

HOGESCHOOL ROTTERDAM / CMI

Project 3 – Data analyse

INFPRJ00-3



Aantal studiepunten: 4

Modulebeheerder: Stelian Paraschiv



Modulebeschrijving

Project 3 - Data analyse Modulenaam:

INFPRJ00-3 Modulecode:

Aantal studiepunten

studiebelastinguren:

Dit studieonderdeel levert je vier (4) studiepunten op, hetgeen overeenkomt met een studielast van 112 uren.

De aanbevolen verdeling van deze 112 uren over de collegeweken is als volgt:

Begeleide colleges:

Kick-off: 2 * 50 minuten 2 uur 8 * 180 minuten Gedurende 8 weken projectles: 24 uur 1 * 50 minuten Presentatie product: 1 uur

Onbegeleide uren per week:

Tijd werken aan project incl. literatuurstudie 85 uur Totaal 112 uur

Voor deze module is geen voorkennis vereist Vereiste voorkennis:

Projectonderwijs (groepswerk) Werkvorm:

Toetsing vindt plaats op basis van opgeleverd beroepsproduct en toetsing van Toetsina:

projectproces

Franken, M (2013), Scrum voor Dummies. (1e druk), Amsterdam: Pearson Leermiddelen:

Education Benelux B.V., ISBN 978-90-430-2403-7

Draagt bii aan competentie:

Leerdoelen:

Algemene beroepsvaardigheden, zie bijlage 2 modulewijzer INFSKL01-1:

Ondersteunen en samenwerken

Omgaan met anderen en presenteren

Analyseren en interpreteren

Creëren en conceptualiseren

Organiseren en uitvoeren

Je kunt in allerlei situaties feedback op de juiste wijze geven en ontvangen en omzetten in acties ten aanzien van eigen functioneren

Je kunt een korte overtuigende presentatie houden/instructie geven of kennis delen met aandacht voor doel(groep), structuur, inhoud, (non)verbale aspecten en hulpmiddelen

Je bent in staat de specifieke kennis uit je vakgebied eigen te maken via reguliere lessen en zo nodig via zelfstudie uit te breiden

Je bent in staat om met behulp van het creatieve proces een klantenvraag om te zetten in een creatieve, innovatieve oplossing die past bij de klantenvraag

Je kunt met een projectgroep een planning opstellen voor een complex project, een MoSCoW daarin meenemen en tussentijds de planning adequaat bijstellen zonder einddoel uit het oog te verliezen

Je bent in staat, met input van de Product Owner, een sprint - en product backlog op te stellen volgens de (Scrum) richtlijnen

Je bent in staat een sociale media API aan te roepen en deze informatie op te slaan in een database

Je kunt een (portal) applicatie maken met een front-end en een gekoppelde database waarmee data analyses uitgevoerd kunnen worden

Je leert in groepsverband werken (proces) en een projectopdracht voor een Inhoud: opdrachtgever realiseren (product). In een projectgroep ga je werken aan een

innovatieve opdracht die gericht is op het analyseren van sociale media gegevens voor de Gemeente Rotterdam. Je geeft advies over de mogelijkheden en beperkingen van deze innovatieve -, nieuwe techniek. Aan bod komen allerlei aspecten van het ontwikkel-/projectmethodiek Scrum.

Aanwezigheid is verplicht. Groepen worden door tutor tijdens de eerste les Opmerkingen:

samengesteld. 4 à 5 studenten per groep.

Modulebeheerder: Stelian Paraschiv

9 februari 2015 Datum:



1. Algemene omschrijving

1.1. Inleiding

Het thema van deze onderwijsperiode is 'analyse van social media data met andere datasets'. 'Data' staat officieel voor gegevensverzameling en 'analyseren' staat voor (iets) grondig onderzoeken en erover nadenken. Je gaat dus deze onderwijsperiode simpel gezegd gegevens verzamelen vanuit onder andere de sociale media en deze grondig onderzoeken. De reden voor de opleiding om voor dit onderwerp te kiezen is dat wij geloven dat het McKinsey rapport terecht stelt dat "big data de volgende stap is in innovatie, competitie en productiviteit". Verder voorspelt McKinsey dat er in de VS in 2018 meer dan 150.000 vacatures zullen zijn voor mensen met 'deep analytical talent' dat wil zeggen met kennis en vaardigheden op het gebied van data analyse¹.

