

Projectplan

Digital Twin 3.0

Handpicked Agencies

Breda

Datum: 08-02-2021
Versie: 0.4
Auteur: Koen Pijnenburg

Introductie

Tijdens het achtste semester van de studie ICT & Software engineering zal mijn afstudeerstage plaatsvinden bij Handpicked Agencies. Het project waar ik aan zal werken is de digital twin genaamd [Twindle](#). Dit product is vorig jaar ontwikkeld door Handpicked Labs in samenwerking met Tectenna. Het doel van de applicatie was om de luchtkwaliteit, energieverbruik en brandveiligheid van gebouwen in kaart te brengen door middel van een digitale kopie van het gebouw.

Twindle meet, onder andere, de temperatuur, luchtvochtigheid en CO2 en bepaalt of er voldaan wordt aan de gestelde luchtkwaliteitseisen. Mede door COVID-19 is de luchtkwaliteit in een ruimte steeds belangrijker geworden. Slechte luchtkwaliteit kan ernstige gezondheidsproblemen veroorzaken. Door dit project uit te voeren willen we de leefbaarheid van ruimtes verbeteren.

Momenteel geeft Twindle alleen de huidige situatie weer. Om tijdig te kunnen ingrijpen bij verslechterende luchtkwaliteit zou ook de toekomstige situatie zichtbaar moeten zijn. Het is aan mij om technieken zoals machine learning toe te passen om dit te kunnen realiseren.

In dit document zal de basis voor dit project gelegd worden door middel van een projectplan waarin, onder andere, de projectopdracht, doelen en het tijdplan worden toegelicht.

Inhoudsopgave

1.	Projectopdracht	6
1.1	Context	6
1.2	Doelstelling	7
1.2.1	SMART	8
1.3	Begrenzing en Randvoorwaarden	10
1.4	Strategie	12
1.5	Onderzoeksvragen	13
1.5.1	Hoofdvraag	14
1.5.2	Deelvragen	14
1.6	Eindproducten	17
2.	Projectorganisatie	19
2.1	Teamleden	19
2.2	Communicatie	20
3.	Activiteiten en tijdplan	21
3.1	Aanpak	21
3.2	Tijdplan	22
4.	Testaanpak en Configuratiemanagement	23
4.1	Testaanpak/strategie	23
4.2	Testomgeving en benodigdheden	24
4.3	Configuratiemanagement	25
5.	Risicomanagement	26
6.	Bronnen	27

Versiebeheer

Versie	Datum	Wijzigingen
0.1	08-02-2021	Toegevoegd: <ul style="list-style-type: none">- 1.1 Context- 1.2 Doelstelling
0.1	09-02-2021	Toegevoegd: <ul style="list-style-type: none">- 1.3 Begrenzing en Randvoorwaarde- 1.4 Strategie- 1.5 Onderzoeksvragen
0.1	10-02-2021	Toegevoegd: <ul style="list-style-type: none">- 2.1 Teamleaden- 2.2 Communicatie- 1.6 Eindproducten
0.1	11-02-2021	Toegevoegd: <ul style="list-style-type: none">- 3.1 Opdeling en aanpak- 3.2 Tijdplan
0.1	15-02-2021	Toegevoegd: <ul style="list-style-type: none">- Introducties Aangepast: <ul style="list-style-type: none">- 1.2 Doelstellingen -> SMART methode.- 1.6 Eindproducten -> toelichtingen.- PBS Business Proposal -> Project plan- PBS toelichting toegevoegd.
0.1	17-02-2021	Toegevoegd: <ul style="list-style-type: none">- 4.1 Teststrategie
0.1	18-02	Toegevoegd: <ul style="list-style-type: none">- 5: Risicomanagement Afronding voor eerste feedbackmoment
0.2	02-03-2021	Feedback toegepast: <ul style="list-style-type: none">- Probleemstelling in introductie- Hoofd- en deelvragen herzien<ul style="list-style-type: none">- Specifieker gemaakt- Zijn er ethische... -> Welke ethische...
0.2	03-03-2021	Feedback toegepast: <ul style="list-style-type: none">- 1.6 Model evaluatie beschreven.- 1.6 Integratie verduidelijkt.- 3.2 Producten en onderzoeken aan planning gekoppeld.
0.2	04-03-2021	Feedback toegepast:

		<ul style="list-style-type: none"> - 4.1 Test uitleg toegevoegd aan 4.2 - 4.3 Versiebeheer, deployment configuratie, change management toegevoegd.
0.2	05-03-2021	Aangepast: <ul style="list-style-type: none"> - 1.3 Begrenzing aan de hand van story map. - 1.3 SW aspecten uitgebreid. - 1.3 Gelimiteerd react kennis verplaats naar risico's - 1.4 naar hoofdstuk 3.1 verplaatst
0.3	09-03-2021	Aangepast: <ul style="list-style-type: none"> - Deelvraag 4 - Scope uitgebreid met data- verzameling en voorbereiding. - Scope randvoorwaarden verduidelijkt. - SMART aangepast; luchtkwaliteit kennis is er nog niet. - Deelvraag 3 specifiek op luchtkwaliteit gericht.
0.4	22-03-2021	Aangepast: <ul style="list-style-type: none"> - Twindle bepaald -> bepaalt - Bronnen hoofdstuk nummer verwijderd - Oplever datum jaartal toegevoegd

1 Projectopdracht

Tijdens de stage zal gewerkt worden aan het afstudeerproject. In dit hoofdstuk wordt dit project verduidelijkt door, onder andere, de context te beschrijven, doelen te stellen en hoofd- en deelvragen op te stellen.

