



## **Data Engineering Project II - Documentatie**

Faculteit IT en digitale innovatie  
Valentin Vaerwyckweg 1  
9000 GENT

ACADEMIEJAAR 2022-2023

OPLEIDING TOEGEPASTE INFORMATICA

GROEP: 03

**STUDENTEN:** Daems Benjamin, Beeckman Jarno, Bjarne Van den Abbeelee,  
Montaine Tanguy, Van Campenhout Siebe

**BEGELEIDER:** Johan Decorte, Guy Dekoning

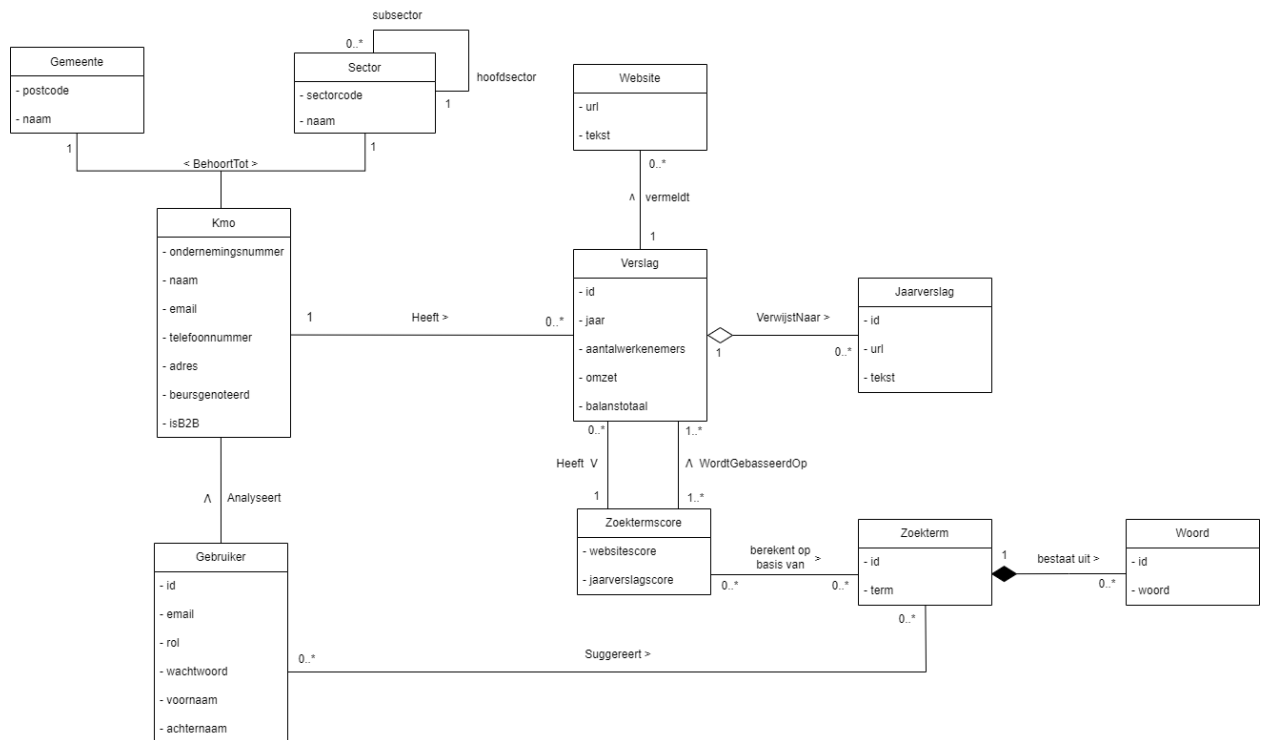
## Inhoudsopgave

Inhoudsopgave .....	1
Analyse.....	3
Use Case Diagram.....	3
Entity Relationship Diagram.....	3
Domain Model.....	4
Functional Requirements.....	4
Use case 1.....	4
Use case 2.....	4
Use case 3.....	5
Use case 4.....	5
Use case 5.....	5
Use case 6.....	6
Use case 7.....	6
Non Functional Requirements .....	7
NFR 1:.....	7
NFR 2:.....	7
NFR 3:.....	7
NFR 4:.....	7
NFR 5:.....	8
NFR 6:.....	8
NFR 7:.....	8
NFR 8:.....	8
Technische Keuzes.....	9
Databank.....	9
Backend.....	9
Frontend.....	9
Machine Learning .....	9
Testing .....	10
Functional Changes .....	11
Applicatie.....	12
Aanmelden en registreren .....	13
Account.....	14
Dashboard .....	15

Grafieken – Vergelijking met andere bedrijven in de sector .....	17
Sector overzicht.....	18
KMO overzicht.....	19
Moderator Tools - Zoektermen.....	20
Moderator Tools - Voorspellen .....	21
Admin Tools – Rollen Wijzigen.....	22
Ethische Reflectie.....	23
Repositories.....	25



