

# Methodische bijlage: inzet op maatschappelijke thema's

Het ministerie van Economische Zaken (EZ) en andere stakeholders vragen regelmatig om verschillende thema's in kaart te brengen en inzicht te geven in welke mate deze worden ondersteund binnen de EZ-instrumenten zoals de WBSO en de PPS-innovatieregeling. In zo'n thematische analyse wordt onderzocht hoeveel projecten betrekking hebben op één of meerdere specifieke thema's. In opdracht van het ministerie van EZ heeft RVO in beeld gebracht in hoeverre projecten die ondersteund zijn vanuit de innovatie-instrumenten bijdragen aan vier maatschappelijke thema's: 1) Energietransitie & Duurzaamheid; 2) Landbouw, Water & Voedsel; 3) Gezondheid & Zorg, en 4) Veiligheid. Deze analyse wordt jaarlijks gepubliceerd op de website [bedrijvenbeleidinbeeld.nl](http://bedrijvenbeleidinbeeld.nl), via deze pagina: [Versterken van innovatie in het bedrijfsleven | Bedrijvenbeleid in beeld](#). Dit document zet de methode uiteen hoe tot dit resultaat gekomen is.

**Let op:** dit document is een bijlage van de website [bedrijvenbeleidinbeeld.nl](http://bedrijvenbeleidinbeeld.nl) van het ministerie van EZ.

## Methodische achtergrond

### Scope

RVO heeft de analyse uitgevoerd over de volgende innovatieprogramma's en -regelingen van het ministerie van EZ:

- de Wet Bevordering Speur- en Ontwikkelingswerk (WBSO);
- de PPS-innovatieregeling (voorheen PPS-toeslag Onderzoek en Innovatie);
- Mkb-innovatiestimulering Regio en Topsectoren (MIT)<sup>1</sup>;
- Vroegefasefinanciering (VFF)<sup>2</sup>;
- SEED Capital;
- Small Business Innovation Research (SBIR), tegenwoordig Innovation Impact Challenge;
- Innovatiekrediet;
- Eureka;
- Eurostars;
- Globalstars;
- het Nationaal Groeifonds; en
- subsidies gericht op energie-innovatie binnen de Topsector Energie (EI) voor zover dit vanuit EZ-budget is gefinancierd.

### Toedeling

RVO heeft de gecommitteerde projecten binnen genoemde programma's en regelingen op basis van projectomschrijvingen<sup>3</sup> toebedeeld aan een of meer van de vier maatschappelijke thema's: 1) Energietransitie & Duurzaamheid; 2) Landbouw, Water & Voedsel; 3) Gezondheid & Zorg; en 4) Veiligheid. Met uitzondering van de WBSO zijn de

---

<sup>1</sup> Betreft de MIT R&D-samenwerkingsprojecten en Haalbaarheidsstudies. De MIT Kennisvouchers zijn niet gelabeld.

<sup>2</sup> Voor zover in uitvoering bij RVO.

<sup>3</sup> Met uitzondering van SEED Capital en (deels) SBIR, waar is gelabeld op respectievelijk fonds- en tenderniveau.

projecten handmatig geclassificeerd op basis van een uitgebreide omschrijving van de vier maatschappelijke thema's en de missies die binnen deze thema's vallen (zoals uiteengezet in de kamerbrief '[Missiegedreven topsectoren en innovatiebeleid](#)' van 26 april 2019).

Voor alle instrumenten zijn de uiteindelijke labels toegekend door RVO-adviseurs. Voor de PPS-Innovatieregeling (en voorheen de PPS-toeslagregeling) hebben de Topconsortia voor Kennis en Innovatie (TKI's) zelf een voorzet gedaan voor een categorisatie, en voor de MIT-regeling is dit door indieners gedaan. Om tot een consistente beoordeling te komen, is de beoordeling van de PPS-toeslag-inzetprojecten en MIT-projecten gecheckt en waar nodig aangepast door adviseurs van RVO. Voor alle instrumenten geldt dat eventuele twijfelgevallen binnen het team van projectbeoordelaars zijn besproken en onderling zijn afgestemd.

