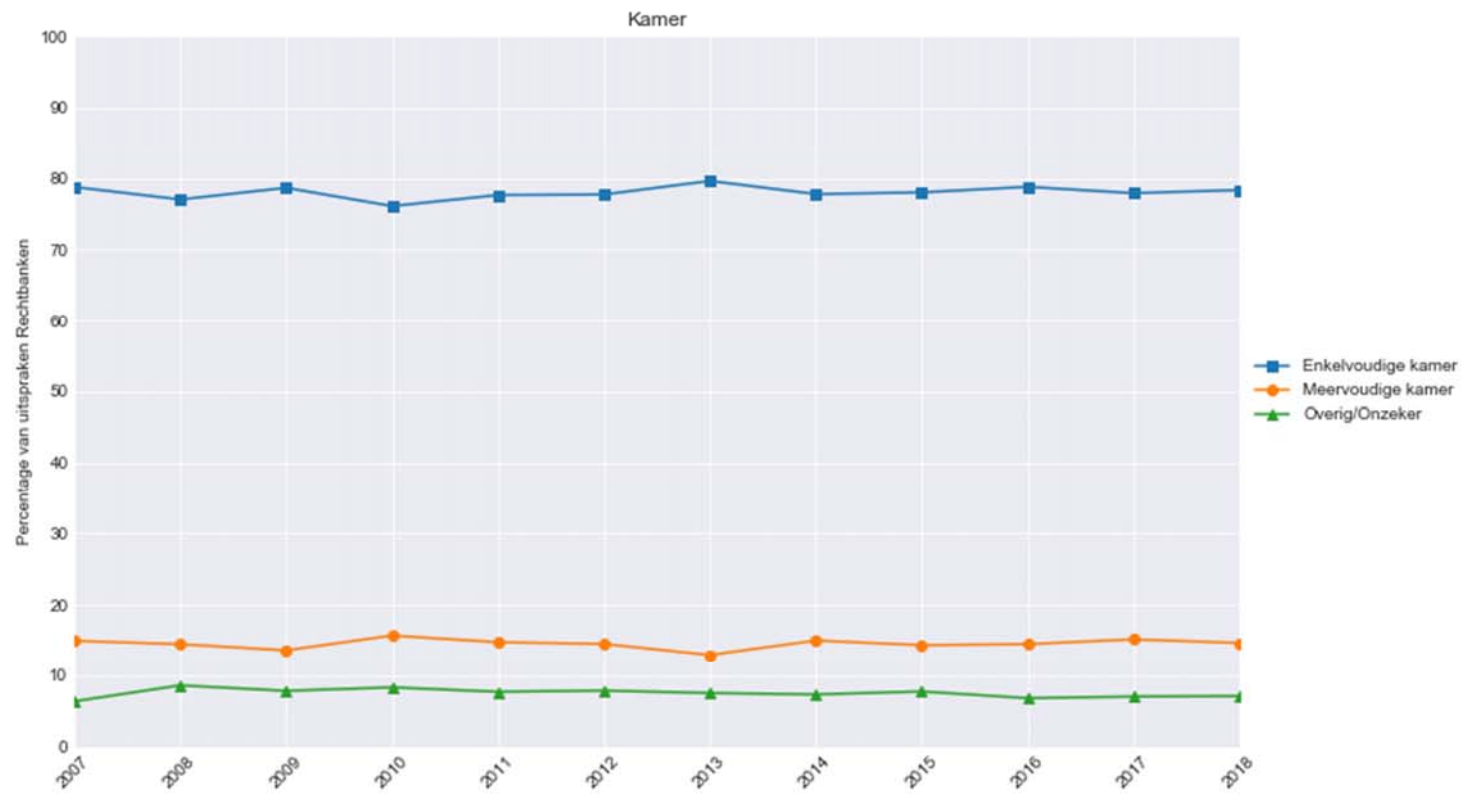


Soort kamer

Figuur 4.6 toont de schatting van het geclassificeerde kenmerk 'soort kamer'. De hierbij onderscheiden categorieën zijn *enkelvoudig* of *meervoudig*. Informatie over dit kenmerk is in het algemeen te vinden in de Kop van de uitspraken.

Geschat wordt dat (bijna) 8 op de 10 gedigitaliseerde rechtbankuitspraken door de enkelvoudige kamer worden gedaan.

FIGUUR 4.6. Soort kamer: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

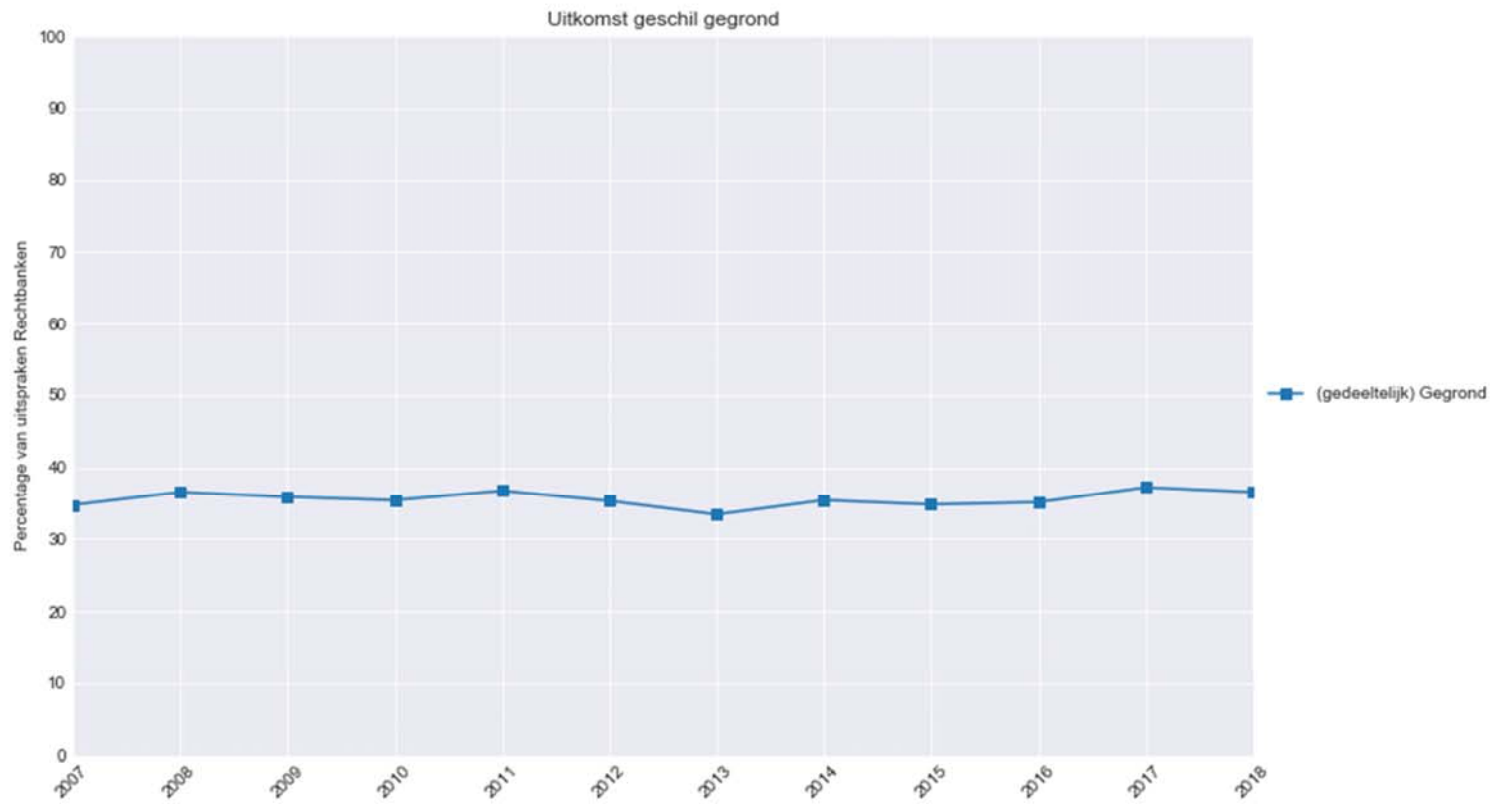


Uitkomst geschil

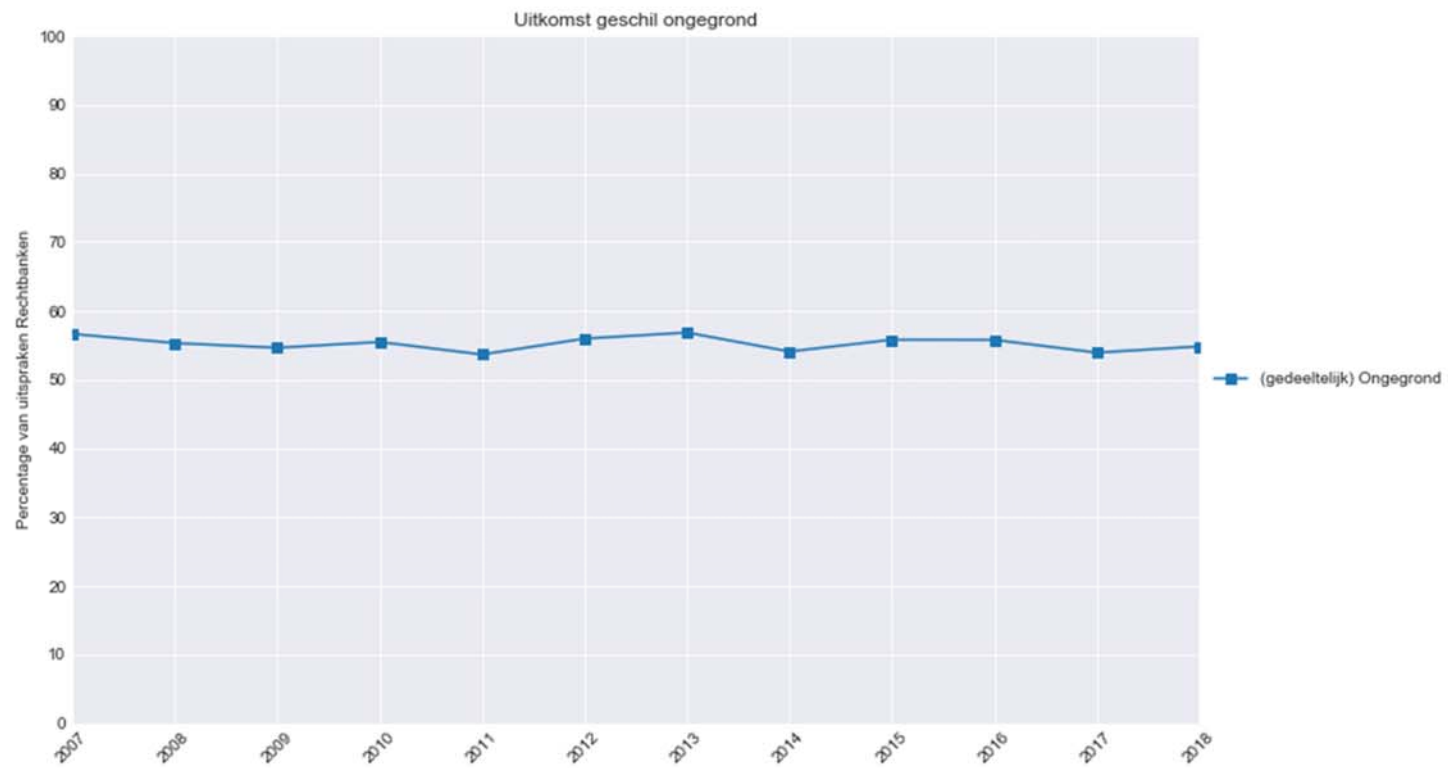
Figuur 4.7a tot en met c tonen de schatting van het geclassificeerde kenmerk 'uitkomst geschil'. Omdat hetgeen wordt aangevochten in een zaak tegelijk gedeeltelijk *gegrond* en gedeeltelijk *ongeground* kan worden verklaard en de eiser(s) tegelijk gedeeltelijk *niet-ontvankelijk* kan/kunnen worden verklaard, zijn deze kenmerken separaat geschat. Het zijn dus eigenlijk drie kenmerken, te weten 'geheel of gedeeltelijk *gegrond*', 'geheel of gedeeltelijk *ongeground*' en 'geheel of gedeeltelijk *niet-ontvankelijk*' en elk kenmerk kan wel of niet van toepassing zijn (0 of 1). Dit zorgt er ook voor dat de percentages in de navolgende drie figuren optellen tot meer dan 100 procent. Wanneer enkel het eerste kenmerk 'geheel of gedeeltelijk *gegrond*' van toepassing is, is hetgeen is aangevochten dus geheel *gegrond*. Is tegelijk sprake van 'geheel of gedeeltelijk *gegrond*' en 'geheel of gedeeltelijk *ongeground*', dan betekent dit dus dat hetgeen is aangevochten deels *gegrond* en deels *ongeground* is bevonden door de rechter(s). Het was de bedoeling hier ook 'onbevoegd' te onderscheiden, maar dit bleek te weinig voor te komen om goed te kunnen schatten. Informatie over dit kenmerk is in het algemeen te vinden in de Conclusie van de uitspraken.

Geschat wordt dat bij 3 tot 4 op de 10 van de gedigitaliseerde rechtbankuitspraken sprake is van een geheel of gedeeltelijke *gegrondverklaring* door de rechter(s). Een aanvullende analyse liet zien dat bij een groot deel van deze uitspraken (ruim 3 van de 10 gedigitaliseerde uitspraken) sprake was van een volledige *gegrondverklaring*. Bij 5 tot 6 op de 10 gevallen is sprake van een geheel of gedeeltelijke *ongegroundverklaring*. Ook hier bleek het merendeel te gaan om volledige-*ongegroundverklaringen* (4 tot 5 op de 10 gedigitaliseerde uitspraken). In ruim 1 op de 10 uitspraken is sprake van (gedeeltelijke) *niet-ontvankelijkheid*.

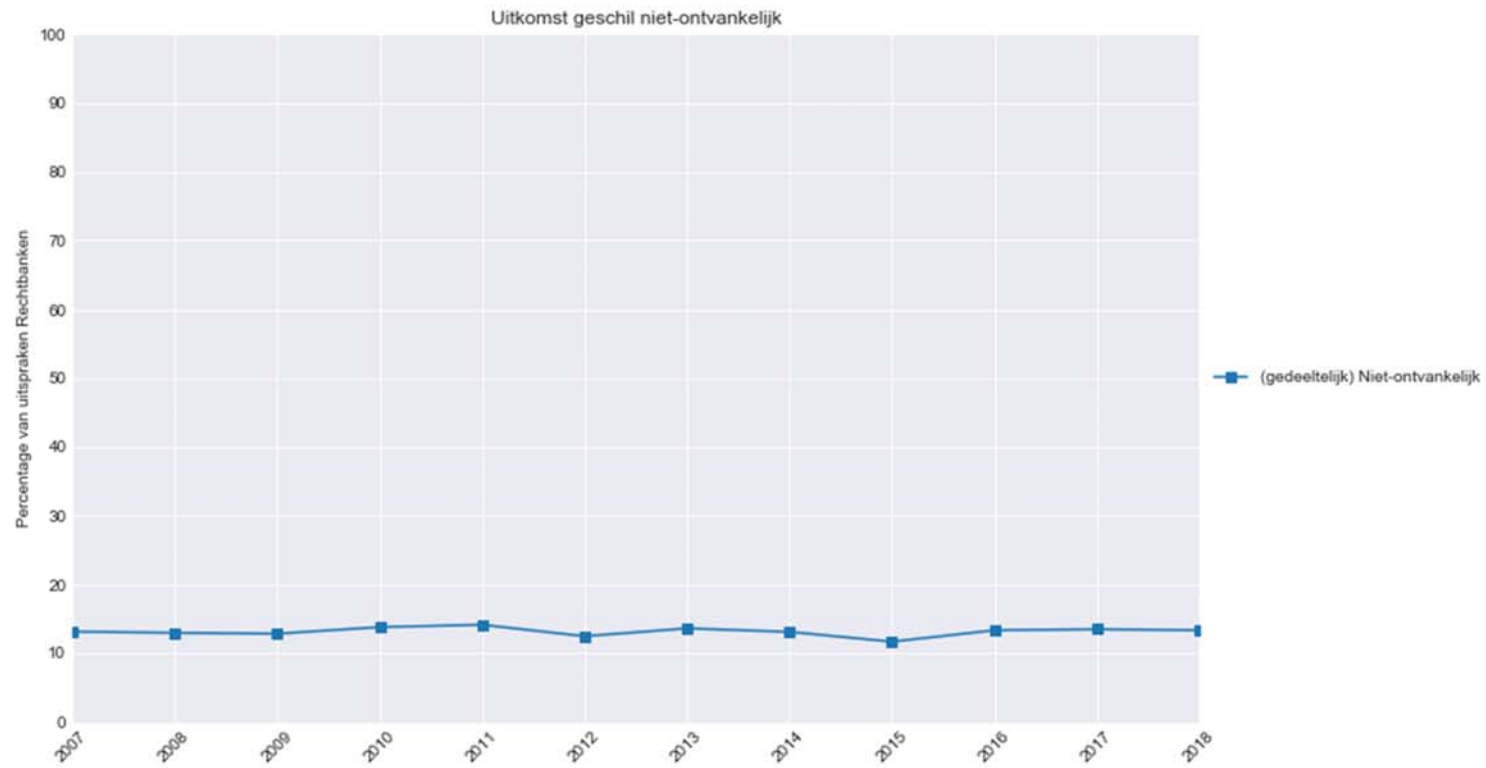
FIGUUR 4.7a. Uitkomst geschil geheel of gedeeltelijk gegrond: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018



FIGUUR 4.7b. Uitkomst geschil geheel of gedeeltelijk ongegrond: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018



FIGUUR 4.7c. Uitkomst geschil geheel of gedeeltelijk niet-ontvankelijk: geschatte ontwikkelingen gedigitaliseerde uitspraken omgevingsrecht rechtbanken 2007-2018

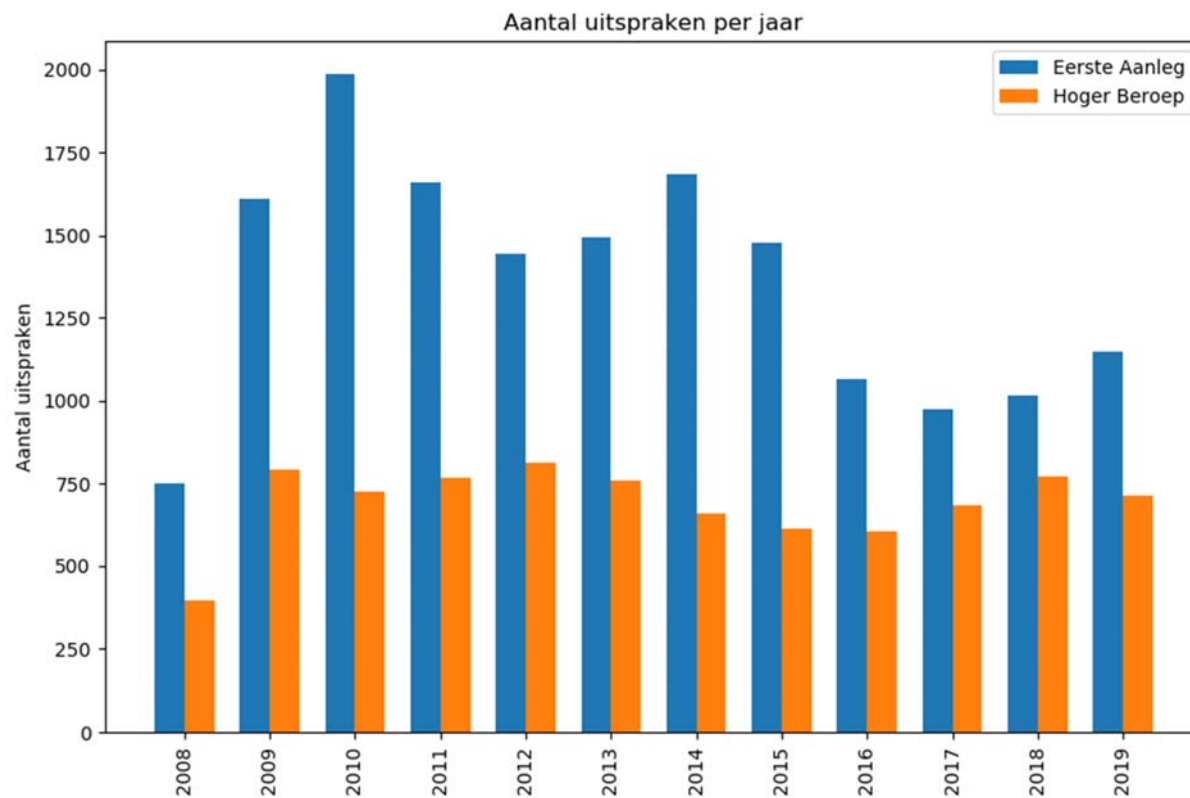


4.2 Geschatte kenmerken AbRvS-uitspraken omgevingsrecht naar jaar

Eerste aanleg en hoger beroep

Anders dan bij de rechtbanken kan bij de AbRvS zowel sprake zijn van beroepszaken in eerste en enige aanleg, als van hoger beroepszaken waarin de eerdere uitspraak van de rechtbank wordt aangevochten. In Figuur 4.8 wordt het aantal AbRvS-uitspraken per jaar uitgesplitst naar eerste (en enige) aanleg (EA) en hoger beroep (HB) weergegeven. Omdat dit kenmerk is aangeleverd als meta-informatie, zijn dit geen schattingen, maar gaat het om feitelijke tellingen op basis van de geleverde data.