## Domain Model



## Functional Requirements

### Use case 1

**Use case:** Gebruiker aanmelden

**Primaire actor:** gebruiker (not logged-in)

**Stakeholders:** /

**Precondities:** /

**Postcondities:** De gebruiker is aangemeld

**Normaal verloop:**

1. De gebruiker logt in via email-adres en paswoord
2. Systeem valideert gebruiker volgens DR\_login

**Alternatieve verlopen:**

1A. De gebruiker maakt een nieuw account aan

1A1. De gebruiker registreert via naam, email-adres en paswoord

1A2. Keer terug naar stap 1 in het verloop

2A. De gebruiker voert ongeldige gegevens in

2A1. Het systeem geeft een gepaste melding

2A2. Keer terug naar stap 1 in het verloop

**Domeinspecifieke regels:**

DR\_login

De inloggegevens moeten correct zijn

### Use case 2

**Use case:** Gebruiker afmelden

**Primaire actor:** gekende gebruiker/beheerder/admin

**Stakeholders:** /

**Precondities:** De gebruiker is reeds aangemeld

**Postcondities:** De gebruiker is afgemeld

**Normaal verloop:**

1. De gebruiker meldt zich af
2. Systeem verwijst gebruiker naar startpagina

**Alternatieve verlopen:**

/

**Domeinspecifieke regels:**

/

### Use case 3

**Use case:** Sectorenlijst bekijken

**Primaire actor:** gekende gebruiker/beheerder/admin

**Stakeholders:** /

**Precondities:** De gebruiker is reeds aangemeld

**Postcondities:** De gebruiker ziet score's van KMO's in eenzelfde sector

**Normaal verloop:**

1. De gebruiker filtert KMO's uit sectorenlijst
2. Het systeem toont lijst KMO's met score's

**Alternatieve verlopen:**

/

**Domeinspecifieke regels:**

/

### Use case 4

**Use case:** Bedrijven opzoeken

**Primaire actor:** gekende gebruiker/beheerder/admin

**Stakeholders:** /

**Precondities:** De gebruiker is reeds aangemeld

**Postcondities:** De gebruiker ziet alle gegevens omtrent opgezocht bedrijf

**Normaal verloop:**

1. De gebruiker zoekt via KMO
2. Het systeem geeft overzicht data, duurzaamheidsscore en coding tree

**Alternatieve verlopen:**

/

**Domeinspecifieke regels:**

/

### Use case 5

**Use case:** Duurzaamheidsscore voorspellen

**Primaire actor:** gekende gebruiker/beheerder/admin

**Stakeholders:** /

**Precondities:** De gebruiker is reeds aangemeld

**Postcondities:** De gebruiker verkrijgt voorspelling aan de hand van ingevoegde gegevens

**Normaal verloop:**

1. De gebruiker voorspelt duurzaamheidsscore via aantal werknemers, omzet, balanstotaal, verstedelijkheidsgraad en hoofdsector
2. Het systeem toont duurzaamheidsscore

**Alternatieve verlopen:**

/

**Domeinspecifieke regels:**

/

## Use case 6

**Use case:** Duurzaamheidsscore voorspellen

**Primaire actor:** gekende gebruiker/beheerder/admin

**Stakeholders:** /

**Precondities:** De gebruiker is reeds aangemeld

**Postcondities:** De lijst met zoektermen is aangepast

**Normaal verloop:**

1. De gebruiker voegt zoektermen toe/verwijdert zoektermen
2. Het systeem toont lijst zoektermen

**Alternatieve verlopen:**

/

**Domeinspecifieke regels:**

/

## Use case 7

**Use case:** Rollen beheren

**Primaire actor:** admin

**Stakeholders:** /

**Precondities:** De admin is reeds aangemeld

**Postcondities:** De admin heeft rol(len) toegevoegd of verwijderd van een gebruiker

**Normaal verloop:**

1. De admin voegt rol toe aan gebruiker via email gebruiker
2. Systeem valideert volgens DR\_rol
3. Systeem toont toegevoegde rol gebruiker

**Alternatieve verlopen:**

2A. De gebruiker voegt rol toe dat al in lijst "huidige rollen" staat

2A1. Het systeem verwijdert rol gebruiker

2A2. Keer terug naar stap 1 in het verloop

**Domeinspecifieke regels:**

## DR\_rol

De rol mag niet in lijst "huidige rollen" staan

## Non Functional Requirements

### NFR 1:

NFR	Onderhoudbaarheid
Indicator	Wijzigbaarheid
Meetvoorschrift	De moderator of admin moet zelf zoektermen kunnen wijzigen.
Norm	De tijd dat nodig is de zoektermen aan te passen bedraagt hoogstens twee minuten.

### NFR 2:

NFR	Betrouwbaarheid
Indicator	Beschikbaarheid
Meetvoorschrift	De connectie met de databank mag niet verstoord worden tijdens het gebruik van de applicatie.
Norm	De up-time van de databank moet meer dan 99.0% zijn doorheen het hele jaar

### NFR 3:

NFR	Beveiligbaarheid
Indicator	Authenticiteit
Meetvoorschrift	De admin en moderator krijgen andere permissies die bij deze rollen behoren.
Norm	Het system moet de correcte rollen associëren met de gebruiker op het moment van inloggen.

### NFR 4:

NFR	Bruikbaarheid
Indicator	Bedienbaarheid
Meetvoorschrift	Een onervaren internetgebruiker moet zich makkelijk kunnen navigeren door de applicatie zonder IT voorkennis te hebben.
Norm	De onervaren internetgebruiker kan binnen één minuut een bedrijf naar keuze opzoeken en data ervan ophalen.



### NFR 5:

NFR	Bruikbaarheid
Indicator	Herkenbare geschiktheid
Meetvoorschrift	De gebruiker moet kunnen beoordelen op de applicatie voldoet aan zijn vereisten.
Norm	De gebruiker start de applicatie voor het eerst op en kan binnen tien seconden beoordelen waarover de applicatie gaat.

