

MASTER'S THESIS

Managing data science projects to gain business value

Management van data science projecten voor waardecreatie

Dreschler, D (Dirk)

Award date:

2020

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 09. Sep. 2021

Open Universiteit
www.ou.nl



Managing data science projects to gain business value

Management van data science projecten voor waardecreatie

Opleiding: Open Universiteit, faculteit Management, Science & Technology
Masteropleiding Business Process Management & IT
Variant Data Science Management

Programme: Open University of the Netherlands, faculty of Management, Science & Technology
Master Business Process Management & IT

Cursus: IM0602 Voorbereiden Afstuderen Business Process Management and IT
IM9806 Afstudeeropdracht Business Process Management and IT

Student: Dirk Dreschler

Identiteitsnummer:

Datum: 09-07-2020

Afstudeerbegeleider Prof.dr.ir. Remko Helms

Meelezer Dr.ir. Harry Martin

Versie nummer: 3.0

Status: Definitief

Abstract

Organisaties voeren data science projecten uit om blijvende waarde te creëren. Ondanks de potentie van data science projecten mislukt meer dan 50% van deze projecten. De oorzaak hiervan kan voornamelijk gevonden worden in de organisatorische en culturele aspecten van organisaties. Naast een inzicht in de manier waarop organisaties data science projecten succesvol kunnen managen geeft dit case study onderzoek inzicht in de invloed die de data-analysecapaciteiten, en in het bijzonder de datastrategie en data gedreven cultuur, hebben op de verschillende projectmanagementfactoren. Ook wordt inzicht gegeven in de manier waarop organisaties bij de totstandkoming en het management van de data science projecten rekening houden met het belang van de datastrategie en data gedreven cultuur. Stakeholdermanagement en communicatie zijn daarbij van groot belang en zouden moeten worden ondersteund door een goede projectmanagementmethode. Tot slot laat deze case study zien dat kleinere organisaties met een aspirante volwassenheid van hun data-analysecapaciteiten zouden moeten werken het creëren van tijd, budget en het in huis halen van data science kennis voor het verbeteren van die data-analysecapaciteiten. Grotere organisaties met een ervaren volwassenheid in hun data-analysecapaciteiten zouden zich moeten richten op het vrijmaken van capaciteit en prioriteit bij data scientists en stakeholders, en het organiseren van een goede databeschikbaarheid en data-architectuur.

Sleutelbegrippen

Big data, data science, waardecreatie, data-analysecapaciteiten, projectmanagement

Samenvatting

Het toepassen van big data-analyses zorgt voor 5% meer productiviteit en 6% meer winstgevendheid ten opzichte van concurrerende organisaties die geen gebruik maken van big data. Het is dan ook niet verwonderlijk dat organisaties steeds meer big data willen gebruiken om processen te optimaliseren, kosten te besparen, nieuwe inzichten te verkrijgen of zelfs nieuwe producten en businessmodellen te ontwikkelen. Om deze blijvende waardecreatie te realiseren voeren organisaties data science projecten uit en moeten organisaties beschikken over data-analysecapaciteiten en projectmethoden en -vaardigheden. Ondanks de grote potentie weten organisaties maar slechts beperkt structureel te profiteren van big data. Van de uitgevoerde data- science projecten wordt 55% niet afgerond en vele anderen organisaties stoppen de toepassing nadat het project is afgerond. De oorzaak hiervan ligt voornamelijk bij de data-analysecapaciteiten organisatie en mensen, meer specifiek de datastrategie en een data gedreven cultuur, en minder bij de technologische of proces gerelateerde data-analysecapaciteiten.

In dit case study onderzoek wordt antwoord gegeven op de vraag hoe organisaties data science projecten succesvol moeten managen om tot blijvende waardecreatie te komen vanuit big data. Daarbij wordt achtereenvolgens gekeken naar wat blijvende waardecreatie voor organisaties inhoudt, hoe data science projecten tot stand komen en worden gemanaged en welke projectfactoren daarbij een rol spelen, welke data science projectmethoden worden gebruikt en wat hun voor- en nadelen zijn en welke data-analysecapaciteiten van invloed zijn en in welke mate. Extra aandacht is er daarbij voor de organisatorische en culturele aspecten, de invloed daarvan op het succesvol managen van data science projecten en de manier waarop organisaties rekening houden met deze aspecten bij de totstandkoming en het managen van de data science projecten.

Vanuit de literatuurstudie is een wetenschappelijk beeld gevormd van het succesvol managen van data science projecten en de data-analysecapaciteiten en projectfactoren die daarop van invloed zijn. De resultaten van de literatuurstudie zijn geoperationaliseerd in een theoretisch raamwerk dat de basis vormt voor het vervolgonderzoek. Via het afnemen van semigestructureerde interviews bij zes organisaties die onlangs data science projecten hebben uitgevoerd is empirische informatie verzameld. De interviews zijn geanalyseerd aan de hand van een 10-stappenplan en met behulp van een op de literatuur gebaseerde boomstructuur gecodeerd en geanalyseerd. Daarbij is ook onderscheid gemaakt naar de verschillende kenmerken van de organisaties. Zo is een verdiepend beeld gekregen van de manier waarop data science projecten succesvol zouden moeten worden gemanaged en de impact die de organisatorische en culturele aspecten daarin hebben.

Om het management van data science projecten als succesvol te bestempelen vinden de organisaties dat blijvende waarde gecreëerd moet zijn door de resultaten van die projecten structureel te integreren in de vaste processen van de organisatie. Tegelijk geven de organisaties aan dat ze dit succes beperkt kunnen meten. Organisaties met een ervaren volwassenheid van hun data-analysecapaciteiten ervaren vooral dat de projectresultaten niet structureel worden toegepast in de organisatie. Organisaties met een aspirante volwassenheid van hun data-analysecapaciteiten ervaren voornamelijk moeite met het meten van het (structurele) effect van de projectresultaten.

Analyse van de interviewresultaten bevestigt dat de data-analysecapaciteiten mensen, datastrategie en een data gedreven cultuur de grootste impact hebben op het succesvol managen van een data science project. In het geval van mensen gaat het om de analytische, technische en communicatieve vaardigheden van de data scientists maar ook om een brede kennis van de mogelijkheden en toepassing van data science binnen de hele organisatie. De data-analysecapaciteiten mensen en organisatie hebben vooral invloed op de projectfactoren project medewerking en teamcompetenties.

Om binnen de organisatie rekening te houden met de organisatorische en culturele aspecten worden stakeholders al vroeg betrokken in verschillende projectfasen van de een data science project. Zo wordt afstemming gezocht in de organisatie over de behoefte en prioriteit van de data science projecten en worden stakeholders betrokken bij het formuleren van de probleemstelling en het bepalen van de mogelijke oplossingen en benodigdheden daarvoor. Gedurende het project wordt bij voorkeur in het zicht van de organisatie gewerkt aan het data science project en wordt regelmatig gecommuniceerd en om feedback gevraagd voor het creëren van draagvlak en een data gedreven cultuur. De Agile Kanban en Agile Scrum methoden faciliteren met de interactie en iteratie principes de planning, beheersing en communicatie van het data science project beter dan de CRISP-DM methode die op projectmanagement vlak tekortschiet. De CRISP-DM methode daarentegen is weer meer geschikt voor het data mining proces.

Voor het vergroten van het data science projectsucces zouden organisaties vooral moeten werken aan het verhogen van de volwassenheid van de data-analysecapaciteiten mensen en organisatie. Drie organisaties hebben hiervoor hun kennis en vaardigheden op directieniveau uitgebreid door invulling te geven aan de functie van Chief Digital Officer (CDO). Deze CDO is verantwoordelijk voor de datastrategie en de ontwikkeling van de benodigde data-analysecapaciteiten. Bij het verbeteren van de data-analysecapaciteiten zouden kleinere aspirant organisaties zich vooral moeten richten op het creëren van tijd, budget en het in huis halen van data science kennis. Grotere ervaren organisaties zouden zich meer moeten richten op het vrijmaken van capaciteit en prioriteit bij data scientists en stakeholders, en het organiseren van een goede databeschikbaarheid en data-architectuur. De organisaties geven aan dat voor het succesvol managen van een data science project de projectscope klein gehouden moet worden en beperkt moet worden tot de bestaande data-analyse capaciteiten van de organisatie.

Summary

The application of big data analyzes provides 5% more productivity and 6% more profitability compared to competing organizations who do not use big data. Therefore, it is not surprising that organizations want to use more and more big data to optimize processes, save costs, gain new insights, or even develop new products and business models. To achieve this lasting value creation, organizations implement data science projects. To do so, organizations must have data analysis capabilities, project methods and skills. Despite the great potential, organizations are only able to profit structurally from big data to a limited extent. 55% of the implemented data science projects are not completed and many organizations stop the application after the project has been completed. This is mainly due to the data analysis capabilities organization and people, more specifically the data strategy and data culture, and less to the technological or process-related data analysis capabilities.

This case study research answers the question of how organizations should successfully manage data science projects to achieve lasting value creation from big data. In doing so, we look successively at what lasting value creation means for organizations, how data science projects are created and managed and which project factors play a role in this, which data science project methods are used and what their advantages and disadvantages are and which data analysis capabilities influence this and to what extent. Extra attention is given to the organizational and cultural aspects, the influence of this on the successful management of data science projects and the way in which organizations take these aspects into account in the creation and management of the data science projects.

The literature study provides a scientific view on successful management of data science projects and the data analysis capabilities and project factors that influence this. The results of the literature study have been operationalized in a theoretical framework that forms the basis for the empirical research. Empirical information was collected by conducting semi-structured interviews with six organizations that have recently carried out data science projects. The interviews were analyzed using a 10-step plan and coded and analyzed using a literature-based tree structure. A distinction was also made according to the different characteristics of the organizations. This has provided an in-depth view of the way in which data science projects should be managed successfully and the impact of the organizational and cultural aspects.

To label the management of data science projects as successful, organizations believe that lasting value must be created by structurally integrating the results of those projects into the regular processes of the organization. At the same time, the organizations indicate that they can hardly measure this success. Organizations with an experienced maturity of their data analysis capabilities mainly experience that the project results are not applied structurally in the organization. Organizations with an aspiring maturity of their data analysis capabilities mainly have trouble in measuring the (structural) effect of the project results.

Analysis of the interview results confirms that the data analysis capabilities people, data strategy and data culture have the greatest impact on the successful management of a data science project. In the case of people, it concerns the analytical, technical and communication skills of the data scientists, but also a broad knowledge of the possibilities and application of data science throughout the organization. The data analysis capabilities people and organization mainly influence the project factors project cooperation and team competences. In order to take into account the organizational and cultural aspects within the organization, stakeholders are involved early on in various project phases of a data science project. For example, coordination is sought in the organization about the need and priority of the data science projects and stakeholders are involved in formulating the problem definition and determining the possible solutions and requirements for it. During the project, the data science project will preferably be worked on within the organization's view, and regular

communication will take place and feedback will be requested to create support and a data culture. The Agile Kanban and Agile Scrum methods, with the interaction and iteration principles, facilitate the planning, management, and communication of the data science project better than the CRISP-DM method that is lacking in project management. The CRISP-DM method, on the other hand, is more suitable for the data mining process.

To increase the data science project success, organizations should mainly work on increasing the maturity of the data analysis capabilities people and organization. Three organizations have expanded their knowledge and skills at board level by fulfilling the position of Chief Digital Officer (CDO). This CDO is responsible for the data strategy and the development of the required data analysis capabilities. In improving data analysis capabilities, smaller aspiring organizations should focus primarily on creating time, budget, and in-house data science knowledge. Larger experienced organizations should focus more on freeing up capacity and priority among data scientists and stakeholders and should organize good data availability and data architecture. The organizations indicate that to successfully manage a data science project, the project scope must be kept small and limited to the existing data analysis capabilities of the organization.

INHOUDSOPGAVE

1. Introductie	1
Aanleiding	1
Probleemstelling	1
Onderzoeksvraag	2
Relevantie	2
Onderzoeksaanpak.....	2
2. Theoretisch kader	3
Onderzoeksaanpak.....	3
Uitvoering.....	5
Resultaten en conclusies.....	6
Waardecreatie	6
Data-analysecapaciteiten.....	6
Data science projecten.....	8
Conclusie.....	10
3. Methodologie.....	11
Conceptueel ontwerp: keuze van onderzoeksmethode(n).....	11
Technisch ontwerp: uitwerking van de methode	11
Gegevensanalyse.....	12
Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten	13
4. Resultaten	14
De caseorganisaties.....	14
Codering en opbouw.....	15
Waardecreatie.....	16
Projectfasen	17
Projectmanagementmethode.....	19
Data-analysecapaciteiten.....	21
5. Discussie, conclusies en aanbevelingen.....	25
Conclusies.....	25
Discussie	26
Reflectie	27
Aanbevelingen voor de praktijk	28
Aanbevelingen voor verder onderzoek.....	28
Bijlage 1: interview protocol.....	31
Bijlage 2: interviewmatrix	35
Bijlage 3: omschrijving caseorganisaties.....	36
Bijlage 4: codering	37

1. INTRODUCTIE

Aanleiding

Steeds vaker worden geavanceerde data-analysetechnieken ingezet om op een kosteneffectieve manier blijvende waarde te creëren uit grote hoeveelheden data. Organisaties verwachten hiermee te kunnen innoveren, hun productiviteit te verbeteren en een voordeel te behalen op de concurrentie (Vidgen, 2017). Meer specifiek verwachten zij processen te optimaliseren, kosten te besparen, nieuwe inzichten te verkrijgen, producten en diensten uit te breiden en mogelijk nieuwe businessmodellen te ontwikkelen (Schüritz, 2017). Het toepassen van big data-analyses zorgt voor 5% meer productiviteit en 6% meer winstgevendheid ten opzichte van concurrerende organisaties die geen gebruik maken van big data (McAfee, 2012). Het is dan ook niet verwonderlijk dat organisaties steeds meer data willen gebruiken voor het nemen van beslissingen.

Een organisatie die haar beslissingen in grote mate wil baseren op big data moet beschikken over data die voldoet aan invulling van de 4 V's. Daarbij gaat het om de omvang van de data (Volume), het aantal databronnen en de structuur van de data (Variety), de snelheid waarmee de data wordt gegenereerd en verzameld (Velocity) en de betrouwbaarheid van de data (Veracity) (Saltz J. S., 2017). Om big data om te zetten in waarde moet een organisatie beschikken over data-analysecapaciteiten en voeren organisaties data-analyseprojecten uit. Deze projecten verlopen via het proces van data verzameling, naar het omzetten van data in inzichten, tot en met het vertalen van de inzichten naar actie. Gezien de kenmerken van big data vraagt de uitvoering van het analyseproces een breed scala aan analyse-, communicatie- en managementvaardigheden (Koronios, 2015). Personen die over dergelijke vaardigheden beschikken worden vaak data scientists genoemd (Watson, 2014).

Voor het managen van data science projecten worden vele technieken, tools en methoden gebruikt. De meest bekende en gebruikte methode is CRISP-DM (Mariscal, 2010). Sinds enkele jaren worden de Agile methoden steeds meer gebruikt om data science projecten te managen (Saltz J. S., 2017). Gezien de afname in het gebruik van de CRISP-DM methodologie en de wisselende successen van verschillende modellen en methoden (Mariscal, 2010) lijken zowel de wetenschap als organisaties op zoek naar nieuwe inzichten om hun data science projecten succesvol te managen.

Probleemstelling

Ondanks de grote potentie van big data, de aanwezigheid van analysecapaciteiten, specifieke projectmanagementmethoden en hooggekwalificeerde data scientists weten organisaties maar slechts beperkt structureel te profiteren van big data. Van de uitgevoerde data science projecten wordt 55% niet afgerond en vele anderen stoppen de toepassing nadat het project is afgerond (Koronios, 2015). Ook geeft 32% van het senior management in organisaties slechts een 3 of lager (op een 5 puntenschaal) voor de datagedrevenheid van hun organisatie (McAfee, 2012). De oorzaak van een beperkte succesvolle toepassing van big data lijkt niet te liggen bij het verkrijgen en beheren van kwalitatief goede data of bij de benodigde technologie, maar veel meer in de organisatorische aspecten management en cultuur. De meest genoemde belemmeringen die organisaties ervaren in hun ambitie om data gedreven te worden zijn het gebrek aan inzicht hoe big data te gebruiken om te komen tot verbeteringen, het gebrek aan prioriteit bij het management en het gebrek aan data-analyse vaardigheden bij de interne afdelingen (Lavelle, 2011).