FIGUUR 4.8. Eerste (en enige) aanleg en hoger beroep: aantal uitspraken omgevingsrecht AbRvS 2008-2019

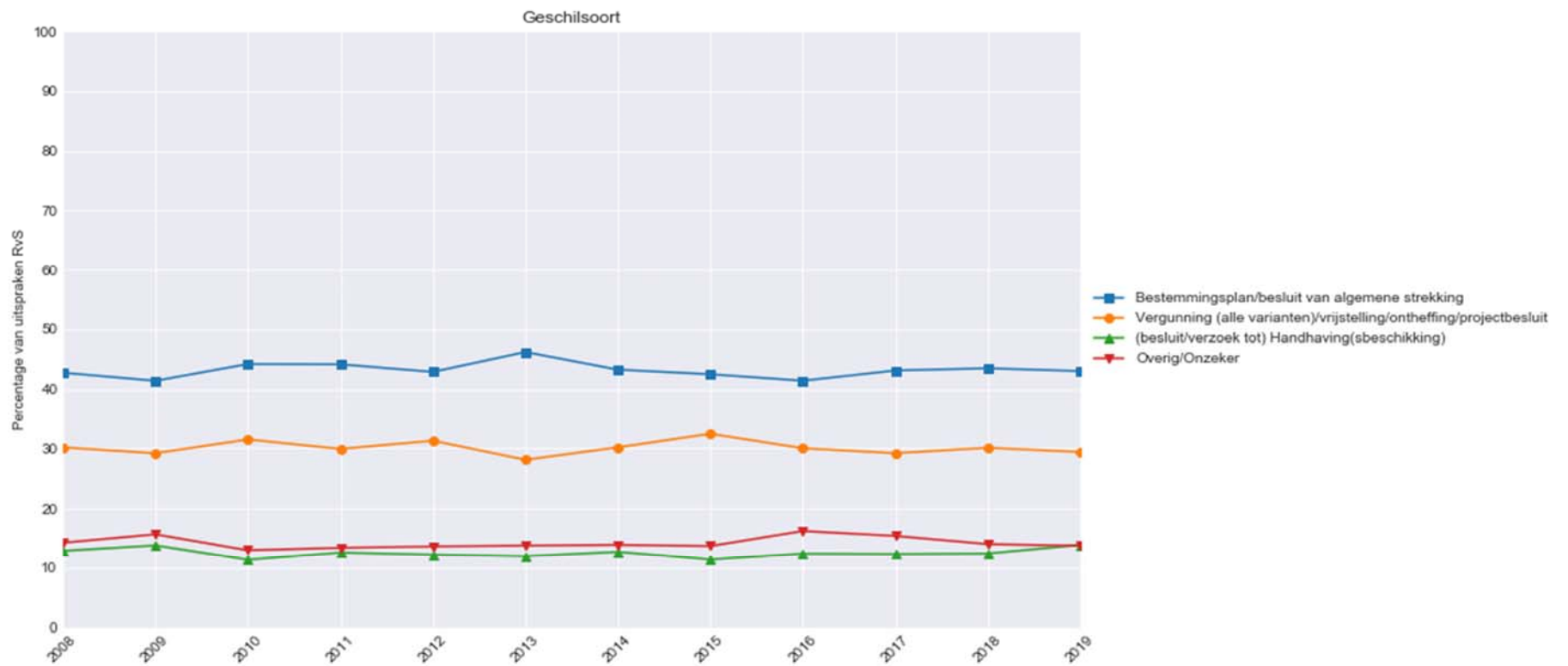


Geschilsoort

Figuur 4.9 toont voor alle AbRvS-uitspraken in het omgevingsrecht in de periode 1-7-2008 tot en met 31-12-2019 de schatting van het geclassificeerde kenmerk 'geschilsoort'. Het gaat hier om de aard van het onderliggend besluit dat door de eiser wordt bestreden. De hierbij onderscheiden categorieën zijn net als bij de rechtbanken *vergunningbeschikkingen* (alle varianten en inclusief vrijstellingen, ontheffingen, projectbesluiten, milieumeldingen en aanwijzingsbesluiten ORAC) welke in het algemeen – maar hierop zijn uitzonderingen – uitspraken in HB betreffen, *handhavingsbeschikkingen (ook vooral uitspraken in HB)* en daarnaast bovendien bestemmingsplannen/besluiten van algemene strekking (uitspraken in EA). Tot slot is ook hier sprake van een categorie *overig/onzeker*. Onder *overig* vallen onder andere planschadebeschikkingen en tracébesluiten (deze zouden oorspronkelijk separaat worden geschat, maar het aantal waarnemen in de geannoteerde datasets bleek te laag om dit betrouwbaar te kunnen doen), beheerplannen, natuurbeschermingswet, aanwijzingsbesluiten nieuwbouw, besluiten hogere waarden, wet geluidhinder, bodembescherming, projectplannen en waterwet. Informatie over het soort geschil is in het algemeen te vinden in het Procesverloop.

Geschat wordt dat ruim 4 op de 10 AbRvS-uitspraken betrekking heeft op conflicten over bestemmingsplannen/besluiten van algemene strekking en dat dit voor vergunningbeschikkingen circa 3 op de 10 is. Opgemerkt moet worden dat het kenmerk geschilsoort voor een relatief groot deel (ca. 15%) van deze uitspraken het label *overig/onzeker* heeft gekregen. Bij de rechtbankuitspraken was dit aandeel nog iets hoger (ca. 20%).

FIGUUR 4.9. Geschilsoort: geschatte ontwikkelingen uitspraken AbRvS 2008-2019

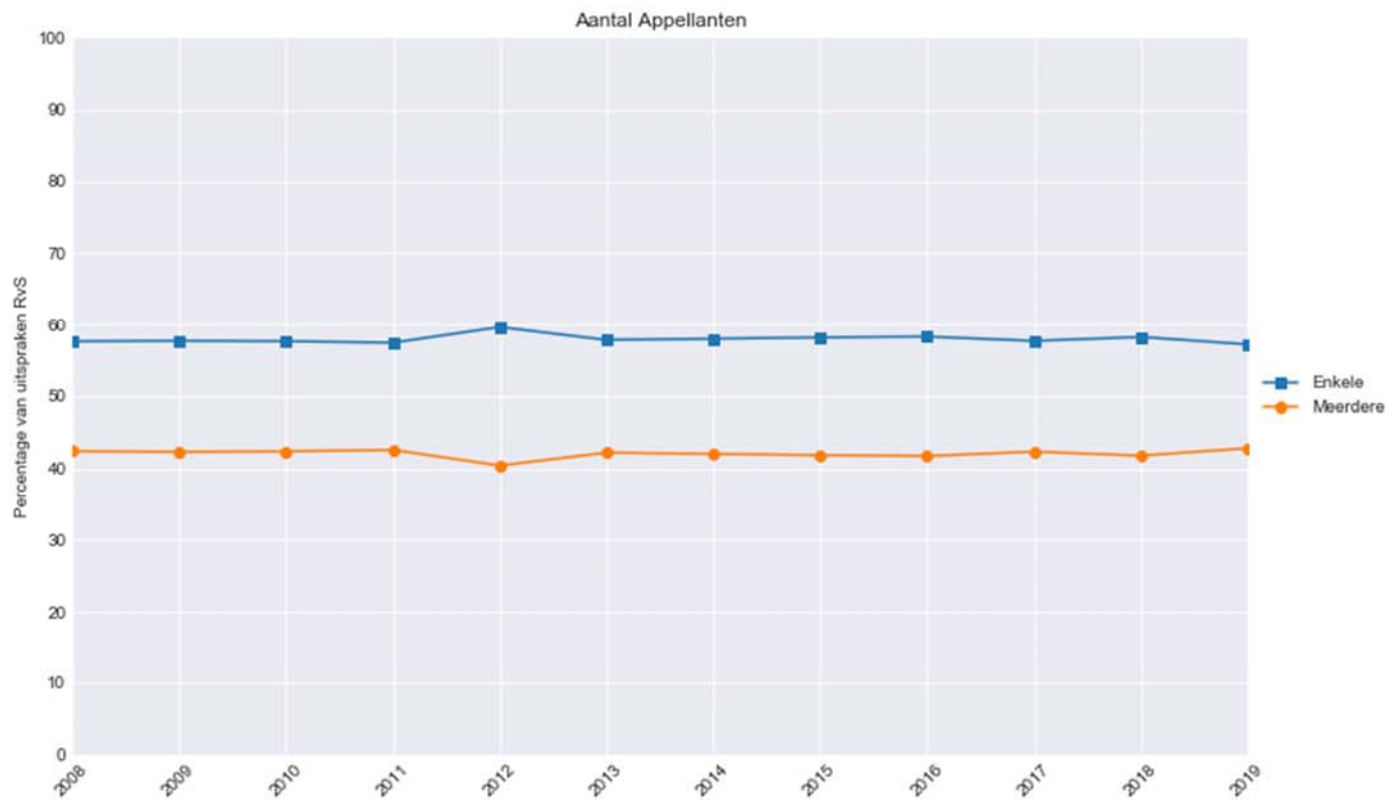


Aantal appellanten

Figuur 4.10 toont de schatting van het geclassificeerde kenmerk 'aantal appellanten'. De hierbij onderscheiden categorieën zijn *één* of *meerdere*. Informatie over dit kenmerk is soms te vinden in de Kop van de uitspraken, soms in het Procesverloop.

Geschat wordt dat bij bijna 6 op de 10 AbRvS-uitspraken sprake is van een enkele eiser. Deze schatting is lager dan bij de gedigitaliseerde rechtbankuitspraken (circa 7 op de 10).

FIGUUR 4.10. Aantal appellanten: geschatte ontwikkelingen uitspraken AbRvS 2008-2019

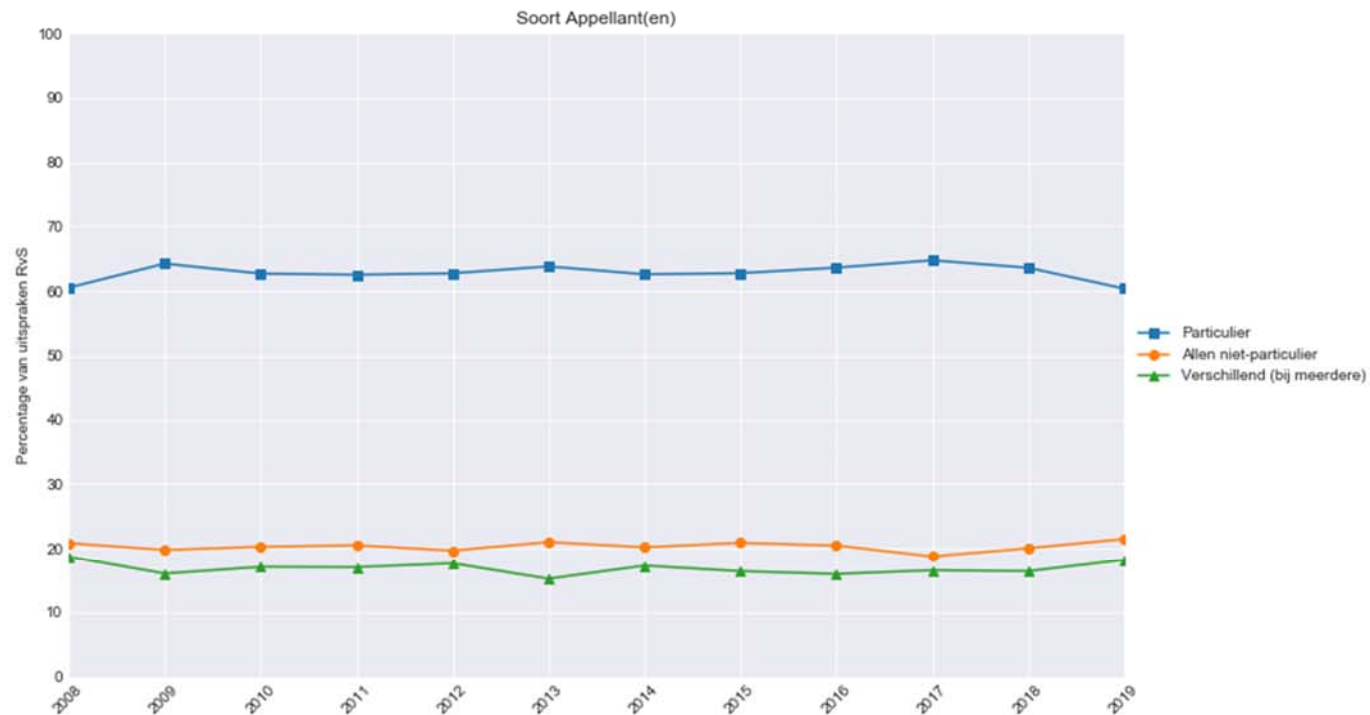


Soort appellant(en)

Figuur 4.11 toont de schatting van het geclassificeerde kenmerk 'soort appellant(en)'. De hierbij onderscheiden categorieën zijn *(allen) particulier*, *(allen) niet particulier* of *verschillend* (bij zowel particuliere en niet-particuliere appellanten; gemengd dus). Informatie over dit kenmerk is te vinden in de Kop van de uitspraken. Alleen wanneer enkel eigennamen werden genoemd bij de appellant(en), werd dit kenmerk door de annoteurs beschouwd als *particulier*. Wanneer bij een eigenaam ook een bedrijfsnaam of bijvoorbeeld een stichting werd genoemd, waar de betreffende persoon bijvoorbeeld directeur van was, dan werd dit gezien als *niet particulier*. Dit kenmerk meet dus niet het onderscheid tussen natuurlijke personen en rechtspersonen.

Geschat wordt dat bij ruim 6 op de 10 van de AbRvS-uitspraken sprake is van (een) particuliere appellant(en). Deze schatting is lager dan bij de gedigitaliseerde rechtbankuitspraken (meer dan 8 op de 10).

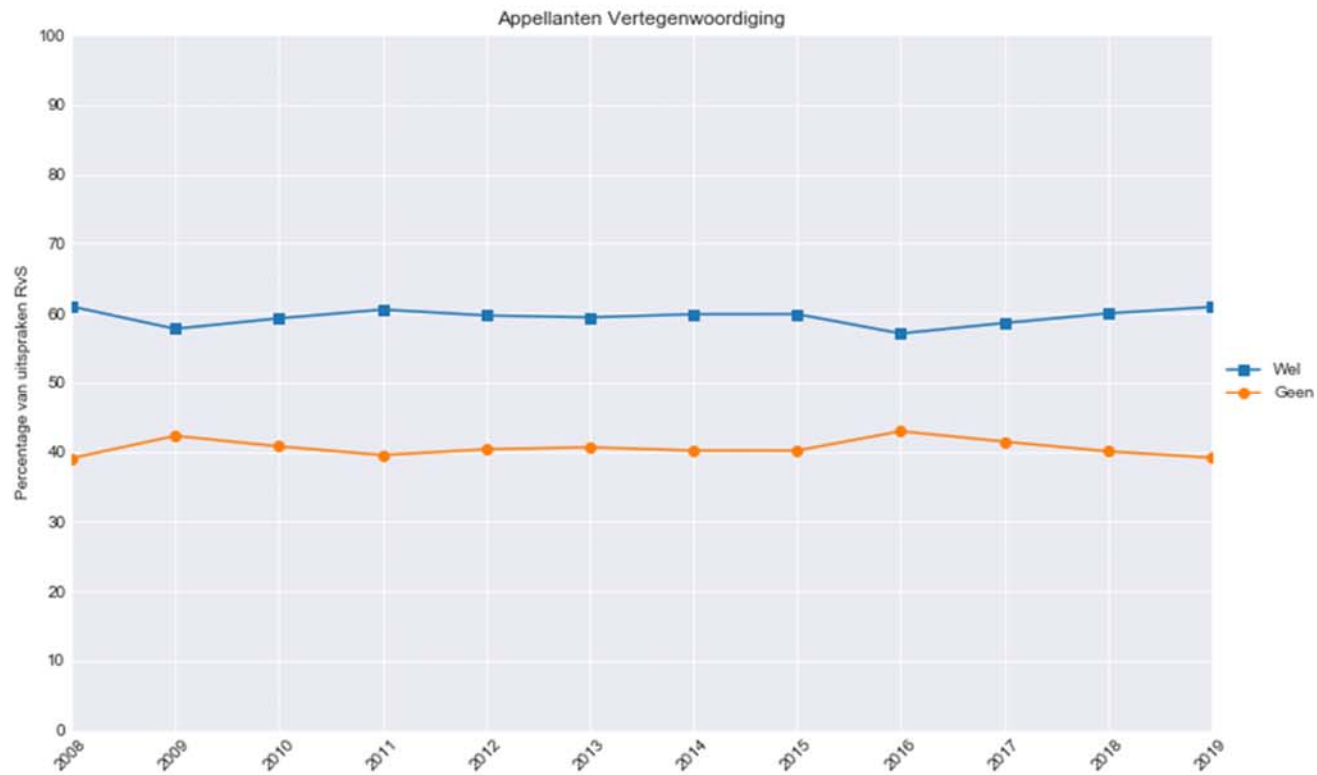
FIGUUR 4.11. Soort appellant(en): geschatte ontwikkelingen uitspraken AbRvS 2008-2019



Vertegenwoordiging appellant(en)

Figuur 4.12 toont de schatting van het geclassificeerde kenmerk 'vertegenwoordiging appellant(en)'. De hierbij onderscheiden categorieën zijn *wel* of *geen*. Een vertegenwoordiger kan een advocaat zijn, maar dit hoeft niet het geval te zijn. Ook wanneer sprake was van een gemachtigd familielid, werd dit door de annoteurs gescoord als *wel* vertegenwoordiging. Informatie over dit kenmerk is soms te vinden in de Kop van de uitspraken onder de appellant, soms is het te vinden in het Procesverloop. Geschat wordt dat bij circa 6 op de 10 AbRvS-uitspraken de appellant zich door iemand heeft laten vertegenwoordigen. Deze schatting is vergelijkbaar met de gedigitaliseerde rechtbankuitspraken.