### NFR 6:

NFR	Prestatie-efficiëntie
Indicator	Snelheid
Meetvoorschrift	De applicatie moet aan de hand van de nieuwe ingegeven zoektermen scores kunnen herbepalen voor alle bedrijven.
Norm	Het systeem herberekent de zoekterm scores binnen één minuut.

### NFR 7:

NFR	Functionele geschiktheid
Indicator	Functionele correctheid
Meetvoorschrift	De voorspellingsscores moeten binnen drie seconden berekend worden.
Norm	De nauwkeurigheidsgraad van de voorspellingen is minstens 99.5%.

### NFR 8:

NFR	Functionele geschiktheid
Indicator	Functionele compleetheid
Meetvoorschrift	Het ophalen van de gegevens voor een bedrijf laadt direct na het ingeven van een bedrijf.
Norm	De data van alle bedrijven moeten voor minstens 90% vervuld zijn.

## Technische Keuzes

### Databank

Als databank hebben we net zoals aanbevolen gekozen voor een Postgres databank. Deze ondersteunt full-text search, en dit was net wat we nodig hadden.

De full-text search wordt gebruikt bij het berekenen van de scores voor al de KMO's. De volledige tekst van de website en jaarverslagen wordt opgeslagen in de databank.

Voor de websites wordt er rekening gehouden met welke taal het onderdeel is. Deze wordt daarna gesplitst in Nederlands en Engels. Ook al de pdf-bestanden worden ingelezen.

Het jaarverslag, dat we ophaalden van de Nationale Bank van België, analyseerde we ook volledig. In tegenstelling tot de website kan dit maar in één, taal zijn, het Nederlands. Al de afbeeldingen die voorkomen in het jaarverslag worden ook gelezen met OCR. Zodat we zeker al de tekst hebben.

Wanneer al de tekst opgeslagen is wordt deze omgezet naar een tekst-vector waarop je de full-text search kan uitvoeren. Wanneer de scores berekend moeten worden wordt voor elke KMO gekeken welke taal het meest voordelige is. Deze score wordt genomen.

### Backend

Dit jaar hebben we in tegenstelling tot vorig jaar een Python backend gemaakt. Deze werd met behulp van FastAPI gemaakt. Dit is een framework om snel stabiele Python API's te maken.

Om een databank connectie te maken gebruiken we SQLAlchemy. Met deze package kan je query schrijven in pure Python code.

Al de functies van vorig jaar werden ook overgenomen in deze API. Er zijn ook nog functies toegevoegd. Zoals de searchterms, rollen, ...

### Frontend

We hebben voor de frontend van dit jaar verder gebouwd op de frontend van vorig jaar. Deze werd gemaakt in React.js. Een JavaScript framework om eenvoudig dynamische websites mee te ontwikkelen.

Om de gevraagde grafieken te tonen maakten we gebruik van het Recharts package. Hiermee kan je verschillende soorten grafieken tonen aan de hand van de data die je ingeeft.

### Machine Learning

Voor ons model hebben we gekozen voor Tensorflow Keras. Tensorflow is open-source, en de Keras-API maakt de creatie van een model zeer gemakkelijk.

We hebben voor een neuraal netwerk gekozen, omdat deze goed zijn voor regressie-problemen (score bepalen). Daarbij is de data zelf vrij complex (700 sectoren, 3 soorten verstedelijking, al dan niet beursgenoteerd).

We maken gebruik van GridSeachCV met een Keras wrapper class om de beste parameters voor het model te bepalen, en maken gebruik van de EarlyStopping-callback om te zorgen dat er geen overfitting plaats vindt.

Om een gelijke verdeling in de dataset te maken, hebben we besloten dat een minimum van 200 bedrijven per afgeronde score een redelijk aantal is. Dit wil zeggen dat we de bedrijven gegroepeerd hebben op basis van hun score, afgerond naar een integere waarde. Nadien hebben we alle groepen met minder dan 200 bedrijven laten vallen, en nadien alle groepen verkleint naar het kleinste overblijvende aantal. Dit gaf ons een resultaat van 18 groepen met 221 bedrijven.

### **Testing**

We gebruiken testing in de backend. De eerste testen die we schreven waren geschreven in Pytest. Dit kan je gemakkelijk uitvoeren in je IDE. Wat ook handig is aan Pytest is dat je fixtures kan toevoegen waardoor je voor elke test nieuwe data kan genereren. Ook mocking van methodes in de code was een mogelijkheid.

De API request zelf testen deden we via Postman. Hiermee kan je ook testsuites bouwen waarmee je direct al de functionaliteiten van je API kan testen.

Ook de scratches bevatten een aantal testen zodat er verzekerd kan worden dat de code werkt.

## Functional Changes

Er zijn een aantal wijzigingen geweest aan de scope doorheen het semester. Hier gingen we meestal al snel te werk om te kijken hoe we deze nieuwe uitdagingen konden aanpakken.

### Zoeken in de woorden

Dit is het meeste veranderd over heel het project. De klant wist zelf niet goed wat al de mogelijkheden waren en wat het beste was. Wij hebben gekozen om een berekening uit te voeren waardoor de beste KMO een score heeft van 1 op één onderwerp en de slechtste 0. De rest valt hiertussen naarmate hoeveel zij gescoord hebben. Daarna werd er een som gemaakt van al de verschillende soorten scores.

