

XXX

Projectbeschrijving en -verantwoording, v. 0.9



let's change
YOU. US. THE WORLD.

Inhoudsopgave

1 Inleiding	3
1.1 Wat hebben opleidingen en studenten aan dit onderzoeksproject?	3
1.2 Wat heeft De HHs aan dit onderzoeksproject?	3
1.3 Positie van het project in de onderzoekslijnen van het lectoraat, het kenniscentrum en De HHs	3
1.4 Doel en opbouw van dit document	3
2 Onderbouwing van het belang	4
2.1 Theoretische onderbouwing	4
2.2 Ethische overwegingen	4
3 Methode	5
3.1 Onderzoeksvragen en operationalisering	5
3.2 Afbakening	5
4 Databronnen	6
4.1 Toelichting op bewerkingen per bron	6
4.1.1 Schematische weergave bronnen en koppelingen	6
5 Proces van levering en bewerking	7
5.1 Aanvraag aan levering of download	7
5.2 Datamanagement	7
5.3 Bewerking en verrijking	7
6 Methoden van analyse	8
7 Planning	9
8 Verwachte resultaten en beperkingen	10
8.1 Resultaten	10
8.2 Beperkingen	10
9 Reproduceerbaarheid	11
10 Budget en financiering	12
Referenties	13
Versiegeschiedenis	14
Repository	14

Bijlage 1 - Data Science Ethics Checklist	15
10.1 A. Dataverzameling	15
10.2 B. Data-opslag	15
10.3 C. Analyse	16
10.4 D. Modelleren	17
10.5 E. Inzet	18

1 Inleiding

Het onderzoeksproject [Naam] van het lectoraat Learning Technology & Analytics (LTA) heeft tot doel

Het onderzoeksproject wordt uitgevoerd door [naam + functie + contactgegevens onderzoeker(s)] en begeleid door de lector, Dr. Theo C. Bakker (t.c.bakker@hhs.nl).

1.1 Wat hebben opleidingen en studenten aan dit onderzoeksproject?

Studenten ...

Opleidingen

1.2 Wat heeft De HHs aan dit onderzoeksproject?

Dit onderzoeksproject draagt bij aan het instellingsplan van De HHs '[Onderzoekend leren met impact](#)'. In het bijzonder aan de strategische thema's ...

1.3 Positie van het project in de onderzoekslijnen van het lectoraat, het kenniscentrum en De HHs

Het project draagt bij aan de onderzoekslijnen ... van het lectoraat. ...

Het onderzoek valt verder in de onderzoekslijn Transformative Technology van het Kenniscentrum Global & Inclusive Learning en sluit vanwege het snijvlak Machine Learning en gelijke kansen aan bij de onderzoeksthema's Digitale Toekomst en Rechtvaardige Samenleving van De HHs.

1.4 Doel en opbouw van dit document

Dit document beschrijft de verdere onderbouwing van het belang van het project (hoofdstuk 2), de operationalisering en onderzoeksvragen van het project (hoofdstuk 3), de bronnen die we daarvoor gebruiken (hoofdstuk 4), het proces voor levering en bewerking (hoofdstuk 5), de methoden van analyse (hoofdstuk 6), de planning (hoofdstuk 7), de verwachte resultaten en eventuele beperkingen (hoofdstuk 8), de reproduceerbaarheid van het onderzoek (hoofdstuk 9) en tot slot het budget en financiering (hoofdstuk 10).

2 Onderbouwing van het belang

Dit hoofdstuk bevat een theoretische en ethische onderbouwing van het belang van het onderzoeksproject.

2.1 Theoretische onderbouwing

...

2.2 Ethische overwegingen

Vanuit het principe van academische integriteit is dit document opgesteld om transparant te zijn in het onderzoek en verantwoording af te leggen over de verzamelde data en toegepaste methoden. In deze paragraaf lichten we op hoofdlijnen toe welke ethische overwegingen zijn gemaakt. In bijlage 1 is aan de hand van de [deon checklist voor data scientists](#) een breed scala aan ethische vragen behandeld.

...

3 Methode

3.1 Onderzoeksvragen en operationalisering

[Beschrijf de onderzoeksvragen en de operationalisering daarvan.]

...

3.2 Afbakening

Dit onderzoek beperkt zich tot...

4 Databronnen

...

4.1 Toelichting op bewerkingen per bron

...

4.1.1 Schematische weergave bronnen en koppelingen

Samenvattend zijn de bronnen en koppelingen als volgt geordend:

...