FIGUUR 4.12. Vertegenwoordiging appellant(en): geschatte ontwikkelingen uitspraken AbRvS 2008-2019

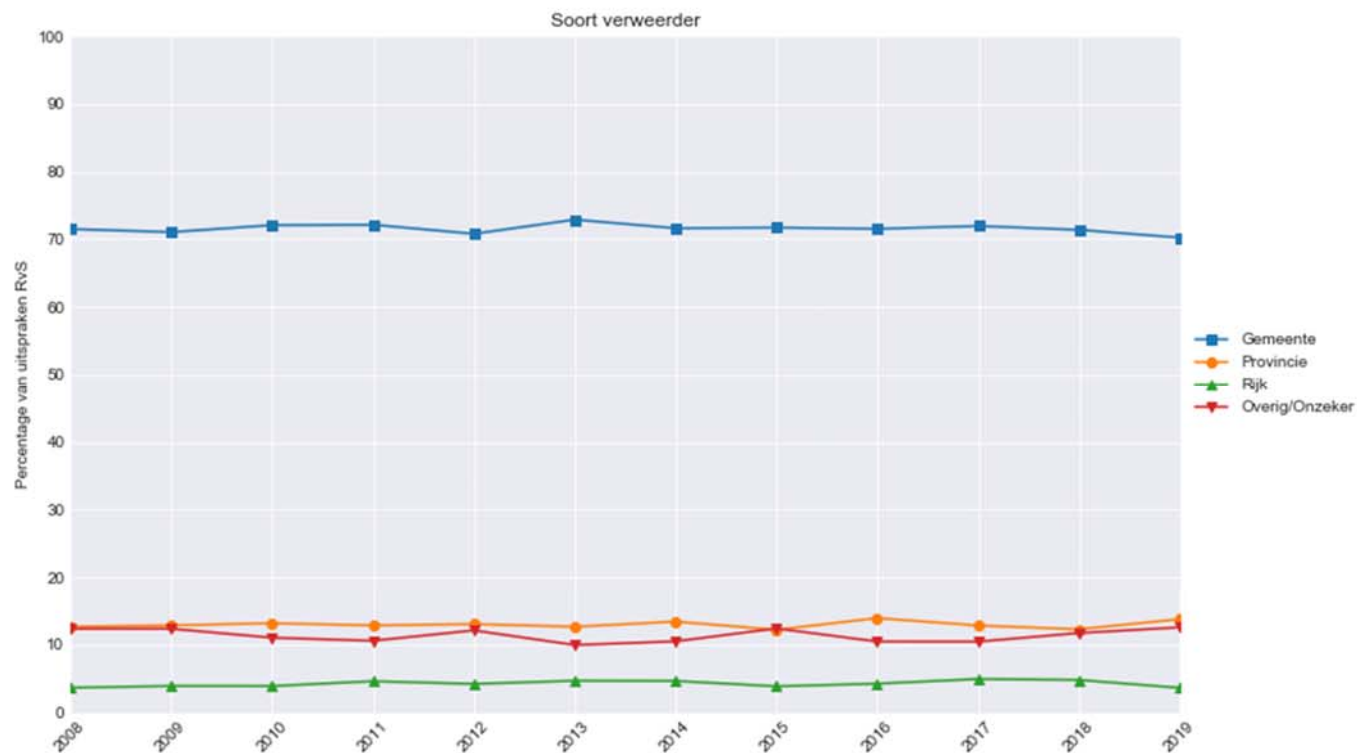


Soort verweerder

Figuur 4.13 toont de schatting van het geclassificeerde kenmerk 'soort verweerder'. De hierbij onderscheiden categorieën zijn *gemeente*, *provincies*, *rijk* en *overig/onzekeer*. Het was de bedoeling hier ook *waterschappen* als aparte categorie te onderscheiden, maar deze bleek te weinig voor te komen om goed te kunnen schatten. Dit is ook het geval voor de categorie *particulier(en)* (bij HB-zaken zijn particulieren in sommige gevallen de verweerder). Deze zijn derhalve geschaard onder *overig/onzekeer*. Informatie over dit kenmerk is in het algemeen te vinden in de Kop van de uitspraak.

Geschat wordt dat bij ruim 7 van de 10 AbRvS-uitspraken de verweerder een gemeente is. Dit is lager dan bij de gedigitaliseerde rechtbankuitspraken, waar dit op ruim 9 van de 10 wordt geschat.

FIGUUR 4.13. Soort verweerder: geschatte ontwikkelingen uitspraken AbRvS 2008-2019

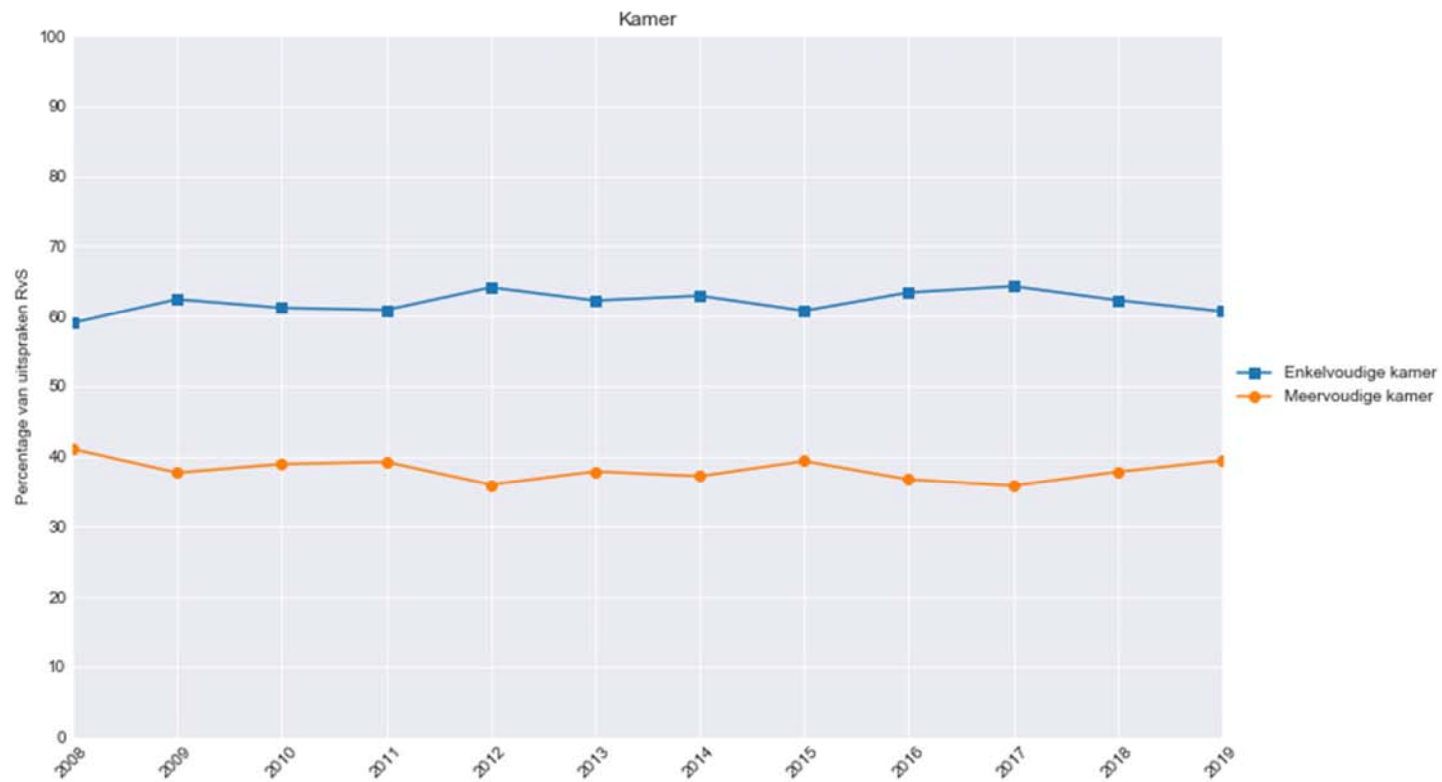


Soort kamer

Figuur 4.14 toont de schatting van het geclassificeerde kenmerk 'soort kamer'. De hierbij onderscheiden categorieën zijn *enkelvoudig* of *meervoudig*. Informatie over dit kenmerk is in het algemeen te vinden in de Conclusie van de uitspraken.

Geschat wordt dat ruim 6 op de 10 AbRvS-uitspraken gedigitaliseerde rechtbankuitspraken door de enkelvoudige kamer worden gedaan. Dit is lager dan bij de gedigitaliseerde rechtbankuitspraken, waar dit op (bijna) 8 van de 10 werd geschat.

FIGUUR 4.14. Soort kamer: geschatte ontwikkelingen uitspraken AbRvS 2008-2019



Uitkomst geschil

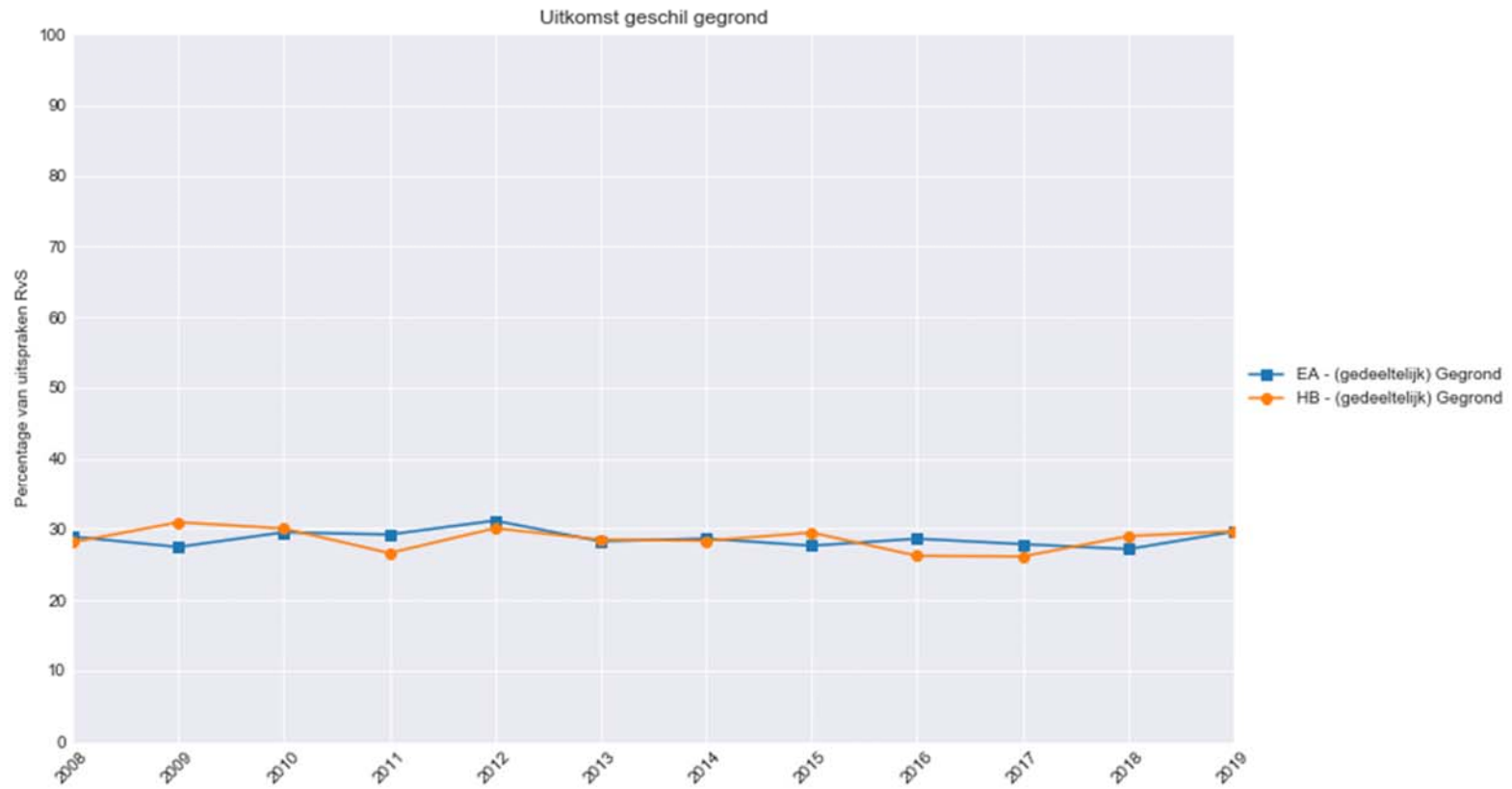
Figuur 4.15a tot en met c tonen de schatting van het geclassificeerde kenmerk 'uitkomst geschil'. Omdat hetgeen wordt aan gevochten in een zaak tegelijk gedeeltelijk *gegrond* en gedeeltelijk *ongeground* kan worden verklaard en de appellat(en) tegelijk gedeeltelijk *niet-ontvankelijk* kan/kunnen worden verklaard, zijn deze kenmerken separaat geschat. Het zijn dus eigenlijk drie kenmerken, te weten 'geheel of gedeeltelijk gegrond', 'geheel of gedeeltelijk ongegrond' en 'geheel of gedeeltelijk niet-ontvankelijk' en elk kenmerk kan wel of niet van toepassing zijn (0 of 1). Dit zorgt er ook voor dat de percentages in de navolgende drie figuren optellen tot meer dan 100 procent. Wanneer enkel het eerste kenmerk 'geheel of gedeeltelijk gegrond' van toepassing is, is hetgeen is aangevochten door de AbRvS dus geheel gegrond verklaard. Een (gedeeltelijke) gegrondverklaring in HB betekent dat de eerdere uitspraak van de rechtbank (gedeeltelijk) wordt vernietigd. Afhankelijk van de inhoud van die eerdere uitspraak, kan dit een gegrondverklaring van het oorspronkelijke beroep van de appellant betekenen (wanneer de appellant in hoger beroep is gegaan) of een ongegrondverklaring (wanneer de overheidsinstantie in hoger beroep is gegaan).

Is tegelijk sprake van 'geheel of gedeeltelijk gegrond' en 'geheel of gedeeltelijk ongegrond', dan betekent dit dus dat hetgeen is aangevochten deels gegrond en deels ongegrond is bevonden door de rechter(s). Het was de bedoeling hier ook 'onbevoegd' te onderscheiden, maar dit bleek te weinig voor te komen om goed te kunnen schatten. Informatie over dit kenmerk is in het algemeen te vinden in de Conclusie van de uitspraken.

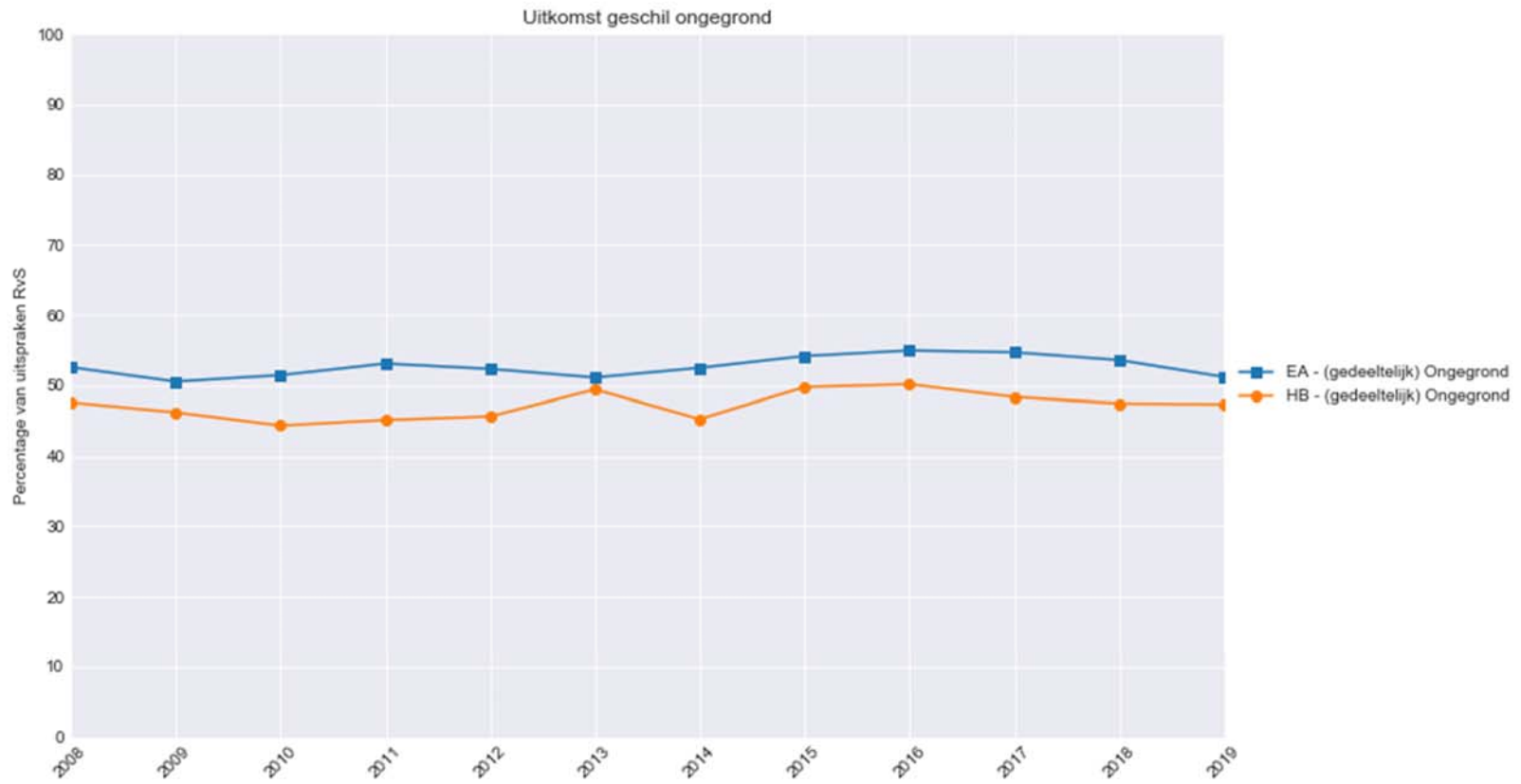
De uitkomsten staan hieronder apart weergegeven voor zaken in EA en HB en wat opvalt is dat de schattingen dichtbij elkaar liggen. Deze kenmerken konden echter met een hoge mate van nauwkeurigheid worden geschat, zoals onder meer blijkt uit de *F1*-scores in Tabel 4.2 (zie Paragraaf 4.3), dus er is geen indicatie dat dit een artefact van het gebruikte model is. De correctiemethode voor uitkomst geschil (gedeeltelijk) gegrond en uitkomst geschil (gedeeltelijk) niet-ontvankelijk, maakt geen onderscheid tussen EA en HB. Bij uitkomst geschil (gedeeltelijk) ongegrond bleek uit een nadere analyse dat het wel noodzakelijk was dit te doen. Figuur 4.15b is dus de resultante van separaat (voor EA en HB) gecorrigeerde aantallen, terwijl de lijnen in Figuur 4.15a en c *achteraf* zijn gesplitst met behulp van de beschikbare meta-informatie over EA en HB. Het voordeel van een dergelijke splitsing is dat de correctiemethode rekening houdt met de samenhang tussen beide kenmerken, dat leidt tot minder vertekening. Een nadeel is dat de schattingen dan worden gebaseerd op een kleiner aantal onderliggende waarnemingen in de trainingsset, waardoor de onzekerheid toeneemt (ofwel de betrouwbaarheidsintervallen op basis van de standaarddeviatie breder worden).

Geschat wordt dat bij circa 3 op de 10 van de AbRvS-uitspraken sprake is van een geheel of gedeeltelijke gegrondverklaring door de rechter(s) en dat bij circa 5 op de 10 gevallen sprake is van een geheel of gedeeltelijke ongegrondverklaring. In bijna 3 op de 10 uitspraken is sprake van (gedeeltelijke) niet-ontvankelijkheid. Bij de gedigitaliseerde rechtbankuitspraken lagen deze schattingen respectievelijk op ruim 3 op de 10, bijna 6 op de 10 en ruim 1 op de 10.