Ruim 50% van de data science projecten mislukt door organisatorische en culturele aspecten

Onderzoeksvraag

Om de oorzaak te achterhalen van het mislukken van een groot deel van de data science projecten en om manieren te ontdekken die leiden tot succes, is het belangrijk te onderzoeken welke kennis in de wetenschappelijke literatuur bekend is. De focus ligt daarbij op de organisatorische en culturele aspecten in het waardecreatie proces vanuit big data en de dimensies en succesfactoren die daarbij een rol spelen. De onderzoeksvraag van dit onderzoek luidt dan ook:

Hoe moeten organisaties data science projecten succesvol managen om tot blijvende waardecreatie te komen vanuit big data?

Verschillende elementen spelen een rol bij de het succesvol managen van data-science projecten. Allereerst gaat het daarbij om ontwerp gerelateerde zaken zoals de totstandkoming en afbakening van het project, de doelstellingen en verwachtingen en het analyiseniveau. Daarnaast gaat het om meer projectmanagement gerelateerde zaken zoals de toegepaste methoden, processen en succesfactoren. Tot slot spelen de data-analysecapaciteiten van de organisatie een belangrijke rol. Dit leidt tot de volgende sub-onderzoeksvragen:

- Wat houdt blijvende waardecreatie vanuit big data in bij het managen van data science projecten?
- Hoe komen data science projecten tot stand, worden de verschillende projectfasen gemanaged en welke projectfactoren spelen daarbij een rol?
- Welke data science projectmanagementmethoden en -processen kunnen worden gebruikt en wat zijn hun voordelen en nadelen?
- Welke data-analysecapaciteiten van de organisatie zijn van invloed op het managen van data science projecten en in welke mate?

Relevantie

Het grootschalig gebruik van big data is pas sinds enkele jaren mogelijk als gevolg van technologische ontwikkelingen. De bestaande wetenschappelijke kennis daarover is dan ook nog beperkt. Dit onderzoek draagt bij aan het beter begrijpen van dimensies en factoren die een rol spelen in het succesvol managen van data science projecten. In de literatuurstudie is een weergave gegeven van de huidige kennis hoe organisaties data science projecten succesvol moeten managen. Deze kennis is verrijkt met een kwalitatief onderzoek naar de manier waarop data science projecten in de praktijk worden gemanaged en in welke mate rekening wordt gehouden met de in de literatuur gevonden voorwaarden voor succes. Omdat steeds meer organisaties verwachten in de toekomst überhaupt of intensiever gebruik te zullen maken van big data toepassingen is het belangrijk inzicht te hebben hoe organisaties data science projecten succesvol moeten managen zodat het aantal mislukte projecten tot een minimum wordt beperkt en waardecreatie zo effectief en efficiënt mogelijk wordt gerealiseerd.

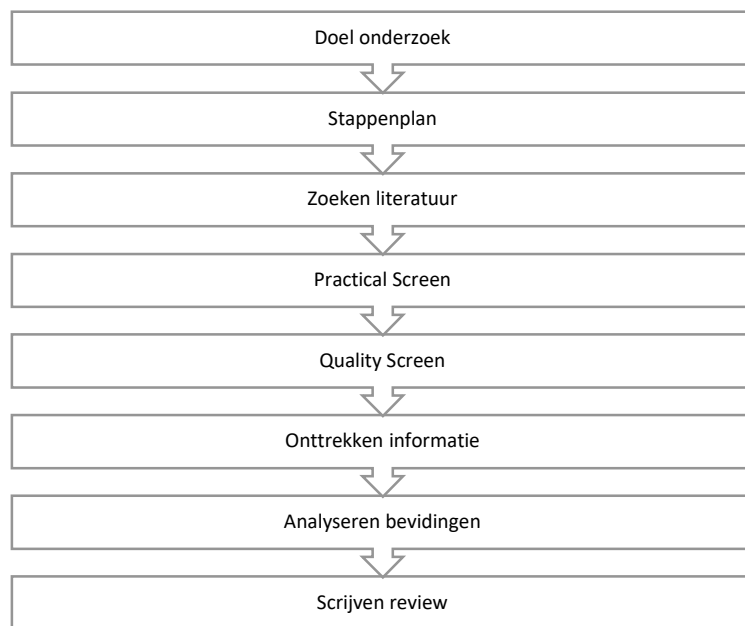
Onderzoeksaanpak

Vanuit de erkenning van de organisatorische en culturele problemen bij het uitvoeren van data science projecten is de daaropvolgende onderzoeksvraag gedefinieerd. Doormiddel van een literatuurstudie, waarvan het ontwerp hierna te lezen is, zal antwoord worden gegeven hoe organisaties hun data science projecten het best uit kunnen voeren. Deze bevindingen zijn te lezen in hoofdstuk 2. Hoofdstuk 3 beschrijft de manier waarop het verdiepende case study onderzoek naar de manier hoe zij hun data science projecten managen is uitgevoerd bij verschillende organisaties. De resultaten van dit onderzoek zijn beschreven in hoofdstuk 4 van waaruit conclusies zijn getrokken en vergeleken met de bestaande wetenschappelijke kennis in hoofdstuk 5.

2. THEORETISCH KADER

Onderzoeksaanpak

Het theoretische onderzoek wordt uitgevoerd aan de hand van acht stappen die zijn opgesteld voor het uitvoeren van literatuuronderzoek binnen het domein van informatiesystemen (Okoli, 2010), welke visueel zijn weergegeven in figuur 1. Allereerst is een onderzoeksplanning opgesteld waarin doel, stappen en afbakening worden vastgesteld. Daarna wordt verder ingegaan op het zoeken en selecteren van relevante literatuur. Tot slot worden criteria voor de practical screen en quality screen uiteengezet om te waarborgen dat het aantal geraadpleegde bronnen praktisch handelbaar en van hoge kwaliteit is en voldoet aan de beantwoording van de onderzoeksvraag. In deze paragraaf wordt het stappenplan verder toegelicht. In de paragraaf uitvoering is het verloop van het onderzoek toegelicht. In de paragraaf resultaten en conclusies wordt de gevonden informatie uiteengezet.



Figuur 1: stappenplan van het literatuuronderzoek (Okoli, 2010)

Om antwoord te krijgen op de onderzoeksvraag wordt allereerst gezocht in de bestaande wetenschappelijke literatuur. Deze literatuurstudie wordt uitgevoerd aan de hand van zoekquery's, toegepast binnen wetenschappelijke bronnen. Door de Open Universiteit zijn negen relevante artikelen ter beschikking gesteld. Vanuit praktisch oogpunt, passend binnen de scope en duur van dit onderzoek, worden de door de Open Universiteit geselecteerde artikelen uitgebreid met maximaal zeven tot tien relevante nieuwe artikelen, gevonden in de universiteitsbibliotheek van de Open Universiteit.

Voor het opstellen van de zoekquery is gebruik gemaakt van Boolean logic waarbij de onderzoeksvraag is opgesplitst in de onderdelen vraagtype, context, onderwerp en actie.

[Hoe moeten = vraagtype] [organisaties = context] [data science projecten = onderwerp] [succesvol managen = actie] om tot [blijvende waardecreatie te komen vanuit big data = context]?

Voor de onderdelen onderwerp en actie zijn synoniemen en/ of antoniemen genoteerd en gebruikt in de zoekquery, zie tabel 1. Bij het beoordelen van de gevonden literatuur is rekening gehouden met beide context onderdelen. Het onderdeel vraagtype heeft invloed op het soort vervolgonderzoek.

Door de synoniemen samen te voegen met “AND” en “OR” en () is de zoekquery vastgesteld. Tussen de kolommen wordt met “AND” gezocht, tussen de rijen met “OR”. Met het * symbool wordt naar meerdere woordvormen gezocht (Saunders, 2016). Dit resulteert in deze zoekquery:

("data science project" OR "data analytics project" OR "data mining project") AND (manag) AND ("challenge" OR "failure" OR "success" OR "performance" OR "problem")*

Tabel 1: zoekquery in de bibliotheek van de Open Universiteit

Data science project	Manage	Uitdaging/ Succes
Data science project	Management	Challenge
Data analytics project	Managing	Failure
Data mining project		Success
		Performance
		Problem

Voor het terugbrengen van de zoekresultaten naar een zo relevant en praktisch werkbaar aantal, wat de practical screen wordt genoemd (Janssen, 2013), wordt gebruik gemaakt van de volgorde selectiecriteria:

1. Zoekopdracht met de zoekquery in de universiteitsbibliotheek van de Open Universiteit
2. Zoekopdracht verbreden met ‘vinkje aan’ bij het uitbreiden met andere bronnen
3. Verfijnen van de resultaten van in alle talen geschreven artikelen naar enkel Engelstalige literatuur vanwege de leesbaarheid en beste begrip van definities
4. Verfijnen van de resultaten tot enkel peer reviewed voor kwaliteitsborging
5. Verfijnen van de resultaten naar het managementwetenschapsdomein ‘business’ voor relevantie
6. Verfijnen van de resultaten met onderzoek relevantie op basis van titel en abstract
7. Via de snowballing techniek zoeken naar aanvullende en/of verdiepende artikelen
8. Een peer review van de afstudeerbegeleider op de geschiktheid van de artikelen aan de hand van een gedeelde lijst met daarin de publicatietitel en -bron van de geselecteerde artikelen

Met een quality screen wordt gewaarborgd dat de gevonden en geselecteerde literatuur daadwerkelijk bijdraagt aan de beantwoording van de onderzoeksvraag (Janssen, 2013). Om te komen tot een selectie van de in de literatuur gevonden artikelen zijn de volgende stappen doorlopen:

1. Lezen van de inleiding, conclusies en suggesties voor verder onderzoek
2. Beoordelen van de relevantie voor de beantwoording van de onderzoeksvraag
3. Lezen van de relevante paragrafen en diagonaal lezen van gehele artikel voor beter begrip
4. Beoordelen van de relevantie voor beantwoording van de onderzoeksvraag
5. Definitieve selectie van het artikel voor verdere analyse

Van de geselecteerde artikelen wordt een samenvatting gemaakt van 300-500 woorden. Deze samenvattingen worden in één Word document onder elkaar gezet en via verschillende kleuren van elkaar onderscheiden. Door het rangschikken van de teksten naar verschillende thema’s wordt een overzicht gemaakt van de gevonden informatie waarmee de synthese wordt geschreven. Deze synthese geeft nieuwe informatie zodat het onderzoek idee, de probleemstelling en deelvragen en het theoretische en empirische raamwerk kunnen worden vormgegeven (Janssen, 2013).

Uitvoering

Het literatuuronderzoek is in december 2019 uitgevoerd aan de hand van het opgestelde stappenplan. Met de gedefinieerde zoekquery zijn in de onlinebibliotheek van de Open Universiteit artikelen gezocht. De resultaten van de artikelselectie zijn in tabel 2 weergegeven. De zoekquery leverde 981 resultaten op. Uitbreiding naar andere bronnen gaf 4.389 artikelen. Na selectie op Engelstalig, peer reviewed en business bleven 121 artikelen over. Van deze artikelen zijn de titel en abstract gelezen en beoordeeld op relevantie ten aanzien van de onderzoeksvraag. De negentien overgebleven artikelen zijn ter beoordeling met publicatietitel en -bron voorgelegd aan de afstudeerbegeleider. Na deze peer review zijn deze artikelen allemaal diagonaal en deels geheel gelezen. De meest relevante artikelen voor het beantwoorden van de onderzoeksvraag zijn daarna definitief opgenomen voor de analyse. Uiteindelijk zijn van de gevonden artikelen er acht additioneel gebruikt, aan de door de Open Universiteit beschikbaar gestelde artikelen, voor beantwoording van de onderzoeksvraag.

Tabel 2: resultaten van de artikelselectie

Onderdeel	Resultaat
1. Zoekquery	981 artikelen
2. Vinkje 'andere bronnen'	4.389 artikelen
3. Selectie 'taal = Engels'	4.367 artikelen
4. Selectie 'peer review'	716 artikelen
5. Selectie 'vakgebied = business'	121 artikelen
6. Selectie relevantie titel en abstract	19 artikelen
7. Selectie relevantie kwaliteit	8 artikelen

Via backward en forward snowballing in de uiteindelijk geselecteerde acht artikelen is een additioneel artikel gevonden dat nog aanvullende informatie opleverde. Daardoor is nog een extra artikel toegevoegd waarmee het totaal additionele artikelen negen is geworden. Van de relevante achttien artikelen, inclusief de door de Open Universiteit beschikbaar gestelde artikelen, is een samenvatting geschreven van 300-500 woorden die met kleurcodes zijn samengevoegd en zijn uitgeschreven tot een synthese welke is weergegeven in de paragraaf resultaten en conclusies.

Resultaten en conclusies

Waardecreatie

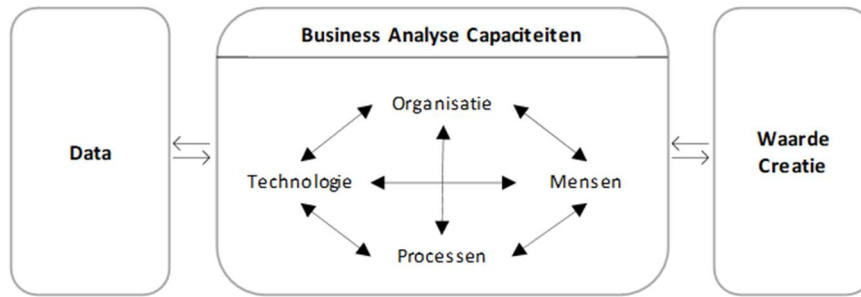
Het creëren van waarde vanuit data-analyse vraagt om een datastrategie die is afgestemd op de algehele strategie van de organisatie. Data waardecreatie is niet iets magisch maar vergt inspanning en tijd (Provost, 2017).

In een organisatie met een data gedreven strategie is de mening van de HiPPO (Highest Paid Person's Opinion) niet van doorslaggevend belang bij het nemen van beslissingen maar zijn dit de inzichten die worden gegenereerd vanuit data-analyse. Leaders moeten het voortouw nemen bij het identificeren van data-analyse mogelijkheden en deze vertalen naar doelen en projecten (McAfee, 2012). De focus zou bij de formulering van data-analyseprojecten moeten liggen op business problemen waarop data-analyse eenvoudig kan worden toegepast en op projecten die de hoogste impact zullen hebben. Omdat data-analyseprojecten een hoge mate van onzekerheid hebben zullen deze ook op die manier moeten worden ingericht (Provost, 2017). Voor het realiseren van data gedreven beslissingen vanuit data-analyseprojecten zullen organisaties eerst hun data-analyseafdeling vorm moeten geven met daarin de juiste mensen en vaardigheden (Vidgen, 2017). Ook moet op leiderschapsniveau de fundamentele basiskennis van data-analyse aanwezig zijn zodat samen met de data-analyseafdeling de juiste beslissingen ten aanzien van de organisatie inrichting worden genomen (Provost, 2017). De volgende stap is dat de resultaten van de data-analyses eenvoudig en visueel aantrekkelijk worden weergegeven voor de eindgebruikers en beslissers, bijvoorbeeld doormiddel van dashboards. Uiteindelijk moet de informatie worden geïntegreerd in de operationele processen zodat de juiste actie kan worden ondernomen op de juiste momenten (Lavalle, 2011). Hiermee wordt een data gedreven cultuur gecreëerd waarin analytische werknemers kunnen excelleren, en beslissingen worden genomen op basis van data-analyse, afgestemd op de algehele organisatiestrategie (Vidgen, 2017; Provost, 2017).

Om te komen tot het identificeren van data-analyse kansen, het genereren van waardevolle inzichten en het succesvol toepassen daarvan is het nodig dat organisaties beschikken over een bepaalde mate van data-analyse capaciteiten (Mikalef, 2019). Succesvolle toepassing kan worden gemeten door verbeteringen in winstgevendheid, marktaandeel, groei, innovatie, kostenleiderschap en levertijd af te zetten tegen de prestaties van de concurrentie (Mikalef P. B., 2019). Het is van belang dat organisaties hun bestaande data-analysecapaciteiten behouden en deze stap voor stap uitbreiden richting de lange termijn doelen van de datastrategie. Bij het definiëren van de datastrategie en doelen moeten organisaties flexibel blijven om continue aansluiting te houden met veranderende organisatie strategieën en doelen (Lavalle, 2011).

Data-analysecapaciteiten

Een belangrijke factor in het omzetten van data naar succesvolle waardecreatie zijn de data-analysecapaciteiten van de organisatie. Data-analysecapaciteiten kunnen worden gedefinieerd als het vermogen van een organisatie om data te verzamelen en te analyseren zodat inzichten worden verkregen door het effectief organiseren en inzetten van data, technologie en talent. Vidgen (2017) noemt dit businessanalysecapaciteiten en onderscheidt de vier dimensies technologie, organisatie, mensen en processen. Naast hun invloed op het omzetten van data in waardecreatie hebben die ook onderling invloed op elkaar (Vidgen, 2017). Deze relatie is visueel weergegeven in figuur 2.