1.1 Context

Bedrijf

Het afstudeerproject zal ontwikkeld worden voor Handpicked Agencies. Dit is een samenvoeging van zeven bedrijven gevestigd in Breda en Eindhoven met ongeveer 170 personen in dienst.

Binnen Handpicked Agencies is er een team, Handpicked Labs, wat werkt aan het onderzoeken van nieuwe trends, technieken en innovaties. Tijdens mijn stage zal ik onderdeel van dit team uitmaken.

Structuur

Ieder bedrijf binnen Handpicked Agencies is gericht op een ander vakgebied. Per bedrijf zal kort toegelicht worden waarin zij zich specialiseren en voor wie zij werken of welke producten er ontwikkeld zijn.

1. Bluebird Day; Innovatieve e-commerce voor, bijvoorbeeld, Nespresso, JohnBeerens.com en VIA VAI.
2. Boldy XR; AR & VR experiences zoals 360° live streams, fotografie en AR product visualisaties.
3. E-sites; Web- en app ontwikkeling voor, bijvoorbeeld, CZ, Amarant en Paaspop.
4. Fingerspitz; Datagedreven digitale marketing voor, onder andere, Hubo, Bruynzeel en Kenwood.
5. TDE; Digitale sportmarketing door middel van concepting & strategy, UX design, customer profiling en meer.
6. Twentyseven; Enterprise technologie en digitale transformaties.
7. Weekend; Branding en strategy voor, onder andere, Heineken en Nespresso.

1.2 Doelstelling

Momenteel wordt er veel data verzameld door Twindle die betrekking heeft op de luchtkwaliteit. Dit zijn meetwaarden zoals temperatuur, CO2, luchtvochtigheid en dergelijke. Het doel is om voor deze meetwaarde een voorspellend model te ontwikkelen zodat slechte luchtkwaliteit voorkomen kan worden.

In de onderstaande afbeelding kan een mock-up gevonden worden van de potentiële eindsituatie.



Afbeelding 1: Uitbreiding Twindle Mockup

De voornaamste aanpassing is de toevoeging van een tijdlijn. Door middel van deze tijdlijn moeten gebruikers in staat gesteld worden om, onder andere, waarschuwingen, metingen en opmerkingen uit het verleden en toekomst in te kunnen zien.

Om deze eindsituatie effectief te realiseren zullen doelen worden opgesteld. Dit zal gebeuren via de SMART-methode (Benders, 2020) om tot concrete en haalbare doelen te komen.

1.2.1 SMART

Met de gemeten data is meest toepasselijk om beter inzicht te geven in toekomstige luchtkwaliteit. De volgende doelstelling is opgesteld:

1. Vanaf 18-06-2021 zullen de gebruikers van Twindle meldingen kunnen ontvangen wanneer slechte luchtkwaliteit wordt verwacht. Het machine learning model wat hiervoor wordt toegepast zal een “recall” en “precision” score hebben van minimaal 95%.

Specifiek

Vragen	Antwoorden
Wat wil je bereiken?	Twindle uitbreiden zodat de toekomstige luchtkwaliteit van een ruimte inzichtelijk gemaakt wordt. Deze inzichten kunnen gebruikt worden om problemen die door slechte luchtkwaliteit veroorzaakt worden te voorkomen.
Wie zijn erbij betrokken?	Stagiaire, Handpicked Labs team & Techtenna.
Wanneer vindt het plaats?	08-02-21 t/m 09-07-2021
Waarom wil je dit bereiken?	Mede door COVID-19 is de luchtkwaliteit in een ruimte belangrijker geworden. Slechte luchtkwaliteit kan ernstige gezondheidsproblemen veroorzaken. Door dit project uit te voeren willen we de leefbaarheid van ruimtes verbeteren.

Meetbaar

Vragen	Antwoorden
Wat ga je doen?	Een systeem ontwikkelen wat een melding geeft wanneer verwacht wordt dat de luchtkwaliteit niet voldoet aan de eisen.
Hoe kun je dit meten?	Door te berekenen wanneer de luchtkwaliteit niet voldeed aan de eisen en hier een model voor te ontwikkelen wat 95% accuraat is.
Welk resultaat levert het op?	Beter inzicht- en controle over de luchtkwaliteit.

Acceptabel

Vragen	Antwoorden
Is het doel logisch?	Ja, luchtkwaliteit is, mede door COVID-19, een steeds belangrijker geworden.
Is er draagvlak voor het doel?	Ja, vanuit Handpicked Agencies was dit al eerder bedacht en uitgezet als een stageproject.

Realistisch

Vragen	Antwoorden
Is het plan uitvoerbaar?	De stappen in het plan, zie hoofdstuk 4.2, zijn besproken en aanvaard door de stagebegeleiders.
Zijn de kennis en middelen aanwezig?	Er is momenteel nog niet voldoende kennis van voorspelling van luchtkwaliteit. Dit houdt in dat er onderzoek gedaan moet worden naar, onder andere, de data benodigheden en validatie technieken van dit soort modellen.
Is het doel haalbaar?	Ja mits de kennis en/of data verzameld kan worden.

Tijdgebonden

Vragen	Antwoorden
Wanneer start het project?	08-02-21
Wanneer stopt het project?	09-07-2021
Wanneer is het doel behaald?	18-06-2021

1.3 Begrenzing en Randvoorwaarden

Samen met de projectbegeleiders vanuit Handpicked Agencies en Fontys zal vastgesteld worden welke onderdelen tot het project behoren en welke buiten de scope van het project vallen.

