Discussion paper

Vrijwilligerswerk en de Sustainable Development Goals

In samenwerking met Deedmob

Alexander Harms

Maart 2020

Inhoudsopgave

1. Inleiding 5
1.1 Doelstelling 6
1.2 Probleemstelling/vraagstelling 6

2. Vooronderzoek 7
2.1 Meten en rapporteren van vrijwilligerswerk 7
2.2 Meetinstrumenten van sociale impact 10
2.3 Gebruik van de Sustainable Development Goals 11

3. Data en methode 13
3.1 Data 14
3.2 Genereren van de kernwoorden 15
3.3 Generalisatie van kernwoorden 16
3.4 Indelen van vacatures naar SDG's 17
3.5 Classificatie met machine learning 18

4. Resultaten 20
4.1 Opstellen van de kernwoordenlijsten per SDG 20
4.2 Classificatie met machine learning 24

5. Conclusie 27
5.1 Toekomstig werk 28

6. Referentielijst 29

7. Bijlage 1. Gevonden kernwoorden in de vacatures 34

Samenvatting

De UN Sustainable Development Goals (SDG's) zijn doelen waarop de samenleving zich kan richten zodat deze zich op een duurzame manier ontwikkelt. De SDG's vormen hierdoor een ideaal raamwerk waarin de inzet voor of de impact van een maatschappelijk initiatief weergegeven kan worden. Uit interviews blijkt echter dat de SDG's niet aansluiten op de beleefwereld van (regionale) vrijwilligers en vrijwilligersorganisaties. Om de bijdrage van dit vrijwilligerswerk op de SDG's in kaart te brengen, zonder noodzaak voor aanvullende administratie door de vrijwilligers, moet gebruik gemaakt worden van de data die al voor handen is. Een voorbeeld van dergelijke data zijn de vacatures in vrijwilligersvacaturebanken. In dit rapport wordt een methode uitgewerkt waarmee het mogelijk is om op basis van steekwoorden een verbinding te maken tussen vacature-teksten en de SDG's.

In dit onderzoek wordt op basis van interviews de staat opgemaakt van de manieren waarop de maatschappelijke sector omgaat met de SDG's en de verzameling en analyse van data van het uitgevoerde vrijwilligerswerk. Uit deze interviews blijkt dat het gebruik van en de interesse in de SDG's binnen een organisatie meegroeit naarmate het operatiegebied meer richting nationaal en internationaal neigt. De SDG's nemen ook steeds een grotere rol aan binnen een organisatie op het moment dat financiering en verantwoording over de toepassing van deze financiering meer in beeld komen.

Tevens wordt in dit project als *proof-of-concept* een lijst met kernwoorden opgesteld waarmee het mogelijk is om vacatures van de Rijksoverheid te classificeren naar de SDG's.

Dankwoord

Dit paper is tot stand gekomen in samenwerking met Deedmob. Graag bedank ik Boudewijn Wijnands (Deedmob) voor zijn hulp, Frank Pijpers (CBS) voor de begeleiding en Sofie de Broe (CBS) en Peter Struijs (CBS) voor het opzetten van de verbinding met Deedmob en het initiëren van dit project. Daarnaast gaat mijn dank uit naar de mensen die ik mocht interviewen voor dit project: Judith Deynen (VC Utrecht), Joost van Alkemade (NOV), Micha Reusen (Bouwinvest), Steffen de Wolff (VSBFonds), Joringel den Hoedt en Gerben van der Linde (Gemeente Den Haag), Melanie Vaessen (Oranje Fonds) en Bob Groot (CBS). Ook bedank ik graag Arnout van Delden (CBS) voor zijn advies en het reviewen van dit paper op 16 januari 2020.

Keywords

Vrijwilligerswerk, UN Sustainable Development Goals, automatic keyword extraction.

1. Inleiding

In dit rapport wordt het resultaat beschreven van het onderzoek naar een koppeling tussen de 'UN Sustainable Development Goals' (SDG's) en vrijwilligerswerk. Dit onderzoek is uitgevoerd in samenwerking met Deedmob. Als input voor deze koppeling wordt gebruik gemaakt van vacatures voor vrijwilligerswerk zoals deze aangeboden worden op het Deedmob-platform of een andere vacaturebank. Een dergelijke koppeling opent de deuren voor het CBS om nauwkeurigere rapportages van vrijwilligerswerk en de bijdrage van dit vrijwilligerswerk aan de SDG's te produceren.

De Sustainable Development Goals zijn 17 doelen opgesteld door de Verenigde Naties (VN) en ondertekend door alle lidstaten in 2015 [1]. Een aantal voorbeelden van deze doelen zijn SDG 1 'Geen armoede', SDG 5 'Gendergelijkheid', SDG 8 'Eerlijk werk en economische groei' en SDG 13 'Klimaatactie'. Een overzicht van de doelen is te vinden in Figuur 1. De VN richt zich er op dat deze doelen voor 2030 behaald worden. De SDG's zijn een ideologisch vervolg van de 'UN Millienium Development Goals' waar de lidstaten van de VN van 2000 tot 2015 aan gewerkt hebben [2].

Figuur 1 De VN Duurzame Ontwikkelingsdoelstellingen, de Nederlandse vertaling van de 'UN Sustainable Development Goals' [3].

Bij het uitvoeren van vrijwilligerswerk, streeft men vaak de Sustainable Development Goals na, bijvoorbeeld door het verminderen van ongelijkheid of armoede in de samenleving. Het zou daarom interessant zijn om de bijdrage van vrijwilligerswerk aan de SDG's in kaart te brengen. Vanuit deze gedachte is in dit onderzoek in samenwerking met Deedmob gekeken naar de mogelijkheden van het gebruiken van de SDG's bij het rapporteren van vrijwilligerswerk.

Deedmob is in 2017 opgericht en is op het moment het grootste platform waar vrijwilligers en vrijwilligersorganisaties elkaar kunnen vinden in Nederland [4]. Zij bedienen ongeveer 1500 goede doelen en hebben een bereik van ongeveer 850.000 vrijwilligers [4], [5]. Andere organisaties die een nationale vrijwilligersvacaturebank aanbieden zijn onder andere 'NLvoorelkaar' en 'IkWilVrijwilligerswerkDoen.nl' [6], [7]. Ook op regionaal niveau zijn er initiatieven om vrijwilligers en vrijwilligersorganisaties aan elkaar te verbinden. Voorbeelden hiervan zijn de vacaturebanken van VC Utrecht of Vrijwilligerscentrale Amsterdam [8], [9].

In dit onderzoek is gekeken naar een methode om de vacatures in te delen naar SDG's. Om deze methode in de juiste context te plaatsen is ook gekeken naar de methoden om over vrijwilligerswerk te rapporteren en is onderzocht hoe men op dit moment omgaat met de SDG's in verschillende organisaties. Voor dit tweede deel van het onderzoek zijn onder andere ook interviews uitgevoerd.

1.1 Doelstelling

Het doel van dit project is tweeledig. In dit project wordt hoofdzakelijk gezocht naar een methode om vrijwilligerstaken in te delen naar de Sustainable Development Goals. Hiernaast wordt door middel van interviews met mensen uit de vrijwilligerssector onderzocht in hoeverre de SDG's op dit moment gebruikt worden bij het rapporteren over vrijwilligerswerk en andere activiteiten.

1.2 Probleemstelling/vraagstelling

De hoofdvraag van dit project is: "In hoeverre kunnen vrijwilligerstaken ingedeeld worden bij één of meerdere Sustainable Development Goals?". Om deze verdeling tot stand te laten komen wordt zo veel mogelijk gebruik gemaakt van voorgaand onderzoek en expertkennis.

Hiervoor zijn verschillende deelvragen gedefinieerd:
- Welk onderzoek is er al gedaan naar het rubriceren van vrijwilligerswerk of andere arbeid in de verschillende SDGs?
- Welke set van woorden karakteriseren de individuele SDGs?
- Is de informatie die beschikbaar is vanuit de omschrijvingen van de vrijwilligerstaken en –organisaties voldoende om de taken te classificeren naar de SDGs?

In sectie 2 wordt ingegaan op de context voor de methode die ontwikkeld is tijdens dit onderzoek. Als onderdeel hiervan worden ook de bevindingen uit de gevoerde interviews beschreven. In sectie 3 wordt de gebruikte data en worden de toegepaste methoden beschreven. In sectie 4 zijn de resultaten te zien van de toepassing van de methoden op de gebruikte data. Wegens problemen met beschikbaarheid van teksten van vrijwilligersvacatures voor dit onderzoek, wordt in deze sectie als resultaat de verdeling van een set vacatures van de Rijksoverheid over de SDG's gegeven.

2. Vooronderzoek

In deze sectie wordt ingegaan op de huidige stand van zaken omtrent het meten en rapporteren van vrijwilligerswerk en de voortgang richting de SDG's. Naast het raadplegen van literatuur en andere bronnen zijn er voor dit vooronderzoek ook interviews gehouden met verschillende mensen die werkzaam zijn in de maatschappelijke sector of die in hun werk zich bezighouden met vrijwilligerswerk of de SDG's.

Voor deze sectie is gebruik gemaakt van informatie uit interviews met:

Naam	Functie	Organisatie	Datum Interview
Judith Deynen	Directeur	VC Utrecht	1 november 2019
Joost van Alkemade	Directeur	Nederlandse Organisaties Vrijwilligerswerk (NOV)	1 november 2019
Micha Reusen	Head of Sustainability and Innovation	Bouwinvest	30 oktober 2019
Steffen de Wolff	Medewerker Externe betrekkingen	VSBFonds	15 november 2019
Joringel den Hoedt	Domeinmedewerker Onderwijs, Cultuur en Welzijn	Gemeente Den Haag	5 december 2019
Gerben van der Linde	Opgavenmanager werkplaats Kennis en Innovatie	Gemeente Den Haag	5 december 2019
Melanie Vaessen	Projectadviseur	Oranje Fonds	13 december 2019

2.1 Meten en rapporteren van vrijwilligerswerk

Uit een onderzoek van het CBS uit 2018 bleek dat in de 12 maanden ervoor iets minder dan de helft van de Nederlanders vrijwilligerswerk heeft uitgevoerd voor een organisatie of een vereniging [10]. Alles opgeteld vormt dit een grote bijdrage aan de maatschappij. Dit zorgt ervoor dat verdere inzichten in de wereld van de vrijwilligerssector zeer interessant zouden zijn. Bij het accuraat rapporteren van vrijwilligerswerk zijn echter een aantal hindernissen die genomen moeten worden.

De uitdagingen van meten en rapporteren van vrijwilligerswerk begint al bij de term 'vrijwilligerswerk'. Het CBS houdt de volgende definitie aan: *"Vrijwilligerswerk is werk dat in enig georganiseerd verband onverplicht en onbetaald wordt verricht ten behoeve van anderen of de samenleving."* [11]. Veel vrijwilligers zien echter hun activiteiten namelijk niet als vrijwilligerswerk. Ouders die helpen bij de sportclub van hun kind zien dit bijvoorbeeld al gauw als iets gewoons en zullen waarschijnlijk niet bewust hun taken als

vrijwilligerswerk bestempelen. Vanwege de brede definitie van de term 'vrijwilligerswerk' geeft het produceren van nauwkeurige cijfers op dit gebied de nodige uitdagingen. Door middel van een enquête produceert het CBS een statistiek voor het aandeel vrijwilligers in Nederland. In deze enquête gaat het CBS om met de hiervoor genoemde uitdaging door ook verder te vragen naar bijvoorbeeld de hulp die mensen bijdragen aan een sportclub, school of religieuze organisatie [10]. Ook in verschillende gemeentes, bijvoorbeeld Den Haag en Utrecht, wordt door middel van enquêtes het aandeel vrijwilligers in de gemeente gemeten [12], [13].

Met het uitvoeren van deze enquêtes komen echter ook aanvullende moeilijkheden. Naast de kosten van het enquêteren is het in de vrijwilligerssector ook belangrijk om de enquête- en administratiedruk zo laag mogelijk te houden. Het gebruik van de data van vacaturebanken of vrijwilligersplatforms zoals Deedmob zou hierbij een uitkomst kunnen zijn. Dit zorgt ervoor dat een deel van de gegevens over de sector verzameld kan worden zonder dat vrijwilligers of vrijwilligersorganisaties daar last van ondervinden.

