Wrakingsprocedures

Een iteratieve rechtspraakanalyse

Raimond Dufour

17-1-2021

### 1. Inleiding

Kwantitatieve analyse van uitspraken betreft het “tellen” van en binnen jurisprudentie. Het biedt een interessante mix tussen technologische en juridische onderwerpen. Met name de tekstuele analyse, ofwel het *natural language processing*, is de afgelopen jaren behoorlijk ontwikkeld, ook voor de Nederlandse taal, zodat de tech-savvy jurist zijn hart kan ophalen.

Daar komt bij dat het aantal gepubliceerde (Nederlandse) uitspraken in de afgelopen jaren fors is gegroeid, en dan doel ik op de database van rechtspraak.nl . Helaas worden evenwel nog lang niet alle uitspraken gepubliceerd. Sommige schrijvers menen dat daarmee de representativiteit ontbreekt zodat elk kwantitatief onderzoek zinloos is, maar ik meen dat onderzoek wel degelijk zinvol kan zijn waar het gaat om hun onderlinge variabelen[[1]](#footnote-1).

Om de kloof tussen technici en juristen wat kleiner te maken zal ik hierna een onderzoek beschrijven dat ik heb uitgevoerd en ook de uitkomsten daarvan bespreken. Ik beoog hiermee bij te dragen aan de verbetering van de verschillende onderzoeksmethoden. Daarvoor is interactie tussen juristen en technici onmisbaar, dus ik sta zeer open voor kritiek en suggesties ter verbetering van de door mij beschreven methoden.

#### Onderwerp van onderzoek: de wrakingsprocedure

Dit artikel gaat over het onderzoek naar de wrakingsprocedure. Ik heb hiervoor gekozen omdat het [publicatieprotocol](https://www.rechtspraak.nl/Uitspraken/Paginas/Selectiecriteria.aspx) voorschrijft dat alle wrakingsbeslissingen worden gepubliceerd. De wrakingsprocedure is bovendien een eenvormige procedure zonder veel complicaties, betrekkelijk weinig rechtsontwikkeling en met veel subjectieve toetsen, terwijl de wettelijke kaders voor de drie rechtsgebieden bestuursrecht, strafrecht en civiel recht min of meer woordelijk gelijk zijn. Het aantal wrakingsuitspraken is bovendien voldoende voor analyse, terwijl het eveneens behapbaar is voor het bestek van dit onderzoek. Tot slot is de wrakingsprocedure een belangrijk onderdeel van de rechtspleging, zodat een analyse ervan kan bijdrage aan de praktijk van alledag.

Het doel van dit onderzoek is tweeledig: ik probeer inzicht te geven in de praktijk van wrakingsprocedures en een onderzoeksmethode te presenteren die zou kunnen worden toegepast ter vervanging van of in aanvulling op de bestaande praktijk van handmatige codering van uitspraken.

#### De cijfers van rechtspraak.nl

Ook rechtspraak.nl heeft een [aparte themapagina](https://www.rechtspraak.nl/Themas/Wraking) gewijd aan wraking. En in het [jaarverslag](https://jaarverslagrechtspraak.nl/wp-content/uploads/sites/2/2021/04/Jaarverslag-Rechtspraak-2020.pdf#page=66) van de rechtspraak komt de wrakingsprocedure ook terug.

In de [kengetallen 2019](https://www.rechtspraak.nl/SiteCollectionDocuments/kengetallen-2019.pdf) van rechtspraak.nl staan overigens de cijfers van ingediende wrakingsverzoeken bij de rechtbanken:

Jaar AantalVerzoeken RechtspraakGehonoreerd Dataset DatasetDeelsToegewezen  
1 2016 516 17 381 16  
2 2017 480 10 339 11  
3 2018 563 15 426 17  
4 2019 559 16 373 13

De uitspraken-dataset van rechtspraak.nl lijkt een andere te zijn dan waar de kengetallen van de rechtspraak op zijn gebaseerd. Dat zou betekenen dat niet alle wrakingsprocedures worden gepubliceerd. We kunnen dus concluderen dat veel wrakingsuitspraken niet worden gepubliceerd. Dat ligt niet aan een vertraging van publicatie van wrakingsuitspraken: Gemiddeld duurt het ongeveer 60 dagen voordat een wrakingszaak wordt gepubliceerd.

#### Iteratief onderzoek

Het onderzoek dat ik heb uitgevoerd is iteratief: Na enkele experimentele queries en korte data-analyses heb ik gekozen voor het onderwerp wraking, de (voorlopige) dataset bepaald, diverse uitspraken bestudeerd, trefwoorden genoteerd en classificaties bedacht. Hoe meer uitspraken je bekijkt, hoe meer (tekst)patronen opvallen. Daar heb ik vervolgens mijn onderzoek weer op aangepast. Gewoon beginnen is het devies, het onderzoek gedurende de rit aanpassen, en weer doorgaan met de nieuwe inzichten.

Dit document is opgesteld met R-markdown: hier is de broncode beschikbaar waarmee in RStudio data kunnen worden opgevraagd en bewerkt.

#### Dataset

De door mij gehanteerde dataset is, zoals gezegd, afkomstig van de totale uitspraken-dataset van rechtspraak.nl. Deze totale dataset wordt maandelijks bijgewerkt en aangevuld. Enkele jaren geleden heb ik de dataset gedownload en periodiek een update doorgevoerd waarbij ik rekening heb gehouden met wijzigingen als de wijzigingsdata in de door rechtspraak.nl meegeleverde metadata daartoe aanleiding gaven.