Tot het project behoort:	Tot het project behoort niet:
1. Exploratory data analyse (EDA)	1. Brandveiligheid modellen
2. Data verzameling.	
3. Data voorbereiding.	
4. Model experimenten voor: <ul style="list-style-type: none">- Luchtkwaliteit- Energiebesparing	
5.. Implementatie in Twindle d.m.v. <ul style="list-style-type: none">- Proof of Concept implementatie- Machine learning pipeline API	

** machine learning*

Randvoorwaarde

De onderdelen die tijdens het project gerealiseerd zullen worden kunnen randvoorwaarde zoals beschikbaarheid van data en technologie keuze bevatten. In dit gedeelte zullen deze per onderdeel toegelicht worden.

Exploratory data analyse

De huidige Twindle applicatie maakt gebruik van MongoDB. In deze database is ongeveer drie maanden aan gegevens opgeslagen. Deze zullen onderzocht worden in een exploratory data analyse door middel van Python Notebooks.

De inzichten uit de exploratory data analyse zullen gebruikt worden om meer inzicht in de applicatie te verkrijgen en verder onderzoek te verrichten naar de data requirements.

Data verzameling

De inzichten van de EDA zullen gebruikt worden om de verdere data requirements in kaart te brengen. Mocht er nog data verzameld moeten worden dan zal deze beschikbaar moeten zijn uit openbare of interne bronnen. Eventueel kan Tectenna worden gevraagd extra sensoren te plaatsen.

Data voorbereiding

De data vanuit de Twindle applicatie en extra bronnen zal worden voorbereid op het gebruik in modellen. Dit houdt in dat er data cleaning en feature engineering technieken toegepast zullen worden.

Model experimenten

Onderzoek zal verricht worden naar de meest geschikte modellen. Met deze modellen zal geëxperimenteerd worden in Python Notebooks gebruik makend van, onder andere, Sci-Kit, Pandas en Numpy.

De experimenten zullen met elkaar vergeleken worden door de resultaten te valideren en te vergelijken. Uiteindelijk zal dit resulteren in een model wat gekoppeld kan worden aan de bestaande Twindle applicatie

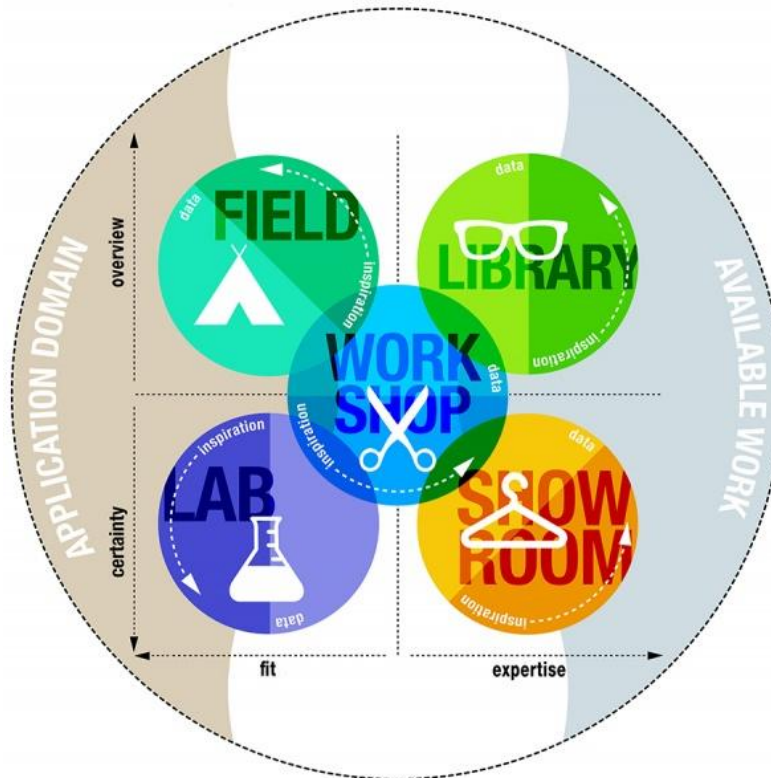
Implementatie in Twindle

De modellen zullen gebruikt worden om het Twindle product uit te breiden. Dit betekent dat het front- en backend uitgebreid dient te worden. Dit zal aan de volgende voorwaarden moeten voldoen:

- Backend implementatie; Machine learning API. Gebruik makend van Python.
- Frontend implementatie; Uitbreiding met tijdlijn. Ontwikkeld in React.
- Backend testing; Applicatie moet 90% test coverage hebben.

1.4 Onderzoeksvragen

Tijdens het gehele project zal onderzoek verricht worden volgens het [Development Oriented Triangulation \(DOT\) framework](#). Dit framework bestaat uit verschillende onderzoeks categorieën die gecombineerd dienen te worden om tot een valide conclusie te komen. In de onderstaande afbeelding valt een overzicht van het DOT-framework te zien.



Afbeelding 3: DOT-framework

1.4.1 Hoofdvraag

Hoe kan Twindle uitgebreidt worden om een hoge luchtkwaliteit in ruimten te garanderen?

1.4.2 Deelvragen

In dit hoofdstuk worden deelvragen opgesteld om zo de hoofdvraag te kunnen beantwoorden. Meerdere onderzoekstechnieken uit het DOT-framework zullen worden toegepast om tot een valide antwoord te komen.

1. Wie zijn de gebruikers van Twindle?

Om uit te vinden wie precies de gebruikers van Twindle zijn en wat zij belangrijk vinden kunnen Field, Library en Workshop methodes toegepast worden.

Field

Een stakeholders analyse zal uitgevoerd worden om alle belanghebbende en hun behoefte in kaart te brengen.