Vb.

KMO1: Score = 0.75 -> 1.0

KMO2 Score = 0.6 -> 0.5

KMO3: Score = 0.25 -> 0.0

### Meertaligheid Applicatie

In het begin moest er geen rekening gehouden worden met de meertaligheid van de websites. Wij hadden dit dus ook niet gedaan.

Met het laatste gesprek komen we te weten dat we de Engelse en Nederlandse delen van de website apart moeten opslaan om zo correct te zoeken. We moesten dus nog last minute al de data opnieuw ophalen zodat we dit konden toepassen.

### Grafieken in de applicatie

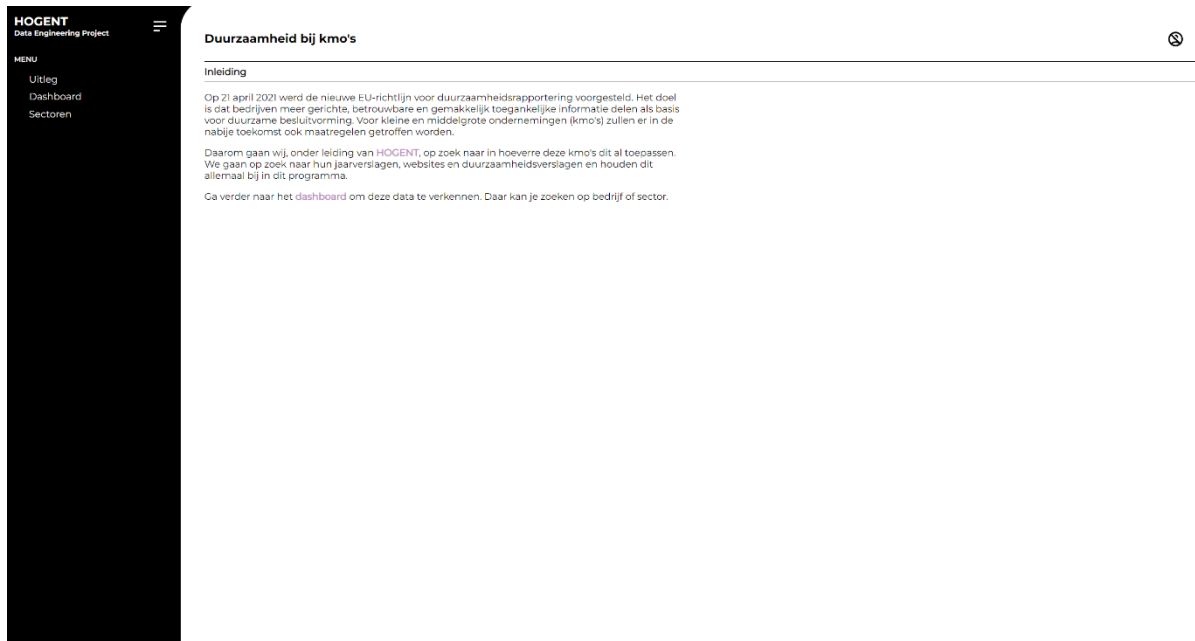
Ook met het laatste contactmoment komen we te weten dat de klant graag extra grafieken te zien krijgt in de applicatie. Deze hebben we dus ook lastminute moeten toevoegen. Hiermee kan je nu een vergelijking maken met afgelopen jaren en hoe goed ze bezig zijn ten opzichten van de andere kmos in hun sector.

## Applicatie

Hieronder staat een stappenplan van al de functies de applicatie bevat.

### Hoofdscherm

Wanneer je surft naar de website krijg je een uitleg-scherm te zien. Hier wordt uitgelegd wat de bedoeling is van de applicatie.



## Aanmelden en registreren

In de applicatie is het mogelijk om aan te melden en te registreren. Zonder een account kan je geen gegevens raadplegen.

### Aanmelden



WELKOM BIJ HO Gent - DATA ENGINEERING PROJECT

EMAIL\*

WACHTWOORD\*

**Aanmelden**

NOG GEEN ACCOUNT? [REGISTREER HIER.](#)