Voor de WBSO is de classificatie naar maatschappelijke thema's uitgevoerd door middel van een gewogen-trefwoordenmodel en sinds 2025 ook een deeplearning-model. Hoe dit precies werkt wordt hieronder bij de sectie 'Uitgelicht' toegelicht. Gezien het grote aantal jaarlijkse projecten is het niet haalbaar om met beperkte middelen alle WBSO-projecten te lezen en handmatig te classificeren.

## Bedragen

In de analyses wordt afwisselend gesproken over middelen, budgetten en committeringen. Hiermee wordt in principe altijd het gecommiteerde bedrag bedoeld voor de projecten, met de volgende toelichting of uitzonderingen:

- Bij de MIT betreft het enkel de landelijke committeringen, vanuit het EZ-budget, en niet de aanvullende budgetten van de regio's.
- Bij de PPS-toeslag betreft het de door de TKI's gerapporteerde ingezette PPS-toeslag.
- Bij de SBIR betreft het budgetten van SBIR-calls. Het betreft ook SBIR's waar het ministerie van EZ geen opdrachtgever van is. Het ministerie van EZ is wel verantwoordelijk voor het SBIR-instrument als geheel.
- Bij de SEED Capital-regeling betreft het een participatiebedrag op fondsniveau.

## Aanvullende analyses missiegedreven innovatiebeleid

Specifiek voor instrumenten binnen het missiegedreven innovatiebeleid, de PPS-Innovatieregeling, de MIT en de SBIR-regeling, zijn er aanvullende analyses beschikbaar op deze pagina: [Innovatie voor maatschappelijke vraagstukken | Bedrijvenbeleid in beeld](#). Voor deze instrumenten is er gelabeld op zes mogelijke Kennis- en Innovatie Agenda's (KIA's) en onderliggende Meerjarige Missiegedreven Innovatieprogramma's (MMIP's). De KIA's bestaan uit vier thematische KIA's die overeenkomen met de maatschappelijke thema's, en daarnaast de KIA Sleuteltechnologieën en de KIA Maatschappelijk Verdienvermogen. Deze laatste twee zijn dwarsdoorsnijdend en van belang voor het zowel realiseren van innovatie als het creëren van impact. Elke KIA heeft een aantal missies. Voor de toedeling naar missies is gebruik gemaakt van de toedeling naar de MMIP's. Deze zijn op dezelfde wijze door RVO gecontroleerd als de toedeling naar de KIA's.

## Uitgelicht: Methode WBSO

### Achtergrond

Via de WBSO stimuleert het ministerie van EZ ondernemers om meer te investeren in spur- en ontwikkelingswerk (S&O). Ondernemers die gebruik maken van de WBSO kunnen hiermee hun S&O-uitgaven verlagen. Jaarlijks worden door RVO ruim 80.000 WBSO-projecten (niet uniek) toegekend. Van deze projecten zijn alle omschrijvingen opgeslagen in een database van RVO<sup>4</sup>.

Gezien de omvang van het aantal WBSO-projecten is het niet haalbaar om met beperkte middelen alle WBSO-projecten te lezen en handmatig te classificeren. Hiervoor zijn verschillende modellen in gebruik. Vanaf 2025 worden ernaast gewogen-trefwoorden-modellen ook deep learning-modellen gebruikt om WBSO-projectteksten te classificeren naar de bovengenoemde maatschappelijke thema's. Tijdens de ontwikkeling van deze modellen werd door RVO geconstateerd dat ze beter presteerden dan de bestaande gewogen-trefwoorden-modellen, met uitzondering van het deep learning-model voor het thema Veiligheid<sup>5</sup>. In overleg met EZ is daarom besloten de methodiek voor dat thema niet aan te passen. Voor rapportages over de budgetjaren 2023 en verder wordt voortaan voor de thema's Energietransitie en Duurzaamheid; Landbouw, Water & Voedsel en Gezondheid & Zorg wel gebruik gemaakt van deep learning-modellen.

