# Checklist Systematisch Literatuuronderzoek

De tekst weergegeven in *blauw cursief* is ingevuld door de onderzoeker.

***Deelvraag 1:*** *Welke AI-/ML-technieken en -modellen worden er in de chemische sector gebruikt om het proces voor nieuwe chemische producten te ontwikkelen te ondersteunen?*

***Deelvraag 2:*** *Welke data-, hardware- en software-vereisten worden er gesteld om die technieken en modellen te kunnen gebruiken?*

## Formulering zoekvraag

Bepaal de hoofdvraag. Verdeel de hoofdvraag in deelvragen die een aspect van het onderwerp betreffen.

|  |  |
| --- | --- |
| **Hoofdvraag** | *Welke AI-/ML-technieken/-modellen worden er in het materials science-domein ingezet, en hoe zijn deze technieken geïmplementeerd?* |
| **Deelvragen** | *Voor welke toepassing is de AI-/ML-techniek ingezet?* |
| *Binnen welke component van materials science, past deze toepassing?* |
| *Welke data is gebruikt om het ML-model te trainen?* |
| *Hoe is het ML-model ontwikkeld; Welke stappen zijn genomen om het algoritme en model te ontwikkelen?* |
|  | *Welke softwarepakketten zijn gebruikt om het model te implementeren en welke versie van die softwarepakketten?* |
|  | *Zijn er bepaalde hardware-restricties of benodigdheden opgelegd die de inzet van het ontwikkelde model mogelijk te maken dan wel beperken?* |
|  | *Op welke wijze is het ML-model gevalideerd; zijn er fysieke experimenten uitgevoerd om de resultaten van het model te verifiëren?* |

## Globale afbakening van het onderwerp

Beperk je onderzoek tot een bepaalde periode (tijd), plaats en taal

|  |  |
| --- | --- |
| **Periode/Tijd** | *Geen restrictie met betrekking tot periode. Verwachting is wel hoe recenter, hoe relevanter en meer bruikbaar.* |
| **Plaats** | *Chemische Industrie. Indien mogelijk; specifiek petrochemische industrie* |
| **Taal** | *Nederlands en Engels* |

## Adviesgroep/ -panel

[Geef een lijst van de personen die je van plan bent te raadplegen, hun titels en organisaties, en hun potentiële rol in het literatuuronderzoek. Denk aan praktijkbeoefenaars, wetenschappers, bibliothecarissen en literatuuronderzoekspecialisten]

|  |  |  |  |
| --- | --- | --- | --- |
| **Naam** | **Titel** | **Organisatie** | **Rol in literatuuronderzoek** |
| *Sukrit* | *Research Scientist* | *Dow* | *Peer review van de opzet en resultaten van het onderzoek*  Is de opzet gericht op het vinden van:   1. Juiste chemische processen/reacties? 2. Juiste productsoorten en materialen?   Komen de resultaten overeen met het doel en de verwachting van het onderzoek? Zijn de resultaten te generaliseren naar de producten die Dow ontwikkelt? |
| *R Gerreman* | *IT Director* | *Dow* | *Peer review van de opzet en resultaten van het onderzoek* |
| *A Wongwall* | *IT Leader* | *Dow* | *Peer review van de opzet en resultaten van het onderzoek* |

Tabel 1 – Details leden adviesgroep

[Hoe denk je deze mensen te gaan betrekken in jouw onderzoek? Het is belangrijk om hierover voldoende informatie te verschaffen]

*Tabel 1 bevat de gegevens van de personen die zullen worden geraadpleegd om de opzet en de resultaten van het literatuuronderzoek te beoordelen. Na de eerste iteratie van het literatuuronderzoek (verzamelen en analyseren van de bronnen, analyse en weergeven van de resultaten), zullen de personen worden gevraagd om de resultaten te beoordelen op volledigheid en bruikbaarheid.*

*Het doel van de review en beoordeling is om te waarborgen dat de gebruikte opzet en methoden, en de resultaten die door middel van het onderzoek zijn verzameld, volledig zijn en dat de resultaten zijn te gebruiken in het verdere vervolg van het onderzoek.*

*De personen zullen voorafgaand aan de review, informatie krijgen over de opzet van het literatuuronderzoek; de methoden die gebruikt zijn en de invulling daarvan.*

*Naast de beoordeling van de volledige adviesgroep, is er tweewekelijks een overleg met de Research Scientist. In dit overleg zal de uitvoering en voortgang van het volledige onderzoek worden besproken. Deze momenten zullen ook worden gebruikt om het onderzoek waar nodig, bij te sturen.*

## Persoonlijke verklaring

Maak voorafgaande aan het onderzoek duidelijk:

• *Waarom doe je dit literatuuronderzoek?*

*Er is gekozen voor een literatuuronderzoek omdat dit de mogelijkheid biedt om de inzet van AI/ML in het materials science domein vanuit een breed, academisch perspectief te beschouwen. Daarnaast draagt het bestuderen van de literatuur met betrekking tot dit onderwerp, bij aan de ontwikkeling van het kennisniveau van de onderzoeker. Dit zal het verdere onderzoek ten goede komen.*

