

MASTER'S THESIS

De houdbaarheid van data analyse modellen

den Ouden, Gideon

Award date:
2024

[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: 16. Feb. 2025

Open Universiteit
www.ou.nl



BPMIT graduation project

Shelf life of data analytics models

De houdbaarheid van data analyse modellen

BPMIT graduation assignment preparation (IM0602)

Business Process Management and IT Graduation Assignment (IM9806)

Opleiding:	Open Universiteit, faculteit Management, Science & Technology Masteropleiding Business Process Management & IT
Programma:	Open University of the Netherlands, faculty of Management, Science & Technology Master Business Process Management & IT (MSc)
Cursus:	IM9806 Afstudeeropdracht Business Process Management and IT
Student:	Gideon den Ouden
Studentnummer:	
Datum:	23 februari 2024
Afstudeerbegeleider	Samaneh Bagheri – Remko Helms
Meelezer	Ben Roelens
Versie nummer:	1.1
Status:	AF Definitve

Open Universiteit
www.ou.nl



Abstract

Data gedreven werken is cruciaal in het huidige bedrijfsleven, echter het succesvol onderhouden van data analyse modellen blijft een uitdaging. Dit onderzoek richt zich op de houdbaarheid van deze modellen, met aandacht voor de aanhoudende effectiviteit en kwaliteit.

Centraal staat de beperkte kennis van aanhoudende effectiviteit van data analyse modellen. Hoofzakelijk richt dit onderzoek zich op bijdrage aan succesvol onderhoud en kwaliteit Er wordt een literatuuronderzoek omschreven op basis van bestaande wetenschappelijke kennis, waarna deze resultaten vergeleken worden met de resultaten uit verschillende interviews en een aantal case studies. Daarbij wordt onderzocht welke toepassing de kwaliteit van data analyse modellen beïnvloed en welk effect dit op de houdbaarheid heeft. Daarbij wordt gezocht welke factoren het succesvolle onderhouden van data analyse modellen bepalen. Centraal daarin staan de mens, het proces en de techniek. Daar om heen worden thema's van data gedefinieerd. .

Dit onderzoek benadrukt de cruciale factoren voor het succesvol onderhouden van data analyse modellen: doelmatigheid, evenwicht tussen continuïteit en flexibiliteit, voortschrijdend inzicht, eenvoud, en effectieve communicatie. Het toont aan dat de kwaliteit en effectiviteit van deze modellen afhankelijk zijn van hun vermogen om zich aan te passen aan veranderende omstandigheden, ondersteund door een heldere strategie en adaptief vermogen.

Key words

1. **Data Analysis Models (Data analyse modellen):** Modellen die worden gebruikt om verschillende vormen van gegevens te interpreteren.
2. **Flexibility (Flexibiliteit):** Het vermogen om zich aan te passen aan veranderende omstandigheden en eisen.
3. **Continuity (Continuïteit):** Het vermogen om zich aan te passen aan veranderende omstandigheden en eisen.
4. **Sustainability (Houdbaarheid):** De mate waarin een data analyse model in staat is om zijn functionaliteit en waarde in stand te houden over een langere periode, rekening houdend met veranderende omstandigheden, technologische ontwikkelingen en gebruikersbehoeften.
5. **Iterative Process (Iteratief proces):** Een herhalend proces van verfijning en verbetering van datamodellen.
6. **Technical Robustness (Technische Robuustheid):** De sterkte en betrouwbaarheid van de technische infrastructuur die datamodellen ondersteunt.
7. **Quality Assurance (Kwaliteitsborging):** Het waarborgen van de kwaliteit en integriteit van modellen en gegevens.
8. **Stakeholder Engagement (Stakeholderbetrokkenheid):** De betrokkenheid en verantwoordelijkheden van verschillende stakeholders in het proces van data analyse.
9. **Operational Synchronization (Operationele afstemming):** Het afstemmen van operationele gegevens met analyse modellen.
10. **Scalability (Schaalbaarheid):** Het vermogen om te groeien en om te gaan met veranderingen in data volume en complexiteit.
11. **Data Pipeline (Data Pipeline):** Reeks verwerkingsstappen voor data vanaf bron tot met bestemming.
12. **Continuous Improvement (Continue Verbetering):** Doorlopende inspanningen om de prestaties en efficiëntie van modellen te verbeteren.
13. **Data Quality (Datakwaliteit):** De nauwkeurigheid en integriteit van de gegevens die in modellen worden gebruikt.
14. **Monitoring and Optimization (Monitoring en Optimalisatie):** Prestaties volgen en aanpassingen maken voor betere resultaten.
15. **Data Governance (Data Governance):** Het beheren en toezicht houden op data-assets om toegankelijkheid, betrouwbaarheid en veiligheid te waarborgen.
16. **Change Management (Verandermanagement):** De benadering om individuen, teams en organisaties te begeleiden naar een gewenste toekomstige staat in data analyse.
17. **Decision-Making (Besluitvorming):** Het proces van het maken van keuzes door een beslissing te identificeren, informatie te verzamelen en alternatieve oplossingen te beoordelen met behulp van gegevens.
18. **User Experience (Gebruikerservaring):** De algehele ervaring van een persoon bij het gebruik van een bepaald product, vooral in termen van gebruiksgemak en plezier.
19. **Data Integration (Data-integratie):** Het combineren van gegevens uit verschillende bronnen om een uniform beeld te bieden.
20. **Performance Metrics (Prestatiemetriecken):** Maten die worden gebruikt om de prestaties van datamodellen en processen te beoordelen.
21. **Agile Development (Agile Ontwikkeling):** Een iteratieve benadering van projectmanagement en softwareontwikkeling die teams helpt sneller en met minder problemen waarde te leveren aan hun klanten.
22. **Compliance (Naleving):** Het naleven van wetten, regelgeving, richtlijnen en specificaties die relevant zijn voor bedrijfsprocessen.
23. **Predictive Analytics (Predictieve Analyse):** Het gebruik van gegevens, statistische algoritmen en machine learning-technieken om de kans op toekomstige uitkomsten op basis van historische gegevens te identificeren.

Samenvatting

Dit onderzoek richt zich op de houdbaarheid van data analyse modellen, van cruciaal belang voor blijvend concurrentievoordeel in een tijd waarin data gedreven werken essentieel is voor essentiële besluitvorming. De centrale opgave richt zich op de factoren die de kwaliteit en houdbaarheid van deze modellen beïnvloeden. Daarbij wordt nader onderzocht welke factoren deze aspecten beïnvloeden. Centraal daarbij staat de houdbaarheid van data analyse modellen.

Data analyse modellen zijn methoden en technieken die worden gebruikt om ruwe data om te zetten in bruikbare inzichten. Ze helpen bij het vereenvoudigen van complexe data voor interpretatie en zijn toepasbaar op zowel gestructureerde als ongestructureerde data. Deze modellen maken het mogelijk om inzichten te genereren door processen zoals visualisaties voor patroonherkenning en het Extractie, Transformatie en Laad (ETL) proces, dat diverse bronnen integreert voor gegevensanalyse en besluitvorming. Basis uitgangspunt dat hiervoor genomen is, is het CRISP-DM methodiek. Er worden over het algemeen drie verschillende doeleinden van data analyse modellen onderscheiden. Descriptieve modellen beantwoorden de vraag "Wat is er gebeurd?" en omvatten producten zoals dashboarding, reporting, scorecards en datawarehousing. Predictieve modellen beantwoorden de vraag "Wat gaat er gebeuren?" en omvatten producten zoals data-mining, tekst mining, web/media mining en machine learning. Prescriptieve modellen beantwoorden de vraag "Wat zal ik doen?" en omvatten producten en handelingen zoals simulaties, beslismodellen en optimalisaties.

Data analyse modellen spelen een cruciale rol in besluitvormingsprocessen en volgen een levenscyclus, vergelijkbaar met de productlevenscyclus. Deze datalevenscyclus omvat processen zoals de creatie, verplaatsing, transformatie, opslag, onderhoud, deling en uiteindelijke verwijdering van data. Het doel is ervoor te zorgen dat data worden gemaakt, verplaatst, getransformeerd, opgeslagen, onderhouden, gedeeld en verwijderd op een manier die voldoet aan de behoeften van de gebruikers en de organisatie. Data analyse modellen zijn dus essentieel voor het genereren van inzichten uit data en ondersteunen organisaties bij het nemen van weloverwogen beslissingen. Ze spelen een centrale rol in het proces van dataanalyse en dragen bij aan het succes van data-gedreven organisaties. Belangrijk om er voor te zorgen dat deze kwalitatief en goed onderhouden blijven.

CRISP-DM is de meest geaccepteerde methodiek voor het ontwikkelen van data analyse modellen, dit wordt vanuit de meest onderzochte wetenschappelijke artikelen beaamd. Hoewel slechts beperkte toelichting is binnen CRISP-DM methodiek over het onderhouden van data analyse modellen, wordt in recent onderzoek benadrukt dat een grondig monitoring- en onderhoudsplan essentieel is om problemen te voorkomen tijdens de operationele fase van een model. Dit onderhoud vereist niet alleen het volgen van een vooraf opgezet plan, maar ook flexibiliteit en voortdurende aandacht. Hoewel er slechts zeer beperkte literaire bronnen zijn die specifiek ingaan op het onderhoud van data analyse modellen, kunnen de principes van CRISP-DM organisaties helpen bij het identificeren en aanpakken van factoren die van invloed zijn op de houdbaarheid van deze modellen. De basisbeginselen van het CRISP-DM methodiek, blijken dan ook een belangrijk fundament voor het behoud van kwaliteit en houdbaarheid in data analyse modellen.

Het onderzoek benadrukt een wetenschappelijke lacune in bestaande frameworks voor ontwikkeling van data analyse modellen, en de noodzaak voor onderzoek naar kwaliteitsbehoud en succesvolle toepassing van data analyses in een snel veranderende technologische omgeving. Het theoretisch kader richt zich op duurzaamheid in data analyse modellen, met een systematische aanpak om relevante thema's te identificeren. Data analyse modellen transformeren ruwe data in inzichten via technieken zoals visualisaties en ETL-processen, met onderscheid in descriptieve, predictieve en prescriptieve doeleinden.

Als referentie wordt hierbij gebruik gemaakt van het DAMA DMBOK-framework. DAMA International is een wereldwijde non-profitorganisatie die zich richt op de bevordering en ondersteuning van de principes en praktijken van gegevensbeheer, voornamelijk gefocust op data-gedreven werken. Deze biedt een gestructureerde aanpak voor het beheren van data, en ook mogelijkheden om deze toe te passen op data analyse modellen. Het benadrukt het belang van menselijke, procesmatige en technologische aspecten bij het genereren van inzichten uit data. Door dit raamwerk te integreren met de thema's van het onderzoek, wordt de focus gelegd op continue afstemming met de operatie, duidelijke taken en verantwoordelijkheden, flexibiliteit en iteratieve processen, een robuuste data-infrastructuur en het waarborgen van datakwaliteit. Dit helpt organisaties bij het effectief beheren en onderhouden van data analyse modellen, waardoor ze betrouwbare inzichten kunnen blijven genereren die cruciaal zijn voor besluitvorming.

De onderzoek aanpak omvat casestudies en diepte-interviews, met zorgvuldige selectie van cases en geïnterviewden. Het onderzoek volgt een iteratieve aanpak, beginnend met verkennende interviews, gevolgd door analyse en codering van resultaten, en eindigend met validatie via casestudy-onderzoek. De gegevensanalyse is uitgevoerd door middel van hybride codering om patronen en verbanden te identificeren. De betrouwbaarheid en validiteit werden gewaarborgd door methodische benaderingen en ethische richtlijnen.

Belangrijke thema's zoals aanpassingsvermogen, retentiebeleid en de actualiteitswaarde van informatie kwamen naar voren tijdens de interviews. De resultaten bieden inzicht in aspecten van duurzaamheid van data analyse modellen, met een nadruk op aanpassingsvermogen, afstemming met belanghebbenden en kwaliteitsborging. Effectieve communicatie met belanghebbenden werd als essentieel beschouwd, met nadruk op interactie tussen ontwikkelaars en eindgebruikers. De veranderlijke omgeving en de noodzaak hierdoor niet onnodig complexe modellen te maken hadden een sterk verband. Flexibiliteit is daarbij ondergeschikt aan continuïteit, afhankelijk ook van de mate en doel waarvoor de data analyse modellen worden gebruikt en toegepast.

Dit onderzoek belicht essentiële factoren voor het behoud van kwaliteit en houdbaarheid in data-analysemodellen, cruciaal voor bedrijfsstrategieën en besluitvormingsprocessen. Het identificeert communicatie, continuïteit, en voortschrijdend inzicht als sleutelaspecten voor het succesvol onderhouden van deze modellen. Hierbij wordt benadrukt dat modellen als producten gezien moeten worden, die continue moeten evolueren om aan veranderende eisen te voldoen, waarbij eenvoud en doelmatigheid voorop staan. Het onderzoek onderstreept het belang van strategische afstemming, de integratie van technische en organisatorische facetten, en een proactieve houding tegenover marktontwikkelingen voor het waarborgen van hun effectiviteit. De bevindingen suggereren dat een holistische benadering, die zowel de menselijke, procesmatige als technische aspecten omvat, fundamenteel is voor het duurzaam ontwikkelen en onderhouden van data-analysemodellen. Dit vraagt om regelmatige evaluatie, aanpassingen op basis van feedback, en een focus op duidelijke, eenvoudige modellen die makkelijk te onderhouden zijn. Het framework dat uit dit onderzoek voortkomt, biedt een strategische blauwdruk voor het ontwikkelen van duurzame modellen, voorbereid op de toekomst, en benadrukt de noodzaak voor data professionals om een gecoördineerde strategie te hanteren die specialisatie combineert met een geïntegreerde visie.

Implicaties voor de praktijk benadrukken de noodzaak om theoretische fundamenten beter af te stemmen op de praktijk. Reflectie van literatuurbevindingen in praktijksituaties en onderzoek naar praktijkfundamenten zijn essentieel om een geïntegreerd begrip te ontwikkelen en het gat tussen theorie en praktijk te overbruggen.

Summary

This study focuses on the sustainability of data analytics models, which is crucial for maintaining a competitive advantage in an era where data-driven operation is essential for critical decision-making. The central challenge addresses the factors influencing the quality and sustainability of these models, with a particular focus on the factors affecting these aspects. At the core is the sustainability of data analytics models.

Data analytics models are methods and techniques used to transform raw data into actionable insights. They aid in simplifying complex data for interpretation and are applicable to both structured and unstructured data. These models enable insight generation through processes such as visualizations for pattern recognition and the Extraction, Transformation, and Loading (ETL) process, integrating various sources for data analysis and decision-making. The foundational approach taken here is the CRISP-DM methodology. Generally, three different purposes of data analytics models are distinguished. Descriptive models answer the question "What happened?" and include products like dashboarding, reporting, scorecards, and data warehousing. Predictive models answer the question "What will happen?" and include products like data mining, text mining, web/media mining, and machine learning. Prescriptive models answer the question "What should I do?" and include products and actions like simulations, decision models, and optimizations.

Data analytics models play a crucial role in decision-making processes and follow a lifecycle similar to the product lifecycle. This data lifecycle includes processes such as creation, movement, transformation, storage, maintenance, sharing, and ultimate deletion of data. The aim is to ensure that data are created, moved, transformed, stored, maintained, shared, and deleted in a manner that meets the needs of the users and the organization. Data analytics models are thus essential for generating insights from data and supporting organizations in making informed decisions. They play a central role in the data analysis process and contribute to the success of data-driven organizations. It is important to ensure that these are of high quality and well-maintained.

CRISP-DM is the most commonly accepted methodology for developing data analytics models, as affirmed by the most researched scientific articles. Although there is only limited explanation within the CRISP-DM methodology about maintaining data analytics models, recent research emphasizes that thorough monitoring and maintenance plans are essential to prevent issues during the operational phase of a model. This maintenance requires not only following a pre-established plan but also flexibility and ongoing attention. Although there are very limited literary sources that specifically address the maintenance of data analytics models, the principles of CRISP-DM can help organizations identify and address factors affecting the sustainability of these models. The fundamental principles of the CRISP-DM methodology thus prove to be an important foundation for maintaining quality and sustainability in data analytics models.

The research highlights a scientific gap in existing frameworks for the development of data analytics models, and the need for research into quality maintenance and successful application of data analyses in a rapidly changing technological environment. The theoretical framework focuses on sustainability in data analytics models, with a systematic approach to identifying relevant themes. Data analytics models transform raw data into insights through techniques such as visualizations and ETL processes, with distinctions in descriptive, predictive, and prescriptive purposes.

As a reference, the DAMA DMBOK framework is used. DAMA International is a global non-profit organization dedicated to promoting and supporting the principles and practices of data management, primarily focused on data-driven operation. It offers a structured approach to data

management and also possibilities for applying this to data analytics models. It emphasizes the importance of human, procedural, and technological aspects in generating insights from data. By integrating this framework with the themes of the research, the focus is placed on continuous alignment with operations, clear tasks and responsibilities, flexibility and iterative processes, a robust data infrastructure, and ensuring data quality. This aids organizations in effectively managing and maintaining data analytics models, thereby enabling them to continue generating reliable insights crucial for decision-making.

The research approach includes case studies and in-depth interviews, with careful selection of cases and interviewees. The research follows an iterative approach, starting with exploratory interviews, followed by analysis and coding of results, and concluding with validation through case study research. Data analysis was conducted using hybrid coding to identify patterns and connections. Reliability and validity were ensured through methodical approaches and ethical guidelines.

Key themes such as adaptability, retention policy, and the current value of information emerged during the interviews. The results provide insights into aspects of the sustainability of data analytics models, with an emphasis on adaptability, alignment with stakeholders, and quality assurance. Effective communication with stakeholders was deemed essential, with an emphasis on interaction between developers and end-users. The changing environment and the necessity thereof not to make unnecessarily complex models had a strong correlation. Flexibility is subordinate to continuity, also depending on the extent and purpose for which the data analytics models are used and applied.

This research highlights essential factors for maintaining quality and sustainability in data analysis models, crucial for business strategies and decision-making processes. It identifies communication, continuity, and progressive insight as key aspects for successfully maintaining these models. Emphasizing that models should be viewed as products that need to continuously evolve to meet changing requirements, where simplicity and effectiveness are prioritized. The study underscores the importance of strategic alignment, the integration of technical and organizational facets, and a proactive stance towards market developments to ensure their effectiveness. The findings suggest that a holistic approach, encompassing human, procedural, and technical aspects, is fundamental for sustainably developing and maintaining data analysis models. This requires regular evaluation, adjustments based on feedback, and a focus on clear, simple models that are easy to maintain. The framework emerging from this research provides a strategic blueprint for developing sustainable models, prepared for the future, and highlights the need for data professionals to adopt a coordinated strategy that combines specialization with an integrated vision.

Implications for practice emphasize the need to better align theoretical foundations with practice. Reflecting literature findings in practical situations and researching practical foundations are essential to develop an integrated understanding and bridge the gap between theory and practice.

Inhoud

Abstract.....	2
Key words.....	3
Samenvatting.....	4
Summary.....	6
Bijchriften	10
H1 introductie.....	11
1.1 Aanleiding	11
1.2 Probleemstelling.....	11
1.3 Onderzoeksvraag.....	12
1.4 Relevantie	12
1.5 Leeswijzer	12
H2 Theoretisch kader	13
2.1 Opzet literatuur onderzoek	13
2.2 Begrip data analyse modellen.....	13
2.3 Onderzoeksaanpak	13
2.4 Uitvoering	16
2.5 Resultaten.....	17
2.6 Doel van het vervolgonderzoek.....	21
H3 Methodologie	22
3.1 Conceptueel ontwerp: keuze van onderzoeksmethode(n)	22
3.2 Technisch ontwerp: uitwerking van de methode.....	22
3.3 Gegevensanalyse	27
3.4 Reflectie validiteit, betrouwbaarheid en ethische aspecten	27
4. Resultaten.....	28
4.1 Uitvoer van het onderzoek	28
4.2 Resultaten onderzoek	29
4.3 Resultaat onderzoek vs. theorie.....	33
4.4 Resumé en conclusies	33
5. Discussie, conclusies en aanbevelingen	34
5.1 Discussie - reflectie	35
5.2 Conclusies	35
5.3 Aanbeveling voor de Praktijk	37
5.4 Suggesties en richting voor toekomstig onderzoek.....	37
5.5 Framework houdbaarheid van data analyse modellen.....	38
6. Tot slot	40

B1 Bibliografie.....	41
Appendix.....	44
A1 CRISP-DM.....	44
A2 Literatuuronderzoek referentie	45
A3 Toepassing onderzoeksmethodiek Okoli	46
A4 Toelichting thema's uit literatuuronderzoek.....	49
A5 Interviews.....	52
A6 Case studies.....	76
A6.1 Case 1 Huisvesting analyse	77
Embedded units & verbanden.....	77
Theoretische en praktische bijdrage onderzoek.....	78
A6.2 – Case 2 Planningsanalyse	79
Embedded units & verbanden.....	79
Theoretische en praktische bijdrage onderzoek.....	80
A6.3 – Case 3 QHSE dashboard.....	81
Embedded units & verbanden.....	81
Theoretische en praktische bijdrage onderzoek.....	82
A6.4 – Case 4 Productie- en contract monitorings model	83
Embedded units & verbanden.....	83
Theoretische en praktische bijdrage onderzoek.....	84
A6.5 Belangrijkste conclusies cases	85

Bijschriften

Figuur 1 - Relevance tree gebruikt voor zoekstrategie	15
Figuur 2 - Dama-dmbok raamwerk voor functioneel data management.....	17
Figuur 3 - Coderingschema ter voorbereiding op analyse en trekken van verband tussen kernbegrippen, indicatoren en bijbehorende wetenschappelijke bron	20
Figuur 4 - Toepassing van multiple embedded cases.....	25
Figuur 5 - Tactics van Yin bij toepassing case studies	27
Figuur 6 - Resultaten codering interviews	30
Figuur 7 - Framework: Houdbaarheid van data analyse modellen.....	39
Figuur 8 - CRISP-DM model (Saltz, 2021)	44
Figuur 9 - gebruikte codes voor coderen interviews	75
Figuur 10 - Overzichtskaart met daarin gehuisveste medewerkers, project- en kantoorlocaties	78
Figuur 11 - Gebruikte visualisatie in de planningsanalyse.....	80
Figuur 12 - Gebruikte visualisatie in het QHSE dashboard	82
Figuur 13 - Overzichten die inzicht bieden in behaalde productie- en contract monitoring	84

H1 introductie

1.1 Aanleiding

Het realiseren van data-gedreven oplossingen binnen het bedrijfsleven is in toenemende mate van cruciaal belang voor het maken van strategische-, tactische- en operationele beslissingen (Verma & Voids, 2016). Met deze oplossingen willen bedrijven onderscheidend en concurrerend kunnen blijven (Verma & Voids, 2016). De term “data-gedreven” is eerder regel dan uitzondering volgens het mission statement van diverse bedrijven (Lavallo, et al., 2011). Momenteel worden alle stappen en handelingen van activiteiten van uiteenlopende bedrijfsprocessen in diverse systemen vastgelegd (Vidgen, 2017), het is haast onmogelijk om niet data-gedreven te werken binnen een organisatie. Organisaties verzamelen data voor analyse doeleinden, en analyseren dit in zogenoemde data analyse modellen (Lavallo, Lesser, Schockley, & Hopkins, 2011). Hiermee wordt data op praktische wijze toegepast om het mogelijk te maken actiegerichte besluiten te kunnen maken (Verma & Voids, 2016). Bedrijven leggen voornamelijk focus op het ontwikkelen van data analyse modellen in plaats van op het leveren van lang houdbare kwaliteit. Slechts zeer beperkt aantal data mining- en Knowledge Discovery in Databases (KDD) methodieken schenken aandacht aan monitoring en in stand houden van data analyse modellen (Mariscal, Marbán, & Fernández, 2010). De focus van organisaties lijkt daarmee te liggen op data-gedreven werken als doel op zich, in plaats van aandacht te besteden aan de houdbaarheid en robuustheid van een product (Yeoh, 2015).

Onder data analyse modellen wordt verstaan het toepassen en interpreteren van diverse datamining oplossingen (Rahul & Banyal, 2020). Data analyse modellen ondersteunen het maken van gegronde beslissingen op basis van feiten (Mariscal, Marbán, & Fernández, 2010). Datamining technologieën kennen een lange historie. Kurgan & Musilek (2006) lichten toe dat de KDD (Knowledge Discovery en Data) initieel begon als het bedenken van een proces voor het realiseren van data analyse modellen (Verma & Voids, 2016). Terwijl adoptie en in stand houding slechts zeer beperkt aandacht aan besteed wordt in de data analyse methodieken (Mariscal, Marbán, & Fernández, 2010).

