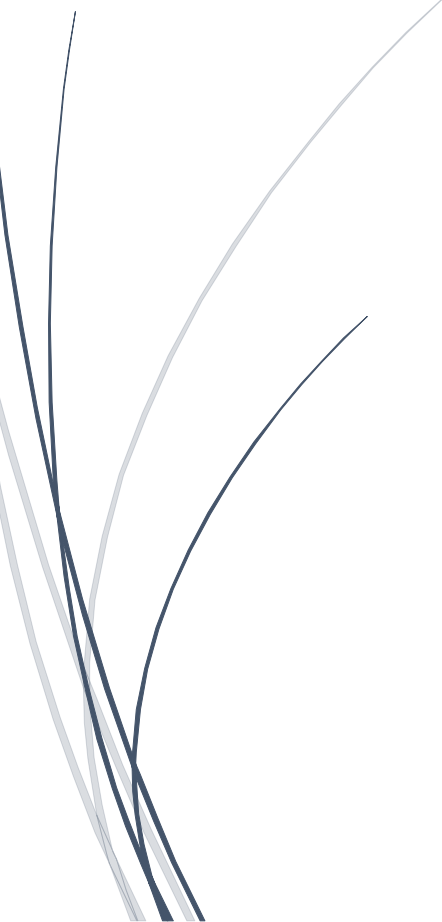




14-4-2022

Extended Abstract

Digital Readiness lokaliseren bij
Chemelot



Randy van Breugel 1620126

Stage Lectoraat Data Intelligence

Samenvatting

Duurzaamheid en circulariteit gaan hand in hand. Maar deze concepten zijn hedendaags ver te zoeken. De Sustainable Development Goals gelden nu als een groot struikelblok voor de maakindustrie. Het lectoraat Data Intelligence heeft besloten om projecten te initiëren die maatschappelijke problemen zoals beschreven te helpen oplossen. Een mogelijk oplossing voor dit probleem zou AI kunnen zijn. Om dit te kunnen testen, zal de huidige situatie eerst in kaart gebracht moeten worden. Dit wordt gedaan door een webapplicatie die gerealiseerd is. Deze applicatie meet op basis van DMI (Data-driven Management Index) hoe ver bedrijven staan op het gebied van Digital Readiness. Om een meer accurate meting te kunnen doen, is er onderzoek gedaan naar een NLP model om de antwoorden te classificeren in de juiste trap van de index.

Aanleiding

De maakindustrie is altijd al een innovatieve en vooruitstrevende bron geweest voor omgang met nieuwe uitdagingen en trends. Denk bijvoorbeeld aan de eerste industriële revolutie, maar meer recent ook de digitale revolutie (Thiede, 2020). Daarnaast zorgen ook verduurzamingsdoelen zoals de Sustainable Development Goals (SDGs) (Sustainability Development Goals, sd) en de opkomende 'circulaire economie' voor de nodige uitdagingen in deze sector. Daarom is er een flexibel, schaalbaar en nieuw productieproces nodig om met deze toekomstige ontwikkelingen om te kunnen gaan. (Thiede, 2020)

Het lectoraat Data Intelligence van Zuyd Hogeschool heeft besloten om projecten te initiëren op dit snijvlak tussen slimme maakindustrie, circulariteit en data. Het doel is om bedrijven die zijn gevestigd op de Brightlands Chemelot Campus te helpen met de grote maatschappelijke opgaven van deze tijd.

Het lectoraat wil onderzoeken op welke manieren Artificial Intelligence (AI) een oplossing kan bieden voor de ontwikkeling van nieuwe circulaire productieprocessen. Binnen het brede domein van AI wordt er specifiek gekeken naar concepten zoals 'predictive maintenance' (voorspellen aan de hand van machine-learning modellen wanneer een machine onderhoud nodig heeft), 'resource-management', 'supply-chain management' op basis van technieken zoals 'deep learning' of 'neural networks'. (Luis Romeral Martínez, 2020) (Jamwal, 2021)

Omdat het lectoraat op dit vlak in de opstartfase zit is het noodzakelijk om eerst in kaart te brengen wat de huidige situatie is bij de bedrijven op de Chemelot campus. Daarnaast zal het lectoraat een gewenste situatie schetsen en onderzoeken of deze ook daadwerkelijk past bij de praktijk.