Library

Vanuit de stakeholders analyse kan een interview opgesteld worden met een domeinexpert. Hij kan, onder andere, richting geven aan het project en veel voorkomende valkuilen aanduiden.

Workshop

Een brainstorming sessie in de vorm van story mapping zal gebruikt worden om de behoefte van de gebruikers te verduidelijken en te prioriteren. Op deze manier wordt er voor gezorgd dat de meest belangrijke onderdelen als eerst worden toegevoegd.

2. Welke data is benodigd om luchtkwaliteit te voorspellen?

Er is historische data beschikbaar om modellen te trainen. Het zal waarschijnlijk zo zijn dat er externe databronnen gekoppeld moeten worden. Deze bronnen kunnen in kaart gebracht worden door Library en Lab onderzoekstechnieken toe te passen.

Library

Via een best- good- and bad practices onderzoek kan bekeken worden of de uitdagingen van het project al door andere projectgroepen zijn opgelost. Mocht dit zo zijn dan kunnen de gevonden succesfactoren worden doorgezet tijdens dit project.

Lab

Door middel van een exploratory data analysis kan onderzocht worden welke data er momenteel beschikbaar is en hoe deze toegepast kan worden om aan de behoefte van de stakeholder te voldoen.

3. Hoe kan machine learning gebruikt worden om luchtkwaliteit te voorspellen?

Wanneer de databehoeftes en machine learning targets zijn vastgesteld kan onderzocht worden welke modellen gebruikt kunnen worden en hoe zij moeten worden gevalideerd. Dit zal gedaan worden door Library, Workshop en Showroom te combineren

Library

Een combinatie tussen available product- en SWOT analyse kan uitgevoerd worden om te bepalen wat de mogelijkheden en hun sterke en zwakke punten zijn. Uit dit onderzoek zal een lijst met meest toepasbare modellen komen.

Workshop

Om te valideren dat de modellen naar behoren werken zijn zullen er prototypes worden ontwikkeld en worden gevalideerd. Per model zal onderzocht moeten welke validatie technieken toegepast moeten worden.

Showroom

Om te valideren dat het juiste probleem wordt aangepakt zal er een product review plaats vinden. Hierbij worden de bemoeide stakeholders een demo gepresenteerd en gevraagd om feedback.

4. Met welke ethische aspecten dient rekening gehouden te worden? En op welke manier?

Het ontwikkelen van applicaties en het gebruik van data kan ethische problemen opleveren. Deze moeten onderzocht worden en rekening mee gehouden worden tijdens de ontwikkeling en gebruik van het product.

De [TICT-tool](#) kan gebruikt worden om de impact op de gebruikers en eventueel andere stakeholders van applicaties in te schatten.

5. Hoe kunnen machine learning modellen gekoppeld worden aan Twindle?

Wanneer de machine learning modellen zijn ontwikkeld moeten deze geïntegreerd worden met de bestaande Twindle applicatie. Om dit succesvol uit te voeren zullen Field, Lab en Showroom technieken worden toegepast.

Field

Het bestaande domein moet worden onderzocht door middel van domain modelling. Hierbij kunnen bestaande modellen zoals klasse- en database diagrammen worden uitgebreid met de benodigde componenten.

Lab

De volgende stap is om functionele- en niet functionele tests te beschrijven voor de componenten die geïmplementeerd dienen te worden. Wanneer met de stakeholders is

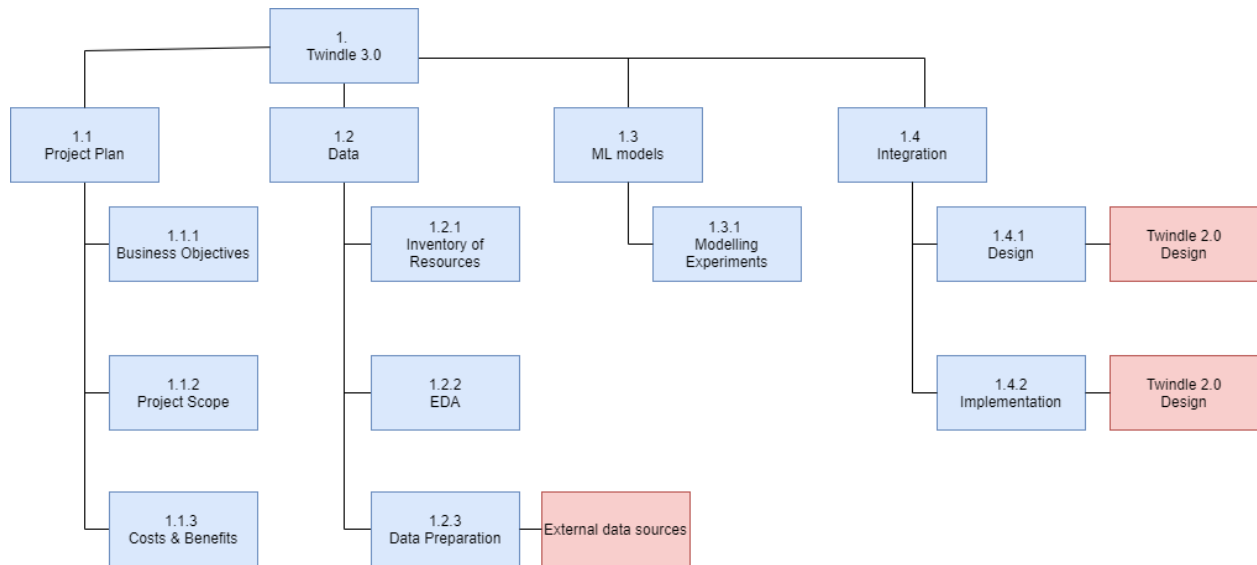
vastgesteld dat de tests goed zijn kunnen de componenten daadwerkelijk worden geïmplementeerd.