Figuur 2: Business Analyse Capaciteiten (Vidgen, 2017)

Mikalef (2019) benoemt in zijn onderzoek naar data-analysecapaciteiten dezelfde vier factoren als Vidgen (2017) en ontdekte dat ook de contextuele omstandigheden, zoals de dynamiek van de industrie waarin de organisatie opereert en de omvang van de organisatie van invloed zijn op de prestaties van de data-analysecapaciteiten. In de eerdere paragraaf waardecreatie is beschreven wat waardecreatie uit het model van Vidgen (2017) voor organisaties betekend. Onderstaand zullen de betekenis van data en de data-analysecapaciteiten verder worden toegelicht.

Data

Een organisatie moet in staat zijn big data te verzamelen en zowel interne als externe data met elkaar te integreren (Mikalef P. B., 2019). De kwaliteit van de data hoeft niet perfect te zijn maar wel geschikt voor het succesvol toepassen van de data-analyse. Ook de (realtime) beschikbaarheid en de consistentie van data zijn belangrijk. De vele verschillende systemen waarin data zijn opgeslagen vragen om het definiëren van één plek waarin bepaalde informatie wordt bewaard. Wanneer de kwaliteit van de data beter is reduceert dit de tijd die nodig is voor het ontsluiten en opschonen van de data. Deze tijd kan dan besteed worden aan het doen van analyses en creëren van inzichten (Vidgen, 2017).

Organisatie

De transformatie naar een data gedreven organisatie vraagt om een radicaal herontwerp van de organisatie inrichting en een cultuurverandering die jaren in beslag zal nemen (Vidgen, 2017; Mikalef, 2019). Structuur zal moeten worden aangebracht in wie welke rol en verantwoordelijkheid heeft bij het beheren en gebruiken van data. Ook zullen relationele afspraken moeten worden gemaakt om kennis te delen, medewerkers op te leiden en voor de strategische planning (Mikalef P. B., 2019).

Processen

Het verzamelen, bewerken en toepassen van grote hoeveelheden data vraagt van organisaties dat daarvoor procedures worden opgesteld, waarbij er ook rekening wordt gehouden met de ethische aspecten rondom data (Vidgen, 2017; Mikalef P. B., 2019). Voor het analyseren en managen van data-analyseprojecten zijn formele methodes nodig. De complexiteit van big data vraagt om flexibele en cross functionele processen en data-analysesmethoden. De keuze voor een methode is afhankelijk van het type data en het type inzicht waaraan wordt gewerkt (Mikalef P. B., 2019). In softwareprojecten wordt al jaren succesvol gebruik gemaakt van Agile principes, een iteratieve manier waarmee waarde wordt gecreëerd. Van deze Agile aanpak kan veel worden geleerd in data-analyseprojecten (Vidgen, 2017).

Mensen

Om het waardecreatie proces uit te voeren zijn gedreven en intrinsiek gemotiveerde data scientists nodig met de juiste persoonlijke eigenschappen (nieuwsgierig, probleemgericht, zelfstandig, samenwerken, communicatief), technische vaardigheden (programmeren, statistiek) en managementvaardigheden (kennis van toepassingsgebieden, business behoeften, functionele bijdrage

en evalueren output) (Vidgen, 2017; Mikalef P. B., 2019). Ook moet de data scientists in staat zijn zich nieuwe analysetools en -technieken snel eigen te maken. Data-analyseteams bestaan over het algemeen uit data scientists, business analisten en IT-specialisten die zich respectievelijk richten op het verwerken en analyseren van data, het creëren van businesswaarde en het permanent integreren van de datamodellen in de producten en processen (Vidgen, 2017).

Technologie

Om big data te verzamelen en te verwerken is nieuwe dataopslagtechnologie nodig, evenals een cloud gebaseerde data infrastructuur en visualisatie tools (Mikalef P. B., 2019). De data-analyseresultaten moeten aantrekkelijk en begrijpelijk worden gevisualiseerd zodat de ze een integraal onderdeel worden van het communicatieproces en ook niet technische mensen mee kunnen discussiëren over de implicaties (Vidgen, 2017).

Hoe volwassener de ontwikkeling van de dimensies, hoe beter de organisatie in staat is succesvol waarde te creëren, al is deze curve niet altijd lineair (Mikalef P. &, 2019; Mikalef P. B., 2019). De volwassenheid waarin organisaties over data-analysecapaciteiten beschikken kan opgedeeld worden in drie levels (Lavallo, 2011):

Aspirant: deze organisaties gebruiken data primair voor het automatiseren en efficiënter maken van processen en het verlagen van kosten. Hiervoor zijn beperkte data-analysecapaciteiten nodig.

Ervaren: deze organisaties gaan een stap verder en zoeken actief naar manieren om data te verzamelen en toe te passen zodat de organisatie wordt geoptimaliseerd. Hiervoor zijn uitgebreidere data-analysecapaciteiten nodig.

Getransformeerd: deze organisaties hebben de data-analysecapaciteiten automatisch geïntegreerd in alle processen, daarmee een concurrentievoordeel behaald en zetten deze in om de klantwinst verder te vergroten.

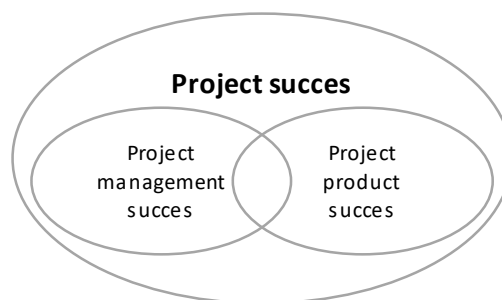
Organisaties kunnen hun data-analysecapaciteiten gefragmenteerd of gecentraliseerd ontwikkelen en borgen. Een gecentraliseerde aanpak is die via een Analytics Competency Center (ACC) met een efficiëntie of innovatie gedreven strategie. Efficiëntie gedreven ACC's ondersteunen voornamelijk interne afdelingen bij het vergroten van hun efficiëntie. Innovatie gedreven ACC's richten zich met name op het verbeteren van de producten en service voor de klanten. Een ACC, en in het bijzonder wanneer die innovatie gedreven is, ondersteunt de transformatie naar een data gedreven organisatie (Schüritz, 2017). Organisaties met een hogere volwassenheid in hun data-analysecapaciteiten maken vaker gebruik van een ACC (Lavallo, 2011). Organisatie die te weinig competenties hebben in hun data-analysecapaciteiten kunnen deze het beste buiten hun organisatie zoeken wanneer zij een datastrategie en data-analyseprojecten succesvol willen implementeren (Nemati, 2003).

Alhoewel datakwaliteit, databeheer en technologische tools absolute voorwaarden zijn voor het succes van data gedreven waardecreatie liggen daar niet de grootste uitdagingen. De grootste obstakels waaraan organisaties moeten werken zijn hun datastrategie, een data gedreven cultuur (Vidgen, 2017; Mikalef P. B., 2019; Lavallo, 2011) en de ethische aspecten (Vidgen, 2017; Mikalef P. B., 2019), die allen onderdeel zijn van de data-analysecapaciteit organisatie.

Data science projecten

Om te komen tot waardecreatie vanuit big data worden data-analyseprojecten uitgevoerd. Bij data-analyseprojecten wordt met geavanceerde methoden en principes informatie en kennis uit grote volumes data met een grote variëteit aan data gehaald door het ontwikkelen van een hoge snelheid in het verzamelen, onderzoeken en analyseren van die data (Vidgen, 2017; Li, 2016). Data science projecten zijn vergelijkbaar met data-analyseprojecten en hebben eveneens gelijkenis met

softwareprojecten. Het succes van een softwareproject is afhankelijk van het projectmanagement succes en het project product succes (Sudhakar, 2012). Deze relatie is visueel weergegeven in figuur 3. Product succes wordt onder meer gerelateerd aan de accuraatheid en betrouwbaarheid van de output. Projectmanagement succes wordt gelinkt aan doelen en transparantie op de factoren kosten, tijd en kwaliteit. Projecten die eerder worden opgeleverd met minder omvang en functionaliteit, en minder resources dan verwacht, worden als meer succesvol ervaren (Sudhakar, 2012). Dit bevestigt bevindingen dat succesvolle data-analyseprojecten korter moeten zijn met minder resources en een gereduceerde projectscope (Nemati, 2003).



Figuur 3: Project succes (Sudhakar, 2012)

Net als bij het proces om big data om te zetten in waardecreatie spelen de aanwezigheid en volwassenheid van de dimensies data, organisatie, processen, mensen en technologie een rol in het succes van een data science project (Koronios, 2015; Sudhakar, 2012; Dutta, 2015; Saltz J. S., 2017). Daarnaast spelen bij een aan data science gerelateerd softwareproject ook de factoren communicatie, team, product en projectmanagement een belangrijke rol (Sudhakar, 2012; Dutta, 2015).

Data-analyseprojecten zouden zich voornamelijk moeten richten op het oplossen van echte business problemen, en slechts beperkt op het willekeurig experimenteren met en verkennen van data (Vidgen, 2017; Provost, 2017). De vorming van een data science project komt vanuit de business of wordt geïnitieerd vanuit een gespecialiseerde ACC. Nadat besloten is een project te starten wordt een projectteam samengesteld, mogelijk bestaande uit meerdere personen met verschillende achtergronden en vanuit verschillende locaties (Saltz J. S., 2017; Schüritz, 2017). De projectsamenstelling, methoden en doelen hangen sterk af van de analysecontext van het project. Het kan daarbij gaan om hypothese generatie, data generatief onderzoek en het testen van hypotheses. Aan de hand van de karakteristieken ontdekken en infrastructuur kunnen er vier soorten data science projecten worden onderscheiden (Saltz J. S., 2017):

1. Hard to justify. Er is nog onduidelijkheid over het gewenste resultaat maar op voorhand is het wel nodig grote investeringen te doen om het project uit te kunnen voeren.
2. Exploratory. De verwachte uitkomsten zijn niet helemaal duidelijk bij dit soort projecten maar de investering is laag waardoor het makkelijker is gewoon iets te proberen.
3. Well-defined. Deze projecten vragen om significante effort maar de doelen zijn ook duidelijk waardoor bedrijven beter kunnen inschatten of de investering het waard is.
4. Smaller data. Deze projecten vergen niet veel van de infrastructuur en zijn vaak duidelijk over de gewenste uitkomsten.

Nadat het project is gevormd wordt het uitgevoerd aan de hand van een data mining proces (Schüritz, 2017). Het meest gebruikte proces om kennis te onttrekken aan data is het CRISP-DM model dat bestaat uit de zes fasen business understanding, data understanding, data preparation, modelling, evaluation en deployment. Het model beschrijft de taken die gedaan moeten worden in de verschillende fasen van een data mining proces, maar niet hoe die taken moeten worden gedaan. Door de toegenomen complexiteit van big data voldoet het CRISP-DM model niet meer op taken gerelateerd

aan projectmanagement processen, integrale processen en organisatie processen (Mariscal, 2010). Het is nodig om aanvullend op het CRISP-DM proces meer aandacht te besteden aan de probleemformulering fase, zodat realistische verwachtingen worden geschapen ten aanzien van de analyse uitkomsten. Ook is het nodig de modelmanagementfasen te monitoren en flexibeler te zijn in het heen en weer bewegen tussen de verschillende fasen van het proces (Li, 2016). Wanneer de doelstellingen van het project zijn gehaald wordt het product geïmplementeerd in de vaste producten en processen van de organisatie (Schüritz, 2017).

De ontwikkeling in de modellen en processen waarop data science projecten worden aangepakt heeft parallellen met software development. Ook daar was er eerst een focus op taken en daarna op andere zaken zoals de samenwerking tussen personen. Op den duur zijn in software development steeds meer Agile methoden toegepast. Een duidelijk verschil met softwareprojecten is dat dataprojecten veel meer gefocust zijn op het type data, de beschikbaarheid, kwaliteit en de tijdigheid van data. Het zomaar overnemen van deze Agile methoden zou dan ook niet verstandig zijn (Saltz J. S., 2017). De Agile principes interactie en iteratie, die veelvuldig gebruikt worden in software development, dragen ook bij aan het succes van een data mining project (Schmidt, 2018). Wel is er in de toepassing van verscheidende varianten van de Agile methode, zoals Kanban en Scrum, verschil in de succesvolle toepassing op data science projecten (Saltz J. S., 2017).

Conclusie

Om blijvende waarde te creëren uit big data moeten organisaties beschikken over een datastrategie en -cultuur die is afgestemd op de algehele strategie van de organisatie. In organisaties met een data gedreven cultuur worden beslissingen genomen op basis van informatie verkregen uit data-analyses. Voor het verkrijgen van deze informatie worden data science projecten ingezet. Organisaties moeten hun data science projecten focussen op business problemen met de meeste impact en de scope van de projecten beperken. Bij de totstandkoming van de projecten worden verschillende fasen doorlopen en moeten duidelijke doelstellingen worden opgesteld ten aanzien van kosten, tijd en kwaliteit en tijdens het project moet de status hiervan doorlopend worden geëvalueerd. Het projectteam moet bestaan uit verschillende disciplines met verschillende vaardigheden. Voor het managen van de data science projecten is de CRISP-DM methode goed bruikbaar, mits deze is uitgebreid met elementen die voorzien in meer aandacht voor probleemformulering, modelmanagement en flexibiliteit, bijvoorbeeld door toepassing van de Agile principes interactie en iteratie. Het analysetype, de datakwaliteit en -beschikbaarheid en de volwassenheid van de organisatie in haar data-analysecapaciteiten hebben grote impact op het succesvol managen van een data science project. Daarom moeten deze als zodanig een belangrijke rol spelen bij de totstandkoming en het management van die data science projecten. De organisatorische data-analysecapaciteiten datastrategie, data gedreven cultuur en ethiek vormen de grootste obstakels om blijvende waarde te creëren vanuit big data.

Ondanks deze kennis en het inzicht in een grote verscheidenheid aan succesfactoren en -dimensies mislukt toch meer dan de helft van de data science projecten. Om het inzicht in de oorzaken en mogelijke oplossingen te vergroten wordt een vervolgonderzoek uitgevoerd met verdiepende onderzoeksvragen. Allereerst is het van belang te weten hoe organisaties in de praktijk aankijken het succesvol managen van een data science project en op welke manier dat succes wordt gemeten. Wetende dat de organisatorische en culturele aspecten de grootste oorzaak vormen voor het mislukken van data science projecten is het van belang te onderzoeken welke invloed die aspecten hebben op het succesvol managen van data science projecten, hoe organisaties rekening houden met deze aspecten bij de totstandkoming en het management van data science projecten, op welke wijze de projectmanagementmethoden rekening houden met die aspecten en op welke manier gewerkt kan worden aan het verbeteren van de prestaties op die aspecten.

3. METHODOLOGIE

Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Het doel van het vervolgonderzoek is om een diepgaand inzicht te krijgen hoe organisaties data science projecten in de praktijk managen om te komen tot blijvende waardecreatie vanuit big data. De invloed van de organisatorische en culturele aspecten hebben daarbij extra aandacht. Om dit inzicht te verkrijgen is het nodig te begrijpen wat er gebeurt binnen de context van een organisatie. De meest geschikte vorm om hier inzicht in te krijgen is het uitvoeren van een casestudie onderzoek (Kamminga, 2013). Een casestudie kan gericht zijn op een enkele casus of op meerdere cases en zich toespitsen op de case als geheel of verschillende subonderdelen. Om met dezelfde methodische aanpak bewijs te verzamelen vanuit verschillende bronnen richt dit onderzoek zich op meerdere cases (organisaties). Het onderzoekt het managen van data science projecten vanuit het gehele perspectief van de organisatie en niet enkel vanuit een specifieke afdeling of rol. Daarmee is de het data science project dan ook het object of study. Een studie bij één caseorganisatie is minder geschikt omdat het onderzoek niet primair gericht is op het testen van een theorie, ontdekken van afwijkingen, vinden van een voorbeeld of veranderingen over tijd (Adolphus, 2013).