Doelstelling

Het doel van het lectoraat is om op basis van data voorspellende, slimme en structurele informatie te komen tot inzicht in de manier waarop productiesystemen duurzaam en circulair zijn. De doelstelling van het project Digital Readiness lokaliseren bij Chemelot is echter om de huidige situatie in kaart brengen bij lokale bedrijven uit de maakindustrie.

- Hoe wordt data verzameld?
- Hoeveel en welke data wordt verzameld?
- Hoe staat het bedrijf erin om eventueel een AI-toepassing te gaan gebruiken?
- Is er al een hoofdsysteem of bestaan er allemaal losse systemen?

Om antwoorden op deze vragen te krijgen zal er data worden verzameld door middel van een enquête. De vraag is tevens of de data uit deze enquêtes op een automatische manier te verwerken is zodat er een tussenresultaat uit gehaald kan worden voor het vervolg van het project.

Theoretisch kader

Digital readiness is het concept dat als fundament dient voor deze stageopdracht. Er is veel geschreven over digital readiness, maar in brede zin kunnen we stellen dat het gaat om het niveau van gereedheid van een organisatie om over te stappen naar gedigitaliseerde workflows die mogelijk worden gemaakt door software en technologie. (Oden Technologies, sd) Deze gereedheid van organisaties kan op verschillende manieren worden gemeten wordt per industrie verschillend opgepakt. (Hermann Lödding)

Digital readiness is dus een meetmaat die beschrijft of bedrijven/organisaties klaar zijn voor een digitale transformatie (Industrie 4.0). Elke meetmaat heeft andere criteria behorend bij het doel.

Oden Technologies beschrijft dat hun model bestaat uit drie onderdelen: Cultuur, processen en technologie. Elk onderdeel bevat een assesment met vragen met betrekking tot het onderwerp. Van alle drie de componenten samen wordt een oordeel geveld. (Oden Technologies, sd)

Anna De Carolis schrijft in haar rapport dat digital readiness beoordeeld kan worden aan de hand van een maturity model op basis van "maturity levels", beschreven en ontwikkeld door haarzelf. Deze levels duiden aan hoe "volwassen" het bedrijf is en schets een huidige situatie op het vlak van digitaliseren. (Hermann Lödding)

Binnen het lectoraat zijn er andere gedachten namelijk: een KPI-gedreven index die functioneert aan de hand van een trappensysteem. Dit systeem sluit meer aan bij de doelstelling.

Deze index (Data-driven Management Index) bestaat uit een zestal trappen:

- Trap 1: KPI's goed in beeld (1 kpi is geen duidelijk beeld)
- Trap 2: Er is data (sensoren aanwezig) + KPI
- Trap 3: Invloed van data op KPI's is duidelijk (Beeld welke sensoren/informatiebronnen beschikbaar zijn en welke invloed hebben op kpi's)
- Trap 4: controlemechanisme op vertraagde data
 - Trap 4.5: met strategische voorspellingen
- Trap 5: controlemechanisme op realtime data
- Trap 6: strategische beslissingen op basis van voorspellingen

De DMI geeft weer hoe ver bedrijven zijn op het gebied van data-gedreven management op basis van hun beschreven KPI's.

Om een bedrijf te classificeren in de juiste trap, is er contact gezocht met de bedrijven. Aangezien interviews te gedetailleerd en te tijd/geld-rovend zijn en enquêtes te oppervlakkig en niet tot de kern van de essentie komen, is er gekozen voor een semigestructureerde vragenlijst. Hiermee is de essentie te vinden in een handbereik.

Het struikelblok van deze manier is dat er meerdere vormen van vragen zijn. Open vragen wordt door elke deelnemer uit de doelgroep anders ingevuld en zal daarom ook van invloed zijn op het resultaat.

Methode

Dit project is volgens de stappen van de Design-Science-Research (DSR) methode uitgevoerd. DSR richt zich op de ontwikkeling en prestaties van (ontworpen) producten/programma's met de expliciete bedoeling om de functionele prestaties van het producten/programma's te verbeteren. (Hevner) Doorheen dit project zijn de relevance en rigor cycle cycli op iteratieve wijze doorlopen.

In de relevance cycle is er gekeken waaraan de webapplicatie moet voldoen volgens de wensen/eisen van de opdrachtgever. Om dit in kaart te brengen, is er gebruikt gemaakt van een elicitatieproces met de opdrachtgever aan de hand van interviews. Deze interviews zijn doormiddel van vrije gesprekken tot stand gekomen. Die hebben geleid tot een merendeel van de eisen. Aan de hand van een stuk grounding (uitwerken van een data-mineproces zodat vele gebruikers er toegang tot hebben, verwerken van opgehaalde data, data indelen in de een gegeven maat van het lectoraat), zijn er meer functionele en ontwerpeisen bijgekomen.