### Registreren



WELKOM BIJ HO Gent - DATA ENGINEERING PROJECT

VOORNAAM\*

ACHTERNAAM\*

EMAIL\*

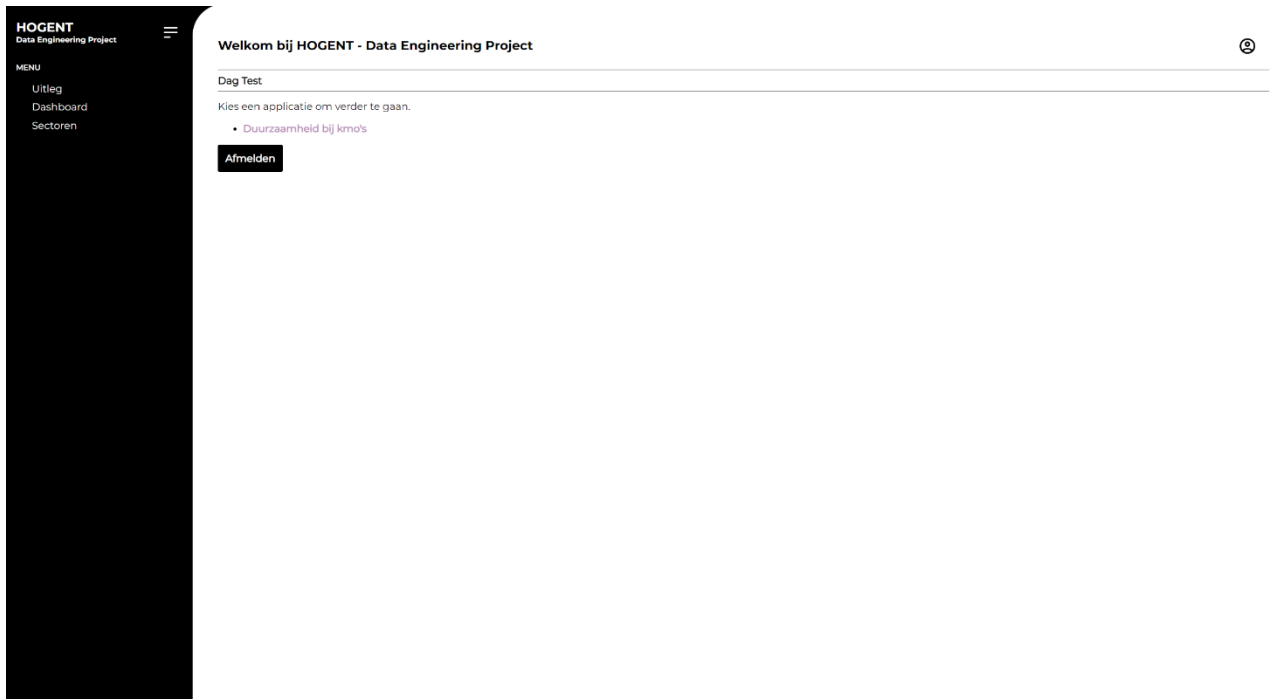
WACHTWOORD\*

**Registreer**

HEB JE AL EEN ACCOUNT? [MELD HIER AAN.](#)

## Account

Wanneer er een account gecreëerd wordt krijg je een account scherm te zien. Hier is niet veel te beleven. Het enige wat je kan doen is doorklikken naar het dashboard.



## Dashboard

Het dashboard is het hart van de applicatie. Hier kan je zoeken naar bedrijven om al hun informatie weer te geven.

**HOGENT**  
Data Engineering Project

MENU

[Uitleg](#)  
[Dashboard](#)  
[Sectoren](#)

veneco

Q

INTERGEMEENTELIJKE VERENIGING VENECO

**Venootschapsnaam**

INTERGEMEENTELIJKE VERENIGING VENECO

**Ondernemingsnummer**

BE0200065765

**Telefoonnummer**

+32 9 343 81 57

**Email**

[info@veneco.be](mailto:info@veneco.be)

**Adres**

Panhuysstraat 1  
9070, Destelbergen

**Website**

<https://www.veneco.be>

**Jaarverslag**

[Download Jaarverslag](#)

**Omzetcijfer**

€ 5439

**Balanstotaal**

€ 78782159

**beursgenoteerd**

Niet Beursgenoteerd

**Aantal werknemers**

27

**Hoofdsector**

Openbaar bestuur en defensie; verplichte sociale verzekeringen

**Subsector**

Openbaar bestuur op het gebied van het bedrijfsleven; stimuleren van het bedrijfsleven

Scores

Totale score: 32.3 punten

Website score: Geen score gevonden

Jaarverslag score: 64.6 punten

Grafieken

HIERONDER ZIE JE DE SAMENVATTENDE GRAFIEKEN VAN "INTERGEMEENTELIJKE VERENIGING VENECO"

SCORE PER CATEGORIE

Deze grafiek toont de score per categorie in vergelijking met andere kmos in de hoofdsector "Openbaar bestuur en defensie; verplichte sociale verzekeringen".

Term	Website Score	Jaarverslag Score
Resource Use and Circular Economy	0/100	64.9/100
Own Workforce	0/100	62.8/100
Workers in the Value Chain	0/100	44.7/100
Affected Communities	0/100	94/100
Environment	0/100	49.6/100
Climate Change	0/100	76.5/100
Pollution	0/100	79.3/100
Water and Marine Resources	0/100	55.7/100
Biodiversity	0/100	90/100
Business Conduct	0/100	62.5/100
Governance, Risk Management, Internal Control	0/100	57.2/100
Consumers	0/100	38/100

## Informatie Bedrijf

### INTERGEMEENTELIJKE VERENIGING VENECO

<b>Venootschapsnaam</b>	INTERGEMEENTELIJKE VERENIGING VENECO
<b>Ondernemingsnummer</b>	BE0200065765
<b>Telefoonnummer</b>	+32 9 343 81 57
<b>Email</b>	<a href="mailto:info@veneco.be">info@veneco.be</a>
<b>Adres</b>	Panhuysstraat 1 9070, Destelbergen
<b>Website</b>	<a href="https://www.veneco.be">https://www.veneco.be</a>
<b>Jaarverslag</b>	<a href="#">Download Jaarverslag</a>
<b>Omzetcijfer</b>	€ 5439
<b>Balanstotaal</b>	€ 78782159
<b>beursgenoteerd</b>	Niet Beursgenoteerd
<b>Aantal werknemers</b>	27
<b>Hoofdsector</b>	Openbaar bestuur en defensie; verplichte sociale verzekeringen
<b>Subsector</b>	Openbaar bestuur op het gebied van het bedrijfsleven; stimuleren van het bedrijfsleven