*• Wat hoop je te bereiken?*

*Het literatuuronderzoek dient een overzicht op te leveren van de verschillende ML-technieken gegroepeerd naar de componenten waaruit de materials science-discipline bestaat. Dit overzicht, met de bijbehorende gegevens op het gebied van data, software en hardware, zal dienen als basis voor het praktijkonderzoek.*

*• Welke eigen en professionele vooronderstellingen heb je?*

*De verwachting is dat er veel literatuur beschikbaar is over het onderwerp. Eén van de zaken die de dataverzameling en -analyse kunnen bemoeilijken is het technische karakter van artikelen. De beschrijving van de chemische processen, materialen en machine learning algoritmen, kunnen het kennisniveau van de onderzoeker overstijgen waardoor een goed begrip van het artikel niet vanzelfsprekend is.*

*Om dit probleem te minimaliseren zal er op regelmatige basis afstemming plaatsvinden met de research scientist van Dow, die de nodige technische achtergrond kan inbrengen. Op deze manier zal worden gewaarborgd dat de artikelen op een juiste manier geanalyseerd worden.*

*• Welke belangentegenstellingen (denk bijv. aan onderzoekssponsors) kunnen het literatuuronderzoek beïnvloeden?*

*Geen*

~~Maak gedurende het literatuuronderzoek aantekeningen over:~~

*~~• Alle belangrijke beslissingen, alsmede de rechtvaardiging voor die besluiten.~~*

*~~• Alle bijeenkomsten met de adviesgroepleden: wat werd er besproken? Wat is er besloten?~~*

*~~• De genomen maatregelen om vooronderstellingen te minimaliseren.~~*

*~~• Hoe jouw begrip over het onderwerp zich gaandeweg ontwikkelt.~~*

~~Geef na afloop van het literatuuronderzoek verklaringen voor:~~

*~~• Wat je hebt geleerd.~~*

*~~• Wat je anders zou doen (als je het onderzoek opnieuw zou moeten doen)~~*

*~~• De beperkingen van het literatuuronderzoek~~*

*~~• Hoe je het literatuuronderzoek gaat presenteren in jouw onderzoek~~*

*~~• De additionele literatuurdomeinen die moeten worden bestudeerd voor jouw dissertatie/scriptie.~~*

## Doel van het literatuuronderzoek

Wat is het doel van het literatuuronderzoek? Welke specifieke vragen denk je te gaan beantwoorden?

*Het doel van dit onderzoek is het verkrijgen van een overzicht van de toepassingen van AI/ML in materials science die vandaag de dag gebruikt worden in de chemische industrie, en de AI/ML-technieken die daarvoor ingezet worden.*

*Tevens dient het onderzoek een overzicht op te leveren, van de verschillende aspecten van de ontwikkeling van het ML-model of ML-techniek:*

1. *Gebruikte datasets.*
2. *Stappen voor het ontwikkelen van het algoritme.*
3. *Gebruikte software, inclusief versienummers.*
4. *Gebruikte hardwarecomponenten.*
5. *Wijze waarop het ML-model is gevalideerd.*

*De bedoeling van dit literatuuronderzoek is, om in eerste instantie, een algemeen beeld te ontwikkelen over de inzet van AI/ML in het materials science domein. De nadruk ligt op het beschouwen van dit onderwerp vanuit een zo breed mogelijk perspectief. Er dient in eerste instantie een overzicht ontwikkeld te worden waarbij zoveel mogelijk aspecten van machine learning én materials science, beschouwd worden.*

*Er is veel literatuur beschikbaar waarbij ingezoomd wordt op slechts één enkel aspect van dit onderwerp. Bijvoorbeeld een artikel over de inzet van een Neural Network voor het voorspellen van de elektronische structuur van een metaalachtig materiaal. Deze soort artikelen zullen in het tweede gedeelte van het literatuuronderzoek worden geanalyseerd (zie hoofdstuk 6 – Zoekstrategie voor een verdere uitleg over de uitvoering van het onderzoek).*

## Zoekstrategie

*• Hoe ga je het systematisch literatuuronderzoek uitvoeren?*

*• Welke informatiebronnen worden inbegrepen in je zoekacties?*

*• Maak een lijst van de zoekwoorden en -zinnen die je in het onderzoek gaat gebruiken:*

### **6.1 Uitvoering**

*Het literatuuronderzoek is opgesplitst in twee gedeelten. Het eerste gedeelte bestaat uit de zoektocht naar ML-technieken en -toepassingen in het domein, vanuit een zo breed mogelijk perspectief. Dat wil zeggen zonder restricties op het gebied van machine learning (ML-technieken of – algoritmes), of restricties op het gebied van materiaalsoorten en chemische processen.*

*Het tweede gedeelte van het onderzoek bestaat uit een verdere analyse van de ML-technieken en -toepassingen, die zijn gevonden in de eerste stap.*