1.2 Probleemstelling

Organisaties geven slechts beperkte invulling aan de houdbaarheid van data analyse oplossingen (Vidgen, 2017). Weinig is bekend over het consequent in stand houden van het succes van data analyse modellen binnen diverse productcycli (Obeidat, North, & Richardson, 2015) (Dedehayir & Steinert, 2016). Vanuit het CRISP-DM (verdere toelichting zie bijlage A1 CRISP-DM) worden data analyse modellen beschouwd als product (Saltz, 2021), bijvoorbeeld een rapportage, dashboard, data-mining, simulatie of andere toepassing ter ondersteuning van besluitvorming (Lavallo, Lesser, Schockley, & Hopkins, 2011). Factoren die de houdbaarheid beïnvloeden zijn onder andere de data lifecycle (Dama-international, 2017) flexibiliteit, robuustheid en schaalbaarheid (Yeoh, 2015).

Pete Chapman, et al. (2000) geeft beperkte toelichting over de beheerfase binnen het CRISP-DM. Er worden slechts enkele woorden besteed over het onderhouden van reeds geproduceerde data analyse modellen (Vidgen, 2017). Saltz (2021) omschrijft in zijn onderzoek: “Develop a thorough monitoring and maintenance plan to avoid issues during the operational phase (or post-project phase) of a model.” Het onderhouden data analyse modellen beperkt zich niet alleen tot het uitvoeren van een voor opgezet plan, maar vergt flexibiliteit en continue aandacht (Lavallo, Lesser, Schockley, & Hopkins, 2011). Er is géén enkele literaire bron die expliciet aandacht besteed aan het onderhouden van data analyse modellen (Mariscal, Marbán, & Fernández, 2010). Indien bekend welke factoren dit beïnvloedt, kan hier gericht aandacht aan worden gegeven (Vidgen, 2017).

1.3 Onderzoeksvraag

Organisaties gebruiken in toenemende mate data analyse modellen (Lavalle, Lesser, Schockley, & Hopkins, 2011). Echter lijkt de toegevoegde waarde binnen wetenschap en praktijk meer toegespitst op de ontwikkeling van data analyse modellen, dan op het leveren van kwaliteit (Saltz, 2021). Over het in stand houden van reeds opgeleverde producten in een continu veranderende omgeving is slechts beperkt antwoord op gegeven. Onder kwaliteit wordt in dit geval verstaan, de mate waarop een product effectief, efficiënt, betrouwbaar en duurzaam wordt ingezet voor het doel waarmee het is gemaakt, en hoe lang. Daarmee luidt de onderzoeksvraag: “Welke factoren bepalen het succesvol in stand houden van data analyse modellen, en hoe lang is de kwaliteit hiervan in stand te houden?”

Er ontbreekt een lifecycle van data analyse modellen, van initiatie tot support (Solarte, 2002), en het in stand houden (Dedehayir & Steinert, 2016). Data analytics voegt bij succesvolle toepassing enorme waarde toe (Lavalle, Lesser, Schockley, & Hopkins, 2011) , echter is het belangrijk de oplossing te doorlopen en kritisch te blijven toetsen (Obeidat, North, & Richardson, 2015). Géén datamining framework omschrijft een best practice over het beheren en in standhouden (Kurgan & Musilek , 2006), of een controle fase (Saltz, 2021).

Hieruit komt duidelijk naar voren dat enkele kernbegrippen als kwaliteit en houdbaarheid zeer beperkt aan bod komen (Mariscal, Marbán, & Fernández, 2010), terwijl deze van essentieel belang zijn voor een vitaal en levensvatbaar product (Brodsky & Luo, 2015).

Op basis van de theorie zijn de deelvragen als volgt geformuleerd:

1. *Welke business enablers beïnvloeden de houdbaarheid van data analyse modellen?*
2. *Welke strategieën helpen om data analyse modellen succesvol houdbaar te maken?*
3. *In welke mate is het van belang om de houdbaarheid van data analyse modellen in stand te houden, en verschilt dit per specifieke toepassing?*

1.4 Relevantie

Wetenschappelijk onderzoek over de ontwikkeling van data analyse modellen, omschrijft weinig over de houdbaarheid en succesvol in stand houden van data analyse modellen (Mariscal, Marbán, & Fernández, 2010). Er is een wetenschappelijke lacune over meest gangbare en geaccepteerde frameworks die uitweiden over de beheerfase (Saltz, 2021). De wetenschap is gebaat bij een antwoord op de vraag welke factoren bijdragen aan voornamelijk de kwaliteit en houdbaarheid van een data analyse model (Verma & Volda, 2016). In tegenstelling tot ruim beschikbare bronnen die uitweiden over de technieken en processen voor het komen tot data analyse modellen (Mariscal, Marbán, & Fernández, 2010). Met oog op nieuwe ontwikkelingen en technologieën is het belangrijk het product in stand te houden (Dedehayir & Steinert, 2016), en het succesvol toepassen van data analytics binnen bedrijven (Lavalle, Lesser, Schockley, & Hopkins, Big Data, Analytics and the Path From Insights to Value, 2011). Voor data-gedreven beslissingen maken zullen organisaties wendbaar en flexibel moeten zijn (Kutzias, Dukino, & Kett, 2021). Er zijn echter beperkte middelen beschikbaar die het mogelijk maken om goede data analyse oplossingen in stand te houden (Saltz, 2021).

1.5 Leeswijzer

In dit onderzoek zal in hoofdstuk twee de theorie van het onderzoek en verdiepende kennis worden omschreven op basis van het literatuuronderzoek. Hierna wordt uitgebreid stilgestaan bij de in hoofdstuk drie omschreven methodologie van het onderzoek, en hoe de empirische resultaten tot stand komen vanuit de bedachte methodologie en technische ontwerp. Hierna zullen de resultaten uiteen worden gezet op basis van de gekozen onderzoeksmethodieken in hoofdstuk 4. Op het eind volgt de discussie, conclusie en aanbevelingen in hoofdstuk 5.

H2 Theoretisch kader

2.1 Opzet literatuur onderzoek

Het literatuuronderzoek is essentieel als fundament voor dit onderzoek, met nadruk op bronnen die waarde toevoegen aan de theoretische basis (Okoli & Schabram, 2010). Gezien de snel evoluerende technieken binnen de sector, is recente literatuur van belang, vooral gericht op de houdbaarheid van data analyse modellen (Luo, et al., 2020). Dit onderzoek onderzoekt relevante thema's en studies rondom de levenscyclus van deze modellen, gezien als producten, om de beïnvloedende factoren te identificeren (Brodsky & Luo, 2015) (Lavalley, et al., 2011).

2.2 Begrip data analyse modellen

Data analyse modellen zijn methoden en technieken voor het omzetten van ruwe data naar bruikbare inzichten, vereenvoudigend complexe data voor interpretatie (Dama-international, 2017) (Sharda, 2017). Deze modellen, toepasbaar op zowel gestructureerde als ongestructureerde data, faciliteren het genereren van inzichten via processen zoals visualisaties en patroonherkenning. Het Extractie, Transformatie en Laad (ETL) proces integreert diverse bronnen voor gegevensanalyse en ondersteunend bij besluitvorming (Laudon & Laudon, 2019). Data analyse modellen maken het mogelijk om inzichten te genereren door bijvoorbeeld visualisaties te maken waarmee patronen kunnen worden herkend (Dama-international, 2017).

Er wordt over het algemeen verschil gemaakt tussen drie verschillende doeleinden van data analyse modellen die de inzichten kunnen generen (Sharda, 2017):

- **Descriptieve modellen:** Wat is er gebeurd?
Producten zoals dashboarding, reporting, scorecards en datawarehousing
- **Predicatieve modellen:** Wat gaat er gebeuren?
Producten zoals data-mining, tekst mining, web/media mining en machine learning
- **Prescriptieve modellen:** Wat zal ik doen?
Producten en handelingen zoals simulaties, beslismodellen, optimalisaties

Data analyse modellen, beschouwd als producten, spelen een cruciale rol in besluitvormingsprocessen (Lavalley, et al., 2011). De data lifecycle, geïnspireerd door de product lifecycle, omvat de creatie, verplaatsing, transformatie, opslag, onderhoud, deling, en uiteindelijke verwijdering van data, weerspiegeland de gelijkenissen tussen data- en productlifecycle (Dama-international, 2017). "Het omvat processen die data maakt of genereert, het verplaatsen hiervan, transformeren en opslaan, en ervoor te zorgen dat data onderhouden en gedeeld worden met degenen die het gebruiken, toepassen, evenals zich ervan ontdoen" (Dama-international, 2017).

2.3 Onderzoeksaanpak

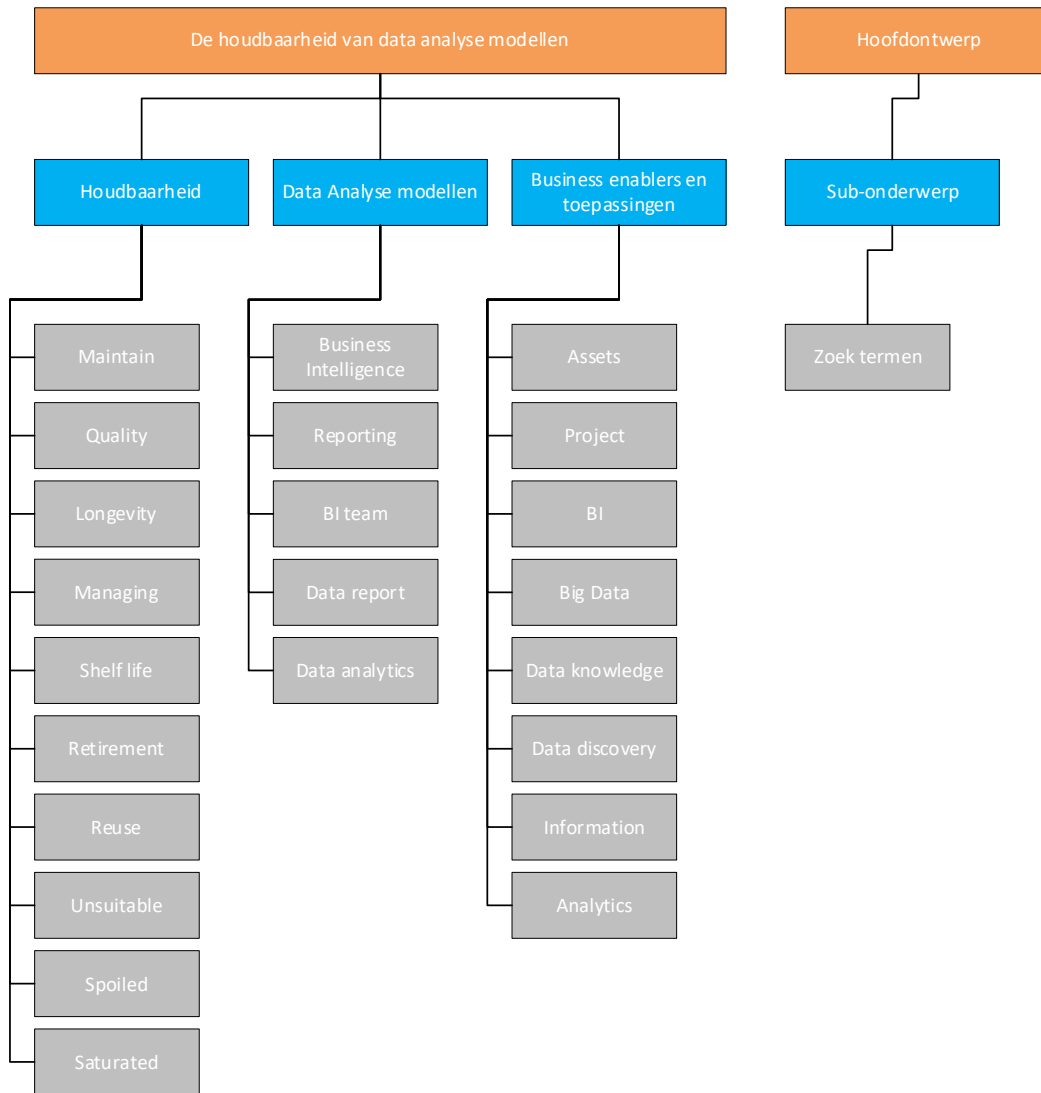
Het literatuuronderzoek streeft naar een systematische en kritische evaluatie van bestaande wetenschappelijke bronnen, met het doel een fundament te leggen voor vervolgonderzoek en het beantwoorden van de onderzoeksvraag (Okoli & Schabram, 2010). Het belang van het selecteren van relevante bronnen om een consistente lijn te identificeren wordt benadrukt (Knopf, 2006), evenals het bieden van een basis voor vervolgonderzoek via een deductieve onderzoeksbenadering (Saunders, et al., 2017, pp. 77-79). De focus ligt op het verzamelen van informatie uit relevante wetenschappelijke artikelen, met nadere uitleg en onderbouwing in de appendix (zie ook A2 Literatuuronderzoek referentie). Wetenschappelijke artikelen worden gezocht via databanken, met

een methodische benadering van theorie evaluatie en samenvatting (Webster, 2002). Aanvullend worden drie boeken gebruikt voor theorievalidering en kennisverwerving:

1. "Management Information Systems" door Laudon & Laudon (2019)
Theoretisch basisboek bedoeld voor informatiesystemen binnen het hogere onderwijs
2. "Dama-dmbok" (2e editie) door Dama International (2015)
Framework over het in stand houden data governance en -management binnen organisaties
3. "Business Intelligence, Analytics and Data Science" door Ramesh Sharda, Dursun Delen, en Efraim Turban (2017)
Management benadering voor toepassing business intelligence binnen organisaties

De zoekstrategie, gedefinieerd via een Relevance Tree, helpt bij het systematisch vinden van bruikbare bronnen door zowel synoniemen als antoniemen te overwegen, waardoor het zoekgebied wordt verbreed. Deze aanpak ondersteunt het structureel zoeken naar relevante literatuur, gericht op het identificeren van een consistente rode lijn in het onderzoeksthema (Saunders, Lewis, & Thornhill, 2017, pp. 93-94) (Okoli & Schabram, 2010) (Knopf, 2006). De Relevance Tree visualiseert hoe het hoofdonderwerp en de subonderwerpen zich verhouden tot de hoofd- en subvragen van het onderzoek. In Figuur 1 is de relevance tree uiteengezet.

Voor het zoeken naar geschikte literatuur binnen dit onderzoek, is er gezocht met verschillende zoektermen binnen het onderwerp, en sub-onderwerpen. De zoek-query ziet er als volgt uit: `[((maintain OR quality OR longevity OR managing OR "shelf life" OR retirement OR reuse OR unsuitable OR spoiled OR saturated) AND ("business intelligence" OR reporting OR "bi team" OR "data report" OR "data analytics")) AND (assets OR project OR bi OR "big data" OR "data knowledge" OR "data discovery" OR information OR analytics))]`. De zoektermen zijn opgesteld in verschillende concepten op basis van In Figuur 1 weergegeven relevance tree. Daarbij is iedere zij-vertakking (middelste niveau) uiteengezet en als concept opgenomen in de zoek query (laagste niveau). Binnen de concepten zijn individuele termen gezocht (OR) en combinatie tussen de drie concepten (AND).



FIGUUR 1 - RELEVANCE TREE GEBRUIKT VOOR ZOEKSTRATEGIE

2.4 Uitvoering

Voor de uitvoering van het literatuuronderzoek is de methodiek van Okoli en Schabram toegepast. Okoli en Schabram (2010) omschrijven acht stappen in de uitvoering van een literatuur onderzoek, die voor dit onderzoek als volgt zijn gebruikt. Zie hieronder belangrijkste benoemde stappen. Zie A3 Toepassing onderzoeksmethodiek Okoli voor uitgebreide omschrijving van stappen.

1. Doel van een literatuuronderzoek

Het doel van dit literatuuronderzoek is om de wetenschappelijke lacune die ontstaan is in de context "de houdbaarheid van data analyse modellen" verder te kunnen verkennen.

2. Protocol & training literatuuronderzoek omschrijven

Er is voornamelijk gezocht naar literatuur vanuit de relevance tree (Figuur 1):

- Houdbaarheid – *hoe kan een product zo lang mogelijk in stand worden gehouden?*
- Data – analyse modellen – *welke producten vloeien voort uit data analyse modellen?*
- Business enablers en toepassingen - *wat zijn de eigenschappen van data analyse modellen en hoe worden deze succesvol toegepast?*

3. Selectiefase van de literatuur omschrijven

Er is gezocht naar relevante wetenschappelijke artikelen via de Open Universiteit Bibliotheek, Google Scholar, en databases zoals EBSCO, Science Direct, en Springer. Boeken van Laudon & Laudon (2019), Dama International (2015), en Sharda (2017) vormden het fundament.

4. Praktische toets omschrijven voor keuze literatuur

Het beschreven zoek- en selectieproces voor relevante artikelen volgt een systematische aanpak, waarbij initieel circa 50 peer-reviewed artikelen van de afgelopen 20 jaar zijn geselecteerd op basis van titel, context en beschrijving uit vooraanstaande databases. Na verdere evaluatie op relevantie voor het specifieke onderwerp "de houdbaarheid van data analyse modellen", zijn er uiteindelijk 12 artikelen overgebleven. Deze selectie is gemaakt door de artikelen te beoordelen op inhoud, de voor- en nadelen ervan te wegen, en ze te vergelijken op basis van thema's of hoofdpunten die overeenkomen met het framework van Dama-dmbok (2017), resulterend in 5 volledig passende en 7 gedeeltelijk of beperkt passende artikelen. Zie "Praktische toets omschrijven voor keuze literatuur" voor toelichting.

5. Kwalitatieve toets maken

Een kwalitatieve toets, gebaseerd op thematische analyse en het Dama International (2015) framework, werd ontwikkeld voor latere codering. Dit coderingsschema, samengevat in Figuur 3, dient als basis voor het analyseren van de gevonden resultaten.

6. Data extractie omschrijven

Om tendens uit de 12 geselecteerde artikelen te halen, zijn de eigenschappen onderzocht:

- Wat zijn de belangrijkste focuspunten?
- Welke onder- en overbelichte thema's worden aangehaald?
- Welke eigenschappen van de belangrijkste thema's worden behandeld?

7. Analyse van de bevindingen geven

In dit onderzoek zijn functionele en technische overeenkomsten tussen gevonden artikelen, die zullen helpen met het schetsen van een concept van het onderzoeksonderwerp.

8. Schrijven van het review

Het fysiek schrijven van de literatuur review is uitgebreid toegelicht bij en te vinden in de appendix (zie

2.5 Resultaten

In dit hoofdstuk worden de resultaten van de gebruikte literaire artikelen beschouwd. Buiten de scope van het literatuuronderzoek was kwantitatief onderzoek of statistische resultaten. Er is gekozen voor een kwalitatieve en deductieve benadering, zoals voorgeschreven door Okoli & Schabram (2010). Na de literatuur te hebben gelezen en samengevat, werd thematische Analyse toegepast om patronen te identificeren binnen de kwalitatieve literatuur, zoals beschreven door Saunders (2019). Dit proces volgt de benadering van Knopf (2006) om een samenhangende lijn te vinden in kwalitatief literatuuronderzoek, waardoor artikelen effectief gesynthetiseerd kunnen worden.

2.5.1 Raamwerk en rode draad artikelen

Om structuur te geven aan de bevindingen uit het literatuuronderzoek, is er in de basis gekeken naar een erkend raamwerk die houvast geeft aan de gevonden resultaten. Dama-International (2017) omschrijft een methodiek waarbij mens, proces en techniek centraal staan. Zie Figuur 2 hieronder.

Dit raamwerk maakt een duidelijk onderscheid tussen rollen en verantwoordelijkheden, activiteiten, tools, organisatie en cultuur, technieken en overdracht (deliverables). Lavallo (2011) onderschrijft deze factoren ook als meest optimale vorm of level waarbij een organisatie capabel is om inzichten te genereren op basis van data analyse modellen. Tijdens de selectie van specifieke theoretische bronnen, is er gekeken naar de laatste overgebleven wetenschappelijke bronnen, dit waren circa 12 bronnen. Vanuit de thematische analyse zijn zes thema's onderscheiden:

1. Continue afstemming operatie
2. Taken en verantwoordelijkheden
3. Flexibiliteit
4. Iteratief proces
5. Technisch robuuste data pipeline
6. Verantwoorden kwaliteit



FIGUUR 2 - DAMA-DMBOK RAAMWERK VOOR FUNCTIONEEL DATA MANAGEMENT

Wanneer dit gecombineerd wordt met bovenstaande thema's ontstaat de volgende categorisering:

- Mens = Taken en verantwoordelijkheden + continue afstemming operatie
- Proces = Flexibiliteit + iteratief proces
- Techniek = Technisch robuuste data pipeline + verantwoorden kwaliteit

2.5.3 Alternatieve raamwerken

Er zijn ook alternatieve raamwerken overwogen. Zo is er gekeken naar het duivels kwadrant van Dumas, et al. (2018), Time/Place framework van DeSanctis & Gallupe (1987) en Framework for Information Quality van Danette McGilveray (2020). Deze waren echter beperkt gericht op het omschrijven van een product en leggen elders nadruk op. Benoemde frameworks omschrijven respectievelijk de kwaliteit van het proces, onderlinge samenwerking en omgevingsfactoren. De frameworks kennen relatief veel diepgang waardoor beperkte mogelijkheden zijn om deze breed toe te passen voor data analyse modellen. Het Dama framework is flexibel toe te passen, en is een internationaal erkende en gestandaardiseerde taal (Dama-international, 2017).

2.5.2 Thema toelichting

De gesynthetiseerde resultaten uit het literatuuronderzoek richten zich op thema's en het framework, gebaseerd op de opbouw van de deelvragen uit de onderzoeksvraag (zie 1.3 Onderzoeksvraag), gedetailleerd beschreven in de appendix (A4 Toelichting thema's uit literatuuronderzoek.). De kernthema's omvatten:

1. Continue afstemming operatie (mens)

Het is voor de houdbaarheid van data analyse modellen belangrijk continu de wensen en behoeften met de operatie af te stemmen (Lavalley, Lesser, Schockley, & Hopkins, 2011). De afstemming moet structureel gedaan worden (Yeoh, 2015), waarbij sommige elementen van het model moeten worden uitgefaseerd (Zhaojing, et al., 2020). Continu afstemmen van data analyse modellen kan bijvoorbeeld door frequent af te spreken of enquêtes te sturen om uit te vragen naar potentiële verbeteringen (Kutzias, Dukino, & Kett, 2021).

2. Taken en verantwoordelijkheden (mens)

Organisatie moeten onderscheid maken tussen het onderhouden en ontwikkelen van van data analyse modellen (Zhao, 2022) (Ransbotham, 2015). Data science medewerkers hebben vaak een breed takenpakket (Saltz & Hotz, 2020). Er moet juist focus worden aangebracht (Vidgen, 2017). Regelmatig afstemmen tussen de ontwikkelaars en gebruikers van de data analyse modellen kan daarbij helpen (Staples, 2016). Een community tussen de stakeholders kan enorm een boost geven aan de efficiëntie en effectiviteit van het product (Yeoh, 2015).

3. Flexibiliteit (proces)

Belangrijk uitgangspunt bij het CRISP-DM model is een flexibele aanpak (Saltz, 2021). Een flexibele werkmethode, zoals Agile-SCRUM kan dit helpen waar te maken (Verma & Voida, 2016), ook in de CRISP-DM werkmethode omschreven (Saltz & Hotz, 2020). Voor de houdbaarheid van data analyse modellen is dit van belang om efficiënt en effectief te veranderen (Zhao, 2022). Een gedegen database of datawarehouse helpt om reproductie en schaalbare data analyse modellen te realiseren (Brodsky & Luo, 2015).

4. Iteratief proces (proces)

Het belang van een iteratieve werkwijze om continue aanpassing te faciliteren, wordt onderstreept door Chapman, Clinton, & Randy (2000) en Yeoh (2015). In het beheer moet daarom ook mogelijkheid blijven om continu te verbeteren (Saltz, 2021). Er moet ingespeeld worden op de technische ontwikkeling (Kutzias, Dukino, & Kett, 2021).

5. Technisch robuuste data pipeline (techniek)

Een technische robuuste data pipeline is nodig om een continue stroom van data te faciliteren voor data analyse modellen (Zhao, 2022) (Verma & Voida, 2016). Een robuuste pipeline is essentieel voor succesvolle data analyseprojecten (Saltz, 2021). Hergebruik van de data (analyse modellen) draagt bij aan de houdbaarheid (Brodsky & Luo, 2015).