## Scores

---

### Scores

---

Totale score: 32.3 punten

Websitescore: Geen score gevonden

Jaarverslagscore: 64.6 punten

## Subscores

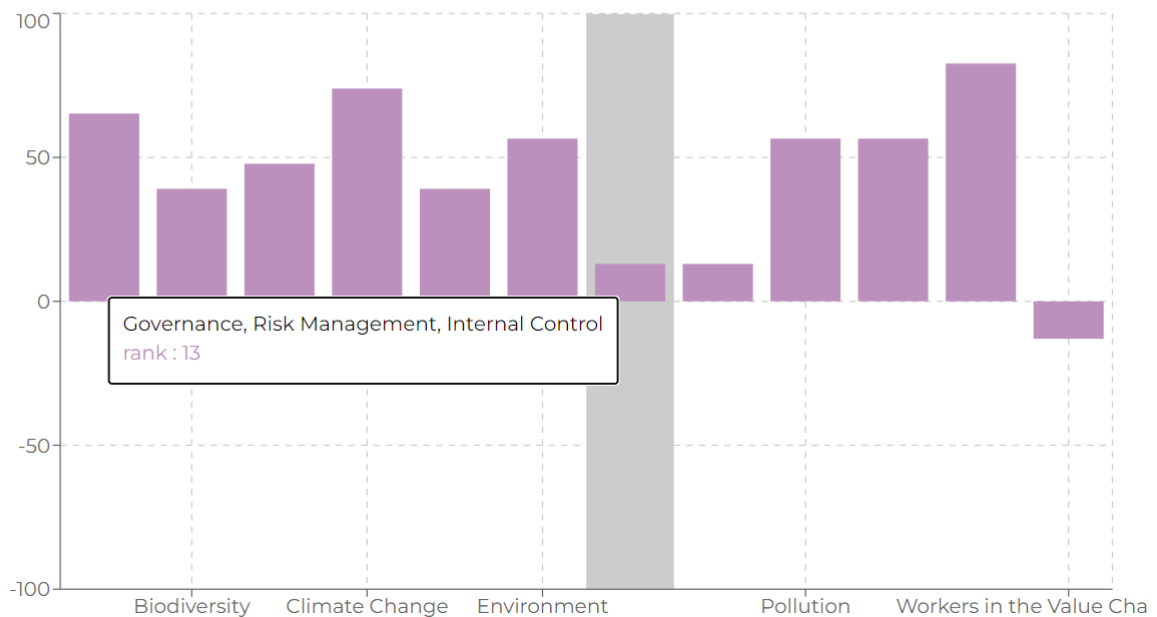
Term	Website Score	Jaarverslag Score
Resource Use and Circular Economy	0/100	64.9/100
Own Workforce	0/100	62.8/100
Workers in the Value Chain	0/100	44.7/100
Affected Communities	0/100	94/100
Environment	0/100	49.6/100
Climate Change	0/100	76.5/100
Pollution	0/100	79.3/100
Water and Marine Resources	0/100	55.7/100
Biodiversity	0/100	90/100
Business Conduct	0/100	62.5/100
Governance, Risk Management, Internal Control	0/100	57.2/100
Consumers	0/100	38/100

## Grafieken – Vergelijking met andere bedrijven in de sector

HIERONDER ZIE JE DE SAMENVATTENDE GRAFIEKEN VAN "INTERGEMEENTELIJKE VERENIGING VENECO"

### SCORE PER CATEGORIE

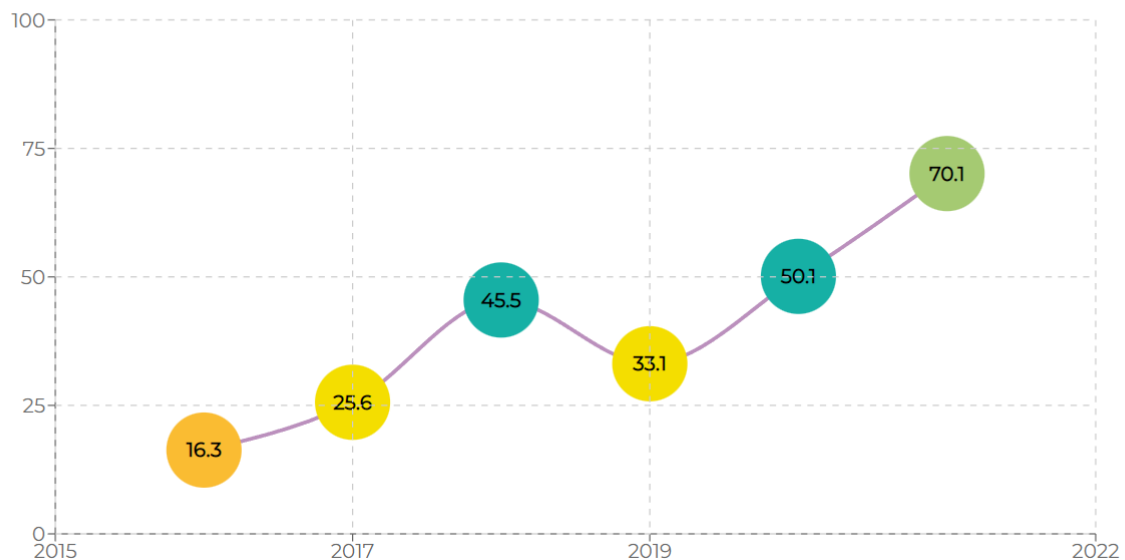
Deze grafiek toont de score per categorie in vergelijking met andere kmos in de hoofdsector "Openbaar bestuur en defensie; verplichte sociale verzekeringen".