De dataset van rechtspraak.nl behelst ca. 600.000 uitspraken en deze uitspraken zijn voorzien van metadata, zoals een uniek ecli-nummer, de uitspraakdatum, een inhoudsindicatie en nog veel meer. Voor de eerste selectie heb ik als criterium gehanteerd dat de tekenreeks “wraking” voorkomt in de inhoudsindicatie en/of de uitspraak zelf, waarbij geen onderscheid wordt gemaakt tussen hoofdletters en kleine letters, en ook de omringende tekens niet relevant zijn (als in de inhoudsomschrijving slechts het woord “wrakingsprocedure” voorkomt, dan wordt ook die uitspraak geselecteerd).

Het tijdvak dat ik heb gehanteerd betreft de vijf jaren 2016 tot en met 2020. Tuchtrechtuitspraken heb ik niet geselecteerd, omdat deze niet vallen onder het publicatieprotocol van rechtspraak.nl. Ook conclusies van de procureur-generaal in wrakingszaken heb ik buiten de dataset gelaten. Als variabelen heb ik uitsluitend het ecli-nummer, de inhoudsindicatie en de uitspraaktekst opgevraagd.

Met de selectie op basis van deze criteria resteert een dataset van 2914 uitspraken. Enkele uitspraken zijn niet interessant, omdat die geen tekst bevatten. De uitspraakteksten in de dataset van rechtspraak.nl wordt aangeleverd inclusief html-codes, dat zijn kort gezegd de layout-codes van een uitspraak. Die codes staan doorgaans tussen vishaken (“<” en “>”) en zijn dus ook gemakkelijk te verwijderen. Die heb ik uit de dataset verwijderd. Om een idee te krijgen van de dataset gaat hierbij een voorbeeld (de tekstvelden zijn afgebroken i.v.m. de lengte):

# A tibble: 5 x 4  
 ecli wrakingszaak inhoudsindicatie uitspraak   
 <chr> <fct> <chr> <chr>   
1 ECLI:NL:RBROT:2016:6188 wrakingszaak Wrakingsverzoek afgewez~ "Rechtbank Rott~  
2 ECLI:NL:RBROT:2017:7036 wrakingszaak Wrakingsverzoek afgewez~ "Rechtbank Rott~  
3 ECLI:NL:RBROT:2016:5883 wrakingszaak Wrakingsverzoek afgewez~ "Rechtbank Rott~  
4 ECLI:NL:RBROT:2016:5910 wrakingszaak Wrakingsverzoek afgewez~ "Rechtbank Rott~  
5 ECLI:NL:RBROT:2016:5912 wrakingszaak Wrakingsverzoek afgewez~ "Rechtbank Rott~

De lijst met eclinummers is hier te downloaden als tekstbestand.

#### Foutmarge

Bij rechtspraakanalyse is een foutmarge onvermijdelijk. Fouten ontstaan bij de invoer door rechtspraak.nl, bij het aanleveren/ontvangen van de database van rechtspraak.nl, bij het verwerken ervan in een eigen database, bij de selectie van subsets, bij het onderzoek en bij het interpreteren van het onderzoek. De foutmarge kan problematisch zijn waar je hem niet ziet: ontbrekende data en ontbrekende resultaten. Maar zeker met machinelearning-algoritmen heb je nu eenmaal te maken met foutmarges. Dat is geen probleem als die foutmarges binnen acceptabele grenzen blijven en daar ligt dan ook de focus bij het ontwikkelen van deze algoritmen.

Voor wat betreft dit onderzoek bestaat de mogelijkheid dat er in de dataset, ondanks het publicatieprotocol, uitspraken ontbreken over wrakingsprocedures. Dat maakt de database dan niet meer volledig, maar nog wel representatief. Of en hoeveel uitspraken ontbreken, valt niet goed in te schatten. Ook is mogelijk dat er wrakingsprocedures zijn geweest waarin de tekenreeks “wraking” niet in de uitspraak staat vermeld. De kans hierop is zeer klein, zo zullen wij hierna zien. Overigens worden nog maandelijks wrakingsuitspraken gepubliceerd door rechtspraak.nl, dus dat er enige uitspraken ontbreken in de dataset is niet onwaarschijnlijk. Het gaat echter om te lage aantallen om significante invloed te hebben op de uitkomsten van het onderzoek. Dat maakt de door mij gestelde foutmarges (tot twee cijfers achter de komma) heel betrekkelijk.

Wat de door mij gestelde foutmarges ook betrekkelijk maakt, is het feit dat algoritmen hun berekeningen vaak baseren op zeer veel willekeurig gekozen selecties van data. De ene selectie is de andere niet en de ene uitkomst zal de andere daarom niet zijn. De kunst is om de foutmarge bij algoritmen zoveel mogelijk te beperken, maar een zekere variatie is hier onvermijdelijk.

## 2. Wrakingsprocedure of niet?

Van de 2914 gevonden uitspraken zal bepaald moeten worden of het hier daadwerkelijk gaat om een wrakingsprocedure. De tekenreeks “wraking” kan immers ook in een ander verband voorkomen. Er zijn verschillende methoden om te bepalen of het hier gaat om een wrakingsprocedure of niet. Ik bespreek de volgende:

1. Handmatige selectie;
2. Selectie op basis van reguliere expressies;
3. Selectie op basis van een machine-learning algoritme.