Een belangrijke reden waarom er door de vrijwilligersorganisaties geen data verzameld wordt over de werkzaamheden is dat zij geen belang aan hechten aan deze gegevens. Binnen een vrijwilligersorganisatie bekommert men zich vooral om de noodzaak in de directe omgeving. Zij spelen in op een behoefte die zij zien en zien de relevantie niet van het verzamelen van gegevens die niet noodzakelijk zijn of waar verder geen actie op ondernomen wordt. Zelfs wanneer de data voor handen is wordt het om deze redenen vaak niet verwerkt en geanalyseerd. Een voorbeeld: Bij VC Utrecht zien zij het gelijk behandelen van iedere vrijwilligersorganisatie en vacature op hun vacaturebank als één van de pijlers van hun organisatie. Het is dus niet verwonderlijk dat zij geen prioriteit hebben om uit hun data te destilleren hoeveel vacatures er per categorie op hun website verschijnen. Er wordt namelijk niet gestuurd op deze informatie.

Het registreren en verwerken van data in de vrijwilligerssector wordt gestimuleerd als de vrijwilligers en vrijwilligersorganisaties zelf nut zouden ondervinden van de geproduceerde rapportages. Een reden voor vrijwilligersorganisaties om te rapporteren op het vrijwilligerswerk zou kunnen zijn dat het inspirerend werkt. Het is immers erg mooi als een organisatie kan zeggen dat 200 vrijwilligers aan hen verbonden zijn of dat al 400 uur aan vrijwilligerswerk naar het helpen van eenzame ouderen is gegaan. Het accuraat rapporteren van het verrichte vrijwilligerswerk kan ook helpen bij het indienen van aanvragen bij een particulier fonds of voor het aanvragen van subsidies. Het Oranje Fonds en het VSBFonds hechten bijvoorbeeld veel waarde aan sociale cohesie en de participatie van mensen in de samenleving. Zij zijn daarom bij een financieringsaanvraag ook benieuwd hoeveel vrijwilligers er betrokken zijn bij een organisatie. Een organisatie die het verrichte vrijwilligerswerk registreert kan bijvoorbeeld gemakkelijk informatie verzamelen over het aantal vrijwilligers in de organisaties en de eventuele groei in de afgelopen jaren.

Als het vrijwilligerswerk gerapporteerd zou worden, dan blijft de vraag welke grootheid hiervoor gebruikt wordt. De meest voor de hand liggende optie zou het aantal vrijwilligers zijn. Echter werkt niet iedere vrijwilliger dezelfde hoeveelheid uren. Normaliter is het rapporteren van het aantal uren verrichte vrijwilligerswerk niet mogelijk omdat de meeste vrijwilligersvacaturebanken of -organisaties niet bijhouden hoeveel uren een vrijwilliger werkt. Bij Deedmob wordt deze informatie echter wel opgeslagen en is het dus mogelijk om deze gegevens te rapporteren. Idealiter zou men willen weten wat het resultaat is van

het verrichte vrijwilligerswerk, maar dit is zeer gecompliceerd. In sectie 2.2 zal hier verder op ingegaan worden.

2.1.1 Categorisering binnen vacaturebanken

In de vorige sectie is besproken dat een toegankelijk beginpunt van dataverzameling binnen de vrijwilligerssector bij de vacaturebanken ligt. Om vervolgens te kunnen differentiëren in de hoeveelheid vrijwilligerswerk die op verschillende manieren ingezet wordt, is een eerste logische stap om te kijken naar de categorieën die de vacaturebank al gebruikt voor de vacatures. De vrijwilligersorganisaties bepalen zelf in welke categorieën hun vacatures terecht komen. Deze categorieën zijn dusdanig ingericht dat een vrijwilliger makkelijk een geschikte vacature kan vinden.

Bij Deedmob zijn de vacatures ingedeeld in een ‘Doel’, bijvoorbeeld ‘Ouderenhulp’, ‘Vluchtelingen & Minderheden’ en ‘Milieu & Duurzaamheid’ [14]. Deze termen geven een indicatie van het doel waar men als vrijwilliger aan mee helpt. Een vacature kan thuishoren bij meerdere doelen. Door het gebruik van deze categorieën kan snel een onderscheid gemaakt worden tussen de verschillende manieren waarop vrijwilligerswerk toegepast wordt. De verwachting is echter dat voor het produceren van een statistiek een andere onderverdeling beter geschikt is. In de onderverdeling zoals Deedmob die gebruikt zijn namelijk zowel doelen, bijvoorbeeld ‘Dierenwelzijn’, ‘Gezondheid & Welzijn’ en ‘Sociale rechtvaardigheid en ongelijkheid’, als doelgroepen, bijvoorbeeld ‘Mensen met beperkingen’, ‘Vluchtelingen & Minderheden’ en ‘Vrouwen’, vertegenwoordigd. Hier ziet men een verschil tussen een categorisering die opgezet is voor gebruiksvriendelijkheid en een indeling die bedoeld is voor de productie van statistieken. Voor een gebruiker van het platform maakt deze inconsistentie niet uit maar bij statistiekproductie spelen deze details een grotere rol.

In plaats van de vacatures te selecteren op basis van een 'Doel' kunnen zij ook op basis van 'Activiteit' gefilterd worden [14]. Op deze manier kan men aangeven hoe men wilt helpen. De activiteiten zijn opgedeeld in groepen zoals ‘Kantoor’, ‘Onderwijs & Opleiding’ en ‘Praktisch’. Door middel van de ‘Activiteit’ kan men bijvoorbeeld beter in kaart brengen welke vaardigheden er gevraagd worden in de vrijwilligerssector. De vacatures kunnen ook op zowel een doel als een activiteit geselecteerd worden.

De categorisering van de vacatures verschilt per vrijwilligersvacaturebank. Bij de VC Utrecht worden de activiteiten ook ingedeeld in een activiteit. De selectie is hier ‘Wat wil je doen?’ genoemd, waaronder keuzes zoals ‘Collecteren & Fondsenwerving’, ‘Koken / maaltijdverzorging’ en ‘Vervoer en Begeleiding’ schuil gaan. Naast een activiteit is er ook doelgroep te selecteren. Deze selectie is ‘Voor wie/met wie?’ genoemd en bevat keuzes zoals ‘Buurt & Wijk’, ‘Mensen met een beperking’ en ‘Mensen uit een andere cultuur’ [8]. In de vacaturebank van Den Haag Doet worden de vacatures ook geordend volgens de activiteit en de doelgroep [15]. De Vrijwilligerscentrale Amsterdam hanteert ook de verdeling in een activiteit en een doelgroep maar voegt hier ook de keuze van een sector, bijvoorbeeld ‘Evenementen’, ‘Kunst en cultuur’ en ‘Zorg’, aan toe [9].

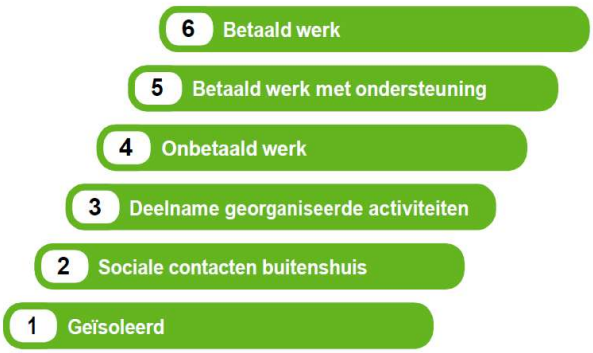
Doordat de gebruikte categorieën sterk verschillen met de SDG's is het bij meerdere vacaturebanken niet mogelijk om direct een link te leggen tussen de vrijwilligersvacatures en de SDG's.

2.2 Meetinstrumenten van sociale impact

Het meten van de impact op de samenleving van een project is een veel bestudeerd en betwist onderwerp. Om deze reden bestaan vele meetinstrumenten met ieder hun eigen methoden en gebieden van expertise. Een instrument dat in enkele interviews terug kwam is de 'Social Return On Investment' (SROI) [16]. Deze maat werkt op een gelijke manier als de 'reguliere' 'Return On Investment', de rendement op een investering. Het verschil is dat bij de SROI alle maatschappelijke effecten omgezet worden in een economische waarde en ook meegenomen worden in het bepalen van de grootte van het rendement. De uitdaging bij de SROI is dat deze conversie voor veel maatschappelijke effecten moeilijk te schatten is.

Daarnaast bieden vele bedrijven producten aan om sociale impact te meten, bijvoorbeeld Sinzer [17], Cigarbox [18] en MAEX [19]. Van deze drie bieden Sinzer en MAEX een product aan dat de impact van een project of organisatie meet op de SDG's. De impact van een vrijwilligersorganisatie wordt bij deze producten door de organisaties zelf toegeschreven aan de verschillende SDG's. Hierin schuilt nog steeds een verhoogde administratiedruk en een deel subjectiviteit. Het berekenen van het impact en het indelen bij de SDG's gebeurt vaak onder begeleiding van deze bedrijven of software die zij ontwikkelen.

Naast het meten van 'sociale impact' in algemene zin, zijn er ook specifieke meetinstrumenten. Uit de gevoerde interviews zijn enkele voorbeelden naar voren gekomen. Voor het meten van de participatie van mensen in de maatschappij heeft men de beschikking over onder andere de participatieladder [20], zoals te zien in Figuur 2, en de participatiematrix [21]. Op gezondheidsvlak kan men de vooruitgang van mensen meten met de Zelfredzaamheid-Matrix [22] en de vragenlijst Positieve Gezondheid [23]. Alle vier deze meetinstrumenten worden door de mensen zelf of door een behandelaar ingevuld.



Figuur 2 De participatieladder [20].

Een ogenschijnlijk simpel alternatief voor deze verschillende meetinstrumenten zou het berekenen van de 'replacement cost' of de vervangingskosten van een vrijwilliger zijn; de kosten die men zou moeten betalen als de vrijwilliger door een betaalde kracht vervangen wordt. Dit is een veelgebruikte maat als men de voordelen van vrijwilligerswerk in geld wil uitdrukken. Deze maat verliest echter zijn kracht in economische zin bij het uitvoeren van sommige vormen van 'corporate volunteering', vrijwilligerswerk als bedrijfsactiviteit. Het is natuurlijk aan te moedigen als de medewerkers van een advocatenkantoor zich, buiten hun reguliere werk, in willen zetten voor de maatschappij. Zij kunnen bijvoorbeeld door middel van corporate volunteering de gemeenschap helpen door onder werktijd de

plaatselijke kinderboerderij op te ruimen. Uit een economisch oogpunt zijn echter vraagtekens te zetten bij een advocaat die voor zijn gebruikelijke uurloon de geitenstal uitmest, waar waarschijnlijk lagere vervangingskosten voor gelden. Als deze zelfde advocaat echter op 'pro bono'-basis in deze zelfde tijd, of zelfs in onbetaalde tijd, iemand helpt met een juridisch probleem dan wordt er echter wel de economische waarde van de tijd behouden of wordt er zelfs economische waarde gecreëerd.

Een ander punt van kritiek op de vervangingskosten is dat bij het wegvallen van de vrijwilligers men vaak niet evenveel betaalde krachten in dienst neemt. In ref. [24] wordt de casus van vrijwilligers in een verzorgingstehuis genoemd. Deze vrijwilligers helpen de verzorgers door bijvoorbeeld met bewoners mee te gaan naar doktersbezoeken, te helpen bij excursies of door voor gezelligheid te zorgen in het tehuis. Een manager geeft in deze casus aan dat als alle vrijwilligers zouden stoppen dat het geen gevolgen zou hebben voor het verzorgingstehuis als organisatie, maar dat het wel een nadelig effect zou hebben op het leven van de bewoners. Er worden in deze situatie dan niet evenveel betaalde krachten aangenomen maar zal de organisatie simpelweg accepteren dat er bijvoorbeeld minder excursies mogelijk zijn. In dit geval is de bewering dat deze vrijwilligers hun vervangingskosten besparen of opleveren strikt gezien niet correct. Dit neemt niet weg dat deze mensen natuurlijk een positieve invloed hebben op de bewoners en daarmee de samenleving. In situatie zoals deze komt het verschil tussen de economische impact en de totale impact van een vrijwilliger naar voren.