## Grafieken – Geschiedenis

### SCORE GESCHIEDENIS

Een geschiedenis van de scores van afgelopen jaren.



## Sector overzicht

Vanop het dashboard kan je doorklikken naar “het sectoroverzicht” hier krijg je een overzicht van al de sectoren. Je kan zoeken en sorteren op hun naam.

Welke hoofdsectoren presteren het best?		
N°	Sector	Gemiddelde Score
1	Winning van delfstoffen	31.6
2	Openbaar bestuur en defensie; verplichte sociale verzekeringen	23.7
3	Exploitatie van en handel in onroerend goed	16.1
4	Informatie en communicatie	15.8
5	Productie en distributie van elektriciteit, gas, stoom en gekoelde lucht	15.6
6	Distributie van water; afval- en afvalwaterbeheer en sanering	14.9
7	Financiële activiteiten en verzekeringen	14.9
8	Vrije beroepen en wetenschappelijke en technische activiteiten	14.7
9	Onderwijs	14.7
10	Administratieve en ondersteunende diensten	14.5
11	Menselijke gezondheidszorg en maatschappelijke dienstverlening	13.9
12	Kunst, amusement en recreatie	13.4
13	Overige diensten	13.2
14	Groot- en detailhandel; reparatie van auto's en motorfietsen	13.1
15	Vervoer en opslag	12.7
16	Industrie	12.7
17	Verschaffen van accommodatie en maaltijden	12.1
18	Huishoudens als werkgever; niet-gedifferentieerde productie van goederen en diensten door huishoudens voor eigen gebruik	12
19	Bouwnijverheid	11.7
20	Landbouw, bosbouw en visserij	11.6

## KMO overzicht

Op het sector overzicht kan je weer doorklikken naar het KMO overzicht. Hier krijg je een overzicht hoe de KMO's presteren onder deze bepaalde sector.

HOAGENT  
Data Engineering Project

MENU

Uitleg

Dashboard

Sectoren

MODERATOR TOOLS

Zoektermen

Voorspellen

ADMIN TOOLS

Rollen

Productie van elektriciteit

BESTE KMO'S IN 'PRODUCTIE VAN ELEKTRICITEIT'

N°	KMO	Score
1	BE0445389356 - ECOPOWER - Antwerpen	42.7
2	BE0552775977 - ELICIO - Oostende	37.9
3	BE0819431848 - ENECO SOLAR BELGIUM - Gent	21.6
4	BE0861775912 - BIOPOWER OOSTENDE - Oostende	14.4
5	BE0883912203 - BIOSTOOM OOSTENDE - Oostende	13.8
6	BE0891347450 - A&S ENERGIE - Oostrozebeke	13.5
7	BE0471509971 - ROVAK - Hoogstraten	8.5

## Moderator Tools - Zoektermen

Wanneer je aangemeld bent als moderator of admin, krijg je speciale rechten. De moderator kan de zoektermen en de woorden wijzigen. Alsook de scores herberekenen.

HOAGENT  
Data Engineering Project

MENU

Uitleg  
Dashboard  
Sectoren

MODERATOR TOOLS

Zoektermen  
Voorspellen

ADMIN TOOLS

Rollen

Zoektermen

DE ZOEKWOORDEN VAN "ENVIRONMENT"

Ga Terug

Nieuw woord... +

Woord	Acties
impact op het milieu	
milieu	
milieucriteria	
milieu-impact	
milieunormen	
milieubeleid	
planeet	

HOAGENT  
Data Engineering Project

MENU

Uitleg  
Dashboard  
Sectoren

MODERATOR TOOLS

Zoektermen  
Voorspellen

ADMIN TOOLS

Rollen

Zoektermen

OVERZICHT VAN ALLE ZOEKTERMEN.

Hoofdtitel	Subtitel	Acties
Environment		
	Climate Change	
	Pollution	
	Water and Marine Resources	
	Biodiversity	
	Resource Use and Circular Economy	
Social		
	Own Workforce	
	Workers in the Value Chain	
	Affected Communities	
	Consumers	
Governance		
	Business Conduct	
	Governance, Risk Management, Internal Control	

Herberekenen van de scores

Door op de knop hieronder te klikken kan je al de scores herberekenen. Dit kan enkele minuten duren. De scores zullen enkel veranderen indien er een aanpassing is geweest in de woorden. Indien er geen aanpassing is geweest, zullen de scores niet veranderen. Gebruik deze knop niet te veel want de berekening is vrij intensief.

Scores Herberekenen

## Moderator Tools - Voorspellen

Het is mogelijk om als moderator of admin in de applicatie een voorspelling te doen naar een score van een kmo. Er moeten bepaalde gegevens ingegeven worden en het machine learning model zal dan een corresponderende score terug geven.