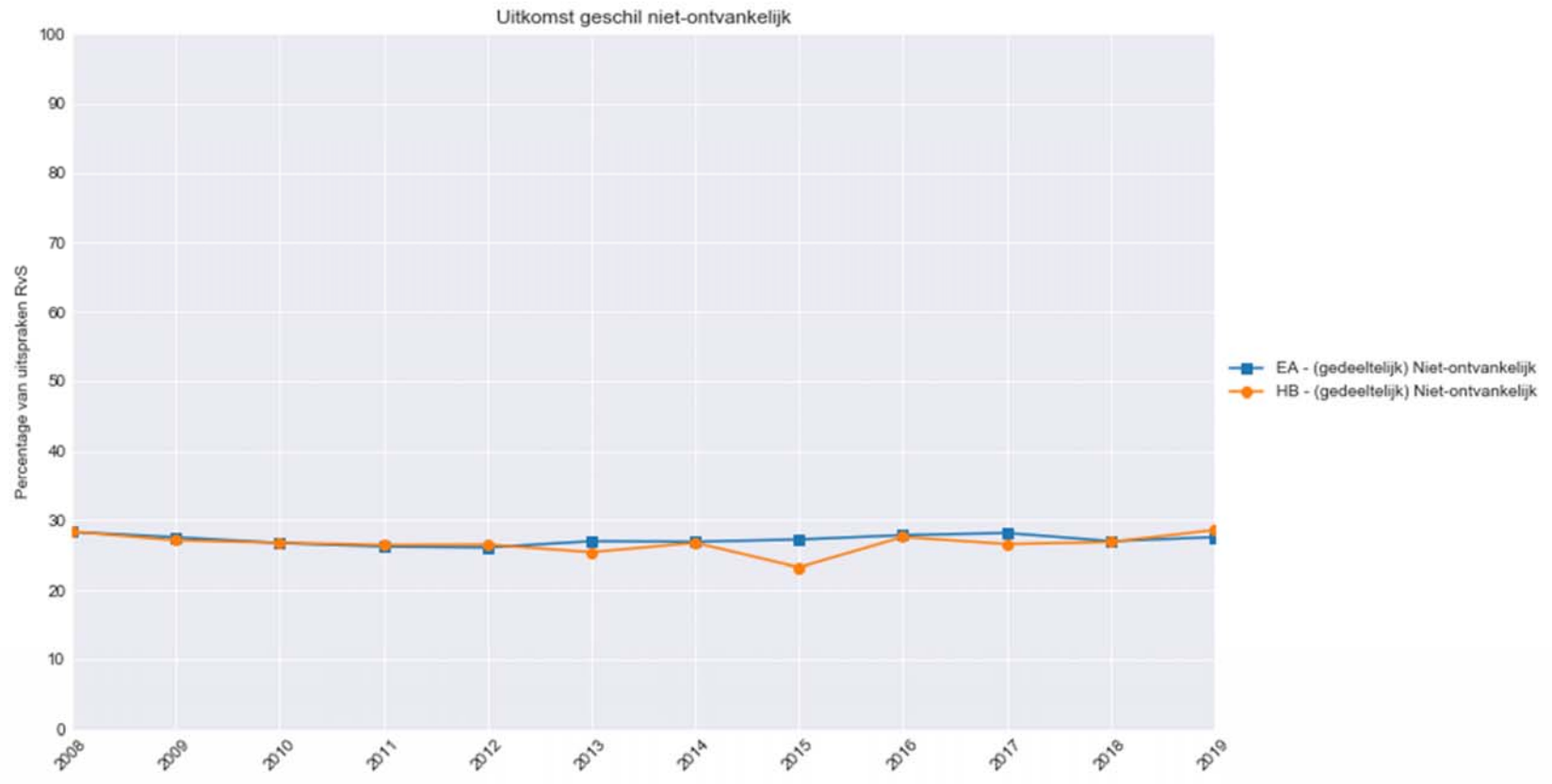
FIGUUR 4.15a. Uitkomst geschil geheel of gedeeltelijk gegrond: geschatte ontwikkelingen uitspraken AbRvS 2008-2019



FIGUUR 4.15b. Uitkomst geschil geheel of gedeeltelijk ongegrond: geschatte ontwikkelingen uitspraken AbRvS 2008-2019



FIGUUR 4.15c. Uitkomst geschil geheel of gedeeltelijk niet-ontvankelijk: geschatte ontwikkelingen uitspraken AbRvS 2008-2019



4.3 Nauwkeurigheid van de schattingen

Correcties van schattingen

Modelschattingen zijn bijna nooit perfect; er is vrijwel altijd sprake van een (kleine) vertekening. Op de originele modelschattingen is in dit onderzoek een correctiemethode toegepast om deze vertekening te verminderen. De methode om deze gecorrigeerde schattingen te maken staat beschreven in Van Delden, Scholtus, & Burger (2015).¹⁴ De aanname bij deze methode is dat de verhouding tussen categorieën in de trainingset overeenkomt met de verhouding in de totale dataset. In de Tabellen 4.1 en 4.2 zijn de voor vertekening gecorrigeerde schattingen van de frequenties van elke categorie binnen de onderzochte kenmerken weergegeven.

Nauwkeurigheid van schattingen

Omdat de weergegeven ontwikkelingen zijn gebaseerd op geschatte kenmerken, bevatten de geschatte aandelen van categorieën een bepaalde mate van onzekerheid. In de Tabellen 4.1 en 4.2 zijn daarom, naast de geschatte frequenties van de verschillende categorieën, de bijbehorende standaarddeviaties (SD's) opgenomen, als indicator van de mate van onzekerheid van de schattingen. De SD's van de gecorrigeerde schattingen zijn gebaseerd op de methode beschreven in Van Delden, Scholtus en Burger (2015), die recentelijk is uitgebreid om rekening te kunnen houden met de onzekerheid van de confusion matrix zelf. Globaal genomen moet bij het beoordelen van de gecorrigeerde schattingen rekening worden gehouden met marges van tussen de $-2 \times SD$ en $+2 \times SD$ om de schattingen, voor een 95%-betrouwbaarheidsinterval. Onder de assumptie dat de annotatiesets representatief zijn voor de totale onderzoeksbestanden, én onder de assumptie dat de schattingsfouten normaal verdeeld zijn, ligt de werkelijke waarde met 95 procent zekerheid binnen de genoemde marges. De normaliteitsassumptie is aannemelijk, aangezien de totale dataset een groot aantal uitspraken bevat. Het aantal annotaties is gelijk verdeeld over de jaren, om een zo representatief mogelijke annotatieset te krijgen.¹⁵ Van lijnen in de grafieken die dicht bij elkaar liggen kan dus niet met voldoende zekerheid worden gesteld dat ze daadwerkelijk op de getoonde wijze van elkaar verschillen. Om in statistische termen te spreken: deze verschillen mogelijk niet statistisch significant van elkaar.¹⁶

Ook zijn in Tabel 4.1. en 4.2 de *F1*-scores per categorie weergegeven. In Tabel 3.4 gebeurde dit reeds voor de kenmerken als geheel. Uit deze uitgesplitste *F1*-scores, tezamen met de weergegeven SD's, valt op te maken dat binnen kenmerken niet alle categorieën even nauwkeurig en betrouwbaar kunnen worden geschat.

¹⁴ Delden, A. van, Scholtus, S. en Burger, J. (2015). *Quantifying the effect of classification errors on the accuracy of mixed-source statistics*. Technical Report. The Hague: Statistics Netherlands.

¹⁵ Tijdens het samenstellen van de annotatieset van Rvdr zijn de filters die in Hoofdstuk 2 staan een aantal keren aangescherpt. Het aantal annotaties van Rvdr is daardoor niet gelijk verdeeld over de jaren, maar bevat relatief meer uitspraken vanaf 2015.

¹⁶ Hier moet aan worden toegevoegd dat in figuren met geschatte ontwikkelingen voor uitkomst geschil bij de AbRvS-uitspraken onderscheid is gemaakt naar EA en HB. Hierdoor zijn de uitkomsten gebaseerd op een kleiner onderliggend aantal waarnemingen, met als gevolg een groter betrouwbaarheidsinterval rondom de geschatte ontwikkelingen. Deze opsplitsing is niet gemaakt in Tabel 4.2 en de hier weergegeven standaarddeviaties kunnen derhalve niet worden gebruikt voor globale betrouwbaarheidsintervallen rondom de geschatte ontwikkelingen.

TABEL 4.1. Correcties van schattingen rechtbanken

Kenmerk	Categorie	F1-score	Gecorrigeerde schatting	SD
Geschildsoort	Handhavingsbeschikking	0,91	4.710	253
	Planschadebeschikking	0,69	3.824	465
	Vergunningsbeschikking ^a	0,90	12.757	421
	Overig/onzeker	0,54	5.048	368
Aantal eisers	Enkele	0,99	18.936	160
	Meerdere	0,97	7.403	161
Soort eiser(s)	Allen niet-particulier	0,79	4.226	419
	Particulier	0,93	21.870	502
	Verscheidend (bij meerdere)	0,67	243	268
Vertegenwoordiging eiser(s)	Geen	0,81	7.588	392
	Wel	0,93	18.751	393
Soort verweerder	Gemeente	0,99	24.070	197
	Provincie	0,90	448	96
	Rijk	0,94	763	78
	Overig/onzeker ^b	0,75	1.058	151
Soort kamer	Enkelvoudige kamer	1,00	20.513	119
	Meervoudige kamer	1,00	3.829	- ^c
	Overig/onzeker	0,94	1.887	120
Uitkomst geschil gegrond	(gedeeltelijk) Gegrond	0,96	8.307	201
Uitkomst geschil ongegrond	(gedeeltelijk) Ongegrond	0,97	15.798	226
Niet-ontvankelijk	(gedeeltelijk) Niet-ontvankelijk	0,97	3.212	112

^a Alle varianten en inclusief vrijstellingen, ontheffingen, projectbesluiten, milieumeldingen en aanwijzingsbesluiten ORAC.

^b Inclusief Waterschappen.

^c In de testset wordt deze categorie perfect geschat. De SD kan daardoor niet worden geschat, maar deze zal zeer klein zijn.

TABEL 4.2. Correcties van schattingen AbRvS

Kenmerk	Categorie	F1-score	Gecorrigeerde schatting	SD
Geschildsoort	Handhavingsbeschikking	0,72	3.073	416
	Bestemmingsplan/besluit van algemene strekking	0,94	10.678	280
	Vergunningsbeschikking ^a	0,83	7.434	436
	Overig/onzeker ^b	0,76	3.482	324
Aantal appellanten	Enkele	0,99	14.293	110
	Meerdere	0,99	10.374	110
Soort appellant(en)	Allen niet-particulier	0,95	5.014	176
	Particulier	0,98	15.516	172
	Verscheidend (bij meerdere)	0,93	4.138	187
Vertegenwoordiging appellant(en)	Geen	0,86	9.512	425
	Wel	0,94	15.155	424
Soort verweerder	Gemeente	0,98	17.646	233
	Provincie	0,97	3.185	122
	Rijk	0,88	1.047	150
	Overig/onzeker ^c	0,78	2.789	298
Soort kamer ^d	Enkelvoudige	1,00	15.302	18
	Meervoudige	1,00	9.365	18
Uitkomst geschil gegrond	(gedeeltelijk) Gegrond	0,97	7.064	145
Uitkomst geschil ongegrond	(gedeeltelijk) Ongegrond	0,94	12.604	324
Niet-ontvankelijk	(gedeeltelijk) Niet-ontvankelijk	0,98	6.644	113

^a Alle varianten en inclusief vrijstellingen, ontheffingen, projectbesluiten, milieumeldingen en aanwijzingsbesluiten ORAC.

^b Inclusief Tracébesluiten.

^c Inclusief Waterschappen.

^d Alle annotaties van soort kamer zijn goed geschat door het model. Er is daarom geen correctie toegepast. De SD is berekend op basis van de 95%-betrouwbaarheidsniveaus van de true positives en true negatives.

Stabiliteit schattingen over de onderzochte jaren

De ontwikkelingen in de voorgaande paragrafen vertonen een grote mate van stabiliteit over de onderzochte periode. Het onderliggend totaal aantal uitspraken verschilt wel per jaar. Om uit te sluiten dat de hoge mate van stabiliteit in de geschatte ontwikkelingen een artefact is van de gebruikte modellen, zijn hier enkele aanvullende analyses op verricht.¹⁷ Hoewel deze geen 100 procent zekerheid geven, duiden deze analyses erop dat de gevonden stabiliteit overeenkomt met de werkelijkheid. Zo bleken twee van de kenmerken, te weten Soort kamer en Aantal eisers, bijna foutloos met regex vast te stellen. Deze kunnen dus met een hoge waarschijnlijkheid zeer goed worden voorspeld, en ook hier tonen de modelschattingen stabiele ontwikkelingen over de tijd. Ook uit een expertreview met inhoudsdeskundigen van de Rvdr, RvS, de Universiteit van Tilburg, Rechtbank Oost-Brabant en Min BZK is bovendien gebleken dat de geschatte ontwikkelingen in grote mate werden herkend, dan wel aannemelijk werden gevonden.

¹⁷ In de modelschattingen is geen onderscheid gemaakt tussen de jaren. Om te testen of het gerechtvaardigd is te veronderstellen dat de modellen in één keer geschat kunnen worden op alle jaren gezamenlijk, zijn voor enkele geschatte kenmerken getoetst of de nauwkeurigheid van de modelvoorspellingen afhangt van de factor jaar (via een deviance toets met G2 grootheid). Deze tests zijn uitgevoerd voor AbRvS-uitspraken voor alle jaren individueel én (vanwege het lage aantal annotaties per jaar), voor de jaren tot 2014 versus de jaren vanaf 2014. De tests lieten telkens zien dat het niet nodig is voor de schattingen onderscheid te maken naar jaren.

5. Conclusies

5.1 Opbrengst tot nu toe

De afgelopen anderhalf jaar heeft het CBS, op aanvraag van Min BZK/ICTU, in samenwerking met de Rvdr, de RvS en onder begeleiding van een begeleidingsgroep (Zie Bijlage 1), veel nieuwe inzichten verkregen in de mogelijkheden om door middel van text mining en machine learning rechterlijke uitspraken over escalaties in het omgevingsrecht te classificeren. Een eerste beeld van aard en omvang van deze escalaties is bovendien inmiddels gerealiseerd.

Een belangrijke conclusie is dat het mogelijk is om, op basis van informatie in de teksten van rechterlijke uitspraken, op een gestructureerde wijze een aantal vooraf bepaalde kenmerken met categorieën te classificeren.

Er is een basis neergezet voor verder onderzoek naar de aard en omvang in escalaties in het omgevingsrecht:

- De Rvdr en de RvS kunnen documenten uit hun registraties filteren van relevante (niet geanonimiseerde) uitspraken en die aanleveren aan het CBS. Het CBS kan deze vervolgens geautomatiseerd nader filteren en voorbereiden/structureren, zodat ze geschikt worden gemaakt voor annoteren en vervolgens modelleren.
- Het CBS heeft een 'annotatie-tool' gebouwd waarin rechterlijke uitspraken op een efficiënte manier kunnen worden geannoteerd en waarbij de resultaten op een gestructureerde wijze worden opgeslagen.
- Het CBS heeft, met behulp van gangbare modellen, modelschattingen gemaakt voor een set van 9 indicatoren die iets zeggen over de aard van een zaak/uitspraak en daarbij ook gekeken naar de kwaliteit en betrouwbaarheid van modelschattingen. Deze modelschattingen zijn nu uitgevoerd op de voorts *gedigitaliseerde* rechterlijke uitspraken door rechtbanken in de periode 2007-2018 en alle AbRvS-uitspraken in de periode 2008-2019. Uit een expertreview met inhoudsdeskundigen van de Rvdr, RvS, de Universiteit van Tilburg, Rechtbank Oost-Brabant en Min BZK is gebleken dat de modeluitkomsten, in ieder geval op hoofdlijnen, overwegend aansluiten bij de beelden en verwachtingen van deskundigen.
- De door het CBS geschatte kenmerken geven een relatief stabiel beeld in de tijd; de verhoudingen tussen categorieën van variabelen veranderen slechts marginaal. Hoewel het niet met 100 procent zekerheid te zeggen is, zijn er voldoende aanwijzingen om aan te nemen dat deze stabiele resultaten een reëel beeld geven en dat dit dus geen model-artefact is. De ontwikkelingen voor de rechtbank-uitspraken en voor de AbRvS-uitspraken zijn separaat, in afzonderlijke modellen, geschat. Desalniettemin komt een deel van de resultaten aardig overeen; bijvoorbeeld als je kijkt naar het aantal eisers/appellanten (één, meerdere), vertegenwoordiging van eisers/appellanten en ook de uitkomst van het geschil (gegrond, ongegrond). Er zijn ook zichtbare verschillen, zoals de verhoudingen naar soort eiser/appellant, soort kamer en soort verweerder.
- Voorbeelden van uitkomsten die stroken met verwachtingen bij rechtbankuitspraken zijn dat er een ontwikkeling is geweest van een toenemend aandeel omgevingsrecht-uitspraken waarbij de eiser zich laat vertegenwoordigen en dat het aandeel rechtbankuitspraken dat (gedeeltelijk) gegrond is verklaard constant tussen de 30 en 40 procent ligt.
- Er zijn, in samenwerking met de Rvdr, eerste inzichten opgedaan met betrekking tot de representativiteit van de selectie van omgevingsrecht-uitspraken uit het E-archief. Er zijn ook suggesties voor nader onderzoek. Nader onderzoek wordt aanbevolen om deze selectiviteit (voor de volledige onderzoeksperiode) beter te kwantificeren.
- De modelschattingen van geclassificeerde kenmerken kunnen relatief eenvoudig worden weggeschreven in een databestand met daarin een regel voor iedere unieke uitspraak, aangevuld met de meta-kenmerken uit de brondata (zoals jaar van uitspraak en rechtbank) en de geclassificeerde kenmerken in de kolommen. Een dergelijk databestand is per definitie ontdaan van identificerende persoonsgegevens en is daarmee, technisch en privacy-technisch gezien, direct geschikt voor nader onderzoek. Voor inhoudelijk gebruik kunnen grofweg de mogelijkheden en beperkingen worden aangegeven.

5.2 Suggesties voor vervolgonderzoek

Het CBS ziet de volgende kansen en uitdagingen om in gezamenlijkheid toe te werken naar een *complete, representatieve, robuuste en toekomstbestendige nulmeting*.

Met *compleet* wordt bedoeld dat de huidige set aan indicatoren nog niet volledig dekkend is om de hypothesen over ontwikkelingen na de invoering van de Omgevingswet, in aard en omvang van escalaties in het omgevingsrecht goed te kunnen toetsen. Zo is het thema complexiteit van de rechtszaak bijvoorbeeld op dit moment nog onderbelicht in de set van indicatoren. Bovendien moet dit thema nog helder worden gedefinieerd. Met input van inhoudelijk experts, bijvoorbeeld de deskundigen in de begeleidingsgroep, kan de set van gewenste indicatoren verder aangevuld en ook aangescherpt worden. Deze nieuwe indicatoren kunnen dan worden toegevoegd aan de annotatietool en de modelschattingen en kunnen vervolgens worden gebruikt bij de beantwoording van onderliggende onderzoeksvragen.

Met *representatief* wordt bedoeld dat er uiteindelijk zicht moet komen op ontwikkelingen in het *totaal* van alle escalaties in het omgevingsrecht, dan wel door meer rechtbankuitspraken te digitaliseren (lange termijn) en op te nemen in bijvoorbeeld het E-archief, dan wel door een methodiek te ontwikkelen om, op basis van een bundeling van de nu beschikbare data, informatie en expertise, de representativiteit te kwantificeren en/of de huidige modelschattingen te corrigeren voor onvolledigheid.

Met *robuust* wordt bedoeld dat de uitkomsten van de nulmeting ongevoelig zijn voor kleine wijzigingen. Als vervolgstappen stelt het CBS concreet voor om 'aan de voorkant' de dataselecties nog verder aan te scherpen (met name uit het E-archief) en daarnaast het proces van annoteren te verbeteren, bijvoorbeeld door meer onderzoekers met juridische achtergrond laten annoteren, of te betrekken voor instructiesessies. Zeker als de nog toe te voegen indicatoren juridisch complexer zijn (zoals verwacht bij het thema complexiteit), lijkt dit een verstandige vervolgstap. Ook wordt aanbevolen om het aantal annotaties nog te vergroten om de ontwikkelingen in relatief kleine, nu lastig te schatten groepen uitspraken (zoals tracébesluiten) te kunnen kwantificeren. Verder zou nader onderzoek kunnen worden gedaan naar de inhoud van categorieën die nu nog vragen oproepen, zoals de categorie *Overig/onzeker* bij het kenmerk Geschilsoort. Met robuust wordt bovendien bedoeld op een verfijning van de modellen 'aan de achterkant', bijvoorbeeld door in de modellen input mee te geven over (cor)relatie tussen verschillende indicatoren. Daarbij hoort bijvoorbeeld ook een methode ontwikkelen om op een consequente en begrijpelijke wijze informatie mee te genereren over de *marges* rondom de modelschattingen. In de huidige nulmeting is daar een voorschot op genomen.

Met *toekomstbestendig*, ten slotte, wordt bedoeld op de verdere toekomst en dan concreet de koppeling tussen de nulmeting en vervolgmetingen, na de invoering van de Omgevingswet, die is gelegen in 2022. Door de invoering van de Omgevingswet zal het omgevingsrecht behoorlijk op de schop gaan en dat zal consequenties hebben voor de wijze van rechtspraak en op de daarbij behorende terminologie. Er zal zorgvuldig moeten worden gekeken naar aansluiting op de indicatoren van de nulmeting.

Alle bovengenoemde aanbevelingen zouden, eventueel als deelprojecten, kunnen worden opgepakt in een gezamenlijk vervolgproject op deze eerste nulmeting. Al met al is een rode draad in bovengenoemde suggesties voor het vervolg een *intensievere bundeling van krachten van de betrokken belanghebbende organisaties*, als belangrijke stap om deze nulmeting verder door te ontwikkelen tot een toekomstbestendige monitor.