2.3 Gebruik van de Sustainable Development Goals

Wanneer de input voor of impact van een maatschappelijk initiatief bekend is, dan kunnen de UN Sustainable Development Goals gebruikt worden om deze input of impact weer te geven als een bijdrage aan een breed spectrum aan doelen [1]. In het geval dat 10 vrijwilligers een werkdag lang helpen bij een initiatief dat zich inzet voor het bevorderen van onderwijs, dan kan men stellen dat er 80 uur naar SDG 4 'Kwaliteitsonderwijs' is gegaan. Naast de doelen op zich heeft de Verenigde Naties ook een lijst met indicatoren opgesteld waarmee lidstaten hun voortgang naar de SDG's kunnen bijhouden [25]. In dit rapport zal alleen gekeken worden naar de input van een maatschappelijk initiatief.

Onafhankelijk van de Sustainable Development Goals heeft het CBS de Monitor Brede Welvaart ontwikkeld [26]. Het doel van de Monitor Brede Welvaart was om naast economische veranderingen ook oog te houden op het welzijn in Nederland. In de Monitor Brede Welvaart wordt tevens gekeken naar de invloed die de staat van Nederland in het hier en nu heeft op de brede welvaart in de landen om ons heen en op de brede welvaart later. Aangezien de Sustainable Development Goals en de Monitor Brede Welvaart zich gezamenlijk bezig houden met zowel welvaart als welzijn, is in 2019 een publicatie verschenen waarin de stand van de brede welvaart vanuit beide ideeën bekeken wordt [27]. Vrijwilligerswerk verschijnt alleen bij SDG 10 'Ongelijkheid verminderen'. Het percentage vrijwilligers wordt gezien als een indicatie voor de sociale samenhang in de maatschappij.

Uit de interviews die voor dit project afgenomen zijn, bleek dat in de meeste organisaties de SDG's niet wijd gebruikt worden. Van de organisaties waarmee gesproken is, werkt alleen Bouwinvest op dit moment actief met de SDG's. Binnen de andere organisaties zijn

de doelen in de bovenlagen vaak bekend maar in de rest van de organisatie kent men de SDG’s niet. Zoals in sectie 2.1 aangegeven zijn vrijwilligers en vrijwilligersorganisaties vooral bezig met het helpen van hun medemens en hebben zij geen interesse in niet-essentiële zaken hier omheen, zoals het produceren van statistieken over het vrijwilligerswerk en de SDG's. Dit kwam naar voren uit gesprekken met de VC Utrecht en het NOV. Het NOV als organisatie ziet in de SDG’s weldegelijk het potentieel om, onder andere, internationaal verbinding te zoeken met andere vrijwilligersnetwerken. Het spreken van eenzelfde ‘taal’ vergemakkelijkt hier de communicatie. Een van de trends die naar voren kwam uit de interviews is dat de hoeveelheid interesse in en kennis van de SDG’s meeloopt met het spectrum tussen lokale en globale operatie. Een vrijwilligersinitiatief in de buurt zal veel minder affiniteit hebben met de SDG’s dan een nationale, of zelfs internationale vrijwilligersorganisatie.

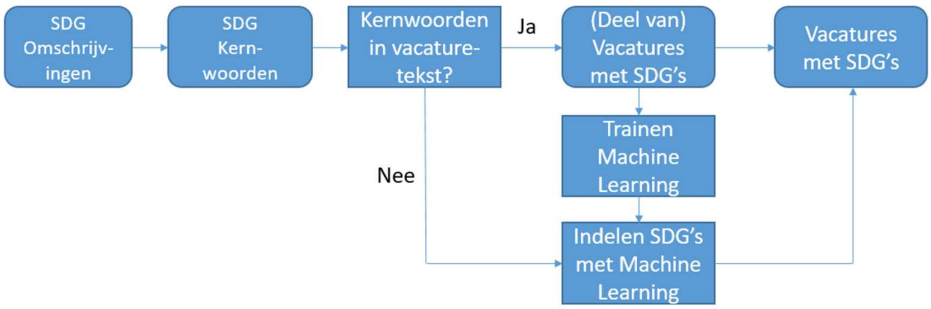
Een vergelijkbare trend is te zien langs het spectrum tussen partijen zonder en partijen mét winstoogmerk. Bij commerciële partijen wordt er vaak gevraagd om rapportage naar de SDG’s, bijvoorbeeld door aandeelhouders, om gemakkelijk maatschappelijke betrokkenheid te laten zien. Ook bij de fondsen, waar men vaak op zoek is naar manieren om makkelijk de maatschappelijke opbrengst als gevolg van hun financiering te tonen, leven de SDG’s meer dan bij partijen die zich niet bezig houden met het aanvragen of verstrekken van subsidies of andere geldstromen. Een gevaar bij het gebruik van de SDG’s om maatschappelijke betrokkenheid aan te tonen is dat er een fenomeen ontstaat vergelijkbaar met ‘green washing’, het ‘groener’ voordoen van de acties van een bedrijf dan dat ze eigenlijk zijn. Bedrijven kunnen initiatieven gaan ontplooien die op de oppervlakte lijken te werken aan een van de SDG’s maar die eigenlijk geen constructieve bijdrage aan de maatschappij in zich hebben.

3. Data en methode

Deze sectie beschrijft de methode waarop de vacature-teksten ingedeeld zullen worden naar één of meerdere Sustainable Development Goals. Ook zal de data die gebruikt wordt in dit onderzoek toegelicht worden. Voor het indelen van de vacature-teksten naar de SDG's zal gebruik gemaakt worden van een kernwoordenlijst. De kernwoorden op deze lijst kunnen uit losse woorden of uit combinaties van twee woorden bestaan. Zowel de losse woorden of de woordcombinaties zullen in dit verslag 'kernwoorden' genoemd worden. Op basis van beschrijvende teksten voor iedere SDG wordt een initiële lijst met kernwoorden opgesteld. Deze kernwoorden zullen vervolgens uitgebreid worden door met behulp van 'word embeddings' en een 'WordNet' te zoeken naar gerelateerde termen. De totale verzameling kernwoorden wordt vervolgens gebruikt om de vacatures in te delen naar SDG. Als een kernwoord van een SDG wordt gevonden in een vacaturetekst, dan wordt deze ingedeeld bij deze SDG. Een vacature kan bij meerdere SDG's thuis horen.

Bij het opstellen van de kernwoordenlijst worden alleen de beschrijvende teksten, de set aan word embeddings en het WordNet als bron voor de kernwoorden beschouwd. Na het opstellen deze lijst kunnen woorden handmatig verwijderd of toegevoegd worden. Aanvullende bronnen of software om alle vervoegingen van de kernwoorden in de lijst te verwerken zijn niet gebruikt.

Het kan zijn dat er geen kernwoorden gevonden worden in de vacaturetekst. Om deze vacatures toch in te delen bij een of meerdere SDG's wordt een machine learning-model gebruikt. Het machine learning-algoritme leert uit de combinaties van de vacatureteksten die wél door middel van de kernwoorden bij de SDG's ingedeeld konden worden en de bijbehorende SDG's het verband tussen de vacatures en de SDG's. Dit wordt het 'trainen' van het model genoemd. Na het trainen van het machine learning-model wordt dit model gebruikt om toch de vacatureteksten zonder kernwoorden in te delen bij de SDG's. In Figuur 3 is een overzicht gegeven van het proces.



Figuur 3 Diagram van de methode voor het indelen van de vacatures per SDG.

Een alternatief voor het gebruik van kernwoorden om de vacatures in de delen bij de Sustainable Development Goals is het handmatig indelen van de vacatures. Een machine learning-model kan dan uit deze handmatig gemaakte set van vacatureteksten en de SDG's leren hoe een mens de vacatures toekent aan de SDG's. Hier zijn echter voldoende vacatureteksten per Sustainable Development Goal voor nodig. Gezien het grote aantal doelen en de verwachting dat sommige doelen erg weinig terug zullen komen in de dataset

met vacatures, was het niet zeker dat aan deze voorwaarde voldaan zou worden voor iedere SDG. Om deze reden is gekozen om kernwoorden te gebruiken om de vacatures in te delen naar de SDG's. Deze voorwaarde voor het correct trainen van een machine learning-model geldt nog steeds voor het model dat gebruikt wordt om de vacatureteksten zonder kernwoorden in te delen bij de SDG's. Dit model zal hierdoor een hogere nauwkeurigheid hebben in het indelen van de vacatures bij veelvoorkomende SDG's dan bij SDG's waar slechts een klein aantal van de vacatures aan toebehoren.

De analyses van dit project zijn uitgevoerd in Python 3.6 [28].

3.1 Data

Voor het rubriceren van de vrijwilligersactiviteiten naar de SDG’s zouden in eerste instantie de vacatureteksten zoals die bekend zijn op het Deedmob-platform gebruikt worden. In verband met de GDPR waren deze teksten echter niet beschikbaar voor dit onderzoek. Als alternatief zijn de vacatureteksten van Werken voor Nederland gebruikt [29]. Dit zijn Nederlandse vacatureteksten voor vacatures binnen de Rijksoverheid. Voor deze set van vacatureteksten is gekozen omdat deze vrij verkrijgbaar was en vanwege de verwachting dat de functies binnen de Rijksoverheid beter aansluiten op een groter deel van de Sustainable Development Goals dan functies in de private sector.

In de periode van 30 oktober tot en met 18 december 2019 zijn via de API gedurende de werkweek op 29 dagen de openstaande vacatures op die data verzameld. Vervolgens zijn de dagelijkse vacaturelijsten samengevoegd en is de totale lijst ontdubbelt op basis van een unieke vacaturecode om de vacatures die meerdere dagen op de website beschikbaar waren terug te brengen tot een enkele vermelding. De dataset waarmee verder gewerkt wordt, bestaat uit 2375 vacatures die gedurende deze periode op de website van Werken voor Nederland geplaatst zijn. Door middel van het Python-package 'langdetect' zijn 10 vacatures met niet-Nederlandstalige omschrijvingen verwijderd [30]. De gebruikte dataset telt daarmee 2365 Nederlandstalige vacatures.

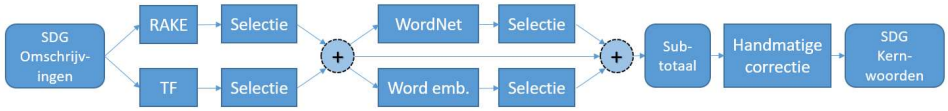
Van de vacatures zijn de volgende gegevens gebruikt:
- Naam en omschrijving van de baan
- Naam en omschrijving van de organisatie die de vacature uitschrijft
- Naam en omschrijving van de afdeling binnen de organisatie

Voor het vergaren van de kernwoorden zijn Nederlandse, beschrijvende teksten van de Sustainable Development Goals gebruikt. Deze omschrijvingen zijn afkomstig van:
- het artikel “Duurzame Ontwikkelingsdoelstellingen” van de Nederlandse Wikipedia [31],
- de Nederlandse omschrijvingen van de SDG’s van UNRIC [32],
- de omschrijvingen van de SDG’s van SDG Nederland [33],
- de omschrijvingen van de SDG’s uit de publicatie “Monitor Brede Welvaart & Sustainable Development Goals 2019” van het CBS [27],
- en de omschrijvingen van de SDG’s in de publicatie “Toepassing van SDG’s door Nederlandse organisaties” door PwC [34].

Zowel de vacatures als de omschrijvende teksten worden voor hun analyses eerst ontdaan van cijfers en interpunctie, waarna alle tekst omgezet wordt naar kleine letters.