#### A. Handmatige selectie

Met handmatige selectie kan op misschien wel de minst-foutloze wijze worden vastgesteld of sprake is van een wrakingsprocedure of niet. Het nadeel van deze methode is vanzelfsprekend de arbeidsintensiviteit. Het kost veel tijd, moeite en concentratie om de selectie handmatig te maken. Maar ook bij handmatige selectie zijn hulpmiddelen te gebruiken die het werk sneller en gemakkelijker maken.

*Gerechtelijke grammatica*: De meeste uitspraken hebben een uniforme opbouw:

1. De aanhef (partijnamen, advocaten/gemachtigden, e.d.);
2. De expositie, d.w.z. de passages in de uitspraak die het procesverloop beschrijven, de geschilpunten, de rechtsvraag, etc.;
3. De overwegingen, waarin de rechter aan het woord is, oordeelt over de feiten en over het recht, en middels motiveringen komt tot zijn
4. Beslissing, ofwel het dictum.
5. De meeste uitspraken eindigen met de legalisatie, waarin verantwoord is van wanneer de uitspraak (in het openbaar) is uitgesproken en welke rechters (en griffier) de uitspraak hebben gewezen.;
6. Sommige uitspraken bevatten nog één of meerdere bijlagen.

Deze verschillende onderdelen noem ik gerechtelijke grammatica. Het is in de meeste gevallen mogelijk door middel van regexen te bepalen welke passages tot welke onderdelen behoort. Per uitspraak kan dan een kolom voor elk onderdeel van de uitspraak worden gemaakt. Het is vervolgens eenvoudig om van alle wrakingsuitspraken alleen de beslissing op te vragen.

*QDA-software*: Er zijn verschillende qda-programma’s waarmee uitspraken “gecodeerd” kunnen worden, dat wil zeggen op een slimme manier van labels kunnen worden voorzien en gecategoriseerd kunnen worden. Voor korte teksten heb ik in excel een eenvoudig [qda-programma](bvo.dds.nl/QDA.xlsm) gebouwd waarmee snel maar doeltreffend teksten kunnen worden gecategoriseerd . Met de beslissing van elke uitspraak kan snel worden bepaald of het hier gaat om een wrakingsprocedure en zo ja, wat de uitkomst daarvan is.

De uitkomst van het handmatige onderzoek naar de vraag welke uitspraken een wrakingsprocedure betreft is [hier](github.com/) te vinden. In het navolgende zal ik de handmatige uitkomsten gebruiken als *benchmark* voor de overige twee methoden, waar ik niet elke uitspraak zelf heb bestudeerd, maar dit heb overgelaten aan de computer.

De dataset ziet er na de handmatige codering als volgt uit:

ecli wrakingszaak inhoudsindicatie uitspraak   
 Length:2914 geen wrakingszaak: 583 Length:2914 Length:2914   
 Class :character wrakingszaak :2331 Class :character Class :character

#### B. Selectie op basis van reguliere expressies;

Het gebruik van *regular expressions* (ook wel regex genoemd) betreft het zoeken op tekstpatronen, en dus niet slechts op een specifieke tekenreeks. Er bestaat een min of meer uniforme syntaxis van voor verschillende programmeertalen. Het is zeg maar een meer geavanceerde manier van zoeken met wildcards (zoals \*) . De regex “verzoek(st|)er” geeft uitspraken die zowel verzoeker als verzoekster bevatten.

Het is mogelijk om een lijst te maken van reguliere expressies en per uitspraak die regexen te tellen. Het gaat dan om woorden of tekstpatronen die veel voorkomen in wrakingsprocedures en niet of veel minder in andere soorten procedures. Voor de hand ligt om woorden/tekstpatronen te gebruiken die ook in de wetteksten met betrekking tot wraking worden gehanteerd, maar dat hoeft niet per se. Als hulpmiddel om je woorden te kiezen kun je ook een lijst van meest gebruikte woorden generen (zie hier voor deze dataset), en daaruit de meest onderscheidende woorden selecteren. Hoe meer woorden je selecteert, hoe beter je resultaten zullen zijn.

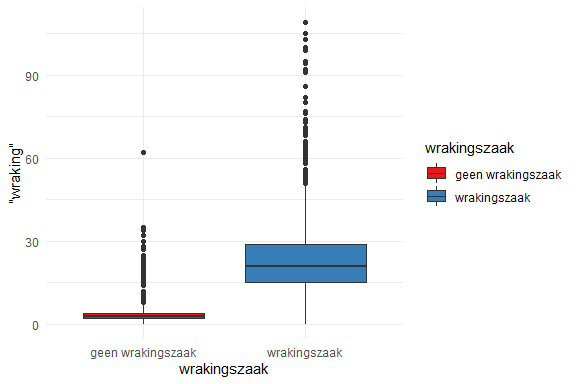
Het is belangrijk om te onderkennen dat regex-methode al snel een foutmarge impliceert. Het is de kunst die foutmarge te verkleinen door meerdere en “betere” regexen te gebruiken. Het vergt wat oefening en creativiteit maar is ook weer niet enorm moeilijk. Het aardige is: hoe meer uitspraken je bekijkt, hoe beter je kunt inschatten welke woorden relevant zijn voor het onderscheid. Een ander nadeel is dat spelfouten niet of verkeerd worden gesignaleerd. Ondanks deze bezwaren kan het gebruik van regexen zeer dienstbaar zijn voor het *pre-processen* van data, zoals wij hierna zullen zien.