Met een casestudie onderzoek kan door direct contact met de werkelijkheid een diepgaand inzicht in complexe relaties tussen het te onderzoeken fenomeen en de variabelen in zijn natuurlijke context worden verkregen zonder dat deze moedwillig wordt beïnvloed. De primaire onderzoeksmethode is het houden van diepgaande interviews met direct betrokkenen bij het managen van data science projecten. Een enquête is hiervoor minder geschikt omdat daarmee vooral beweringen kunnen worden getoetst en deze methode minder geschikt is voor het verkrijgen van een diepgaand inzicht. Bij de op semigestructureerde wijze vormgegeven interviews wordt gebruik gemaakt van een vast interviewscript, zoals gebruikelijk bij een gestructureerd interview, gecombineerd met de mogelijkheid tot doorvragen voor extra verdieping. Waar mogelijk en noodzakelijk geacht, wordt gevraagd om bewijsvoering van de door geïnterviewden gedane uitingen en stellingen. Vanuit de literatuurstudie is een eerste beeld gevormd van de variabelen en de context. Het is daarom niet nodig vooraf extra beeldvormende enquêtes af te nemen (Kamminga, 2013).

De casestudie is kwalitatief van aard. Er wordt data verzameld door intensieve interviews die op een kwantitatieve manier worden geanalyseerd. Het onderzoek streeft naar het beschrijven van een objectief waarneembare werkelijkheid maar heeft niet een puur objectief karakter. Het onderzoek is deels subjectief omdat het nodig is om doormiddel van interpretatie van de interviews te komen tot een dieper inzicht in het succesvol managen van data science projecten. Bij het beoordelen van de informatie wordt een beroep gedaan op het menselijke vermogen van de onderzoeker om patronen en logica te ontdekken. Iedere geïnterviewde heeft een eigen beeld van de werkelijkheid die geïnterpreteerd moet worden in overeenstemming met de eerder bestudeerde wetenschappelijke kennis. De opzet van het onderzoek kan daarmee worden geduid als objectivistisch positivisme zonder formulering van harde hypothesen die getoetst worden maar wel met een afgebakend framework waarbinnen het beschrijvende onderzoek zich afspeelt (Martin, 2013).

Technisch ontwerp: uitwerking van de methode

Het onderzoek wordt uitgevoerd bij organisaties die tijdens de afgelopen twaalf maanden een data science project hebben afgerond met als doel het creëren van blijvende waarde vanuit big data. Deze organisaties moeten enige mate van omvang hebben zodat de omgevingsfactoren voldoende complex zijn. Ook moeten organisaties bereid zijn kennis te delen in interviews, aangevuld met het mogelijk inzage geven in verschillende documenten.

Om een beeld te vormen vanuit meerdere perspectieven worden minimaal drie caseorganisaties onderzocht en worden semigestructureerde interviews gehouden met personen in verschillende

rollen binnen dezelfde organisatie. Per caseorganisatie één interview op organisatieniveau (directie) en één interview op projectniveau (management) wat resulteert in een totaal van minimaal zes interviews zodat inzage wordt verkregen vanuit organisatieperspectief en vanuit het projectperspectief en de mogelijke verschillen daartussen.

Het benaderen van organisaties gebeurt in het persoonlijke netwerk van de onderzoeker. Wanneer onderzoeksresultaten bij meerdere caseorganisaties dezelfde theorie ondersteunen leidt dat tot een betere analytische generaliseerbaarheid (Gibbert, 2010). Daarom zullen minimaal drie verschillende caseorganisaties worden onderzocht aan de hand van hetzelfde semigestructureerde interview. Het uiteindelijke aantal organisaties dat wordt bestudeerd is afhankelijk van het aantal beschikbare en geschikte organisaties en de beschikbare tijd van de onderzoeker. Voordat het interview plaatsvindt wordt de geschiktheid van de organisatie verkend.

De interviews worden gehouden op de reguliere arbeidslocatie van de geïnterviewde en vinden plaats in een aparte vergaderlocatie. Dit omdat het belangrijk is dat de geïnterviewde geconcentreerd is op het interview en niet wordt afgeleid door andere (dagelijkse) werkzaamheden. De interviews duren 60-90 minuten en worden na toestemming opgenomen zodat gedane uitspraken later goed kunnen worden geanalyseerd. Het semigestructureerde interview, dat te vinden is in bijlage 1, is gebaseerd op de in de literatuurstudie gevonden data-analysecapaciteiten en data science projectfactoren. Hiermee wordt gestreefd naar een zo zuiver mogelijke chain of evidence. Tijdens het interview wordt eerst een globale toelichting gegeven op het onderwerp en het onderzoeksdoel. De eerste vragen gaan over het creëren van waarde uit big data en de inzet van data science projecten daarbij. Aansluitend worden meer data science projectmanagement vragen gesteld. Daarna volgt een set vragen over de data-analysecapaciteiten van de organisatie en de invloed daarvan op het succesvol managen van data science projecten. Tijdens het interview is de onderzoeker extra alert op de houding van de geïnterviewde om mede te beoordelen of doorgevraagd moet worden op een bepaald onderwerp.

Voordat het interview start wordt de mogelijkheid aangegeven anoniem deel te nemen aan het interview en wordt toestemming gevraagd om het interview op te nemen voor verdere analyse. Afsluitend wordt gevraagd of er documenten zijn die de uitspraken van de geïnterviewde ondersteunen en of daar inzicht in verkregen kan worden (Adolphus, 2013).

Gegevensanalyse

De analyse van de interviews wordt uitgevoerd aan de hand de 10-stappenmethode in Excel en Word (Ose, 2016). Deze stappen zijn:

1. Verzamelen van de data
2. Transcriberen van de audiobestanden
3. Omzetten tekst van Word naar Excel
4. Voorbereiden Excel bestand voor coderen
5. Coderen in Excel
6. Voorbereiden van code voor sorteren
7. Sorteren van de data
8. Omzetten tekst van Excel naar Word
9. Sorteren tekst in een logische manier
10. Analyseren van de data

Nadat de interviews zijn gehouden worden deze getranscribeerd en opgeslagen als Word document. De gemaakte notities tijdens het interview en de eventueel ontvangen aanvullende documenten worden bestudeerd om te beoordelen of deze de transcripties ondersteunen, aanvullen of ontkrachten. De transcripties worden gesorteerd, gestructureerd en gecodeerd in een boomstructuur. De boomstructuur bestaat uit thema's, categorieën en items verkregen uit de literatuurstudie. Er zijn

vier thema's, te weten; de kenmerken van de organisaties en geïnterviewden, de projectfasen en methoden, de projectfactoren en de data-analysecapaciteiten. De kenmerken van de organisaties worden gebaseerd op de industrie, omvang en data science ervaring. De gedetailleerdere structuur van de projectfasen wordt gebaseerd op de gegeven antwoorden van de geïnterviewden. De categorieën en items van de projectfactoren en data-analysecapaciteiten zijn ontleed aan respectievelijk de businessanalysecapaciteiten van Vidgen (2017) en de projectfactoren van Sudhakar (2012).

De teksten van de verschillende interviews worden gesorteerd op onderwerp in een Word document waardoor het eenvoudiger is deze te analyseren (Ose, 2016). Een voordeel van dit 10-stappenplan is dat het pragmatisch gebruik maakt van de al voor de onderzoeker bekende tools Word en Excel. Er hoeft geen nieuwe software eigen gemaakt te worden en de methode is geavanceerder dan het enkel toepassen van kleurcodering in Word. Het analyseren van de data gebeurt met behulp van pattern matching, waarbij de uitkomsten van het onderzoek worden vergeleken met de studieresultaten, en cross-case analysetechnieken, waarbij de uitkomsten van de verschillende cases met elkaar worden vergeleken (Adolphus, 2013).

Reflectie t.a.v. validiteit, betrouwbaarheid en ethische aspecten

Voor het bewaken van een objectief positivistisch onderzoek in haar natuurlijke context bij de toepassing van een casestudie onderzoek is het natural science model het meest invloedrijk. Dit model onderscheidt meerdere onderzoekacties verdeeld over de vier criteria: construct validiteit, interne validiteit, externe validiteit en betrouwbaarheid. De construct validiteit beschrijft in hoeverre het onderzoek daadwerkelijk heeft onderzocht wat het zegt te hebben onderzocht. De interne validiteit geeft aan in hoeverre er een causale relatie aanwezig is tussen de dimensies en het resultaat en is vooral van toepassing op kwantitatief onderzoek. Bij de externe validiteit gaat het erom in hoeverre de resultaten van het onderzoek ook waarde hebben in een andere context. Deze is bij casestudie onderzoek vaak beperkt. Tot slot gaat betrouwbaarheid over de mate waarin andere onderzoekers tot dezelfde conclusies komen wanneer zij hetzelfde onderzoek uitvoeren en analyseren (Gibbert, 2010).

Voor het vergroten van de construct validiteit is een chain of evidence opgebouwd. Op basis van de wetenschappelijke literatuurstudie zijn onderzoeksvragen en een framework ontworpen, is een interview script geoperationaliseerd en is de boomstructuur voor codering vormgegeven. Vanuit meerdere caseorganisaties, perspectieven en onderzoeksmethoden is de onderzoeksvraag onderzocht en is gevraagd om aanvullende en ondersteunende documentatie zodat triangulatie kan worden toegepast. Voor het uitvoeren van de interviews is een protocol opgesteld waarbij alle relevante zaken zoals tijden, locatie en andere interview omstandigheden zijn geregistreerd.

Om te zorgen voor een hogere externe validiteit wordt onderzoek gedaan bij minimaal drie organisaties die geselecteerd zijn op basis van vooraf opgestelde criteria. De mate van verscheidenheid en overeenkomsten tussen de kenmerken van de organisaties geeft eveneens een indicatie over de mate van generaliseerbaarheid. Ook draagt het vergelijken van de onderzoeksresultaten met de eerder gevonden resultaten uit het literatuuronderzoek tot een hogere generaliseerbaarheid.

Er is een onderzoeksprotocol opgesteld dat consequent wordt toegepast zodat onderzoeksfouten worden geminimaliseerd en de aanpak en uitvoer transparant zijn gedocumenteerd.

Met verschillende ethische aspecten van onderzoek is rekening gehouden. Zo worden geïnterviewden niet onder druk gezet en ontvangen ze geen beloning voor deelname aan het onderzoek. Geïnterviewden hebben te allen tijde de mogelijkheid het interview te stoppen en duplicatie te verbieden. Een zorgvuldige omgang met de verkregen data wordt toegezegd. Zowel organisaties als geïnterviewden hebben de mogelijkheid anoniem deel te nemen aan het onderzoek. Hun privacy is daarmee gerespecteerd.

4. RESULTATEN

De caseorganisaties

Vanwege de beperkingen door het coronavirus is de initiële onderzoeksmethode niet volledig toegepast. Uiteindelijk zijn interviews gehouden bij zes verschillende organisaties met uiteenlopende kenmerken qua omvang en activiteiten. Eveneens zijn door de beperkingen de interviews online afgenomen via Microsoft Teams in plaats van in een real life setting en zijn de interviews met behulp van een AI-tool getranscribeerd. De duur van de interviews betrof tussen de 60 en 90 minuten. Tijdens het interview gaven de geïnterviewden blijk van een goede kennis van big data, data science en data science projectmanagement, door de betekenis van deze begrippen in eigen termen te verwoorden in lijn met de in dit onderzoek gehanteerde omschrijvingen. Aan het begin van het interview is gevraagd naar kenmerken van de organisatie, mede om nogmaals te toetsen of organisaties daadwerkelijk voldoen aan de vooraf opgestelde criteria. Per e-mail zijn in een later stadium aanvullende kenmerken verzameld. Van twee organisaties is inzicht verkregen in projectdocumentatie. Een overzicht van de organisatiekenmerken is te vinden in tabel 3. Daarbij zijn fictieve bedrijfsnamen gebruikt. In bijlage 2 is een interviewmatrix opgenomen waarbij inzicht wordt gegeven welke vragen zijn gebruikt voor het bepalen van de kenmerken van de organisaties. In bijlage 3 is per organisatie een korte beschrijving van de activiteiten, de geïnterviewde, recent uitgevoerde data science projecten, de algehele strategie van de organisatie en de rol van data daarbij, opgenomen.

Tabel 3: kenmerken van de caseorganisaties

Bedrijfsnaam <i>(fictief)</i>	EduTrans	MobiSoft	DataForce	LogiPlan	MultiTech	HighBasics
Industrie	Commerciële opleider	Software ontwikkelaar	Data science consultancy	ERP SaaS aanbieder	Technologie concern	Producent grondstoffen
Rol geïnterviewde	Directie	Management	Directie	Directie	Management	Management
Organisatie typering	MKB	MKB	MKB	MKB	Multinational	Multinational
Jaaromzet 2019 (euro)	< 5 miljoen	< 5 miljoen	< 5 miljoen	< 5 miljoen	> 1 miljard	> 1 miljard
Aantal medewerkers	10 - 50	10 - 50	< 10	10 - 50	> 1000	> 1000
Data science ervaring	< 1 jaar	< 1 jaar	> 10 jaar	< 1 jaar	> 5 jaar	> 15 jaar
Data science projecten	< 5 per jaar	5 - 10 per jaar	5 - 10 per jaar	< 5 per jaar	5 - 10 per jaar	> 10 per jaar
Volwassenheid DAC	Aspirant	Aspirant	Ervaren	Aspirant	Ervaren	Ervaren
Data afdeling	Centraal	Centraal	Centraal	Centraal	Decentraal	Decentraal
Projectmethode	Agile Kanban	Agile Kanban	CRISP-DM	Agile Scrum	Agile Kanban	Agile Scrum

** De weergegeven informatie heeft betrekking op de Nederlandse divisie.*

DataForce is een data science consultancy bedrijf dat voornamelijk voor grotere organisaties, zoals multinationals, data science projecten uitvoert. Ondanks dit gegeven is bij de organisatietypering gekeken naar de feitelijke eigenschappen van DataForce en heeft het daarom de typering 'MKB' gekregen. Ondanks dat DataForce 1,5 jaar bestaat is hun data science ervaring op >10 jaar ingeschaald vanwege de ruime ervaring van de gespecialiseerde medewerkers. Aan de hand van de door de geïnterviewden gegeven toelichting op de doelen waarmee data science projecten worden ingezet en de beschrijving van de aanwezigheid van data-analysecapaciteiten, zijn de volwassenheidsstadia van Lavalley (2011) toegekend aan de caseorganisaties.

Tussen de kenmerken van de organisaties vallen in hoofdlijnen vijf verschillen op. De rol van de geïnterviewden is 50/50 verdeeld tussen directie en management. De omvang van de omzet en het aantal medewerkers is sterk verschillend tussen de MKB-bedrijven en de multinationals. De organisaties met meer dan 5 jaar ervaring in data science projecten, inclusief DataForce, hebben een ervaren volwassenheid van hun data-analysecapaciteiten. Bij de multinationals is de data science

afdeling decentraal ingericht op divisie of clusterniveau. Bij de MKB-bedrijven is deze afdeling centraal vormgegeven. En bij het managen van de data science projecten wordt er gebruik gemaakt van drie verschillende primaire methoden. Tijdens de analyse van de interviewresultaten is met een cross case analyse extra aandacht geschonken aan het vinden van herkenbare patronen in verschillen of overeenkomsten tussen de organisaties op aspirant en ervaren niveau en ook specifiek naar de vijf hiervoor genoemde kenmerken van de organisaties. Waar patronen zijn gevonden en deze relevant lijken, zijn deze opgenomen en toegelicht bij de verdere analyses. De relatie tussen de omvang van de organisatie waar, en het volwassenheidsniveau van, de data-analysecapaciteiten waarmee de data science projecten zijn uitgevoerd, is sterk. Datzelfde geldt voor het aantal jaren dat al data science projecten worden uitgevoerd. Wanneer er in dit onderzoek wordt gesproken over ervaren data-analysecapaciteiten heeft dit betrekking op data science projecten die bij multinationals zijn uitgevoerd, met al meer dan 5 jaar ervaring daarin. Bij aspirante data-analysecapaciteiten betreft het data science projecten die bij MKB-bedrijven zijn uitgevoerd die daar minder dan 1 jaar ervaring in hebben.

Codering en opbouw

Vanuit de onderzoeksvragen en het literatuuronderzoek is het semigestructureerde interview opgebouwd waarin doelgerichte vragen zijn gesteld over het blijvend creëren van waarde uit big data, de totstandkoming en het managen van data science projecten, de data-analysecapaciteiten van een organisatie en hun invloed op het succesvol managen van data science projecten. De project gerelateerde factoren zijn gecodeerd zoals die door Sudhakar (2012) zijn omschreven en ingedeeld naar de categorieën product, projectmanagement, team en communicatie. De verschillende projectfasen zijn ontleend aan de omschrijving zoals die door de geïnterviewden zijn gegeven en de tijdvolgelijkheid daarin. Aan de hand van de door Vidgen (2017) beschreven items onder de businessanalysecapaciteiten organisatie, mensen, proces en technologie, aangevuld met de beschrijvingen van de elementen data en waardecreatie zijn de categorieën en items onder het thema data-analysecapaciteiten gecodeerd. De opbouw van de codes is achtereenvolgend thema, categorie, item. In figuur 4 is een lijst van thema's en categorieën weergegeven. Achter de codering staat het aantal keer dat deze code voorkomt. In bijlage 3 is een longlist aan gebruikte coderingen, inclusief items, opgenomen.