3.2 Genereren van de kernwoorden

Door middel van de technieken 'Term Frequency' (TF) en 'Rapid Automatic Keyword Extraction' (RAKE) krijgen de woorden en combinaties van twee woorden uit de omschrijvingen van de SDG's genoemd in sectie 3.1 een score toebedeeld. Op basis van deze score worden de woorden geordend. Uit deze geordende lijst worden vervolgens handmatig passende kernwoorden gekozen. Deze set van geselecteerde woorden worden daarna door de technieken in sectie 3.3 uitgebreid. In Figuur 4 is een diagram weergegeven van het proces van de omschrijvingen van de SDG's tot de complete lijst van kernwoorden per SDG.



Figuur 4 Grafisch overzicht van het opstellen van de kernwoordenlijsten per Sustainable Development Goal.

3.2.1 Term Frequency (TF)

Als eerste bepaling van de kernwoorden van de beschrijvende teksten wordt er gekeken naar de meest voorkomende woorden in de tekst. Bij het berekenen van de Term Frequency wordt simpelweg geteld hoe vaak een term voor komt in een tekst.

Gezien het feit dat per SDG teksten uit meerdere bronnen beschikbaar zijn, zijn de volgende aanpakken toegepast:

- Neem per SDG de teksten uit alle bronnen samen en bereken de TF-scores op deze samengevoegde tekst.
- Neem per SDG de teksten uit alle bronnen samen en bereken de TF-scores voor alle combinaties van twee woorden uit deze samengevoegde tekst.
- Houdt per SDG de teksten uit de bronnen apart en bereken de TF-scores per tekst, en vergelijk de woorden met de hoogste scores. De tien hoogst scorende woorden krijgen een waarde toegekend van 1 tot en met 10. De waardes van de woorden uit verschillende bronnen worden vervolgens opgeteld.

Bij de eerste twee woordenlijsten worden alleen woorden of woordcombinaties meegenomen met een TF-score hoger dan 2. Woorden of woordcombinaties die minder dan 2 keer voorkomen in de tekst worden niet meegenomen in deze lijsten. Uit de drie woordenlijsten, gerangschikt op basis van de bijbehorende score, worden handmatig geschikte kernwoorden gekozen.

3.2.2 Rapid Automatic Keyword Extraction (RAKE)

Naast Term Frequency wordt er ook gezocht naar kernwoorden in de omschrijvingen door middel van het Rapid Automatic Keyword Extraction-algoritme [35]. Dit algoritme geeft door middel van een score per woord aan hoe belangrijk het woord is in het gegeven

document. Deze score wordt berekend door eerst de tekst op te splitsen in een vector van zinnen. Deze zinnen worden vervolgens opgedeeld op de plekken van interpunctie en stopwoorden. De woorden in de overgebleven tekstdelen krijgen een punt als zij voorkomen in een tekstdeel en een aanvullend punt voor ieder ander woord dat in hetzelfde tekstdeel voorkomt. Deze punten worden opgeteld, dit wordt de 'word degree' genoemd, en gedeeld door het aantal keer dat het woord voorkomt in de tekst, de 'word frequency' genoemd. Voor ieder tekstdeel worden deze score per woord opgeteld. In dit rapport wordt alleen gekeken naar tekstdelen van 1 of 2 woorden lang. Door middel van het RAKE-algoritme worden deze tekstdelen geordend op basis van de RAKE-score. De beschrijvende teksten van de SDG's uit de vijf bronnen worden samengevoegd tot één tekst waarop het RAKE-algoritme is toegepast. Op basis van de geordende lijst van tekstdelen wordt een handmatige selectie van kernwoorden gemaakt.

Voor het toepassen van het RAKE-algoritme is binnen Python gebruik gemaakt van de software ‘rake_nltk (1.0.4)’ [36]. Als parameters zijn de waardes "language = 'dutch'", 'min_length = 1' en 'max_length = 2' bovenop de standaard parameters gebruikt.

3.3 Generalisatie van kernwoorden

Om met deze kernwoorden een zo groot mogelijke hoeveelheid vacatures in de kunnen delen naar de juiste SDG's, zullen deze kernwoorden worden aangevuld met woorden met soortgelijke betekenissen. Hiervoor wordt gebruik gemaakt van een WordNet, specifiek het Open Dutch WordNet, en van word embeddings [37], [38], [39]. Zie Figuur 4 voor de plaats van deze stappen in het totale proces.

3.3.1 WordNet

Een WordNet is een collectie van woorden die ieder verbonden zijn door semantische relaties [40]. Voorbeelden hiervan zijn de relaties met hypero- en hyponiemen, overkoepelende termen en meer specifieke woorden. Het woord ‘kleur’ is bijvoorbeeld een hyperoniem van ‘rood’ en ‘blauw’; ‘rood’ en ‘blauw’ zijn op hun beurt weer hyponiemen van het woord ‘kleur’. Deze en andere relaties, zoals synoniemen en antoniemen, kunnen nu ook geëxploiteerd worden om vergelijkbare woorden te vinden. In dit onderzoek wordt alleen gekeken naar woorden in het WordNet die een directe relatie hebben met de gegenereerde woorden uit de omschrijvingen van de SDG's.

Als Nederlands WordNet wordt het Open Dutch WordNet gebruikt [38]. Dit WordNet wordt verkend door middel van de software in Python die hiervoor geschreven is door de onderzoekers achter het Open Dutch WordNet project.

3.3.2 Word embeddings

Om de 'word embedding' van een woord te verkrijgen wordt een specifiek algoritme toegepast op een tekst. Dit algoritme geeft aan ieder woord in de tekst een vector in een multidimensionale ruimte op basis van de context van het woord binnen deze tekst [39]. Woorden die op een vergelijkbare manier gebruikt worden zullen vaak in dezelfde context voorkomen en zullen hierdoor vectoren in de multidimensionale ruimte toebedeeld krijgen die eenzelfde kant op gericht zijn. Een vector wordt genoteerd als een reeks getallen die

even lang is als het aantal dimensies in de gebruikte ruimte. Als de word embeddings behandeld zouden worden als vectoren in een driedimensionale ruimte dan zouden de word embeddings uit drie getallen bestaan.

In dit rapport wordt een verzameling 160-dimensionale word embeddings gebruikt die zijn gegenereerd op basis van het SoNaR-500 corpus [41]. Dit is een verzameling van 500 miljoen hedendaagse Nederlandse woorden verzameld door de Nederlandse Taalunie [42]. De word embeddings zijn niet verder getraind op de teksten in de vacatures. Van de gegenereerde woorden uit sectie 3.2 worden eerst de bijbehorende word embeddings uit de verzameling gezocht waarna per woord de tien woorden met de meest vergelijkbare word embeddings genoteerd worden. De gelijkenis tussen de word embeddings wordt bepaald door middel van de cosinusgelijkenis. Dit is de cosinus van de hoek tussen de vectoren. Vectoren die dezelfde richting op wijzen hebben een cosinusgelijkenis van 1; vectoren die loodrecht op elkaar staan hebben een cosinusgelijkenis van 0.

Voor deze stap in de analyse is in Python de software ‘gensim (3.8.1)’ gebruikt [43].

3.4 Indelen van vacatures naar SDG's

De gebruikte vacatures worden uitgeschreven door een organisatie, bijvoorbeeld een ministerie of een uitvoeringsorganisatie. In de meeste gevallen behoren de vacatures ook toe aan een afdeling binnen deze organisatie. Naast dat iedere baan waarvoor de vacature uitgeschreven wordt een beschrijvende tekst heeft, heeft ook iedere organisatie en afdeling een eigen beschrijvende tekst.

Voordat de vacatures toegekend kunnen worden aan een bepaald Sustainable Development Goal moet er eerst bepaald worden wanneer een activiteit precies bijdraagt aan een bepaald doel. Een activiteit zoals het ruimen van zwerfafval uit het plaatselijke meer is duidelijk toe te schrijven aan SDG 14 ‘Leven in het water’, maar bij het werk als boekhouder voor de Voedselbank is dit al moeilijker te bepalen. Van de activiteit ‘boekhouden’ op zich is niet direct te zeggen dat het helpt bij de SDG’s 1 en 2, ‘Geen armoede’ en ‘Geen honger’, maar met zijn of haar activiteit helpt de boekhouder de Voedselbank wél om deze doelen te bereiken. Hierbij wordt eigenlijk erkent dat men met iedere activiteit die ondersteuning biedt aan een organisatie die naar het behalen van de SDG’s streeft, ook zelf meewerkt aan dezelfde SDG’s. In lijn met deze gedachte worden de vacatures als volgt ingedeeld bij de SDG's:

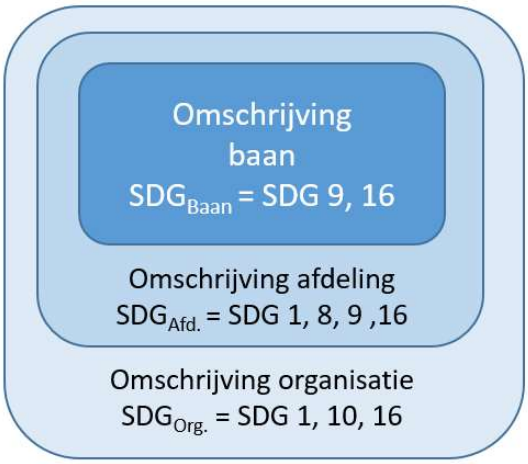
- Iedere baan kan van zichzelf SDG’s nastreven.
- De SDG’s die een afdeling nastreeft zullen ook toegekend worden aan de posities binnen deze afdeling.
- Iedere afdeling en werknemer binnen een organisatie streeft ook de SDG's na die de organisatie in zijn geheel nastreeft.

De vacature in totaal behoort toe aan de combinatie van de SDG's van de baan, de afdeling en de organisatie. Deze aanpak is grafisch weergegeven in Figuur 5. In deze afbeelding staat ter voorbeeld de indeling van een vacature uitgewerkt.

Een risico van deze aanpak is dat hier ‘het doel heiligt de middelen’ in verborgen zit. In de context van vrijwilligersactiviteiten zal de kans klein zijn dat een tak van een vrijwilligersorganisatie activiteiten uitoefent die tegen het hoofddoel van de organisatie in

gaan. Bij het toepassen van deze methode op banen in de private sector zal deze kans groter zijn en zal men dit in ogenschouw moeten nemen.

Een kernwoord wordt alleen herkend in de tekst als het losstaand voorkomt. Dit betekent dat kernwoorden niet opgemerkt worden als zij voorkomen in een samengesteld woord. Om te controleren of het indelen van de vacatures naar de SDG's correct gebeurt worden de ingedeelde vacatures steekproefsgewijs bekeken.



$$SDG_{Vacature} = SDG_{Baan} \ \& \ SDG_{Afd.} \ \& \ SDG_{Org.} = SDG\ 1, 8, 9, 10, 16$$

Figuur 5 Diagram van de methode voor het indelen van de vacatures naar de SDG's. De genoemde SDG's zijn een voorbeeld.

3.5 Classificatie met machine learning

Om vacatures waar geen van de kernwoorden in gevonden zijn in te delen bij een of meerdere SDG's, kan gebruik gemaakt worden machine learning. Aangezien een vrijwilligerstaak aan meer dan één SDG gekoppeld kan worden is het noodzakelijk om multi-label classificatie toe te passen.

Het machine learning-model wordt getraind op alle teksten die met behulp van kernwoorden ingedeeld zijn bij een of meerdere SDG's. De aanname wordt gemaakt dat deze teksten perfect zijn ingedeeld bij de SDG's. Bij het trainen worden alle omschrijvingen van de banen, afdelingen en organisaties apart meegenomen. Deze teksten worden door middel van 'Term Frequency - Inverse Document Frequency' (TF-IDF) getransformeerd waarna een Random Forest-model wordt getraind op de data. Voor het meten van de prestaties van het model wordt gebruik gemaakt van ‘k-folding’ met vijf segmenten. De data wordt op een aselecte manier in vijf delen opgesplitst. Per test wordt 4/5e van de data, de train-dataset, gebruikt om het model te trainen waarna de bijbehorende SDG's voor de overige 1/5e van de teksten, de test-dataset, voorspeld wordt. Deze voorspelling wordt vervolgens gestaafd aan de SDG's die door middel van de aan de teksten verbonden zijn. Op basis van deze twee sets SDG's wordt de ‘Hamming loss’ bepaald; dit is het percentage aan Sustainable Development Goals dat verkeerd is voorspeld.