5 Proces van levering en bewerking

5.1 Aanvraag aan levering of download

...

5.2 Datamanagement

...

5.3 Bewerking en verrijking

...

6 Methoden van analyse

In dit hoofdstuk beschrijven we per onderzoeksvraag van het onderzoek de methode van analyse.

...

7 Planning

In dit hoofdstuk beschrijven we per onderdeel van het onderzoek de planning van activiteiten en resulterende producten.

[Deel de planning in met een tabel met de volgende kolommen Datum, Activiteit, Door/met wie, Product]

...

8 Verwachte resultaten en beperkingen

8.1 Resultaten

De resultaten van het onderzoek zullen op de volgende manieren beschikbaar worden gesteld:

[Beschrijf de resultaten die je verwacht, hoe je die ter beschikking gaat stellen en aan wie]

...

8.2 Beperkingen

Het onderzoek kent mogelijk de volgende beperkingen:

[Geef een opsomming van beperkingen en hoe daarmee wordt omgegaan]

...

9 Reproduceerbaarheid

Het onderzoek volgt de [FAIR principles](#) voor reproduceerbaarheid van onderzoeksresultaten: **F**indability (vindbaarheid), **A**ccessability (toegankelijkheid), **I**nteroperability (uitwisselbaarheid), en **R**euse (herbruikbaarheid) van digitale assets. We betrekken dit op broncode, metadata, data en beroepsproducten. ...

10 Budget en financiering

[Beschrijf per activiteit de verwachte inzet en hoe deze wordt bekostigd.]

Referenties

Versiegeschiedenis

- [voor elke versie een regel met de datum, de aanpassing en - indien van toepassing - de verspreiding]

Repository

De broncode voor dit document kan bewerkt worden via [GitHub](#). Leden van de kenniskring van het lectoraat kunnen op verzoek toegang krijgen en meewerken aan de inhoud.

Bijlage 1 - Data Science Ethics Checklist

ethics checklist deon

10.1 A. Dataverzameling

Aandachtspunt	Overwegingen
A.1 Informed consent	Als er menselijke proefpersonen zijn, hebben deze dan geïnformeerde toestemming gegeven, waarbij de proefpersonen er zelf voor kiezen en een duidelijk begrip hebben van het gegevensgebruik waarvoor ze toestemming geven? ...
A.2 Bias in data verzameling	Hebben we bronnen van vooringenomenheid overwogen die geïntroduceerd zouden kunnen worden tijdens het verzamelen van gegevens en het ontwerpen van enquêtes en hebben we stappen ondernomen om deze te beperken? Zie Databronnen
A.3 Beperkte onthulling van persoonlijk identificeerbare informatie	Hebben we manieren overwogen om de blootstelling van persoonlijk identificeerbare informatie (PII) te minimaliseren, bijvoorbeeld door anonimisering of door geen informatie te verzamelen die niet relevant is voor analyse? ...
A.4 Mitigatie van downstream bias	Hebben we manieren overwogen om downstreamresultaten te testen op bevooroordeelde uitkomsten (bijv. het verzamelen van gegevens over beschermde groepsstatus zoals etniciteit of geslacht)? ...

10.2 B. Data-opslag

Aandachtspunt	Overwegingen
B.1 Data beveiliging	<p>Hebben we een plan om gegevens te beschermen en te beveiligen (bijv. encryptie in rust en in transit, toegangscontroles op interne gebruikers en derde partijen, toegangslogs en up-to-date software)?</p> <p>De gegevens worden opgeslagen conform de richtlijnen van de hogeschoolbibliotheek van De HHs via SURF Research Drive. Eventuele overdracht van data vindt plaats met behulp van SURF Filesender, waarbij de bewijzen van transacties separaat worden opgeslagen. De code wordt separaat opgeslagen in een private omgeving van github voor leden van de kenniskring van het lectoraat Learning Technology & Analytics.</p> <p>Het versiebeheer op de software (R) vindt plaats via het renv package in R.</p>
B.2 Het recht om vergeten te worden	<p>Hebben we een mechanisme waarmee een persoon kan vragen dat diens persoonlijke gegevens worden verwijderd?</p> <p>Omdat de data is geanonimiseerd is het niet mogelijk om een student naderhand te verwijderen omdat de student niet geïdentificeerd kan worden.</p>
B.3 Data bewaar plan	<p>Is er een schema of plan om de gegevens te verwijderen als ze niet langer nodig zijn?</p> <p>De data wordt na afloop van het onderzoek en publicatie van de resultaten na 1 jaar gearchiveerd in DARK-store. De HHs heeft deze voorziening nog niet, maar hierover zal het lectoraat in gesprek gaan met de data stewards van de hogeschoolbibliotheek.</p>