Daarnaast heeft Facebook 1 miljard leden, waarvan er 552 miljoen dagelijks actief zijn op Facebook. Twitter heeft 517 miljoen leden, waarvan er 6,9 miljoen dagelijks actief zijn op Twitter². Momenteel worden elke 2,5 dag 1 miljard tweets verstuurd. Verder heeft Facebook 2,5 biljoen 'pieces of content' en meer dan 500 terrabytes aan data³. Deze data zijn ideaal om analyses op uit te voeren.

1.2. Relatie met andere onderwijseenheden

De skills in een project zijn onlosmakelijk verbonden met het project zelf. Denk daarbij aan samenwerken, leiding geven, overleggen, elkaar feedback geven en mondeling en schriftelijk communiceren. De flankerende modulen die je naast het project krijgt, hebben allemaal te maken met de kennis en vaardigheden die je in het project nodig hebt. Het onderwerp van het project is daarbij leidend. Bovendien trainen we je om een professionele beroepshouding te laten zien.

Wat je tijdens de flankerende modulen leert, pas je direct toe tijdens dit project. Hier kun je zien of je voldoende achtergrondkennis hebt of dat je je extra moet voorbereiden om deze module te kunnen volgen. De flankerende modulen zijn:

- <u>INFANL01-3</u>: Bij de module Analyse 3 leer je hoe je een informatiebehoefte vertaalt naar een gegevensmodel. Vervolgens wordt dit model geïmplementeerd in een (relationele) database waarin je data gaat toevoegen, wijzigen en verwijderen. Met deze leerstof doe je kennis en ervaring op voor het gebruik van relationele databases.
- INFDEV01-3: In deze module ga je verder met de concepten van objectgeoriënteerd programmeren. OO-programmeren is een belangrijke vaardigheid waar je je tijdens je studie in bekwaamt en dat vergt veel oefeningen. De meeste software die vandaag de dag beschikbaar is, wordt op deze manier gemaakt: van eenvoudige websites en de apps op je mobiel tot de interactieve televisie thuis. Het aanleren van het objectgeoriënteerde paradigma (manier van denken) geeft je dus de mogelijkheden om vrijwel alle denkbare software te schrijven
- INFSKL02-1 (OP3): Skills is een module die het gehele jaar naast de projecten loopt en waarin je dit kwartaal kennis maakt met het creatief proces van iets nieuws bedenken. Hoe pak je dat aan? En daarnaast oefen je om een goede presentatie te kunnen geven. Iets wat je veel zult doen in je opleiding, maar ook zeker daarna in de beroepspraktijk

1.3. Leermiddelen

Verplicht:

- Franken, M. (2013), Scrum voor Dummies (1e druk), Amsterdam: Pearson Education Benelux B.V., ISBN 978-90-430-2403-7
- Trello: https://trello.com (project Scrum)
- Box.com: https://box.com (project dossier)

Facultatieve bronnen:

https://confluence.hro.nl/display/CMIP/S1314+-+INFPRJ003+-+Project+3

¹ Big Data: The Next Frontier for Innovation. McKinsey Global Institute, Mei 2011. www.mckinsey.com/insights/mgi/research/

technology_and_innovation/big_data_the_next_frontier_for_innovation
2 www.frankwatching.com/wp-content/uploads/2012/05/Social-mediagebruik-wereldwijd-een-update-infographic.png

³ http://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day



2. Programma en inhoud

2.1.

Sociale media zijn niet meer weg te denken uit het dagelijks leven. Mensen twitteren en facebooken over dingen die ze tegenkomen en geven hun mening over gebeurtenissen. Op die manier verspreiden ze informatie naar het internet in vorm van berichten, foto's en video's. Dit is informatie die vaak beschikbaar wordt voor veel meer mensen dan alleen hun volgers of vrienden.