In Tabel 1 is ter illustratie van de Hamming loss een voorbeeld uitgewerkt voor drie teksten. Als een tekst ingedeeld wordt bij een SDG maar hier niet thuis hoort, dan wordt dit geteld als een verkeerde voorspelling. Omgekeerd als een tekst niet ingedeeld wordt bij een SDG terwijl dit wel zou moeten, dan wordt dit bij het aantal verkeerde voorspellingen opgeteld. Het aantal verkeerde voorspellingen voor alle teksten wordt vervolgens gedeeld door het aantal doelen die correct voorspeld hadden kunnen zijn. Dit zijn 17 doelen per tekst.

Voor het trainen en testen van het machine learning-model is de software 'scikit learn (0.20.2)' gebruikt [44]. Voor de TF-IDF-transformatie zijn de standaard parameters gegeven door scikit learn gebruikt. Het Random Forest-model gebruikt de hyperparameter 'n_estimators = 100' als aanvulling op de standaard hyperparameters van het model.

Tabel 1 Voorbeeld ter illustratie van de berekening van de Hamming loss.

Tekst	Voorspelling SDG's	Echte SDG's	Aantal verkeerde voorspellingen
1	1, 2, 8, 10	1, 8, 10	1
2	13, 14	6, 14, 17	3
3	8, 9, 12	8, 9, 12	0
Totaal			4
Hamming loss			4/51 = 7.8 %

4. Resultaten

In dit hoofdstuk worden de resultaten gepresenteerd van het genereren van de kernwoordenlijsten per SDG en het indelen van de vacatures naar de SDG's.

4.1 Opstellen van de kernwoordenlijsten per SDG

In Tabel 2 is een overzicht te zien van de aantallen kernwoorden die gevonden zijn per analyse-stap. Kernwoorden kunnen bestaan uit één of twee woorden. Het proces van het produceren van de kernwoordenlijsten is weergegeven in Figuur 4. In de kolom 'TF & RAKE' ziet men onder 'Gev.', van 'Gevonden', het aantal kernwoorden dat door de gebruikte technieken beschreven in sectie 3.2 zijn gerangschikt op basis van de TF- of RAKE-score. Van deze kernwoorden zijn de meest geschikte kernwoorden handmatig geselecteerd. In de kolom 'Sel.', van 'Selectie', onder 'TF & RAKE' is te zien hoeveel kernwoorden deze selectie beslaat. Op basis van deze geselecteerde termen zijn vervolgens door het gebruik van word embeddings en een WordNet, zoals beschreven in sectie 3.3, lijsten met gerelateerde woorden gegenereerd. De aantallen woorden, niet voorkomend in de selectie van de 'TF & RAKE'-stap, die gegenereerd zijn met word embeddings of het WordNet zijn te vinden in de kolommen 'Gegen.', van 'Gegenereerd', onder respectievelijk 'Word Emb.' en 'WordNet'. Van deze lijsten zijn weer handmatig de meest geschikte woorden geselecteerd. Het aantal woorden is weergegeven in de kolommen 'Sel.'. Het aantal unieke woorden per SDG dat geselecteerd is uit woordenlijsten geproduceerd door middel van Term Frequency en RAKE, word embeddings en het WordNet is weergegeven in de kolom 'Subtot.', van 'Subtotaal'. Door de overlap tussen de verschillende SDG's komt het voor dat kernwoorden op de kernwoordenlijsten van meerdere SDG's terugkomen. Een voorbeeld hiervan is het kernwoord 'arbeid' dat terug te vinden is bij SDG 1 'Geen Armoede' en SDG 8 'Eerlijk werk en economische groei'.

Op basis van de detectie van de kernwoorden in de baan-, afdeling- en organisatieomschrijvingen, zijn de omschrijvingen ingedeeld bij de SDG's. Van deze omschrijvingen zijn ieder 100 exemplaren gecontroleerd en op basis van deze controle zijn handmatig woorden verwijderd uit de lijst van kernwoorden of juist toegevoegd. De aantallen hiervan zijn in de kolom 'Handmatig' van Tabel 2 te zien. De reden voor het verwijderen van woorden zat meestal in de ambiguïteit van de woorden; bij SDG 3 'Goede gezondheid en welzijn' is bijvoorbeeld het woord 'zorg' verwijderd aangezien dit woord veelvuldig voorkomt buiten de context van gezondheidszorg. Een omschrijving van Rijkswaterstaat werd bijvoorbeeld toegekend aan SDG 3 door de zin
*"… De **zorg** voor onze veiligheid op weg en water vereist een intensieve en goede samenwerking. ..."*.
Om deze verkeerde classificatie te voorkomen is het woord 'zorg' verwijderd van de lijst met kernwoorden voor SDG 3.

Woorden die juist toegevoegd zijn, zijn vaak zeer specifieke woorden die niet in een algemene omschrijving van een SDG voorkomen. Voorbeelden in het geval van SDG 16 "Vrede, justitie en sterke publieke diensten" zijn 'cybersecurity', 'mededinging' en

‘kartelvorming’. De omschrijving van het Nationaal Cyber Security Centrum, onderdeel van de Nationaal Coördinator Terrorismebestrijding en Veiligheid, luidt als volgt:
*"Het Nationaal **Cyber Security** Centrum (NCSC) is onderdeel van de Nationaal Coördinator Terrorismebestrijding en Veiligheid en helpt de weerbaarheid van de Nederlandse samenleving in het digitale domein verbeteren. Dan doen we via samenwerkingen tussen bedrijfsleven, overheid en **wetenschap**. … NCSC is daarmee het centrale meld- en informatiepunt voor **ICT**-dreigingen en -veiligheidsincidenten. Het NCSC kenmerkt zich door de talentvolle, ambitieuze en betrokken mensen die er werken. Zij zetten zich 24/7 in om Nederland digitaal veiliger te maken. Werken in het hart van ‘**cybersecurity** Nederland’ is zeer dynamisch en uitdagend. …"*
Zonder de toevoeging van de termen 'cyber security' en 'cybersecurity' zou deze omschrijving op basis van de termen 'ict' en 'wetenschap' alleen bij SDG 9 'Industrie, innovatie en infrastructuur' ingedeeld worden. Het nastreven van veiligheid op het digitale vlak, en daarmee ook het NCSC, sluit echter naadloos aan op SDG 16. Om dit recht te zetten was het noodzakelijk om de termen 'cybersecurity' en 'cyber security' toe te voegen aan de lijst met kernwoorden.

Tabel 2 Overzicht van het aantal kernwoorden die gevonden zijn per stap van het proces van het opstellen van de kernwoordenlijsten.

	TF & RAKE		Word Emb.		WordNet			Handmatig		
SDG	Gev.	Sel.	Gegen.	Sel.	Gegen.	Sel.	Subtot.	Verw.	Toeg.	Totaal
1	418	42	164	38	189	13	93	3	2	92
2	582	72	314	69	428	73	214	1	0	213
3	558	104	378	155	420	88	345	4	0	341
4	397	41	146	36	143	34	111	1	0	110
5	469	30	118	17	191	10	57	1	0	56
6	446	76	243	53	254	7	136	0	0	136
7	400	54	149	51	238	23	128	1	1	128
8	529	51	232	46	347	13	110	3	2	109
9	667	84	251	49	756	24	156	2	0	154
10	504	37	76	22	45	2	61	0	2	63
11	591	50	193	34	421	3	87	1	0	86
12	460	50	121	23	116	27	100	0	0	100
13	404	39	140	17	49	0	56	0	1	57
14	550	65	184	30	156	5	100	1	0	99
15	459	63	149	27	257	12	100	0	0	100
16	530	63	297	64	340	19	145	2	8	151
17	531	35	96	14	40	2	50	2	1	49

Het handmatig aanpassen van de kernwoordenlijsten is noodzakelijk aangezien het taalgebruik tussen de omschrijvingen en vacatureteksten erg kan verschillen. Ook bij het overzetten van deze kernwoordenlijst naar vrijwilligersvacatures zullen handmatige aanpassingen nodig zijn.

In Bijlage 1 is een overzicht te zien van de kernwoorden die herkend zijn in de teksten.

In Tabel 3 zijn te zien hoeveel beschrijvingen van de organisaties, afdelingen of de baan op zich ingedeeld worden volgens welke SDG wanneer de kernwoorden voorkomen in deze

omschrijvingen. In de kolom ‘Totaal Vacatures’ staan het aantal vacatures dat toebedeeld wordt aan een SDG als de procedure uit sectie 3.4 wordt toegepast.

Tabel 3 Overzicht van het aantal omschrijvingen van de baan, afdelingen of organisaties dat ingedeeld wordt naar een bepaalde SDG. In de kolom ‘Totaal Vacatures’ zijn de hoeveelheid vacatures die in ingedeeld naar een SDG zoals beschreven in sectie 3.4. Bij 'Geen' zijn het aantal ingevulde omschrijvingen zonder SDG genoteerd. Een baan, afdeling, organisatie, of vacature kan bij meerdere SDG's ingedeeld zijn.

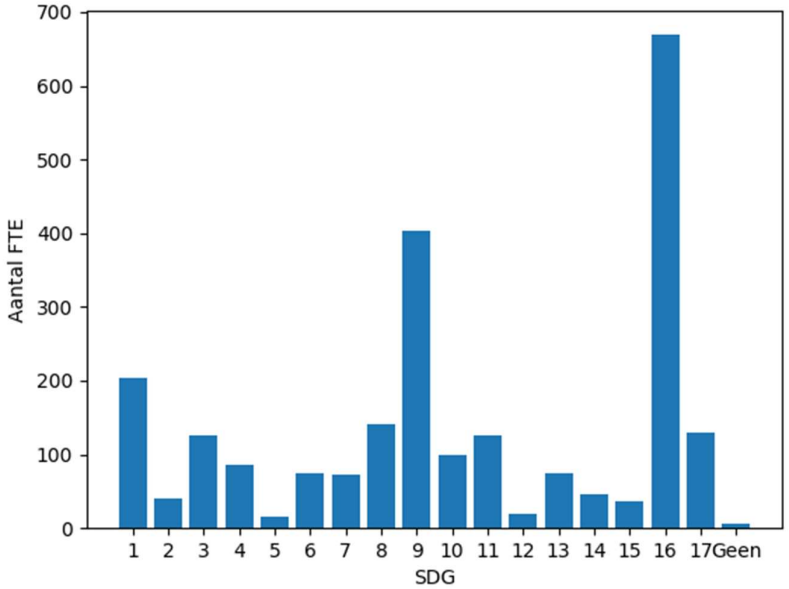
SDG	Omschrijving Baan	Omschrijving Afdeling	Omschrijving Organisatie	Totaal Vacatures
1	371	317	633	891
2	138	111	108	215
3	246	286	167	445
4	183	95	148	317
5	30	47	2	69
6	125	128	325	402
7	130	64	203	344
8	348	256	236	591
9	1028	784	629	1467
10	42	60	414	451
11	179	203	393	598
12	42	60	48	130
13	115	74	408	463
14	26	84	264	296
15	102	142	83	227
16	1097	843	1529	1887
17	77	65	320	399
Geen	317	681	74	6

Bij het indelen van teksten in groepen met kernwoorden, zeker wanneer deze met de hand geselecteerd worden, kan de verdeling vaak gestuurd worden door de selectie van de kernwoorden of door de keuze van de brontekst voor de kernwoorden. De kernwoorden 'belastingen' en 'subsidies' zijn bijvoorbeeld toegevoegd aan de kernwoordenlijst voor SDG 10 'Ongelijkheid verminderen' aangezien het doel van het belasting- en subsidiestelsel is om inkomensongelijkheid te verminderen. Het is echter ook een verdedigbaar standpunt dat deze kernwoorden bij SDG 1 ‘Geen Armoede’ thuishoren. Een positie met een invloed op de belastingen kan immers ook de financiële thuissituatie van de burger veranderen en daarmee armoede verminderen. De overlap tussen de SDG's maakt het een uitdaging om strikt objectief de vacatures in te delen bij de verschillende Sustainable Development Goals. In overleg met Deedmob is gekozen om 'belastingen' en 'subsidies' bij SDG 10 in te delen.