<p>Data science projectfactoren (139)</p> <ul style="list-style-type: none"> Communicatie (11) Product (13) Projectmanagement (77) Team (38) 	<p>Projectfasen (44)</p> <ul style="list-style-type: none"> Aanleiding (11) Afstemming organisatie (5) Probleem begrijpen (7) Middelen verkennen (9) Project vorming (12) 	<p>Data-analysecapaciteiten (126)</p> <ul style="list-style-type: none"> Data (13) Mensen (25) Organisatie (44) Processen (8) Technologie (3) Waardecreatie (18)
---	---	---

Figuur 4: interviewcodering naar thema's en categorieën

Allereerst wordt beschreven hoe organisaties in de praktijk aankijken tegen het creëren van waarde uit big data, welke rol data science projecten daarin hebben en op welke manier het succes van deze data science projecten wordt gemeten. Ondanks dat waardecreatie in de codering onderdeel is van het thema data-analysecapaciteiten is dit in essentie het doel waarvoor data science projecten worden ingezet en wordt het daarom apart uitgelicht in de eerste paragraaf. In de paragraaf projectfasen worden de verschillende fasen beschreven waarin de caseorganisaties hun data science projecten hebben gevormd en welke activiteiten ze daarbinnen uitvoeren. De paragraaf projectmanagementmethoden beschrijft de gebruikte projectmanagementmethoden, de motivatie voor deze methode en hun voor- en nadelen. In de laatste paragraaf data-analysecapaciteiten zal dieper worden ingegaan op de impact die data en de data-analysecapaciteiten afzonderlijk hebben op het succesvol managen van een data science project en op welke onderdelen van de verschillende

projectfactoren deze impact het grootste is. Aan de hand van deze opbouw worden de sub-onderzoeksvragen en verdiepende vragen, gebaseerd op de literatuurstudie, in logische volgorde één voor één beantwoord vanuit het empirische onderzoek. Tijdens het analyseren van de resultaten zijn cross case analyses gemaakt met betrekking tot de organisatiekenmerken en gebruikte projectmanagementmethoden. Waar relevant wordt in de betreffende paragraaf een toelichting gegeven.

Waardecreatie

Alle caseorganisaties erkennen het belang van big data binnen de algehele bedrijfsstrategie. Voor de één speelt het een grote en belangrijke rol, voor de ander is dit in toenemende mate het geval. Bij DataForce is “data de brandstof voor onze klanten en de eigen organisatie”. Voor MobiSoft geldt dat “data is de basis en daar bouwen we bovenop”. Van de zes geïnterviewde organisaties zijn er vier bedrijven die data-analyses commercieel inzetten en waarbij data een klantproduct is.

Om blijvende waarde te creëren vanuit big data worden data science projecten ingezet met als doel processen efficiënter te maken, informatie voor beslissingen te ontsluiten of nieuwe producten en diensten te ontwikkelen. Bij aspirant organisaties ligt de nadruk meer op informatiewinning en bij ervaren organisaties op procesoptimalisatie.

Het succes van een data science project wordt door de organisaties op verschillende manieren bepaald, wat is weergegeven in tabel 4. Waardecreatie is daarbij de meest benoemde categorie (8x), gevolgd door product (4x) en projectmanagement (3x). Het belangrijkste criterium is of het project daadwerkelijk heeft geleid tot blijvende waardecreatie in de vorm van waarde voor de klant, een business case of daadwerkelijke toepassing in de bestaande processen. EduTrans noemt het project niet gelukt als het een eenmalige exercitie is geweest die niet is geïntegreerd in vaste processen. Daarnaast wordt het succes van een data science project ook bepaald aan de hand van de mate waarin het is gelukt te voldoen aan de product en projectmanagement uitgangspunten. Voorbeelden hiervan zijn dat DataForce, MultiTech en HighBasics vooraf de performance meten van het betreffende proces (0-meting) en dit herhaalden na het project om zo het effect te kunnen bepalen. HighBasics berekent daarbij alles terug naar geld. “Dat is de makkelijkste manier om te valideren of iets succesvol is of niet”.

Tabel 4: manieren waarop caseorganisaties data science projectsucces bepalen

Thema's, categorieën & items	Aspirant	Ervaren	Totaal
Data science projectfactoren	4	3	7
Product	3	1	4
Producteisen	1	1	2
Productkwaliteit	2		2
Projectmanagement	1	2	3
Projectdoelen		2	2
Projectplanning	1		1
Data-analysecapaciteiten	5	3	8
Waardecreatie	5	3	8
Businesscase	1	1	2
Klantwaarde	2	2	4
Toepassing	2		2

In het bepalen van het succes van een data science project ervaren organisaties verschillende knelpunten. Deze knelpunten zijn weergegeven in tabel 5. Het grootste knelpunt bevindt zich duidelijk in de data-analysecapaciteiten (14x) en niet in de data science projectfactoren (2x). Items in de categorie waardecreatie zijn het meest genoemd (8x), gevolgd door processen (6x) en projectmanagement (2x). Voor ervaren organisaties is de uiteindelijke toepassing van het resultaat uit de data science projecten het grootste probleem waardoor waardecreatie, en dus projectsucces

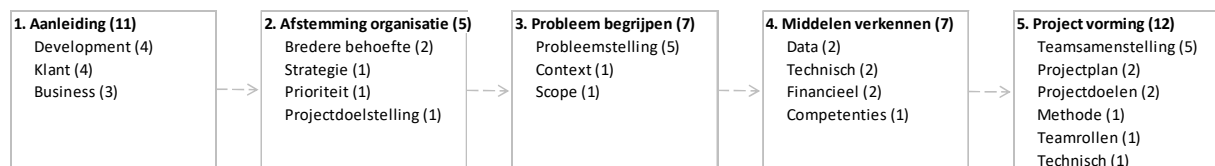
uitblijft, typerend verwoord door MultiTech met “als zij niet bereid zijn een oplossing toe te passen dan heeft het nog geen resultaat”. Een ander belangrijk knelpunt, voornamelijk bij aspirant organisaties, is het ontbreken van prestatieprocesmanagement om überhaupt de prestaties te meten. LogiPlan geeft aan dat het hebben van de juiste meetinstrumenten een probleem is voor het meten van succes. Bij DataForce ontbreekt het bij het meten van projectsucces regelmatig aan een benchmark en is er afhankelijkheid van externe factoren, zoals andere afdelingen. MultiTech geeft aan dat het niet altijd mogelijk is om een 0-meting te doen.

Tabel 5: knelpunten bij het meten van data science projectsucces

Thema's, categorieën & items	Aspirant	Ervaren	Totaal
Data science projectfactoren	1	1	2
Projectmanagement	1	1	2
Projectdoelen	1	1	2
Data-analysecapaciteiten	8	6	14
Processen	4	2	6
Prestatiemanagement	4	2	6
Waardecreatie	4	4	8
Businesscase	1		1
Klantwaarde	1		1
Toepassing	2	4	6

Projectfasen

De totstandkoming van data science projecten verloopt via vijf vergelijkbare processtappen bij de verschillende organisaties. In figuur 5 zijn de processtappen visueel weergegeven, inclusief de genoemde activiteiten die daarin plaatsvinden. Het aantal keer dat die items zijn genoemd is tussen haakjes weergegeven. Na figuur 5 zal per processtap een verdere inhoudelijke toelichting worden gegeven.



Figuur 5: processtappen in de totstandkoming van data science projecten

Allereerst is er een aanleiding. Nadat er een verzoek is gekomen wordt dit eerst getoetst in de organisatie. Vervolgens wordt met verschillende stakeholders de verdieping gezocht om het probleem tot in detail te begrijpen en te verkennen welke middelen nodig zijn om tot een oplossing te komen. Wanneer dit duidelijk is wordt het project gevormd en gestart. Er is voor deze stappen geen generiek toegepaste methode gevonden. DataForce en HighBasics gebruiken een door henzelf ontwikkeld model voor dit proces. Bij HighBasics wordt dit de ‘project intake funnel’ genoemd. “Je moet dat zien als een grote trechter wat steeds nauwer toeloopt waar heel veel uit wordt gefilterd”.

1. Aanleiding: het ontstaan van een data science project start met een verzoek vanuit de development afdeling, de klant (in het geval van de consultancy organisaties) of vanuit de business. De reden voor een development afdeling om te starten met een data science project is in het geval van MobiSoft, DataForce, LogiPlan en HighBasics de (door)ontwikkeling van (klant)producten gebaseerd op data. Bij een aanleiding vanuit de business geven EduTrans, MultiTech en HighBasics aan dat projecten in eerste instantie zijn ontstaan vanuit het management of de development afdeling en dat later steeds meer verzoeken vanuit de operatie kwamen. Door HighBasics is dat verwoord als “In het begin kwamen veel projecten

bij ons vandaan. Door het halen van heel veel succes gaat het rond bij de business. Dan gaan ze zich op een gegeven moment zelf aanmelden”.

2. Afstemming organisatie: nadat er een verzoek is gekomen tot een data science project kijken EduTrans en HighBasics of deze behoefte breder in de organisatie leeft en toetsen EduTrans en LogiPlan of dit past bij de strategie en prioriteiten. MultiTech bepaalt in deze fase de doelstelling voor een eventueel project.

3. Probleem begrijpen: DataForce geeft aan dat business understanding heel erg belangrijk is binnen de door hen gebruikte CRISP-DM methode. “We doen altijd een eerste scan met de organisatie om dit (context, scope en centrale probleemstelling) helder te hebben”. Deze inventarisatie wordt later gebruikt om een inschatting te maken van de oplossing en wat daarvoor nodig is. Om het probleem goed te begrijpen worden in deze fase door alle organisaties al diverse stakeholders (klanten, business) en dataspecialisten betrokken.

4. Middelen verkennen: zodra de probleemstelling duidelijk is en bekend is welke oplossingsrichtingen er zijn, worden de benodigdheden op data, technisch en menselijk vlak geïnventariseerd. DataForce en LogiPlan werken in deze fase een business case uit om te bepalen of het data science project (financieel) haalbaar is. HighBasics onderzoekt met een Proof of Concept de kwaliteit van de beschikbare data en bepaalt op basis van de resultaten of het project haalbaar is en wat de doelstellingen worden.

5. Project vorming: voordat het project definitief van start gaat wordt het projectteam samengesteld, worden de rollen verdeeld, wordt de projectmethode bepaald en waar nodig de ontbrekende tooling aangeschaft. De exacte invulling hiervan is afhankelijk van de in processtap drie en vier geïnventariseerde behoeften en benodigdheden. In de projectvormingsfase stellen de organisaties ook een projectplan en projectdoelen op. De analysecontext heeft invloed op een data science project, vooral op de keuzes die worden gemaakt in vormingsfase. Bij hypothese gedreven onderzoek is het stellen van doelen makkelijker, geven EduTrans en MultiTech aan, en exploratief onderzoek zorgt voor een iets andere invulling van de methode aldus de ervaren organisaties.

Bij de totstandkoming van data science projecten ervaren organisaties knelpunten in hun data-analysecapaciteiten organisatie (8x), mensen (6x) en data (5x), zoals weergegeven in tabel 6. In het geval van mensen gaat het om kennis en vaardigheden van data science in de breedte van de organisatie en bij stakeholders, maar ook over de meer specialistische analytische en technische kennis en vaardigheden. De ervaren organisaties merken op dat het niet goed georganiseerd hebben van de beschikbaarheid en kwaliteit van data voor hen een belangrijk obstakel is. Aspirant organisaties ervaren juist meer problemen in de beschikbare tijd die kan worden vrijgemaakt en het ontbreken van voldoende budget.

Tabel 6: knelpunten in de totstandkoming van data science projecten

Thema's, categorieën & items	Aspirant	Ervaren	Totaal
Data-analysecapaciteiten	10	9	19
Data	1	4	5
Data beschikbaarheid		2	2
Dataarchitectuur	1	1	2
Datakwaliteit		1	1
Mensen	4	2	6
Analysevaardigheden	1		1
Datavaardigheden	2	2	4
Technischevaardigheden	1		1
Organisatie	5	3	8
Datacultuur		1	1
Financieel	2		2
Projectscope		1	1
Tijd	3	1	4

Gevraagd naar de belangrijkste adviezen voor organisaties om hun data science projecten succesvol te vormen en managen is het advies van DataForce kenmerkend: “Denk groot, maar begin klein. Heb een vergezicht maar begin met kleine stapjes waarmee je vandaag resultaat kunt boeken. Maak het meetbaar en durf fouten te maken”. Door andere organisaties aangevuld dat de probleemstelling goed uitgezocht moet worden en er een duidelijk projectdoel dient te zijn en “niet een pot mensen oentrekken en dan het beste er van hopen”, aldus HighBasics.

Projectmanagementmethode

Voor het managen van data science projecten worden drie verschillende methodes door de organisaties gebruikt, te weten: Agile Kanban, Agile Scrum en CRISP-DM. Twee organisaties, die respectievelijk Agile Scrum en CRISP-DM gebruiken, hebben dit aangevuld met een eigen ontwikkeld projectmanagementmodel. Een overzicht van de gebruikte methoden per organisatie is eerder getoond in tabel 3. In tabel 7 zijn de genoemde motieven voor het gebruik van een bepaalde projectmethode weergegeven. De data science projectfactoren (11x) zijn daarbij duidelijk vaker genoemd dan de data-analysecapaciteiten (1x). Bij de keuze voor een projectmethode is het team (7x) het meest genoemde motief. Naast specifieke voor- en nadelen, zoals het iteratief zijn bij Agile Kanban, motiveren de organisaties de keuze voor een bepaalde methode ook omdat ze als specialist of team daarin eerdere positieve ervaring hadden opgedaan of omdat binnen de organisatie wordt gewerkt met die methode. MobiSoft en MultiTech hebben bewust voor Agile Kanban gekozen, en niet voor Agile Scrum vanwege de tijdsintensieve registratie en het gebruik van harde deadlines die de Agile Scrum methode met zich meebrengt. Opmerkelijk is dat slechts één organisatie een methode (CRISP-DM) heeft gekozen omdat die aansluit bij het product data.

Tabel 7: motivatie voor de gebruikte data science projectmethode

Thema's, categorieën & items	Agile Kanban	Agile Scrum	CRISP-DM	Totaal
Data science projectfactoren	6	3	2	11
Product			1	1
Productmanagement			1	1
Projectmanagement	3			3
Projectplanning	2			2
Review & Feedback	1			1
Team	3	3	1	7
Teamcommitment	3	2		5
Teamcompetenties		1	1	2
Data-analysecapaciteiten	1			1
Organisatie	1			1
Datacultuur	1			1

De voordelen van de gekozen methoden zijn weergegeven in tabel 8. De methoden kennen vooral projectmanagement gerelateerde (16x) voordelen. De organisaties geven aan dat de projectmethoden hen helpen het proces te plannen (4x), controle te houden over de voortgang (4x), risico's te bewaken (3x) en (tussentijds) te reviewen en feedback te krijgen (3x). Er is hierin weinig onderscheid of overeenkomst tussen de organisatiekenmerken en de gekozen projectmethode. De organisaties die voor de Agile Kanban hebben gekozen herhalen wederom het iteratieve planningsvoordeel. Datzelfde geldt voor het product aspect in het geval van DataForce. LogiPlan en MultiTech benoemen het geven van tweewekelijkse presentaties aan intern en extern betrokken stakeholders een belangrijk pluspunt. "Er komt dus erg snel en hele goede feedback" aldus LogiPlan.