6. Verantwoorde kwaliteit (techniek)

Het definiëren van kwaliteit binnen data analyse is complex (Kutzias, Dukino, & Kett, 2021). Betrouwbare informatie is essentieel voor het nemen van weloverwogen beslissingen op

basis van feiten (Laudon & Laudon, 2018). Het monitoren van de datakwaliteit vanuit de bron is een belangrijke manier om de kwaliteit van data analyse modellen te waarborgen.

Dama International (2015) onderscheidt zeven essentiële factoren voor data kwaliteit: *Nauwkeurigheid, Integriteit, Consistentie, Volledigheid, Validiteit, Tijdigheid en Toegankelijkheid.*

Naast data kwaliteit, is het voor data analyse modellen als product belangrijk dat deze als product gevalideerd en verifieerd zijn door de organisatie (Staples, 2016). Daarmee moet de data analyse model voldoende kunnen voorzien in de informatie behoefte (Yeoh, 2015), en continu worden gemonitord op gebruik en potentiële verbeteringen (Zhao, 2022). Monitoring op de kwaliteit van data analyse modellen, draagt bij aan de houdbaarheid en succes hiervan (Ransbotham, 2015).

2.5.3 Coderingsschema

Om duidelijkheid te kunnen krijgen over de theorie en bijbehorende dimensies van het onderzoek, is onderstaand coderingsschema uiteengezet (zie Figuur 3). Dit schema maakt het mogelijk om een verbinding te maken tussen het Dama international (2017) framework, de thematische analyse en de indicatoren vanuit de samenvatting van onderzochte artikelen. Dit maakt je in staat ook diepere betekenis en begrip te krijgen van de theorie die je wilt ontwikkelen omtrent dit onderzoek (Saunders, et al., 2017, pp. 400-407). De codering zal hybride worden toegepast: Bij verwerking van de interviews wordt gekeken in hoeverre de thema's en dimensies een relevantie hebben en welke relatie daarmee kan worden gemaakt. Door de resultaten van de interviews middels onderstaand coderingsschema te coderen, zal het ook mogelijk moeten maken om de subvragen uit alinea 1.3 Onderzoeksvraag te beantwoorden. Hiermee wordt het mogelijk de empirische resultaten uit het interview en de theorie vanuit de bestaande wetenschap beter met elkaar te vergelijken.

Zie ook aanvullend A5.3 Overzicht gebruikte coderingen.

Kernbegrip (raamwerk)	Dimensie (thema)	Indicator - meetwaarde	Bron – verdiepende theorie
Mens	Taken en verantwoordelijkheden	Duidelijk omschreven rol m.b.t. data analyse modellen - beheer en ontwikkeling	(Saltz & Grady, 2017)
	Taken en verantwoordelijkheden	Taken en verantwoordelijk volgen uit herleidbaar procesdata	(Davenport, 2006)
	Taken en verantwoordelijkheden	Alle gebruikstoepassing worden op juiste manier begrepen en toegepast	(Lavalle, Lesser, Schockley, & Hopkins, Big Data, Analytics and the Path From Insights to Value, 2011)
	Taken en verantwoordelijkheden	Eenvoudige taken en scheiding van verantwoordelijkheden	(Francesco, Demi, Magrini, Marzi, & Papa, 2021)
	Continue afstemming operatie	Sturing op data kwaliteit en rapporteren op impact informatiebehoefte	(Yeoh, 2015)
	Continue afstemming operatie	Gebruikersperspectief en informatiebehoefte wordt frequent onderzocht	(Kutzias, Dukino, & Kett, 2021)
	Continue afstemming operatie	De actuele informatiebehoefte is afgestemd tussen de techniek en de operatie	(Zhaojing, et al., 2020)
	Continue afstemming operatie	Gebruiksfunctionaliteit wordt frequent (1x/ week of dag) gecontroleerd	(Staples, 2016)
	Continue afstemming operatie	De business wordt continu gevraagd feedback te geven en de data analyse modellen te observeren	(Staples, 2016)
	Continue afstemming operatie	Regelmatische afstemming met eigenaren en beslisser die de data analyse modellen gebruiken	(Chang, Abdel-Basset, & Ramachandran, 2018)
	Continue afstemming operatie	Afstemming met de operatie over kansen en mogelijkheden	(Zhao, 2022)
	Continue afstemming operatie	Gebruik maken van nieuwe technieken en mogelijkheden voor verbetering	(Francesco, Demi, Magrini, Marzi, & Papa, 2021)
Proces	Flexibiliteit	Hergebruik van eerder werk wordt efficiënt en effectief gedaan	(Brodsky & Luo, 2015)
	Flexibiliteit	Regelmatische afstemming met eigenaren en beslisser die de data analyse modellen gebruiken	(Chang, Abdel-Basset, & Ramachandran, 2018)
	Flexibiliteit	Toepassen van Efficiënte Data Analytics	(Zhao, 2022)
	Iteratief proces	Continue iteratie bij data analyse modellen	(Yeoh, 2015)
	Iteratief proces	Proces wordt continu geoptimaliseerd	(Kutzias, Dukino, & Kett, 2021)
Techniek	Technisch robuuste data pipeline	Gedocumenteerde en toegankelijke dataflow bron naar data analyse model	(Zhaojing, et al., 2020)
	Technisch robuuste data pipeline	Er wordt gekeken naar meest optimale wijze waarop een data pipeline wordt ingericht	(Zhaojing, et al., 2020)
	Technisch robuuste data pipeline	Gebruik van eenvoudige codes zijn key	(Brodsky & Luo, 2015)
	Verantwoorden kwaliteit	Consistentie in gebruik van de producten binnen data analyses	(Chang, Abdel-Basset, & Ramachandran, 2018)
	Verantwoorden kwaliteit	Validatie continu op de behoefte vanuit de operatie	(Yeoh, 2015)
	Verantwoorden kwaliteit	Integriteit is getest en beoordeeld	(Yeoh, 2015)
	Verantwoorden kwaliteit	Continue verbinding tussen techniek en business	(Yeoh, 2015)
	Verantwoorden kwaliteit	Bepalen van betrouwbaarheid en validiteit op basis van bedrijfsbrede standaarden	(Chang, Abdel-Basset, & Ramachandran, 2018)
	Verantwoorden kwaliteit	Toegankelijkheid ingeregeld voor meerdere lagen binnen de organisatie	(Ransbotham, 2015)
	Verantwoorden kwaliteit	Eenvoudige inzichten leiden tot verbeterde beslissingen	(Ransbotham, 2015)
Verantwoorden kwaliteit	Er wordt een trade-off gemaakt tussen complexiteit en kwaliteit van de oplossing	(Zhaojing, et al., 2020)	

FIGUUR 3 - CODERINGSSCHEMA TER VOORBEREIDING OP ANALYSE EN TREKKEN VAN VERBAND TUSSEN KERNBEGRIPPEN, INDICATOREN EN BIJBEHORENDE WETENSCHAPPELIJKE BRON

2.6 Doel van het vervolgonderzoek

Voor vervolgonderzoek zal er gekeken moeten worden in welke mate de thema's zich gedragen binnen het fenomeen: De houdbaarheid van data analyse modellen. De resultaten zullen met een expert, die operationeel opereert en expertise vanuit hands-on praktijk heeft, besproken moeten worden (Yeoh, 2015). Dit zal er ook toe leiden dat de resultaten steeds concreter en bruikbaarder zullen worden (Saunders, et al., 2017, pp. 39-44). Daarbij zal onderzocht moeten worden hoe dit binnen verschillende contexten plaats zal gaan vinden. Wanneer gesproken wordt over de lange termijn en houdbaarheid van specifieke producten en data analyse modellen, zal er onderscheid worden gemaakt tussen enerzijds projectmatige werken ten opzichte van het onderhouden van producten (Dama-international, 2017). Projectmatig werken is namelijk een wezenlijk andere discipline dan het onderhouden van een product (Rahul & Banyal, 2020). Daarnaast zal de term "houdbaarheid" ook breder onderzocht moeten worden en onderzocht worden wat de "levensvatbaarheid" is van data analyse modellen.

De centrale onderzoeksvraag voor dit onderzoek luidt: "Welke factoren bepalen het succesvol in stand houden van data analyse modellen, en hoe lang is de kwaliteit hiervan in stand te houden?" Het uiteenzetten van deze vraag, zal de toelichting hiervan moeten vereenvoudigen (Saunders, et al., 2017). Om deze vraag te beantwoorden, zal in dit onderzoek de aandacht gegeven worden aan de factoren die het succes van een data analyse model definiëren, namelijk: Kwaliteit, houdbaarheid, bruikbaarheid en flexibiliteit (KD Nuggets, 2023). Aansluitend zal er voornamelijk gezocht worden op de toepassing van het CRISP-DM model in de laatste kritische fase van adoptie (Saltz, 2021). Gezien de beperkte omschrijving van een beheerfase, zal het vervolgonderzoek in staat moeten zijn om in lijn met het CRISP-DM model een volledige uitspraak te kunnen doen (Saltz & Grady, 2017). Daarmee wordt het mogelijk nog meer waarde toe te voegen aan het CRISP-DM model, daar deze in de beheerfase tekort schiet (Saltz, 2021).

H3 Methodologie

3.1 Conceptueel ontwerp: keuze van onderzoeksmethode(n)

Het onderzoek verkent de houdbaarheid van data analyse modellen kwalitatief en contextueel, gericht op het identificeren van factoren die dit beïnvloeden. Saunders, et al. (2017, pp. 78-79) onderstrepen het belang van representatieve cases en context in dit soort onderzoek. Gibbert & Ruigrok (2010) voegen toe: "Case studies enable a researcher to study contemporary phenomena in a real-life setting, where boundaries between context and phenomenon tend to be blurred (Stake, 1995; Yin, 1994)", wat de methodologische aanpak illustreert.

3.2 Technisch ontwerp: uitwerking van de methode

Het onderzoek gebruikt casestudies en diepte-interviews om de houdbaarheid van data analyse modellen te onderzoeken, met nadruk op zorgvuldige selectie van cases en geïnterviewden. Kennis over het fenomeen wordt eerst verzameld, waarna experts helpen bij het identificeren van relevante cases. De resultaten worden gecodeerd om de invloed van business enablers en hun interacties te analyseren.

De onderzoeksmethodiek zal als volgt worden opgebouwd:

1. Verkennende interviews (semi-gestructureerd)
2. Analyseren en coderen
3. Cases study (toetsing)

Minder geschikte onderzoeksmethodieken, die ook meegenomen zijn in de afweging, worden hieronder kort benoemd. Op basis van het boek van Saunders, et al. (2017) over onderzoeksmethoden voor wetenschappelijk onderzoek zijn de volgende keuzes gemaakt:

Verschillende methoden zoals enquêtes, de Delphi-methode, focusgroepen, uitgebreide literatuurstudie, observaties, en documentanalyse werden overwogen voor dit onderzoek, maar uiteindelijk als minder geschikt bevonden. Enquêtes en focusgroepen bieden beperkte diepgang en flexibiliteit, terwijl de Delphi-methode en observaties beperkt zijn door tijd en toegankelijkheid van experts. Literatuurstudies missen actualiteit door snelle technische ontwikkelingen, en documentanalyse wordt beperkt door de beschikbaarheid en subjectiviteit van documenten.

3.2.1 Interviews

De sectie over interviews beschrijft het proces van het afnemen van semigestructureerde interviews met experts om theoretische bevindingen te toetsen en inzicht te krijgen in de houdbaarheid van data analyse modellen. Saunders, et al. (2017, p. 443) en Vidgen (2017) ondersteunen dit proces door het belang van deze interviews te benadrukken voor het verkennen van thema's en het bereiken van theoretische saturatie. Het interviewprotocol is ontworpen om diepte en structuur te bieden aan de gesprekken, gericht op het verkrijgen van relevante kennis en ervaringen van de experts. Ook zal er aan de hand van het framework van Dama international (2015) een verband en vergelijk worden gemaakt met de uitkomsten van de resultaten. Door de resultaten volgens het raamwerk uiteen te zetten zal het eveneens een uitkomst zijn om deze te toetsen aan de theorie.

Aan de hand van een gedegen interviewprotocol zal er structuur en benodigde diepgang aangebracht kunnen worden aan het afnemen en verzamelen van gegevens tijdens de interviews (Saunders, et al., 2017). Het interview protocol is als volgt opgebouwd:

1. Inleiding en toelichting onderzoek
2. Eigenschappen expert en
3. Hoofdvragen vanuit thematische analyse en raamwerk
4. Verdiepende vragen op basis van de antwoorden
5. Feedback en discussie
6. Afsluiten interview

In de appendix (zie A5.1 Interviewprotocol) een uitgebreid interviewprotocol. Merk ook op dat slechts zeer beperkte voorkennis over dit onderzoek te geven alvorens het interview is afgenomen.

3.2.2 Analyseren en coderen

Nadat de interviews zijn gevoerd en theoretische saturatie is opgetreden, is het belangrijk om de resultaten van de interviews met elkaar te vergelijken. Dit kan wordt bereikt door op basis van de codering verbanden te trekken. Door structureel te coderen, zullen de overeenkomsten vergeleken worden met de resultaten uit het literatuuronderzoek. De codering zal gebeuren aan de hand van het coderingsschema (zie Figuur 3). Basis hiervoor is de uitkomst van de thematische analyse en het theoretische raamwerk.

Voor het trekken van de verbanden zal hybrid coding worden toegepast. De stappen die in hybrid coding worden toegepast zijn: (Saunders, et al., 2017, pp. 234-242)

1. *Data verzameling*
De gegevens vanuit zowel theoretisch onderzoek als de interviews worden samengevoegd.
2. *Gegevenspreparatie*
Beide gegevensbronnen zullen gecodeerd worden aan de hand van het coderingsschema, waardoor het mogelijk is de gegevens kwantitatief te benaderen.
3. *Samenvoegen gegevens*
De gegevens van verschillende bronnen zullen worden toegepast om overeenkomsten te vinden in context, fenomeen en eigenschappen daarvan. Aan hand hiervan kunnen categorieën en thema's worden verbonden aan de eigenschappen.
4. *Onderzoeken*
Zodra de gecodeerde gegevens gereed zijn, is het mogelijk om aan de hand van categorieën en thema's patronen te herkennen en analyseren. Dit zal overzicht genereren. Er zal daarbij voornamelijk gelet worden op hoe vaak specifieke codes zijn gebruikt en in welke context.
5. *Bevindingen samenvoegen*
Aan de hand van de thematische analyse uit de literatuuronderzoek en gegevens van de expertinterviews wordt het mogelijk verbanden te leggen en diepere inzichten te krijgen. In dit onderdeel zullen eerder gevonden theorieën geëvalueerd worden.
6. *Conclusie*
Uiteindelijk zullen de bevindingen worden samengevoegd, zodat er conclusies getrokken kunnen worden aan de hand van de methodiek. Ook zal er gereflecteerd worden of eventueel andere omstandigheden tot andere resultaten had kunnen leiden. Dit resulteert in een uitgebreide rapportage die reflecteert met de probleemstelling en hoofdvraag.

De analyse zal klaar zijn op het moment dat er analytische saturatie is bereikt. Dit wil zeggen dat er met bovenstaande stappen net zo lang gecodeerd en geanalyseerd zal worden, totdat er geen nieuwe conclusies meer getrokken kunnen worden.

3.2.3 Case study toetsing

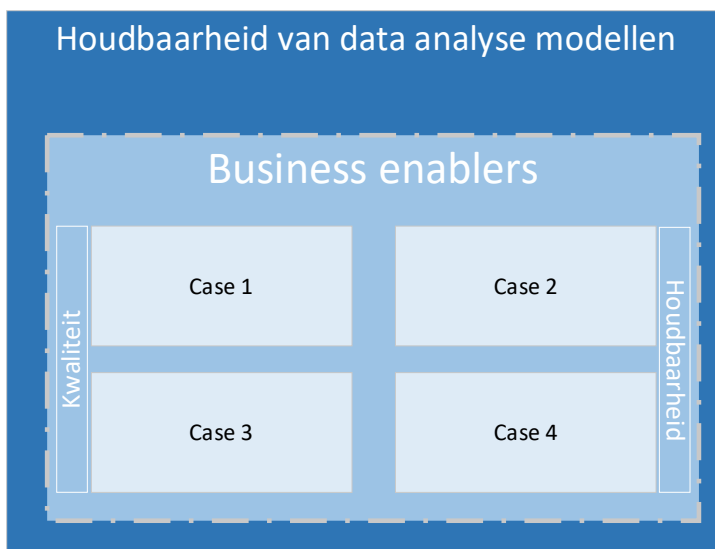
Door de resultaten van een gecodeerde analyse te toetsen met meerdere cases in een multiple embedded cases onderzoek, zoals besproken door Saunders, et al. (2017, pp. 197-199) en geïllustreerd in Figuur 4, wordt het fenomeen van de houdbaarheid van data analyse modellen in diverse contexten onderzocht. Dit onderzoek, gericht op het bevestigen van interviewresultaten, vereist een minimum van vier cases om homogene groepen te analyseren en de bevindingen te valideren tegen het coderingsschema en framework, met als doel een valide en gedifferentieerd beeld te vormen van de onderzoeksbevindingen. Hiermee wordt gestreefd naar het onderzoeken van overwegend homogene onderzoeksgroepen. Verwachting is dat er niet heel veel meer cases onderzocht kunnen worden, gegeven het tijdsbestek van dit onderzoek.

Hieronder enkele criteria- en aandachtspunten voor de case studies die mee worden genomen in het onderzoek (Saunders, et al., 2017, pp. 196-208):

- *Relevant*
Case studies in dit onderzoek moeten direct relevant zijn voor het fenomeen van houdbaarheid van data analyse modellen en actuele, diepgaande informatie bieden, bij voorkeur van de laatste vijf jaar. Het verzamelingsproces van data voor deze case studies wordt gestructureerd volgens het protocol in A6.0.1 Casestudy protocol, waarbij gerichte interviews worden ingezet om de nauwkeurigheid en relevantie van de bevindingen te verzekeren.
- *Diversiteit*
De context van de diverse gecodeerde schema's zullen toegepast moeten kunnen worden vanuit de case studies. Daarbij moet het mogelijk worden de context duidelijk te identificeren, maar ook voldoende variëteit hebben.
- *Representatief*
De case studies zullen theoretisch relevant moeten zijn en een bijdrage leveren aan de probleemstelling en onderzoeksdoelstelling. Hiernaast moeten dit geen uitzonderlijke situaties zijn, maar vooral ook generaliseerbaar en representatief voor binnen de branche.
- *Contrast en variatie*
Er moet voldoende variatie aanwezig zijn in de cases. Hiernaast moet er zonder al te veel diepgang een vergelijking gemaakt kunnen worden tussen de verschillende cases.
- *Toegankelijk*
De case studies moeten een bepaalde minimale hoeveelheid informatie bevatten om te kunnen gebruiken voor dit onderzoek. Deze gegevens moeten beschikbaar zijn in een toegankelijk format zodat deze op juiste wijze kunnen worden ingezet.
- *Omgevingsfactoren*
Er moet een verband getrokken kunnen worden tussen de eigenschappen van het bedrijf en de resultaten vanuit het fenomeen. De eigenschappen van de cases zullen ook duidelijk moeten worden vastgesteld om replicatie mogelijk te kunnen maken.

Proces case studies

In het onderzoek is gekozen voor een benadering met meerdere ingebedde casestudies om het fenomeen van de houdbaarheid van data analyse modellen te onderzoeken, gegeven de beperkte tijd en middelen. Dit proces omvatte het selecteren van vier cases op basis van de eigenschappen die omschreven zijn in het fenomeen: kwalitatieve en houdbare voorbeelden die de expert binnen de beschikbare organisatie kon aandragen. Deze case zijn in algemeenheid beschouwd en vervolgens geanalyseerd op menselijke, procesgerichte, technische, en cross-functionele aspecten. De analyse richtte zich op het gedrag van de cases ten opzichte van het fenomeen zonder diepgaand onderzoek, met het doel bij te dragen aan de onderzoeksdoelstellingen. Aanvullende inzichten werden verkregen door gekozen expert. De gekozen cases werden onderzocht op relevantie, diversiteit, representativiteit, contrast, variatie, toegankelijkheid, en omgevingsfactoren om een valide en gedifferentieerd beeld van de bevindingen te vormen. Dit proces was gericht op het bevestigen van interviewresultaten en het valideren van de bevindingen tegen een coderingschema en framework, binnen het gestelde tijdsbestek van het onderzoek.



FIGUUR 4 - TOEPASSING VAN MULTIPLE EMBEDDED

3.2.4 Synthese onderzoeksresultaten

Centraal binnen dit onderzoek willen we het hoofddoel bereiken door te verklaren: “hoe houdbaar zijn data analyse modellen” en “welke business enablers beïnvloeden dit”. Er wordt gestreefd om een diepgaand inzicht te verkrijgen in factoren die de houdbaarheid van data analyse modellen beïnvloeden, evenals de strategieën (business enablers) die organisaties kunnen inzetten om ze succesvol te handhaven. Het begrijpen van de houdbaarheid van data analyse modellen is van essentieel belang, omdat het niet alleen de effectiviteit van zakelijke besluitvorming beïnvloedt, maar ook de kosten en inspanningen die organisaties moeten investeren om hun modellen up-to-date en relevant te houden. In de synthese zullen de subvragen beantwoord worden:

1. Welke business enablers beïnvloeden de houdbaarheid van data analyse modellen?

Er zal onderzocht worden welke externe en interne factoren, zoals technologische ontwikkelingen, en organisatorische dynamiek, van invloed zijn op de levensduur van data analyse modellen.

2. Welke strategieën helpen om data analyse modellen succesvol houdbaar te maken?

Er zullen verschillende strategieën en benaderingen worden onderzocht die organisaties kunnen inzetten om de levensduur en effectiviteit van hun data analyse modellen te verbeteren.

3. In welke mate is het van belang om de houdbaarheid van data analyse modellen in stand te houden, en verschilt dit per specifieke toepassing?

Er zal onderzocht worden in welke mate de levensduur van data analyse modellen van belang is in verschillende toepassingen, en identificeren welke variabelen deze specifieke relevantie bepalen.

Door te luisteren naar de inzichten van deskundigen en door de praktische toepassing in echte organisaties te onderzoeken, wordt er gestreefd om een holistisch begrip te ontwikkelen van de houdbaarheid van data analyse modellen en de business enablers die dit beïnvloeden. Dit onderzoek is van belang voor organisaties die streven naar het maximaliseren van de waarde van hun gegevens en analytische inspanningen. Het biedt mogelijkheid om houdbaarheid van data analyse modellen te versterken en diens toepassing te verbeteren in een wereld die voortdurend in beweging is.

3.4.5 Ethische verantwoording

In dit onderzoek wordt ethische verantwoording serieus genomen, in lijn met de richtlijnen van Laudon & Laudon (2018) en Saunders, et al. (2017, pp. 232-279). Er wordt zorgvuldig omgegaan met persoonlijke en vertrouwelijke informatie, waarbij deze waar nodig wordt gemaskeerd en er duidelijke afspraken worden gemaakt over wat gedeeld mag worden. Discriminatie en onrechtvaardig handelen worden vermeden, met constante aandacht voor het naleven van ethische principes in alle fasen van het onderzoek. Er zal duidelijk gecommuniceerd worden met de organisaties en betrokkenen wat er wel en niet mag worden gedeeld. Er wordt verantwoord omgegaan met de informatie en de rechten van personen niet geschonden worden. Ook zal ruggenspraak worden gehouden om te voorkomen dat ethische principes worden overtreden (Laudon & Laudon, 2018) (Sharda, 2017).

3.3 Gegevensanalyse

Middels hierboven benoemde stappen zal naar verwachting de hoofvraag en deelvragen beantwoord kunnen worden. Er zijn echter wel enkele nadelen vanuit het inductieve karakter van dit onderzoek. Het zal namelijk enige tijd duren voor het opsporen van de juiste experts om te interviewen en de juiste selectie voor cases te maken. Daarnaast zal de volgorde uitmaken en in de weg kunnen zitten indien niet alles in hierboven benoemde volgorde kan worden uitgevoerd. Daarentegen is het nodig hiervoor tijdig te beginnen met de selectie van experts voor interviews en selectie van geschikte cases. Hiernaast is een groot nadeel dat bij het ontwikkelen van de theorieën het mogelijk is dat er tegenstrijdigheden uit de brede selectie aan experts en cases met zich meebrengt. Hiervoor is als tegenmaatregel ervoor gekozen tussentijds de resultaten te toetsen voor het maken van een volgende stap.

3.4 Reflectie validiteit, betrouwbaarheid en ethische aspecten

Ten aanzien van de validiteit en betrouwbaarheid binnen dit onderzoek, wordt voornamelijk de voorgeschreven methodiek van Saunders (2017) gevolgd. Hierin wordt uitgebreid stilgestaan Yin's tactics (2003) binnen het uitvoeren van case study onderzoek. Figuur 5 hieronder staat een overzicht.