10.3 C. Analyse

Aandachtspunt	Overwegingen
C.1 Missende perspectieven	<p>Hebben we geprobeerd om blinde vlekken in de analyse aan te pakken door samenwerking met relevante belanghebbenden (bijv. het controleren van aannames en het bespreken van implicaties met betrokken gemeenschappen en materiedeskundigen)?</p> <p>...</p>

Aandachtspunt	Overwegingen
C.2 Dataset bias	<p>Hebben we de gegevens onderzocht op mogelijke bronnen van vooroordelen en stappen ondernomen om deze vooroordelen te verminderen of aan te pakken (bijv. bestendiging van stereotypen, bevestigingsvooroordelen, onevenwichtige klassen of weggelaten beïnvloedende variabelen)?</p> <p>...</p>
C.3 Eerlijke representatie	<p>Zijn onze visualisaties, samenvattende statistieken en rapporten ontworpen om de onderliggende gegevens eerlijk weer te geven?</p> <p>...</p>
C.4 Privacy in de analyse	<p>Hebben we ervoor gezorgd dat gegevens met PII niet worden gebruikt of weergegeven tenzij dit noodzakelijk is voor de analyse?</p> <p>...</p>
C.5 Controleerbaarheid	<p>Is het proces voor het genereren van de analyse goed gedocumenteerd en reproduceerbaar als we in de toekomst problemen ontdekken?</p> <p>Ja. Zie hiervoor het hoofdstuk over Beperkingen</p>

10.4 D. Modelleren

Aandachtspunt	Overwegingen
D.1 Proxy discriminatie	<p>Hebben we ervoor gezorgd dat het model niet berust op variabelen of benaderingen voor variabelen die oneerlijk discriminerend zijn?</p> <p>...</p>
D.2 Eerlijkheid tussen groepen	<p>Hebben we de modelresultaten getest op eerlijkheid met betrekking tot verschillende getroffen groepen (bijv. getest op ongelijke foutpercentages)?</p> <p>...</p>
D.3 Variabele selectie	<p>Hebben we de effecten van het optimaliseren voor onze gedefinieerde variabelen overwogen en hebben we aanvullende variabelen overwogen?</p> <p>...</p>
D.4 Uitlegbaarheid	<p>Kunnen we in begrijpelijke termen een beslissing uitleggen die het model heeft genomen in gevallen waarin een rechtvaardiging nodig is?</p>

Aandachtspunt	Overwegingen
	...
D.5 Communicatie over bias	<p>Hebben we de tekortkomingen, beperkingen en vooroordelen van het model gecommuniceerd aan relevante belanghebbenden op een manier die algemeen begrepen kan worden?</p> <p>De beperkingen van het onderzoek zullen worden meegenomen in de rapportages/artikelen en verschillen vormen waarin we de uitkomsten van het onderzoek zullen bekend maken. Zie Planning</p>

10.5 E. Inzet

Aandachtspunt	Overwegingen
E.1 Monitoring en evaluatie	<p>Hoe zijn we van plan het model en de effecten ervan te bewaken nadat het is ingevoerd (bijv. prestatiebewaking, regelmatige controle van steekproefvoorspellingen, menselijke beoordeling van beslissingen die veel op het spel staan, beoordeling van downstreameffecten van fouten of beslissingen met een lage betrouwbaarheid, testen op concept drift)?</p> <p>...</p>
E.2 Het repareren en voorkomen van schade	<p>Hebben we met onze organisatie een plan van aanpak besproken voor het geval gebruikers schade ondervinden van de resultaten (bijv. hoe evalueert het datascience team deze gevallen en hoe werkt het analyses en modellen bij om schade in de toekomst te voorkomen)?</p> <p>...</p>
E.3 Terugdraaien	<p>Is er een manier om het model in productie uit te schakelen of terug te draaien als dat nodig is?</p> <p>...</p>
E.4 Onbedoeld gebruik	<p>Hebben we stappen ondernomen om onbedoeld gebruik en misbruik van het model te identificeren en te voorkomen en hebben we een plan om dit te controleren zodra het model is ingevoerd?</p> <p>Zie het antwoord bij vraag E.2</p>

Bron: Data Science Ethics Checklist, gemaakt met [deon](#) en vertaald door T. Bakker met behulp van Deepl; zie de site van deon voor de [originele Engelse versie](#).