Showroom

De implementatie kan worden gevalideerd door code reviews en demo's te houden waarin wordt aangetoond dat aan de tests wordt voldaan.

1.5 Eindproducten

Tijdens het project zullen er meerdere eindproducten worden opgeleverd. Deze worden op hiërarchische wijze weergegeven in het onderstaande “*Product Breakdown Structure*” diagram (APM, z.d.).



Afbeelding 4: Product Breakdown Structure

Data

Het data gedeelte is opgedeeld in de volgende onderdelen:

- Inventory of resources; Een lijst met beschikbare data en een korte beschrijving.
- Exploratory data analysis; Een rapport waarin de data wordt bekeken en eventuele uitdagingen worden opgelost.
- Data preparation; De data wordt opgeschoont en voorbereid op de verwerking naar machine learning modellen.

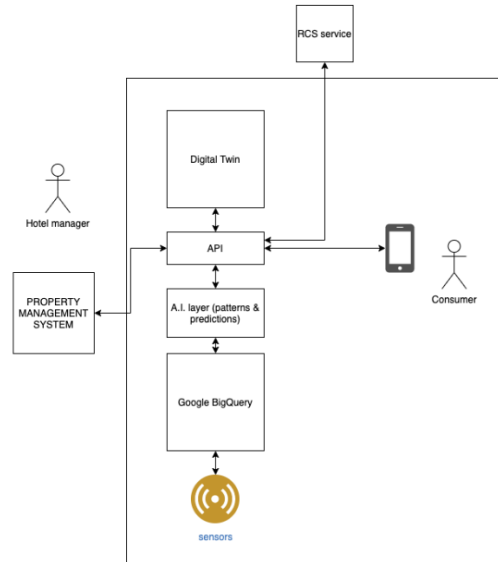
De resultaten van deze onderzoeken worden samengevoegd in een document waar alle inzichten en uitdagingen worden besproken.

Machine learning models

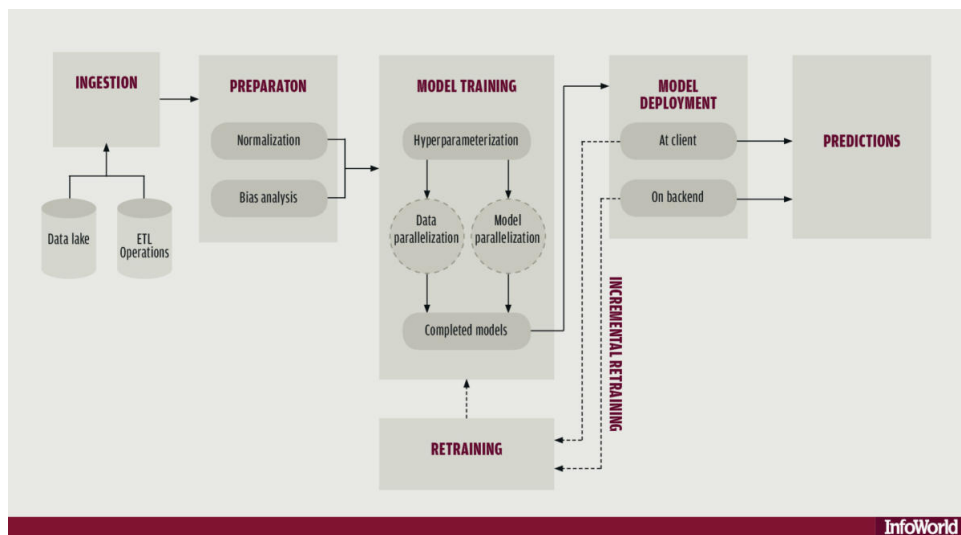
Om te bepalen welke machine learning modellen het beste presteren zullen experimenten in Python Notebooks worden uitgevoerd. Dit resulteert in een lijst met potentiële keuzes. Gebaseerd op gegevens zoals F1-, precision- en recall score zal geconcludeerd worden welke model het meest geschikt is en verder geïmplementeerd zal worden.

Integration

Wanneer een model ontwikkeld is wat aan de gestelde eisen voldoet zal deze worden geïntegreerd met de bestaande Twindle applicatie. Hiervoor zullen eerst de bestaande ontwerpen, zie onderstaande afbeelding, worden uitgebreid met de toe te voegen onderdelen. Na goedkeuring zullen deze worden geïmplementeerd.



De *A.I layer (predictions & patterns)* zal tijdens de stage ontwikkelt worden. Waarschijnlijk zal de volgende architectuur gebruikt worden. Deze code zal getest en gedeployed worden volgens de beschrijving in hoofdstuk 4.



2 Projectorganisatie

Tijdens het project zal er gecommuniceerd moeten worden met verschillende partijen. Om dit effectief te kunnen doen zullen de teamleden, hun rol bij het project en communicatiewijze worden vastgelegd.

2.1 Teamleden

In het onderstaande overzicht valt te zien welke personen er bij het project betrokken zijn, wat hun rol of taak is binnen het project en hoeveel dagen zij beschikbaar zijn voor het project.