Rotterdam is de stad waar we allemaal iets mee hebben. De stad ligt aan de Nieuwe Maas en wordt gekenmerkt door veel moderne architectuur. Dat is vooral het gevolg van het vrijwel platbombarderen van de Rotterdamse binnenstad tijdens de Tweede Wereldoorlog. Vooral als je langs de kades van de Nieuwe Maas wandelt zie je hoe sterk de skyline van Rotterdam gedomineerd wordt door moderne bouwwerken. Rotterdam is door het deels ontbreken van een historisch centrum anders dan de gemiddelde West-Europese stad. De gedwongen nieuwe start van Rotterdam na de Tweede Wereldoorlog heeft ervoor gezorgd dat moderne architectuur volop de ruimte heeft gekregen binnen het hart van de stad. Dat heeft geleid tot een stad boordevol prachtige moderne gebouwen waar menig andere Europese stad een puntje aan kan zuigen.

Dat Rotterdam een mooie stad is, hoeven we dus niet in twijfel te trekken. De moderne bouwwerken zoals Centraal Station, Markthal, Blijdorp, Ahoy, Euromast, SS Rotterdam, Hotel New York, enz... zijn daar het zichtbare bewijs van. Wel vragen de beheerders van dergelijke gebouwen zich af hoe Rotterdammers, bezoekers en/of toeristen tegen deze gebouwen aankijken? Sociale media is daarbij een uitkomst. De gedeelde informatie kan opgezocht - , uitgelezen en geanalyseerd worden, om dergelijke vragen te kunnen beantwoorden.

De Product Owner (PO) vertegenwoordigt één van de moderne bouwwerken van Rotterdam. Samen met een marketingbureau zijn ze van plan een grote reclamecampagne op touw te zetten, met als doel meer bezoekers aan te trekken. Daarnaast willen ze de grootste ergernissen van bezoekers elimineren. Om dit voor elkaar te krijgen, willen ze antwoord krijgen op vragen als: "Wat zijn de grootste ergernissen van bezoekers? (sentiment analyse)", "Wat vinden bezoekers juist zo fijn aan het bouwwerk? (sentiment analyse)", "Hoe oud zijn de bezoekers? (demografische analyse)", "Zijn de meeste bezoekers vrouw of man? (demografische analyse)", "Waar komen de meeste bezoekers vandaan? (geografische analyse)" en "Hoe snel reageert het bedrijf op berichten?". De mening van mensen, en dus ook bezoekers, verandert constant. Om deze real-time informatie te structureren, wil de PO verschillende betrokkenen (eigenaar, admin, medewerkers, marketingbureau(s)) middels een portal hiertoe toegang geven. Zo kan een admin nieuwe gebruikers/gebouwen toevoegen en kunnen medewerkers rapporten uitdraaien. Met behulp van deze portal kan de PO niet alleen nu, maar ook in de toekomst inspelen op de behoeften van bezoekers.

Ook komt er steeds meer andere data (ook wel open data genoemd) beschikbaar via bijvoorbeeld overheden4. Nu zijn er initiatieven die beide databronnen combineren5. Zo zou het kunnen dat weerdata gecombineerd met tweets een voorspellende waarde kunnen hebben voor bezoekersaantallen. Ook dit is zeer interessant voor de PO.

2.2. **Opdracht**

Realiseer aan de hand van bovenstaande beschrijving een gebruiksvriendelijk systeem waarmee analyses uitgevoerd kunnen worden op zowel sociale media data en andere open data. Het systeem moet niet alleen nu -, maar ook in de toekomst inzetbaar zijn. Houd hier dus rekening mee!

Scrum planning richtlijnen

In project 3 ga je verder met het toepassen van Scrum. Na project 1 en project 2 heb je een basis gelegd en nu ga je verder met het opdoen van ervaring. Sprint 1 zou de research en analyse fase kunnen bevatten. Sprint 2 het ophalen van de juiste data en deze (ongestructureerd) opslaan. Sprint 3 het gestructureerd opslaan van data en deze visueel weergeven. Sprint 4 het rapporteren en dus beantwoorden van de uitgezette vragen. Elke sprint backlog wordt naar eigen inzicht samengesteld en wordt definitief gemaakt na afstemming en overeenstemming met PO.