Tevens is onderzoek gedaan naar de manieren waarop de uiteindelijke oplossing technisch uitgebeeld zou kunnen worden. Aan de hand van deskresearch (literatuur-quickscan) en expert-gesprekken kwam er uiteindelijk een beeld bij hoe de applicatie zou kunnen functioneren. De quickscan is uitgevoerd van een aantal zoektermen. De zoektermen die hiervoor gebruikt werden zijn:

- NLP
- NLP-methode
- Tekst classificatie
- Tekst summarization
- RNN/CNN

Deze informatie is continue afgestemd met de opdrachtgever zodat het project binnen de projectgrenzen bleef.

Op basis van de resultaten is er in de ontwerpcyclus een ontwerp gemaakt en gerealiseerd. Dit ontwerp is gemodelleerd met behulp van UML (flow-chart, use-case diagrammen, sequentiediagram) en met regelmaat bekeken door verschillende experts ter validatie. In overleg met opdrachtgever en het eisenpakket als leidend criteria is het ontwerp goedgekeurd en gerealiseerd.

Resultaten

Om de Data-driven Management Index bij bedrijven te meten zal er een heel proces doorlopen worden. Van het data verzamelen bij het bedrijf tot aan het verwerken van de data. Het proof of concept maakt inzichtelijk hoe dit mogelijk is.

Requirements

De opdrachtgever wou een online omgeving creëren waarbij hij de door het lectoraat ontwikkelde digital readiness maat (DMI) kon meten. Om dit te kunnen meten, is er een semi-structureerde vragenlijst ontwikkeld. De online omgeving moest input automatisch verwerken en de output ervan visualiseren. Het design moest modern zijn, aangezien de webapplicatie het visitekaartje is van het hele project. Aan de hand van een quick-scan werd duidelijk dat de open vragen die bedacht zijn, verwerkt kunnen worden met NLP.

Onderzoek

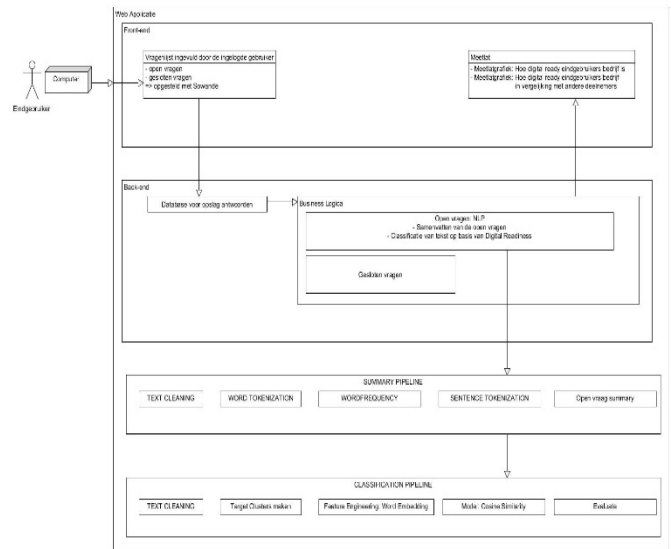
Om binnen de scope van het project te blijven, is er deskresearch verricht. Aangezien er in de applicatie gewerkt wordt met gepersonaliseerde tekst en deze automatisch verwerkt zal moeten worden, is er gekeken of Natural Language Processing van toepassing kon zijn. Om deze techniek toe te passen, was er wel een uitdaging. Er was trainingsdata nodig om NLP toe te passen.

Het eerste onderzoek resulteerde in een richting op het gebied van Convolutionele Neurale Netwerken en Recurrent Neurale Netwerken. Recurrent Neurale Netwerken zijn netwerken die sequentiele informatie te interpreteren. Ze doen dit door steeds terug te keren naar eerdere mogelijke datapunten zodoende ze betere voorspellingen kunnen maken. Terwijl Convolutionele Neurale netwerken bestaan uit lagen die data transformeert aan de hand van filters. Als de input is getransformeerd dan gaat hij pas verder naar de volgende laag. Deze methoden hebben veel tijd en data nodig om optimaal te werken. Dit valt daarom buiten de scope.