Naam Tel Email	Rol/taken	Beschikbaarheid (dagen)
Koen Pijnenburg koen.pijnenburg@student.fontys.nl 06 260 492 74	Stagiaire	5
Bartosz Paskowski b.paszkowski@fontys.nl 0885072632	Docentbegeleider	0.5
Sjoerd van Oosten sjoerd@handpickedagencies.com 076 204 3063	Conceptueel begeleider Productowner	1
Samet Yilmaz syilmaz@handpickedagencies.com	Technisch begeleider	1

2.2 Communicatie

De communicatie tijdens het project kan opgedeeld worden in twee categorieën. Docentbegeleider en Stagebegeleiders. Met beide zal de communicatie op een andere manier verlopen.

Docentbegeleider

De stagiaire en docentbegeleider hebben iedere twee weken contact via Microsoft Teams. Tijdens deze meetings zijn de andere stagiaires die Bartosz begeleid aanwezig en zullen er korte demo's gegeven worden.

Daarnaast zal via Canvas contact zijn voor feedback en beoordelingen op gemaakte stukken.

Stagebegeleiders

De communicatie met de stagebegeleiders zal verlopen via Google Meet, E-mail en Slack . Iedere maandag is er een afspraak gepland om de vooruitgang van de voorgaande week te bespreken. Daarnaast is er iedere dag om 9:30 een stand-up meeting samen met de andere stagiaires en, afwisselend, Samet of Sjoerd.

Om de twee weken zullen er demo presentaties plaatsvinden met de andere stagiaires en stagebegeleiders. Tijdens deze demo's zal feedback gegeven en verzameld worden.

3.2 Aanpak

Het project zal de SCRUM (What is Scrum?, z.d.) aanpak hanteren. Dit betekent dat de SCRUM onderdelen, sprints, stand-ups, opleveringen, retrospectives en planningen, iteratief uitgevoerd zullen worden.

In het onderstaande overzicht kunnen deze onderdelen, een toelichting en hun frequentie gevonden worden.

Onderdeel	Toelichting	Frequentie
Sprint	Iedere twee weken begint een sprint. Het doel is om de sprint backlog uit te werken.	Continu
Stand-up	Om op de hoogte te blijven van andermans werk en eventueel uitdagingen te bespreken zijn er korte meetings met de andere stagiaires en begeleiders.	Iedere dag om 9:30
Sprint oplevering	Op het eind van een sprint wordt een demo gegeven van het gemaakte werk. Tijdens deze demos is er ruimte voor feedback van stakeholders.	Om de week op vrijdag om 9:30
Sprint retrospective	Na iedere sprint wordt bekeken wat er goed ging en wat er minder goed ging. Aan de hand van deze informatie kan de werkwijze worden aangepast.	Na review
Sprint planning	Samen met de product owner wordt bepaalt welke taken van de product backlog naar de sprint backlog worden geplaatst.	Na retrospective

3.3 Tijdplan

In het volgende hoofdstuk zal het tijdplan worden opgesteld. Dit wordt gedaan door iedere sprint, zoals beschreven in het voorgaande hoofdstuk, te voorzien van werkzaamheden zoals beschreven in *hoofdstuk 3.1 Strategie* en een start- en einddatum.

1. Van probleem naar aanpak.
2. Van behoeften naar verzameling.
3. Van begrip naar voorbereiding.
4. Van modelleren naar evaluatie.
5. Van deployment naar feedback.

Deze vijf stappen kunnen iteratief uitgevoerd worden. Het plan is om twee van deze iteraties uit te voeren over 7 sprints. Hierna is er eventueel nog ruimte om een derde iteratie gedeeltelijk uit te voeren.

Sprint verdeling

#	Fase	Onderzoeken	Producten	Start	Eind
0	Opstart		Projectplan	08-02	12-02
1.	1	Exploratory data analyse stakeholdersanalyse	EDA report Backlog	15-02	26-02
2.	2	Data requirements analyse	Data requirements report Dataset	01-03	12-03
3.	3 - 4	Modelling experimenten	Model prototype	15-03	26-03
4.	5	Model integratie	Ontwerp Proof of concept	29-03	9-04
5.	1 - 2	Model evaluatie Exploratory data analyse	Dataset	12-04	23-04
6.	3 - 4	Modellering experimenten	Model	26-04	07-05
7.	5	Machine learning pipeline architecture	ML pipeline API proof of concept.	10-05	21-05
8.			Portfolio	24-05	04-06
9.			Presentatie voor oplevering	07-06	18-06