1. *Stap 1: Algemene zoektocht en analyse*

*Voor het eerste gedeelte van het literatuuronderzoek wordt gezocht naar artikelen die een overzicht geven van de verschillende toepassingen van machine learning in het domein, met de daarbij ingezette ML-technieken. Er wordt specifiek gezocht naar artikelen die niet ingaan op één enkel onderzoek of onderwerp, maar waar juist wordt verwezen naar specifieke onderzoeken voor de verschillende toepassingen en technieken. De artikelen dienen in de beschrijving in te gaan op de toepassingen van machine learning en de daarbij gebruikte technieken en algoritmes.*

*Uit de eerste verkenning van de literatuur is gebleken, dat er een gigantische hoeveelheid literatuur beschikbaar is over dit onderwerp. De meeste artikelen gaan echter in op een specifiek onderzoek, waarbij slechts één toepassing van ML wordt beschouwd.*

*Het is praktisch niet uitvoerbaar om op basis van die gigantische hoeveelheid literatuur, een overzicht te compileren van verschillende ML-toepassingen. Daarom is ervoor gekozen om de eerste stap te richten op die onderzoeken, die een poging doen om een algemeen beeld te geven van de inzet van AI/ML in het materials science domein.*

*Vanuit de probleem- en doelstelling, en de onderzoeksvragen, zijn zoekwoorden gedefinieerd die zullen worden gebruikt om relevante artikelen te vinden. Zie tabel 2, paragraaf 6.3, voor een overzicht van de zoekwoorden met daarbij een korte uitleg per zoekwoord.*

*De zoekwoorden zijn ingedeeld naar categorie; onderwerp en context. Elk deelwoord van de categorie ‘onderwerp’ wordt gecombineerd met een zoekwoord uit de categorie ‘context’ (bouwsteenmethode). Deze combinatie levert de zoekquery op.*

*Notitie naar aanleiding van review uitgevoerd op 21 Maart 2022*

1. *Stap 2: Specifieke zoektocht en analyse*

*Het tweede gedeelte van dit literatuuronderzoek bestaat uit het analyseren van artikelen die ingaan op onderzoek naar een specifieke toepassing van machine learning en de daarvoor gebruikte technieken en algoritmes.*

*In stap 1 is een selectie gemaakt van artikelen die een breder overzicht geven over de inzet van AI/ML. In deze tweede stap wordt ingezoomd op de toepassingen waarnaar wordt verwezen, om op die manier te achterhalen wat de benodigdheden zijn, om de beschreven ML-technieken te kunnen implementeren.*

*De artikelen die zijn geselecteerd in stap 1 van dit onderzoek, zullen verder worden geanalyseerd op zoek naar verwijzingen naar die specifieke onderzoeken. Dit kan worden gezien als een variant van de sneeuwbalmethode.*

*De gevonden verwijzingen worden vastgelegd en zullen worden gebruikt als zoekquery, om het gerefereerde artikel te kunnen vinden.*

### **6.2 Informatiebronnen**

*Voor stap 1 zal er gebruik worden gemaakt van de volgende databanken:*

* *Google Scholar*
* *Arxiv.org*

*Beide informatiebronnen bevatten een grote hoeveelheid academische artikelen, waarvan het de verwachting is dat deze voldoende materiaal zullen opleveren.*

*Voor de uitvoering van stap 2 zal in eerste instantie alleen Google Scholar gebruikt worden. De verwachting is dat Google Scholar, op basis van de referentie uit de in stap 1 gevonden artikelen, het gezochte artikel zal opleveren. Indien Google Scholar het gezochte artikel niet kan vinden, zal dezelfde query (de referentie naar het artikel) gebruikt worden in Arxiv.org.*

### **6.3 Kern-/ Zoekwoorden**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Kern-/ zoekwoord*** | ***Beschrijving*** | ***Relevantie Onderzoeksvraag*** | ***Categorie*** |
| *Artificial Intelligence* | *Het door een computer simuleren van menselijke intelligentie* | *Het algemene onderwerp van het onderzoek* | *Onderwerp* |
| *AI-model* | *Een softwareprogramma dat is getraind om patronen te herkennen op basis van een dataset* | *Welke modellen er gebruikt kunnen worden en de hoeveelheid data die benodigd is voor het trainen ervan* | *Onderwerp* |
| *Machine learning* | *Een vorm van AI die gericht is op het kunnen leren van de verwerkte data om zo beter te presteren* | *Welke modellen er gebruikt kunnen worden en de hoeveelheid data die benodigd is voor het trainen ervan* | *Onderwerp* |
| *Deep learning* | *Een discipline van Machine Learning waarin het systeem kan leren met behulp van neurale netwerken* | *Welke modellen er gebruikt kunnen worden en de hoeveelheid data die benodigd is voor het trainen ervan* | *Onderwerp* |
| *Arificial Neural Network* | *Een discipline van Machine Learning waarin het systeem kan leren met behulp van neurale netwerken* | *Welke modellen er gebruikt kunnen worden en de hoeveelheid data die benodigd is voor het trainen ervan* | *Onderwerp* |
| *AI-algorithm* | *De set instructies die het AI-model gebruikt om tot een bepaald resultaat te komen* | *Welke modellen er gebruikt kunnen worden en de hoeveelheid data die benodigd is voor het trainen ervan* | *Onderwerp* |
| *Predictive Analytics* | *Serie analytische statistische technieken om toekomstige acties en/of gedragingen te voorspellen* | *Mogelijke techniek voor het inzetten van AI/ML* | *Onderwerp* |
| *Predictive Intelligence* | *Synoniem voor Predictive Analytics* | *Mogelijke techniek voor het inzetten van AI/ML* | *Onderwerp* |
| *Materials Science* | *Materiaalwetenschappen is het vakgebied dat zich bezighoudt met de samenstelling en eigenschappen van materialen* | *De context waarin het onderzoek wordt uitgevoerd* | *Context* |
| *Materials Design and Production* | *Zie 'Materials Science'* | *De context waarin het onderzoek wordt uitgevoerd* | *Context* |
| *Petrochemical Industry* | *Petrochemie is een sector die zich bezighoudt met het verwerken van aardolie* | *De context waarin het onderzoek wordt uitgevoerd* | *Context* |
| *Chemical Industry* | *Sector die zich bezighoudt met het vervaardigen van producten dmv chemische processen* | *De context waarin het onderzoek wordt uitgevoerd* | *Context* |
|  |  |  |  |