⁴Alvarez-espinar, M. (2013). What is Open Data? Make the most of public assets. http://www.w3c.es/Presentaciones/2013/0514openData_WSIS-MA/0514-openData_WSIS-MA.pdf

Kalampokis, E., Tambouris, E., Hausenblas, M., & Tarabanis, K. (2012). Augmenting Open Government Data with Social Media

Data.http://www.w3.org/2012/06/pmod/pmod2012_submission_29.pdf



2.3. Mijlpaalproducten

In de onderstaande tabel staan de op te leveren mijlpaalproducten per week. Hiervoor geldt: lever ze uiterlijk 24 uur vóór aanvang van de betreffende les in. Het inleveren gebeurt middels de systemen Trello (project Scrum) en Box.com (project dossier). Eindpresentatie +/- 15 minuten.

Week	Sprint	Activiteit	Mijlpaalproduct PO (cijfer)	Mijlpaalproduct Tutor (inleveren ter afvinken)*	Individuele bijdrage Tutor (cijfer)
1	1	Kick-off			
2		Sprint review	- Epics - Product backlog (opstellen) - Sprint backlog (voorstel Sprint 2) - User stories (voor Sprint 2)	- Samenwerkingscontract - Scrum board	- Retrospective
3	2	Toelichting ver	wachtingen sprint, afsten	nming met PO en Tutor	
4		Sprint review	- Product backlog (overzicht) - Sprint backlog (voorstel Sprint 3) - User stories (voor Sprint 3) - Shippable product	- Retrospective - Scrum board - Burndown chart Sprint 2	- Retrospective
5	3	Toelichting verwachtingen sprint, afstemming met PO en Tutor			
6		Sprint review	- Product backlog (overzicht) - Sprint backlog (voorstel Sprint 4) - User stories (voor Sprint 4) - Shippable product	- Retrospective - Scrum board - Burndown chart Sprint 3 - DoD	- Retrospective
7	4	Toelichting verwachtingen sprint, afstemming met PO en Tutor			
8		Opleveren gehele project	- Product backlog (overzicht) - Eindproduct (code**, database** & applicatie**) - Eindpresentatie	- Retrospective - Scrum board - Burndown chart Sprint 4 - DoD	- Retrospective
9		Inloop voor feedback beoordeling			

^{*} Consequenties niet afgevinkte criteria (b.v. samenwerkingscontract, Scrum board, enz...) heeft uiteindelijk invloed op het missen van punten voor het teamcijfer!
** Telt mee voor eindcijfer, zie criteria bijlage 3 en 4



3. Toetsing en beoordeling

3.1. Voorwaarden

Om de punten voor het team per sprint te ontvangen, moeten de onderstaande mijlpaalproducten voldoende zijn. Ook moet worden voldaan aan de aanwezigheidsplicht en moeten de (tussen)producten op tijd zijn ingeleverd.

Sprint	Onderdelen afgevinkt door tutor	Doel
1	Retrospective, Samenwerkingscontract en Scrum board	Afvinken (moet voldoende zijn)
2	Retrospective, Scrum board en burndown chart sprint 2	Afvinken (moet voldoende zijn)
3	Retrospective, Scrum board en burndown chart sprint 3	Afvinken (moet voldoende zijn)
4	Retrospective, Scrum board, burndown chart sprint 4 en eindpresentatie	Afvinken (moet voldoende zijn)
Elke sprint	Aanwezigheid (mede afgevinkt door PO)	Afvinken (moet voldoende zijn)
Elke sprint	Op tijd inleveren producten (mede afgevinkt door PO)	Afvinken (moet op tijd zijn)

3.2. Eindcijfer

Het project wordt beoordeeld op een combinatie van groepsprestatie en individuele inzet. Het teamcijfer telt voor 40% - en het individuele cijfer telt voor 60% mee.

Voor elke sprint zijn punten te behalen. De toekenning van deze punten wordt tijdens de reviews bepaald, waarbij de tutor de individuele bijdrage beoordeeld en de product owner het groepsresultaat. Hieronder staat de verdeling van de punten per sprint en de bijbehorende onderdelen.