4 Testaanpak en Configuratiemanagement

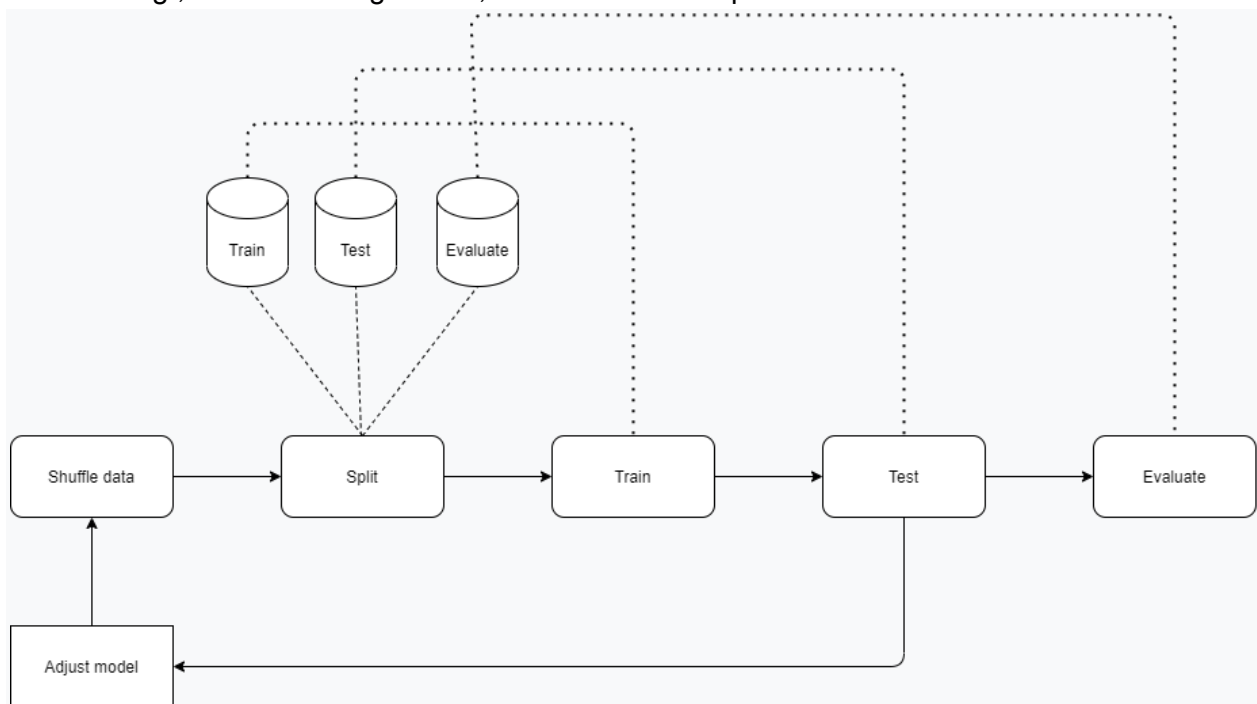
Om te verifiëren of de geschreven software juist werkt zullen tests worden uitgevoerd. Hoe deze tests worden opgezet zal in dit hoofdstuk worden toegelicht door een testaanpak vast te stellen en te bepalen wat voor deze aanpak benodigd is.

4.1 Testaanpak/strategie

Tijdens het project zullen twee verschillende teststrategieën worden gebruikt; Model- en integratie ontwikkeling.

Modelontwikkeling

De machine learning modellen die ontwikkeld gaan worden zullen worden geëvalueerd. Deze validatie volgt, in de meeste gevallen, het onderstaande proces.

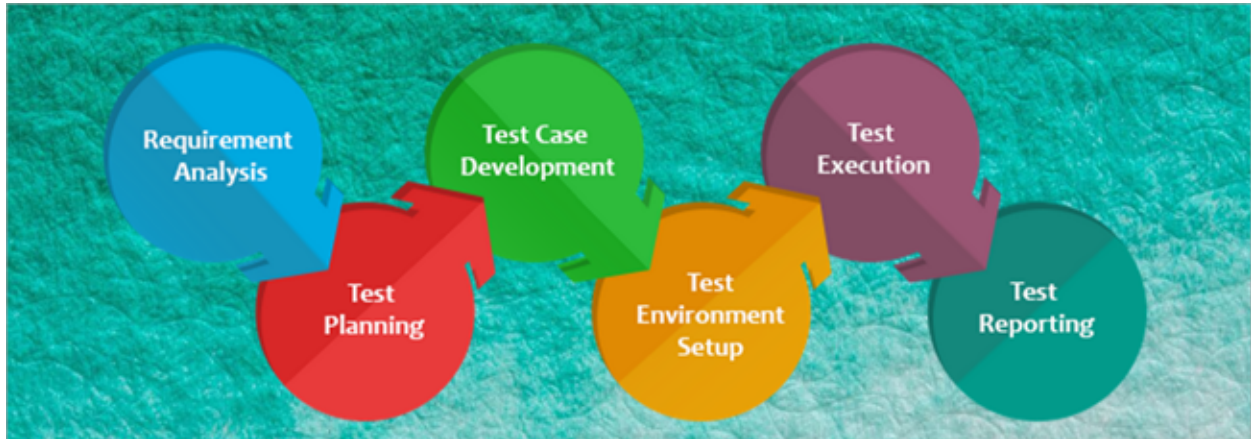


Afbeelding 5: Modelontwikkeling test strategie

In het bovenstaande diagram kan gezien worden dat ieder model eerst wordt getest. Het kan zijn dat het model nog niet aan de eisen voldoet. In dat geval wordt het model aangepast en opnieuw getraind. Dit kan herhaald worden totdat het aan de eisen voldoet waarna het nog een laatste keer wordt getest op de evaluatie set. Dit met de reden om problemen zoals under- en overfitting te voorkomen.

Integratie ontwikkeling

Voor de ontwikkeling van de integratie tussen de bestaande applicatie en de machine learning modellen zal de Software Testing Life Cycle (Software Testing Help, 2021) strategie gebruikt worden. Zie onderstaande afbeelding voor een overzicht van de onderdelen.

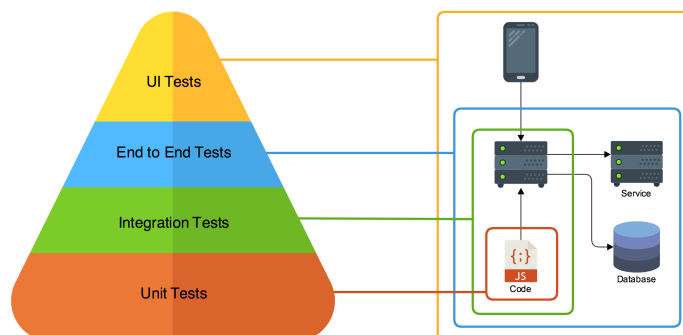


Afbeelding 6: Software Testing Life Cycle

Voor deze strategie is gekozen omdat het goed past in de agile werkhouding die voor de rest van het project wordt gebruikt. Iedere sprint kan de cyclus worden uitgevoerd. Dit leidt dan weer tot inzichten die tijdens de volgende sprint gebruikt kunnen worden.