Hoe meer regexen in de uitspraak voorkomen, hoe groter de kans is dat het gaat om een wrakingsprocedure en op deze manier kan ook de uitkomst van een wrakingsprocedure worden “ingeschat”. Ik heb zes regexen gebruikt: “wraking”, “wrakingskamer”, “wrakingsverzoek”, “partijdig”, “rechter” en “verzoek(st|)er”. Ik realiseer mij daarbij dat er een overlap bestaat van de eerste regex met de tweede resp derde regex, want de tekenreeks “wraking” komt ook voor in de tweede en derde tekenreeks.

Als laatste kolom heb ik de scores van de zes regexen bij elkaar opgeteld.

De eenvoudigste methode om de uitkomst te bepalen is als volgt. De frequentie per regex is een score die je ook ten opzichte van elkaar kunt wegen. Hoe hoger de overall-score, des te hoger de kans dat het een wrakingsprocedure betreft. Steeksproefsgewijs zou je uitspraken kunnen bekijken en onderzoeken bij welke score het omslagpunt voor de kans of de uitspraak een wrakingsprocedure betreft of niet.

Voor de tekenreeks “wraking” is het verschil in incidentie binnen de dataset als volgt weer te geven:



Uit de handmatig gecodeerde set volgt dat ongeveer 20% van de uitspraken waar de tekenreeks “wraking” in voorkomt niet een wrakingsprocedure betreft. Als we voor elke regex en de totaalscore de hoogste 80% score een punt geven en vervolgens per uitspraak de punten optellen, dan krijgen we een totaalscore met een maximum van 7. Als we er vanuit gaan dat dat ook geldt voor de andere regexen, dan krijgen de uitspraken met een score van 4 de kwalificatie “wrakingsprocedure”. Deze werkwijze is overigens betrekkelijk eenvoudig te *tunen*, met betere resultaten tot gevolg.

De resultaten van deze werkwijze zijn af te zetten tegen de resultaten uit het handmatige onderzoek in een *confusion matrix*: De rijen geven de voorspellingen weer, de kolommen de referentiewaarde (handmatige codering):

Reference  
Prediction geen wrakingszaak wrakingszaak  
 geen wrakingszaak 466 117  
 wrakingszaak 113 2218

Uit deze confusion matrix blijkt een foutmarge van 7.89%.

#### C. Selectie op basis van een machine-learning algoritme.

De derde methode is op basis van een door de computer te ontwikkelen algoritme. Een algoritme leert op basis van bekende uitkomsten. We zullen dus een dataset moeten creëren waarin de uitkomsten bekend zijn (de trainingset), op die dataset het algoritme bouwen en dat algoritme toepassen op de overige uitspraken (de testset). We pakken willekeurig 70% van de dataset die wij handmatig hebben bekeken en maken daarvan een testset, maar zien er wel op toe dat de verhouding in uitkomsten gelijk blijft. De data waarop wij (in eerste instantie) gaan trainen bestaat uit de frequenties van de regexen zoals wij die in de vorige paragraaf hebben geteld. Waar wij handmatig de frequenties van de verschillende regexen optelden, laten wij nu de computer kijken naar andere dan lineaire verbanden tussen de regexen onderling. De computer ziet, kort gezegd, verbanden tussen variabelen die mensen niet kunnen zien.

We hebben de dataset van de regexen en kennen de uitkomsten van het handmatige onderzoek. Die dataset is verdeeld in een trainingset (70%) en een testset (30%). Elke uitspraak met variabelen kan als een beslisboom leiden tot de uitkomst (wrakingszaak of niet). De computer selecteert willekeurig ruim honderdduizend maal enkele variabelen uit de trainingset en checkt de uitkomst (de random forest-methode). Zo worden de variabelen in meerdere dimensies met elkaar vergeleken en afgezet tegen de bekende uitkomst. Dit vereist rekenkracht: Mijn computer is hier met zes cores (zeg maar chips in de computer) ongeveer een minuut mee bezig geweest.

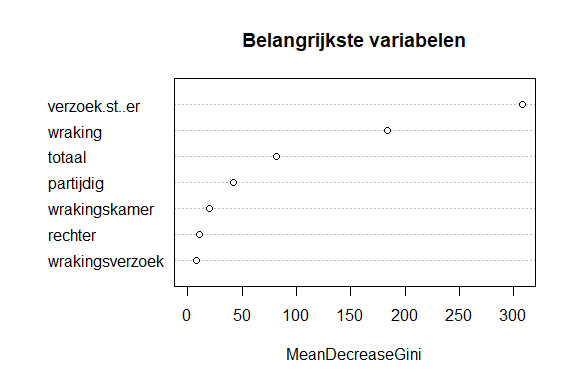
De beslisbomen (het algoritme) die zijn gegenereerd met behulp van de trainingset worden vervolgens toegepast op de testset. Het is belangrijk om te weten dat de testset dus niet betrokken is geweest bij het genereren van het algoritme, zodat de computer de testset uitsluitend benadert met de “kennis” van de trainingset.

Het algoritme toont een sterke verbetering ten opzichte van de Regex-methode:

Reference  
Prediction geen wrakingszaak wrakingszaak  
 geen wrakingszaak 171 9  
 wrakingszaak 3 690

De *accuracy* is hier 0.9863, oftewel 1.37% is foutief vastgesteld.

Welke regexen zijn nu het belangrijkst bij het algoritme? De computer maakt dat heel mooi inzichtelijk:



De meeste bepalende regex zijn “verzoek(st|)er” en “wraking”.