Een uitschieter in Tabel 3 is SDG 16 “Vrede, justitie en sterke publieke diensten”; deze hoge waarden zijn waarschijnlijk te danken aan de inclusie van de kernwoorden ‘justitie’ en ‘defensie’ voor dit Sustainable Development Goal. Dit zorgt ervoor dat alle vacatures van het Ministerie van Justitie en Veiligheid en het Ministerie van Defensie aangezien worden als bijdragend aan SDG 16. Een andere uitschieter is SDG 9. De reden hiervoor is dat woorden zoals 'onderzoek' voorkomen in de kernwoordenlijst voor deze SDG. Bij een groot deel van de banen wordt het uitvoeren van onderzoek genoemd in de omschrijving.

Iedere vacature kan aan meerdere SDG's toebehoren. In dit onderzoek is de keuze gemaakt dat een vacature die ingedeeld is bij meerdere SDG's evenveel bijdraagt aan iedere SDG. Als aangenomen wordt dat ieder vacature geldt voor 1 FTE, een 'full-time equivalent' ofwel een voltijd positie, dan is met deze verdeling het aantal FTE te berekenen dat binnen deze vacature-dataset naar iedere SDG gaat. Als een vacature bijdraagt aan twee Sustainable Development Goals, dan wordt deze bijdrage gerekend als een halve FTE richting het ene en een halve FTE richting het andere doel. Hierbij is verondersteld dat alle vacatures uit de dataset ingevuld zijn en dat de mensen in deze functies niet opnieuw van baan gewisseld zijn. Deze verdeling is te zien in Tabel 4 en in Figuur 6. In Tabel 4 zijn het aantal FTE per SDG afgerond op gehele getallen. Door deze afronding valt de som van het aantal FTE in de tabel met 2 FTE hoger uit dan het totaal aantal vacatures. In Figuur 6 is te zien dat de meeste menskracht naar SDG 16, 9 en 1 gaan. In Tabel 4 ziet men dat meer FTE aan vacatures besteed worden aan SDG 16, 669 FTE, dan aan SDG 1 en 9 samen, welke gelijk zijn aan 606 voltijd-functies.

Met de aanname dat iedere FTE geldt voor 36 uur per week is ook een aantal uren te koppelen aan de SDG's. De uren zijn afgerond naar de dichtstbijzijnde tiental uren. In Tabel 4 ziet men het aantal uren dat, bij het invullen van alle vacatures, extra per week besteed worden bij de Rijksoverheid aan de verschillende SDG's. In Tabel 4 ziet men dat onder deze aanname de 669 nieuwe FTE voor SDG 16 gelijk zijn aan ongeveer 24 duizend uur per week.



Figuur 6 Overzicht van het aantal FTE verdeeld naar de Sustainable Development Goals.

Tabel 4 Overzicht van het aantal FTE en uren per week per SDG op basis van de vacatures in de dataset. De aanname is gemaakt dat iedere vacature evenveel bijdraagt aan iedere SDG waar een vacature is ingedeeld. Ook is aangenomen dat iedere vacature telt voor één FTE van 36 uur per week.

SDG	Naam SDG	Aantal FTE	Uren per week [36 uur/FTE]
1	Geen armoede	203	7.310
2	Geen honger	40	1.440
3	Goede gezondheid en welzijn	125	4.500
4	Kwaliteitsonderwijs	86	3.100
5	Gendergelijkheid	16	580
6	Schoon water en sanitair	74	2.660
7	Betaalbare en duurzame energie	72	2.590
8	Eerlijk werk en economische groei	141	5.080
9	Industrie, innovatie en infrastructuur	403	14.510
10	Ongelijkheid verminderen	99	3.560
11	Duurzame steden en gemeenschappen	126	4.540
12	Verantwoorde consumptie en productie	20	720
13	Klimaatactie	74	2.660
14	Leven in het water	46	1.660
15	Leven op het land	37	1.330
16	Vrede, justitie en sterke publieke diensten	669	24.080
17	Partnerschappen om doelstellingen te bereiken	129	4.640
Geen		6	220
Totaal		2.365 + 1	85.140 + 40

4.2 Classificatie met machine learning

Uit 2365 vacatures worden 6786 teksten gehaald die fungeren als omschrijving van de baan, het departement of de organisatie waar de vacature uitgeschreven wordt. Van deze teksten zijn er 4641 uniek. Binnen deze groep van 6786 omschrijvingen zijn er 6023 toegekend aan één of meerdere Sustainable Development Goals. Van deze 6023 teksten zijn 4031 uniek.

Bij het toepassen van de k-folding-techniek op de dataset van teksten die ingedeeld zijn bij één of meerdere SDG's, ziet men een gemiddelde Hamming loss van 4,30 %. Dit komt neer op gemiddeld één fout per 23 voorspelde SDG's.

Na deze test is het machine learning-model getraind op alle omschrijvingen met minimaal één toegekende SDG en zijn met dit model alle teksten zonder kernwoorden voor een SDG, ingedeeld bij één of meerdere SDG's. Dat wil zeggen dat ook omschrijvingen, bijvoorbeeld organisatie-omschrijvingen, waar geen kernwoorden in gevonden worden, alsnog door het machine learning-model worden ingedeeld bij de SDG's. Dit betekent dat dit ook gebeurt als de vacature in zijn geheel wél ingedeeld is bij één of meerdere SDG's. Het kan hierdoor gebeuren dat een vacature bij extra SDG's ingedeeld wordt, waardoor de FTE van die

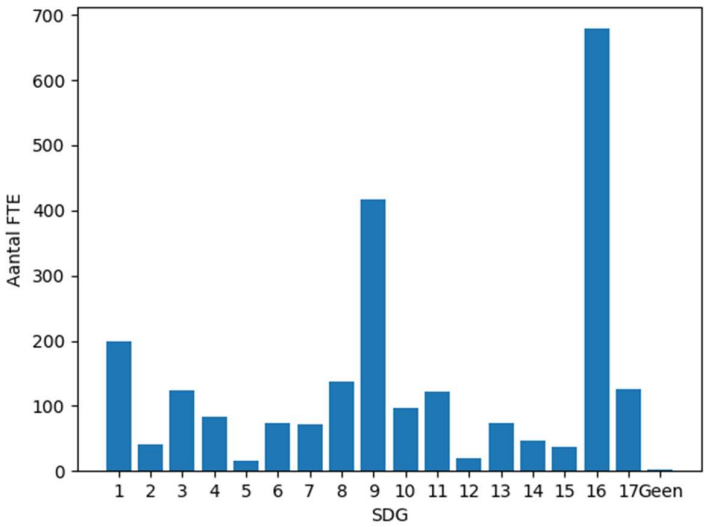
vacature over meer SDG's verdeeld wordt. Dit heeft als gevolg dat het aantal FTE's bij een SDG kan dalen na het toepassen van het machine learning-model.

Na het indelen van de omschrijvingen zonder kernwoorden met behulp van het machine learning-model, is de methode zoals die beschreven is in sectie 3.4 gebruikt om deze indelingen van de omschrijvingen naar de SDG's te vertalen naar de bijbehorende SDG's bij de vacature in zijn geheel.

In Tabel 5 ziet men de uitkomsten van deze methode in vergelijking met de methode waarin alleen kernwoorden worden gebruikt. De waardes voor het aantal FTE's zijn afgerond op hele waarden en het aantal uur per week op tientallen uren. In deze tabel is te zien dat door het gebruik van het machine learning-model vier van de zes vacatures zonder SDG's ingedeeld zijn bij de SDG's. Door het toevoegen van het machine learning-model in het proces van het indelen van de vacatures naar de SDG's is te zien dat het aantal FTE bij SDG's 9 en 16 is toegenomen en bij SDG's 1, 8, 10, 11 en 17 met meer dan twee FTE is afgenomen. In Figuur 7 zijn het aantal FTE per SDG weergeven als zowel de kernwoorden als het machine learning-model gebruikt worden bij het indelen van de vacatures bij de SDG's.

Tabel 5 Overzicht van het aantal FTE dat vacatures uit de gebruikte dataset toevoegen aan de SDG's als zij alleen door middel van de kernwoorden of door middel van kernwoorden en het machine learning-model ingedeeld worden bij één of meerdere SDG's.

SDG	Alleen kernwoorden		Kernwoorden en machine learning		Verschil	
	FTE	*Uren/week*	*FTE*	*Uren/week*	*FTE*	*Uren/week*
1	203	7.310	200	7.200	-3	-110
2	40	1.440	40	1.440	0	0
3	125	4.500	123	4.330	-2	-170
4	86	3.100	84	3.020	-2	-80
5	16	580	16	580	0	0
6	74	2.660	74	2.660	0	0
7	72	2.590	71	2.560	-1	-30
8	141	5.080	138	4.970	-3	-110
9	403	14.510	418	15.050	+15	+540
10	99	3.560	96	3.460	-3	-100
11	126	4.540	122	4.390	-4	-150
12	20	720	20	720	0	0
13	74	2.660	73	2.630	-1	-30
14	46	1.660	46	1.660	0	0
15	37	1.330	37	1.330	0	0
16	669	24.080	679	24.440	+10	+360
17	129	4.640	126	4.540	-3	-100
Geen	6	220	2	70	-4	-150
Totaal	2.365 + 1	85.140 + 40	2.365	85.140 + 10	-1	-130



Figuur 7 Overzicht van het aantal FTE verdeeld naar de Sustainable Development Goals bij gebruik van kernwoorden en machine learning.

5. Conclusie

Gezien de grote hoeveelheid werk die verzet wordt door vrijwilligers is het vanuit een statistisch oogpunt interessant om meer informatie te vergaren over deze maatschappelijke sector. Vooral de rapportage naar de Sustainable Development Goals zou een mooie blik geven in de manieren waarop de mensen streven naar een betere wereld. Uit interviews blijkt echter dat op het moment vrijwilligersorganisaties niet of nauwelijks bezig zijn met dataverzameling of -analyse. Vrijwilligers bekommeren zich immers over behoeften en problemen die zij zien in hun omgeving en zien geen noodzaak voor cijfermatige rapportages in.

Het rapporteren van het verrichte vrijwilligerswerk kan echter ook nuttig zijn voor de vrijwilligersorganisaties. Het bijhouden van de prestaties van de organisatie kan een grote hulp zijn ten tijden van subsidie- en financierings-aanvragen. De verstrekkende partijen zijn ook steeds meer op zoek naar manieren om in het oog te krijgen welke impact hun financiering heeft op de samenleving. Daarnaast werkt het ook motiverend voor de betrokken vrijwilligers wanneer zij kunnen inzien hoeveel werk zij tezamen al verricht hebben.

Ook het rapporteren over de indeling naar de Sustainable Development Goals worden weinig toegepast in de maatschappelijke sector. Uit interviews blijkt dat de SDG's beter bekend zijn en meer gebruikt worden naar mate de organisatie op een grotere schaal, bijvoorbeeld (inter-)nationaal, opereert en naar mate financiering en commercie een grotere rol spelen. Lokale vrijwilligersinitiatieven zonder subsidies of geld uit fondsen houden zich dus niet bezig met de SDG's.

Om toch de bijdrage van deze sector aan de SDG's te meten is gekeken naar de gegevens die voor handen zijn, zonder de administratiedruk op de vrijwilligers te vergroten. Als databron is daarom gekeken naar vrijwilligersvacaturebanken, zoals Deedmob. Om uit vacatures informatie te halen over de relevantie tot één of meerdere Sustainable Development Goals is per SDG een lijst met kernwoorden opgesteld om vacatureteksten te verbinden met de SDG’s. Doordat er geen vacatureteksten van een vrijwilligersvacaturebank beschikbaar waren is een dataset met 2365 Nederlandstalige vacatures voor functies bij de Rijksoverheid afkomstig van Werken voor Nederland onderzocht. Op basis van een kernwoordenlijst die handmatig afgestemd is op deze dataset zijn de vacatures ingedeeld bij één of meerdere SDG’s.