Sprint	Onderdelen team (40%), beoordeeld door product owner	Aantal punten
1	Epics, product backlog, sprint backlog (voorstel Sprint 2) en user stories (voor Sprint 2)	0 of 1
2	Product backlog (overzicht), sprint backlog (voorstel Sprint 3), user stories (voor Sprint 3) en shippable product	0 of 1
3	Product backlog (overzicht), sprint backlog (voorstel Sprint 4), user stories (voor Sprint 4) en shippable product	0, 1 of 2
4	Product backlog (overzicht), eindproduct (code* & applicatie) en eindpresentatie (mede beoordeeld door tutor)	0, 1 of 2

1	Sprint	Onderdelen team (40%), beoordeeld door analyse docent	Aantal punten
	4	Eindproduct code (SQL)***, database***	0, 1 of 2

^{***} Wordt beoordeeld door analyse docent, zie criteria bijlage 4

Sprint	Onderdelen team (40%), beoordeeld door development docent	Aantal punten
4	Eindproduct code****	0, 1 of 2

^{****} Wordt beoordeeld door development docent, zie criteria bijlage 3



Sprint	Onderdelen individueel (60%), beoordeeld door tutor	Aantal punten
1	Retrospective	0, 1 of 2
2	Retrospective	0, 1 of 2
3	Retrospective	0, 1 of 2
4	Retrospective	0, 1 of 2

3.3. Formule

Om het eindcijfer te kunnen bepalen wordt de volgende formule gebruikt:

$$eindcijfer = \frac{(4 \cdot punten team) + (6 \cdot (punten individueel + 2))}{10}$$

Voorbeeld voldoende (geslaagd):

eindcijfer =
$$\frac{(4\cdot5)+(6\cdot(4+2))}{10}$$
 = $\frac{56}{10}$ = 5,6

Voorbeeld onvoldoende (herkansen):

eindcijfer =
$$\frac{(4\cdot6)+(6\cdot(3+2))}{10}$$
 = $\frac{54}{10}$ = 5,4

3.4. Herkansing

Bij een onvoldoende voor INFPRJ00-3 dient het project in zijn geheel te worden herkanst in de volgende onderwijsperiode. Dit houdt in dat er een vergelijkbare opdracht opnieuw in een samengestelde projectgroep gemaakt moet worden (OP-beheerder deelt in), zonder begeleiding en zonder ondersteunend onderwijs. Alleen eindproduct(en) worden beoordeeld.



Bijlage 1: Toetsmatrijs

De toetsmatrijs geeft de relatie weer tussen de leerdoelen en de opdrachten en criteria van een toets.

Leerdoelen	Verwijzing naar producten / criteria
Je kunt een korte overtuigende presentatie houden/instructie geven of kennis delen met aandacht voor doel(groep), structuur, inhoud, (non)verbale aspecten en hulpmiddelen	Eindpresentatie
Je bent in staat de specifieke kennis uit je vakgebied eigen te maken via reguliere lessen en zo nodig via zelfstudie uit te breiden	Retrospective
Je bent in staat om met behulp van het creatieve proces een klantenvraag om te zetten in een creatieve, innovatieve oplossing die past bij de klantenvraag	Epics, user stories, shippable product en eindproduct
Je kunt met een projectgroep een planning opstellen voor een complex project, een MoSCoW daarin meenemen en tussentijds de planning adequaat bijstellen zonder einddoel uit het oog te verliezen	Product backlog, sprint backlog, user stories, Scrum board en burndown chart
Je bent in staat, met input van de Product Owner, een sprint - en product backlog op te stellen volgens de (Scrum) richtlijnen	Product backlog en sprint backlog
Je bent in staat een sociale media API aan te roepen en deze informatie op te slaan in een database	Shippable product en eindproduct (code)
Je kunt een (portal) applicatie maken met een front-end en een gekoppelde database waarmee data analyses uitgevoerd kunnen worden	Eindproduct



Bijlage 2: Format retrospective (door tutor)