### Deep learning-model (thema's Energietransitie & Duurzaamheid; Landbouw, Water & Voedsel en Gezondheid & Zorg)

Voor het classificeren van WBSO-teksten naar de KIA's is eerst een bepaald tekstclassificatiemodel gekozen en ontwikkeld. Vervolgens is dit model getraind aan de hand van teksten en zijn de resultaten geëvalueerd. Ook is het model geoptimaliseerd, om te zorgen voor een efficiënte training en de beste resultaten. Hieronder wordt elke stap toegelicht.

#### Classificatiemodel

Voor de ontwikkeling van het tekstclassificatiemodel is gekozen voor een vooraf getrainde versie van het zogenaamde RoBERTa-model<sup>6</sup>. Dit model staat bekend om zijn sterke taalrepresentaties en is uitgebreid met een laag<sup>7</sup> die de tekstrepresentaties vertaalt naar concrete classificaties. Het model wordt verder getraind op de dataset met RVO-projecten, bestaande uit voornamelijk WBSO-projecten.

#### Training en evaluatie

De teksten worden voorbereid<sup>8</sup>, waarna de dataset wordt gesplitst in 90% trainingsdata en 10% testdata. Binnen elke trainingsronde wordt 10% van de trainingsdata gebruikt als validatieset, om prestaties tijdens het leerproces te monitoren en over fitting te voorkomen. Het model kan maximaal 1000 epochs (één volledige trainingscyclus) trainen. Echter, als er gedurende twintig opeenvolgende epochs geen significante verbetering meer optreedt, wordt de training automatisch beëindigd<sup>9</sup>.

---

<sup>4</sup> Deze informatie wordt gebruikt om de WBSO uit te voeren en valt onder de algemene wet inzake Rijksbelastingen (AWR).

<sup>5</sup> Meer technische informatie is op verzoek beschikbaar.

<sup>6</sup> Hugging Face Transformers

<sup>7</sup> De Softmax multi-class-classificatielaag

<sup>8</sup> Via RobertaTokenizer. De voorbereiding houdt in dat tekst wordt omgezet naar tokens waarmee gerekend kan worden door de modellen.

<sup>9</sup> Via een early-stopping-criterium.

## Optimalisatie

Om te zorgen voor een efficiënte training en om de kans te vergroten dat het model stabiel convergeert naar de beste prestaties, wordt het model geoptimaliseerd<sup>10</sup>.

## Gewogen-trefwoordenmodel WBSO (thema Veiligheid)

### Methode

Voor het thema Veiligheid wordt gebruik gemaakt van een zogenaamde 'gewogen-zoektermenmethodiek'. De gewogen-zoektermenmethodiek is door RVO ontwikkeld en kan worden gebruikt om WBSO-projecten te classificeren. Bij deze methode worden de projecten niet handmatig gelezen en geclassificeerd, maar wordt er in de projectomschrijvingen gezocht naar gerelateerde zoektermen, waarbij er een bepaalde gewing aan deze zoektermen wordt gegeven.

De methodiek bestaat uit vier primaire stappen die in figuur 1 worden weergegeven.



*Figuur 1 - Schematische weergave van de gewogen zoektermenmethodiek van RVO binnen de WBSO, een iteratief proces*

### Zoektermen & weging

Om thema's af te bakenen worden voorafgaand aan de analyse voor een thema een lijst met zoektermen opgesteld door adviseurs van RVO, met per zoekterm een wegingsfactor. Dit is een zeer belangrijke stap in het totale proces die in grote mate de kwaliteit van de uitkomst bepaalt. Deze stap wordt dan ook in samenwerking met een vakinhoudelijk-expert op het gebied van het gevraagde thema uitgevoerd, daardoor kan zoveel mogelijk naar de juiste (technische) vaktermen worden gezocht. Ieder trefwoord krijgt een gewing op basis van relevantie in relatie tot het gezochte thema. Zoektermen die kenmerkender zijn voor het thema dat wordt gezocht, krijgen een hogere waarde dan meer algemene woorden die wellicht raakvlak hebben met het gezochte thema, maar mogelijk ook in een andere context een andere betekenis kunnen hebben. In de lijst met zoektermen wordt rekening gehouden met linguïstiek (taal, spellingsvarianten en lemmatisering - het gebruik van woordstammen). Zoektermen die aangrenzende thema's kunnen identificeren die juist niet meegenomen mogen worden, kunnen een (sterk) negatieve gewing meekrijgen om bepaalde projecten uit te sluiten.