Tabel 2 – Kern- en zoekwoorden literatuuronderzoek

## Selectiecriteria

Hoe ga je het onderzoeksdomein of –onderwerp inperken? Geef eerst criteria voor het doorzoeken van titels en samenvattingen, gevolgd door selectiecriteria voor de volledige tekst doorzoekingen.

Criteria kunnen bestaan uit:

*• Sector (primaire, productie, diensten, hoogtechnologisch, enz.)*

*• Landen en/of regio’s (Nederland, Europa, VS, Azië, etc.)*

*• Publicatiejaar of –datum (na 1980, 1990 of vóór 2010, etc.)*

*• Populaties (praktijkbeoefenaars, middenmanagers, CEOs, beleidsmakers, enz.)*

*• Onderwerpfocus (organisatieontwerp, publiek-private samenwerking, strategische besluitvorming, etc.)*

### **7.1 Selectiecriteria**

Gebruik deze criteria bij het onderzoeken van titels en samenvattingen: *(Opmerking: deze criteria worden alleen gehanteerd voor stap 1 van het literatuuronderzoek. Voor stap 2 worden geen titels of samenvattingen geanalyseerd)*

|  |
| --- |
| ***Context/Sector:*** *Materials science/materiaalwetenschappen. Chemische processen. Artificial Intelligence en Machine Learning.* |
| ***Publicatiedatum:*** *na 01-01-2017* |
| ***Regio:*** *wereldwijd* |
| ***Taal:*** *Artikelen in het Engels en Nederlands* |
| ***Toegang tot artikel:*** *het artikel dient gratis toegankelijk te zijn. Hetzij in PDF of online leesbaar.* |

Selectiecriteria voor volledige teksten van artikelen:

|  |
| --- |
| ***Onderwerpfocus:***   * *Artikelen stap 1:*   *De tekst moet de toepassingen van AI/ML in het materials science-werkveld bevatten, met de daarvoor gebruikte AI-/ML-methoden en -technieken.*  *Het artikel wordt als relevant beschouwd als er beschreven wordt voor welke toepassingen machine learning is ingezet, en de daarvoor gebruikte ML-technieken en -algoritmes.*   * *Artikelen stap 2:*   *De artikelen moeten een antwoord geven op de volgende vier vragen:*   * *Is de dataset beschikbaar gesteld?* * *Zijn de stappen voor het ontwikkelen van het algoritme beschreven?* * *Zijn de omgevingsvariabelen (o.a. versienummers van de gebruikte software) genoemd?* * *Zijn de uitkomsten van het ML-model gevalideerd tegen de uitkomsten van fysieke experimenten?* |
|  |
|  |

### **7.2 Exclusiecriteria**

Criteria om artikelen uit te sluiten:

|  |
| --- |
| ***Onderwerpfocus:***   * *Artikelen stap 1:*   *Het artikel wordt uitgesloten als er geen duidelijke relatie wordt beschreven tussen de toepassing van AI/ML, en de daarvoor gebruikte ML-technieken of -algoritmes.*  *Als het artikel een onderzoek beschrijft, van slechts één specifieke toepassing met specifiek daarvoor ontwikkeld ML-model of -algoritme, zal het artikel als niet relevant worden beschouwd*. |
| ***Toegang tot artikel:*** *het artikel is niet gratis toegankelijk.* |
|  |

## Zoeken in catalogi, databanken, Internet

Schrijf hieronder per doorzochte databank de zoektermen op die je gebruikt hebt en de gegevens van de publicaties die je nodig denkt te hebben om je hoofdvraag te kunnen beantwoorden.

*Voor een overzicht van de gebruikte zoektermen, en gegevens die benodigd zijn om een antwoordt te geven op de onderzoeksvragen, wordt gerefereerd naar de documenten ‘Werkblad Deelvraag 1’ en ‘Werkblad Deelvraag 2’.*

*Zie hoofdstuk 10 – Gegevensonttrekking van dit document voor een uitleg over beide documenten.*

## Kwaliteitsbeoordeling

Geef aan op welke wijze de kwaliteit van de gevonden artikelen worden beoordeeld: [geef opsomming]

|  |
| --- |
|  |
|  |
|  |

### **9.1 Publicaties/ informatie**

Toets de kwaliteit van de publicaties door bij iedere publicatie die je gevonden hebt van de volgende criteria aan te geven of de publicatie hieraan voldoet.