Test	Omschrijving	Case Study Tactiek	Fase onderzoek
<i>Construct validiteit</i>	Vaststellen juiste operationele meetinstrumenten	Gebruik van meerdere onderzoeksbronnen, keten van bewijs vaststellen	Data collectie
<i>Interne validiteit</i>	Vaststellen causale relatie en onderscheiden van niet causale verbanden	Cognitieve en opbouwende verklaringen gebruiken	Data analyse
<i>Externe validiteit</i>	Vaststellen domein waartoe resultaten generaliseerbaar zijn	Gebruik van replicatie logica in multiple case studies	Onderzoeksofzet
<i>Betrouwbaarheid</i>	Aantonen dat iemand anders met dezelfde middelen hetzelfde resultaat behaald	Gebruik van case studie protocol en ontwikkelen case study database	Data collectie

FIGUUR 5 - TACTICS VAN YIN BIJ TOEPASSING CASE STUDIES

3.4.1 Construct validiteit

Construct validiteit omschrijft of het onderzoek meet wat het beoogt te meten, benadrukt door Saunders, et al. (2017, p. 517). Uniforme interviewinstellingen en duidelijke communicatie over gebruikte definities dragen bij aan de validiteit.

Het verkennende en iteratieve karakter van dit onderzoek introduceert het risico dat de methodiek niet volledig aansluit bij de onderzoeksvragen, wat kan leiden tot beperkte behandeling van deelvragen. Dit risico wordt geminimaliseerd door focus en herhaalde duidelijkheid in definities binnen het onderzoek.

3.4.2 Interne validiteit

Construct validiteit zorgt ervoor dat wat gemeten wordt in het onderzoek daadwerkelijk overeenkomt met het onderzoeksdoel, volgens Saunders, et al. (2017, p. 517). Dit wordt bereikt door uniforme interviewsettings en duidelijke communicatie over definities.

Het risico in dit verkennende onderzoek is dat de methodiek de onderzoeksvraag niet volledig kan dekken, wat kan leiden tot beperkte of incorrecte behandeling van deelvragen. Dit wordt geminimaliseerd door een heldere focus en het consequent hanteren van duidelijke definities.

3.4.3 Externe validiteit

Externe validiteit betreft de generaliseerbaarheid van onderzoeksresultaten naar andere settings of contexten, zoals beschreven door Saunders, et al. (2017, pp. 216-217). Het onderzoek streeft naar representativiteit door variëteit in de onderzochte bedrijven en contexten en maakt gebruik van replicatie logica voor case studies.

Een risico binnen dit onderzoek op het gebied van externe validiteit is het behoud van de mogelijkheid om de onderzoeksresultaten voldoende te generaliseren. Dit risico van beperkte generaliseerbaarheid wordt aangepakt door gegevens uit diverse bronnen te verzamelen tot theoretische saturatie is bereikt, hoewel dit in verkennend onderzoek complex kan zijn (Saunders, et al., 2017, pp. 239-242).

3.4.4 Betrouwbaarheid

Betrouwbaarheid in dit onderzoek, gedefinieerd door Saunders, et al. (2017, pp. 363-366), betreft de reproduceerbaarheid van resultaten met dezelfde gegevens door anderen. Er wordt gestreefd naar transparantie en documentatie voor inzage, terwijl de consistentie van resultaten uit diepte-interviews wordt gewaarborgd door een gestructureerd protocol en het systematisch opbouwen van een database via hybride codering.

In dit onderzoek vormt het risico op beperkte externe validiteit een uitdaging voor de generaliseerbaarheid van de resultaten. Dit risico wordt aangepakt door gegevensverzameling uit een breed spectrum van organisaties, cases, en individuen, strevend naar theoretische saturatie zoals Saunders, et al. (2017, pp. 239-242) adviseren. Het juiste aantal bedrijven moet worden onderzocht, hoewel de nadruk op generaliseerbaarheid in verkennend onderzoek beperkt blijft door de intrinsieke moeilijkheid om brede toepasbaarheid te waarborgen.

4. Resultaten

Om gedegen verslag te doen van de interviewresultaten, zal allereerst worden toegelicht hoe het afnemen van de interviews is gegaan. Hierna zullen de algehele conclusie worden samengevat, waarbij de resultaten van de coderingen worden toegevoegd. Tot slot zal een gereflecteerd worden op eerdere bevindingen vanuit de theorie en dus hoe de resultaten van het interview zich verhouden tot het literatuuronderzoek.

4.1 Uitvoer van het onderzoek

In dit gedeelte van het onderzoek wordt de methodologie voor het verkrijgen van resultaten uiteengezet, met een specifieke focus op interviews gericht op de centrale onderzoeksvraag: "Wat bepaalt de houdbaarheid van data analyse modellen?" Door interviews met vier ervaren professionals uit drie verschillende organisaties, met een gecombineerde ervaring van circa 75 jaar in data analyse, werd getracht diepgaande inzichten te verkrijgen in de factoren die de levensduur van

data analyse modellen beïnvloeden. Deze interviews, uitgevoerd in vertrouwde settings en variërend in duur van 1 tot 1,5 uur, streefden naar het verzamelen van praktijkgerichte kennis zonder te verdwalen in theoretische discussies, zoals gedetailleerd in A5.2 Eigenschappen geïnterviewden.

Theoretische saturatie werd bereikt na het vierde interview, wat aangaf dat verdere gesprekken geen nieuwe inzichten zouden opleveren. Dit leidde tot een verschuiving van de focus van het onderzoek naar een meer praktijkgerichte benadering, ondanks initiële plannen om de literatuur centraal te stellen. Belangrijke thema's zoals aanpassingsvermogen, retentiebeleid, en de actualiteitswaarde van informatie kwamen naar voren.

De interviews werden opgenomen, getranscribeerd en handmatig gecodeerd, aangezien een poging tot automatisch coderen niet succesvol was. Dit proces bevestigde de aanwezigheid van data-saturatie. Het onderzoek benadrukte het onderscheid tussen theoretische inzichten en praktijkervaring, en onderzocht de invloed van verschillende soorten data analyse modellen en stakeholders.

Een extra interview binnen een van de deelnemende organisaties bood aanvullende inzichten door het delen van kennis over vier specifieke casestudies, voornamelijk gericht op rapportagemodellen die de besluitvorming ondersteunen. Deze casestudies, geselecteerd op basis van de expertise van de geïnterviewde, benadrukten de praktische toepasbaarheid en het gedrag van data analyse modellen in praktijkscenario's, en verrijkten het begrip van de dynamiek achter de houdbaarheid van deze modellen. Deze casestudies zijn uitgebreid beschreven in bijlage A6 Case studies.

4.2 Resultaten onderzoek

De resultaten van de interviews benadrukken een centrale thematiek rond de houdbaarheid van data analyse modellen, met een focus op continuïteit, betrouwbaarheid, validiteit, en kwaliteit. Een fundamentele bevinding is dat regelmatige afstemming met stakeholders cruciaal is voor het waarborgen van de betrouwbaarheid en validiteit van deze modellen. Om de toegankelijkheid en bruikbaarheid te vergroten, moet de techniek op een laagdrempelige en eenvoudige wijze beschikbaar worden gesteld aan gebruikers. Interessant is dat de technische aspecten niet als de meest beperkende factoren worden gezien; de grootste uitdaging ligt bij de borging van kwaliteit.

De interviews wijzen op het belang van een heldere definitie van de essentie van een data analyse model voordat er met het model gewerkt wordt. Dit benadrukt niet alleen het belang van kwaliteit maar vooral de noodzaak van continuïteit in het gebruik en onderhoud van deze modellen. De nadruk ligt dus niet louter op kwaliteit in de zin van technische specificaties of nauwkeurigheid, maar vooral op het vermogen van een model om consistent en betrouwbaar waarde te blijven leveren over tijd. Deze bevindingen suggereren dat voor het duurzaam inzetten van data analyse modellen, een continue dialoog met en betrokkenheid van stakeholders essentieel is, samen met een duidelijke focus op de praktische toepasbaarheid en het onderhoud van de modellen om hun levensduur te verlengen.

Ter onderbouwing van de resultaten, zie Figuur 6 - Resultaten codering interviews. Dit is het resultaat van de codering, waarin is aangegeven hoe vaak een bepaald thema is benoemd. Herkenbaar zijn de dimensie en indicator die in 2.5.3 Coderingsschema zijn toegelicht, en samengevat als indicator.

Resultaten interviews: Hoe vaak thema benoemd en toegelicht in interviews

	1	2	3	4	Eindtotaal
Continue afstemming operatie	23,08%	38,03%	37,50%	31,67%	33,60%
Actuele informatie behoefte afgestemd	1,28%	4,23%	3,57%	3,33%	3,34%
Afstemming doorontwikkeling	1,28%	5,63%	5,36%	3,33%	4,14%
Business geeft feedback	5,13%	0,00%	0,00%	0,00%	0,64%
Gebruik wordt gecontroleerd	1,28%	0,00%	3,57%	1,67%	1,75%
Gebruikersperspectief duidelijk	0,00%	1,41%	1,79%	8,33%	3,98%
Regelmatige afstemming data (analyse) eigenaren	11,54%	16,90%	14,29%	10,00%	12,90%
Sturen op data kwaliteit en impact	1,28%	5,63%	7,14%	1,67%	3,98%
Verbeteren met nieuwe technieken	1,28%	4,23%	1,79%	3,33%	2,87%
Verantwoorden kwaliteit	34,62%	30,99%	26,79%	33,33%	31,21%
Betrouwbaarheid en validiteit gewaarborgd	6,41%	7,04%	5,36%	5,00%	5,73%
Consistentie gebruik codes	2,56%	4,23%	0,00%	1,67%	1,91%
Continue optimalisatie business en techniek	3,85%	4,23%	5,36%	5,00%	4,78%
Continue verbinding techniek en business	5,13%	2,82%	5,36%	3,33%	3,98%
Duidelijke inzichten helpen bij besluitvorming	3,85%	2,82%	5,36%	5,00%	4,46%
Integriteit getest en beoordeeld	7,69%	2,82%	1,79%	6,67%	4,62%
Kwaliteit van de oplossing afgewogen tegenover complexiteit	3,85%	4,23%	1,79%	5,00%	3,82%
Toegankelijkheid ingeregeld in organisatie	1,28%	2,82%	1,79%	1,67%	1,91%
Iteratief proces	16,67%	12,68%	7,14%	10,00%	10,67%
Continu verbeteren data analyse modellen	7,69%	5,63%	3,57%	6,67%	5,73%
Proces continu verbeterd	8,97%	7,04%	3,57%	3,33%	4,94%
Felxibiliteit	8,97%	7,04%	8,93%	13,33%	10,19%
Hergebruik eerdere werkzaamheden	3,85%	5,63%	1,79%	5,00%	4,14%
Toepassen Efficiënte Data Analytics	5,13%	1,41%	7,14%	8,33%	6,05%
Taken en verantwoordelijkheden	7,69%	8,45%	14,29%	6,67%	9,24%
Duidelijk rol medewerker beheer	2,56%	0,00%	0,00%	0,00%	0,32%
Duidelijke proces taken en verantwoordelijkheden	1,28%	5,63%	7,14%	0,00%	3,34%
Gebruikerstoepassingen duidelijk toegepast	2,56%	1,41%	1,79%	1,67%	1,75%
Versimpelde manier van weergeven	1,28%	1,41%	5,36%	5,00%	3,82%
Technisch robuuste data pipeline	8,97%	2,82%	5,36%	5,00%	5,10%
Data lineage duidelijk inzichtelijk	2,56%	0,00%	0,00%	1,67%	0,96%
Gebruik eenvoudige codes	1,28%	0,00%	1,79%	0,00%	0,64%
Meest optimale data pipeline wordt geambieerd	5,13%	2,82%	3,57%	3,33%	3,50%
Eindtotaal	100,00%	100,00%	100,00%	100,00%	100,00%

FIGUUR 6 - RESULTATEN CODERING INTERVIEWS

4.2.1 Resultaten interviews

Continue afstemming operatie (mens)

De noodzaak van constante samenwerking en communicatie wordt benadrukt om modellen relevant te houden. Dit wordt geïllustreerd door citaten zoals: “Continu toetsen of realiteit nog in lijn is met het model” (interview 1) en de voorkeur voor niet-intensieve communicatiemethoden: “Kiezen voor een communicatie model die niet mens intensief is: Zo makkelijk mogelijk” en “Niet constant contact maar laagdrempelig contact- bijv. met een help knop” (interview 4). Het belang van regelmatige besprekingen wordt aangehaald: “Op de agenda staat bij de vaste overleggen. Rapport gemapt in het overleg structuur” (interview 2), met een waarschuwing tegen te veel flexibiliteit: “té flexibel geweest waardoor de houdbaarheid slecht is geworden” (interview 4), en de noodzaak van voortdurende aanpassing: “Afstemming is continu door af te stemmen met je data model en te toetsen aan voorspellende waarden van een model” (interview 1).

Taken en verantwoordelijkheden (mens)

Duidelijke rollen zijn essentieel voor efficiëntie en effectiviteit, met een focus op de gebruiker en eigenaar: “degenen die de data analyse modellen maakt of onderhoudt middelpunt, terwijl een zeer belangrijke taak weggelegd bij de gebruiker en eigenaar” (interview 4). De noodzaak voor soft skills wordt benadrukt: “Taak van het data team moet zijn om het voor de eindgebruik duidelijk en laagdrempelig toe te passen” (interview 4), met het belang van menselijke interactie: “Begint bij de mensen en inrichten van alles. Velden gebruiken waarvoor ze bestemd zijn” (interview 3).

Flexibiliteit (proces)

Aanpassingsvermogen is cruciaal, zoals blijkt uit: “Zodra er een nieuwe regelgeving komt moeten er systemen worden ingevoerd om dit te hanteren” (interview 3), met een waarschuwing voor de risico's van te veel flexibiliteit: “Welk risico aanvaard ik bij de eindgebruiker om dit toe te gaan passen” (interview 2), en het belang van focus: “als je géén afwijking hebt, wat is dan het nut van het model” (interview 1).

Iteratief proces (proces)

De waarde van iteratie wordt onderstreept voor continue verbetering: “Houdbaarheid staat of valt bij hoe dit er uit ziet” (interview 2), met een nadruk op het balanceren tussen ontwikkelingssnelheid en continuïteit: “snelspoor - borgspoor flexibel genoeg - vs. hoe moet dit beheersbaar blijven” (interview 2), en het belang van consistente controle: “Maatregelen vastleggen in een register” (interview 4).

Technisch robuuste data pipeline (techniek)

De essentie van technische robuustheid wordt benadrukt voor betrouwbare analyses: “Hoe hangen de parameters op elkaar af? - juist simpel houden en naar materiële drivers afpellen” (interview 1), met een focus op gebruiksvriendelijkheid: “Vertalen van technische lading naar een boodschap die bij hen is omschreven” (interview 2), en het belang van standaardisatie: “Generiek data-model gebruiken” (interview 2).

Verantwoorden kwaliteit (techniek)

De noodzaak van transparantie en verantwoording over de kwaliteit van data analyse modellen wordt benadrukt, waarbij de kwaliteit van de onderliggende data van primair belang is. Dit wordt onderstreept door citaten zoals: “In principe check je altijd op missing values, outliers en dergelijke” (interview 4), wat wijst op het belang van een gestandaardiseerde aanpak in de ETL-processen. De verantwoording van technische kwaliteit vereist een proactieve benadering, zowel in de ontwikkeling als het beheer van systemen, om betrouwbare, veilige, en effectieve oplossingen te waarborgen. Dit wordt verder verduidelijkt met: “Dit zal in je ETL gebouwd moeten worden op gestandaardiseerde manier” (interview 4), waarbij het definiëren van standaarden essentieel is: “Belangrijk om vast te houden aan bepaalde kwaliteitsnormen”, “normaliseren van de tabellen” (interview 1 en -3).

Essentie van kwaliteit

Kwaliteit binnen data analyse modellen wordt gezien als een combinatie van nauwkeurigheid, relevantie, en aanpasbaarheid aan veranderende omstandigheden. Het essentiële aspect van kwaliteit omvat niet alleen de technische dimensies maar ook de gebruikerservaring en de mate waarin modellen voldoen aan en blijven voldoen aan de verwachtingen. Dit wordt versterkt door de focus op wat werkelijk belangrijk is: “welke bollen zijn nu van glas” (interview 1), wat de noodzaak benadrukt om de kernvragen te begrijpen: “De vraag achter de vraag doorvoeren. Waarom zal je

naar specifieke uitvoering willen kijken?” (interview 1). Het belang van de menselijke factor wordt bekrachtigd: “het is een product beheren en het is dus belangrijk om bijvoorbeeld vertrouwen op te bouwen” (interview 4), met een nadruk op het destilleren van waarde uit data: “Data to diamonds - data onder hoge druk zetten en alleen waardevolle gegevens hier uit te halen” (interview 3). Dit benadrukt het belang van een mensgerichte benadering en het vermogen om door de technische complexiteit heen te kijken om de werkelijke waarde van data analyse modellen te ontsluiten en te onderhouden.

4.2.2 Resultaten case study

De case studies verrijken de bevindingen uit de interviews door de cruciale rol van flexibiliteit, technische robuustheid, en de noodzaak van continue afstemming met operationele behoeften te benadrukken. Ze tonen aan dat volledige voorbereiding op alle mogelijke scenario's uitdagend is, met problemen die vaak pas zichtbaar worden na de implementatie van data analyse modellen, voornamelijk door onvoorziene omstandigheden en een gebrek aan inzicht in veranderende externe factoren. Het strikte volgen van het CRISP-DM model wordt als essentieel beschouwd voor het succes van deze modellen. Zie uitgebreide toelichting over de case in A6.5 Belangrijkste conclusies cases.

De cases onderstrepen het belang van constante communicatie met stakeholders om de kwaliteit en relevantie van data analyse modellen te handhaven, wijzend op de noodzaak van een solide data-infrastructuur en effectief beheer en onderhoud voor de duurzaamheid van de modellen op lange termijn. Deze bevindingen bevestigen en verdiepen de inzichten verkregen uit de interviews, en belichten het management van de levenscyclus van data analyse modellen als een kritisch, doch onderbelicht, onderzoeksveld.

4.2.3 Overeenkomsten en verschillen

Terwijl alle thema's de nadruk leggen op de integratie van menselijke expertise, procesflexibiliteit en technische bekwaamheid, variëren de benaderingen en prioriteiten tussen de verschillende respondenten. Sommigen leggen meer nadruk op technische aspecten, terwijl anderen de menselijke interactie of procesefficiëntie benadrukken.

Overeenkomsten

De interviews tonen een uniforme nadruk op het belang van continue procesverbetering, efficiënte data analyse voor besluitvorming, en de essentie van samenwerking en afstemming tussen verschillende afdelingen en stakeholders. Dit duidt op het universele belang van interdisciplinaire coördinatie en plaatst menselijke expertise als primair, gevolgd door procesafstemming en technische ondersteuning. De CRISP-DM methodiek en de case studies bevestigen het belang van een gestructureerde voorbereiding voor succes.

Verschillen

De diverse interviews belichten verschillende perspectieven en prioriteiten, variërend van de noodzaak van operationele afstemming en realiteitschecks (Interview 1), strategische implementatie en multidisciplinaire samenwerking (Interview 2), organisatorische dynamiek versus modelwaarde (Interview 3), tot communicatiegemak en technische hulpmiddelen voor procesafstemming (Interview 4). Deze verschillen onderstrepen de unieke invalshoeken en expertisegebieden van de respondenten, leidend tot een rijke variatie aan thematische inzichten.

Houdbaarheid en Balans tussen Mens en Techniek

Een cruciale bevinding is de nadruk op de houdbaarheid van data analyse modellen voor langdurige efficiëntie en effectiviteit binnen dynamische bedrijfsomgevingen. Opmerkelijk is het algemene belang van een evenwichtige relatie tussen menselijke expertise en technologische capaciteiten, met een speciale focus op de capaciteit om snel op veranderingen te reageren en de datakwaliteit voortdurend te waarborgen.

Techniek als verbindende element

Techniek fungeert als brug tussen mens en proces, gericht op het bevorderen van efficiëntie. De complexiteit van de menselijke factor onderstreept het belang van eenvoud en gebruiksvriendelijkheid in technische implementaties. Interviews wijzen op een breed gedragen overtuiging van het belang van een evenwicht tussen menselijke inzicht, procesaanpassingsvermogen, en technische vaardigheid. Deze bevindingen pleiten voor een holistische aanpak in de ontwikkeling en het beheer van data analysemodellen, met een constante focus op aanpassing en verbetering om te voldoen aan veranderende bedrijfseisen.

4.3 Resultaat onderzoek vs. theorie

In dit hoofdstuk zal duidelijke feedback worden gegeven over het verschil tussen de theoretische kader, zoals omschreven in hoofdstuk 1 en 2, en de bevindingen vanuit de interview en cases.

Verschillen Theorie versus Praktijk:

Het onderzoek benadrukt de kloof tussen de theoretische levenscyclus van data analysemodellen en de praktische implementatie ervan binnen dynamische bedrijfsomgevingen, waarbij praktische uitdagingen en onvoorziene marktontwikkelingen de complexiteit verhogen. Deze discrepantie, gedetailleerd in de CRISP-DM verwijzing binnen de probleemstelling (CRISP-DM toelichting zie 1.2 Probleemstelling), vereist een brug tussen theorie en praktijk. Het voorstel is om toekomstige modelontwikkelingen en praktijkfundamenten te verkennen, en door reflectie en evaluatie een geïntegreerde aanpak te vormen die theoretische modellen praktisch toepasbaar maakt, ondersteund door een constante dialoog tussen beide domeinen.

Nuanceverschillen:

In de theorie wordt flexibiliteit en iteratieve processen vaak idealistisch en gestructureerd gepresenteerd. In de praktijk, zoals blijkt uit de interviews, zijn deze processen vaak meer adaptief en reagerend op onvoorziene omstandigheden. Dit nuanceverschil onderstreept het belang van een pragmatische context en het niet kunnen voorzien van alle omstandigheden zoals in theorie is omschreven.

Belang van Theorie en Praktijk Vergelijken:

Het vergelijken van theorie met praktijk is essentieel om de relevantie en toepasbaarheid van theoretische concepten te verifiëren. Dit biedt een realistisch beeld van hoe concepten werken in de dynamische en vaak onvoorspelbare bedrijfswereld. Het helpt ook bij het aanpassen van theoretische modellen om ze effectiever en bruikbaar te maken voor praktische toepassingen.

4.4 Resumé en conclusies

De onderzoeksresultaten bieden inzicht in de cruciale aspecten van de houdbaarheid van data analyse modellen, waarbij de nadruk ligt op continuïteit, betrouwbaarheid, validiteit, en kwaliteit.

Deze aspecten worden benaderd vanuit verschillende perspectieven: de afstemming met stakeholders, de flexibiliteit en iteratieve ontwikkeling van modellen, en de verantwoording van kwaliteit en technische robuustheid. Het onderzoek belicht het onderscheid tussen theorie en praktijk en identificeert belangrijke leer- en verbeterpunten. De belangrijkste conclusies die getrokken kunnen worden zijn:

Houdbaarheid van modellen

De levensduur van data analyse modellen is sterk afhankelijk van hun vermogen om zich aan te passen aan veranderende omstandigheden. Regelmatige afstemming binnen multidisciplinaire teams speelt hierin een sleutelrol.

Theorie versus Praktijk

Er wordt een duidelijk verschil waargenomen tussen theoretische benaderingen en hun praktische toepassingen. De realiteit vereist een adaptieve en responsieve benadering, waarbij voorbereiding op onvoorziene omstandigheden cruciaal is, maar ook erkend wordt dat niet op alles voorbereid kan worden.

Belang van Stakeholderafstemming

Effectieve communicatie met stakeholders is essentieel om de relevantie en toepasbaarheid van data analyse modellen te behouden. Dit onderstreept het belang van laagdrempelige interactie tussen ontwikkelaars en eindgebruikers.

Flexibiliteit en Iteratieve Ontwikkeling

De noodzaak voor flexibiliteit en iteratie in ontwikkelingsprocessen wordt benadrukt, met een focus op de capaciteit om snel aan te passen aan zowel kortetermijnveranderingen als langetermijntrends.

Kwaliteitsverantwoording en Technische Robuustheid

De nadruk ligt op het belang van transparantie in kwaliteitsverantwoording en het behoud van een robuuste technische infrastructuur. Dit duidt op het belang van begrijpelijke en toegankelijke technische oplossingen.

Leer- en Verbeterpunten

De vergelijking tussen theorie en praktijk levert waardevolle inzichten op voor het aanpassen van theoretische modellen aan de praktische bedrijfscontext. Dit proces benadrukt het belang van wederzijds leren en continue verbetering.