Door het toepassen van deze methode kan weergegeven worden welke input er in de vorm van arbeid of vrijwilligerswerk geleverd wordt aan de SDG's. Commerciële methodes gaan vaak een stap verder en passen methodes toe om de impact van een maatschappelijk initiatief te meten. De methode die in dit rapport beschreven wordt heeft echter als voordeel dat deze compleet transparant is, terwijl dit bij commerciële methodes vaak niet het geval is.

In vergelijking met de zelf-rapportages die vaak toegepast worden om input of impact te meten richting de SDG's beperkt de methode in dit rapport de menselijke invloed tot het opstellen van de kernwoordenlijsten. Waar de uitkomsten van zelf-rapportages relatief

gemakkelijk kunnen worden gekleurd door de (vrijwilligers-)organisaties zelf, is dit bij de methode uit dit rapport lastiger. Het is nog steeds mogelijk voor een organisatie om hun bijdrage aan de SDG's rooskleuriger weer te geven dan het in werkelijkheid is, maar zij zullen hiervoor de vacatureteksten grondig aan moeten passen. Dit hindert het uiteindelijke doel van de vacature: het etaleren van de functie om geschikte mensen te werven. Als de organisatie de vacature dusdanig veel aanpast om een hoge bijdrage aan de SDG's te portretteren zodat de vacature niet meer aansluit op de daadwerkelijke functie, dan zullen zij niet meer slagen in het vinden van de juiste mensen. Door de rapportage van de input richting de SDG's te baseren op een bron die een andere functie dient, zal de methode uit dit rapport beter bestand zijn tegen beïnvloeding dan een zelf-rapportage.

5.1 Toekomstig werk

Als alternatief voor indelen van vacatureteksten bij de SDG's door middel van kernwoorden kan men ook handmatig de vacatureteksten indelen bij de Sustainable Development Goals. Deze handmatig ingedeelde teksten kunnen vervolgens met behulp van machine learning gebruikt worden om de SDG’s van nieuwe vacatures mee te bepalen. Het is wel noodzakelijk dat er voldoende teksten per SDG beschikbaar zijn.

Om daadwerkelijk te testen of deze methodes ook werken op vrijwilligersvacatures zal extra onderzoek nodig zijn. De kernwoordenlijsten zullen aangepast moeten worden omdat de tekst van de vrijwilligersvacatures zal verschillen van die van Werken voor Nederland.

Tijdens het genereren en selecteren van de kernwoorden zijn voor sommige kernwoorden vervoegingen opgenomen in de lijst met kernwoorden als deze voorkwamen in de gebruikte bronnen. Bij aanvullend onderzoek kan gekeken worden naar het effect van het toevoegen van alle vervoegingen van de kernwoorden aan de lijst.

Om de subjectiviteit van het handmatig selecteren van de kernwoorden te vermijden kan men experimenteren met methoden om automatisch kernwoorden te selecteren. Het is bijvoorbeeld mogelijk om alle woorden met een RAKE-score boven een drempelwaarde als kernwoorden van de omschrijvingen van de SDG's te nemen. De effectiviteit van deze automatische selectie kan gemeten worden door de kernwoorden te staven aan vooraf handmatig geselecteerde kernwoorden. Als één van de alternatieven voor het RAKE-algoritme kan men kijken naar het TextRank-algoritme [45]. Uit experimenten beschreven in de publicatie van het RAKE-algoritme, ref. [35], blijkt dat het RAKE-algoritme in kortere tijd een betere nauwkeurigheid behaald dan het TextRank-algoritme. Een doorontwikkeling van het RAKE-algoritme is te vinden in de vorm van het 'Totally Automated Keyword Extraction' (TAKE)-algoritme [46]. Wegens de benaderbaarheid van het RAKE-algoritme is dit algoritme eerst toegepast; door de beperkte tijd voor dit onderzoek is de gelegenheid niet geweest om verdere algoritmes, zoals het TAKE-algoritme, te implementeren.

In plaats van het gebruiken van de vacatureteksten kan er ook aanvullend onderzoek gedaan worden naar het vertalen van de gebruikte categorieën van de verschillende vacaturebanken naar de SDG's of naar een gestandaardiseerd systeem.

6. Referentielijst

[1] United Nations, „Sustainable Development Goals," 25 November 2019. [Online]. Available: https://sustainabledevelopment.un.org/?menu=1300.

[2] United Nations, „Background," 25 November 2019. [Online]. Available: https://www.un.org/millenniumgoals/.

[3] SDG Nederland, „Communicatiemateriaal," Juli 2018. [Online]. Available: https://www.sdgnederland.nl/communicatie-toolkit/. [Geopend 8 Januari 2020].

[4] B. Wijnands, Interviewee, *Intakegesprek met Boudewijn Wijnands, Deedmob.* [Interview]. 9 September 2019.

[5] E. Venema, „Deedmob: pragmatisch idealisme übert het vrijwilligerswerk," De Dikke blauwe, 7 Juli 2018. [Online]. Available: https://www.dedikkeblauwe.nl/news/deedmob-ubert-vrijwilligerswerk. [Geopend 13 December 2019].

[6] NLvoorelkaar, „Vrijwilligerswerk & vrijwilligersvacatures in de buurt," [Online]. Available: https://www.nlvoorelkaar.nl. [Geopend 13 December 2019].

[7] IkWilVrijwilligerswerkDoen.nl, „Vrijwilligerswerk," [Online]. Available: https://www.ikwilvrijwilligerswerkdoen.nl. [Geopend 13 December 2019].

[8] Vrijwilligerscentrale Utrecht, „Vacatures," [Online]. Available: https://vcutrecht.nl/vacatures. [Geopend 13 December 2019].

[9] Vrijwilligerscentrale Amsterdam, „Vrijwilligerswerk vacatures in Amsterdam," [Online]. Available: https://vacaturebank.vca.nu/. [Geopend 13 December 2019].

[10] Centraal Bureau voor de Statistiek, „StatLine: 82249NED - Sociale contacten en maatschappelijke participatie," 26 Maart 2019. [Online]. Available: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82249NED/table?ts=1578321947252. [Geopend 6 Januari 2020].

[11] Ensie, „Vrijwilligerswerk - de betekenis volgens CBS," 30 December 2016. [Online]. Available: https://www.ensie.nl/cbs/vrijwilligerswerk. [Geopend 13 Januari 2020].

[12] Gemeente Den Haag, „Den Haag in Cijfers - % Hagenaars dat in het afgelopen jaar wel eens vrijwilligerswerk heeft gedaan 2018," [Online]. Available: https://denhaag.incijfers.nl/Jive?workspace_guid=c764d6bf-afe2-4bbe-aa69-ce6fba97a296. [Geopend 7 Januari 2020].

[13] WistUdata, „Gemeente Utrecht - Vrijwilligerswerk % 2017 - Subwijken," [Online]. Available: https://wistudata.nl//jive?workspace_guid=eae15f4c-0c0f-4ec3-a5b5-cbd07cebbe4c. [Geopend 7 Januari 2020].

[14] Deedmob, „Discover volunteering near you," 25 November 2019. [Online]. Available: https://www.deedmob.com/.

[15] Den Haag Doet, „Vrijwilligerswerk in Den Haag," [Online]. Available: https://denhaagdoet.nl/vacaturebank. [Geopend 16 December 2019].

[16] R. Millar en K. Hall, „Social Return on Investment (SROI) and Performance Measurement: The opportunities and barriers for social enterprises in health and social care," *Public Management Review,* vol. 15, nr. 6, pp. 923-942, 2013.

[17] Sinzer, „Sinzer, Impact Meten," [Online]. Available: https://www.sinzer.org/. [Geopend 13 Januari 2020].

[18] Cigarbox, „Cigarbox," [Online]. Available: http://www.cigarbox.nl/. [Geopend 13 Januari 2020].

[19] MAEX, „MAEX," [Online]. Available: https://maex.nl. [Geopend 13 Januari 2020].

[20] VSBFonds, „Meer informatie over de Participatieladder," [Online]. Available: https://www.vsbfonds.nl/mens-maatschappij/de-participatieladder. [Geopend 13 Januari 2020].

[21] Participatiehulp, „Participatiematrix," Enspiratie, [Online]. Available: http://www.participatiehulp.nl/participatiematrix-1. [Geopend 13 Januari 2020].

[22] Zelfredzaamheid-Matrix, „Zelfredzaamheid-Matrix," GGD Amsterdam, [Online]. Available: https://www.zelfredzaamheidmatrix.nl/. [Geopend 13 Januari 2020].

[23] MijnPositieveGezondheid.nl, „Home," iPH.nl, 2020. [Online]. Available: https://mijnpositievegezondheid.nl/. [Geopend 13 Januari 2020].

[24] W. Bowman, „The Economic Value of Volunteers to Nonprofit Organisations," *Nonprofit Management & Leadership,* vol. 19, nr. 4, pp. 491-506, 2009.

[25] United Nations, „Work of the Statistical Commission Pertaining to the 2030 Agenda for Sustainable Development (A/RES/71/313)," in *UN General Assembly*, New York, NY, USA, 2017.

[26] Centraal Bureau voor de Statistiek, „Monitor Brede Welvaart 2018," CBS, Den Haag/Heerlen/Bonaire, 2018.

[27] Centraal Bureau voor de Statistiek, „Monitor Brede Welvaart & Sustainable Development Goals 2019," CBS, Den Haag/Heerlen/Bonaire, 2019.

[28] Python Software Foundation, „The Python Language Reference - Python 3.6.10 documentation," [Online]. Available: https://docs.python.org/3.6/reference/index.html. [Geopend 8 Januari 2020].

[29] Rijksoverheid, „Werken voor Nederland," 25 November 2019. [Online]. Available: https://www.werkenvoornederland.nl/.

[30] M. ". Danilak, „PyPI - langdetect 1.0.7," 3 Oktober 2016. [Online]. Available: https://pypi.org/project/langdetect/. [Geopend 2 Februari 2020].

[31] Wikipedia, „Duurzame Ontwikkelingsdoelstellingen," Wikimedia, 9 September 2019. [Online]. Available: https://nl.wikipedia.org/wiki/Duurzame_Ontwikkelingsdoelstellingen. [Geopend 23 September 2019].

[32] Verenigde Naties, „Duurzame Ontwikkelingsdoelstellingen," VN Regionale Informatiecentrum, [Online]. Available: https://unric.org/nl/duurzame-ontwikkelingsdoelstellingen. [Geopend 23 September 2019].

[33] SDG Nederland, „Alles over de Duurzame Ontwikkelingsdoelen," [Online]. Available: https://www.sdgnederland.nl/. [Geopend 3 Oktober 2019].

[34] PwC, „Toepassing van SDG's door Nederlandse organisaties, tweede editie," 2019.

[35] S. Rose, D. Engel, N. Cramer en W. Cowley, „Automatic Keyword Extraction from Individual Documents," in *Text Mining: Applications and Theory*, Wiley, 2010.

[36] V. B. Sharma, „PyPI - rake-nltk 1.0.4," [Online]. Available: https://pypi.org/project/rake-nltk/. [Geopend 8 Januari 2020].

[37] C. Fellbaum, WordNet: An Electronic Lexical Database, Cambridge, MA: MIT Press, 1998.

[38] M. Postma, E. van Miltenburg, R. Segers, A. Schoen en P. Vossen, „Open Dutch WordNet," in *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania, 2016.

[39] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado en J. Dean, „Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems,* pp. 3111-3119, 2013.

[40] Princeton University, „WordNet - A Lexical Database for English," [Online]. Available: https://wordnet.princeton.edu. [Geopend 30 Januari 2020].

[41] S. Tulkens, C. Emmery en W. Daelemans, „Evaluating unsupervised Dutch word embeddings as a linguistic resource," *arXiv preprint arXiv:1607.00225,* 2016.