Bijlage 1. Samenstelling Begeleidingsgroep

Voorzitter: Prof.dr.mr. S. Zouridis, UvT

Leden: Prof.dr. B.F.M. Bakker, VU/CBS
Prof.dr. F.J. Bex, UvT /UU
Mr. V. van Dorst, RvS
Mr. M.M. van Driel, RvS
F.J. Herkes MSc., UvT
Mr. Dr. M. Hoogwout, Min BZK/ICTU
Mr. J. Huijben, Min BZK
Mr. B. Rademaker, Min BZK
Drs. J. Smits, Rvdr
Dr. F.P. van Tulder, Rvdr
Mr. M.J.H.M. Verhoeven, RB Oost Brabant
Ing. T. Willemsen, RvS

Bijlage 2. Lijst met zoektermen t.b.v. selectie uitspraken omgevingsrecht uit E-archief

Het gaat om rechterlijke uitspraken die gedaan zijn in de periode 1-1-2007 tot en met 31-12-2018 in de categorie bestuursrecht en waarin minimaal één van de volgende termen in de tekst van de uitspraak voorkomen (indien hoofdlettergevoelig dan zowel de trefwoorden met als zonder hoofdletters):

- Omgevingsrecht
- Omgevingsvergunning
 - o Bouwvergunning
 - o Monumentenvergunning
 - o Reclamevergunning
 - o Milieuvergunning
 - o Gebruiksvergunning
 - o Kapvergunning
 - o Inrit of uitwegvergunning
- Watervergunning
- Wabo
- Waterwet
- Wet ruimtelijke ordening (Wro)
- Wet natuurbescherming (Wnb)
 - o Natuurbeschermingswet
 - o Boswet
 - o Flora en faunawet
- Woningwet
- Wet Milieubeheer
- Ontgrondingenwet
- Wet geluidhinder
- Monumentenwet
- Wet bodembescherming (Wbb)
- Tracéwet
- Crisis en herstelwet (Chw)
- Wet voorkeursrecht gemeenten of Wvg
- Ontheffing

Voor de genoemde wetten kan het voorkomen dat de uitspraken alleen op hun afkorting te vinden zijn.

Er wordt rekening mee gehouden dat de selectie op basis van bovenstaande trefwoorden mogelijk te ruim is voor het doel van het onderzoek. Afgesproken is dat de onderzoekers van het CBS de niet relevante uitspraken eruit zullen filteren zodra ze worden herkend.

Bijlage 3. Tabellen bij Hoofdstuk 4

Onderstaande twee tabellen bevatten de geschatte ontwikkelingen van de gelabelde kenmerken van escalaties in het omgevingsrecht in deze nulmeting. De cijfers worden zowel getoond als relatieve ontwikkelingen (per indicator is weergegeven wat het geschatte procentuele aandeel is van elke categorie in het totaal aantal uitspraken) als absoluut geschatte frequenties, naar jaar van uitspraak. De relatieve cijfers laten een stabiel beeld zien, de absolute aantallen variëren per jaar, doordat het aantal rechterlijke uitspraken jaarlijks fluctueert. De relatieve aantallen corresponderen met de figuren in Paragraaf 4.1 (ontwikkelingen getoond van de *gedigitaliseerde* rechtbankuitspraken in 2007-2018) en 4.2 (ontwikkelingen van de AbRvS-uitspraken in 2008-2019).

Zoals toegelicht in Paragraaf 4.3 bevatten de geschatte aandelen van categorieën een bepaalde mate van onzekerheid. Deze moeten derhalve met voorzichtigheid worden gehanteerd. Om een indicatie te geven van de mate van onzekerheid van de schattingen, zijn in de Tabellen 4.1 en 4.2, naast de geschatte totaal frequenties van de verschillende categorieën, de bijbehorende standaarddeviaties (SD's) opgenomen.

TABEL I. Geschatte aantallen (absoluut en relatieve percentages) rechtbanken, naar jaar van uitspraak

Kenmerk	Jaar																							
	2007	%	2008	%	2009	%	2010	%	2011	%	2012	%	2013	%	2014	%	2015	%	2016	%	2017	%	2018	%
Geschilsoort																								
Handh.besch.	463	19	400	17	453	17	551	18	571	17	466	17	402	19	348	19	273	18	255	19	284	20	243	18
Plnsh.besch.	361	14	354	15	351	13	431	14	471	14	390	14	356	16	248	14	233	16	182	13	209	15	239	18
Verg.besch.	1.235	49	1.158	48	1.280	49	1.494	48	1.611	48	1.296	48	1.030	48	900	50	732	49	680	49	693	49	647	47
Overig/onz.	441	18	503	21	552	21	634	20	674	20	546	20	372	17	321	18	265	18	260	19	242	17	239	17
Aantal eisers																								
Enkele	1.788	72	1.698	70	1.910	72	2.223	72	2.398	72	1.967	73	1.576	73	1.319	73	1.072	71	960	70	1.015	71	1.010	74
Meerdere	712	28	717	30	727	28	886	28	929	28	732	27	584	27	499	27	430	29	416	30	413	29	358	26
Soort eiser(s)																								
Allen n.-part.	353	14	367	15	440	17	512	16	529	16	411	15	364	17	331	18	230	15	216	16	240	17	233	17
Particulier	2.118	85	2.023	84	2.180	83	2.569	83	2.753	83	2.272	84	1.787	83	1.463	80	1.269	84	1.147	83	1.169	82	1.118	82
Versch.	28	1	24	1	17	1	28	1	45	1	15	1	9	0	24	1	3	0	13	1	19	1	17	1
Vert.w. eiser(s)																								
Geen	712	28	716	30	807	31	893	29	990	30	779	29	662	31	510	28	434	29	351	26	374	26	359	26
Wel	1.788	72	1.699	70	1.830	69	2.216	71	2.337	70	1.920	71	1.498	69	1.308	72	1.068	71	1.025	74	1.054	74	1.009	74
Soort verw.																								
Gemeente	2.302	92	2.198	91	2.422	92	2.873	92	3.042	91	2.473	92	1.961	91	1.628	90	1.363	91	1.266	92	1.291	90	1.249	91
Provincie	37	1	30	1	38	1	56	2	67	2	46	2	45	2	40	2	22	1	12	1	28	2	27	2
Rijk	78	3	68	3	83	3	67	2	88	3	80	3	71	3	59	3	34	2	45	3	45	3	45	3
Overig/onz.	83	3	119	5	94	4	114	4	129	4	100	4	83	4	91	5	83	6	52	4	63	4	47	3
Soort kamer																								
Enkelv.	1.968	79	1.859	77	2.073	79	2.364	76	2.582	78	2.097	78	1.719	80	1.413	78	1.171	78	1.084	79	1.112	78	1.071	78
Meerv.	373	15	349	14	358	14	487	16	490	15	391	14	279	13	272	15	215	14	199	14	216	15	200	15
Overig/onz.	159	6	207	9	206	8	258	8	255	8	211	8	162	7	133	7	116	8	93	7	100	7	97	7
Uitk. geschil (gedlt.) gegr.																								
Uitk. geschil (gedlt.) ongegr.	1.417	57	1.336	55	1.441	55	1.725	55	1.786	54	1.511	56	1.228	57	984	54	838	56	768	56	770	54	750	55
(gedlt.) Niet-ontv.	327	13	312	13	337	13	427	14	469	14	335	12	293	14	237	13	175	12	183	13	192	13	182	13

TABEL II. Geschatte aantallen (absoluut en relatieve percentages) AbRvS, naar jaar van uitspraak

Kenmerk	Jaar																							
	2008 ^a	%	2009	%	2010	%	2011	%	2012	%	2013	%	2014	%	2015	%	2016	%	2017	%	2018	%	2019	%
Geschilsoort																								
Handh.besch.	148	13	331	14	308	11	306	13	276	12	270	12	298	13	238	11	207	12	204	12	222	12	258	14
Best.pln/besl.alg.str.	491	43	994	41	1199	44	1074	44	968	43	1044	46	1015	43	890	42	692	41	715	43	777	43	801	43
Verg.besch.	346	30	700	29	854	31	727	30	706	31	634	28	707	30	679	32	502	30	484	29	537	30	547	29
Overig/onz.	163	14	376	16	352	13	325	13	307	14	311	14	326	14	286	14	271	16	255	15	250	14	256	14
Aantal appellanten																								
Enkele	661	58	1384	58	1563	58	1396	57	1344	60	1305	58	1359	58	1217	58	974	58	956	58	1040	58	1064	57
Meerdere	487	42	1017	42	1150	42	1036	43	913	40	953	42	986	42	877	42	698	42	702	42	747	42	797	43
Soort appellant(en)																								
Allen n.-part.	239	21	474	20	550	20	498	20	444	20	474	21	473	20	437	21	342	20	311	19	358	20	399	21
Particulier	695	61	1542	64	1701	63	1520	63	1416	63	1441	64	1467	63	1314	63	1063	64	1073	65	1136	64	1123	60
Versch.	214	19	385	16	463	17	414	17	398	18	344	15	405	17	343	16	267	16	274	17	293	16	338	18
Vert.w. appellant(en)																								
Geen	426	37	964	40	1050	39	912	38	865	38	871	39	894	38	798	38	682	41	653	39	680	38	692	37
Wel	722	63	1437	60	1663	61	1520	62	1392	62	1387	61	1451	62	1296	62	990	59	1005	61	1107	62	1169	63
Soort verw.																								
Gemeente	820	71	1704	71	1953	72	1752	72	1596	71	1644	73	1678	72	1500	72	1195	71	1192	72	1274	71	1305	70
Provincie	145	13	307	13	356	13	311	13	293	13	285	13	313	13	255	12	232	14	212	13	219	12	255	14
Rijk	42	4	94	4	106	4	112	5	95	4	106	5	109	5	81	4	71	4	82	5	86	5	68	4
Overig/onz.	141	12	296	12	298	11	256	11	272	12	224	10	245	10	259	12	174	10	173	10	209	12	233	13
Soort kamer																								
Enkelv.	677	59	1497	62	1658	61	1479	61	1446	64	1404	62	1474	63	1271	61	1059	63	1065	64	1112	62	1128	61
Meerv.	471	41	904	38	1055	39	953	39	811	36	854	38	871	37	823	39	613	37	593	36	675	38	733	39
Uitk. geschil (ged.) gegr EA	217	29	441	27	585	29	485	29	449	31	422	28	481	29	408	28	304	29	270	28	275	27	339	30
Uitk. geschil (ged.) gegr HB	111	28	245	31	218	30	204	27	244	30	217	28	187	28	181	29	159	26	179	26	224	29	212	30
Uitk. geschil (ged.) ongegr EA	395	53	813	51	1021	51	882	53	755	52	764	51	883	52	800	54	585	55	532	55	543	54	587	51
Uitk. geschil (ged.) ongegr HB	188	48	366	46	323	44	347	45	371	46	376	49	300	45	306	50	304	50	331	48	367	47	338	47
(ged.) Niet-ontv. EA	213	28	443	28	530	27	435	26	376	26	403	27	453	27	403	27	296	28	274	28	274	27	316	28
(ged.) Niet-ontv. HB	112	28	215	27	195	27	204	26	215	26	193	25	177	27	143	23	167	28	182	27	208	27	204	29

^a Uitspraken vanaf 1-7-2008.

Bijlage 4. Rapportage haalbaarheidsonderzoek



Rapportage haalbaarheidsonderzoek Monitor Omgevingswet

Dr. Heike Goudriaan

Nick de Wolf, M.Sc.

Drs. Luc Verschuren

projectnummer

180900

24 mei 2020

Inhoudsopgave

1. Inleiding	4
1.1 Achtergrond	4
1.2 Doel	5
1.3 Aanpak	5
1.4 Methode	6
2. Data rechterlijke uitspraken	8
2.1 Filtering uitspraken op domein omgevingsrecht	8
2.2 Ontbreken niet geanonimiseerde RvS-uitspraken	10
2.3 Onvolledige selectie uit E-archief en selectiviteit	11
2.4 Voorzieningenrechterzaken	12
3. Haalbaarheidsanalyses	13
3.1 Verkenning tekststructuur en betrouwbaarheid van splitsing van de data	13
3.2 Het maken van een geannoteerd databestand	14
3.3 Haalbaarheidsbevindingen per kenmerk	15
3.4 Toelichtingen per kenmerk	17
3.5 Overall betrouwbaarheid geclassificeerde kenmerken	19
3.6 Betrouwbaarheid classificaties binnen kenmerken	20
4. Conclusies en plan van aanpak vervolg	21
4.1 Conclusies	21
4.2 Plan van aanpak vervolg	22
BIJLAGE I Haalbaarheidsbevindingen per kenmerk	25

1. Inleiding

De Omgevingswet, die naar verwachting in op 1 januari 2021 in werking treedt, heeft onder andere als doel om de besluitvorming op het gebied van omgevingsrecht en ruimtelijke ordening te versoepelen, zodat bedrijven, burgers en professionals van overheden makkelijker en efficiënter zaken kunnen doen op dit gebied. De nieuwe wet zorgt voor minder en overzichtelijkere regels, een samenhangende benadering van de leefomgeving en snellere besluitvorming. Zo komt er meer ruimte voor initiatieven van burgers en bedrijven, zijn er minder vergunningen nodig en worden de procedures korter. Voor bedrijven wordt het gemakkelijker te innoveren. Een van de verwachtingen vanuit het beleid is dat de invoering van de Omgevingswet zal leiden tot een afname van het aantal conflicten op het gebied van ruimtelijke ordening dat door de rechter moet worden beslecht en in een wijziging van de aard en complexiteit van die conflicten. Het ministerie van BZK wil (laten) onderzoeken in welke mate dit overeenkomst met de verwachtingen.

1.1 Achtergrond

Bij het bepalen van de aard en omvang van de impact van de invoering van de Omgevingswet op de rechtspraak spelen de volgende vragen een belangrijke rol: Wat is de huidige stand van zaken rondom rechterlijke uitspraken op het gebied van omgevingsrecht? Over hoeveel/welke(?) uitspraken hebben we het, wat weten we over de aard en complexiteit? En wat is daadwerkelijk de impact van de Omgevingswet op de omvang en complexiteit van rechtszaken op het gebied van omgevingsrecht? Om inzicht te krijgen in deze vragen, die beantwoord kunnen worden met statistische informatie over rechterlijke uitspraken, heeft BZK-ICTU aan het CBS gevraagd te onderzoeken in hoeverre een zogenoemde nulmeting van de huidige stand van zaken (en in de toekomst mogelijk uitbreiding met een monitoring van de impact van de Omgevingswet) kan worden uitgevoerd met behulp van beschikbare rechterlijke uitspraken.

Omdat bij het archiveren van rechterlijke uitspraken in het bestuursrecht maar een zeer beperkt aantal kenmerken wordt geclassificeerd (bijvoorbeeld of überhaupt sprake is van omgevingsrecht), is onvoldoende gestructureerde informatie beschikbaar om dit onderzoek te kunnen baseren op meer traditionele analyses van registerdata. BZK-ICTU heeft derhalve het CBS gevraagd om in dit project gebruik te maken van innovatieve big data-technieken, zoals tekst mining en, waar nodig, machine learning, zodat gestructureerde analyseerbare informatie onttrokken kan worden aan teksten van rechterlijke uitspraken om inzicht te verkrijgen in ontwikkelingen in de aard en omvang in escalaties in het omgevingsrecht.

Als eerste stap is een haalbaarheidsstudie uitgevoerd, waarin het CBS heeft onderzocht in hoeverre het mogelijk is om, uit de tekstdocumenten van gerechtelijke uitspraken op een gestructureerde wijze kenmerken te herleiden die iets zeggen over de aard van de uitspraken, zoals het type eiser, type bestreden besluit en de uitkomst, maar bijvoorbeeld ook over de complexiteit van een zaak. Het haalbaarheidsonderzoek (en vervolgens ook de nulmeting) richt(en) zich op rechterlijke uitspraken in de periode 1 januari 2007 – 31 december 2018.¹

¹ De datum 1 januari 2007 is gekozen omdat dit de invoeringsdatum was van de Wet Ruimtelijke Ordening (WRO).

1.2 Doel

Het doel van onderhavig **haalbaarheidsonderzoek** (ook wel: **vooronderzoek** of **Fase 1**) is om de mogelijkheden voor het onttrekken van gestructureerde specifieke inhoudelijke informatie over rechterlijke uitspraken op het domein van het omgevingsrecht aan de rechterlijke uitspraken in kaart te brengen en daarmee statistisch informatie te genereren over aard en omvang. Ook levert het een beter beeld op van de methodologische, praktische en juridische implicaties, opbrengsten en kosten voor en van het opzetten van een nulmeting en vervolgmetingen. Dit haalbaarheidsrapport is een belangrijke basis voor het bepalen van de prioriteiten en het inschatten van de benodigde middelen voor de uitvoering van de volgende fase, namelijk een nulmeting.

Het uiteindelijke projectdoel (d.i., incl. nulmeting en vervolgmetingen) is om de betrokken belanghebbenden, Min BZK, ICTU, Raad voor de rechtspraak (Rvdr), Raad van State (RvS), universiteiten en andere belanghebbende organisaties op de meest optimale, (wettelijk) geoorloofde wijze, inzicht te geven in aard en omvang van rechterlijke uitspraken binnen het domein van de Omgevingswet gedurende meerdere jaren.

1.3 Aanpak

Het haalbaarheidsonderzoek is opgedeeld in de volgende 4 deelfases:²

Deelfase 1: Verzamelen van rechterlijke uitspraken m.b.t. omgevingsrecht uit het E-archief van de Rvdr, inclusief het maken van procesafspraken (praktisch en juridisch) hieromtrent.³ In hoofdstuk 2 van deze rapportage wordt nader ingegaan op de in deze deelfase geleverde data, het filteren (alleen de voor het onderzoek relevante rechterlijke uitspraken zijn behouden) hiervan en voor het onderzoek relevante kenmerken van de geleverde data.

Deelfase 2: Haalbaarheidsanalyse gericht op óf en hoe de vragen in het kader van de monitor geautomatiseerd kunnen worden beantwoord op basis van rechterlijke uitspraken. In hoofdstuk 3 wordt ingegaan op de haalbaarheidsbevindingen. Hierbij wordt in een schematisch overzicht – per aan de data te onttrekken indicator – informatie verschaft over:

- Hoe goed er in het haalbaarheidsonderzoek naar is gekeken;
- Hoe goed het naar verwachting uit de data te halen is met de gekozen methodiek;
- In hoeverre de data scientists hier naar verwachting ondersteuning door inhoudelijk experts bij nodig zullen hebben;
- Hoeveel inspanning het naar verwachting kost de gewenste informatie uit de data te halen;
- Wat de te verwachten kwaliteit van de resulterende dataset is.

² Voor een op een aantal punten meer uitgebreide toelichting op deze deelfases wordt verwezen naar het 'Projectvoorstel Monitor Omgevingswet – vooronderzoek.pdf', dd. 8-7-2019. In dit eerdere document is het plan van aanpak voor het vervolg van de nulmeting overigens opgenomen als zijnde onderdeel van deelfase 3 en is sprake van een deelfase 4 die betrekking heeft op het presenteren van de resultaten (dit rapport).

³ De gemaakte procesafspraken vallen buiten het bestek van deze rapportage.

Verder is bekend dat een deel van de rechterlijke uitspraken ontbreekt in het E-archief. Er is echter nog geen goed beeld van de mate van onvolledigheid en selectiviteit hierin. Derhalve zou in dit haalbaarheidsonderzoek een analyse worden uitgevoerd van het aantal ontbrekende rechtspraken m.b.t. omgevingsrecht in het E-archief. Ten behoeve van deze analyse zou de Rvdr een selectie/telling maken van omgevingsrecht-gerelateerde uitspraken uit de financiële administratie. Daarnaast zou worden onderzocht in hoeverre het theoretisch mogelijk is om vanuit de financiële administratie van de Rvdr relevante kenmerken te halen en (op zaaknummer) door het CBS te laten koppelen aan de extractie uit het E-archief, die kunnen helpen bij het beantwoorden van de inhoudelijke vragen. Deze exercities zijn ultimo november 2019 nog niet uitgevoerd en zijn derhalve niet opgenomen in deze rapportage.

Deelfase 3: Trendanalyse van de rechterlijke uitspraken: op basis van de uitkomsten van Deelfase 2 zijn één of enkele kenmerken gekozen die relatief het minst complex te bepalen zijn. Hiervoor is een basisindicator berekend, waarmee een eerste voorlopig beeld van de trend in de onderzochte periode gegeven kan worden. Vanwege het voorlopige karakter hiervan is dit niet opgenomen in de haalbaarheidsrapportage.