5. Discussie, conclusies en aanbevelingen

In dit deel van het onderzoeksrapport zal worden gekeken naar de lessons learned van het onderzoek en de eerder behaalde resultaten. Een kritische analyse op het verloop van de interviews, de methodologische aanpak en thematische bevindingen zullen hier worden uiteengezet. Dit biedt waardevolle inzicht over het verloop en waarde van de resultaten. Eerst zal er gekeken worden naar de conclusies van het onderzoek, op basis van eerder gerapporteerde resultaten. Daarbij gekeken naar de resultaten van het literatuur onderzoek ten opzichte van de resultaten van de data verzameling, en wat hiervan geleerd kan worden. Hierna zal de reflectie volgen op basis van de gekozen methodologische aanpak en discussie over de kwaliteit van het geleverde werk.

5.1 Discussie - reflectie

De discussie benadrukte het belang van voorbereiding en literatuurstudie voor dit onderzoek, ondanks moeilijkheden in termdefinitie, vraagstelling en interviewopzet. Er werd vaak gefocust op specifieke punten zonder diepgang. De analyse toonde aan dat praktijkervaring tot "best practices" leidt, die kunnen verschillen van academische theorieën. Het onderzoek onthulde het belang van relevantie, eenvoud en focus op doelstellingen voor data analyse succes, ondanks tijd- en ervaringsbeperkingen en generalisatie-neigingen, wat een kloof tussen praktijk en theorie aantoont en het nut van een gestructureerde aanpak benadrukt.

Het identificeren van patronen was lastig door inconsistente antwoorden, vooroordelen, en een gebrek aan diepgang, maar er was tevredenheid over de resultaten door antwoord op onderzoeksvragen, consensus met literatuur, en enthousiasme van betrokkenen. Het bereiken van saturatie en de mogelijkheid tot triangulatie door literatuur, interviews, en case studies te combineren benadrukken de waarde van een diverse aanpak in onderzoek.

De thema's in het theoretisch kader lijken goed gekozen, aangezien ze belangrijke aspecten van data analyse en -management omvatten. De interviews reflecteren deze thema's, hoewel de praktische uitvoering ervan varieert. Wat betreft de keuze van geïnterviewden, lijkt het erop dat zij relevante inzichten en ervaringen hebben geboden die de thema's van het theoretisch kader weerspiegelen. Dit suggereert dat de juiste personen zijn gekozen om een breed en diepgaand begrip van de onderwerpen te bieden.

Beperkingen van het onderzoek

Dit onderzoek werpt licht op zowel overeenkomsten als verschillen met eerdere literatuur, door de nadruk te leggen op praktische toepasbaarheid en eenvoud. Echter, de studie kent beperkingen, zoals een beperkte diversiteit in de geïnterviewde organisaties, wat een bias in de resultaten kan introduceren door specifieke werkwijzen. De beperkte reikwijdte, door het interviewen van slechts vijf personen uit drie organisaties, en hun focus op reporting en business intelligence, verminderde de breedte van de verkregen perspectieven. Dit, samen met een soms willekeurige selectie van deelnemers en een subjectieve focus, beperkte diepgang en diverse inzichten.

Deze factoren leiden tot gemiste kansen in het verkrijgen van een vollediger begrip van de houdbaarheid van data analyse modellen en benadrukken het belang van diepgaander onderzoek. Er is een duidelijke noodzaak voor een meer omvangrijke en diepgaande benadering, onderbouwd met kwantitatieve resultaten, om de complexiteit en dynamiek van data analyse modellen volledig te begrijpen. Dit vereist een breder spectrum aan perspectieven en een gestructureerd onderzoeksdesign, vrij van de beperkingen die werden ervaren in dit onderzoek, om de kloof tussen theorie en praktijk te overbruggen en een rijkere diversiteit aan inzichten te bieden.

5.2 Conclusies

In dit wetenschappelijk onderzoek wordt de focus gelegd op het identificeren van cruciale factoren die bijdragen aan het succesvol onderhouden en waarborgen van de kwaliteit van data analyse modellen. De centrale onderzoeksvraag luidt: "Welke factoren bepalen het succesvol in stand houden van data analyse modellen, en hoe lang is de kwaliteit hiervan in stand te houden?" Deze vraagstelling is van essentieel belang voor de datawetenschappen, waarbij de levensduur en bruikbaarheid van modellen directe invloed hebben op bedrijfsstrategieën en besluitvormingsprocessen. Uit het onderzoek blijkt dat de volgende factoren van fundamenteel belang zijn voor het onderhouden van de effectiviteit van data analyse modellen, en de mate van houdbaarheid van data analyse modellen sterk beïnvloeden:

- **Communicatie:** Een open en effectieve communicatielijntussen modelontwikkelaars, gebruikers, onderhoudsteams en stakeholders is essentieel voor tijdig adresseren problemen en het aanbrengen van noodzakelijke wijzigingen. De communicatie moet bijdragen aan:
 - o **Continuïteit:** Continuïteit in modelaanpassing en -ontwikkeling is cruciaal om relevant te blijven onder veranderende omstandigheden. Dit gaat verder dan alleen flexibiliteit; het omvat een proactieve houding ten aanzien van evoluerende markttrends, technologische vooruitgang en organisatorische veranderingen.
 - o **Voortschrijdend inzicht:** Het vermogen om modellen te verfijnen en te verbeteren door opgedane ervaringen en inzichten is van onschatbare waarde voor het behoud van hun relevantie en effectiviteit.
- **Product:** Cruciaal voor succes is om data analyse modellen niet alleen als technische constructies te beschouwen, maar ook producten die continu moeten evolueren om te voldoen aan veranderende bedrijfsbehoeften en technologische vooruitgang. Daarvoor dienen deze de volgende eigenschappen te hebben:
 - o **Eenvoud:** Complexiteit kan een barrière vormen voor effectief onderhoud en aanpassingen. Eenvoudigere modellen zijn vaak gemakkelijker bij te werken, te onderhouden en te communiceren naar belanghebbenden. Ofwel “keep it simpel”. Om het product eenvoudig te maken, moet het een duidelijk doel dienen én in staat zijn zich continu te verbeteren
 - o **Doelmatigheid:** De belangrijkste indicator voor kwaliteit van een data analyse model is de mate waarin het blijft voldoen aan de beoogde doelstellingen. Dit vereist een regelmatige afstemming van de modelprestaties tegen de oorspronkelijke vereisten en verwachtingen.

Ten aanzien van de deelvragen, kunnen de volgende zaken worden geconcludeerd:

1. *Welke business enablers beïnvloeden de houdbaarheid van data analyse modellen?*
Business enablers, zoals strategische afstemming, eenvoud van modellen, en effectieve communicatie, spelen een significante rol in de houdbaarheid van data analyse modellen. Het is belangrijk om modellen te benaderen als zelfstandige producten met duidelijke doelstellingen en verwachtingen, die meehelpen in de besluitvorming.
2. *Welke strategieën helpen om data analyse modellen succesvol houdbaar te maken?*
Het waarborgen van kwaliteit vereist een strategische benadering waarbij vragen gesteld worden over de huidige relevantie, behoeften, en de mate waarin het model antwoord geeft op de centrale vraagstelling. Dit proces helpt bij het identificeren van noodzakelijke aanpassingen of de behoefte aan een nieuw model.
3. *In welke mate is het van belang om de houdbaarheid van data analyse modellen in stand te houden, en verschilt dit per specifieke toepassing?*
De relevantie van het onderhoud van data analyse modellen varieert afhankelijk van de specifieke toepassing. Elk model heeft zijn eigen unieke kenmerken, gebruikers en doeleinden, waardoor maatwerk een vereiste is. Sommige modellen zijn bestemd voor eenmalig gebruik, terwijl andere een langdurige inzet vereisen. Dit vereist maatwerk.

Dit onderzoek benadrukt dat de effectiviteit van data analyse modellen in een constant veranderende omgeving afhangt van hun vermogen tot aanpassing, effectieve communicatie, en het behoud van focus op kern doelstellingen. Een holistische benadering, die zowel technische als organisatorische facetten integreert, is cruciaal voor het managen van de complexiteit en het waarborgen van continuïteit in deze dynamische context. Het onderstreept de noodzaak voor data

professionals en bedrijfskundigen om een coherente strategie te hanteren die specialisatie combineert met een brede, geïntegreerde visie voor duurzaam succes in datamanagement.

5.3 Aanbeveling voor de Praktijk

Organisaties moeten hun data analysemodellen continu herzien voor verbetering en aanpassing aan veranderende markten. Belangrijk zijn stakeholderafstemming, flexibiliteit, eenvoud, en technische robuustheid om kwaliteit te garanderen. Deze aanpak verlengt de levensduur van modellen en versterkt het concurrentievoordeel. Hieronder enkele implicaties:

Ontwikkel een Strategie voor Stakeholderbetrokkenheid: Regelmatige communicatie en afstemming met stakeholders zijn essentieel voor relevantie en effectiviteit van data analyse modellen.

Organisaties dienen processen op te zetten voor continue interactie tussen ontwikkelaars, gebruikers en stakeholders. Daarnaast ook reviews en feedbackloops, om gebruikersfeedback effectief in het ontwikkelingsproces te integreren. Implementeer mechanismen voor feedback en betrek stakeholders actief bij het reviewen en aanpassen van modellen.

Flexibiliteit en Continuïteit in Modellen: Flexibiliteit in data analysemodellen is cruciaal om in te spelen op veranderingen en nieuwe inzichten. Organisaties moeten een cultuur van continu leren & verbeteren promoten, met regelmatige updates van modellen op basis van vernieuwde data, technologieën, en doelen. Bij voorkeur volgens een agile aanpak in ontwikkeling en onderhoud.

Vereenvoudig Modellen waar Mogelijk: Eenvoud in data analyse modellen is cruciaal voor onderhoud en effectiviteit. Organisaties moeten modellen simpel houden, gericht op hoofddoelen, met duidelijke documentatie en standaarden/templates.

Technische Robuustheid en Kwaliteitsverantwoording: Eenvoud in data analysemodellen is cruciaal voor onderhoud en effectiviteit. Organisaties moeten modellen simpel houden, gericht op hoofddoelen, met duidelijke documentatie en standaarden/templates.

5.4 Suggesties en richting voor toekomstig onderzoek

Voor toekomstige studies is de inzet van kwantitatieve methoden en uitgebreide enquêtes essentieel, vooral wanneer deze worden toegepast over een breed scala aan bedrijven. Dit benadert niet alleen de diepte en breedte van inzichten in de duurzaamheid en effectiviteit van data analyse modellen, maar verhoogt ook de schaalbaarheid en de praktische toepasbaarheid van de onderzoeksresultaten. Cruciaal is het verder verkennen van de continuïteit van deze modellen en hun vermogen om zich aan te passen aan veranderende omstandigheden, waarbij een holistische benadering die zowel de technische als organisatorische facetten insluit, onmisbaar is. Het onderzoek moet zich eveneens richten op de diversificatie van de geïnterviewde functionarissen, met name die in specifieke ontwikkel- en beheerrollen, om een breder perspectief te verkrijgen en de benodigde competenties en specialisaties aan het licht te brengen. De integratie van maatschappelijk verantwoorde elementen en de uitbreiding naar een internationale context zijn ook van groot belang om de universele relevantie en toepasbaarheid van de bevindingen te verzekeren. Door deze meervoudige benaderingen te volgen, kan het onderzoek diepgaandere en praktisch toepasbare kennis opleveren, essentieel voor de vooruitgang en implementatie van data analyse modellen in een snel evoluerende wereld.

5.5 Framework houdbaarheid van data analyse modellen

De houdbaarheid van data analyse modellen is niet alleen afhankelijk van de inherente kwaliteit van de modellen, ook mate waaraan deze aan veranderende inzichten en omstandigheden kunnen wijzigen. Dit adaptieve vermogen wordt beïnvloed door doelmatigheid, eenvoud van het product, de oplossing, en effectieve communicatie waarmee continuïteit kan worden geborgd.

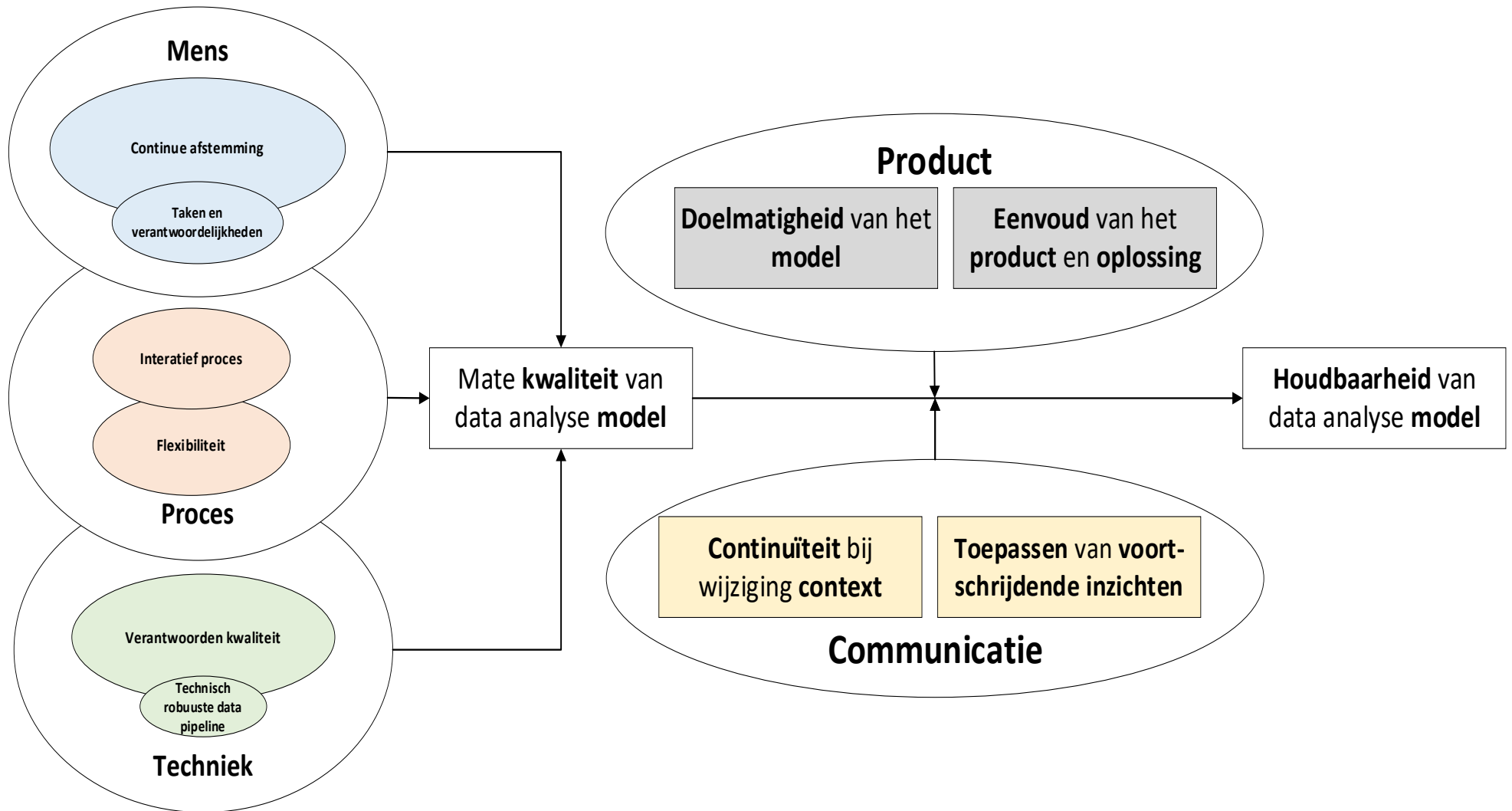
In het verlengde van DAMA DMBOK zijn de menselijke aspecten, zoals de continue afstemming van de operatie en heldere taken en verantwoordelijkheden, en de procesmatige aspecten, zoals flexibiliteit en een iteratief proces, cruciaal. Deze worden gecombineerd door technische factoren, zoals een robuuste data pipeline en het verantwoorden van de kwaliteit. Het is deze synergie tussen mens, proces en techniek die het fundament vormt voor een duurzame aanpak in het ontwikkelen en onderhouden van data analyse modellen. Het formaat van de gekleurde, gelijk aan de thema's uit Figuur 6 - Resultaten codering interviews, weergeven de mate van invloed aan.

In Figuur 7 - Framework: Houdbaarheid van data analyse modellen is een visuele weergave van de conclusies van het onderzoek weergegeven, waarbij het belang van een geïntegreerde benadering van data analyses wordt benadrukt, die technische vaardigheden combineert met organisatorische inzichten. De uitkomst toont dat de effectiviteit van data analyse modellen in een constante veranderende omgeving afhangt van aanpassingsvermogen, effectieve communicatie en het behoud van focus op belangrijkste doelstellingen. Hierbij wordt een holistische benadering aanbevolen, waarbij zowel technische als organisatorische facetten combineert. Dit is essentieel voor het borgen van continuïteit in een dynamische en complexe omgeving, en benadrukt een structurele aanpak.

Het framework visualiseert de benadering en dient als een strategische blauwdruk voor ontwikkeling van veerkrachtige data analyse modellen, die voorbereid zijn op de toekomst. Het accentueert de noodzaak voor data professionals en bedrijfskundigen om een gezamenlijke strategie te hanteren die specialisatie combineert met oog op kwalitatief goede en houdbare data analyse modellen.

Deze aanpak stimuleert een cultuur van continue evaluatie en verbetering, waarbij aanpassingen worden aangemoedigd op basis van regelmatige feedback, met een focus op duidelijke en eenvoudige modellen die makkelijk te onderhouden zijn. In de praktijk betekent dit de ontwikkeling van data analysemodellen die niet alleen theoretisch onderbouwd zijn, maar die ook praktische relevantie hebben en de toets van constante verandering kunnen doorstaan.

Samengevat, het framework dat uit dit onderzoek naar voren komt, biedt een uitgebreide weergave van de verbanden en dynamieken tussen de factoren die de houdbaarheid van data analyse modellen beïnvloeden. Het dient als een visuele conclusie die de bevindingen van het onderzoek integreert en biedt een kader voor data-professionals om modellen te ontwikkelen en te onderhouden die robuust en relevant blijven in een snel evoluerende wereld.



FIGUUR 7 - FRAMEWORK: HOUDBAARHEID VAN DATA ANALYSE MODELLEN

6. Tot slot

Na een avontuurlijke reis vol data, interviews en ontdekkingen, lig alhier mijn finale scriptie. Dit onderzoek, gericht op de houdbaarheid en kwaliteit van data-analysemodellen, was als het in de lucht houden van 10 bollen, waarvan drie van glas – een delicate balans tussen theorie en praktijk, waarbij elke misstap de integriteit van de bevindingen tot scherven kon leiden. Toch, met een mix van voorzichtigheid en durf, ben ik erin geslaagd géén glazenbol te laten vallen. Zonder wrijving géén glans, als al in één van de interviews werd besproken: “Data tot the diamonds”, die “onder druk” wordt gerealiseerd. De diamanten zijn in dit geval een mooie metafoor voor hetgeen waar data analyse modellen voor staan: De kwaliteit en houdbaarheid is in géén van alle gevallen goed. Het is altijd belangrijk voldoende context te schetsen en relevantie te geven: “Snijdt het nog hout?” moet je bij de houdbaarheid en kwaliteit van data analyse modellen constant afvragen. Daarnaast is het van belang dat niet gekeken wordt naar “te flexibele data analyse modellen, die zijn niet goed”, maar vooral naar de continuïteit.

Deze reis heeft mij persoonlijk veel geleerd, niet alleen over het opstellen van wetenschappelijk onderzoek maar ook over de waarde van toenadering zoeken bij de geïnterviewden. Het blijkt dat het probleem – of beter gezegd, de uitdaging – van het behouden van de relevantie en kwaliteit van data-analysemodellen, een universele ervaring is. Echter, kennis over de praktijk bleek zowel een zegen als een vloek; het biedt inzicht, maar ook een potentiële bias.

Een speciaal woord van dank gaat uit naar mijn begeleiding, die prettige kennis en begeleiding heeft kunnen leveren. Ondanks de switch van begeleider vlak voor het inleveren van de voorbereiding en aanpak van dit onderzoek, heb ik veel waardevolle input gekregen.

Als ik terugkijk op dit traject, is "Keep it simpel" misschien wel het meest waardevolle advies gebleken, ook voor mij persoonlijk. In de complexe wereld van data-analyse is eenvoud vaak de sleutel tot begrip en bruikbaarheid. Dit onderzoek heeft getracht de complexiteit te doorgronden, zonder de eenvoud uit het oog te verliezen, en te delen met de lezers.

Ik hoop dat deze scriptie toekomstige onderzoekers inspireert met de boodschap om de fragiele bollen zorgvuldig te jongleren, en wens dat hun onderzoek vruchtbaar zal zijn, met scherpe inzichten en data die continu transformeren in waardevolle kennis.

Dank allen.

B1 Bibliografie

- Solarte, J. (2002). *A Proposed Data Mining Methodology and Its Application to Industrial Engineering - Master's thesis*. Knoxville: University of Tennessee.
- Brodsky, A., & Luo, J. (2015). Decision Guidance Analytics Language (DGAL) - Toward Reusable Knowledge Base Centric Modeling. doi:<https://doi.org/10.5220/0005349600670078>
- Chang, V., Abdel-Basset, M., & Ramachandran, M. (2018). Towards a Reuse Strategic Decision Pattern Framework – from Theories to Practices. *Information Systems Frontiers*, 21(1), pp. 27–44. doi:<https://doi.org/10.1007/s10796-018-9853-8>
- Chapman, P., Clinton, J., & Randy, K. (2000). *CRISP-DM 1.0*. Members of the CRISP-DM consortium.
- Dama-international. (2017). *Dama-dmbok: Data management body of knowledge (2nd edition)*. (S. Earley, D. Henderson, & Data Management Association, Red.) Denville, NJ, USA: Technics Publications, LLC.
- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review* 84, pp. 98-107, 134.
- Dedehayir, O., & Steinert, M. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change* 108, pp. 28–41. doi:<https://doi.org/10.1016/j.techfore.2016.04.005>
- Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of Business Process Management*. Berlin, Germany: Springer . doi:10.1007/978-3-642-33143-5
- Francesco, C., Demi, S., Magrini, A., Marzi, G., & Papa, A. (2021). Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation. *Journal of Business Research Volume 123*, pp. 1-13. doi:<https://doi.org/10.1016/j.jbusres.2020.09.023>
- KD Nuggets. (2023, maart 16). *Your Checklist to Get Data Science Implemented in Production*. Opgehaald van KD Nuggets: <https://www.kdnuggets.com/2017/06/dataiku-checklist-data-science-implemented-production.html>
- Knopf, J. (2006). Doing a Literature Review. *PS Political Science & Politics*, 39(1), 127–132. doi:<https://doi.org/10.1017/s1049096506060264>
- Kurgan, L., & Musilek, P. (2006). A survey of knowledge discovery and data mining process models. *Knowledge Engineering Review* 21(1), pp. 1-24.
- Kutzias, D., Dukino, C., & Kett, H. (2021). Towards a Continuous Process Model for Data Science Projects. *Springer eBooks (pp. 204–210)*. doi:https://doi.org/10.1007/978-3-030-80840-2_23
- Laudon, K., & Laudon, J. (2018). Ethical and Social Issues in . In K. Laudon, & J. and Laudon, *Management Information Systems: Managing the Digital Firm* (15 ed., pp. 150-162). UK: Pearson: Harlow.
- Laudon, K., & Laudon, J. (2019). *Management Information Systems: Managing the Digital Firm*. Pearson.
- Lavalle, S., Lesser, E., Schockley, R., & Hopkins, M. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*, 52(2), pp. 21-31.