Dit resultaat kan nog worden verbeterd. Er zijn vele machine learningmethodes die allemaal kunnen worden *getuned*. Voor de hier gebruikte random forest-methode zou je bijvoorbeeld het aantal beslisbomen per uitspraak kunnen aanpassen, of het aantal berekeningen dat de computer uitvoert. Maar wij zouden ook meer regexen kunnen gebruiken, of andere variabelen toevoegen. Bijvoorbeeld de instantie, of de wraking speelt in een civiele, bestuursrechtelijke of strafrechtelijk procedure. Het moeten niet per sé juridische variabelen zijn, maar het is wel belangrijk om variabelen te gebruiken die enig begrijpelijk causaal verband kunnen hebben met de uitkomst, om toevallige correlatie te voorkomen.

## 3. Uitkomst van de wraking

Op dezelfde wijze gaan we kijken naar de uitkomst van de wraking. We gebruiken de uitspraken waarvan we handmatig hebben vastgesteld dat het gaat om een wrakingsprocedure. In onze dataset zijn dat 2331 uitspraken. Handmatig heb ik vastgesteld dat er 80 zaken (al dan niet deels) zijn toegewezen, 1428 zaken zijn afgewezen en 666 zaken niet-ontvankelijk zijn verklaard. Bovendien zijn er 109 zaken waar de rechter de wraking deels niet-ontvankelijk heeft verklaard en deels heeft afgewezen, bij de handmatige codering heb ik geconstateerd dat het veelal gaat om een niet-ontvankelijkverklaring, waarbij de rechtbank “voor het overige” de wraking afwijst. Die heb ik daarom bij die categorie gevoegd. Tot slot zijn er 9 wrakingsuitspraken waar geen wrakingsbeslissing is genomen, maar bijvoorbeeld een uitstel is gelast. In mijn onderzoek ben ik uitgegaan van drie categorieën:

* afwijzing,
* (gedeeltelijke) toewijzing en
* niet-ontvankelijk.

De uitspraken waar de rechtbank zowel niet-ontvankelijkheid aanneemt en voor het overige afwijst, heb ik geschaard onder de rubriek “niet-ontvankelijk”. Als wij weer gaan regexen, dan kijken we naar de meest voorkomende woorden in de uitspraken:

feature frequency docfreq  
1 verzoeker 33619 2024  
2 rechter 32704 2206  
3 verzoek 22873 2327  
4 mr 20389 2252  
5 beslissing 18143 2328  
6 wrakingskamer 17563 2222  
7 wraking 16647 2326  
8 zitting 15096 2068  
9 wrakingsverzoek 14828 2246  
10 rechtbank 12572 2081  
11 verzoekster 11360 664  
12 behandeling 11351 2214  
13 ter 10383 2053  
14 zaak 10110 2219  
15 omstandigheden 8536 2160  
16 oordeel 6895 1925  
17 vooringenomenheid 6856 1838  
18 grond 6756 2192  
19 2018 6475 825  
20 artikel 6382 2012

Deze woorden zijn weinig onderscheidend waar het gaat om de uitkomst van de procedure. Bijkomend probleem is dat de uitkomsten niet evenwichtig zijn verdeeld: slechts 3,5% van de uitspraken leiden tot een toewijzing. Wat zou een goede benadering kunnen zijn?

De database van rechtspraak.nl bevat zoals gezegd niet alleen de uitspraaktekst, maar ook metadata. Bij veel uitspraken is een inhoudsindicatie gegeven: een korte omschrijving van de zaak. Die inhoudsindicatie zie je ook als je een uitspraak zoekt op de website van rechtspraak.nl. Vaak wordt in die beschrijving al aangegeven wat de uitkomst is. Maar soms ontbreekt de inhoudsindicatie of is die te kort om daar informatie uit te krijgen. Dan biedt het dictum van de uitspraak uitkomst: het dictum bevat de beslissing van de rechter. Het dictum staat aan het eind van de uitspraaktekst en bevat doorgaans niet meer dan 500 tekens. Dus als wij in die gevallen de te doorzoeken tekst kleiner maken door de laatste 500 tekens van de uitspraak te nemen dan zoeken wij in een meer specifieke “corpus”. De meest voorkomende worden geven zo al een betere indicatie over de uitkomst van de procedure:

feature frequency docfreq  
1 rechter 1879 774  
2 verzoek 1876 1240  
3 wrakingsverzoek 1785 1314  
4 verzoeker 1697 856  
5 beslissing 1684 848  
6 wraking 1650 1301  
7 mr 1241 403  
8 wrakingskamer 1091 804  
9 afgewezen 863 820  
10 behandeling 760 591  
11 zitting 725 450  
12 griffier 692 408  
13 voorzitter 662 343  
14 zaak 633 453  
15 gegeven 605 569  
16 vooringenomenheid 580 485  
17 verzoekster 551 250  
18 niet-ontvankelijk 442 431  
19 rechtbank 435 321  
20 grond 430 335

Het probleem van de verdeling is op een slimme manier op te lossen: we kopieren de uitspraken die in de minderheid zijn zodanig, dat de uitkomsten in aantal min of meer gelijk zijn (*upsample*). We zouden ook de uitspraken die in de meerderheid kunnen verwijderen totdat de uitkomst-verhoudingen gelijk zijn (downsample), maar dan gooi je in dit geval wel erg veel data weg. We regexen de belangrijke woorden op inhoudsindicatie en uitspraak en bouwen daarmee een algoritme. Mijn computer doet daar ongeveer 3 minuten over. De uitkomsten zijn als volgt:

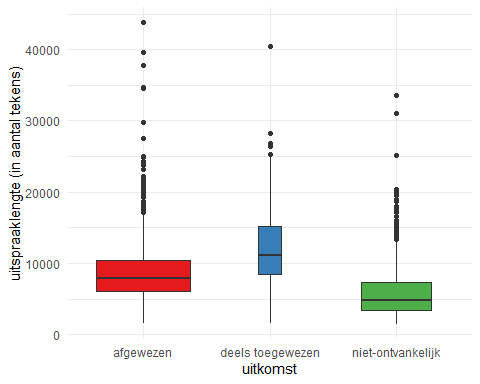
Reference  
Prediction afgewezen deels toegewezen geen wrakingsbeslissing  
 afgewezen 418 4 1  
 deels toegewezen 1 19 0  
 geen wrakingsbeslissing 0 1 0  
 niet-ontvankelijk 17 0 1  
 Reference  
Prediction niet-ontvankelijk  
 afgewezen 18  
 deels toegewezen 0  
 geen wrakingsbeslissing 0  
 niet-ontvankelijk 218

Hier zien we een foutmarge van 6.16%. De belangrijkste variabelen zijn:

Afbeelding met tafel

Automatisch gegenereerde beschrijving

Wat opvalt is dat de lengte van de uitspraak als belangrijke variabele wordt aangemerkt voor het bepalen wat de uitkomst van de uitspraak is. Hebben wij hier te maken met causaliteit? Je zou denken: zonder causaliteit geen voorspellende waarde. Dit is een steeds terugkerende vraag: zijn de gebruikte variabelen geschikt om als voorspeller te fungeren?



De lengte van een toewijzende uitspraak is gemiddeld langer dan een afwijzende uitspraak of een niet-ontvankelijkverklaring. Dat valt te verklaren doordat ook een wrakingsrechter weet dat een wraking zeldzamer is en, gelet op de gevolgen, geneigd zal zijn om meer uitleg te geven. Bovendien toetst de rechter doorgaans eerst of het wrakingsverzoek ontvankelijk is. Als het verzoek niet ontvankelijk is, kan inhoudelijke bespreking achterwege blijven, met een kortere uitspraak tot gevolg. Hetzelfde zal gelden voor de lengte van de inhoudsindicatie.

Wie op zoek is naar een nog beter resultaat, die kan de computer vragen om de uitkomst van de testset te specificeren naar waarschijnlijkheid van de uitkomst. Je krijgt dan per uitspraak vier kolommen: afgewezen, toegewezen, niet-ontvankelijk en geen beslissing. Zo kun je zien bij welke uitspraken de computer meer twijfelde. Die zou je dan weer handmatig kunnen nagaan. Maar ook hier weer: het is goed om te weten dat het hier gaat om schattingen, het algoritme is niet foutloos. Een uitspraak die (volgens de computer) duidelijk een afgewezen wrakingszaak betreft, hoeft dat niet per sé te zijn. Het gaat om (de wet van) de grote getallen.

## 4. Wrakingsredenen

Na bestudering van een aantal uitspraken kwam ik tot de constatering dat er grofweg 3 redenen zijn om een rechter te wraken:

1. De **hoedanigheid** van de rechter; Bijvoorbeeld: Heeft de rechter een eigen (financieel) belang in deze zaak? Of is hij in privé gelieerd aan een betrokken partij?
2. Een onwelgevallige **beslissing**; Veel wrakingsverzoeken zien op een (doorgaans: procedurele) beslissing waaruit de vrees voor partijdigheid/vooringenomenheid wordt gesteld.
3. De bejegening door de rechter. Het **optreden** van een rechter (meestal ter zitting) kan aanleiding geven om de rechter te wraken.

Het verschil tussen deze drie redenen is overigens niet in alle gevallen heel duidelijk. En het komt regelmatig voor dat meerdere redenen voor de wraking worden opgegeven. Ook komt het voor dat de redenen van de verzoeker te onduidelijk zijn om te behandelen of dat de redenen niet duidelijk worden geformuleerd in de uitspraak. De handmatige codering van deze variabele is tijdsintensief en zeker niet foutloos. Toch valt er heel goed een analyse te maken.

Welke woorden zijn kenmerkend voor de verschillende redenen? Dat is niet zo moeilijk te bepalen: we maken een lijst van alle woorden in de uitspraken en kijken welke woorden het meest voorkomen. Vervolgens maken we een subset van de uitspraken per reden en kijken weer naar de woordfrequentie. Het verschil tussen de twee lijsten geeft de meest onderscheidende woorden per reden. Van die woorden kiezen wij de woorden die een te verwachten relatie met de reden hebben. Ik kom dan op de volgende woorden en combinaties:

* Hoedanigheid: “eerder betrokken”, “andere procedure”, “betrokken”, “nevenfunctie”, “lid van”, “werkzaam”
* Optreden: “opmerking”, “proces( |-)verbaal”, “kritisch”, ” woord”, “bejegen”, “vra(a|)g”, “comparitie”, “zitting”, “behandeling”, “wederhoor”
* Beslissing: “beslissing”, “getuige”, “aanhouding|uitstel”, “horen”, “motivering”, “procedurele|processuele”

### Bag of words

Als wij per ecli elk woord van de inhoudsindicatie in een aparte kolom tellen, dan krijgen wij een *bag of words*. Elk nieuw woord in de volgende inhoudsindicatie voegt weer een kolom toe. Uiteindelijk krijg je dan een tabel met per rij de inhoudsindicatie en per kolom alle woorden die in alle inhoudsindicaties stonden. In de cel staat het aantal keer dat het woord in die inhoudsindicatie is gevonden.