Ter illustratie: Bij het zoeken naar het thema Veiligheid wordt de zoekterm 'virus' gebruikt. Door te zoeken naar deze term zullen ook veel medische projecten worden

<sup>10</sup> Voor de optimalisatie wordt gebruik gemaakt van Bayesian Optimization (TPE) via de Hyperopt-library. Hiermee worden relevante parameters van het model geoptimaliseerd, waaronder batchgrootte, optimizer-keuze (Adam of Adadelta) en bijbehorende parameters (zoals beta\_1, beta\_2, rho). Ook algemene instellingen zoals leersnelheid, epsilon, decay en clipping worden geoptimaliseerd.

gevonden die niet geclassificeerd mogen worden tot het thema Veiligheid. Door zoektermen uit de medische technologie toe te voegen met een sterk negatieve weging, wordt er onderscheid gemaakt tussen het bestrijden van een 'medisch virus' en een 'ICT-virus'.

#### Doorzoeken & 'scoren'

Nadat de zoektermenlijst en wegingen zijn bepaald, worden met een query alle WBSO-projecten in de RVO-database doorzocht op deze termen. In de query kan worden aangegeven welke WBSO-jaren worden meegenomen en welke projectvelden worden doorzocht. Komt een zoekterm voor in één van de projectvelden, dan krijgt het project de weging mee. Komt dezelfde term in hetzelfde project nogmaals voor, dan telt de weging slechts één keer mee. Het script telt vervolgens alle wegingen per project op en rangschikt alle WBSO-projecten met de gevonden trefwoorden op projectscore van hoog naar laag.

#### Vaststellen grenswaarde voor selectie

Om de uitkomsten te valideren, wordt door middel van steekproeven bij verschillende projectcores de inhoud van de WBSO-projecten periodiek getoetst of de gevonden WBSO-projecten tot het gezochte thema behoren en welke niet. De validatie wordt uitgevoerd door een vakinhoudelijk expert. De steekproeven vinden daarbij plaats van een hoge projectscore naar een steeds lagere projectscore. Op deze manier wordt gezocht naar een zekere grenswaarde, waar tenminste 2/3 van het aantal WBSO-projecten uit de steekproef tot het thema behoren. Behoren minder dan 2/3 van het aantal projecten met een bepaalde projectscore tot het thema, dan vallen deze buiten de analyseresultaten. Dit is geen wetenschappelijk vastgesteld aandeel, maar ervaring leert inmiddels dat het een grens is die de thema's over het algemeen correct schat in aantal, mits de trefwoordenlijst zorgvuldig is samengesteld en wanneer er intensief contact is met een vakinhoudelijk-expert. Indien er bijvoorbeeld van de WBSO-projecten met een projectscore van 11 meer dan 2/3 tot het gezochte thema behoort, maar bij een projectscore van 10 nog maar 1/3 tot het thema behoort dan wordt de grenswaarde van de projectscore vastgesteld op 11. Alle WBSO-projecten met een projectscore van 11 en hoger worden dan meegeteld in de analyseresultaten, de projecten met een lagere projectscore niet. Deze methode wordt toegepast op uiteenlopende thema-analyses binnen de WBSO. Voor elke nieuwe thema-analyse wordt dit proces herhaald en heeft de grenswaarde per thema een andere waarde.