Tabel 8: voordelen van de gebruikte data science projectmethoden

Thema's, categorieën & items	Agile Kanban	Agile Scrum	CRISP-DM	Totaal
Data science projectfactoren	10	7	3	20
Product	1		1	2
Productkwaliteit	1			1
Productmanagement			1	1
Projectmanagement	8	6	2	16
Project monitoren	1			1
Projectcontrole	1	2	1	4
Projectdoelen		1		1
Projectplanning	3	1		4
Review & Feedback	1	1	1	3
Risicobeheersing	2	1		3
Team	1	1		2
Taakorientatie	1			1
Teamcommitment		1		1

De nadelen van de methoden zijn vooral team (6x) en projectmanagement (4x) gedreven zoals in tabel 9 is weergegeven. Het belangrijkste nadeel van de gekozen Agile Scrum methode is de complexiteit ervan wat afleidt van de taken. Ook moeten deelnemers en stakeholders goed op de hoogte zijn van de werking van de methode en hierin competenties bezitten. Bij EduTrans wordt de Agile Kanban methode door sommige stakeholders als bureaucratisch ervaren en MultiTech merkt op dat Agile Kanban niet voor iedereen werkt, "sommige mensen hebben gewoon deadlines nodig" waardoor er binnen het team beperkt commitment is. MobiSoft en MultiTech merken op dat de Agile Kanban methode soms onoverzichtelijk is en vooraf niet gezegd kan worden hoeveel tijd iets kost en hoe het projectschema eruitziet. Van zowel Agile Kanban als Agile Scrum wordt door respectievelijk EduTrans en LogiPlan aangegeven dat het tijds- en arbeidsintensieve methoden zijn. DataForce heeft voor het

projectmanagement een eigen aanpak ontwikkeld omdat CRISP-DM “niet een methode is om projecten mee te managen maar een methode is om het proces van datamining in kaart te brengen”.

Tabel 9: nadelen van de gebruikte data science projectmethoden

Thema's, categorieën & items	Agile Kanban	Agile Scrum	CRISP-DM	Totaal
Data science projectfactoren	5	4	1	10
Projectmanagement	2	1	1	4
Projectkosten		1		1
Projectschemata	2		1	3
Team	3	3		6
Taakoriëntatie	1	2		3
Teamcommitment	2			2
Teamcompetenties		1		1

Om te borgen dat tijdens het project gewerkt blijft worden aan de projectdoelstellingen worden bij de organisaties projectleiders aangesteld die de voortgang monitoren en hierover rapporteren. In de Agile methoden is hierin voorzien met een vast vergaderritme aan voortgangsbijeenkomsten en een continue review & feedback loop. Agile Scrum, met rollen zoals de scrummaster en product owner, en rituelen zoals de stories, de diverse sprints, de daily standups en presentaties (aan stakeholders) is daarin uitgebreider dan Agile Kanban.

Data-analysecapaciteiten

De organisaties geven aan dat data en de data-analysecapaciteiten grote invloed hebben op het succesvol managen van een data science project. In tabel 10 zijn hiervan de meest genoemde items weergegeven. De data-analysecapaciteiten met de grootste impact zijn de kennis en vaardigheden van data science bij de mensen (8x), zowel met betrekking tot de mogelijkheden als de daadwerkelijke toepassing en het projectmanagement ervan. Naast het menselijke aspect speelt ook de organisatie (7x) een belangrijke rol, vertegenwoordigd door een duidelijke en gedragen datastrategie (4x) en een data gedreven cultuur (2x) in de organisatie. Technologie en processen zijn door de organisaties niet benoemd als capaciteit met veel invloed op het succesvol managen van data science projecten. Een voldoende hoge datakwaliteit (2x) en het beschikken over de data (1x) zijn een voorwaarde om überhaupt data science projecten uit te kunnen voeren.

Tabel 10: data en data-analysecapaciteiten met de meeste impact op het managen van data science projecten

Thema's, categorieën & items	Aspirant	Ervaren	Totaal
Data-analysecapaciteiten	9	9	18
Data	1	2	3
Dataarchitectuur		1	1
Datakwaliteit	1	1	2
Mensen	5	3	8
Analysevaardigheden	2		2
Datavaardigheden	2	3	5
Technischevaardigheden	1		1
Organisatie	3	4	7
Datacultuur	1	1	2
Datastrategie	1	3	4
Tijd	1		1

De data en data-analysecapaciteiten hebben allemaal een andere impact op het succesvol managen van een data science project. In tabel 11 zijn de door de organisaties genoemde invloed van data en de data-analysecapaciteiten op het succesvol managen van data science projecten ingedeeld naar de project gerelateerde factoren communicatie, product, projectmanagement en team. De invloed van data en de data-analysecapaciteiten is het meest in verband gebracht met projectmanagement (29x)

en team gerelateerde factoren (21x). De invloed op product gerelateerde factoren (5x) is slechts beperkt benoemd. De manier waarop data en de data-analysecapaciteiten invloed hebben op het succesvol managen van data science projecten en de verschillende projectfactoren wordt onder tabel 11 verder toegelicht.

Tabel 11: invloed van data en de data-analyse capaciteiten (boven) op data science projectfactoren (links)

Categorieën & Items	Data	Organisatie	Processen	Mensen	Technologie	Totaal
Communicatie		5	1	3		9
Project medewerking		5	1	3		9
Product	4				1	5
Producteisen	3				1	4
Productkwaliteit	1					1
Projectmanagement	8	2	7	6	6	29
Review & Feedback					2	2
Risicobeheersing				1		1
Managementvaardigheden			1			1
Projectdoelen				2		2
Projecteisen	1					1
Projectmethode			3			3
Project monitoren			1	1		2
Projectplanning	4			1	1	6
Projectschema	3	2	2	1	3	11
Team	2	4	4	7	4	21
Empowerment			4	1	2	7
Teamsamenstelling	2	1		1		4
Teamcompetenties		3		4	2	9
Teamcoördinatie				1		1
Totaal	14	11	12	16	11	64

Data: “data is de brandstof van data science projecten”, aldus DataForce. HighBasics bevestigt het onmisbare belang van data als “Als er geen data is kun je niks, heb je niks”. De caseorganisaties zijn eenduidig over het belang van data en de invloed daarvan op het managen van een data science project. Slechte datakwaliteit zorgt ervoor dat het schema vertraging oploopt en projecten langer duren. De data moet dan eerst opgeschoond worden. Ook komt het voor dat de data architectuur niet goed ingericht en toegankelijk is waardoor de data slecht ontsloten kan worden. In beide gevallen zijn er in de projectplanning extra data engineering uren, -kennis en tools nodig. Ook kan een slechte beschikbaarheid of kwaliteit van data zorgen voor beperkingen in de eisen die aan analyses gesteld kunnen worden en de betrouwbaarheid van de resultaten.

Organisatie: alle organisaties geven aan dat de mate waarin een organisatie een data gedreven cultuur heeft, invloed heeft op de bereidheid om mee te werken en te denken aan een data science project. MultiTech verwoordt dat treffend met “Als je geen data gedreven cultuur hebt ben je gedoemd te mislukken. Aan de voorkant ben je dan enthousiast en ga je aan de slag maar de mensen die het moeten gebruiken willen er niet mee werken of veranderen. Het is belangrijk om mensen vanaf het begin goed te betrekken voor die cultuurverandering”. In het project moet dan meer aandacht worden gegeven aan het creëren van draagvlak en overtuigen van de organisatie, wat extra tijd kost en andere competenties van teamleden vraagt. Het kan daarbij helpen open te communiceren en de resultaten uit data-analyses te delen. Ook kan een beperkte ervaring in data gedreven projecten zorgen voor verkeerde inschattingen tijdens het project. EduTrans zegt hierover “Als dat nog een leercurve is zie je dat ook terug in hoe zo’n project loopt”.

Processen: hierbij benoemen de caseorganisaties de gebruikte projectmethode, en in mindere mate de databeschikbaarheid (DataForce, MultiTech). De gekozen projectmethode moet aansluiten op de bestaande manier van werken en bestaande processen binnen een organisatie. Als dat niet het geval is dan beperkt dat de interne samenwerking en afstemming. Ook is het prettig voor de teamleden als er gewerkt wordt met een vertrouwde methode. Vaak vergt dit extra training en opleiding binnen het data science project, of in de organisatie, en een duidelijke aansturing van de projectmanager. MultiTech omschrijft deze ontwikkeling als “Hoe vaker je een methode gebruikt, hoe beter je erin wordt. Je krijgt dan een geoliede machine”.

Mensen: “Je hebt niet alleen technische mensen nodig met goede analytische vaardigheden maar communicatie is ook een heel belangrijk element wat een hard requirement is”. Deze quote van HighBasics vertolkt de essentie wanneer de caseorganisaties de impact van mensen beschrijven op het managen van data science projecten. Communicatie speelt een belangrijke rol in betrekken van stakeholders, het overbrengen van de resultaten en het motiveren van zowel de organisatie als het projectteam. Naast inhoudelijke kennis moet een data science projectmanager daarom goede communicatieve vaardigheden hebben. Als er weinig data scientists in een organisatie aanwezig zijn, zoals bij LogiPlan, maakt dat de organisatie erg afhankelijk, wat een verhoogd risico op fouten en beperkte beschikbaarheid in het project met zich mee brengt. Het inhuren van externen kan op specifieke gebieden een oplossing zijn, aldus HighBasics maar MultiTech ziet dat als nadeel omdat kennis dan niet structureel opgebouwd wordt.

Technologie: de hoeveelheid beschikbare tools is de laatste jaren sterk gestegen en vraagt van data scientists dat ze hun competenties daarin blijven ontwikkelen. Voor niet data scientists kunnen die tools een black box zijn, zoals het geval bij LogiPlan. De keuze en inzet van tools is voor EduTrans onderdeel van de projectvorming. Een goede technologie stelt DataForce, MobiSoft en HighBasics in staat hun data-analyse toepassingen enorm te versnellen, maar het ontbreken van goede tools kan ook juist voor vertraging zorgen zoals LogiPlan heeft ervaren. Het afstemmen en breder gebruiken van een in de organisatie bekende tooling zorgt er in het geval van MobiSoft voor dat reviewen beter mogelijk is, en bij MultiTech en HighBasics dat kennis kan worden gedeeld.

Om de data-analysecapaciteiten in de organisatie te vergroten hebben drie organisaties hun kennis en vaardigheden op directieniveau uitgebreid door invulling te geven aan de functie van Chief Digital Officer (CDO). Deze CDO is verantwoordelijk voor de datastrategie en de ontwikkeling van de benodigde data-analysecapaciteiten. Om mensen mee te nemen in een data gedreven cultuur investeren DataForce en MobiSoft bewust in communicatie. DataForce verwoordt de reden voor hun extra aandacht voor kennisontwikkeling en -overdracht als volgt “Mensen zijn gewend het zo te doen, dus waarom zouden ze het ineens anders gaan doen. Zo worden vaak beslissingen op onderbuikgevoel genomen en niet op basis van data. Ze gaan er dan vanuit dat dat de beste manier is terwijl je uit data kunt kijken of dat inderdaad zo is”.

In tabel 12 is weergegeven welke knelpunten organisaties ervaren bij het op het gewenste niveau brengen van hun data en data-analysecapaciteiten. Aspirant organisaties ervaren daarbij vooral knelpunten in budget (3x), kennis en vaardigheden van data science (3x) en technologische tools (2x) door LogiPlan treffend geformuleerd als “we zijn geen multinational en moeten het met beperkte resources doen, zowel qua mensen als budget en de tools die voor handen zijn”. Ervaren organisaties ervaren daarentegen een tijdsprobleem (2x). De mensen met kennis van zaken van de business en/of data zijn dusdanig druk. “Er is altijd meer vraag dan dat er capaciteit is om die vraag de beantwoorden” aldus DataForce. Ervaren organisaties geven aan dat de cultuur van de organisatie (2x) invloed heeft

op de mate waarin professionals tijd vrij kunnen maken voor data science projecten. Bij de ervaren organisatie HighBasics zijn “de data-analysecapaciteiten op niveau” maar werden de technologische beperkingen (1x) een knelpunt. Daarom wordt de computing power nu verhoogd door het inrichten van een dedicated platform.

Tabel 12: knelpunten bij de ontwikkeling van data-analyse capaciteiten

Thema's, categorieën & items	Aspirant	Ervaren	Totaal
Data-analysecapaciteiten	11	6	17
Mensen	3		3
Datavaardigheden	3		3
Organisatie	6	5	11
Datacultuur	1	2	3
Financieel	3		3
HR beleid		1	1
Projectscope	1		1
Tijd	1	2	3
Technologie	2	1	3
IT beperkingen	2	1	3

De belangrijkste adviezen van de organisaties om data science projecten succesvol te managen, die betrekking hebben op de data en data-analysecapaciteiten van een organisatie, gaan over mensen, data en organisatie. Bij mensen wordt het beschikken over mensen met de juiste data science kennis benadrukt en ook het ontwikkelen van interne data science kennis en ervaring. Voor wat betreft data geven de organisaties aan dat de data beschikbaar moet zijn, van een goede kwaliteit en dat gelijk een schaalbare architectuur neergezet moet worden. EduTrans adviseert dat in het bedrijf niet “in een donker hoekje met data aan de slag” gegaan moet worden en volgens MultiTech wordt “een data gedreven cultuur gecreëerd door mensen mee te nemen en training te geven en show cases te delen”. Deze quotes zijn typerend voor de adviezen rondom de data-analysecapaciteit organisatie. Opvallend is dat technologie weinig wordt genoemd, enkel met als strekking gebruik de goede tools, en dat processen gerelateerd aan de projectmethode niet worden genoemd. Tot slot adviseren twee caseorganisaties dat data science projecten en het aanstellen van data scientists onderdeel moeten zijn van een grotere datastrategie, inclusief ontwikkeling van de benodigde data-analysecapaciteiten.

5. DISCUSSIE, CONCLUSIES EN AANBEVELINGEN

Conclusies

De organisaties geven aan dat data science projecten pas succesvol zijn wanneer daadwerkelijk blijvende waarde is gecreëerd vanuit big data door structurele toepassing en integratie van de resultaten in de vaste processen van de organisatie. Tegelijkertijd geven de ervaren organisaties aan dat het juist ontbreekt aan deze structurele integratie van de resultaten na afronding van het data science project. De aspirant organisaties geven aan niet te beschikken over methoden waarmee het (blijvende) effect van een data science project op de prestaties gemeten kan worden. Een structurele integratie van de resultaten in de bestaande processen van een organisatie zal dan ook integraal onderdeel moeten zijn van het managen en beoordelen van een data science project.

De data-analysecapaciteiten mensen en organisatie hebben de grootste impact op het succesvol tot stand komen en managen van data science projecten, voornamelijk op de projectfactoren project medewerking en teamcompetenties. De organisaties geven aan dat een goede data gedreven cultuur zorgt voor meer betrokkenheid bij het project en een hogere kans dat de resultaten daadwerkelijk toegepast worden. Tijdens de totstandkoming van een data science project houden enkele organisaties dan ook al rekening met de data-analysecapaciteiten mensen en organisatie. Dit wordt gedaan door te toetsen of het project past bij de datastrategie, de prioriteiten en door bij het bepalen van de probleemstelling al afstemming te zoeken met stakeholders. Gedurende het data science project wordt door de organisaties bewust gewerkt aan het creëren van begrip en draagvlak bij de betrokkenen door regelmatig te communiceren over de voortgang en de resultaten van het data science project. Gezien het belang van de datastrategie en een data gedreven cultuur zouden deze stappen structureel onderdeel moeten zijn van een data science projectmethode. De belangrijkste reden waarom de organisaties voor een projectmethode kiezen is de aansluiting op de organisatieprocessen en de beheersing en ervaring van de teamleden met deze methode. De Agile Kanban en Agile Scrum methoden faciliteren de planning, beheersing en communicatie van een data science project maar schieten tekort op het inhoudelijke data mining proces. CRISP-DM daarentegen voorziet wel in het data mining proces maar is onvolledig als het gaat om projectmanagement. Organisaties doen er daarom verstandig aan een eigen aanvullende projectmethode te ontwikkelen zolang er geen geïntegreerde data science projectmethode voor handen is.

Voor het vergroten van het data science projectsucces zouden organisaties vooral moeten werken aan verbetering van de data-analysecapaciteiten mensen en organisatie. Meer specifiek aan de kennis en vaardigheden van data science en aan een duidelijke datastrategie en data gedreven cultuur in de organisatie. Qua kennis gaat het dan niet alleen om de analytische en technische kennis en vaardigheden bij de mensen op de data science afdeling, maar ook om communicatieve vaardigheden. Daarnaast zal gewerkt moeten worden aan het begrip van data science en de toepassing daarvan bij alle mensen in de organisatie. Bij het verbeteren van de data-analysecapaciteiten zouden kleinere aspirant organisaties zich vooral moeten richten op het creëren van tijd, budget en het in huis halen van data science kennis. Grotere ervaren organisaties zouden zich meer moeten richten op het vrijmaken van capaciteit en prioriteit bij data scientists en stakeholders, en het organiseren van een goede databeschikbaarheid en data-architectuur. De organisaties geven aan dat voor het succesvol managen van een data science project de projectscope klein gehouden moet worden en beperkt moet worden tot de bestaande data-analyse capaciteiten van de organisatie.