1. toonaangevend tijdschrift?

2. toonaangevende uitgever?

3. toonaangevende auteur?

4. inhoud

5. actuele publicatie?

6. relevante publicatie?

Zet een kruisje (X) in de kolom van het corresponderende nummer in de bij stap 5 gevonden publicaties.

*Dit onderdeel is opgenomen in het Excel bestand ‘Werkblad Deelvraag 1’, zie kolommen G t/m K.*

## Gegevensonttrekking

Geef het proces van gegevens onttrekking weer. Geef de categorieën informatie die zullen worden onttrokken. Deze kunnen o.a. zijn:

• *Referentiegegevens (bijv. Auteur, tijdschrift, jaar, enz.)*

*• Beschrijvende informatie (land, bedrijfstak, middenmanagement, enz.)*

*• Methodologische informatie (empirisch/theoretisch, positivistisch/fenomenologisch, kwalitatief/kwantitatief, casusonderzoek/enquête, etc.)*

*• Thematische informatie (belangrijke concepten, ideeën, theorieën, aanpak, e.d.)*

*• [enz.]*

|  |
| --- |
| ***Referentiegegevens:*** *Titel, URL, Database, paginanummer van zoekresultaten, toegang tot artikel (Yes/No) publicatiedatum, Uitgever/Journal, Aantal citaties op Google Scholar.* |
| ***Thematische Informatie****: Zie paragraaf ‘Synthese’.* |
|  |

*De gegevensonttrekking van de artikelen is opgedeeld naar de twee stappen van dit literatuuronderzoek.*

*Voor stap 1 worden de volgende gegevens vastgelegd:*

* ***Zoektocht:***
  + *Gebruikte zoektermen.*
  + *Database waarin het artikel is gevonden.*
  + *Paginanummer van de zoekresultaten.*
* ***Referentiegegevens:***
  + *Titel van het artikel zoals weergegeven in de geraadpleegde database.*
  + *URL naar het artikel.*
  + *Of het artikel gratis toegankelijk is (Y/N).*
  + *Publicatiedatum van het artikel.*
  + *Uitgever van het wetenschappelijk tijdschrift waarin het artikel is gepubliceerd.*
* ***Kwaliteit en relevantie:***
  + *Toonaangevende uitgever (Y/N).*
  + *Het aantal citaties volgens Google Scholar.*
  + *Relevantie volgens de selectiecriteria beschreven in paragraaf 7.1 van dit document (Y/N).*
* ***Thematische informatie:***
  + *De toepassing(en) waarvoor machine learning is ingezet.*
  + *De ML-technieken en algoritmes die voor die toepassing zijn gebruikt.*
  + *Het nummer van de verwijzing naar het artikel waar het onderzoek van de specifieke toepassing is beschreven, in het geanalyseerde artikel.*

*Opmerking: als het artikel niet gratis toegankelijk is, zullen alleen de gebruikte zoektermen, database en paginanummer, URL en titel van de zoekresultaten worden vastgelegd.*

*De thematische informatie (ML-toepassingen en -technieken) dient als input voor stap 2 van het literatuuronderzoek. De artikelen waarvan zowel ML-toepassingen, als de daarvoor gebruikte ML-technieken en -algoritmes, gelden als basis voor stap 2.*

*Het tweede gedeelte van dit onderzoek bestaat uit het verzamelen van de gegevens, waaruit opgemaakt kan worden wat de benodigdheden zijn, om de ML-techniek binnen Dow te kunnen inzetten.*

*Stap 1 van het onderzoek levert een lijst op met daarin ML-toepassingen, de gebruikte ML-technieken, en de verwijzing naar het artikel waarin het onderzoek van die specifieke toepassing is beschreven.*

*In stap 2 worden de artikelen opgezocht die behoren bij de referenties die zijn gevonden in stap 1. Van deze artikelen worden de volgende gegevens vastgelegd:*

* *Nummer van de verwijzing in het originele artikel (dit is het artikel dat in stap 1 is gevonden).*
* *De tekstuele referentie uit dat artikel, gevonden op basis van het nummer.*
* *De titel van het artikel, zoals weergegeven op Google Scholar.*
* *URL naar het artikel.*
* *Of het artikel gratis toegankelijk is (Y/N).*

*Zodra deze lijst is samengesteld, zullen de artikelen worden geanalyseerd aan de hand van de volgende vragen:*

* *Is de dataset beschikbaar gesteld (Y/N)?*
* *Zijn de uitkomsten van het ML-model gevalideerd tegen resultaten van fysieke experimenten (Y/N)?*
* *Is beschreven hoe het algoritme is opgebouwd (Y/N)?*
* *Zijn de omgevingsvariabelen (software-versies, hardware benodigdheden) benoemd (Y/N)?*

*De uitkomsten van deze vragen zullen per artikel, in een tabel in Excel worden vastgelegd.*

## Synthese

Hoe worden de bevindingen en concepten van de verschillende onderzoeken gecompileerd of samengevoegd?