#### Valideren grenswaarde

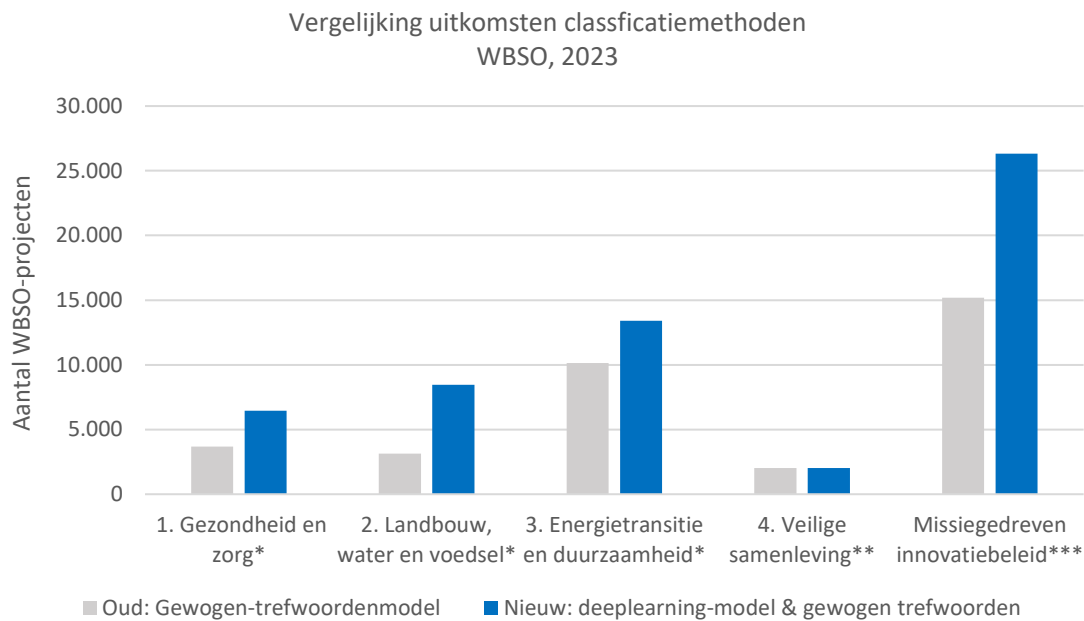
Op basis van verschillende steekproeven per projectscore wordt bepaald of projecten, via de eerdergenoemde grenswaarde, correct zijn geclassificeerd. Doordat het steekproeven betreft, worden niet alle projecten op deze manier handmatig gevalideerd. Met deze steekproeven wordt bepaald of de gekozen grenswaarde de verwachte resultaten toont.

In de jaren 2022 en 2024 zijn daarnaast nog aanvullende, grotere steekproeven uitgevoerd om de kwaliteit van de gewogen-zoektermenmethode, in combinatie met de gekozen grenswaarden, te kunnen vaststellen en te vergelijken met de resultaten van andere typen modellen.

#### Wijzigingen in uitkomsten door gebruik nieuwe methodiek per 2025

Zoals eerder beschreven is in 2025 een overstap gemaakt naar een combinatie van een deep learning-model en een gewogen-trefwoordenmodel. Voor projecten uit 2023 zijn beide modellen vergeleken. De resultaten worden hieronder weergegeven. Wat opvalt is

dat via de nieuwe methode meer projecten worden gevonden op de thema's Gezondheid en Zorg, Landbouw, Water, Voedsel, en Energietransitie en Duurzaamheid. Met name op het thema Landbouw, Water, Voedsel is een zeer grote stijging te zien.



\* Getraind via een deeplearning-model.

\*\*Getraind via een gewogen-trefwoordenmodel.

\*\*\* Missiegedreven innovatiebeleid: alle projecten die tot één van de subthema's behoren worden meegeteld. Het aantal projecten is ontdubbeld.

Vanaf 2024 wordt gebruik gemaakt van de nieuwe methodiek waarbij een combinatie gebruikt wordt van de verschillende modellen zoals hierboven beschreven. Deze resultaten zijn, zoals eerder genoemd, te vinden op [bedrijvenbeleid.nl](https://bedrijvenbeleid.nl), op de pagina [Versterken van innovatie in het bedrijfsleven | Bedrijvenbeleid in beeld](#).