Deelfase 4: Plan van aanpak nulmeting Op basis van de conclusies uit het haalbaarheidsonderzoek wordt een plan van aanpak voorgesteld met betrekking tot de vervolgstappen ('start nulmeting') en, op hoofdlijnen, voor het vervolg van de nulmeting (zie hoofdstuk 5).

Betrokken partijen

De uitvoering van dit haalbaarheidsonderzoek is een gezamenlijke inspanning van Min BZK, ICTU, de Rvdr, de betrokken universiteit(en), de RvS en het CBS. Deze partijen zijn ook allen vertegenwoordigd in de inhoudelijke begeleidingsgroep.

1.4 Methode

Analysemethodes

Voor het classificeren van de in dit onderzoek relevante kenmerken in de rechterlijke uitspraken is gebruik gemaakt van tekstanalyse middels Bag-of-Words (BoW) in combinatie met tf-idf (afkorting van 'Term Frequency – Inversed Document Frequency') en een Support Vector Machine (SVM) classifier. Met BoW wordt een stuk tekst vertaald naar een vector met getallen, waar elke locatie in de vector een woord representeert en zijn waarde het aantal keer dat dit woord voorkwam in de tekst. Met tf-idf wordt bepaald hoe belangrijk bepaalde woorden zijn in een tekst. Op basis van de tf-idf worden woorden uit de vector verwijderd als ze te vaak of te weinig voorkomen. Voor de huidige resultaten zijn woorden die minder dan 5 keer voorkwamen weggelaten, en is de top 1% van de woorden ook verwijderd. De resulterende vector is vervolgens door de SGD classifier gebruikt om de resultaten voor de verschillende kenmerken te voorspellen. Om de classifier te trainen en de parameterinstellingen te kunnen optimaliseren is gebruik gemaakt van 5-fold cross validation.

Tevens is in deze haalbaarheidsfase een beperkte steekproef van de rechterlijke uitspraken (n=360) op een deel van de relevante kenmerken door 3 personen (niet-juristen) geannoteerd, om enerzijds zicht te krijgen op de haalbaarheid van het annoteren van

gerechtelijke uitspraken door niet-juristen en om anderzijds tot een geannoteerd databestand te komen, op basis waarvan uitspraken kunnen worden gedaan over de validiteit van de analyses.

Preprocessing

Omdat de te onderzoeken rechterlijke uitspraken in het domein van het omgevingsrecht middels ruim gekozen steekwoorden zijn onttrokken uit het E-archief van de RvdR⁴ is eerst een filtering toegepast op de geleverde zaken. De hierbij gebruikte methodiek staat beschreven in paragraaf 2.1.

De resulterende – en over het algemeen vergelijkbaar gestructureerde – rechterlijke uitspraken in het domein van het omgevingsrecht, zijn vervolgens met behulp van reguliere expressie (RegEx) gesplitst in verschillende tekstblokken. Een reguliere expressie is een manier om patronen te beschrijven waardoor een computer softwarematig tekst kan herkennen. Deze splitsing is gemaakt omdat kenmerken vaak in specifieke tekstblokken van een uitspraak te vinden zijn en het algoritme zo minder ruis binnen krijgt. Dit staat verder beschreven in paragraaf 3.1.

De tekst van de documenten wordt onderworpen aan nog een aantal preprocessing stappen. Eerst worden zaken zoals witregels, interpunctie en getallen verwijderd. Daarna worden namen van maanden vervangen door het woord “maand”. Dit om te voorkomen dat het algoritme patronen gaat herkennen vanuit datums. Vervolgens wordt er gekeken naar veel voorkomende woorden, ook wel stopwords genoemd. Dit zijn woorden zoals: ik, wij, de, het, een, die, dat, enz. Omdat dit soort woorden in bijna alle documenten voorkomen hebben deze woorden bijna geen voorspellende waarde deze worden verwijderd.

De laatste stap bestaat uit het verwijderen van hoofdletters en het toepassen van een *stemmer*. Een stemmer verwijdert de vervoegingen van een woord. Bijvoorbeeld zowel “uitspraak” als “uitspraken” worden gereduceerd tot “uitsprak”. Hierdoor is het effect van vervoegingen van woorden gereduceerd en worden deze twee woorden ook als hetzelfde woord behandeld in een BoW-vector.

Voor het detecteren van het aantal eisers, is de stemmer achterwege gelaten, aangezien het gebruik van woorden zoals “verzoeker” en “verzoekers” wel degelijk een aanduiding kunnen geven over het aantal eisers.

Om het soort eiser te detecteren, is er ook nog gebruik gemaakt van een Named Entity Recognition (NER) algoritme. Hiermee wordt een poging gedaan om personen, locaties en organisaties te identificeren in een stuk tekst. Op basis van deze methode wordt onderscheid gemaakt tussen personen en bedrijven/stichtingen/overheid.

⁴ Of de zaak waarop een uitspraak betrekking heeft een omgevingsrechtzaak is of niet, valt namelijk niet goed te halen uit de meta-informatie van de rechterlijke uitspraken

2. Data rechterlijke uitspraken

Voor het haalbaarheidsonderzoek ten behoeve van de nulmeting van de effecten van de Omgevingswet op de rechtspraak is gebruik gemaakt van digitaal beschikbare rechterlijke uitspraken in het omgevingsrecht-domein. Hiervoor zijn zowel uitspraken door rechtbanken als uitspraken door de RvS van belang en zowel uitspraken die zijn gepubliceerd op rechtspraak.nl, als niet-gepubliceerde uitspraken. Voor dit onderzoek is gekozen uitspraken te analyseren uit de periode 1-1-2007 tot en met 31-12-2018.

Inmiddels weten we dat zowel de RvdR als de RvS documenten met een relatief vaste structuur hebben en ook beschikbaar kunnen stellen voor dit onderzoek. Dit is een mooi en belangrijk startpunt voor het verdere onderzoekstraject. In het haalbaarheidsonderzoek hebben we ook gekeken naar de hiaten en aandachtspunten die we op dit moment nog zien in de beschikbare data. Op dit punten zal in dit hoofdstuk vooral nader worden ingegaan.

De rechterlijke uitspraken waarop dit haalbaarheidsonderzoek feitelijk betrekking heeft, zijn eind juli 2019 aan het CBS aangeleverd door de RvdR. De RvdR heeft hiertoe, met behulp van een vooraf samengestelde steekwoordenlijst, rechterlijke uitspraken onttrokken aan haar E-archief die betrekking zouden kunnen hebben op het omgevingsrecht. Dit leverde bijna 284.000 documenten met rechterlijke uitspraken op uit de periode 1-1-2007 tot en met 31-12-2018. Elk document bevat de tekst van een rechterlijke uitspraak en bijbehorende metadata (bijv. datum uitspraak, rechtelijke instantie, uniek identificatienummer). Door de wijze van onttrekken aan het E-archief bevatte de geleverde dataset nog documenten dubbel en bovendien was de verwachting dat een deel van de documenten geen betrekking zou hebben op het omgevingsrecht. Omgekeerd kan worden verwacht dat de wijze van onttrekken heeft geleid tot het ontbreken van een deel van de uitspraken die wel degelijk betrekking hebben op het omgevingsrecht. Op zowel deze 'overdekking' als 'onderdekking' wordt in de navolgende paragrafen nader ingegaan.

2.1 Filtering uitspraken op domein omgevingsrecht

Omdat relevante rechterlijke uitspraken uit het E-archief te onttrekken is gebruik gemaakt van een trefwoordenlijst. De RvdR heeft het E-archief per zoekterm apart bevestigd en de resulterende bestanden met uitspraken aan het CBS geleverd. Dit is een waardevolle eerste 'filtering' op alle rechterlijke uitspraken in het E-archief. Vooraf werd er echter al wel van uit gegaan dat de selectie op basis van enkelvoudige⁵ trefwoorden te ruim zou zijn voor het doel van het onderzoek. De term 'ontheffing' is meermaals als voorbeeld gebruikt voor een woord dat ook veel niet relevante uitspraken op zou kunnen leveren.

De aangeleverde set van documenten bleken inderdaad nog een substantiële hoeveelheid uitspraken te bevatten die geen betrekking hebben op het omgevingsrecht. Documenten met enkel de zoekterm ontheffing bleken bijvoorbeeld regelmatig betrekking te hebben op civielrecht en meer specifiek familierecht (ontheffingen uit het ouderlijk gezag).

⁵ De RvdR heeft helaas geen mogelijkheden om te filteren op basis van combinaties van zoektermen; dus bijvoorbeeld alleen uitspraken mee te nemen op het moment dat twee of meer zoektermen in combinatie voorkomen.

De bijna 284.000 aangeleverde documenten met rechterlijke uitspraken zijn allereerst ontdebeld op basis van hun unieke ECLI-nummer (European Case Law Identifier ; dit is bij alle uitspraken opgenomen in de meta-informatie). Dit was nodig doordat uitspraken die meerdere van de trefwoorden bevatten ook meermaals zijn aangeleverd. Vervolgens is stapsgewijs verder gefilterd op relevante zaken binnen de doelpopulatie. Zo worden omgevingsrecht-zaken altijd behandeld bij een reguliere rechtbank of bij de RvS en nooit bij een andere rechtelijke instantie. Informatie over de rechtelijke instantie is voor elke uitspraak bekend via de meta-data, dus dit kon eenvoudig worden gefilterd. Ook zijn alleen rechterlijke uitspraken bewaard die gecategoriseerd waren als behorend tot het bestuursrecht; meer specifiek die één van de volgende ‘rechtsgebieden_identifiers’ bevatten in de meta-data: ‘bestuursrecht_geen’, ‘bestuursrecht_bestuursprocesrecht’ en ‘bestuursrecht_omgevingsrecht’.⁶ Onbruikbare uitspraken zijn er eveneens uit gefilterd. Dit zijn uitspraken met het type “jurisprudentie” (deze documenten bevatten geen rechterlijke uitspraken) en documenten waar de tekst minder dan 150 tekens bevat (dit zijn documenten waar vaak foutmeldingen vermeld staan of de mededeling dat een zaak gesloten is zonder uitspraak).⁷

TABEL 2.1 Filtering van de geleverde rechterlijke uitspraken

Stap	Omschrijving	Resterend aantal uitspraken
0.	Initieel geleverde documenten Rvdr	283.612
1.	1. Ontdubbelen van dubbel geleverde documenten op basis van ECLI 2. Filteren op instantietype (alleen Rechtbank en RvS behouden) 3. Verwijderen jurisprudentie-stukken 4. Verwijderen bestanden zonder tekst of met tekstveld <150 tekens	88.769
2.	Behoud alleen documenten die volgens de meta-informatie betrekking hebben op omgevingsrecht, bestuursrecht algemeen en bestuursprocesrecht	69.977
3.	Filter zaken op woorden zoals “erfrecht”, “kinderbescherming”, etc. ⁸	67.618
4.	Filter lastig geschreven proces-verbalen eruit, om in een later stadium apart te behandelen	65.768
5.	Verwijder Anonieme documenten (ofwel: filter alle RvS-zaken uit)	35.475

Na nadere bestudering van de overgebleven zaken (middels handmatig controleren van een beperkt aantal uitspraken) is met behulp van RegEx nog een nadere uitfiltering toegepast op de termen als ‘erfrecht’ en ‘kinderbescherming’. Deze filtering resulteerde in een bestand met ruim 67.000 relevante rechterlijke uitspraken. Zie Tabel 2.1 voor een overzicht van dit filteringsproces.

⁶ Op advies van prof. Floris Bex, is ervoor gekozen deze filtering tenminste voor het haalbaarheidsonderzoek toe te passen.

⁷ Indien deze documenten wel relevant worden geacht voor de nulmonitor omdat het relevant zou zijn om te weten hoe vaak er geen uitspraak is gedaan in een zaak, dan kunnen deze zaken alsnog worden betrokken in het onderzoek

⁸ Volledige woordenlijst: "proces-verbaal van het verhandelde", "proces-verbaal van de verhandelde", "proces-verbaal van de zitting van", "proces-verbaal van de behandeling ter zitting", "proces-verbaal van de behandeling", "de rechtbank sluit het onderzoek ter zitting en deelt mee dat op", "bijlage, inhoudende de bewijsmiddelen in de zaak", "aanvulling bewijsmiddelen", "samenvatting van de stappen voor het opslaan van een", "opslaan van uitspraken in het e-archief", "fout! bladwijzer niet gedefinieerd", "kinderbescherming", "sector familie- en jeugdrecht", "deze uitspraak is vervallen verklaard", "proces-verbaal van de comparitie", "proces-verbaal van comparitie", "schuldsanering", "erfrecht", "kinderrechter", "team jeugdrecht", "team handel - voorzieningenrechter", "kantonrecht", "kanton", "kopje", "deze uitspraak is ambtshalve vervallen", "nakomen koopovereenkomst", "faillissementswet", "belastingdienst", "vreemdelingenzaken", "hoogheemraadschap", "het college van dijkgraaf"

Voor het haalbaarheidsonderzoek, waar deze rapportage betrekking op heeft, zijn vervolgens nog bijna 2.000 proces-verbalen uitgefilterd omdat deze een nogal afwijkende tekststructuur kennen⁹ en uiteindelijk zijn eveneens **de uitspraken van de RvS vooralsnog buiten beschouwing gelaten**, omdat deze geanonimiseerd bleken te zijn (zie ook paragraaf 2.2). De haalbaarheidsanalyses zijn verricht op de resulterende ruim 35.000 zaken.

In dit hoofdstuk wordt voorts ingegaan op de eigenschappen van de resulterende dataset met gefilterde data en op implicaties hiervan voor de haalbaarheid.

2.2 Ontbreken niet geanonimiseerde RvS-uitspraken

De data voor het onderzoek zijn geleverd door de bronhouder – de Rvdr. Pas in de loop van het haalbaarheidsonderzoek (na de filtering en splitsing van uitspraken), eind sept 2019, werd duidelijk dat het E-archief waar de Rvdr uit put wel alle gedigitaliseerde uitspraken van de rechtbanken bevat, maar van de RvS-zaken enkel de (geanonimiseerde) uitspraken voor zover deze zijn gepubliceerd op rechtspraak.nl.

Omdat het type zaak bepaalt of deze bij een rechtbank wordt behandeld of bij de RvS (beroep versus hoger beroep, maar ook het type onderliggend besluit waarover een conflict bestaat), zal het ontbreken van (niet geanonimiseerde) uitspraken van de RvS leiden tot selectiviteit in de data. Omdat de geanonimiseerde uitspraken niet de voor het onderzoek benodigde informatie bevatten over eisers en verweerders zijn ze buiten beschouwing gelaten worden voor nadere analyses. Het haalbaarheidsonderzoek waar deze rapportage betrekking op heeft is dus gebaseerd op alleen de rechtbankzaken. Bevindingen op basis van de huidige data kunnen daarom in deze fase nog niet worden gegeneraliseerd naar alle uitspraken in het omgevingsrecht.¹⁰

Voor de volledigheid van een nulmeting is het van belang dat de rechterlijke uitspraken van RvS-zaken alsnog worden toegevoegd aan de onderzoeksdata. Op 12-11-2019 heeft de RvS aan BZK/ICTU laten weten in te gaan op dit verzoek en op 27 november zijn hierover ook al nadere afspraken gemaakt. Hiertoe zullen, net als eerder tussen BZK/ICTU-CBS-Rvdr, procesafspraken (praktisch en juridisch) tussen BZK/ICTU-CBS-RvS worden gemaakt.

In hoofdstuk 5 en ook de oplegger bij dit document voor BZK wordt extra aandacht besteed aan het toevoegen van deze data aan het onderzoek.

⁹ Idem op advies van prof. Floris Bex.

¹⁰ Het is tot op heden niet precies bekend welk aandeel de RvS-zaken hebben in de totale bulk van uitspraken binnen het omgevingsrecht. Echter: van de aan het CBS geleverde uitspraken in het omgevingsrecht-domein bleek na filtering ca. 46% geanonimiseerd. Dit zijn de RvS-uitspraken die zijn gepubliceerd op rechtspraak.nl. Marcel Hoogwout gaf in een gesprek op 9-10-2019 aan dat van de RvS-zaken naar verwachting ca. 90% ontsloten is via rechtspraak.nl. Als deze percentages specifiek ook binnen het omgevingsrecht-domein bij benadering zouden kloppen, zou dit betekenen dat grofweg de helft van alle rechterlijke uitspraken binnen het omgevingsrecht RvS-zaken zouden betreffen.

2.3 Onvolledige selectie uit E-archief en selectiviteit

Zoals eerder genoemd heeft de RvdR voor dit onderzoek een eerste waardevolle selectie kunnen maken uit hun E-archief met daarin alle gerechtelijke uitspraken. Er zijn wel enkele kanttekeningen bij deze wijze van selecteren.

De selectie op basis van een trefwoordenlijst uit het E-archief heeft enerzijds geleid tot de aanlevering van niet relevante zaken (zie paragraaf 2.1), maar het is niet onwaarschijnlijk dat het omgekeerd ook heeft geresulteerd in het – los van ontbrekende RvS-zaken (zie paragraaf 2.2) – missen van een deel van de relevante uitspraken. Zo is de trefwoordenlijst samengesteld met actuele woorden en is mogelijk niet voldoende stilgestaan bij trefwoorden die in de eerste jaren, voor inwerkingtreding van de Waterwet in 2009 en de Wabo in 2010, relevant zouden zijn geweest (een vorm van ‘concept drift’). Als dat zo is, zouden niet alleen relevante zaken zijn gemist bij de samenstelling van het onderzoeksbestand, maar zou dit bovendien selectief zijn over de tijd, met meer missende zaken in de eerdere jaren. Dit zou een vertekend kunnen opleveren van trends.

Bij de RvdR is bovendien bekend dat niet alle rechterlijke uitspraken worden opgenomen in het E-archief. Er is dus sprake van onderdekking. Er is echter geen beeld van hoe groot die onderdekking is in termen van percentages en bovendien is onbekend in hoeverre hierbij sprake is van selectiviteit. Het is bijvoorbeeld mogelijk dat de kans dat een uitspraak wordt opgenomen in het e-archief samenhangt met iets als de complexiteit van een zaak. Of dat in de meer recente jaren een relatief groter deel van de zaken in het E-archief wordt opgenomen dan in de eerdere jaren van de onderzoeksperiode.

Om zicht te krijgen op het aantal en de selectiviteit van ontbrekende omgevingsrecht-gerelateerde uitspraken in het E-archief is bij aanvang van het haalbaarheidsonderzoek afgesproken een analyse van het aantal missende uitspraken te doen op basis van een vergelijking met tellingen van omgevingsrecht-gerelateerde rechterlijke uitspraken uit de (wel volledige) financiële administratie van de Rvdr. Op basis daarvan ontstaat een indicatie van het aantal/aandeel missende uitspraken in het E-archief. Dit deel van het haalbaarheidsonderzoek is nog niet uitgevoerd, maar is dus wel zinvol voor het vervolg.

Bovenstaande zou inzicht opleveren in de omvang van ontbrekende uitspraken, maar zou ons nog niets leren over mogelijke selectiviteit hierbij. Om hier enig zicht op te verkrijgen is in de aanloop naar het haalbaarheidsonderzoek de mogelijkheid besproken dat de Rvdr, middels een kwalitatieve uitvraag bij Rechtbanken, navraag zou doen naar de aard en omvang van niet-gedigitaliseerde rechterlijke uitspraken. Hierover zijn op dit moment nog geen concrete afspraken uitgewerkt. Om zicht op mogelijke selectiviteit te krijgen zou eventueel ook voor een kwantitatieve aanpak kunnen worden gekozen: er zou door het CBS een koppeling op zaaks-/uitspraakniveau gemaakt kunnen worden tussen de geleverde zaken uit het E-archief en een lijst van ECLI-nummers van omgevingsrecht-zaken uit de financiële registratie van de Rvdr, er van uitgaande dat het mogelijk is een exacte selectie van omgevingsrecht-zaken uit deze financiële registratie te maken.

Deze selectiviteitsanalyses (zowel kwalitatief als kwantitatief) vielen buiten de scope van het haalbaarheidsonderzoek en dus buiten deze rapportage. Voor de (duiding van) kwaliteit en interpretatie van uitkomsten van een uiteindelijke nulmeting zal dit issue gedresseerd moeten worden. Stel dat vooral weinig complexe uitspraken vaak niet worden geregistreerd in het E-archief, dan zou dat de uitkomsten uit dit onderzoek over (trends in) mate van complexiteit van omgevingsrecht-gerelateerde uitspraken kunnen beïnvloeden.