*Het verzamelen en beoordelen van de data uit de literatuur, is op iteratieve wijze uitgevoerd. Elke stap van de synthese borduurt voort op de voorgaande stap, waarbij in de nieuwe stap de artikelen opnieuw zijn geanalyseerd.*

*In de volgende paragrafen wordt per stap beschreven, welke acties er zijn ondernomen, en hoe de gegevens zijn vastgelegd.*

*De presentatie van de resultaten uit deze synthese, zijn verwerkt in hoofdstuk 3 ‘Theorie’ van het document ‘Onderzoeksopzet en Theorie’.*

### Synthese Stap 1

*Stap 1: Algemene zoektocht en analyse*

*Het doel van deze stap is het verkrijgen van een lijst van ML-toepassingen, geclassificeerd naar de componenten van materials science, en de additionele categorie ‘methodieken’.*

#### **Stap 1.1**

*De artikelen die zijn gevonden voor stap 1, zijn vastgelegd in een tabel in Excel. Deze tabel bestaat uit de volgende kolommen:*

* ***Onderwerp*** *– de combinatie zoekwoorden.*
* ***Database*** *– database waarin gezocht is.*
* ***Naam*** *– naam van het artikel zoals is weergegeven in de database.*
* ***URL*** *– link naar het artikel.*
* ***Toegang*** *– of het artikel gratis toegankelijk is.*
* ***Publicatiedatum*** *– datum waarop het artikel gepubliceerd is.*
* ***Uitgever*** *– naam van de uitgever die het artikel heeft gepubliceerd.*
* ***Aantal citaties*** *– het aantal citaties zoals aangegeven bij het zoekresultaat in Google Scholar.*
* ***Relevantie*** *– beoordeling of het artikel voldoet aan de gestelde selectiecriteria.*
* ***Toepassing van ML*** *– overzicht van de toepassing van ML zoals beschreven in het artikel. Tussen haakjes het paginanummer waar de desbetreffende toepassing is beschreven.*
* ***ML-methoden, -technieken en -algoritmes*** *– overzicht van de gebruikte ML-techniek. Nummering komt overeen met de nummering in de kolom Toepassing.*

*De tabel met de gegevens is opgenomen in tabblad ‘Stap 1.1’ in document Werkblad Deelvraag 1.*

*Additionele gegevens met betrekking tot de zoektocht, zoals onder andere datum van de zoekopdracht, specifieke zoekquery, aantal hits, is beschikbaar op aanvraag bij de onderzoeker.*

#### **Stap 1.2**

*Om een overzicht te krijgen van die artikelen, waarvan zowel ML-toepassingen, als ML-technieken zijn genoemd, is een pivottabel aangemaakt, met als databron de volledige tabel uit ‘Stap 1.1’.*

*In die pivot zijn de volgende tabellen geselecteerd:*

* *Naam*
* *Aantal Citaties*
* *Toepassing van ML*
* *ML-methoden, -technieken en -algoritmes*

*De kolom ‘Toegang?’ is ingesteld als filter voor de pivot. In deze filter is alleen de waarde ‘Y’ geselecteerd zodat alleen die artikelen worden weergegeven, die gratis toegankelijk zijn.*

*In de kolommen ‘Toepassing van ML’ en ‘ML-methoden, -technieken en -algoritmes’ zijn de cellen met de waarden ‘n/a’, ‘(blank)’ en/of lege cellen, niet opgenomen in de filter.*

*Deze pivot inclusief filters heeft een overzicht opgeleverd, van die artikelen waarin verschillende ML-toepassingen met bijbehorende ML-technieken, zijn beschreven.*

#### **Stap 1.3**

*In deze stap is een eerste classificatie van de toepassingen gemaakt. De gevonden artikelen zijn nogmaals geanalyseerd, waarbij in deze iteratie een codering is aangebracht in de artikelen. De uitkomst van die codering is weergegeven in tabblad ‘Stap 1.3 Artikelen Gecodeerd’, kolom A – Artikelen, en kolom B - Codering. De artikelen, inclusief de codering zijn te vinden op* [*https://github.com/braaltje/Artikelen-Gecodeerd/tree/main/Gecodeerd*](https://github.com/braaltje/Artikelen-Gecodeerd/tree/main/Gecodeerd)*.*

*Voor het coderen van de artikelen is gebruikgemaakt van het gratis softwareprogramma PDF-XChange Editor. Om de labels bij de codering te kunnen inzien, is het noodzakelijk om het gecodeerde artikel te openen in PDF-XChange Editor.*

*Vervolgens zijn de coderingen ingedeeld naar verschillende categorieën. Deze categorieën hebben betrekking op het volledige artikel, en geven dan ook de niet-relevante tekstgedeeltes weer. Het doel van deze categorisering, was om eenvoudiger te kunnen filteren op de ML-toepassingen per artikel.*

#### **Stap 1.4**

*De volgende stap in de verwerking en analyse, bestaat uit het onttrekken van de verschillende ML-technieken voor elke toepassing. De resultaten hiervan zijn terug te vinden in tabblad ‘Stap 1.4 Art-Toepas-ML Techn’.*