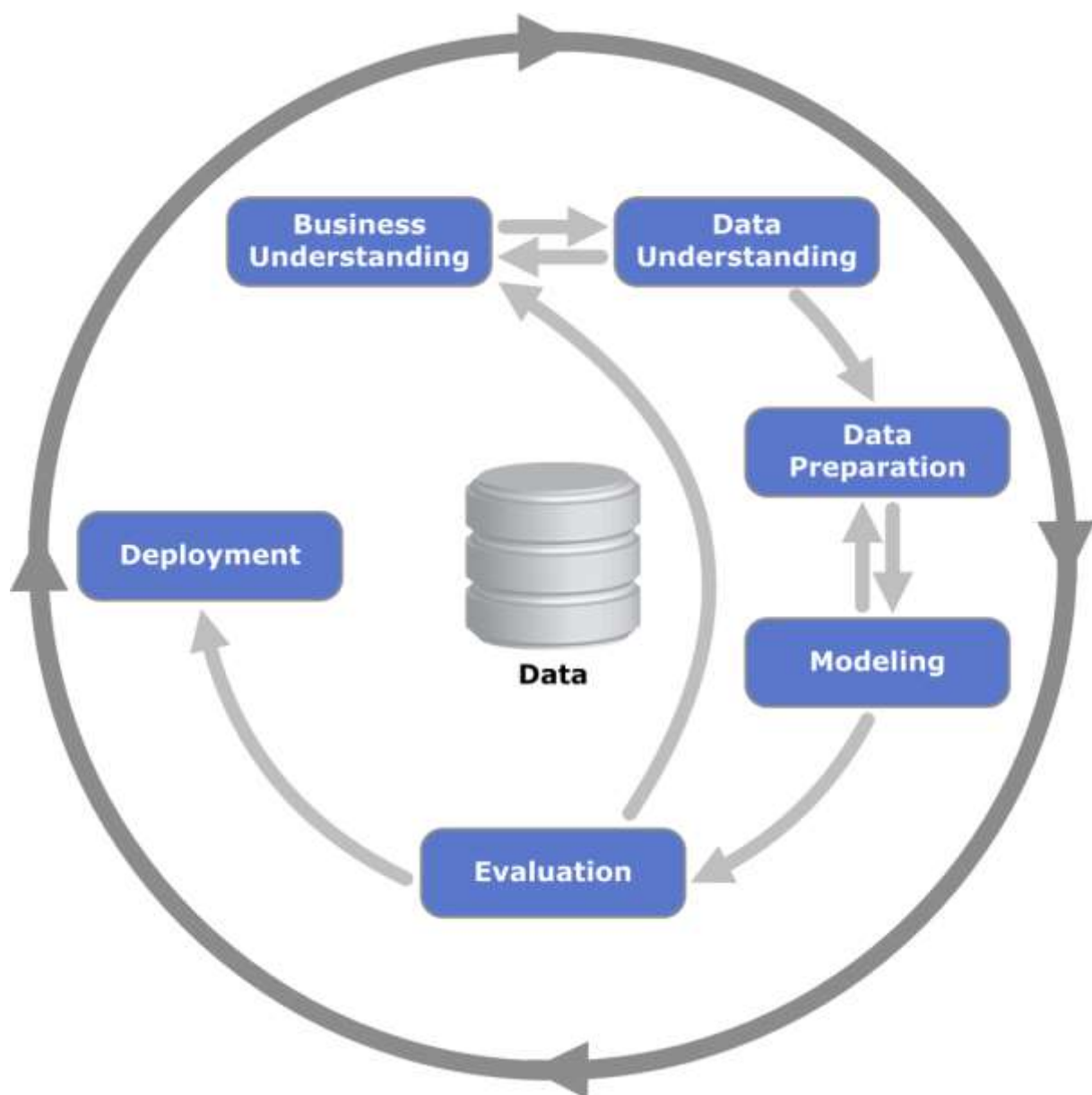
- Lavalle, S., Lesser, E., Schockley, R., & Hopkins, M. (2011). *Big Data, Analytics and the Path From Insights to Value*.
- Luo, Z., Yeung, S., Zhang, M., Zheng, K., & Zh. (2020). MLCask: Efficient Management of Component Evolution in Collaborative Data Analytics Pipelines. *arXiv (Cornell University)*.
doi:<https://doi.org/10.48550/arxiv.2010.10246>
- Mariscal, G., Marbán, Ó., & Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, 25(2), pp. 137–166.
doi:<https://doi.org/10.1017/s0269888910000032>
- Obeidat, M., North, M., & Richardson, R. (2015). Business Intelligence Technology, Applications, and Trends. *International Management Review Vol. 11 No. 2 2015*, pp. 47-55.
- Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Social Science Research Network*.
doi:<https://doi.org/10.2139/ssrn.1954824>
- Rahul, K., & Banyal, R. (2020). Data Life Cycle Management in Big Data Analytics. *Procedia Computer Science*, 173, pp. 364–371. doi:<https://doi.org/10.1016/j.procs.2020.06.042>
- Ransbotham, S. (2015). Minding the analytics gap. *Dialnet*.
doi:<https://dialnet.unirioja.es/servlet/articulo?codigo=5130994>
- Saltz, J. (2021, December). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. *In proceedings: 2021 IEEE International Conference on Big Data*, pp. 2337-2344.
doi:<https://doi.org/10.1109/bigdata52589.2021.9671634>
- Saltz, J., & Grady, N. (2017). The ambiguity of data science team roles and the need for a data science workforce framework. doi:<https://doi.org/10.1109/bigdata.2017.8258190>
- Saltz, J., & Hotz, N. (2020, December). Identifying the most common frameworks data science teams use to structure and coordinate their projects. *In proceedings: 2020 IEEE International Conference on Big Data (Big Data)*, pp. 2038-2042.
doi:<https://doi.org/10.1109/bigdata50022.2020.9377813>
- Saunders, M., Lewis, P., & Thornhill, A. (2017). *Research methods for business students (8th ed.)*. London, England: Pearson Education.
- Sharda, R. (2017). *Business intelligence: A managerial approach, global edition (4th ed.)*. London, England: Pearson Education.
- Staples, M. Z. (2016). Continuous validation for data analytics systems.
doi:<https://doi.org/10.1145/2889160.2889207>
- Verma, N., & Voids, A. (2016). On Being Actionable: Mythologies of Business Intelligence and Disconnects in Drill Downs.
- Vidgen, R. S. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, 261(2), pp. 626-639.
doi:<https://doi.org/10.1016/j.ejor.2017.02.023>
- Webster, J. &. (2002). Analyzing the past to prepare for the future: writing a literature review. *Management Information Systems Quarterly*, 26(2), 3.
doi:https://web.njit.edu/~egan/Writing_A_Literature_Review.pdf

- Yeoh, W. &. (2015). Critical Success Factors for Business Intelligence Systems. *Journal of Computer Information Systems*, 50(3), 23–32. doi:<https://doi.org/10.1080/08874417.2010.11645404>
- Zhao, J. X. (2022). Reuse of knowledge by efficient data analytics to fix societal challenges. *Information Processing and Management: An International Journal Volume 59 Issue 101 January*, 59(1), 102764. doi:<https://doi.org/10.1016/j.ipm.2021.102764>
- Zhaojing, L., Yeung, S., Zhang, M., Zheng, K., Zhu, L., Chen, G., . . . Ooi, B. (2020). MLCask: Efficient Management of Component Evolution in Collaborative Data Analytics Pipelines. *arXiv (Cornell University)*. doi:<https://doi.org/10.48550/arxiv.2010.10246>

Appendix

A1 CRISP-DM

CRISP-DM is de meest geaccepteerde en gebruikte methodiek voor de ontwikkeling van data analyse modellen (Saltz & Hotz, 2020). CRISP-DM, zie Figuur 8 - CRISP-DM model hieronder, biedt eenvoudig omschreven stappen voor het ontwikkelen van data analyse oplossingen. Solarte (2002) vermeldt een fase in zijn onderzoek die is toegewijd aan de continu support van het product, erg belangrijk voor in stand houden van data analyses (Mariscal, Marbán, & Fernández, 2010). Hierin wordt benadrukt dat in de continu veranderende omgeving het belangrijk is initiële specificaties van analyse modellen te toetsen (Solarte, 2002). Een onderbelicht onderwerp gezien toenemende technologische ontwikkelingen die vandaag de dag aan bod komen (Laudon & Laudon, 2019), en toenemend specialismes (Mariscal, Marbán, & Fernández, 2010)



FIGUUR 8 - CRISP-DM MODEL (SALTZ, 2021)

A2 Literatuuronderzoek referentie

In de aanpak van Chitu Okoli en Kira Schabram (2010) wordt uitgebreid stilgestaan bij een methode hoe systematisch literatuuronderzoek te doen binnen een relevante context: Informatiesystemen. In de wetenschap van informatiesystemen, wordt de noodzaak van een structureel uitgevoerd literatuuronderzoek onderkend (Okoli & Schabram, 2010). Daarnaast is het vakgebied relatief jong, en heeft te kampen met hoge complexiteit en ingewikkelde contexten binnen interdisciplinaire vakgebieden (Webster, 2002). Om deze redenen is het belangrijk om kritisch te kijken naar de uitvoering van systematisch onderzoek, met de beperkt aanwezige bronnen.

Deze methodiek is gekozen aangezien dit goed in de context past van het thema en omgeving die belicht worden in dit onderzoek. Relatief weinig empirisch bewijs uit eerder wetenschappelijk is bruikbaar. Binnen de gekozen onderzoeksmethodiek wordt voornamelijk gezocht naar een algehele rode draad en het realiseren van concepten, eveneens als door James Webster (2002) wordt omschreven, waarbij het realiseren van een concept centraal staat. Een korte toelichting van de fases van Okoli en Webster (2010) (2002) wordt hieronder uitgebreider toegelicht. Ook worden de richtlijnen van Jeffrey W. Knopf gevolgd in de zoektocht naar geschikte en relevante theorieën. Knopf (2006) onderschrijft voornamelijk het belang om theorieën in grotere context te omschrijven en focus te leggen op de rode draad en het geheel van bevindingen, alvorens het trekken van conclusies (Knopf, 2006).

A3 Toepassing onderzoeksmethodiek Okoli

Om structuur te geven aan de uitvoering van het literatuuronderzoek, is hieronder uiteengezet op welke wijze invulling is gegeven aan de methodiek van Okoli. Hierin wordt kort stilgestaan bij de theorie die Okoli (2010) omschrijft en hoe er in dit onderzoek concreet invulling aan is gegeven.

1. Doel van een literatuuronderzoek

Het doel van dit literatuuronderzoek is om de wetenschappelijke lacune die ontstaan is in de context “de houdbaarheid van data analyse modellen” verder te kunnen verkennen. Het literatuuronderzoek kan over het algemeen worden gezien als een algehele beschouwing van de bestaande theorie (Knopf, 2006). Okoli (2010) omschrijft het doel van literatuuronderzoek als een startpunt voor de gekozen richting. Daarnaast moet het literatuuronderzoek leiden tot een synthese van bestaande kennis, die gebruikt kan worden voor het realiseren van een specifiek concept of toepassing (Webster, 2002).

2. Protocol & training literatuuronderzoek omschrijven

Voor het beantwoorden van de centrale vraag: “Welke business enablers bepalen het succesvol in stand houden van data analyse modellen, en hoe lang is de kwaliteit hiervan in stand te houden?”, zijn voornamelijk de volgende thema’s uiteengezet:

- Business enablers (eigenschappen van data analyse modellen)
- Succesvolle toepassingen (omschrijving over de wijze van het product)
- In stand houden (hoe kan een product zo lang mogelijk in stand worden gehouden)
- Data – analyse modellen (producten die voortvloeien uit data analyses zoals rapportages, analyses en gebruik)

Middels deze thema’s worden de onderzoeksresultaten van het onderzoek uiteengezet. Daarmee is het de bedoeling om per thema van ieder onderzochte literatuur een samenvatting te maken, om zo kennis op te bouwen over het algehele onderwerp (Okoli & Schabram, 2010).

3. Selectiefase literatuur omschrijving

De wetenschappelijke artikelen worden gezocht via de wetenschappelijke databanken. Na een brede selectie zal de theorie uiteengezet, samengevat en geëvalueerd worden zoals in de tweede stap hiervoor omschreven (Webster, 2002). Naast de resultaten uit wetenschappelijke artikelen, worden een drietal boeken als basis genomen voor het opdoen van relevante kennis en informatie, en het valideren van de theorieën en kennis daarbuiten:

1. Management Information Systems - Laudon & Laudon (2019)
Theoretisch basisboek bedoeld voor informatiesystemen binnen het hogere onderwijs
2. Dama-dmbok (2^e versie) - Dama International (2015)
Framework over het in stand houden data governance en -management binnen organisaties
3. Business Intelligence, Analytics and Data Science - Ramesh Sharda, Dursun Delen en Efraim Turban (2017)
Management benadering voor toepassing business intelligence binnen organisaties

Gezien de beperkte aanwezigheid van relevante wetenschappelijke bronnen, is het belangrijk een gedegen aanpak te definiëren voorafgaand aan het zoeken. Hiervoor is er een zoekstrategie gedefinieerd, gebaseerd op de zogenoemde Relevance Tree. Een Relevance Tree stelt je in staat op structurele wijze een zoekstrategie te definiëren om bruikbare bronnen te vinden (Saunders, et al., 2017, pp. 93-94). In Figuur 1 is de relevance tree uiteengezet. Er wordt niet alleen gekeken naar

synoniemen, maar juist ook naar antoniemen om het zoekgebied te vergroten. Hiermee wordt de filosofie gevolgd om ruimdenkend en structureel te zoeken naar literatuur (Okoli & Schabram, 2010) waarbij de focus wordt gelegd op het vinden van relevante studies die een rode draad vormen voor het onderwerp van dit onderzoek (Knopf, 2006).

4. Praktische toets omschrijven voor keuze literatuur

Aan de hand van de zoekstrategie, zijn er selectiecriteria opgesteld voor het vinden van relevante artikelen (Okoli & Schabram, 2010). Belangrijk in een opvolgende stap is om de meest betrouwbare bronnen uiteen te zetten (Knopf, 2006). Van belang was om te evalueren of de kennislacunes overeenkomstig waren tussen alle artikelen (Webster, 2002), en de thema's terugkwamen in een overeenkomstige rode draad tussen de artikelen (Knopf, 2006).

Tijdens het zoeken naar relevante onderzoeksartikelen zijn de volgende stappen uitgevoerd:

- I. Selectie op basis van context en omschrijvingen in de titel, zoals omschreven in de Relevance Tree (zie Figuur 1). Hiervoor is de query toegepast zoals omschreven in stap 3. Tijdens deze zoektocht is er alleen gebruik gemaakt van peer-reviewed (wetenschappelijke) artikelen. Hiernaast is er gekeken naar recente artikelen die niet ouder zijn dan 20 jaar. Dit zorgt voor een representatief en actueel resultaat (Knopf, 2006). De gebruikte zoekmachines waren naast de Open Universiteit bibliotheek uit vooraanstaande bronnen gezocht zoals Elsevier Sciencedirect, EBSCO, SpringerLink en GoogleScholar. De eerste selectie resulteerde in circa 50 artikelen.
- II. Na de basiselectie is er gekeken in welke mate het specifieke onderwerp "de houdbaarheid van data analyse modellen" behandeld was in het artikel. Daarbij is gekeken of het onderwerp als hoofd- of subonderwerp behandeld werd. Zodra er maar enkele passages benoemd werden met betrekking tot het onderwerp werden de artikelen geschrapt. Na een selectie op basis van inleiding, conclusie en vervolgonderzoek, en discussie van de artikelen resulteerde dit in circa 30 artikelen.
- III. Van de 30 geselecteerde artikelen zijn er korte samenvattingen gemaakt op basis van het volledige artikel. Aan de hand van deze samenvatting zijn voor- en nadelen benoemd van de artikelen en beoordeeld of deze definitief konden worden gebruikt. Dit resulteerde in een selectie van 12 artikelen.
- IV. Op de uiteindelijke selectie van 12 artikelen is er nog een keer nauwkeurig gekeken naar de artikelen en zijn deze met elkaar vergeleken. Uiteindelijk zijn enkele thema's of hoofdpunten benoemd die de toon voerde binnen de artikelen. Deze 12 artikelen konden aan de hand van deze thema's goed met elkaar vergeleken worden waarbij verschillen en overeenkomsten met elkaar konden worden vergeleken. Van deze 12 artikelen waren 5 artikelen die erg goed in het framework van Dama-dmbok (2017) paste, en 7 artikelen die hier deels of beperkt inpasten op basis van de theorieën en conclusies.

Gebruikte literatuur

Hieronder een opsomming van de 12 gebruikte artikelen voor het literatuuronderzoek, tevens gebruikt voor de thematische analyse:

1. Brodsky, A., & Luo, J. (2015). *Decision Guidance Analytics Language (DGAL) - Toward Reusable Knowledge Base Centric Modeling*
2. Chang, V., Abdel-Basset, M., & Ramachandran, M. (2018). *Towards a Reuse Strategic Decision Pattern Framework – from Theories to Practices*
3. Thomas H. Davenport (2006). *Competing on Analytics*

4. Francesco, C., Demi, S., Magrini, A., Marzi, G., & Papa, A. (2021). *Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation*
5. Kutzias, D., Dukino, C., & Kett, H. (2021). *Towards a Continuous Process Model for Data Science Projects*
6. Lavalley, S., Lesser, E., Schockley, R., & Hopkins, M. (2011). *Big Data, Analytics and the Path From Insights to Value*
7. Ransbotham, S. (2015). *Minding the analytics gap*
8. Saltz, J., & Grady, N. (2017). *The ambiguity of data science team roles and the need for a data science workforce framework*
9. Staples, M. Z. (2016). *Continuous validation for data analytics systems*
10. Yeoh, W. &. (2015). *Critical Success Factors for Business Intelligence Systems*
11. Zhao, J. X. (2022). *Reuse of knowledge by efficient data analytics to fix societal challenges*
12. Zhaojing, L., Yeung, S., Zhang, M., Zheng, K., Zhu, L., Chen, G., Ooi, B. (2020). *MLCask: Efficient Management of Component Evolution in Collaborative Data Analytics Pipelines*

5. Kwalitatieve toets maken

In de kwalitatieve onderzoeken moet voornamelijk gekeken worden naar de wijze van argumentatie die aangehouden wordt (Okoli & Schabram, 2010). Daarbij belangrijk: Wat zijn de lessons learned en overlap met de meeste conclusies van de onderzoeken (Knopf, 2006).

Niet alle studies zijn primair geschikt om kwalitatief te beoordelen. Okoli (2010) benadrukt ook dat het belangrijk is juist naar uitsluitingen binnen het onderzoek te kijken. Tijdens de zoektocht naar het vinden van relevante literatuur is er een filter toegepast in het zoekresultaat op ICT- en IT gerelateerde onderwerpen. Relatief veel studies werden gevonden in het vakgebied van de medische zorg die minder relevant waren. De zoekresultaten namen hierdoor echter wel met zo'n 85% af.

6. Data extractie omschrijven

In de fase van data extractie is voornamelijk gekeken naar de praktische toepassing en extractie van bruikbare informatie uit de gevonden bronnen (Okoli & Schabram, 2010). In deze fase is voornamelijk gekeken naar de artikelen, door een samenvatting te maken van de gevonden informatie. Bij het samenvatten van de artikelen kan men zich het beste richten op een opsomming van de belangrijkste bevindingen (Knopf, 2006). In dit onderzoek is dit praktisch toegepast door te kijken naar de volgende eigenschappen van het onderzoek:

- Wat zijn de belangrijkste focuspunten?
- Welke onder- en overbelichte thema's worden aangehaald?
- Welke eigenschappen van de belangrijkste thema's worden behandeld?

Belangrijkste resultaten van de data extractie zijn ook te vinden in het coderingsschema, zie Figuur 3.

7. Analyse van de bevindingen en synthese

Na de data extractie, is het van belang de resultaten van de studies samen te voegen en te synthetiseren (Webster, 2002). Op basis van deze synthese kan een overstap gemaakt worden naar een conceptuele omschrijving en verdere uitwerking van het onderzoeksonderwerp. Op het einde zal de synthese op toepasbaarheid en bruikbaarheid getoetst worden voor vervolgonderzoek (Okoli & Schabram, 2010). In dit onderzoek is voornamelijk gekeken naar functionele en technische overeenkomsten tussen de verschillende gevonden artikelen, die kunnen helpen een concept te schetsen van het onderzoeksonderwerp.

8. Schrijven van het review

Als laatste stap omschrijft Okoli (2010) het werkelijk schrijven van de literatuur review. Deze stap wordt in de volgende paragraaf verder toegelicht. Belangrijk is dat het onderzoek reproduceerbaar is (Okoli & Schabram, 2010). Om deze redenen zijn de gemaakte stappen in vorige fases nader toegelicht. Het resultaat van de synthese en bevindingen worden hier omschreven.

A4 Toelichting thema's uit literatuuronderzoek

A4.1 Flexibiliteit

Belangrijk uitgangspunt bij het CRISP-DM model is een flexibele aanpak (Saltz, 2021). Flexibiliteit in de aanpak en methodiek bij het realiseren van data analyse modellen zijn een belangrijk aandachtspunt om tot het gewenste resultaat te komen binnen een iteratieve werkwijze (Verma & Volda, 2016). Agile SCRUM is een vaak gebruikte methode voor data analyse modellen (Saltz & Hotz, 2020). Een belangrijke voorwaarde om CRISP-DM in een dergelijke omgeving toe te passen, is door flexibel te zijn in de methodiek en proces binnen het framework. Op die manier kan de diepere betekenis en toegevoegde waarde van het op te leveren product worden ontleend (Saltz, 2021). Tevens wordt het flexibel zijn, aangehaald als een belangrijke voorwaarde om binnen het efficiënt hergebruik van kennis vanuit data analyses toegevoegde waarde te bieden (Zhao, 2022). Ook benadrukken Brodsky & Luo (2015) de mate van flexibiliteit in te bouwen in een database, de bron van informatie. Daarin wordt duidelijk benadrukt dat het hergebruik en schaalbaarheid van de databron belangrijk is voor de reproductie van data analyse modellen (Brodsky & Luo, 2015).

A4.2 Technisch robuuste data pipeline

Met pipeline, of pijplijn, wordt de opgezette infrastructuur bedoeld die is opgezet om frequent een continu stroom van nieuwe data in een datawarehouse beschikbaar te maken voor analyse doeleinden (Zhaojing, et al., 2020). Een datawarehouse vormt het fundament voor de werking van data analyse modellen (Verma & Volda, 2016). Het belang van een robuuste pipeline voor aanvoer van data binnen de databases wordt gesteld als belangrijke voorwaarde om technische werking van data analyse projecten in stand te houden (Saltz, 2021). De infrastructuur moet als een semantische onderlaag dienen voor het ontwikkelen en in stand houden van data analyse modellen (Verma & Volda, 2016). Het is een uitdaging om de data pipeline goed te onderhouden (Zhaojing, et al., 2020). Brodsky & Luo (2015) omschrijven een methodiek voor zogenoemde Decision Guidance Analytics Language (of DGAL). Daarmee wordt het mogelijk beschikbare kennis te hergebruiken en de houdbaarheid van data analyse modellen in essentie te verlengen (Saunders, et al., 2017). Er moet echter wel een trade off worden gedaan tussen de complexiteit van de oplossing en volledigheid en kwaliteit die met het bouwen van een pipeline samengaat (Zhao, 2022).

A4.3 Continue afstemming operatie

Het is belangrijk om continu bij te blijven in de business en blijvend de toegevoegde waarde van data analyse modellen af te stemmen met de business (Lavalle, Lesser, Schockley, & Hopkins, Big Data, Analytics and the Path From Insights to Value, 2011). Meerdere bedrijven passen Business Intelligence en data analytics toe zonder te beseffen dat het belangrijk is een gestructureerde wijze iteratief door te ontwikkelen (Yeoh, 2015). Daarbij wordt benadrukt dat het luisteren van de business

belangrijker is dan het op orde brengen van de techniek binnen de organisatie, om data analyse modellen te kunnen realiseren. In een continu veranderende omgeving is het belangrijk voor organisatie om zich te kunnen aanpassen. Daarbij is het belangrijk zelfs te kunnen snoeien in de bestaande bedrijfsprocessen en methodes (Zhaojing, et al., 2020). Voornamelijk het gebruikersperspectief is belangrijk bij de toepassing en het continu aftasten van de oplossingen die geboden worden (Kutzias, Dukino, & Kett, 2021).

A4.4 Iteratief proces

De volledige cyclus van CRISP-DM moet iteratief doorlopen worden (Chapman, Clinton, & Randy, 2000). Dit kan gezien worden als een sterk punt van het CRISP-DM methodiek, echter is het ook een belangrijk aandachtspunt om rekening mee te houden in de ontwikkeling en het beheer (Saltz, 2021). Voor het leveren van een gedegen oplossing in termen van data analyse modellen, is het belangrijk dat er ook een duidelijk business case wordt aangeleverd waarin toegevoegde waarde wordt onderbouwd (Yeoh, 2015). Deze (voor)waarden voor succes dienen continu te worden doorlopen en met enige regelmaat ook met de business worden doorgelopen in de beheerfase. Op basis hiervan kan beoordeeld worden wat de informatiebehoefte is en waarmee een data analyse model ook valide is om te gaan gebruiken. Daarmee is men beter in staat in te spelen op de ontwikkeling en adoptie van nieuwe technieken (Kutzias, Dukino, & Kett, 2021). De iteratie zal allereerst technisch van aard zijn, waarna deze overgedragen zal moeten worden van de technische specialisten naar de operatie.