Door het onderzoek te verschuiven naar tekst classification kwam er een tweede struikelblok. Tekst classification kan in twee vormen gebeuren namelijk: supervised learning (target variabel + trainingsdata) en unsupervised learning (geen target variabel en geen trainingsdata). Door een soort van semi-supervised learning methode te creëren (target variabele, maar zonder trainingsdata), is er een manier gevonden op basis van cosine similarity. Deze similarity bepaalt de afstand tussen twee nul-vectoren. Door woorden te transformereren naar vectoren, kan er een afstand bepaalt worden. Met de techniek kan er in korte tijd en geen trainingsdata, input verwerkt worden en geplaatst worden in een juiste cluster. De deployment van dit model viel buiten de scope van de opdracht.

Ontwerp

Om tot dit PoC (Figuur 1 Ontwerp flow) te komen, moest er eerst een basis gelegd worden. Aan de hand van elicitatieproces bij de opdrachtgever en grounding (deskresearch) is er een ontwerp ontstaan. Dit ontwerp laat zien dat het bedrijf eerst een semi-gestructuurde vragenlijst invult in de webapplicatie. Deze antwoorden worden opgeslagen in een database in afwachting tot verdere verwerking. De verwerking gaat als volgt. De openvragen gaan doorheen een NLP-model (tekst-classificatie) zodat deze geplaatst worden in de juiste cluster (trappen van de DMI). De gesloten vragen worden meegenomen met het resultaat. Tot slot zal er een beeld weergegeven worden waarin aangetoond wordt in welke positie het bedrijf zich bevindt.



Figuur 1 Ontwerp flow

Webapplicatie

De webapplicatie is ontwikkeld aan de hand van het Anvil Framework. Met dit framework kunnen er full-stack en in productie genomen webapplicatie gerealiseerd worden. Anvil werkt aan de hand van een soort MVC-structuur. Het Model/View/Server houdt in dat er drie lagen zijn.

De eerste laag is de serverlaag. Op deze laag worden alle handelingen vanuit de database richting de model/view geregeld. Handelingen betreft deze laag in de applicatie zijn onder andere:

- Ophalen van gegevens uit de database
- Users toevoegen
- Antwoorden toevoegen

Bedrijven kunnen zich registreren zodat ze de vragenlijst kunnen invullen en zij enkel hun resultaten kunnen zien. Bedrijven die zich niet geregistreerd hebben, kunnen de vragenlijst niet invullen en geen resultaten van andere bedrijven zien.

De DMI wordt op dit moment bepaald aan de hand van de gesloten vragen aangezien er geen data beschikbaar is. Dus op basis van of bepaalde vragen Ja/Nee geassocieerd worden, wordt er een trap aangewezen en wordt deze trap uitgelegd.

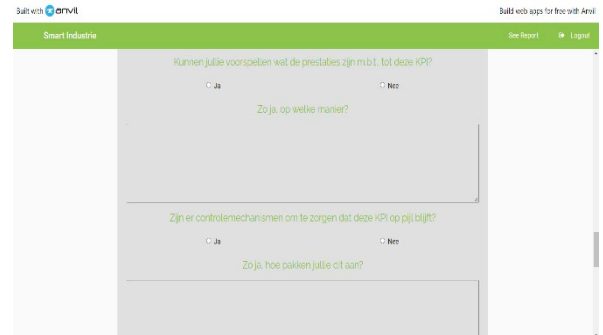
De tweede laag is de Model/Viewlaag. Deze laag bevat twee onderdelen aangezien er ook een stuk logica gedaan kan worden op de viewpage. In deze laag worden de form laten zien. De forms die de webapplicatie heeft is:

- Homepagina
- Surveypagina
- Reportpagina
- Formpagina

De Homepagina is de pagina waar de gebruikers binnen komen als ze op de website binnenkomen (<https://vragenlijst-dataintelligence.anvil.app/>). Hier kunnen gebruikers zich registreren/inloggen. Als de gebruiker ingelogd is, kan hij/zij beginnen met de vragenlijst.

De surveypagina (Figuur 2 Surveypagina) is de pagina waar de vragenlijst is gevisualiseerd. Op deze pagina bevindt zich ook de logica om de data op te slaan richting de database. Deze view bestaat uit een aantal componenten:

- Input (radiobutton's, textarea's)
- Form (opmaak)
- Columnpanels (onderscheid onderdelen van vragenlijst)

The screenshot shows a web application interface for a survey. At the top, there is a green header with the text 'Smart Industrie' and 'Over Report' and 'Log out'. The main content area contains three questions, each with two radio button options: 'Ja' and 'Nez'. The questions are: 1. 'Kunnen jullie voorspellen wat de prestaties zijn m.b.t. tot deze KPI?' with a sub-question 'Zo ja, op welke manier?'. 2. 'Zijn er controlemechanismen om te zorgen dat deze KPI op pijl blijft?' with a sub-question 'Zo ja, hoe pakken jullie dit aan?'. 3. A third question is partially visible at the bottom.