[42] Instituut voor de Nederlandse Taal, „SoNaR-corpus," 12 Maart 2018. [Online]. Available: https://ivdnt.org/downloads/taalmaterialen/tstc-sonar-corpus. [Geopend 30 Januari 2020].

[43] R. Rehurek en P. Sojka, „Software framework for topic modelling with large corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, 2010.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot en E. Duchesnay, „Scikit-learn: Machine learning in Python," *Journal of machine learning research,* vol. 12, pp. 2825-2830, 2011.

[45] R. Mihalcea en P. Tarau, „Textrank: Bringing order into text," in *Proceedings of the 2004 conference on empirical methods in Natural Language Processing*, 2004.

[46] T. Pay, „Totally automated keyword extraction," in *2016 IEEE international conference on big data*, 2016.

7. Bijlage 1. Gevonden kernwoorden in de vacatures

SDG	Kernwoorden
1	'arbeid', 'armen', 'armoede', 'armoedebestrijding', 'bestaansminimum', 'bestaanszekerheid', 'betaalbare', 'crisis', 'economieën', 'economische', 'financiële', 'financiële middelen', 'inkomen', 'inkomenszekerheid', 'kinderopvangtoeslag', 'ontwikkelingsdoelstellingen', 'ontwikkelingslanden', 'oudedagsvoorziening', 'sociale zekerheid', 'uitsluiting'
2	'boeren', 'boerenbedrijf', 'dierenwelzijn', 'duurzame landbouw', 'eten', 'geiten', 'jagen', 'koken', 'landbouw', 'landbouwontwikkeling', 'landbouwproductiviteit', 'landerijen', 'levensmiddelen', 'maaltijden', 'mondiale voedselvoorziening', 'os', 'overgewicht', 'planten', 'platteland', 'pluimvee', 'schapen', 'teelt', 'tuinbouw', 'tuinderij', 'varkens', 'vee', 'veehouder', 'veehouderij', 'veeteelt', 'vis', 'visserij', 'vissers', 'vlees', 'voeden', 'voeder', 'voeding', 'voedingsmiddelen', 'voedsel', 'voedselkwaliteit', 'voedselproductie', 'voedselveiligheid', 'voedselvergiftiging', 'voedselverspilling', 'voedselvoorziening', 'voedselzekerheid', 'zuivel'
3	'aandoeningen', 'aids', 'alcohol', 'alcoholgebruik', 'antibiotica', 'antistoffen', 'apotheek', 'besmetting', 'betaalbare zorg', 'chemicaliën', 'drugsverslaving', 'duurzame ontwikkeling', 'epidemieën', 'epidemiologie', 'ernstige infectieziekten', 'euthanasie', 'geboorte', 'geneesmiddel', 'geneesmiddelen', 'gevaccineerd', 'gewonde', 'gewonden', 'gezondere', 'gezondheid', 'gezondheidsorganisatie', 'gezondheidsrisico', 'gezondheidstoestand', 'gezondheidszorg', 'gezondheidzorg', 'goede gezondheid', 'hiv', 'huisartsenzorg', 'infecties', 'infectieziekten', 'intensive care', 'ketenzorg', 'levensduur', 'medicatie', 'medicijn', 'medicijnen', 'milieuhygiëne', 'nederlandse zorgstelsel', 'neurologie', 'orthopedie', 'ouderenzorg', 'overgewicht', 'pasgeborenen', 'pathologie', 'patiënten', 'persoonlijkheidsstoornissen', 'psychiatrie', 'psychopathologie', 'reproductieve gezondheid', 'resistentie', 'roken', 'serologie', 'tbc', 'therapie', 'traumatologie', 'vaccineren', 'vaccins', 'verkeersveiligheid', 'verpleegkundigen', 'verpleging', 'verslaving', 'verzorging', 'virologie', 'volksgezondheid', 'welzijn', 'ziektelast', 'ziekten', 'ziektes', 'zorgaanbod', 'zorgbeleid', 'zorgsector', 'zorgverleners', 'zorgverlening', 'zwangerschap', 'zwangerschappen'
4	'basisonderwijs', 'beroepsonderwijs', 'beroepsvaardigheden', 'bijscholing', 'cursus', 'didactiek', 'docenten', 'doceren', 'educatie', 'faculteit', 'goed onderwijs', 'hoger onderwijs', 'kennisintensieve', 'lang leren', 'leerkrachten', 'leerlingen', 'leermiddelen', 'leerwegondersteunend', 'leraren', 'les', 'nascholing', 'onderwijs', 'onderwijsinspectie', 'onderwijsinstelling', 'onderwijzen', 'opleiding', 'primair

	onderwijs', 'scholing', 'school', 'studerenden', 'universiteit', 'vervolgopleidingen', 'wetenschappen'
5	'dochter', 'emancipatie', 'gelijke behandeling', 'gelijke rechten', 'gender', 'gendergelijkheid', 'meisjes', 'mensenhandel', 'moeder', 'reproductieve gezondheid', 'seksueel geweld', 'sociale veiligheid', 'vrouw', 'vrouwelijke', 'vrouwenrechten'
6	'afvalwater', 'binnenwateren', 'drinkwater', 'drinkwaterbedrijven', 'drinkwatervoorziening', 'grondwater', 'hemelwater', 'hygiëne', 'kaderrichtlijn water', 'koelwater', 'oppervlaktewater', 'riolering', 'sanitair', 'schoon water', 'toilet', 'verzilting', 'water', 'waterbeheer', 'waterbeleid', 'waterhuishouding', 'waterkwaliteit', 'watertekort', 'zoetwater', 'zoetwaterbronnen', 'zuivering', 'zwemwater'
7	'aandeel hernieuwbare', 'aardgas', 'aardolie', 'aardwarmte', 'benzine', 'biogas', 'biomassa', 'brandstof', 'brandstoffen', 'broeikasgassen', 'duurzame energie', 'elektriciteit', 'energie', 'energiebesparing', 'energiebron', 'energiebronnen', 'energieopwekking', 'energieverbruik', 'energiezuinige', 'gas', 'gaswinning', 'hernieuwbare', 'hernieuwbare energie', 'olie', 'opwarming van de aarde', 'stroom', 'verwarming', 'warmte', 'warmtekrachtkoppeling', 'waterstof', 'windenergie', 'windmolenparken', 'windmolens', 'windturbines'
8	'arbeid', 'arbeids', 'arbeidsmarkt', 'arbeidsomstandigheden', 'arbeidsparticipatie', 'arbeidsplaatsen', 'arbeidsverhoudingen', 'banen', 'beroepsbevolking', 'dienstverband', 'duurzame economische', 'economische', 'economische groei', 'economische zaken', 'mensenhandel', 'mensensmokkel', 'ondernemend', 'ondernemerschap', 'onderneming', 'productiviteit', 'tewerkstelling', 'vacatures', 'vrije tijd', 'werkloosheid', 'werklozen', 'werknemer', 'werknemers', 'werkomstandigheden', 'werkzaamheid'
9	'betere infrastructuur', 'biomedisch', 'biotechnologie', 'bus', 'cargo', 'communicatietechnologie', 'economische ontwikkeling', 'elektrotechniek', 'goederentransport', 'hightech', 'ict', 'industrie', 'industriële', 'infrastructuur', 'innovatie', 'innovaties', 'innovatieve', 'investeringen', 'metro', 'mkb', 'moderniseren', 'mvo', 'nederlandse industrie', 'nederlandse mkb', 'nieuwe technologieën', 'onderzoek', 'openbaar vervoer', 'personenvervoer', 'promotieonderzoek', 'research', 'taxi', 'technologie', 'technologieën', 'technologische', 'toegevoegde waarde', 'tram', 'trein', 'treinen', 'vervoer', 'vooruitgang', 'vrachtverkeer', 'wegtransport', 'wetenschap', 'wetenschappelijk', 'wetenschappelijk onderzoek', 'zonder wegen'
10	'belastingen', 'discriminatie', 'gediscrimineerd', 'gelijke kansen', 'subsidies', 'uitsluiting', 'vergrijzing'
11	'betaalbare woningen', 'dorpen', 'gemeentelijke', 'hoofdstad', 'huurmarkt', 'huurwoningen', 'leefbaarheid', 'leefklimaat', 'leefomgeving', 'openbaar vervoer', 'publieke gezondheid', 'stad', 'stedelijke', 'stedelijke groei', 'steden', 'vastgoedmarkt', 'veilige leefomgeving', 'verstedelijking', 'voorzieningen', 'werklocatie', 'woning', 'woningbouw', 'woningmarkt', 'woonplek', 'woonruimte'
12	'afval', 'afvalstoffen', 'afvalwater', 'circulaire economie', 'consumptie', 'duurzaam toerisme', 'duurzame productie', 'gevaarlijk afval', 'grondstoffen', 'hergebruik', 'levenscyclus', 'natuurlijke hulpbronnen', 'vervuiler', 'voedselproductie', 'voedselverspilling'
13	'broeikasgassen', 'co', 'de opwarming', 'emissies', 'energieakkoord', 'energiebeleid', 'klimaat', 'klimaatactie', 'klimaatbeleid', 'klimaatconferentie',

	'klimaatverandering', 'klimaatveranderingen', 'milieubeleid', 'milieuvriendelijke', 'natuurrampen', 'opwarming', 'overstromingen', 'parijsakkoord', 'stijgende zeespiegel', 'stikstofoxiden', 'temperatuur', 'uitstoot'
14	'kust', 'kustwateren', 'mariene', 'mariene wateren', 'noordzee', 'oceanen', 'verzilting', 'visserij', 'waddenzee', 'wateren', 'zeevisserij', 'zeeën'
15	'biodiversiteit', 'bodem', 'bomen', 'bos', 'dieren', 'diersoorten', 'droogte', 'ecosysteem', 'ecosystemen', 'fauna', 'insecten', 'nationale parken', 'natuur', 'natuurwaarden', 'planten', 'plantensoorten', 'vegetatie', 'woud'
16	'aanbestedingsregels', 'aantal gedetineerden', 'belastingontduiking', 'conflicten', 'corruptie', 'criminaliteit', 'cyber security', 'cybersecurity', 'defensie', 'delicten', 'democratie', 'democratisch', 'democratische', 'diefstal', 'doodslag', 'extremisme', 'fraude', 'gedetineerden', 'geldigheid', 'geweld', 'huiselijk geweld', 'instellingen', 'justitie', 'kartelvorming', 'mededinging', 'mensenhandel', 'mensenrechten', 'misdaad', 'misdrijf', 'misdrijven', 'moord', 'nationale veiligheid', 'ombudsman', 'onveiligheid', 'openbare orde', 'ordeverstoring', 'overheden', 'overheidsdiensten', 'overheidsinstellingen', 'parlement', 'politie', 'publieke dienstverlening', 'rechtbank', 'rechtmatigheid', 'rechtsorde', 'rechtspraak', 'rechtsstaat', 'rechtstaat', 'rechtszaak', 'rechtvaardigheid', 'regelgeving', 'terreur', 'terrorisme', 'veilige samenleving', 'veiligheidszorg', 'vervolging', 'vrede', 'vreedzame', 'wetgeving', 'zorginstellingen'
17	'buitenlandse zaken', 'duurzame ontwikkeling', 'internationale samenwerking', 'kennisuitwisseling', 'ontwikkelingsagenda', 'ontwikkelingsdoelen', 'ontwikkelingslanden', 'ontwikkelingssamenwerking', 'partnerschap', 'partnerschappen', 'partnership', 'partnerships', 'uitwisselingen', 'verenigde naties'

Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopig cijfer
**	Nader voorlopige cijfer
2017–2018	2017 tot en met 2018
2017/2018	Het gemiddelde over de jaren 2017 tot en met 2018
2017/’18	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2017 en eindigend in 2018
2013/’14–2017/’18	Oogstjaar, boekjaar, enz., 2015/’16 tot en met 2017/’18

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

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

Prepress
CCN Creatie en visualisatie

Ontwerp
Edenspiekermann

Inlichtingen
Tel. 088 570 70 70, fax 070 337 59 94
Via contactformulier: www.cbs.nl/infoservice

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2018.
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.