Discussie

Uit het empirische onderzoek blijkt dat data science projecten pas als succesvol worden beschouwd wanneer resultaten daadwerkelijk hebben geleid tot blijvende waardecreatie door een structurele integratie binnen de operationele processen. In de literatuur worden deze twee aspecten eveneens benoemd (Lavallo, 2011; Sudhakar, 2012; Schüritz, 2017) maar wordt een structurele operationele integratie niet benoemd als voorwaarde voor het uiteindelijke succes van een project. Het definiëren van projectsucces blijft beperkt tot voornamelijk de kaders van het projectmanagement, zoals kosten, tijd en kwaliteit, en het product zelf (Sudhakar, 2012). Ook blijkt uit dit onderzoek dat organisaties moeite hebben om vanuit de context van waardecreatie het al dan niet succesvol zijn van een data science project daadwerkelijk cijfermatig in beeld te brengen.

De totstandkoming van een data science project verloopt via verschillende fasen (Saltz J. S., 2017; Schüritz, 2017). In dit onderzoek zijn de projectfasen opgesplitst in aanleiding, afstemming, begrijpen, verkennen en vormen, is de volgorde van de projectfasen inzichtelijk gemaakt en zijn de activiteiten daarbinnen benoemd. De analysecontext heeft invloed op de projectsamenstelling en doelen zoals beschreven door Saltz (2017) maar slechts in beperkte mate op de gekozen projectmethode, zo blijkt uit dit onderzoek. De keuze voor een projectmethode is eerder afhankelijk van bestaande methoden in de organisatie en de voorkeur vanuit het projectteam of de individuele specialisten. Bij de onderzochte caseorganisaties werd CRISP-DM slechts eenmaal als projectmethode toegepast, daarbij werd bevestiging gevonden dat CRISP-DM tekortschiet op van het gebied van projectmanagement-, integratie- en organisatieprocessen zoals door Mariscal (2010) omschreven. In tegenstelling tot de literatuur (Schüritz, 2017) is CRISP-DM in dit onderzoek niet de meest gebruikte methode. De Agile methoden Kanban en Scrum werden vaker gebruikt, mede vanwege de interactieve en iteratie voordelen die eerder in het literatuuronderzoek ook zijn benoemd (Li, 2016; Schmidt, 2018). Daarbij is opgemerkt dat het zomaar toepassen van de Agile methoden niet verstandig zou zijn vanwege de beperkte focus op het dataproduct (Saltz J. S., 2017), wat in dit onderzoek werd bevestigd. Om de tekortkomingen in de projectmanagementmethodes te compenseren hebben organisaties zowel bij Agile Scrum als CRISP-DM eigen aanvullende methoden nodig.

De data, data-analysecapaciteiten en de projectsuccesfactoren hebben invloed op het succesvol managen van een data science project, zoals in eerdere literatuur omschreven (Vidgen, 2017; Sudhakar, 2012; Dutta, 2015). Ook is bevestiging gevonden dat de data-analysecapaciteiten onderling invloed op elkaar hebben (Vidgen, 2017) en afhankelijk zijn van de contextuele omstandigheden, zoals de kenmerken van een organisatie (Mikalef P. B., 2019). De meeste invloed op het succesvol managen van data science projecten hebben de data-analysecapaciteiten organisatie en mensen. Vanuit de literatuur zijn datastrategie, een data gedreven cultuur en ethiek (Vidgen, 2017; Mikalef P. B., 2019; Lavallo, 2011), die allen deel uitmaken van de data-analysecapaciteit organisatie, benoemd als grootste obstakels voor succesvolle waardecreatie. Uit dit onderzoek blijkt dat het hebben van een datastrategie, data gedreven cultuur en mensen met kennis van data science de belangrijkste data-analysecapaciteiten zijn voor het succesvol managen van data science projecten. Hiermee wordt het belang van een datastrategie en data gedreven cultuur bevestigd. Ethiek wordt wel benoemd maar niet als zijnde heel invloedrijk. Ook wordt inzicht gegeven hoe de organisaties bij het managen van data science projecten rekening houden met de datastrategie en (het creëren van) een data gedreven cultuur. Daarnaast laat dit onderzoek zien dat organisaties op een verschillende manier moeten werken aan het ontwikkelen van de data-analysecapaciteiten, afhankelijk van de organisatiekenmerken en huidige volwassenheid van de data-analysecapaciteiten.

In dit onderzoek is een verband beschreven tussen data, de data-analysecapaciteiten en de projectsuccesfactoren, waarbij blijkt dat de factoren projectmanagement en team het meest worden beïnvloed door data en de data-analysecapaciteiten. De beschikking en volwassenheid van data en de

data-analysecapaciteiten lijkt vooral invloed te hebben op de analysedoelen (informatie bij aspirant, optimalisatie bij ervaren) en de scope van de data science projecten. Hierin zou een bevestiging gezien kunnen worden dat succesvolle data science projecten korter moeten zijn, met minder resources en met een gereduceerde projectscope (Nemati, 2003).

Reflectie

Door interviews af te nemen bij zes verschillende organisaties is mogelijk een breder beeld ontstaan hoe data science projecten in de praktijk tot stand komen en worden gemanaged, maar dat beeld is slechts eenzijdig belicht vanuit iedere organisatie. Dat verlaagt de betrouwbaarheid van de resultaten, al is in twee gevallen wel gebruik gemaakt van validatie door het inzien van projectdocumenten. Voor zover de resultaten vanuit een dergelijk kwalitatief onderzoek al generaliseerbaar zijn, is de mogelijkheid hiervan verder afgenomen en dus zeer beperkt. Ook heeft het mogelijk invloed op de kwaliteit van de analyses en leesbaarheid van de scriptie vanwege een verhoogde complexiteit.

De veranderende setting waarin de interviews zijn afgenomen (online via Microsoft Teams in plaats van in een real life setting) hebben mogelijk impact op de interpretatie van de antwoorden door de onderzoeker, al is het observeren van non-verbale signalen ook via video redelijk goed mogelijk.

De geanalyseerde kenmerken van de organisaties komen dusdanig met elkaar overeen dat het moeilijk is een onderscheid te maken tussen de invloed die de omvang van de organisatie heeft en de volwassenheid van de data-analysecapaciteiten op het succesvol managen van data science projecten. Interpretatie van verschillen tussen aspirant en ervaren organisaties zullen dan ook met grote terughoudendheid moeten worden gedaan.

Via iteratie is de uiteindelijke analysemethode ontstaan waarin thema's, categorieën en items zijn gedefinieerd waarmee de interviews zijn gecodeerd. De aanpak om deze codering te baseren op de combinatie van de in het literatuuronderzoek gevonden businessanalysecapaciteiten en de projectsuccesfactoren heeft de relatie tussen het literatuuronderzoek en de empirische studie versterkt. Tegelijk zijn het twee separate, bij elkaar gebrachte, coderingen waarin mogelijk overlap aanwezig is. Bij het ontbreken van een compleet beeld aan factoren die een rol spelen rondom het succesvol managen van data science projecten was dit de meest voor de hand liggende aanpak. Het toepassen van in de mogelijk in de literatuur gedefinieerde data science projectfasen had de codering kunnen versterken. Ook had dat kunnen zorgen voor een mogelijk vergelijk tussen resultaten uit de literatuurstudie en het empirische onderzoek op dit thema.

Het aantal keren dat items, categorieën en thema's zijn benoemd door de respondenten is niet statistisch representatief. De aard van dit kwalitatieve onderzoek is niet gericht op het verzamelen van statistisch bewijs. Het aantal waarnemingen is daarvoor onvoldoende en tussen de kenmerken van de organisaties zijn de verhoudingen ongelijk. Zo is de CRISP-DM methode slechts door één organisatie uitgevoerd en komen de andere methoden vaker voor. Ook met interpretatie van de analyseresultaten uit dit kwalitatieve onderzoek dient daarom terughoudend en kritisch omgegaan te worden.

In het onderzoek worden verschillende termen gebruikt welke mogelijk niet voor iedere lezer even bekend en begrijpelijk zijn zonder de definitie ervan of omdat ze zijn vertaald vanuit het Engels. Bijvoorbeeld data-analysecapaciteiten wat oorspronkelijk (data) analytics capabilities wordt genoemd. Het gebruik van originele Engelse termen en het definiëren van betekenissen had kunnen leiden tot een betere afbakening en begrip voor de lezer. Hetzelfde geldt voor de gevonden benamingen in de literatuur en de overname daarvan in het vervolg van het onderzoek. Zo zijn bijvoorbeeld de businessanalysecapaciteiten van Vidgen (2017) vertaald naar data-analysecapaciteiten zonder een uitgebreide discussie tussen de in de literatuur aanwezige verschillende theorieën en definities en zonder een uitgebreid gemotiveerde keuze van de gebruikte benaming in dit onderzoek. Ook zijn binnen de modellen kleine nuances aangebracht die wel worden gemotiveerd maar mogelijk voor

verwarring zorgen. Zo maakt data geen deel uit van de businessanalysecapaciteiten in het model Vidgen (2017) maar wordt dit in de interviews en de codering wel opgenomen in het thema data-analysecapaciteiten.

In het onderzoek en in de interviews zijn heel veel onderwerpen besproken. Vanuit de focus op blijvende waardecreatie en de rol van de organisatorische en culturele aspecten daarbij, is het onderzoek breed uitgewaaierd. Daardoor zijn elementen binnen het managen van data science projecten oppervlakkiger onderzocht. Een alternatief had geweest om meer te focussen op het daadwerkelijk managen van data science projecten binnen de kaders van die organisatorische en culturele aspecten. Wellicht zelfs al vanaf de literatuurstudie maar dat was zeker mogelijk bij het empirische onderzoek. Dan was mogelijk een minder breed maar rijker beeld gevormd van het managen van data science projecten.

Aanbevelingen voor de praktijk

Bij het vormen en managen van data science projecten doen organisaties er verstandig aan rekening te houden met het uiteindelijke doel van het project, namelijk waardecreatie, en om gelijk een systeem te ontwikkelen om het effect daarop te meten.

De geschetste projectfasen en methoden en de voor- en nadelen daarvan zouden een handig hulpmiddel kunnen zijn dat organisaties kunnen gebruiken bij de overwegingen hoe zij hun data science projecten vormgeven en managen.

Om de kans op het succesvol managen van hun data science projecten te vergroten zouden organisaties de scope van de projecten af moeten stemmen op de volwassenheid van de data-analysecapaciteiten en die stapsgewijs ontwikkelen vanuit een duidelijke datastrategie, voornamelijk de organisatorische en culturele aspecten.

Aanbevelingen voor verder onderzoek

Het breder in de praktijk onderzoeken van de criteria waarop het succes van een data science project wordt gemeten lijkt nodig omdat vanuit het dit onderzoek blijkt dat het daarbij om de uiteindelijke waardecreatie zou moeten gaan en dat vanuit de huidige literatuur de nadruk meer ligt op projectmanagement- en productsucces.

De projectmethoden CRISP-DM, Agile Kanban en Agile Scrum hebben allemaal beperkingen waardoor ze niet als projectmanagementmethode voor het gehele data science proces gebruikt kunnen worden. De ontwikkeling en toepassing van een eenduidige projectmanagementmethode kan mogelijk helpen het inzicht in het succesvol managen van data science projecten verder te vergroten.

De data-analysecapaciteiten en de volwassenheid daarin hebben invloed op het succesvol managen van data science projecten en de projectfactoren die daarvoor nodig zijn. In dit onderzoek zijn daarvoor verschillende verbanden gevonden. Verdere verdieping naar deze relatie doormiddel van een kwalitatieve studie zou daarvoor geschikt zijn. Ook een kwantitatieve studie naar de mate waarin deze invloed aanwezig is kan nieuwe inzichten opleveren, al lijkt het verstandig daarvoor eerst te komen tot een eenduidig model van data-analysecapaciteiten en projectfactoren die een rol spelen.

Een datastrategie, data gedreven cultuur en mensen met data science kennis en vaardigheden zijn de data-analysecapaciteiten die de meeste invloed hebben op het succesvol managen van data science projecten. Het zou daarom van belang zijn om te onderzoeken op welke wijze deze data-analysecapaciteiten het best ontwikkeld kunnen worden.

REFERENTIES

- Adolphus, M. (2013). How to ... undertake case study research. In *Methoden en Technieken voor onderzoek, Reader 1* (pp. 195-208). Heerlen: Open Universiteit.
- Dutta, D. B. (2015). Managing a Big Data project: The case of Ramco Cements Limited. *International Journal of Production Economics*, 165.
- Gibbert, M. &. (2010). The “What” and “How” of Case Study Rigor: Three Strategies Based on Published Work. *Organizational Research Methods*, 13(4), 710–737.
- Janssen, B. (2013). Het uitvoeren van wetenschappelijk literatuuronderzoek. In *Methoden en Technieken voor onderzoek, Reader 1* (pp. 105-112). Heerlen: Open Universiteit.
- Kamminga, P. (2013). Case Study research. In *Methoden en Technieken voor onderzoek, Reader 1* (pp. 168-171). Heerlen: Open Universiteit.
- Koronios, A. &. (2015). Towards A Process View on Critical Success Factors in Big Data Analytics Projects. *Proceedings of the Twenty-First Americas Conference on Information Systems*, (pp. 1–14).
- Lavalle, S. L. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), 21–31.
- Li, Y. T.-B. (2016). A Snail Shell Process Model for Knowledge Discovery via Data Analytics. *Decision Support Systems*, 91.
- Mariscal, G. M. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2), 137–166.
- Martin, H. (2013). Een verkenning van kwalitatief onderzoek. In *Methoden en Technieken voor onderzoek, Reader 1* (pp. 113-120). Heerlen: Open Universiteit.
- McAfee, A. &. (2012). Big Data: the management revolution. *Harvard Business Review*, 90(10), 61–68.
- Mikalef 2, P. B. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, Volume 98, 261-276.
- Mikalef, P. &. (2019). Big data and business analytics: A research agenda for realizing business value. *Information & Management*.
- Nemati, H. B. (2003). Key factors for achieving organizational data-mining success. *Industrial Management and Data Systems*, 282 – 292.
- Okoli, C. S. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, 10(26).
- Ose, S. (2016). Using Excel and Word to Structure Qualitative Data.
- Provost, F. E. (2017). What Managers Need to Know About Big Data. *Research-Technology Management*.
- Saltz, J. S. (2017). Comparing Data Science Project Management Methodologies via a Controlled Experiment. *Proceedings of the 50th Hawaii International Conference on System Sciences*, (pp. 1013–1022).

- Saltz, J. S. (2017). Predicting data science sociotechnical execution challenges by categorizing data science projects. *Journal of the Association for Information Science and Technology*.
- Saunders, M. L. (2016). *Research methods for business students*. Harlow: Pearson.
- Schmidt, C. &. (2018). Synthesizing agile and knowledge discovery: case study results. *Journal of Computer Information Systems*, 58(2), 142–150.
- Schüritz, R. B. (2017). How To Cultivate Analytics Capabilities Within an Organization ? – Design and Types of Analytics Competency Centers. *Proceedings of the 25th European Conference on Information Systems (ECIS)* (pp. 1–15). Guimarães, Portugal: AIS Library.
- Sudhakar, G. (2012). A model of critical success factors for software projects. *Journal of Enterprise Information Management*, 25.
- Vidgen, R. S. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, Volume61, Issue 2, 626-639.
- Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(1), 1247–1268.

Bijlage 1: interview protocol

Interview protocol

Inleiding (5 min.)

Hartelijk dank voor deze afspraak en uw medewerking aan dit onderzoek. Dat waardeer ik erg. Voordat we met het interview beginnen wil ik vragen of ik mag tutoyeren tijdens het interview.

→ Prima, tijdens het interview spreken we elkaar aan met u/jij.

Ik zou het interview willen opnemen zodat ik later het interview terug kan luisteren en beter kan analyseren. Is dat akkoord?

→ Ja, start de opname

→ Nee, maak notities en beoordeel achteraf of het interview bruikbaar is

We starten nu met het interview. Allereerst zal ik de achtergrond van het onderzoek toelichten en het doel van dit interview daarin benoemen. Daarna zullen we starten met enkele algemene vragen en steeds meer de diepte ingaan. Tot slot behandelen we weer wat meer algemene vragen. Het interview duurt ongeveer 60 minuten. Tijdens het interview zal ik niet inhoudelijk op de antwoorden ingaan. Na het interview is daar wel gelegenheid voor. Zijn er vooraf vragen over de opzet van het interview?

→ Ja, beantwoord deze

→ Nee, start met de introductie

Introductie (10 min.)