Het is duidelijk dat een beetje “corpus” leidt tot een hele grote tabel:

Dit is maar een kleine selectie uit de tabel, in werkelijkheid heeft deze tabel 2331 rijen (de uitspraken), en maar liefst 34999 kolommen. In totaal dus 81582669 cellen. En de meeste cellen, bijna 99% zijn leeg omdat er niet veel overlap is in de gehele woordenverzameling. Er zijn evenwel mogelijkheden om die omvang wat te beperken. We zijn immers alleen op zoek naar betekenisvolle woorden. Om te beginnen kunnen alle letters omgezet worden in kleine letters en stopwoorden zoals “de” en “van” uit de teksten worden gehaald. Ook woorden(tekenreeksen) met een cijfer of vreemd teken worden uit de selectie gehaald. Bovendien worden woorden die in totaal minder dan 3 keer voorkomen ook geschrapt uit de lijst.

De computer ziet natuurlijk geen onderscheid tussen vervoegingen van werkwoorden, meervouden van zelfstandig naamwoorden etcetera. Maar taalkundigen hebben een algoritme gebouwd waarmee van woorden de stam kan worden berekend, zodat alle vervoegingen daaruit gefilterd worden. De stam is dan vaak geen bestaand woord, maar je kunt vaak wel begrijpen wat het oorspronkelijke woord was.

Met bovengenoemde aanpassing wordt de zin:

[1] "Op de rechte stukken moest Verstappen het gas loslaten om op deze banden nog de finish te kunnen halen."

omgezet in:

[1] "recht stuk moest verstapp gas loslat band finish hal"

Niet echt een goed leesbare tekst, maar wel enigszins begrijpelijk en voor de computer heel goed te hanteren. En het voordeel is: de computer telt nu de woorden met dezelfde stam bij elkaar op, zonder de verbuigingen, meervouden e.d. over te slaan.

De tabel die de computer op deze manier oplevert is een tabel van 21217 kolommen en 2331 rijen. Alle woordfrequenties in deze tabel gaan we vervolgens wegen: We zetten de woorden af tegen het totaal aantal woorden in de betreffende uitspraak en eveneens tegen het totaal aantal keren dat het woord in alle documenten voorkomt. Daar komt per woord een score uit van 0 tot ca. 5, met veel cijfers achter de komma. Een hogere score betekent dat het woord vaker voorkomt in relatief weinig uitspraken. Een lage score betekent dat het woord weinig voorkomt in veel uitspraken.

De computer kan prima rekenen met deze getallen. Door te kijken naar de verschillen in woordgewicht per categorie reden kan de computer daar verbanden herkennen.

Omdat in een uitspraak meerdere redenen kunnen worden vermeld, heb ik per reden een algoritme gebouwd om de woorden te herkennen. De accuratesse van deze drie algoritmen komen ver in de 90%:

* Algoritme Hoedanigheid: 99,59%
* Algoritme Optreden: 96,45%
* Algoritme Beslissing: 93,67%

## 5. Bevindingen

In een draaitabel kunnen de uitkomsten als volgt worden weergegeven (maar let wel: hier zitten dubbeltellingen in omdat een uitspraak meerdere redenen kan hebben):

# A tibble: 3 x 7  
 uitkomst hoedanigheid `%hoedanigheid` optreden `%optreden` Beslissing  
 <fct> <int> <chr> <int> <chr> <int>  
1 afgewezen 95 79.83 % 402 86.64 % 675  
2 deels toegewezen 17 14.29 % 39 8.41 % 29  
3 niet-ontvankelijk 7 5.88 % 23 4.96 % 51  
# ... with 1 more variable: %beslissing <chr>

[1] "DIT IS NOG NIET AF: NOG VEEL UITSPRAKEN DIE NIET-ONTVANKELIJK ZIJN HEB IK NOG NIET OP HANDMATIG REDEN GECODEERD"

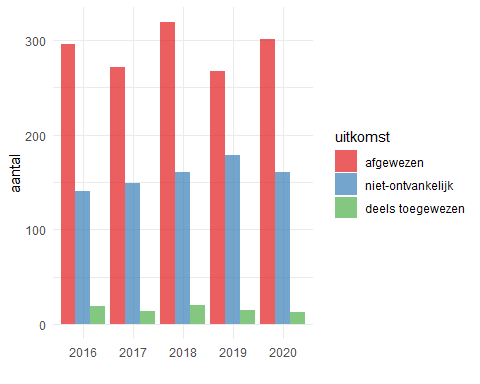
Wat het meest in het oog springt is dat de kans dat een wrakingsverzoek slaagt vanwege de hoedanigheid van een rechter aanzienlijk hoger is (ruim 3x) dan waar het wrakingsverzoek is gestoeld op een procedurele beslissing.

Er zijn echter meer bevindingen te doen. Alleen het ecli-nummer geeft al meer informatie over de uitspraak: de instantie en het uitspraakjaar. Die zijn dus eenvoudig af te zetten tegen de uitkomst en reden van de wrakingsprocedure. En naast het ecli-nummer, de uitspraaktekst en de inhoudsindicatie publiceert rechtspraak.nl nog meer meta-informatie, zoals rechtsgebied, relaties met andere eclis, publicatiedatum en de zittingsplaats. Maar er is nog veel meer te vinden. Want elke analyse biedt weer meer meta-informatie om toe te voegen aan de dataset waarmee de database verrijkt wordt en het onderzoek verbeterd.