Benodigdheden: flipover, diverse geeltjes, stiften, plakband			
Naam groep:		Datum en tijdstip:	
Sprint #:		Procesbegeleider retrospective:	
Notulist:		Klas:	
Fase 1: Set the stage	Formuleer p	per teamlid:	
± 5 min.		rden die kenmerkend zijn voor de afgelopen sprint rden die jouw gedrag weergeven in de afgelopen nt	
en/of die ind positieve ér geeltjes. Gr plak die op Geef aan he schaal van Maak hierva		ak een inventarisatie van gebeurtenissen die zijn bijgebleven of die indruk hebben gemaakt in deze sprint. (Dat kunnen itieve én negatieve punten zijn). Schrijf deze op losse eltjes. Groepeer de geeltjes die over dezelfde items gaan en k die op de flipover. Werk daarbij met symbolen (+ / - / ?). ef aan hoe jouw tijdsbesteding geweest is tot nu toe op een aal van 1-4. ak hiervan een (leesbare) foto complete visuele weergave (foto) dient bijgevoegd te worden bijlage.	
		unten uit fase 2 die prioriteit hebben gekregen. ngen die items wel/ niet goed?	
Fase 4: Decide what to do ± 5 min.	gebeurtenis acties onde	actiepunten die worden opgepakt om de sen aan te pakken uit fase 3. Benoem ook welke rnomen worden om de pluspunten (de dingen die in zijn) te behouden.	
Fase 5: peerassessment ± 15 min.	functionerer ondergemi ieder aan w compliment De uitwerkir elkaar. De tutor nee cijfer: 0= on 1= gemidde Met gemidd afgesproker groepscijfer	alle projectleden kort op hoe je hen vindt in de projectgroep. Je kunt kiezen tussen ddeld/ gemiddeld/ bovengemiddeld. Je geeft voor at je argumenten zijn. (Geef ieder projectlid ook één mee.) ing van alle bevindingen en argumenten komen bij emt dit mee in de becijfering van het individuele dergemiddeld eld, 2= bovengemiddeld. eld bedoel je dat iemand precies werkt zoals jullie in hebben. Dat komt dus overeen met het dat jullie verdienen. De tutor neemt de grootste er mee van de peerassessments.*****	

^{*****} Mocht je van tevoren afspreken dat je ieder bovengemiddeld beoordeelt, dan kun je niet klagen over medeprojectgenoten. Je vindt immers dat ze bovengemiddeld presteren. Leer dus aan je bevindingen eerlijk aan elkaar te geven. Zo geef je elkaar de mogelijkheid om te leren en beter te worden dan daarvoor.



Bijlage 3: Beoordelen van code in het project (door development docent)

Elke groep en elke student moet zijn projectcode in lesweek 3, lesweek 5 en lesweek 7 voorbereiden voor de development lessen. Tijdens deze lesweken kiest de development docent willekeurig groepen uit om de code klassikaal te reviewen en te bespreken. Op deze manier krijgen de groepen tussentijds feedback op hun code, zonder consequenties, zodat zij daar iets van leren en hun code verbeteren voor het eindproduct.

Beoordeling van de code wordt gedaan in lesweek 8 op basis van coding standards. Meer informatie is te vinden in de modulewijzer van INFDEV01-3 en tijdens de development lessen. Maak dus goed gebruik van de tussenweken waarin je feedback krijgt om hieraan te voldoen!

Bijlage 4: Beoordelen van database in het project (door analyse docent)

Deel 1 - lesweek 5

Informatiebehoefte

- Beknopt (maximaal enkele alinea's)
- Duidelijk en eenduidig (SMART)
- Volledia

ERD

- Minimaal 5 entiteiten
- In overeenstemming met de beschrijving van informatiebehoefte
- Verantwoording van keuzes in het ontwerp (op basis van de theorie)
- Consistente naamgevingsconventies (consistent gebruik hoofdletters, enkelvoud/meervoud)
- Betekenisvolle naamgeving (beschrijvend)
- Het model is genormaliseerd volgens de eis van Boyce/Codd-normaalvorm
- Normalisatie is beschreven in stappen
- Voldoet aan de regels van identificatie, kardinaliteit, optionaliteit, domeinwaarden, afhankelijkheid, afleidbaarheid en transitie

Deel 2 - lesweek 7

Database

- Overeenkomstig ERD
- Consistente naamgevingsconventies
- Betekenisvolle naamgeving
- Correct datatype
- 'Nullable', default waarde, auto increment correct bepaald

SQL

- Minimaal 4 queries over meerdere tabellen
- Waarin gebruik wordt gemaakt van de functies HAVING en SET