2.4 Voorzieningenrechterzaken

Van de ruim 35.000 rechterlijke uitspraken die in het haalbaarheidsonderzoek zijn meegenomen, bleken er minstens 8.000 (24,1%) uitspraken door de voorzieningenrechter te zijn. In het haalbaarheidsonderzoek zijn deze uitspraken tezamen met de uitspraken ten principale geanalyseerd. Omdat deze uitspraken deels een afwijkende structuur en kenmerken hebben, kan dit de resultaten van het haalbaarheidsonderzoek enigszins (nadelig) hebben beïnvloed, in die zin dat het classificeren van de gevraagde kenmerken minder goed mogelijk leek dan dat dit het geval zal zijn wanneer de uitspraken door de voorzieningenrechter buiten beschouwing zouden zijn gelaten of apart zouden zijn bestudeerd.¹¹ Overigens nemen we aan dat eventuele effecten op het haalbaarheidsonderzoek vrij beperkt zijn, aangezien de haalbaarheidsanalyses zich met name hebben gericht op uitspraak-onderdelen. Dat zijn onderdelen waarop uitspraken door de voorzieningenrechter weinig af lijken te wijken van uitspraken ten principale.

¹¹ Op 14-11-2019 is een voorlopig besluit genomen door Marcel Hoogwout om deze zaken apart te beschouwen. Wel is de wens geuit om het aantal uitspraken door de voorzieningenrechter in kaart te brengen en te onderzoeken of het mogelijk is om na te gaan welk deel van de uitspraken van de voorzieningenrechter wel/niet uitmondt in een uitspraak ten principale. Dit laatste is in de praktijk lastig, omdat hiervoor een micro-koppeling gemaakt zou moeten worden tussen uitspraken op basis van tekstgedeeltes in deze uitspraken. Wel zal in ieder geval voor de uitspraken ten principale worden onderzocht of er voorafgaand sprake was van een zaak bij de voorzieningenrechter (dit wordt in principe aangegeven in de uitspraak). Dit onderwerp zal nog worden besproken in de begeleidingscommissie.

3. Haalbaarheidsanalyses

In het haalbaarheidsonderzoek zijn verschillende stappen gezet teneinde conclusies te kunnen trekken over de mogelijkheid om uit ongestructureerde data kenmerken te classificeren. In dit hoofdstuk worden deze stappen toegelicht. Eerst is de structuur van de gerechtelijke uitspraken in kaart gebracht, om zo te bepalen waar gezocht moest worden voor bepaalde kenmerken. Vervolgens zijn de uitspraken handmatig voorzien van labels, ook wel annotaties genoemd, voor de verschillende kenmerken. Aan de hand van deze labels kan uiteindelijk gekeken worden of de indicatoren ook de juiste waardes voorspellen. Vanaf hoofdstuk 3.3 worden vervolgens de bevindingen van de haalbaarheidsanalyses besproken.

3.1 Verkenning tekststructuur en betrouwbaarheid van splitsing van de data

Alvorens te starten met het categoriseren van kenmerken, zijn de uitspraken gesplitst in verschillende tekstblokken. Dit is gedaan omdat de te onderzoeken kenmerken vaak in specifieke tekstblokken zijn te vinden en verondersteld wordt dat het de resultaten van algoritmes ten goede komt wanneer zij doelgericht in worden gezet op specifieke tekstblokken.

De rechterlijke uitspraken blijken over het algemeen een onderling vergelijkbare structuur te kennen. Een aantal onderdelen zijn praktisch altijd aanwezig, namelijk:

- De **kop** met bovenaan de datum en rechtsprekende instantie en vaak ook de algemene strekking van de zaak. Vervolgens worden de eisers, verweerders en veelal ook eventuele derden benoemd.
- Het **procesverloop** met informatie over het onderliggende bestreden besluit en over het verloop van het proces (voorgeschiedenis) tot aan de zitting.
- De **overwegingen** met informatie over de twistpunten en de overwegingen van de rechtbank daaromtrent.
- En de **beslissing/conclusie** met daarin de uitspraak/conclusie van de rechtbank en ook wie er (namens eiser(s), verweerder(s) en eventuele derde partij(en) aanwezig was/waren bij de zitting.

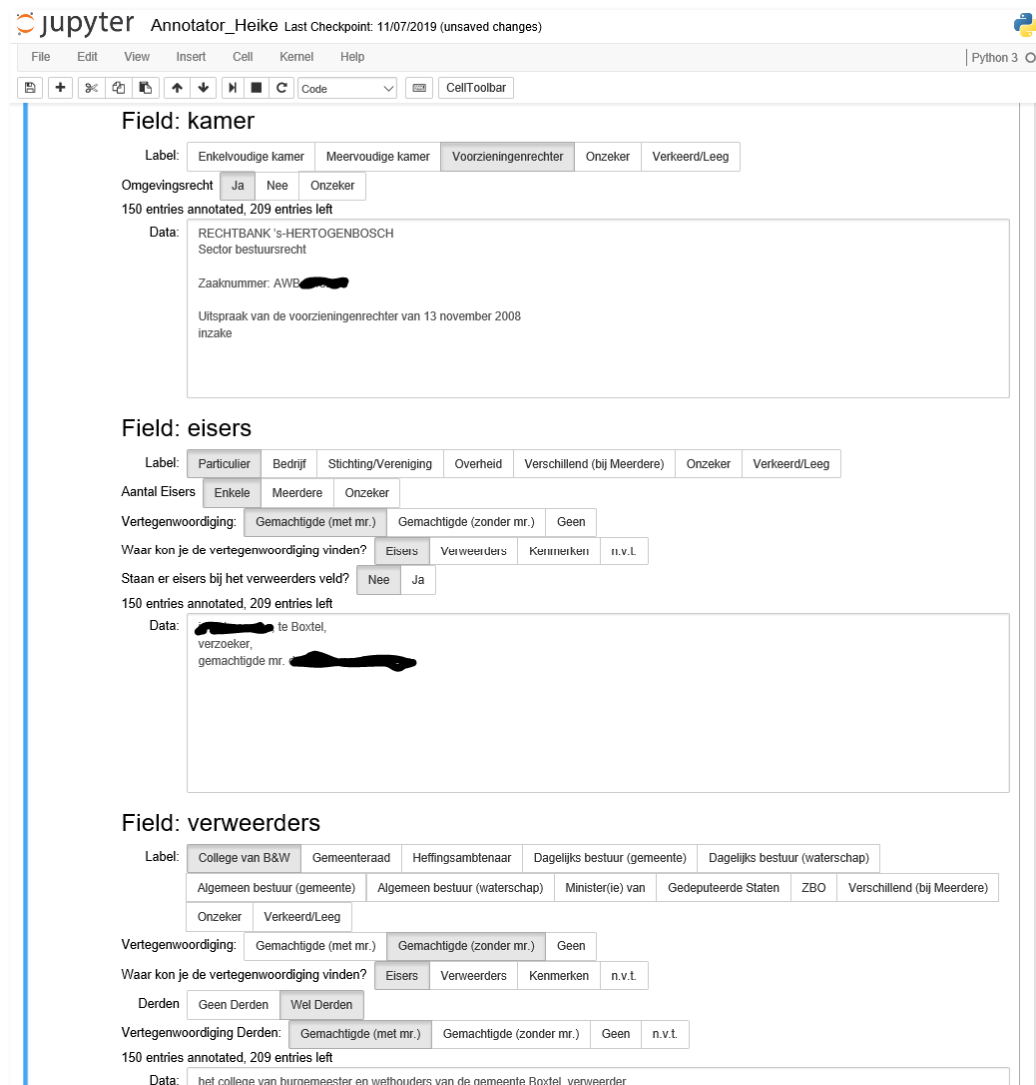
De splitsing kon in het merendeel van de documenten met behulp van reguliere expressies succesvol worden uitgevoerd. De keuze voor de patronen in deze reguliere expressies zijn voornamelijk handmatig gekozen. Om de uitvoering van de reguliere expressies te versnellen zijn deze vervolgens in een boomstructuur opgeslagen. Door een reguliere expressie op te bouwen als een boom, met op elke splitsing een letter, kunnen vrij grote/complexere reguliere expressies snel uitgevoerd worden. Tot slot is de koptekst in drie delen uitgesplitst, zodat de rechtsprekende instantie, de eisers en de verweerders in losse tekstblokken kwamen te staan.

3.2 Het maken van een geannoteerd databestand

Voor deze haalbaarheidsfase is een beperkte steekproef (n=360) van in totaal ruim 35.000 rechterlijke uitspraken in het onderzoek getrokken ten behoeve van het construeren van een geannoteerd databestand. Voor de steekproef is uitgegaan van een willekeurige selectie van 120 uitspraken uit elk van de drie voor het onderzoek geselecteerde rechtsgebieden die in de meta-informatie van de uitspraken is opgenomen, te weten omgevingsrecht, bestuursrecht algemeen en bestuursprocesrecht.

Deze uitspraken zijn op een deel van de voor het onderzoek relevante kenmerken door 3 onderzoekers onafhankelijk van elkaar geannoteerd, d.i. de te onderzoeken kenmerken zijn door deze drie personen handmatig gelabeld. Dit betroffen grofweg de kenmerken zoals die op 1 oktober 2019 door de begeleidingscommissie van het onderzoek zijn vastgesteld (zie paragraaf 3.3 voor meer informatie over deze kenmerken). Het annoteerwerk is verricht in een speciaal hiervoor gemaakte tool (zie figuur 3.1).

FIGUUR 3.1 Screenshot van annotatietool



Daar waar tenminste 2 van de 3 annoteurs het met elkaar eens waren over het classificeren van een kenmerk, is de betreffende waarde direct overgenomen in het geannoteerde databestand. De verdeling van deze consensus per kenmerk is te zien in Tabel 3.1. De uitspraken waar de annoteurs alle drie verschillend op scoorden zijn door hen opnieuw doorlopen om alsnog tot consensus te komen.

Het resulterende geannoteerde databestand is vervolgens gebruikt om een aan de hand van de resultaten van getrainde machine learning algoritmes een indicatie te kunnen geven over hoe goed de gevraagde kenmerken zijn te detecteren.

TABEL 3.1 Annotatie consensus voor de verschillende kenmerken

Kenmerken	Annotatie Consensus (%)
BESTREDEN BESLUIT	
Soort besluit	88,3
Aard besluit	88,9
Reguliere vs. uitgebreide procedure	87,8
EISER(S)	
Soort eiser	98,3
Aantal eisers	98,1
Gemachtigde (eisers)	93,9
VERWEERDER(S)	
Soort verweerder	96,1
Gemachtigde (verweerder)	85
DERDE(N)	
Derden/Belanghebbende	93,9
Gemachtigde (derden)	86,4
COMPLEXITEIT	
Soort kamer	96,9

3.3 Haalbaarheidsbevindingen per kenmerk

In de beoogde monitor worden verschillende thema's (groepen van kenmerken) onderscheiden met daarin (deels nog te bepalen) kenmerken waarop gelabeld dient te worden:

1. Basiskenmerken die nu al in de metadata van de uitspraken zijn opgenomen zoals: rechtsprekende instantie en rechtsgebied.
2. Kenmerken onderliggende overheidsbesluit: wat voor soort besluit wordt bestreden.
3. Kenmerken doorlopen voortraject: welke eerdere instanties zich over de zaak hebben gebogen.
4. Kenmerken eiser(s): wat voor soort eisers beroep instellen en hoe zij zich laten vertegenwoordigen.
5. Kenmerken verweerder(s): tegen wie beroep is ingesteld en hoe zij zich laten vertegenwoordigen.
6. Kenmerken derde(n): of er een derde partij betrokken is en hoe zij zich laten vertegenwoordigen.

7. Kenmerken van hetgeen waar onenigheid over bestaat: bijvoorbeeld in de vorm van wetsartikelen, beroepsgronden, leerstukken.
8. Kenmerken uitkomsten van het geding: in hoeverre het bestreden overheidsbesluit standhoudt
9. Kenmerken die complexiteit van de uitspraak bepalen: bijv. behandeling in enkelvoudige of meervoudige kamer en/of de redenen waarom daarvoor wordt gekozen.

De te labelen kenmerken binnen thema's 2, 4, 5 en 6 zijn vastgesteld na de bijeenkomst van de begeleidingscommissie op 1 oktober 2019. De haalbaarheidsanalyses hebben zich dan ook voornamelijk toegespitst op deze kenmerken. Omdat de overige kenmerken nog vastgesteld dienen te worden, worden hier slechts globale haalbaarheidsinschattingen gegeven.

In schema 3.2 wordt per kenmerk middels kleuren informatie verschaft over (1) hoe goed er in deze haalbaarheidsfase naar is gekeken (*groen = voldoende goed voor betrouwbare uitspraak, oranje = alleen naar bepaalde elementen gekeken, grijs = niet/nauwelijks naar gekeken; buiten scope haalbaarheidsonderzoek*), (2) hoe goed het kenmerk naar verwachting uit data te halen is met de voor dit onderzoek voorgestelde methoden (*groen = dit lijkt betrouwbaar te kunnen, oranje = nog onzeker, grijs = onbekend*), (3) in hoeverre hiervoor advies nodig zal zijn van inhoudelijk/juridisch experts (*groen = niet, oranje = minimaal enig advies nodig, grijs = nog onbekend*) en (4) hoeveel inspanning er naar verwachting nodig is de gevraagde informatie uit de data te halen (*groen = dit kan met beperkte middelen; oranje = dit kost meer inspanning/meer training van de data, grijs = nog onbekend*).

In Bijlage I is voorts een uitgebreider schema opgenomen, waar de informatie per kenmerk nader is uitgesplitst naar categorie.

Voor het bepalen hoe goed kenmerken uit de data te halen zijn en wat de verwachte kwaliteit van de resulterende data zal zijn, zijn voor de reeds onderzochte kenmerken (hiervan is de eerste kolom in het schema groen), per categorie, de precision, recall en F1-scores berekend. De kleur van de betreffende cellen is bepaald op basis van deze scores. De scores zelf zijn voor de overzichtelijkheid niet in het schema opgenomen.

SCHEMA 3.2 Haalbaarheidsbevindingen per kenmerk

KENMERKEN	Hoe goed onderzocht?	Hoe goed uit data te halen met voorgest. methoden?	Advies nodig inh./jur. experts?	Hoeveel inspann. dit uit data te halen?
EISER(S)				
Soort eiser(s)				
Aantal eisers				
Vertegenwoordiging				
VERWEERDER(S)				
Soort verweerder(s)				
Vertegenwoordiging				
BELANGHEBBENDE(N)/DERDEN				
Sprake van belanghebbende(n)/derde partij(en)				
Vertegenwoordiging				
BESTREDEN BESLUIT				
Soort besluit				
Soort procedure				
Aard besluit				
TWISTPUNT(EN) [nog nader vaststellen]				
Materieel recht versus procesrecht (Hoger) beroep heeft betrekking op in te vullen open normen				
Wettelijke normen versus rechtsprincipes				
Aantal beroepsgronden				
DOORLOPEN VOORTRAJECT [nog nader vaststellen]				
<i>Welke eerdere instanties hebben zich over de zaak gebogen?</i>				
COMPLEXITEIT UITSPRAAK [nog nader vaststellen]				
Enkelv. of meerv. kamer				
Aantal rechtsoverwegingen				
Activiteit(en) waarvoor verg. werd aangevraagd (bij vergunning)				
Lengte uitspraak				
Aantal eisers (i.p.v. 1 of meer)				
Doorlooptijd oorspronkelijk besluit en uitspraak				
UITKOMST GEDING [nog nader vaststellen]				
Gegronnd (tenm. op 1 punt)				
Ongegr. (tenm. op 1 punt)				
Niet ontv. (tenm. op 1 punt)				
Overig (bijv. onbevoegd) (tenm. op 1 punt)				

3.4 Toelichtingen per kenmerk

De algemene conclusies in schema 3.2 worden in deze paragraaf voor ieder van de kenmerken nader toegelicht. De kenmerken met een grijze kleur in het schema, worden hier niet behandeld, aangezien deze nog niet in voldoende detail zijn bekeken. De meeste van deze kenmerken zijn ook niet triviaal met bijvoorbeeld een reguliere expressie uit de tekst te halen, waardoor er nog wel extra naar gekeken moet worden.

Soort eiser(s): Voor de soort eiser is het nog lastig de entiteiten te vinden zoals Particulier, Bedrijf, Stichting enz. Het gaat doorgaans om eigennamen en doordat de lijst met mogelijke namen praktisch oneindig is, is het lastig deze goed uit elkaar te halen. Door te werken met Named Entity Recognition, kan er wel een redelijk splitsing tussen personen en organisaties

gemaakt worden, maar ook deze techniek is niet perfect. Dit kenmerk heeft dus nog een aantal haken en ogen waar nog oplossingen voor moeten worden gevonden.

Aantal eisers: Als wordt gekeken naar enkelvoud/meervoud kan het gedeelte tussen de kop en het procesverloop goed worden gebruikt om te bepalen of er één of meerdere eisers zijn. Hiervoor hoeft niet gekeken te worden naar opsommingen, maar meer naar woorden zoals “eiser” vs. “eisers”. In het geannoteerde bestand bleek dit consistent correct genoteerd in de tekst, waardoor zelfs met reguliere expressies goede resultaten behaald worden. Hierdoor zijn er hoge verwachtingen voor de kwaliteit van de indicator voor dit kenmerk.

Vertegenwoordiging eiser(s): Dit kenmerk kan redelijk vaak goed gevonden worden, aangezien de vermelding van de vertegenwoordiging over het algemeen in een vast patroon voor komt in het stuk tekst waar de eiser voor het eerst genoemd wordt. In dit geval kan relatief veel ruis met bijvoorbeeld de verweerder of de derden-partij verwijderd worden, waardoor de uiteindelijke kwaliteit van dit kenmerk hoog geschat wordt. Dit wordt verstrekt door de resultaten uit het haalbaarheidsonderzoek.

Soort verweerder(s): De uitspraken in het geannoteerde bestand bestaan voornamelijk uit College van B&W (meer dan 80%). Een onderscheid in twee categorieën (d.i. college van B&W vs. overig) zou hierdoor nu al goed moeten lukken. Om de overige categorieën te differentiëren zullen echter veel meer geannoteerde uitspraken en dus inspanning/tijd nodig zijn. Door de scheve verdeling in de data is het nu nog lastig hier een exacte schatting voor te geven.

Vertegenwoordiging verweerder(s): Dit kenmerk kan over het algemeen gevonden worden in hetzelfde stuk tekst als de derden-partij (en de bijbehorende vertegenwoordiging). Als het hier niet staat, staat er regelmatig alsnog een vertegenwoordiging in het procesverloop vermeld. Dit zorgt voor veel ruis met de gekozen aanpak, aangezien de vertegenwoordiging van de derden-partij en de verweerder makkelijk door elkaar gehaald kunnen worden. Voor dit kenmerk zal nog meer werk nodig zijn.

Sprake van belanghebbende(n)/derde partij(en): Dit kenmerk wordt afwisselend direct na de verweerder genoemd of anders ergens in het procesverloop (bijv. onder de noemer van “vergunninghouder”). Door de wisselende locaties in de tekst is dit kenmerk relatief lastig te classificeren, met name als de derden pas in het procesverloop genoemd worden, waar meer ruis ontstaat door de aanwezigheid van vele eigennamen. Voor dit kenmerk is dus nog meer werk nodig voor de zaken waar de derden niet gelijk na de verweerder genoemd worden, maar kan nog verbeteren.

Vertegenwoordiging derde(n): Dit kenmerk kan over het algemeen gevonden worden in hetzelfde stuk tekst als de verweerder (en de bijbehorende vertegenwoordiging). Als het hier niet staat, staat er vaak alsnog een vertegenwoordiging in het procesverloop genoemd. Dit zorgt voor veel ruis met de gekozen aanpak, aangezien de vertegenwoordiging van de derden-partij en de verweerder makkelijk door elkaar gehaald kunnen worden. Voor dit kenmerk is nog meer werk nodig, maar kan nog verbeteren.

Soort bestreden besluit: Sommige van de categorieën van dit kenmerk kunnen nog redelijk uiteenlopende zaken bevatten, waardoor het lastiger is voor de classifier om hier in goed te differentiëren. Mogelijk is dit nog een kwestie van de categorieën indelen, anderzijds is de

annotatie-set redelijk ongebalanceerd voor dit kenmerk. Er kan dus waarschijnlijk een betere indicatie gegeven worden zodra er meer geannoteerd is.