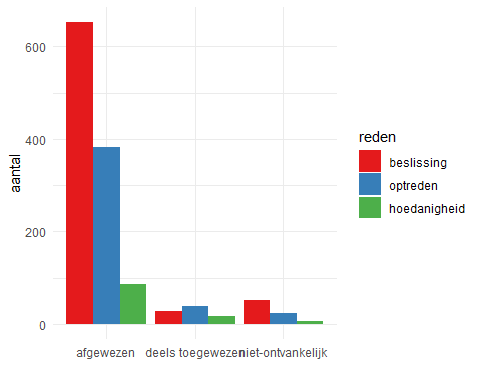
Als we de instanties onderscheiden dan komt het volgende beeld naar voren:

# A tibble: 3 x 7  
 uitkomst rechtbank gerechtshof `Raad van State` `CRVB/CBB` `Hoge Raad` overig  
 <fct> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 afgewezen 1193 155 58 39 7 2  
2 niet-ont~ 663 63 20 34 7 0  
3 deels to~ 66 13 0 1 0 1

Per jaar kunnen de wrakingsuitspraken als volgt worden weergegeven:

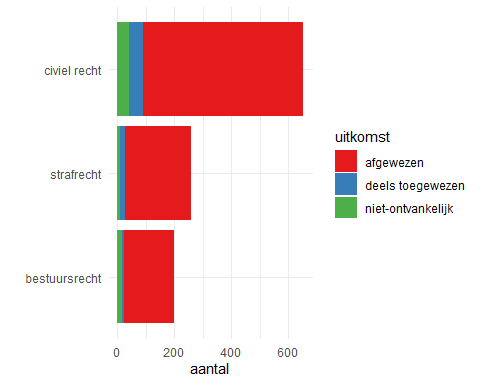


De redenen van wraking bij rechtbanken en gerechtshoven:



Het is duidelijk dat een (procedurele) beslissing de meest voorkomende wrakingsgrond is. Maar deze is ook wel weinig kansrijk. Keer op keer overweegt de wrakingsrechter dat een door de rechter genomen beslissing niet een te honoreren grond is voor wraking, tenzij de beslissing zodanig onbegrijpelijk is dat hierdoor de (schijn van) rechterlijke onpartijdigheid in het geding raakt.

En per (hoofd)rechtsgebied zijn de uitkomsten als volgt:



[1] "DIT IS NOG NIET AF: NOG VEEL UITSPRAKEN DIE NIET-ONTVANKELIJK ZIJN HEB IK NOG NIET OP HANDMATIG REDEN GECODEERD"

We moeten oppassen met het verbinden van conclusies aan deze verhouding: deze is natuurlijk ook gerelateerd aan het aantal procedures dat jaarlijks vatbaar is voor wraking, per rechtsgebied. Omdat lang niet alle uitspraken gepubliceerd worden, maar ook omdat lang niet iedere (voornamelijk civiele) procedures resulteren in een uitspraak, is deze informatie niet heel relevant. Wel is interessant dat het aantal toewijzingen binnen het bestuursrecht relatief hoger is dan binnen de twee andere rechtsgebieden.

## 6. Conclusie met betrekking tot de gehanteerd methoden

Het gebruiken van natural language processing is een uitstekend middel voor analyse van rechterlijke uitspraken. Feit blijft dat een beginset altijd handmatig zal moeten worden gecodeerd om de computer te kunnen leren. Hoe meer handmatig wordt ingevoerd, hoe kleiner de foutmarge van het algoritme zal zijn, maar hier is wel een optimum te vinden.

Er zijn verschillende manieren om een dataset te prepareren: je kunt bijvoorbeeld steekwoorden (liever: tekenreeksen) tellen, alle woorden of sets van woorden analyseren, of juist focussen op de samenhang van woorden. Daarnaast zijn er verschillende wijzen waarop de computer het algoritme bouwt. In dit artikel hebben wij een lineair model en de random forest-methode aan het werk gezien. Het is van tevoren niet goed te voorspellen welke methode de beste prestaties geeft.

Valkuilen zijn:

* De dataset van rechtspraak.nl is niet representatief omdat slechts een klein deel van de uitspraken is gepubliceerd;
* De selectie van uitspraken voor het uit te voeren onderzoek is niet correct en daarmee niet representatief voor het onderzoek;
* De keuze voor variabelen is arbitrair;
* De foutmarge van het ene algoritme werkt door naar het volgende algoritme;
* In productie: (rechts)ontwikkelingen zorgen voor een afwijkend idioom. Dat betekent dat een algoritme van tijd tot tijd geactualiseerd moet worden.

Wij zijn begonnen met het zoeken naar uitspraken in de gehele database van rechtspraak.nl op basis van de tekenreeks “wraking”. Op basis van verschillende reguliere expressies heeft de computer berekend welke uitspraak een wrakingszaak betrof en welke niet. Vervolgens hebben wij de uitkomst bepaald aan de hand van de laatste 500 tekens van de uitspraak en met behulp van een random forest-algoritme. Vervolgens hebben wij gekeken naar de reden van de wraking aan de hand van verschillende algoritmen. Daarmee hebben wij de hoogste score van de in dit artikel weten te vinden. Met de kennis van algoritmen die wij in dit artikel hebben opgedaan, wordt het tijd om de beste algoritmen weer toe te passen op alle coderingen.

1. Zie ook: Pieter Wolters: *Mogelijkheden en beperkingen bij de kwantitatieve analyse van jurisprudentie voor de bestudering van geldend recht* in: Methoden van systematische rechtspraakanalyse, Boomjuridisch 2021 [↑](#footnote-ref-1)