*Elk tekstgedeelte dat in stap 1.3 is geclassificeerd als ‘Toepassing’, is nogmaals geanalyseerd om te achterhalen welke ML-technieken en -algoritmes zijn genoemd.*

*Ter voorbereiding op stap 2 van het literatuuronderzoek, zijn hierbij ook de nummers van de verwijzingen genoteerd. Dit zijn de verwijzingen naar referenties in het desbetreffende artikel, die de specifieke onderzoeken van de ML-toepassing, en de gebruikte ML-technieken beschrijven. In stap 2 zijn deze referenties gebruikt om de bijbehorende artikelen te zoeken.*

*De tabel waarin de gegevens zijn vastgelegd bestaat uit de volgende kolommen:*

1. ***Artikel nummer*** *- het nummer van het artikel dat in stap 1 is geanalyseerd, en waarmee het gecodeerde artikel is opgenomen in de repository* (<https://github.com/braaltje/Artikelen-Gecodeerd/tree/main/Gecodeerd)>.
2. ***Codering*** *- de tekst van het label van de codering in het artikel.*
3. ***Categorie*** *- de categorie van het label dat in stap 1.3 is toegekend (tabel zou alleen regels mogen bevatten met de categorie ‘Toepassing’).*
4. ***ML Techniek*** *- de in het gecodeerde tekstgedeelte genoemde ML-techniek.*
5. ***Verwijzing#*** *- het nummer van de verwijzing in het gecodeerde tekstgedeelte van het desbetreffende artikel.*

#### **Stap 1.5**

*De codering ‘Toepassing’ die in stap 1.3 is toegekend, is gebaseerd op de toepassing zoals die in het desbetreffende artikel is genoemd. In de verschillende artikelen en onderzoeken, zijn verschillende namen gebruikt voor vergelijkbare toepassingen. Dit leidt dan ook tot een lange lijst met toepassingen, waarvan sommige kunnen worden samengevoegd.  
  
In deze stap is voor alle toepassingen uit stap 1.3, een categorie toegevoegd met als doel de toepassingen te kunnen samenvoegen in een kleiner aantal categorieën. In een volgende stap zal deze kortere lijst verder worden onderverdeeld, naar de uiteindelijke categorieën waarmee de resultaten worden gepresenteerd.*

*De tabel uit stap 1.4 is gekopieerd naar tabblad ‘Stap 1.5 Toepassing->Categorie’, waarna de categorieën zijn toegevoegd in kolom F en G. Er is een generalisering van twee niveaus ingebracht in de categorisering. In deze tabel is de categorie die is genoemd in Codering, de meest gedetailleerde. Het 1e en 2e niveau zijn respectievelijk steeds iets algemener. De categorisering die is genoemd in het 2e niveau zal worden gebruikt in het verslag.*

*In de tabel die is gecreëerd in stap 1.5, zijn de rijen verwijderd waarvan de waarde in kolom D – ML-techniek overeenkomt met ‘Niet genoemd’, ‘Geen specifieke techniek, algemeen artikel’ en ‘Te specifiek over medicijnonderzoek, niet relevant voor Dow’.*

*Voor deze rijen zijn geen specifieke ML-technieken bekend, of zijn niet relevant voor Dow vanwege het onderwerp. Deze rijen zijn nog wel in te zien in de tabel uit stap 1.4.*

#### **Stap 1.6**

*De namen van de ML-technieken die zijn genoemd in stap 1.5, zijn weergegeven zoals ze zijn genoemd in de desbetreffende artikelen. Ook hier geldt dat er verschillende benamingen zijn gebruikt voor vergelijkbare, of dezelfde technieken, zodat enige opschoning vereist is.*

*Daarnaast bestaat de lijst met technieken uit verschillende niveaus. Zo zijn in sommige gevallen de namen van de algoritmes zelf genoemd, en in andere gevallen het soort leren dat is toegepast.*

*Door de lijst met ML-technieken te consolideren is geprobeerd de verschillende niveaus te reduceren, en de verschillende technieken en algoritmes, op een vergelijkbaar niveau te kunnen presenteren.*

*De tabel uit stap 1.5 is in zijn volledigheid gekopieerd naar tabblad ‘Stap 1.6 ML-Technieken’, waarna de volgende kolommen zijn toegevoegd: (kolom H uitleggen)*

*Categorieën:*

* *Artificial Neural Network: naar het werk van Leijnen en Veen (2020).* Leijnen, S., & Veen, F. V. (2020). The Neural Network Zoo. *IS4SI 2019 Summit*. https://doi.org/10.3390/proceedings2020047009
* *Combinatie van algoritmes:*
* *ML-Potentials:*
* *Geen ML-algoritme:*

*Verdere uitleg over het creëren v/e tabel voor ML technieken per toepassing, opschoning van de technieken, en redenen waarom bepaalde technieken zijn weggelaten uit de tabel.*