Soort procedure: Voor het classificeren of er een reguliere of een uitgebreide procedure aan de rechtszaak vooraf gaat kan men afgaan op het feit of er door de eiser(s) voorafgaand aan de zaak bezwaar is ingediend, of dat men rechtstreeks in beroep is gegaan. Aangezien dit over het algemeen in de tekst van de uitspraak wordt genoemd, kan de classifier dit relatief goed vinden. Een kleine nuance is alleen dat dit kenmerk geannoteerd is door niet-juristen, waardoor ook niet gegarandeerd kan worden dat de annotaties voldoen aan de werkelijke definitie van een reguliere/uitgebreide procedure. Voor dit kenmerk is een extra controle door een jurist dus nog wenselijk.

Aard bestreden besluit: Vooral de categorie “geweigerd” en “overig” halen de score voor dit kenmerk omlaag. Mede omdat deze relatief uiteenlopende zaken kunnen beschrijven. Tijdens het annoteren werd duidelijk dat er mogelijk nog een keer gekeken moet worden naar de gekozen categorieën voor dit kenmerk. Het lijkt erop dat een verdere opsplitsing (of samenvoeging) wenselijk is. Ook hier is advies van een inhoudsdeskundige gewenst om de resultaten te verfijnen.

Soort kamer: Dit kenmerk is in de meeste gevallen te vinden in de kop-tekst van een gerechtelijke uitspraak. Daarnaast worden de termen zoals enkelvoudige- en meervoudige kamer altijd volledig uitgeschreven. Waardoor dit kenmerk ook al met reguliere expressies makkelijk te classificeren is. De resultaten uit de haalbaarheidsfase bevestigen dit ook. Daarom is het oordeel ook dat dit kenmerk met een hoge nauwkeurigheid gevonden zal kunnen worden.

Word-count: Dit is eenvoudig uit te voeren.

Uitkomst geding: Deze zijn nog niet geannoteerd er is dus nog geen controle-bestand beschikbaar, maar er is wel al vastgesteld dat de vooralsnog te onderscheiden categorieën goed te vinden zijn in de uitspraken en de verwachting is derhalve dat dit met reguliere expressie voldoende goed mogelijk zal zijn

3.5 Overall betrouwbaarheid geclassificeerde kenmerken

Een belangrijk punt van aandacht is dat in deze haalbaarheidsfase de kenmerken één voor één zijn geclassificeerd en er niet tegelijk rekening is gehouden met alle verbanden die daartussen bestaan. Het risico hierbij is dat onderzoekers straks allerlei verbanden vinden die er in werkelijkheid niet zijn, omdat daar bij het classificeren van die kenmerken helemaal geen rekening mee is gehouden. Een model kan namelijk wel een overall score hebben die mooi is, maar dat wil niet zeggen dat het model lokaal ook voldoet, d.w.z. bij bepaalde combinaties van categorieën. Omdat niet op voorhand bekend is welke analyses uiteindelijk op het resulterende bestand uitgevoerd zullen worden of welke verbanden er zullen worden onderzocht, is het lastig hier vooraf goed rekening mee te houden. Voor het uiteindelijk op te leveren bestand voor de nulmonitor zal hiervoor een zo optimaal mogelijke oplossing moeten worden uitgewerkt.

Er wordt bovendien voorgesteld om uiteindelijk, voor gebruik van een onderzoeksbestand met geclassificeerde kenmerken, code mee te leveren, waarmee onderzoekers die totalen bepalen per categorie voor kenmerken die automatisch gelabeld zijn, informatie krijgen over de betrouwbaarheid van die totalen - een bandbreedte om het totaal en de grootte van een mogelijke vertekening.

3.6 Betrouwbaarheid classificaties binnen kenmerken

Wanneer bepaalde waarden binnen een kenmerk zeer weinig voorkomen, kan het lastig zijn deze goed te voorspellen. Zo is verreweg het grootste deel van de verweerders in het haalbaarheidsonderzoek het college van B&W. Als alle verweerders als zodanig geclassificeerd zouden worden dan doet het model het over het geheel genomen voor dat kenmerk niet eens zo slecht. Maar kijken we binnen dat kenmerk, dan zien we dat de overige categorieën altijd fout worden geclassificeerd. Wanneer men over het voorkomen van dit soort kleine categorieën een uitspraak wil doen gaat dit dus mis. In de uiteindelijke modellen zal rekening worden gehouden met dergelijke mogelijke 'disbalansen'.

4. Conclusies en plan van aanpak vervolg

Hierboven zijn de bevindingen van het haalbaarheidsonderzoek beschreven. In deze paragraaf worden de conclusies samengevat en vertaald naar een voorstel voor een vervolg.

4.1 Conclusies

Wat is al bekend/mogelijk?

- Het is mogelijk om data/documenten van rechterlijke uitspraken in het omgevingsrecht uit het systeem (E-archief) van de Rvdr te halen, te verwerken en te ontsluiten, ten behoeve van statistisch onderzoek.¹²
- Het is mogelijk om de uitspraken te ontdebellen, samen te voegen en gestructureerd op te splitsen in delen, voor machine learning, ten einde vooraf bepaalde kenmerken en categorieën te kunnen classificeren.
- Er is een werkbare tool gebouwd (prototype) voor het gericht 'annoteren' van de dataset, ten behoeve van machine learning. Deze tool kan verder worden verfijnd en uitgebreid.
- Van een deel van de kenmerken is de haalbaarheid van automatisch annoteren onderzocht en wordt verwacht dat dit met voldoende kwaliteit bepaald/berekend kan worden op basis van de teksten in de rechterlijke uitspraken. Van een deel van deze kenmerken wordt verwacht dat in de 'start van de nulmeting' een eerste prototype berekend kan worden; het gaat dan in het bijzonder om kenmerken die met behulp van Reguliere Expressies (RegEx) kunnen worden gecategoriseerd (zie paragraaf 5.2). Voor het overige, meer complexe deel zal substantieel meer tijd/capaciteit en/of input van deskundigen nodig zijn om te komen tot een eerste versie van een geclassificeerd kenmerk (zie paragraaf 3.3).

Wat is nog niet bekend/mogelijk, maar in een volgende fase mogelijk wel?

- De 'doelpopulatie' van dit onderzoek, namelijk alle omgevingsrecht-gerelateerde rechterlijke uitspraken, is nog verre van zuiver; een substantieel deel van de uitspraken is nog niet meegenomen in het haalbaarheidsonderzoek, omdat er nog geen niet-geanonimiseerde uitspraken zijn van de RvS. Daarnaast is bekend dat een onbekend deel van de relevante uitspraken nog ontbreekt doordat ze niet in de huidige query uit het E-archief zijn meegenomen, of doordat ze überhaupt niet gedigitaliseerd zijn. Hiervoor is nader onderzoek nodig.
- Een deel van de beoogde kenmerken is nog niet of slechts beperkt onderzocht in het haalbaarheidsonderzoek, omdat ze nog niet definitief zijn vastgesteld in de begeleidingscommissie. Voor deze kenmerken kunnen alleen voorlopige/voorzichtige uitspraken worden gedaan over de haalbaarheid, op basis van de informatie die er nu is m.b.t. de gewenste specificaties.

¹² Hoewel hier nog wel afspraken gemaakt moeten worden met partijen over structurele dataleveringen.

4.2 Plan van aanpak vervolg

Het vervolgplan voor een nulmeting, kan worden ingedeeld in drie fases:

<i>Fase</i>	<i>Tijdpad</i>
Start nulmeting	December 2019 – Q1 ¹³ 2020
Vervolg nulmeting	Tweede helft 2020/begin 2021
Bestending methodiek	2021

Start nulmeting – naar eerste, voorlopige resultaten

Conform gemaakte afspraken tussen BZK-ICTU en het CBS zal eind 2019 de nulmeting worden opgestart. Beoogd resultaat van deze fase was:

- Een geanonimiseerde dataset, met daarin enkele berekende kenmerken;
- Een nader uitgewerkt plan van aanpak voor het vervolg van de nulmeting;
- Een trendanalyse op basis van de kenmerken die op dat moment berekend zijn.

Mede gezien de hierboven beschreven conclusies over de haalbaarheid worden de cijfermatige uitkomst die in deze fase zullen worden opgeleverd nadrukkelijk gezien als een *voorlopig resultaat*. Zo is het niet onwaarschijnlijk dat er maart nog geen volledig zuiver beeld is van de doelpopulatie. Bovendien is de verwachting dat er ook nog verfijning nodig zal zijn van de eerste kenmerken die kunnen worden berekend, bijvoorbeeld omdat er nog verdere stappen kunnen/moeten worden gezet om de kwaliteit/nauwkeurigheid te verbeteren, in de volgende fase.¹⁴ In maart zal dit beeld scherper worden neergezet, in de vorm van een plan van aanpak voor het vervolg van de nulmeting.

In de 'start van de nulmeting' zal voor de volgende kenmerken en categorieën een eerste prototype worden gecategoriseerd en opgenomen in de dataset. Het gaat hier in het bijzonder om de minder complexe kenmerken die grotendeels via 'Reguliere expressies', in combinatie met controle door annoteren, kunnen worden geclassificeerd:

- Jaar uitspraak
- Welke rechtbank
- Soort kamer (enkelvoudig/meervoudig)
- Aantal eisers
- Wel/geen vertegenwoordiging eiser
- Soort verweerder (college van B&W – overig)
- Word count (als eerste proxy voor complexiteit)

Een belangrijk punt is of de data van de RvS in deze fase al kunnen/zullen worden ingebracht. Voordeel van dat wel doen is dat dan eerder in het onderzoek een compleet beeld wordt verkregen van de escalaties in het omgevingsrecht en ook van de eventuele uitdagingen die nog specifiek gelden voor de documenten van de RvS. Het afstemmen, organiseren en verwerken van een datalevering van de RvS was niet voorzien en zal daarom worden gezien als meerwerk.

Ook is relevant om in deze fase meer inzicht te krijgen in de andere factoren die een rol spelen in de onvolledigheid, namelijk het missen van relevante (inmiddels verouderde)

¹³ Eventueel in te schuiven, afhankelijk van de benodigde doorlooptijd van het inspoelen van data van de Raad van State.

¹⁴ Bijvoorbeeld door meer uitspraken te annoteren.

zoektermen in de bij de extractie uit het E-archief gebruikte zoektermenlijst, maar vooral het feit dat niet alle uitspraken bij de RvdR zijn gedigitaliseerd. In deze exercitie zal de RvdR, en mogelijk ook de RvS, een belangrijke rol kunnen spelen.

Belangrijk, ten slotte, in deze fase is om te voorkomen dat voortijdige en onzuivere conclusies zullen worden getrokken uit de voorlopige dataset. Er zal daarom ook een nadrukkelijke disclaimer worden geplaatst bij de status van de dataset aan het einde van deze fase.

Vervolg nulmeting – naar een prototype nulmeting

Na de start van de nulmeting verwachten we dat het volgende nodig zal zijn om te komen tot een volledig dekkende prototype nulmeting:

- Een compleet overzicht van de gewenste kenmerken en categorieën, zoals bepaald door de externe begeleidingsgroep
- Het ontwikkelen van een methode voor het classificeren van de kenmerken die op dit moment nog niet vastgesteld zijn
- Waar nodig, een verfijning van de methode voor het classificeren van de kenmerken die op dit moment al wel vastgesteld zijn
- Methodologische ondersteuning (vanuit CBS) bij bovengenoemde
- (vermoedelijk) juridische input voor het meer juridisch-complexe annoteerwerk
- (mogelijk) verfijnde dataleveringen van de RvdR en de RvS voor het onderzoek

In maart 2020 zal er een plan van aanpak gemaakt worden, waarin bovengenoemde elementen zijn meegenomen, met daarbij ook een overzicht van de benodigde middelen, per onderdeel. Samen met BZK-ICTU zal vervolgens moeten worden gezocht naar een constructie voor een formele opdrachtverstrekking en governance voor het vervolg van de nulmeting. Uitgaande van drie maanden doorlooptijd hiervoor, zou het vervolg van de nulmeting in/rond de zomer kunnen starten en kunnen worden uitgevoerd in de tweede helft van 2020, met uitloop in 2021.

Bestendiging methodiek nulmeting – naar een toekomstbestendige monitor

Een “traditionele statistiek” kan worden bestendigd door een robuust statistisch proces op te zetten, met vaste algoritmen, en bijbehorende standaard controles, om van brondata tot statistische output te komen. Voor deze innovatieve monitor ligt dat wat ingewikkelder, om de volgende redenen.

In de eerste plaats wordt gebruik gemaakt van ongestructureerde data en niet-traditionele statistische methodes (zoals tekst mining en machine learning). De kwaliteit van de output van dit soort processen is niet alleen afhankelijk van de nauwkeurigheid van de broninformatie, maar ook voor een belangrijk deel van wat er ‘handmatig’ wordt ingestopt in de vorm van annotaties. Algemeen gezegd: hoe meer je erin stopt, hoe beter de kwaliteit is van wat eruit komt. Hierbij zou moeten worden gezocht naar een zeker optimum met een acceptabel/gewenst kwaliteitsniveau. Bij het CBS kunnen methodologisch experts zich mede gaan buigen over dit vraagstuk, met als doel om samen toe te werken naar statistische output met een acceptabel kwaliteitsniveau, inclusief kwantitatief inzicht in de mate van nauwkeurigheid (marges). Zeker als in de toekomst beleidsmatige conclusies getrokken gaan worden uit de monitor en daarbij kruisverbanden tussen kenmerken worden gelegd (zie paragraaf 3.4), is het raadzaam om deze vervolgstap te gaan zetten.

lets anders wat deze monitor bijzonder maakt, is dat er naast cijfermatige input (data) ook inhoudelijk-juridische kennis wordt gebruikt met betrekking tot de duiding van de data. Deze inhoudelijk-juridische input speelt een belangrijke rol in de vertaling van informatie uit teksten naar de gewenste classificatie van kenmerken. Omdat de inhoudelijk-juridische werkelijkheid geen constant gegeven is, zal goed moeten worden nagedacht over de wijze waarop dit soort kennis kan worden geborgd in een toekomstbestendige monitor. En dan gaat het niet alleen over de nulmeting, maar ook over de vergelijkbaarheid van de uitkomsten met die van een vervolgmeting, ná de invoering van de Omgevingswet. Na de afronding van de nulmeting, mits die er komt, zal met alle betrokken partijen een plan van aanpak worden bedacht voor de borging van deze kennis-input. Dit zou dan kunnen worden gecombineerd met een opstart richting een vervolgmeting, in 2022.

BIJLAGE I Haalbaarheidsbevindingen per kenmerk

KENMERKEN	Hoe goed onderzocht?	Hoe goed uit data te halen met voorgest. methode?	Advies nodig inh./jur. experts?	Hoeveel inspann. dit uit data te halen?
EISER(S)				
Soort eiser(s)				
<ul style="list-style-type: none"> • Particulier 	Nu relatief simpel model gebruikt. Nog bezig met model voor Named Entity Recognition. dit lijkt aardig te werken voor onderscheid part./bedr./belangenver.			
<ul style="list-style-type: none"> • Bedrijf • Stichting/vereniging <ul style="list-style-type: none"> • Sticht./ver. (regionaal) • Sticht./ver. (landelijk) 	nog geen lijst beschikbaar	indien lijst beschikbaar	voor opstellen landelijke lijst	
<ul style="list-style-type: none"> • Overheid 		onzeker, nog te weinig annotaties om te trainen		
<ul style="list-style-type: none"> • Verschillend 		zorgt voor fouten andere classes		
<ul style="list-style-type: none"> • Overig 		(idem)		
Aantal eisers				
<ul style="list-style-type: none"> • Eén • Meerdere 				
Vertegenwoordiging				
<ul style="list-style-type: none"> • Zonder gemachtigde • Met gemacht.(n), zonder mr. • Met gemacht.(n), met mr. 	Nog niet naar vertegenw. in procesverloop gekeken; accuracy zal verbeteren			
VERWEERDER(S)				
Soort verweerder(s)				
<ul style="list-style-type: none"> • College van B&W 		onzeker, nog te weinig annotaties om te trainen		
<ul style="list-style-type: none"> • Gemeenteraad • Heffingsambtenaar • Dag. bestuur gemeente • Dag. bestuur waterschap • Alg. bestuur gemeente • Alg. bestuur waterschap • Minister(ie) van ... • College van gedeputeerde staten • ZBO 		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		
		(idem)		

KENMERKEN	Hoe goed onderzocht?	Hoe goed uit data te halen met voorgest. methode?	Advies nodig inh./jur. experts?	Hoeveel inspann. dit uit data te halen?
<ul style="list-style-type: none"> Vershillend 				
Vertegenwoordiging <ul style="list-style-type: none"> Met gemacht.(n), zonder mr. Met gemacht.(n), met mr. Zonder gemachtigde 	Nog niet naar vertegenw. in procesverloop gekeken; accuracy zal verbeteren			
BELANGHEBBENDE(N)/DERDEN				
Sprake van belanghebbende(n)/derde partij(en) <ul style="list-style-type: none"> Geen derde(n) Wel derde(n) 				
Vertegenwoordiging <ul style="list-style-type: none"> Met gemacht.(n), zonder mr. Met gemacht.(n), met mr. Zonder gemachtigde 	Nog niet naar vertegenw. in procesverloop gekeken; accuracy zal verbeteren			
BESTREDEN BESLUIT				
Soort besluit <ul style="list-style-type: none"> Vergunning, vrijstelling, ontheffing, projectbesluit (Besluit/verzoek tot) handhaving(sbeschikking) Planschadebeschikking/-besluit Bestemmingsplan, besluit van alg. strekking (excl. tracé) Tracébesluit Overig 				
Procedure <ul style="list-style-type: none"> Regulier Uitgebreid 	<p>is het voldoende precies om te zoeken naar 'bezwaar' in procesverloop?</p> <p>idem naar ontbreken 'bezwaar'?</p>			
Aard besluit <ul style="list-style-type: none"> Verleend/toegekend/opgelegd Geweigerd Afgewezen/Ingetrokken Opgelegd Overig 	<p>controle huidige categorieën</p>			

KENMERKEN	Hoe goed onderzocht?	Hoe goed uit data te halen met voorgest. methode?	Advies nodig inh./jur. experts?	Hoeveel inspann. dit uit data te halen?
TWISTPUNT(EN) [nog nader vaststellen]				
Materieel recht versus procesrecht				
• Materieel				
• Proces				
• Beiden				
(Hoger) beroep heeft betrekking op in te vullen open normen				
• Nee				
• Ja				
• Deels				
Wettelijke normen versus rechtsprincipes				
• ...				
• ...				
• Beiden				
Aantal beroepsgronden				
• ...				
• ...				
DOORLOPEN VOORTRAJECT [nog nader vaststellen]				
<i>Welke eerdere instanties hebben zich over de zaak gebogen?</i>				
COMPLEXITEIT UITSPRAAK [nog nader vaststellen]				
Enkelv. of meerv. kamer				
• Enkelvoudig				
• Meervoudig				
• Voorzieningenrechter				
Aantal rechtsoverwegingen				
• ...				
• ...				
Activiteit(en) waarvoor verg. werd aangevraagd (bij verg.)				
• n.v.t.				
• ...				

KENMERKEN	Hoe goed onderzocht?	Hoe goed uit data te halen met voorgest. methode?	Advies nodig inh./jur. experts?	Hoeveel inspann. dit uit data te halen?
<ul style="list-style-type: none"> • ... (zie lijst van ca. 24 activiteiten op omgevingloket.nl) 				
Lengte uitspraak <ul style="list-style-type: none"> • ... woorden 				
Aantal eisers (i.p.v. 1 of meer) <ul style="list-style-type: none"> • ... eiser(s) 				
Doorlooptijd oorspronkelijk besluit en uitspraak <ul style="list-style-type: none"> • ... dagen 				
UITKOMST GEDING [nog nader vaststellen]				
Gegronnd (tenm. op 1 punt) <ul style="list-style-type: none"> • Nee • Ja 				
Ongegr. (tenm. op 1 punt) <ul style="list-style-type: none"> • Nee • Ja 				
Niet ontv. (tenm. op 1 punt) <ul style="list-style-type: none"> • Nee • Ja 				
Overig (bijv. onbevoegd) (tenm. op 1 punt) <ul style="list-style-type: none"> • Nee • Ja 				

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Inlichtingen

Vragen over deze publicatie kunnen gestuurd worden aan CBS-CvB onder vermelding van het referentienummer 305380. Ons e-mailadres is maatwerk@cbs.nl.

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2020.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.