Figuur 2 Surveypagina

De reportpagina (Figuur 3 Reportpagina) is de pagina waar alle data wordt weergegeven. De index van de ingelogde gebruiker wordt weergegeven, maar tevens de index van andere bedrijven op anonieme basis. Daarnaast worden er nog gedetailleerde data weergegeven met betrekking welke KPI's bedrijven beschreven hebben, in welke sector ze bevinden en hoe groot het bedrijf is. De reportpagina bestaat uit een aantal onderdelen:

- Datagrids (weergeven data uit de database)
- Plots (visualiseren van de data)
- Form (opmaak)



Figuur 3 Reportpagina

De formpagina is de opmaakpagina die geen functionaliteit opzich heeft. Deze pagina fungeert als leidraad voor de opmaak doorheen de hele applicatie. De onderdelen dat deze pagina bat zijn:

- Navigatie
- Kleuren
- Foto's
- Titels

Semi-supervised tekst classification

De vragen die beantwoord worden in de webapplicatie bestaan uit twee soorten vragen. Open en gesloten vragen. De DMI wordt op dit moment bepaald aan de hand van de gesloten vragen. Dit komt door het feit dat het NLP-model(Figuur 3 workflow semi-supervised text classification) door ontwikkeld moet worden. Dit valt buiten de scope van de opdracht.

Om deze antwoorden te classificeren in de juiste trap (target cluster) is er gebruik gemaakt van semi-supervised tekst classification op basis van cosine similarity. Dit komt door het feit dat er geen trainingsdata beschikbaar is, maar er wel target variabelen zijn. Er is gekozen voor cosine, aangezien woord duplicatie belangrijk is. Dit resulteert dat met bijvoorbeeld de Jaccard maat de uitkomst continu blijft wanneer er bijvoorbeeld 50 keer hetzelfde woord geschreven. (Different similarity measures, sd) Dit moet voorkomen worden.

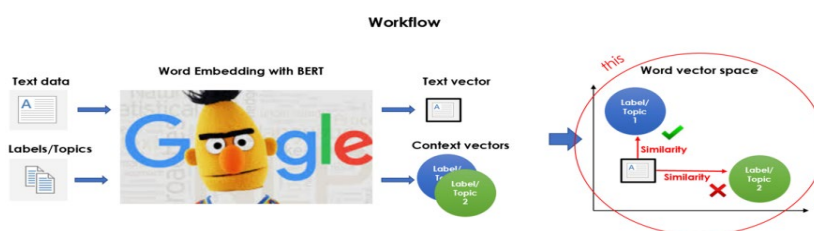
Het model bestaat uit een 7 targetvariabelen. De 7 trappen van de DMI staan beschreven aan de hand van hun kenmerken. Door de input te vergelijken met de kenmerken van de trappen kan er

voorspelling gemaakt worden. Voordat de input klaar is voor een vergelijking, wordt de tekst eerst gecleaned. Dit houdt in dat leestekens, hoofdletters, veel voorkomende woorden (Stopwords) en rare symbolen eruit worden gehaald zodat er enkel de essentie overblijft.

De voorspelling wordt gemaakt aan de hand van twee technieken. Het W2Vec methode houdt in dat de woorden omgevormd worden naar vectoren op basis van een getraind model (Nederlandstalige Fasttext-model). (Vector models, sd) Dit is de basis om de targetclusters te kunnen maken. Tevens worden met de functie `get_similar_words()` vergelijkbare woorden gevonden van de kenmerken van de target variabelen.

De tweede techniek is Word Embedding. Deze techniek houdt in dat de gecleande woorden een reële vector waarden krijgen. Dit wordt gedaan aan de hand van een voorgetraind model BERTje. BERTje is een getraind NLP model van de Universiteit van Groningen die bestaat uit een dataset van 2.4 biljoen tokens. (Wietse de Vries) (Vries)

Al deze vectoren worden in de word vector space geplaatst. In deze space wordt de similarity bepaald. Deze waarde bevindt zich tussen de 0 en de 1. Hoe dichterbij de 1, hoe groter de kans is dat het woord/zin bij die target cluster hoort.