A4.5 Taken en verantwoordelijkheden

In grote organisaties wordt onderscheid gemaakt tussen het onderhouden en ontwikkelen van producten. Taken en verantwoordelijkheden tussen het ontwikkelen en beheren van data analyse modellen verschillen hier ook in (Ransbotham, 2015), het vergt een andere discipline (Zhao, 2022). Aangezien er sprake is binnen de informatie wetenschap van een relatief jong begrip (Okoli & Schabram, 2010), is ook de algehele taakomschrijving van data science medewerkers veelal een breed begrip (Saltz & Grady, 2017). Het definiëren van duidelijke rollen en taken moeten juist leiden tot het geven van focus op het verbeteren van de algehele kwaliteit van een dashboard (Vidgen, 2017). Het is dan ook met name belangrijk om de rol en toepassing van betrokken medewerkers binnen het data science gebied continu ter sprake te stellen in de projecten. Dit kan bijvoorbeeld door frequent feedback te vragen door de toepassing op data analyse modellen (Staples, 2016). Daarbij zal het gebruik van zelf-service methodieken er ook aan bij kunnen dragen. Dit kan bijvoorbeeld met het aanwijzen van key-users voor geboden producten omschreven als een belangrijk voorwaarde voor de iteratie (Yeoh, 2015). Op deze wijze kan er een zelfvoorzienende community in stand worden gehouden (Vidgen, 2017).

A4.6 Verantwoorden kwaliteit

Kwaliteit binnen de data analyse is complex om goed te definiëren (Kutzias, Dukino, & Kett, 2021). De behoefte voor een data gedreven organisatie, vloeit voort vanuit de behoefte van bedrijven om te kunnen voorzien in de juiste informatie. Organisaties zijn bereid om wel overwogen beslissingen te nemen op basis van feiten en data en juist niet op het maken van beslissingen gebaseerd op gevoel (Laudon & Laudon, 2019, pp. 410-412). Hier wordt kwaliteit van informatie, nodig voor het maken van een gedegen beslissing, onderscheiden in de volgende kwaliteitsaspecten voor informatie:

- *Nauwkeurigheid*

Heeft de onderliggende dataset de juiste data kwaliteit en is de verwerking hierop voldoende uitgevoerd om betrouwbare en valide data informatie te kunnen verstrekken?

- *Integriteit*

Is de informatiebehoefte vanuit de data analyse modellen oprecht en schaadt deze behoefte niet individuele personen en conform wet- en regelgeving?

- *Consistentie*

Zijn onderliggende berekeningen consistent genoeg en repeterend zodat iedere keer dezelfde resultaten weergegeven worden en deze voldoen aan de levensduur?

- *Volledigheid*

Biedt het resultaat van de data analyse modellen voldoende grond voor de informatie behoefte en wordt het doel van de voorziening behaald?

- *Validiteit*

Zijn gebruikte datasets voor data analyse modellen volledig genoeg om aan gewenste informatie behoefte te voldoen en komt deze overeen met de vereisten van organisatie?

- *Tijdigheid*

In hoeverre is de gebruikte data actueel genoeg om te kunnen voldoen aan de informatiebehoefte van de organisatie en personen?

- *Toegankelijkheid*

Is het juiste type data analyse model gedefinieerd en geeft deze de informatiebehoefte aan de juiste personen weer in een geschikt format?

De maatstaaf voor kwaliteit van de data analyse modellen wordt ook grotendeels bewaakt door het monitoren op de data kwaliteit vanuit de bron van data (Yeoh, 2015). Eenvoudige data is makkelijker om te onderhouden en daarmee kan kwaliteit beter worden gehandhaafd (Ransbotham, 2015).

A5 Interviews

Bij de selectie van de experts, is er voornamelijk gekeken worden naar de basiskennis over het fenomeen en de context van dit onderzoek. Belangrijk daarbij is dat de experts bekend zijn met de probleemstelling en deze herkennen. Er zullen net zo veel experts worden geïnterviewd tot het bereiken van theoretische saturatie. Verwachting is dat er initieel vier experts geïnterviewd zullen worden en iedere keer dit aantal met twee experts uit te breiden totdat er géén nieuwe inzichten meer worden gegenereerd. De tijdsduur van de interviews zullen initieel één tot twee uur in beslag nemen, om voldoende focus op het onderwerp te behouden (Saunders, et al., 2017, p. 465). Indien er aanvullend nog vragen gesteld dien te worden, dan kan er achteraf altijd nog een vervolg afspraak worden gemaakt. Er zal gekozen worden om de expertinterviews initieel te houden bij een door experts vertrouwde omgeving, zoals bijvoorbeeld de eigen organisatie. Eventuele aanvullende interviews kunnen ook online worden afgenomen. De interviews zullen worden getranscribeerd en gecodeerd. Op deze manier zal de validiteit en betrouwbaarheid vergroot worden en kan duidelijk worden gedeeld welke stappen er zijn ondernomen. Van de resultaten zal een database worden bijgehouden om verbanden tussen alle factoren mogelijk te kunnen maken.

Voor het interviewprotocol zal de thematische analyse vanuit het literatuuronderzoek gebruikt worden als belangrijke basis. Gecombineerd met het raamwerk van Dama-dmbok zal dit een belangrijke basis bieden om gevonden bevindingen uit de interviews te coderen. Er wordt daarmee gekozen voor hybrid codering. Middels hybrid coding kunnen er uitgebreide gegevens worden opgedaan en voordelen van zowel kwalitatieve als kwantitatieve methoden toegepast worden (Saunders, et al., 2017, pp. 636-692). De kwalitatieve resultaten vanuit de interviews zullen gecodeerd worden en kwantitatief worden onderzocht. Ook kan hiermee een basis worden gelegd voor het definiëren van het fenomeen en concepten. Belangrijk om bij de interviewvragen ook uit te vragen naar praktische voorbeelden ter voorbereiding op de case studies, zodat ook de realistische en relevante context kan worden bestudeerd.

Bij het afnemen van interviews wordt gestreefd naar het behalen van theoretische saturatie (Saunders, et al., 2017, p. 207). Door circa zes experts te interviewen over het onderwerp is de verwachting meer inzichten te krijgen in het fenomeen en de context. Door het gegeven tijdsbestek voor uitvoering van dit onderzoek zal het aantal experts ook niet toenemen (Saunders, et al., 2017, pp. 39-44) toegepast moeten worden bij de experts om in meerdere fases de experts mee te nemen in de ontwikkeling van de theorie en conclusies. Om zeker te zijn van de gevonden resultaten zal er op het einde van het onderzoek ook middels “slot” interviews decharge gegeven kunnen worden aan de experts. Daarmee kan direct ook gedaan worden aan cross-reference van de gevonden conclusies (Saunders, et al., 2017, pp. 718-722). Dit dient eveneens als input voor de discussie van dit onderzoek.

A5.1 Interviewprotocol

1. Inleiding en toelichting onderzoek

In dit onderdeel zal de expert begroet worden en zal het onderzoek naar de houdbaarheid van data analyse modellen worden toegelicht in enkele minuten. Belangrijk daarbij is het benoemen van het fenomeen en de opgedane kennis uit het literatuuronderzoek. Ook zal de expert op het gemak worden gesteld met betrekking tot anonimiteit en toestemming. De interviews zullen eveneens worden opgenomen zodat deze terug beluisterd kunnen worden.

2. Eigenschappen expert en organisatie

Hierin zullen de volgende eigenschappen van de expert en organisatie worden genoteerd voor het analyseren van de data achter:

- Hoe groot is de organisatie en wat is de core business?
- Binnen welke branche opereert de expert nu en vanuit verleden?
- Hoe veel ervaring heeft de expert en organisatie met data analyse modellen?
- Op welke wijze is betreffend bedrijf georganiseerd qua data & IT?
- Wat is de visie van de organisatie op het gebruik van data analyse modellen?

3. Hoofdvragen vanuit thematische analyse en raamwerk

Hieronder een uiteenzetten van de interviewvragen, op basis van het coderingsschema uit Figuur 3.

Openingsvraag: Hoe zie jij de houdbaarheid van data analyse modellen in relatie tot kwaliteit en houdbaarheid binnen uw professionele ervaring?

CONTINUE AFSTEMMING OPERATIE (MENS)

1. Hoe wordt de afstemming tussen de operationele processen en de data analyse modellen gewaarborgd?
2. Op welke manier wordt er gezorgd voor een voortdurende synchronisatie tussen de operationele gegevens en de analyse modellen?
3. Welke maatregelen worden genomen om ervoor te zorgen dat de modellen actueel blijven en blijven aansluiten bij de veranderende operationele omgeving?
4. Welke uitdagingen komen experts tegen bij het begrijpen en toepassen van geavanceerde data analysetechnieken?

TAKEN EN VERANTWOORDELIJKHEDEN (MENS)

5. Hoe worden de taken en verantwoordelijkheden gedefinieerd en verdeeld binnen het ontwikkel- en onderhoudsproces van de data analyse modellen?
6. Op welke manier worden eventuele overlappingsen of hiaten in taken en verantwoordelijkheden geïdentificeerd en aangepakt?
7. Welke mechanismen zijn er om ervoor te zorgen dat de betrokken stakeholder hun taken en verantwoordelijkheden op een consistente en effectieve manier uitvoeren?
8. Wat zijn de belangrijkste vaardigheden en competenties die nodig zijn voor het ontwikkelen en onderhouden van data analyse modellen?

FLEXIBILITEIT (PROCES)

9. Hoe wordt flexibiliteit in de aanpak en methodologie geïntegreerd om te kunnen omgaan met veranderende eisen en behoeften?
10. Op welke manier kunnen de data analyse modellen gemakkelijk worden aangepast en afgestemd op nieuwe gegevensbronnen of veranderende bedrijfsomstandigheden?
11. Welke maatregelen worden genomen om ervoor te zorgen dat de modellen flexibel genoeg zijn om te kunnen evolueren en aanpassingen te ondergaan?
12. Op welke manier worden risico's en knelpunten geïdentificeerd en aangepakt bij een proceswijziging?

ITERATIEF PROCES (PROCES)

13. Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?
14. Welke methoden of benaderingen worden toegepast om ervoor te zorgen dat er voortdurende feedback en optimalisatie plaatsvindt tijdens het proces?
15. Welke criteria worden gebruikt om te beoordelen of een data analyse model geschikt is voor implementatie en gebruik?
16. Hoe wordt het ontwikkelings- en onderhoudsproces van data analyse modellen gestructureerd en beheerd?

TECHNISCH ROBUUSTE DATA PIPELINE (TECHNIEK)

17. Welke technologische hulpmiddelen en platforms worden gebruikt bij het ontwikkelen en implementeren van data analyse modellen?
18. Op welke manier wordt er gezorgd voor de robuustheid en betrouwbaarheid van de gebruikte technische infrastructuur, zoals de data pipeline?
19. Welke mechanismen zijn er om de kwaliteit, consistentie en integriteit van de data binnen de pipeline te waarborgen?
20. Op welke manier wordt de pipeline onderhouden en geoptimaliseerd om ervoor te zorgen dat deze efficiënt en betrouwbaar blijft functioneren?

VERANTWOORDEN KWALITEIT (TECHNIEK)

21. Hoe wordt de kwaliteit van de data analyse modellen gemeten en beoordeeld?
22. Op welke manier worden de resultaten van de modellen getoetst aan de gestelde kwaliteitsnormen?
23. Welke rapportage- en communicatiemechanismen zijn er om verantwoording af te leggen over de kwaliteit van de modellen aan belanghebbenden?
24. Welke methoden worden toegepast om de prestaties en efficiëntie van de data analyse modellen te monitoren en te optimaliseren?
25. Hoe worden technische uitdagingen, zoals gegevenskwaliteit, schaalbaarheid en integratie, aangepakt in het ontwikkelingsproces?

4. Verdiepende vragen op basis van de antwoorden

Op basis van de beantwoording van de hoofdvragen, zullen er verdiepende vragen worden gesteld. De verdiepende vragen zullen worden geïmproviseerd en zal voornamelijk het accent of tendens van de resultaten moeten beantwoorden. Indien er bijvoorbeeld erg veel accent gelegd wordt op een specifieke ervaring zal hier verder naar worden uitgevraagd.

5. Feedback en discussie

Na het houden van het interview zal de ervaring van de expert worden uitgevraagd. Daarbij gelet op hoe de setting en vragen werden ervaren. Hiernaast zal gevraagd worden naar de mening over de resultaten van het onderzoek, waarbij alvast een voorlopige conclusie kan worden gedeeld om de mening te peilen. Aanvullend wordt ook gevraagd of de geïnterviewden de resultaten wilde verifiëren. Daarnaast is gevraagd om aanvullingen en informatie op gebied van case studies relevant voor dit onderzoek.

6. Afsluiten interview

De resultaten en beantwoording van de vragen zullen kort doorgelopen worden met de expert. Hierna zal besproken worden hoe het vervolg van het onderzoek zal gebeuren en op welke wijze de resultaten gedeeld zullen worden met de expert.

A5.2 Eigenschappen geïnterviewden

Op basis van de afgenomen interviews, zijn de samenvattingen in de volgende paragrafen te vinden. De interviews zijn conform interview protocol afgenomen en gecodeerd. De volgende personen zijn geïnterviewd.

Geïnterviewde	Opleidingsniveau	Omschrijving functie persoon	Organisatie	Specialisatie en ervaring	Aantal jaren ervaring data analyse modellen
1	WO	Manager data & informatie	Aannemer (ca. 1500 FtE)	Register accountant, financiële transacties, voorspellende modellen en rapportages	25+ jaren
2	WO	Directeur data	IT consultancy (ca. 100 FtE)	Rapportages, datawarehousing en consultancy	25+ jaren
3	WO	Data engineer	IT consultancy (ca. 100 FtE)	Onderhouden van databases en maken van diepgaande analyses	20+ jaren
4	WO	Team lead BI	Recruitment consultant (ca. 3500 FtE)	Data analyst, head of insights en team lead business intelligence	5+ jaren
5 Case study	WO	Back-end specialist	Aannemer (ca. 1500 FtE)	Specialist back-end en voormalig front-end	5- jaren

Om te streven naar volledige openheid is in de volgende paragrafen een samenvatting gegeven per interview. Hierbij zijn bepaalde passage naderhand bewerkt om de anonimiteit van de geïnterviewde te waarborgen. De essentie van de interviews is duidelijk te halen uit de resultaten.

A5.2.1 Interview 1

Openingsvraag: Hoe zie jij de houdbaarheid van data analyse modellen in relatie tot kwaliteit en houdbaarheid binnen het bedrijf?

Kwaliteit van een datamodel richt zicht voornamelijk op:

1. Welke drivers zitten er in
2. Als ik model een andere drive geef, zoals een Excel model
3. Eerst het blad
4. Op invoerblad de impact geeft - bijv. scenario's testen - direct feedback op de impact van de belangrijkste 3 KPI's binnen de context waar het data analyse model antwoord op moet geven

Hoe hangen de parameters op elkaar af? - juist simpel houden en naar materiële drivers afpellen.
Markt omstandigheden veranderen – zoals CO2 berekeningen recentelijk – hoe speel je daar op in?

Continue afstemming operatie (mens)

Continu toetsen of realiteit nog in lijn is met model

1. *Hoe wordt de afstemming tussen de operationele processen en de data analyse modellen gewaarborgd?*

Communicatie, operatie niet alleen leveren wat het moet zijn en wat is je beperkende factor

2. *Op welke manier wordt er gezorgd voor een voortdurende synchronisatie tussen de operationele gegevens en de analyse modellen?*

Afstemming is continu door af te stemmen met je data model en te toetsen aan voorspellende waarden van een model - verleden voorspellen o.b.v. toen - train en validatie model. Het model trainen of je omstandigheden.

3. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen actueel blijven en blijven aansluiten bij de veranderende operationele omgeving?*

Blijven rapporteren, zorg voor een specifieke norm c.q. target en bij het behouden van een target zoals bill of material. Context creëren. Voortdurend meten: waarom wijk ik af van de norm? - als je géén afwijking hebt, wat is dan het nut van het model. Voorspellende waarden voor de toekomstige waardes?

4. *Welke praktische uitdagingen komen experts tegen bij het begrijpen en toepassen van geavanceerde data analysetechnieken? - wat geeft je voortschrijdend inzicht?*

Kies een goede tool en ontwerp het eerst. Maak een ontwerp met de key-drivers en het moet zicht lenen om zicht te onderhouden. Als het géén relevant probleem meer is. Gewoon weggooien.

Taken en verantwoordelijkheden (mens)

5. *Hoe worden de taken en verantwoordelijkheden gedefinieerd en verdeeld binnen het ontwikkel- en onderhoudsproces van de data analyse modellen?*

Verantwoordelijkheid van de data modellen moet liggen bij degene die de besluiten definieert. We moet in staan voor het faciliteren van beslissingen nemen.

Zijn die aannames nog wel correct of zijn bijkomende factoren?

6. *Op welke manier worden eventuele overlappingen of hiaten in taken en verantwoordelijkheden geïdentificeerd en aangepakt?*

Degene die het faciliteert zal vragen moeten stellen, maar is niet de uiteindelijke verantwoordelijk. Degene die wenst die bepaalt. Moet input leveren op degene die het gebruikt. Hij zal altijd moet teruggeven op degene die er een besluit op gaat maken.

7. *Welke mechanismen zijn er om ervoor te zorgen dat de betrokken stakeholder hun taken en verantwoordelijkheden op een consistente en effectieve manier uitvoeren?*
 Ja degene die de data modellen of rapportages maken, de gebruikers steeds moeten challengen. De onzekerheid moet er uit door te toetsen of het model daadwerkelijk zo klopt: In hoeverre waarheid of zijn dit 2 waarheden van de medaille. In andere grootheid geformuleerd? - Rapporten maken challengen - hoe komt het dat het zo is.
 Wat is het doel wat je wilt bereiken en is dit nog actueel? - voortdurend de informatie behoefte kunnen toetsen.
 Adviseren richting de toekomst.
8. *Wat zijn de belangrijkste vaardigheden en competenties die nodig zijn voor het ontwikkelen en onderhouden van data analyse modellen?*
 Niet iedereen hebben juiste vaardigheden en competenties. Vooral aan de vraagkant - men vraagt om lijsten: Het zal leuk zijn als we een grafiekje hebben - de vraag achter de vraag wordt onvoldoende toegepast. Beste rapportenbouwers begrijpen heel weinig van de rapporten. Het is dus altijd een samenspel tussen de analyse persoon en business.

Flexibiliteit (proces)

Mogelijk moet zijn om je aan de veranderende omstandigheden aan te passen zoals:

- Productie
- Logistiek
- Etc.

9. *Hoe wordt flexibiliteit in de aanpak en methodologie geïntegreerd om te kunnen omgaan met veranderende eisen en behoeften?*
 Hoe kan je het beste de flexibiliteit aanpakken - grootste probleem is dat men niet weet waarop gestuurd moet worden. Wat bijv. gisteren van belang was is nu niet meer van toepassing. Het gaat meer om de bevraging en toetsing van de informatie behoefte, en in hoeverre heeft dit zijn weerszijde ten opzichte van de business. Begrijpt hij nog de trends of ziet hij zaken die nooit eerder zijn opgevallen.
10. *Op welke manier kunnen de data analyse modellen gemakkelijk worden aangepast en afgestemd op nieuwe gegevensbronnen of veranderende bedrijfsomstandigheden?*
 We maken heel veel denken weinig na. Mega trends of future watchers hebben grote gemeente deler - informatie verzorger moet afvragen - wat is de impact op mijn organisatie o.i.d. hoe moet ik hier mee om gaan in mijn KPI. Toekomstvisie - wat komt er op af? - en hoe gaan we hier mee om? Dit is deels wel af te vangen o.b.v. de gegevensbron - waterverbruik - zorg dat je deze uitleest en trendanalyse kan toepassen.
11. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen flexibel genoeg zijn om te kunnen evolueren en aanpassingen te ondergaan?*
 Bij het design zal je al moeten nadenken dat de omgevingsfactoren veranderen.
 Zoals parameters e.d. aanpassen zoals PxQ etc. - welke KPI's worden steeds belangrijker en waar moet je op letten? - welke bollen zijn nu van glas
12. *Op welke manier worden risico's en knelpunten geïdentificeerd en aangepakt bij een proceswijziging?*
 De manier van ontwerpen: Helder en invul vakjes een afwijkende kleur. Calculated fields moeten een andere kleur hebben.
 Welke getallen heeft mijn input invloed. Meer analyserende modellen - toetsen aan je markt kennis.
 Wat is er op dit moment belangrijk. Goed de krant lezen

Iteratief proces (proces)

13. *Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?*

Bijvoorbeeld duurzaamheid wordt steeds belangrijker, dus heeft weinig nut om alleen te kijken naar bijv. bruto marge - goed nadenken of het in het specifieke model is opgenomen. Op welk van de
Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?

Steeds voortdurend af te vragen. Welke correlaties zie je tussen de factoren die je inzichtelijk wilt maken. Hoe zorg je er voor dat je de targets meetbaar kan laten maken. Hypothese toetsen aan wat je wilt behalen.

14. *Welke methoden of benaderingen worden toegepast om ervoor te zorgen dat er voortdurende feedback en optimalisatie plaatsvindt tijdens het proces?*

1. Niet echt een methode - voortdurend toetsen of het aan de vereisten voldoet
2. In hoeverre is waar we nu naar kijken over een jaar nog relevant is?

Als je disruptief in de markt wilt zijn, dan moet je rigoureuus anders zijn.

15. *Welke criteria worden gebruikt om te beoordelen of een data analyse model geschikt is voor implementatie en gebruik? - kwaliteitskenmerken*

1. Flexibiliteit
2. Betrouwbaarheid - robuustheid
3. Relevant
4. Transparant opgebouwd en simpel van structuur - te begrijpen

16. *Hoe wordt het ontwikkelings- en onderhoudsproces van data analyse modellen gestructureerd en beheerd?*

Belangrijkste verschillen ontwikkelen en onderhouden. Resp. goed idee vs. in hoeverre is informatie behoefte nog juist?

Het is soms beter en sneller om nieuwe modellen te bouwen.

Technisch robuuste data pipeline (techniek)

In hoeverre wordt het model geaccepteerd. De analyse moet gebruiksvriendelijk zijn.

Beschikbaarheid en betrouwbaarheid - in hoeverre is de data beschikbaar en in het juiste format en medium. Alleen PDF printjes kan je bijvoorbeeld niet bij houden. Kwalitatief goed - weinig 0-en, onderhoudbaar en begrijpbaar

17. *Welke technologische hulpmiddelen en platforms worden gebruikt bij het ontwikkelen en implementeren van data analyse modellen?*

Moet goed begrijpbaar en te hanteren zijn. Resultaat moet duidelijke herleidbaar zijn

18. *Op welke manier wordt er gezorgd voor de robuustheid en betrouwbaarheid van de gebruikte technische infrastructuur, zoals de data pipeline?*

Belangrijkste is de basis:

- Keuze voor gedegen platform - fit for purpose, fit for use E
- Het moet onderhoudbaar zijn
- Release matig werken

19. *Welke mechanismen zijn er om de kwaliteit, consistentie en integriteit van de data binnen de pipeline te waarborgen?*

Release matig werken: testen e.d.

Productie check.

Elke change die je maakt testen en documenteren - performance en betrouwbaarheid. Controleren en uitkomsten vastleggen. Verschillen afdoende verklaart en toegepast.

20. *Op welke manier wordt de pipeline onderhouden en geoptimaliseerd om ervoor te zorgen dat deze efficiënt en betrouwbaar blijft functioneren?*

Als het werkt blijf er met je handen van af. Als je toch nog update moet doen zoals security, dezelfde test e.d. uitvoeren.

Bij wijziging in bronsysteem testen met de OTAP straat. Bronsysteem moet men van af blijven in overleg met alle stakeholders.

Verantwoorden kwaliteit (techniek)

Kwaliteit kan gemeten worden als de gebruiker de analyse ervaart als toereikend

21. *Hoe wordt de kwaliteit van de data analyse modellen gemeten en beoordeeld?*

Je kan de beslissing nemen op basis van je model. Fit for purpose

Simpele zaken als gebruiksvriendelijkheid en snelheid etc. - UI e.d. - de klant / gebruiker bepaalt

22. *Op welke manier worden de resultaten van de modellen getoetst aan de gestelde kwaliteitsnormen?*

Belangrijk om vast te houden aan bepaalde kwaliteitsnormen. Of neem een geaccepteerde standaard om er aan te toetsen: Voldoet deze aan een specifieke standaard / ontwikkeling. Zoals toepassing van een specifieke standaard, zoals IBCS (International Business Communication Standards).

23. *Welke rapportage- en communicatiemechanismen zijn er om verantwoording af te leggen over de kwaliteit van de modellen aan belanghebbenden?*

Er is maar één manier om er achter te komen, door in gesprek te gaan met je eindgebruiker/ klant. Wat is nu echt een goede analyse?