**HOGENT**  
Data Engineering Project

MENU

Uitleg

Dashboard

Sectoren

MODERATOR TOOLS

Zoektermen

Voorstellen

ADMIN TOOLS

Rollen

**Voorstellen**

Voorstel de scores van een onbestaand bedrijf.

Vul de gegevens van een bedrijf in en voorspel de score van het bedrijf. Dit bedrijf hoeft niet te bestaan.

VOORSPELLING

AANTAL WERKNEMERS\*

OMZET\*

BALANSTOTAAL\*

VERSTEDELIJKINGSGRAAD\*

Landelijk

HOOFDSECTOR\*

-- Selecteer --

Voorstel

## Admin Tools – Rollen wijzigen

Als je een administrator bent van de applicatie kan je de rollen wijzigen van de andere gebruikers en hun meer of minder rechten geven. Het enige wat je moet doen is het emailadres opzoeken en de juiste rol toevoegen.

### Beheer Rollen



BEHEER DE ROLLEN VAN DE GEBRUIKERS

EMAIL GEBRUIKER

siebe.vancampenhout@student.hogent.be

Zoeken

#### Gevonden Gebruiker

Naam: Siebe Van Campenhout

Email: siebe.vancampenhout@student.hogent.be

Huidige rollen: standaard, admin, moderator

TE VERWIJDEREN/TOE TE VOEGEN ROL

admin, moderator, ...

Aanpassen

## Ethische Reflectie

Ethiek is een belangrijk onderwerp in de wereld van technologie en informatie. Het gaat om het bepalen van wat juist en verkeerd is in onze omgang met technologie en informatie, en hoe we ervoor zorgen dat we ons aan deze standaarden houden. Dit is vooral relevant als het gaat om het maken van API-verzoeken, webscraping en het gebruik van data.

API-verzoeken zijn verzoeken die we sturen naar een server om informatie op te vragen of om bepaalde acties uit te voeren. Het is belangrijk om ethisch te handelen bij het maken van deze verzoeken, omdat we toegang hebben tot privé-informatie en acties die we niet zonder toestemming zouden mogen uitvoeren. Dit betekent dat we ons aan de voorwaarden en beleid van de server moeten houden, en dat we ons moeten realiseren dat we verantwoordelijk zijn voor het juist gebruik van de informatie die we opvragen.

We hebben gebruik gemaakt van de API server van de nationale bank van België. Hierdoor moesten we minder requests sturen naar de server dan dat we anders hadden moeten sturen indien we een “websurfer” als scraper gebruikten, zoals Selenium. We zorgden er ook voor dat we genoeg tijd tussen al de verschillende request lieten zodat we geen DDOS-attack veroorzaakten.

Webscraping is het automatisch verzamelen van informatie van websites. Dit kan ethische vragen oproepen, omdat we informatie verzamelen zonder toestemming van de eigenaar van de website. Het is belangrijk om ervoor te zorgen dat we geen inbreuk maken op de auteursrechten van de website, en dat we ons aan de voorwaarden en beleid van de website houden. We moeten ook zorgvuldig overwegen of het verzamelen van deze informatie een waarde toevoegt aan ons doel, en of er geen andere manieren zijn om deze informatie te verkrijgen.

Maar aangezien al de data openbaar staat moeten we ons niet veel zorgen maken. Het enige waar we op moesten letten is dat we opnieuw geen DDOS-attack veroorzaakten waardoor de website van de KMO's tijdelijk niet meer werkten. We lieten opnieuw genoeg tijd tussen de verschillende soorten requests, en zorgden ervoor dat we met zo weinig mogelijk requests al de data konden ophalen.

We maakten ook gebruik van artificiële intelligentie, dus is het onze plicht om dit model uit te leggen. Het model dat wij gemaakt hebben houdt rekening met al de gegevens die wij verkregen hebben via de scrapers, en zal dus ook afhankelijk zijn van deze gegevens. Momenteel baseert het model zich te veel op de omzet, en te weinig op sectoren. Dit kan te maken hebben met dat de sectoren weinig correlatie hebben met de scores die de bedrijven behalen.



In samenvatting is het belangrijk om ethisch te handelen bij het maken van API-verzoeken, webscraping en het gebruik van data. Dit betekent dat we ons aan de voorwaarden en beleid van de servers en websites houden, dat we zorgvuldig overwegen of ons doel gerechtvaardigd is

## Repositories

Frontend van het project:

[https://github.com/SiebeVCHoGent/P2DA2022\\_G3\\_Front-end](https://github.com/SiebeVCHoGent/P2DA2022_G3_Front-end)

De volledige frontend van het project zit in deze repository. Deze is gemaakt in react.js. Als je de frontend wil runnen moet je zien dat de backend runt, en daarna voer je “yarn install” en “yarn start” uit.

Backend van het project:

<https://github.com/SiebeVCHoGent/DEP2Backend>

Om de backend te runnen moet je eerst al de python packages installeren via “pip install -r requirements.txt”. Daarna kan je de backend runnen met het commando “uvicorn app.main.main:app”. Deze start een webserver op port 8000.

Scratchpad en Scraper:

<https://github.com/SiebeVCHoGent/DataEngineeringP2>

Dit zijn allemaal verschillende files die we gebruikt hebben om de scraper te maken en om andere toepassingen te maken. (bv het koppelen van geopoint, koppelen NACE codes, ...)

Link naar up-to-date Trello-bord:

<https://trello.com/b/aQvFVTpn/date-engineering-project>