Veel bedrijven werken aan het creëren van toegevoegde waarde door de inzet van big data. Dat wordt gedaan door data science projecten uit te voeren. Uit onderzoek weten we dat ruim 50% van de data science projecten mislukt (niet afgemaakt wordt, of niets oplevert) en dat dat met name licht aan organisatorische en culturele aspecten zoals “hoe moet ik de inzichten gebruiken”, “geen prioriteit bij het management” en “te weinig data-analyse vaardigheden op het verschillende afdelingen”. In dit interview wil ik aan de hand van verschillende thema’s een diepgaand inzicht krijgen hoe data science projecten in de praktijk tot stand komen (1) en gemanaged worden om te komen tot succes (2) en welke elementen daarin een rol spelen (3). Zijn er over de achtergrond en het doel van dit interview vragen?

→ Ja, beantwoord deze

→ Nee, starten met de ‘screener questions’

Vooraf hebben we al verkend of jouw organisatie geschikt is voor dit interview. Om er zeker van te zijn dat dit het geval is en je een waardevolle bijdrage levert aan dit onderzoek wil ik vooraf een aantal korte vragen stellen:

1. Wat zijn de hoofdactiviteiten van jullie organisatie?
2. Wat is jouw rol binnen de gehele organisatie?
3. Wat versta je onder de term big data?

Big data: 4 V's. Omvang van de data (Volume), het aantal databronnen en de structuur van de data (Variety), de snelheid waarmee de data wordt gegenereerd en verzameld (Velocity) en de betrouwbaarheid van de data (Veracity)

4. Wat versta je onder de term data science?

Data science: is een interdisciplinair onderzoeksveld met betrekking tot wetenschappelijke methoden, processen en systemen om kennis en inzichten te onttrekken uit (zowel gestructureerde als ongestructureerde) data.

5. Wat is het laatste data science project dat is uitgevoerd?

6. Wat was het doel van dat data science project?

7. Wat was jouw rol bij dat data science project?

Datastrategie (10 min.)

We hebben de eerste vragen gehad en gaan nu dieper in op het creëren van toegevoegde waarde met behulp van data. Je mag daarbij het data science project van de vorige vragen als referentie nemen, maar dat is niet noodzakelijk. Heb jij tot zover vragen of zijn er dingen onduidelijk?

➔ Ja, beantwoord de vragen

➔ Nee, ga door naar onderstaande vragen

8. Wat kun jij vertellen over de algehele strategie van jullie organisatie?

9. Welke rol speelt data daarin?

10. Hoe zorgen jullie ervoor dat deze op elkaar zijn afgestemd?

11. *Hoe is deze afstemming bij de klanten waarvoor jullie werken? (indien van toepassing)*

12. Wat zijn bij deze onderlinge afstemming de belangrijkste 3 obstakels?

Data science projecten (20 min.)

Intro. Voor het verzamelen, analyseren en toepassen van data worden meestal data science projecten ingezet. Vanuit jullie organisatie voeren jullie deze projecten uit voor en met klanten. Eerst wil ik het hebben over de doelen en het succes van zulke projecten. Daarna gaan we steeds meer de diepte in.

13. Met welke doelen en ambities worden er voornamelijk data-analyses uitgevoerd?

Doorvragen tot duidelijk is welke van onderstaande doelen het meest relevant is

- 1: het automatiseren en efficiënter maken van processen en verlagen van kosten
- 2: het optimaliseren van de organisatie door actief op zoek te gaan naar data
- 3: de totale bedrijfsvoering door automatische data-analyse in alle processen

Het succes van een project (voor een klant) kan op verschillende manieren worden gemeten.

14. Hoe meten jullie het uiteindelijke succes van een data science project?

15. Wat zijn bij het meten van het succes de belangrijkste 3 obstakels?

Doorvragen om erachter te komen of kosten, tijd en kwaliteit worden meegenomen bij het meten

Het succesvol uitvoeren van een data science project is afhankelijk van het projectmanagement en het product succes. Binnen dit onderzoek laten we het product succes buiten beschouwing (bv. hoe

statistisch representatief de uitkomst is) en richten we ons op het succesvol managen van een data science project en elementen die daarbij een rol spelen.

16. Hoe komen data science projecten die jullie uitvoeren tot stand? Wat is de volgende stap?
17. Wat zijn bij het tot stand komen van data science projecten de belangrijkste 3 obstakels?

Doorvragen naar initiatief (business/DS), teamsamenstelling (single/multi/rollen) en formuleren doelstellingen

Als dan eenmaal het data science project is gevormd gaat het team aan de slag. Om de doelen van het data science project te realiseren is het nodig het data science project succesvol te managen. Daarover gaan de volgende vragen.

18. Met welke projectmanagement methode werkt het data science projectteam?
19. Wat was de motivatie om voor deze projectmanagement methode te kiezen?
20. Wat zijn de positieve aspecten van deze methode op het managen van DS-projecten?
21. Wat zijn de negatieve aspecten van deze methode op het managen van DS-projecten?
22. Hoe zorgen jullie ervoor dat er tijdens het project gewerkt blijft worden aan de doelstellingen?

De analyse context van een project kan sterk verschillen. Daarmee bedoel ik dat projecten kunnen gaan over het genereren van een hypothese, het data generatief research (niet direct een doel) doen en het testen van hypothesen.

23. Heeft de analyse context van het DS-project invloed op het projectmanagement? Zo ja, hoe?

Data-analyse capaciteiten (15 min.)

Dat was het derde vragenblok, we zijn over de helft. Om data om te kunnen zetten naar inzichten, door het succesvol uitvoeren van data science projecten, moet een organisatie kunnen beschikken over data-analyse capaciteiten. Dimensies die daarin meespelen zijn bijvoorbeeld; data, organisatie, processen, mensen en technologie. We willen daar eerst generiek naar kijken naar de aanwezigheid en later de relatie tot het succesvol managen van data science projecten bespreken.

24. Wat kun je in het algemeen vertellen over de aanwezigheid van data-analyse capaciteiten in jullie organisatie?
25. *Hoe is dat bij klanten? (indien van toepassing)*
26. Hoe zorgen jullie ervoor dat de data-analyse capaciteiten aansluiten bij de strategie?
27. Wat zijn bij het op niveau krijgen van de data-analyse capaciteiten de belangrijkste 3 obstakels?
28. Hoe is de onderlinge samenhang van deze aanwezige data-analyse capaciteiten tot elkaar?
29. Zijn er nog andere aanwezige data-analyse capaciteiten nodig zijn voor uitvoeren van de datastrategie?

Doorvragen tot de meeste dimensies (data, organisatie, processen, mensen, technologie) zijn benoemd

Eerder hebben we het gehad over de data-analyse capaciteiten van een organisatie en de verschillende dimensies daarvan, namelijk; data (aanwezigheid data en kwaliteit), organisatie (inrichting en cultuur), processen (procedures, methoden), mensen (type, vaardigheden) en technologie (tools). Deze hebben in meer of mindere mate invloed op het succesvol managen van een data science project.

30. Wat kun je vertellen over de invloed van de dimensie “data” op het managen van een DS-project?

31. Wat kun je vertellen over de invloed van de dimensie “organisatie” op het managen van een DS-project?
32. Wat kun je vertellen over de invloed van de dimensie “processen” op het managen van een DS-project?
33. Wat kun je vertellen over de invloed van de dimensie “mensen” op het managen van een DS-project?
34. Wat kun je vertellen over de invloed van de dimensie “technologie” op het managen van een DS-project?

Doorvragen tot de meeste dimensies (data, organisatie, processen, mensen, technologie) zijn benoemd

35. Zijn er andere aanwezige data-analyse capaciteiten binnen organisatie nodig zijn voor managen DS-project?
36. Wat zijn de 3 data-analyse capaciteiten met de meeste impact op het managen van een DS-project?

Afsluiting (10 min.)

We zijn bijna aan het eind gekomen van dit interview. Ter afronding wil ik je vragen om advies uit te brengen aan bedrijven die willen starten met data science projecten of hun data science projecten willen verbeteren.

37. Wat zijn de 3 adviezen die bedrijven juist WEL moeten doen om hun DS projecten succesvol te managen?
38. Wat zijn de 3 adviezen die bedrijven juist NIET moeten doen om hun DS projecten succesvol te managen?

Dan zijn we aan het einde van het interview gekomen. Voordat we formeel afronden wil ik je nog de gelegenheid geven aanvullingen te geven op het interview?

39. Zijn er nog relevante zaken niet benoemd die wel van invloed zijn op het managen van DS projecten?

Dan heb ik nog een laatste vraag:

40. Zijn er documenten, bijvoorbeeld verslagen of plannings, die ik in mag zien en die kunnen bevestigen wat wij zojuist hebben besproken?

Tot slot nog een formele mededeling dat de interviews geanonimiseerd zullen worden verwerkt.

Nogmaals bedankt voor de tijd en antwoorden.

Einde.

Bijlage 2: interviewmatrix

Interview vraagnummer	Organisatie kenmerken	Waarde Creatie	Vormgeving projectfasen	Management methoden	Data-analyse capaciteiten
<i>aantal -></i>	7	16	22	19	25
1	v				
2	v				
3	v				
4	v				
5	v				
6	v				
7	v				
8		v			v
9		v			v
10		v			v
11		v			v
12		v			v
13	v	v	v		v
14		v	v		v
15		v	v		v
16			v	v	
17			v	v	
18			v	v	
19			v	v	
20			v	v	
21			v	v	
22			v	v	
23			v	v	
24	v				v
25	v				v
26		v			v
27		v			v
28		v			v
29		v			v
30			v	v	v
31			v	v	v
32			v	v	v
33			v	v	v
34			v	v	v
35			v	v	v
36			v	v	v
37		v	v	v	v
38		v	v	v	v
39		v	v	v	v
40		v	v	v	v

Bijlage 3: omschrijving caseorganisaties

EduTrans: een MKB-bedrijf dat opleidingen en trainingen verzorgt voor de logistieke sector. Het bedrijf is regionaal gericht en zet vooral in op groei bij bestaande zakelijke klanten. Voor de langere termijn wordt gewerkt aan het opzetten van een schaalbare infrastructuur. Data helpt om te bepalen waar er kansen liggen en inzichtelijk te maken hoe er gepresteerd wordt. Het onderzoeken van marktpotentie en analyseren van de operationele performance zijn onlangs uitgevoerde data science projecten. De algemeen directeur is geïnterviewd en als stakeholder betrokken bij zowel het initiëren van data science projecten als bij het sturing geven aan de voortgang van deze projecten.

MobiSoft: een MKB-bedrijf dat softwareoplossingen ontwikkelt en exploiteert om problemen op te lossen in de mobiliteitssector. De algehele strategie van de organisatie is bij de geïnterviewde data engineer, die ook de rol van data science projectmanager heeft, beperkt bekend. Data vormt de basis van de producten en diensten. Onlangs is het rijgedrag van taxichauffeurs in kaart gebracht en geanalyseerd om het aantal parameters waarop rijgedrag gemeten wordt uit te kunnen breiden. De data op orde krijgen heeft nu veel aandacht zodat daarop dashboards kunnen worden ontwikkeld en aan klanten kunnen worden gepresenteerd.

DataForce: een MKB-bedrijf dat consultancy en projectmanagementdiensten levert op het gebied van data science aan corporate organisaties in onder meer de telecom, energie en verzekeringen branche. Met de ontwikkeling van standaard dataproducten wil dit bedrijf een unieke positie in de markt opbouwen. Data is de brandstof van de klanten en dit bedrijf. Continue wordt gekeken hoe data vertaald kan worden naar inzichten en actie. Onlangs is de marketingmachine van een van de grootste energieleveranciers van Nederland geoptimaliseerd door het personaliseren van de communicatie met klanten, gebaseerd op data science inzichten. De geïnterviewde algemeen directeur is naast zijn commerciële rol ook actief in het coördineren van projecten en draagt waar nodig zijn steentje inhoudelijk bij.

LogiPlan: een MKB-bedrijf dat ERP SaaS software ontwikkelt en onderhoudt voor rijsscholen. Het doel is om rijsscholen te helpen succesvoller te worden door hun werkprocessen te automatiseren, de klantenservice te faciliteren en te helpen hun winstgevendheid te vergroten door de inzet van big data en ontsluiting van inzichten op dashboards. Data speelt daarin een grote rol. Het is een basis om in de te verwachte consolidatie de juiste keuzes te maken aan zowel de commerciële als de operationele kant. Onlangs is met openbare data een marktscan gemaakt en is een datawarehouse gebouwd zodat klanten geholpen kunnen worden met real-time data en analyses om de juiste keuzes te maken. De geïnterviewde is algemeen directeur en vertolkt binnen de projecten het belang van de gebruiker en klant.

MultiTech: een multinational die technologie ontwikkelt voor de sectoren industrie, energie en gezondheidszorg. Een strategie op hoofdkantoorniveau is aanwezig, al ontbreekt het aan een concrete invulling daarvan op landenniveau volgens de geïnterviewde data science projectmanager. Dat uit zich doordat digitalisering wel onderdeel uitmaakt van de strategie maar wat dat inhoudt niet wordt ingekleurd. Voor het inzetten van data geldt iets vergelijkbaars. De mogelijkheden zijn bekend, maar het ontbreekt aan doen. Nadat een CDO is aangetreden verbeterde dit en werden projecten gestart. Een van die projecten was het optimaliseren van het planningsproces aan de hand van data-analyses.

HighBasics: een multinational die uiteenlopende grondstoffen en halffabricaten produceert voor verschillende industrieën. De algehele strategie is gericht op duurzame (winst)groei en het leveren van een bijdrage aan het algemene welzijn van de hele wereld. Digitalisering en automatisering zijn belangrijke pijlers in het realiseren van die strategie, net als de inzichten uit data waardoor efficiënter geproduceerd kan worden. Een van de uitgevoerde data science projecten is het optimaliseren van de output van een fabriek met gelijkblijvende grondstofhoeveelheden. De geïnterviewde heeft de rol van data architect en projectmanager.

Bijlage 4: codering

Thema's, categorien en items			
Data science projecten	138	Kenmerken	80
Communicatie	10	Industrie	6
Project medewerking	9	Commerciele opleider	1
Projectcommunicatie	1	Data science consultancy	1
Product	13	ERP SaaS aanbieder	1
Producteisen	6	Producent grondstoffen	1
Productkwaliteit	5	Software ontwikkelaar	1
Productmanagement	2	Technologie concern	1
Projectmanagement	77	Projectdoelen	12
Managementvaardigheden	1	Efficientie	5
Project monitoren	5	Informatie	4
Projectcontrole	4	Productontwikkeling	3
Projectdoelen	8	Projectmethode	6
Projecteisen	1	Agile Kanban	3
Projectkosten	1	Agile Scrum	2
Projectmethode	6	CRISP-DM	1
Projectplanning	13	Rol	6
Projectschema	15	Directie	3
Projectscope	5	Management	3
Review & Feedback	9	Strategie	18
Risicobeheersing	4	Algemeen	6
Voortgangsbijeenkomsten	5	Data	6
Team	38	Klanten	6
Empowerment	7	Validatie	32
Taakorientatie	4	Big Data	6
Teamcommitment	8	Data Science	6
Teamcompetenties	12	Data Science Projecten	20
Teamcoördinatie	1		
Teamrollen	2		
Teamsamenstelling	4		
Data-analyse capaciteiten	127	Projectfasen	44
Data	13	1. Aanleiding	11
Data beschikbaarheid	5	Business	3
Dataarchitectuur	4	Development	4
Datakwaliteit	3	Klant	4
Security & Privacy	1	2. Afstemming organisatie	5
Mensen	25	Bredere behoefte	2
Analysevaardigheden	3	Datastrategie	1
Datavaardigheden	19	Prioriteit	1
Technischevaardigheden	3	Projectdoelstelling	1
Organisatie	45	3. Probleem begrijpen	7
Compliance	1	Probleem context	1
Datacultuur	9	Probleem scope	1
Datastrategie	9	Probleemstelling	5
Ethiek	3	4. Middelen verkennen	9
Financieel	7	Data beschikbaarheid	2
HR beleid	1	Datakwaliteit	2
Projectscope	1	Financieel	2
Tijd	12	Teamcompetenties	1
Training & Educatie	2	Technisch	2
Processen	8	5. Project vorming	12
Governance	1	Projectdoelen	2
Prestatiemanagement	7	Projectmethode	1
Waardecreatie	18	Projectplan	2
Businesscase	3	Teamrollen	1
Klantwaarde	5	Teamsamenstelling	5
Toepassing	10	Technisch	1
Volwassenheid	10		
Aspirant	5		
Ervaren	5		
Technologie	3		
IT beperkingen	3		