### Synthese Stap 2

*Stap 2: Specifieke zoektocht en analyse*

#### **Stap 2.1**

*Er is begonnen met het vastleggen van de referenties waarnaar wordt verwezen, in de artikelen die in stap 1 zijn geanalyseerd. Er is een kopie gemaakt van de tabel beschreven in stap 1.4, die is aangevuld met de volgende kolommen:*

1. ***Referentie beschrijving*** *– de tekst van de referentie waarnaar wordt verwezen in het desbetreffende artikel.*
2. ***Titel*** *– de titel van het artikel, zoals gevonden en weergegeven in Google Scholar.*
3. ***URL*** *– de URL naar het artikel.*
4. ***Toegang****? – of het artikel gratis toegankelijk is (Y/N).*

*Zie voor de uitwerking hiervan het tabblad ‘Stap 2.1 Artikelen uit DV1’ in het document ‘Werkblad Deelvraag 2’.*

*Opmerking: voor de artikelen gepubliceerd voor 1 Januari 2017 (zie paragraaf 7.1 en 11.2.2), zijn de Titel, URL en Toegang niet vastgelegd.*

#### **Stap 2.2**

*Vervolgens is bepaald of het artikel waarnaar wordt verwezen voldoet aan het selectiecriterium met betrekking tot publicatiedatum. Artikelen dienen gepubliceerd te zijn na 1 januari 2017, om te worden meegenomen in de analyse.*

*De publicatiedatum is bepaald aan de hand van de tekst van referentie. Het jaartal dat is genoemd in die tekst, is beschouwd als het jaartal waarin het artikel is gepubliceerd. De regels waarvan het artikel is gepubliceerd voor 1 Januari 2017, zijn rood gemarkeerd.*

#### **Stap 2.3**

Pivot gemaakt van 2.1/2.2

#### **Stap 2.4**

Vastlegging analyse

## Notities en Aanpassingen

*Naarmate het onderzoek vorderde, en op basis van feedback verkregen uit de review van de adviesgroep, zijn enkele kleine aanpassingen in het onderzoek nodig gebleken.  
In dit hoofdstuk wordt kort beschreven waar de aanpassingen uit bestaan en op basis van welke gronden deze aanpassingen en keuzes zijn gemaakt.*

1. ***Aanpassing in zoekstrategie op basis van een combinatie van de geïdentificeerde zoekwoorden.***

*De combinaties van zoekwoorden leveren heel veel hits op in Google Scholar en Arxiv.org. De analyse van de titels en samenvattingen heeft zich beperkt tot de eerste tien pagina’s met resultaten per combinatie zoekwoorden.*

1. ***Zoek- en kernwoorden aangepast.***

*De zoekwoorden ‘Predictive Analytics’, ‘Predictive Intelligence’ en ‘Artificial Neural Network’ zijn verwijderd. Deze termen beperken de scope doordat ze ingaan op specifieke toepassingen (‘Predictive Analytics’ en ‘Predictive Intelligence’) en een specifieke ML-techniek (‘Artificial Neural Network’).*

*Het doel van de eerste stap in dit onderzoek, was het vinden van toepassingen en technieken vanuit een zo breed mogelijk perspectief. De verwijderde zoekwoorden vormden een beperking op dat perspectief.*

1. ***Aanpassing n.a.v. review plaatsgevonden op 22 Maart 2022.***

*Uit de beoordeling van de resultaten van stap 1 van dit literatuuronderzoek (zie details in paragraaf 6.1), is gebleken dat de categorie ‘Methodologieën’ niet voldoende is afgedekt. De voor deze categorie gevonden toepassingen van ML, hebben vooral betrekking op de ondersteuning en verbetering van bestaande methodologieën voor computationele chemie (DFT en MD).*

*De Dow research scientist, lid van de adviesgroep, heeft een lijst met auteurs aangeleverd waarvan de verwachting is, dat deze actief onderzoek doen naar ML-toepassingen buiten het computationele chemie domein.*

*Op basis van de lijst met namen van de auteurs, zijn aanvullende artikelen gezocht. Voor een deel van de auteurs waren artikelen beschikbaar uit stap 2.2 (zie paragraaf 11.2.2). Deze artikelen zijn overgenomen en aangevuld met een additionele zoektocht. Zie tabblad ‘Aanvullende Artikelen’ in het document ‘Werkblad Deelvraag 2’.*

*De zoektocht naar additionele artikelen is uitgevoerd op Google Scholar, met als zoekterm de naam van de auteur. Artikelen gepubliceerd voor 1 Januari 2017 zijn niet opgenomen in de analyse. Tevens dient het artikel gratis toegankelijk te zijn om te worden geselecteerd voor de analyse. Artikelen die niet voldoen aan deze criteria, zijn niet opgenomen in de lijst.*

*De gevonden artikelen zijn inhoudelijk geanalyseerd om te bepalen of het artikel betrekking heeft op een computationele methode. De resultaten hiervan zijn verwerkt in de tabel in het tabblad ‘Aanvullende Arikelen’.*

*Voor de artikelen die niet ingaan op een computationele methode, is een volgende analyse uitgevoerd om de ML-toepassingen en -technieken, te achterhalen en vast te leggen. Deze zijn ook vastgelegd in de tabel.*