4.2 Testomgeving en benodigdheden

Een testomgeving zal benodigd zijn voor de integratie van het machine learning model. Dit zal gedaan worden door automatisch unit- integratie en end-to-end tests uit te voeren. De onderstaande afbeelding geeft een overzicht van deze omgeving.



Afbeelding 7: Testomgeving

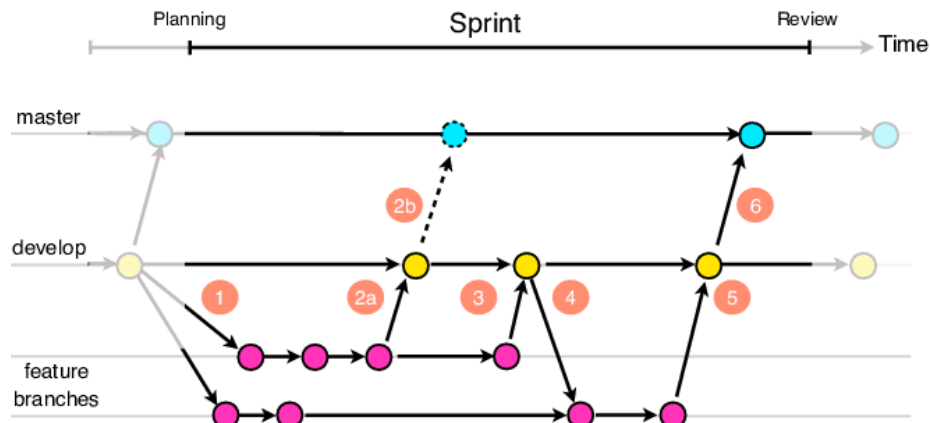
4.3 Configuratiemanagement

Tijdens het project zal gebruik gemaakt worden van versiebeheer, automatische deployments en dergelijke. In dit hoofdstuk zullen de configuraties van deze systemen worden toegelicht.

Binnen Handpicked Labs wordt de versiebeheer tool BitBucket gebruikt. Dit is een Git omgeving waar versiebeheer, testing en deployment samenkomen

Versiebeheer

De onderstaande branching strategie zal toegepast worden tijdens het project. Er is een master branch waar de huidige release op staat vanaf hier wordt een development branch afgeleid. Voor iedere feature die wordt geïmplementeerd wordt een branch gemaakt vanaf development. Wanneer de feature is afgerond wordt deze via een pull-request aan de development branch toegevoegd.



Afbeelding 8: Branching strategie

Deployments

Via BitBucket pipelines zal een automatische deployment straat worden ingericht. Zodra de development branch gemerged wordt naar de master branch zal deze uitgevoerd worden. Tijdens het ontwikkelen van de applicatie zal rekening gehouden worden met omgevings specifieke variabelen via configuratie bestanden.

5 Risicomanagement

In de loop van het project kan het zijn dat er bepaalde situaties voorkomen die invloed hebben op het project. In dit hoofdstuk worden deze situaties beschreven, hoe ze voorkomen kunnen worden en wat de afwijkactiviteit is.

Risico	Activiteiten ter voorkoming opgenomen in plan	Afwijkactiviteiten
1. Feature creep	Zorgen voor juiste verwachtingen bij de stakeholders van het project. D.m.v. dit project plan, story mapping, oplevering, etc.	Mochten er te veel features zijn die toegevoegd moeten worden zal er een meeting gehouden moeten worden waar het aantal features wordt beperkt.
2 Kwaliteitsbehoud	Door goed te testen, code reviews en demo's te houden kan de kwaliteit van het product bewaard worden.	Geplande features pauzeren en code kwaliteit verhogen.
3. Optimistische planning	Het is moeilijk in te schatten hoeveel werk het project daadwerkelijk kost. Om teleurstelling te voorkomen zal iedere sprint bekeken moeten worden of de planning haalbaar is.	Wanneer een planning niet haalbaar is zal met de stakeholders besproken moeten worden hoe de planning moet worden aangepast
4 Ziekte/uitval	n.v.t.	Er zal een vervanger zal gezocht moeten worden.
5. Gelimiteerde kennis React/JavaScript	Op tijd hulp inschakelen wanneer er wordt vastgelopen. Een korte herhalingscursus.	n.v.t.

Bronnen

IBM Analytics. (2015, juni). Foundational Methodology for Data Science (Nr. IMW14828-USEN-00). <https://tdwi.org/~media/64511A895D86457E964174EDC5C4C7B1.PDF>

Logallo, N. (2019, 30 december). Data Science Methodology 101 - Towards Data Science. Medium. <https://towardsdatascience.com/data-science-methodology-101-ce9f0d660336>

APM. (z.d.). Work & Product Breakdown Structure In Project Management | APM. Geraadpleegd op 10 februari 2021, van [https://www.apm.org.uk/resources/find-a-resource/use-of-product-breakdown-structures-and-work-breakdown-structures/#:%7E:text=A%20Product%20Breakdown%20Structure%20\(PBS,form%20of%20a%20hierarchical%20structure.](https://www.apm.org.uk/resources/find-a-resource/use-of-product-breakdown-structures-and-work-breakdown-structures/#:%7E:text=A%20Product%20Breakdown%20Structure%20(PBS,form%20of%20a%20hierarchical%20structure.)

What is Scrum? (z.d.). Scrum.org. Geraadpleegd op 11 februari 2021, van <https://www.scrum.org/resources/what-is-scrum>