Figuur 4 workflow semi-supervised text classification

Discussie

In het begin was er nog al veel onduidelijkheid omtrent de oplossing. Zou dit een oplossing worden die universeel toegepast zou kunnen worden voor meerdere data-mineprojecten of enkel voor dit project? Door de focus eerst te leggen op een universeel product en het niet te valideren met opdrachtgever dreigde dit project buiten de projectgrenzen te komen. Door een gesprek met de opdrachtgever kon tijdig de juiste koers bepaald worden en gewerkt worden naar het gewenste resultaat. Door het daarna te valideren, is de scope weer op het goede pad gekomen en konden er resultaten behaald worden.

Doorheen het project is er weinig/geen contact geweest met potentiële gebruikers van de webapplicatie. Hierdoor is er geen "echte" data beschikbaar om te valideren of de webapplicatie goed een bedrijf indexeert. Hiermee zal ook het NLP-model niet gevalideerd kunnen worden. Een volgende stap zou dus kunnen zijn om in de praktijk met dit PoC te experimenteren.

Conclusie

Duurzaamheid en circulariteit kennen de nodige uitdagingen. Het lectoraat Data Intelligence wilt kijken of deze uitdagingen getackeld kunnen worden met AI. Daarom zal eerst de huidige situatie in kaart gebracht willen worden.

Om dit in kaart te brengen is er gedurende dit project een webapplicatie gerealiseerd die aansluit op de doelstelling van het lectoraat. Deze doelstelling willen ze bereiken aan de hand van digitale readiness. Het lectoraat heeft een eigen index gerealiseerd om dit te bepalen: Data-driven Management Index. Deze DMI is verwerkt in een webapplicatie. Deze webapplicatie is ontwikkeld aan de hand van het Anvil framework. Het framework is gekozen met oog naar de toekomst om verdere tekstanalysemodellen toe te kunnen voegen. Een analysemodel is ontwikkeld aan de hand van NLP.

Aangezien deployment niet binnen de scope valt, is het model nog niet bruikbaar. Het advies hierop is dan ook om dit model verder door te ontwikkelen en te valideren met "echte" data.

Bibliografie

Different similarity measures. (sd). Opgehaald van Towardsdatascience:

<https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50#:~:text=Differences%20between%20Jaccard%20Similarity%20and,term%20frequency%20or%20tf%20idf>)

Hermann Lödding, R. R.-D. (sd). *Advances in Production The Path to Intelligent, Collaborative and Sustainable Manufacturing.*

Hevner. (sd). *A Three Cycle View of Design Science Research.* Opgehaald van Researchgate:

https://www.researchgate.net/publication/254804390_A_Three_Cycle_View_of_Design_Science_Research

Jamwal, A. A. (2021). *Industry 4.0 Technologies for Manufacturing Sustainability: A Systematic Review and Future Research Directions.* .

Luis Romeral Martínez, R. A. (2020). *New Trends in the Use of Artificial Intelligence for the Industry 4.0.*

Oden Technologies. (sd). Opgehaald van What is Digital Readiness: [https://oden.io/blog/what-is-digital-](https://oden.io/blog/what-is-digital-readiness/#:~:text=Digital%20readiness%20is%20defined%20by,enabled%20by%20software%20and%20technology)

[readiness/#:~:text=Digital%20readiness%20is%20defined%20by,enabled%20by%20software%20and%20technology](https://oden.io/blog/what-is-digital-readiness/#:~:text=Digital%20readiness%20is%20defined%20by,enabled%20by%20software%20and%20technology)

Sustainability Development Goals. (sd). Opgehaald van

<https://www.rijksoverheid.nl/onderwerpen/ontwikkelingssamenwerking/internationale-afspraken-ontwikkelingssamenwerking/global-goals-werelddoelen-voor-duurzame-ontwikkeling>

Thiede, P. S.-I. (2020). *Evolutie in de maakindustrie.*

Vector models. (sd). Opgehaald van Fasttext: <https://fasttext.cc/docs/en/crawl-vectors.html>)

Vries, W. d. (sd). *BERTje.* Opgehaald van Github: <https://github.com/wietsedv/bertje>

Wietse de Vries, A. v. (sd). *BERTje: A Dutch BERT Model.* Opgehaald van

<https://arxiv.org/abs/1912.09582>