De vraag achter de vraag doorvoeren. Waarom zal je naar specifieke uitvoering willen kijken? - het denkproces/ vraagvorming begeleiden.

24. *Welke methoden worden toegepast om de prestaties en efficiëntie van de data analyse modellen te monitoren en te optimaliseren?*

Je wordt alleen maar beter met meten (meten = weten, gissen = missen). Belangrijkste criterium is wordt het model nog gebruikt en kunnen we deze nog toepassen?

Voor jezelf achterhalen veranderd er wat waar je nog mee aan de slag wilt?

Zou je willen weten bij welke beslissingen je het zal gaan gebruiken? - Welke beslissingen voor welke modellen?

Frequentie verversing of besluit is ondergeschikt aan de doelmatigheid. Volgen of de performance goed genoeg is.

25. *Hoe worden technische uitdagingen, zoals gegevenskwaliteit, schaalbaarheid en integratie, aangepakt in het ontwikkelingsproces?*

Dan heb je een waardeloos model. Het moet fit for purpose zijn - het moet je de inzichten geven die je toepast. Beter gaan kijken naar een ontwerp. Bijv. hoe krijg ik alle data van de modellen uit elkaar?

- i.p.v. rechtstreeks uit een systeem. Goede afweging om iets te doen. Niet op basis van milliseconden dus is de tijdigheid relevant + betrouwbaarheid: Bijv. 80% van de data kunnen we gebruiken om de analyse te maken. Gaat er juist om: je kan tijdig zijn maar niet volledig. Waar wordt de analyse voor gebruikt? - Grote stappen snel thuis - belangrijk om meer tijd uit te voeren

Nog iets toe te voegen aan de afsluiting?

Belangrijk is: Ben je nog steeds relevant voor de klant? - iets wat nooit meer veranderd hoeft je niet meer te veranderen. Was een interessant model niet meer.

Uitzonderingen i.p.v. zelfde getal te berekenen.

Bepaalde KPI's zijn altijd van belang: Bijv. QHSE - Veiligheid en gezondheid.

A5.2.1 Interview 2

Openingsvraag: Hoe zie jij de houdbaarheid van data analyse modellen in relatie tot kwaliteit en houdbaarheid binnen het bedrijf?

Tussen gelijk hebben en gelijk krijgen zit een heel groot verschil. Lange tijd in een situatie waarbij data gedreven werken binnen organisaties onbetaalbaar was. Veel meer data gedreven binnen de corporates. MKB+ en andere bedrijven, waren meer houtje - touwtje. V.a. 2012 veel opengegooid voor de non-corporates. Met zelfde mindsets binnen corporates. Data voor iedere bedrijf moet benodigd zijn voor de informatie behoefte. Problematiek moet overal hetzelfde zijn. Hoe zorg je dat techniek en kennisniveau aansluit bij een MKB+. De essentie is bij corporate en MKB hetzelfde, alleen schaalgrote is anders. Op een betaalbare en schaalbare manier moet het bereikbaar en bruikbaar zijn: Betaalbaar, schaalbaar en onderhoudbaar en beheersbaar. Het moet aansluiten op het adoptie vermogen van de organisatie - bijvoorbeeld om Excel los te laten. Cultuur en adoptie vermogen is enorm belangrijk. Vijf fasen van volwassenheid - wat is er gebeurt, en wat is er gebeurt - descriptief is eerste (reactief kijken), daarna meer proactief. Het management en directie zit in laatste niveaus te praten - terwijl operatie dit niet eens kan toelaten. Zit een grote discrepantie waarmee dagelijkse operatie mee te maken heeft. Informatie model moet bekeken worden: Vooral management vs. de operationele laag. Afstemmen hoe informatie model moet worden vorm gegeven.

Continue afstemming operatie (mens)

Data oplossing worden vaak op één plek gecreëerd en vaak gescheiden van een IT. Vaak zijn het onafhankelijke oplossingen en deze moeten tijd van elkaar krijgen. Onderhoud aan de systemen - business kan andere belangen hebben.

1. *Hoe wordt de afstemming tussen de operationele processen en de data analyse modellen gewaarborgd?*

Er moet een multidisciplinair model zijn om dit met elkaar waar te maken. Data is het ultieme product wat je oplevert als een bedrijf. In principe ben je altijd data gedreven. Applicaties om processen te ondersteunen begonnen vaak met het proces uit te tekenen. Data volgt altijd vastleggen van de processen en data volgt laatste deel van de proces samenstelling. Juiste vraag moet zijn: Welke data heb ik nodig om continu af te stemmen - data binnenkomt en wordt getransformeerd. Waar moet ik naartoe en wat heb ik nodig.

Ook kijken naar de opzet: Doel-mens-proces techniek, en doel in termen van data. Volledigheid van data en tijdigheid.

2. *Op welke manier wordt er gezorgd voor een voortdurende synchronisatie tussen de operationele gegevens en de analyse modellen?*

Technische afstemming (ontvangende kant) - veel afstemming en veranderingen in applicaties en processen goed te kunnen managen. Technische kan dit niet reactief maar ook proactief worden ondervangen.

3. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen actueel blijven en blijven aansluiten bij de veranderende operationele omgeving?*

Op het menselijke vlak door te zorgen dat je aansluit bij bijv. wijzigingsystematiek - vooral systeem en de integraties. Ook met elkaar in reguliere overleggen. Ondervangen in de reguliere delen. Voordat data wordt ingeladen een query uit te zoeken of metadata uit te lezen. Informatie voorzieningen die daarvoor blijft te ondervangen. Houdbaarheid staat of valt bij hoe dit er uit ziet. Bij de aanbiedende kant - vooral met: Hoe zorg je ervoor dat snelspoor - borgspoor flexibel genoeg - vs. hoe moet dit beheersbaar blijven.

4. *Welke uitdagingen komen experts tegen bij het begrijpen en toepassen van geavanceerde data analysetechnieken?*

Soms ziet men niet duidelijk wat voor een impact het heeft. Continue afstemming binnen dezelfde kaders blijven en periodiek bekijken: Snijdt dit hout? - Voegt het product nog wat toe?

Taken en verantwoordelijkheden (mens)

Ontvangende en bedienende kant: DDA (Data Delivery Agreement) - leverancier intern en externe manier: Hoe lever je bepaalde structuur. Zoals naamgevingsconventies. Structuren en incrypties. Automatisering erop toepassen. Met de hand ingrijpen of technisch ingrijpen. Maakt het moeilijk om waarde te behalen en toe te passen. Standaardisatie toe te passen.

Richting eindgebruiker: IDA (Informatie Delivery Agreement) - vertellen hoe iets in elkaar zit. Met data catalogus en hoe mag de specifieke data gebruikt worden - zodat géén foutieve conclusies getrokken kunnen worden.

5. *Hoe worden de taken en verantwoordelijkheden gedefinieerd en verdeeld binnen het ontwikkel- en onderhoudsproces van de data analyse modellen?*

Zie bovenstaand: Duidelijke afspraken maken over de informatie levering en data levering.

6. *Op welke manier worden eventuele overlappings of hiaten in taken en verantwoordelijkheden geïdentificeerd en aangepakt?*

De vraag die je moet stellen: Wie is eigenaar van de data en de definitie?

Zijn eigenaren van de data platformen onderscheid van de definities? - wat als data afwijkt? - wat en wie moet dan wat doen?

Is een continu proces: Detecteren, mitigeren en doorpakken. Bijv. datumvelden 100+ jaar in de toekomst, kan een tikfout zijn.

7. *Welke mechanismen zijn er om ervoor te zorgen dat de betrokken stakeholder hun taken en verantwoordelijkheden op een consistente en effectieve manier uitvoeren?*

Business en applicatie onderscheiden

De houdbaarheid valt of staat bij de juiste definities. Je hebt vaste code maar de data is variabel

8. *Wat zijn de belangrijkste vaardigheden en competenties die nodig zijn voor het ontwikkelen en onderhouden van data analyse modellen?*

Bedrijfskundig kunnen kijken (abstractie niveau) boven de materie en data tot op bepaald minimum niveau kunnen begrijpen hoe dit werkt. Een goede business analist bijvoorbeeld.

Er is een nieuwe beweging naar exotische termen: Big data, data science, AI e.d. Het zit heel erg aan de techniek kant. Bij andere analisten zoals black belt en six sigma mensen. Nu is data-analisten: Techniek en/of functionele kant. Is een consultancy rol in plaats van de proceskant. Iemand die begrijpt hoe zaken ontstaan zoals ze bestaan. Het is een bijbel - géén functie.

In de functie omschrijving van een persoon zijn data analyse modellen meegenomen. Trends ontwikkelingen en risico's signalering moet in de functie omschrijving staan. - Niet zelf met Excel werkzaamheden uitpluizen.

Flexibiliteit (proces)

9. *Hoe wordt flexibiliteit in de aanpak en methodologie geïntegreerd om te kunnen omgaan met veranderende eisen en behoeften?*

Hoe regide wil je je data model hebben?

Zijn situaties waarbij data heel geraffineerd is, zoals bij fabrieken. Meetdata die zo stabiel is dat het enorm betrouwbaar is, dus kan hele regide aanpak waarmaken. Bij een ERP systeem is dit weer wat lastiger.

10. *Op welke manier kunnen de data analyse modellen gemakkelijk worden aangepast en afgestemd op nieuwe gegevensbronnen of veranderende bedrijfsomstandigheden?*

Definitie minder strak definiëren en bij de gebruiker terugleggen. Benaderbaarheid zien - as-is meenemen. Inconsistent werken. Flexibiliteit zit hem in: Welk risico aanvaard ik bij de eindgebruiker om dit toe te gaan passen.

11. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen flexibel genoeg zijn om te kunnen evolueren en aanpassingen te ondergaan?*

Afhankelijk van de volwassenheid van de processen. Meetbaar maken en vervolgens een oplossing gaan bedenken aan de hand van je logmonitoring.

12. *Op welke manier worden risico's en knelpunten geïdentificeerd en aangepakt bij een proceswijziging?*

Een veranderende pool van stakeholders - bedrijfsvoering en aanpassingen.

Vooraf zit dit in het stakeholders management.

Techniek kant: Data tot laagste niveau vastleggen. Afspraak op hoger niveau - definities worden toegepast op een iets hoger niveau. Onderscheid tussen dimensies en feiten - detecteren van de uitzonderingen. Gemiddelden nemen en out-layers identificeren voor kadering. In de rol van "leverancier van de business" moet dit ingevuld worden.

Iteratief proces (proces)

13. *Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?*

Interactief = iedere keer beter

Continu = in steady state houden

Er zijn wel eens producten geweest in ervaring die verbeterd moesten worden. Rapport is een goed voorbeeld van hoe het doorontwikkeld moest worden.

Moet het interactief zijn altijd?

14. *Welke methoden of benaderingen worden toegepast om ervoor te zorgen dat er voortdurende feedback en optimalisatie plaatsvindt tijdens het proces?*

PDCA cyclus toepassen en implementeren. Dashboard moet onderdeel zijn van de agenda in de frequente meetings. KPI model en dashboard moet bijvoorbeeld in terugkomen. Maatregelen vastleggen in een register. Maatregelen terugkoppelen aan informatie model - wordt je iteratief beter? - technisch of informatief. Belangrijkste is de business.

15. *Welke criteria worden gebruikt om te beoordelen of een data analyse model geschikt is voor implementatie en gebruik?*

Als in je verbeteracties alleen functionele verbeteringen zitten. Alleen in een steady element klaar houden. Is heel erg op gebied.

Validiteit van je architectuur afstemmen met techniek informatie.

16. *Hoe wordt het ontwikkelings- en onderhoudsproces van data analyse modellen gestructureerd en beheerd?*

Onderscheid zal gemaakt moeten worden tussen technisch en functioneel. Hoe kernprocessen e.d. in beweging zijn.

Je moet meebewegen met de business, bijv. bij de wijziging met nieuwe software.

Technisch robuuste data pipeline (techniek)

17. *Welke technologische hulpmiddelen en platforms worden gebruikt bij het ontwikkelen en implementeren van data analyse modellen?*

Technisch robuuste data pipeline moet eerder zijn technisch robuuste data architectuur.

Er moet iets gedaan worden om de verandering op te vangen.

Data vault is een oplossing. Belangrijk om te begrijpen wat er moet worden veranderd aan de techniek om functioneel in bepaalde staat te houden.

18. *Op welke manier wordt er gezorgd voor de robuustheid en betrouwbaarheid van de gebruikte technische infrastructuur, zoals de data pipeline?*

1. Defensief programmeren.
2. Data model dusdanig bepalen
3. Slimme error handling - Error constateren: Doorpakken en ontwikkelen

Data routeren - detecteren en op specifieke manier op de hoogte houden. Business users meenemen en duidelijke taal afstemmen.

19. *Welke mechanismen zijn er om de kwaliteit, consistentie en integriteit van de data binnen de pipeline te waarborgen?*

Hoe ga je met een foutmelding om? - data stopt. Classificeren error en omschrijven hoe je hier mee om moet gaan.

20% van de gevallen data integriteitsproblemen omgaan. Inzichtelijk maken en bespreekbaar - en accepteren we dit of willen we specifieke oplossing.

Consistent werken met elkaar - als iedereen er van af blijft dat gaat het goed. Binnen data zijn we loodgieter en klusjesmannen van de data analyse modellen.

Iedereen hanteert zijn eigen standaarden, ondanks uitgekristalliseerde standaarden.

De snelheid waarmee technologische ontwikkelingen gaan - nieuwe releases e.d. Nieuwe producten en dergelijke. Data Bricks en fabric - De houdbaarheid van de data oplossing valt op staat bij de ontwikkelsnelheid van je gekozen data platform. Bijvoorbeeld AI - open AI in Azure of Co-pilot in PowerBI

20. *Op welke manier wordt de pipeline onderhouden en geoptimaliseerd om ervoor te zorgen dat deze efficiënt en betrouwbaar blijft functioneren?*

Het moet voor de business zo makkelijk mogelijk worden gemaakt om data definities of correcties toe te kunnen voegen. De middelen waarmee je dit bereikt. Hoe bereik je dit met bijv. met eigenaarschap.

Verantwoorden kwaliteit (techniek)

21. *Hoe wordt de kwaliteit van de data analyse modellen gemeten en beoordeeld?*

Als je iets verantwoord moet je weten waar dat vandaan komt. Bijvoorbeeld een 0-meting:

- Waar stonden we?
- Wat hadden we bedacht?
- Hoe moet het werken?

22. *Op welke manier worden de resultaten van de modellen getoetst aan de gestelde kwaliteitsnormen?*

Data kwaliteit meten - en steeds verbeteren

Data kwaliteitscontroles - technisch:

- White space analysis -
- Duplicaten detectie
- Referentiele integriteit
- Stilstaande data - laatste wijzigingsdatum van een jaar geleden maar staat nog open/ actief
- Verwachte format (ID kolom, uniekheid)
- Automatische profilering - count, distinct, average
- Outliners

23. *Welke rapportage- en communicatiemechanismen zijn er om verantwoording af te leggen over de kwaliteit van de modellen aan belanghebbenden?*

Vertalen van technische lading naar een boodschap die bij hen is omschreven.

Bijv. onderzoeken hoe technische documentatie naar specifieke informatie wordt vertaald. Duidelijk afstemmen.

24. *Welke methoden worden toegepast om de prestaties en efficiëntie van de data analyse modellen te monitoren en te optimaliseren?*

Prestaties monitoren door: Gebruik te monitoren. Maar kan 1x / jaar worden gebruikt maar nog wel iets toevoegen. Moet conform de doel.

Gebruik door juiste stakeholders

Op de agenda staat bij de vaste overleggen. Rapport gemapt in het overleg structuur.

Aantal tickets meten m.b.t. rapport zelf - géén tickets vaak géén goed nieuws

25. *Hoe worden technische uitdagingen, zoals gegevenskwaliteit, schaalbaarheid en integratie, aangepakt in het ontwikkelingsproces?*

Labeling - & data classificatie:

1. Autorisatie aan koppelen - niet aan iedereen beschikbaar stellen
2. Retentie beleid kunnen toepassen - hoe lang mag ik iets vasthouden en gebruiken?

A5.2.1 Interview 3

Openingsvraag: Hoe zie jij de houdbaarheid van data analyse modellen in relatie tot kwaliteit en houdbaarheid binnen het bedrijf?

Het wordt zeker wel herkend. Het is vaak ter sprake gekomen - vooral veel verzamelen en dan maar weer nieuwe gegevens op pakken. Het oude wordt vooral gelaten voor wat het is en opnieuw beginnen. Historie mist en verdwijnt al vrij snel. Bedrijven willen veel gegevens verzamelen en deze direct naar de systemen - dat er uit halen wat je verder wilt.

Data to diamonds - data onder hoge druk zetten en alleen waardevolle gegevens hier uit te halen.

Continue afstemming operatie (mens)

Business heeft snel de nijging om te zeggen doe alles maar bewaren en mee aan de slag. Over 10 jaar lopen ze dan vast. Bij een case organisatie bijvoorbeeld - kan je ook verder met geaggregeerde modellen.

1. *Hoe wordt de afstemming tussen de operationele processen en de data analyse modellen gewaarborgd?*

Een spel wat je zou moeten spelen: Wat heb je nu essentieel nodig voor het draaien van je organisatie. Een goede balans vinden en bepalen wat nu echt nodig is: Kwaliteit van je data.

2. *Op welke manier wordt er gezorgd voor een voortdurende synchronisatie tussen de operationele gegevens en de analyse modellen?*

Dit is een evaluatie proces:

- Consequent
- Goed gemodelleerd?
- Hoeveelheid
- Benaming

Grootste probleem zit meestal in het behoud van de stambestanden, transacties gaan meestal wel goed. Uit eerdere ervaring bij verzekeraars: overgenomen en alles door elkaar heen. 4x van ERP gewisseld --> Alles in 1x datawarehouse geplaatst.

3. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen actueel blijven en blijven aansluiten bij de veranderende operationele omgeving?*

Generiek data model gebruiken. Data model compleet los van het ERP systeem, en mogelijkheid om verschillende systeem technisch gezien van elkaar onafhankelijk vorm te geven. Hier wordt dit ook geprobeerd gedaan.

Bij wisselen van ERP programma moet er ook anders gemodelleerd worden. Bij samenvoegen van bedrijven is dit een complexe uitdaging.

Bijv. eerdere ervaring bij een bank: Daar was het niet mogelijk om snel één keer de transactiekosten in te zien.

4. *Welke uitdagingen komen experts tegen bij het begrijpen en toepassen van geavanceerde data analysetechnieken?*

Vooraf voor winstgevendheid. Algoritmes voeden voor gedragsobservaties. In dit geval meestal meer op gerichte acties. Wat meer onder de pet gehouden door de organisaties.

Vroeger data management bij IT bedrijf - bijv. rekenmodel moet doorgezet worden. Analyse op een maatregel die getroffen is. Met data aanduiden dat bepaalde keuzes gericht analyses gedaan kunnen worden. Zoals:

- Een havenbedrijf: In- en uitvoer
- Bankcrisis: Nieuwe modellen zoals Basel methode

Taken en verantwoordelijkheden (mens)

Wet- en regelgeving. Zodra er een nieuwe regelgeving komt moeten er systemen worden ingevoerd om dit te hanteren.

5. *Hoe worden de taken en verantwoordelijkheden gedefinieerd en verdeeld binnen het ontwikkel- en onderhoudsproces van de data analyse modellen?*
Input van wet- en regelgeving.
2013 - 2016 veel met woekerpolissen --> opdrachten waarin gekeken moest worden: Data vraagstukken waarop rapportages gemaakt moest worden: Hebben de klanten voldoende middelen om pensioen en hypotheek af te betalen. Rekenmodellen maken en toepassen. Communicatie wat er is en hoe het gedaan moet worden. Klanten zoals bij verzekeraars, hadden mensen bij 5 verschillende hypotheek/ pensioen verstrekkers allerlei beleggingsproducten.
Specifieke momenten opgehaald en contactmomenten geregistreerd.
6. *Op welke manier worden eventuele overlappings of hiaten in taken en verantwoordelijkheden geïdentificeerd en aangepakt?*
Data klopt niet vs. ik heb het daar vandaan. Als je de data hebt, ben je er verantwoordelijk voor. Vaak verschil tussen de invoer en analyse van de data.
Eigenaarschap en kwaliteit gaan hand in hand.
7. *Welke mechanismen zijn er om ervoor te zorgen dat de betrokken stakeholder hun taken en verantwoordelijkheden op een consistente en effectieve manier uitvoeren?*
Dit heeft bij bank goed gewerkt: Hele duidelijke lijnen bij portfolio management. Schaalbaarheid speelt daarbij een grote rol. Bedrijf in de reisorganisatie: Klein maar heel goed op orde. Werd gewerkt met portals en koppelde dit op één plek
8. *Wat zijn de belangrijkste vaardigheden en competenties die nodig zijn voor het ontwikkelen en onderhouden van data analyse modellen?*
Techniek en de mens - verwachting goed managen. Verantwoording en eigenaarschap. Ene persoon voelt zich verantwoordelijk voor de gegevens en ander verzint allerlei listen om er omheen te komen. Bijv. bij een verzekeraard had medewerker bedacht om hypotheek tot max €100K aan te vragen - werden 3 hypotheek opgevoerd

Flexibiliteit (proces)

Bijvoorbeeld bij het HavenBedrijf: Goed terug te brengen naar een aantal hoogte punten.

9. *Hoe wordt flexibiliteit in de aanpak en methodologie geïntegreerd om te kunnen omgaan met veranderende eisen en behoeften?*
De oplossing moet juist zijn door basic KPI's te definiëren. Duidelijk en universele manier om je gegevens op te slaan zodat je dit ook voor andere doeleinden kan gebruiken.
10. *Op welke manier kunnen de data analyse modellen gemakkelijk worden aangepast en afgestemd op nieuwe gegevensbronnen of veranderende bedrijfsomstandigheden?*
Voornamelijk vervanging van ERP systemen: Leverancier - Dynamics omgeving die 1 op 1 gekoppeld was aan reporting omgeving. Universeel data model zetten.
11. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen flexibel genoeg zijn om te kunnen evolueren en aanpassingen te ondergaan?*
Data model universeel te houden.
12. *Op welke manier worden risico's en knelpunten geïdentificeerd en aangepakt bij een proceswijziging?*
Bedrijfsprocessen kunnen onderkennen en overzien. Specifieke inrichting van één bedrijf is moeilijk. Een proces benadering. Veel de neiging om alles vanuit techniek te benaderen i.p.v. de organisatie. Vijf verschillende systemen voor medewerker gegevens bijv.

Iteratief proces (proces)

13. *Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?*
Door middel van SCRUM en Agile development. Data aan de slag en tonen. Kort cyclische ontwikkelingen met elkaar delen. Meteen en snel inzicht.
14. *Welke methoden of benaderingen worden toegepast om ervoor te zorgen dat er voortdurende feedback en optimalisatie plaatsvindt tijdens het proces?*
Meteen kunnen testen. OTAP test uit kunnen voeren, schifting maken wat wel en niet nodig is. Korte en gedegen stapjes. Met de modellering gestart - en de terugkoppeling met de business. Meerdere iteraties over lange termijn moet mogelijk worden.
15. *Welke criteria worden gebruikt om te beoordelen of een data analyse model geschikt is voor implementatie en gebruik?*
De business moet kritisch en inhoudelijk wel kennis hebben om hierbij te helpen.
16. *Hoe wordt het ontwikkelings- en onderhoudsproces van data analyse modellen gestructureerd en beheerd?*
Het is lastig als de kloof tussen (data) techniek en business. Bedrijven zoals de banken waren ze continu bezig om data te verbeteren en toe te passen en continue zoektocht naar efficiëntie. Bijv. bij een internationaal alle gegevens generiek op te slaan en toe te passen.

Technisch robuuste data pipeline (techniek)

Een goede en duidelijke ontwikkelstraat hebben. Bij grote organisaties is er wat meer een plan en wordt dit beter getest en toegepast.

17. *Welke technologische hulpmiddelen en platforms worden gebruikt bij het ontwikkelen en implementeren van data analyse modellen?*
Een ontwikkelstraat op alle platformen - bron-systeem vs. DWH en data analyse. Bij verandering moet je snel door. Bijv. door verandering toe te passen. Alles aan elkaar gekoppeld.
18. *Op welke manier wordt er gezorgd voor de robuustheid en betrouwbaarheid van de gebruikte technische infrastructuur, zoals de data pipeline?*
Goed testen en communiceren. Een duidelijke OTAP (Ontwikkel, Test, Acceptatie, Productie) straat.
19. *Welke mechanismen zijn er om de kwaliteit, consistentie en integriteit van de data binnen de pipeline te waarborgen?*
Goed inrichten van CI/ CD proces. Begint bij de mensen en inrichten van alles. Velden gebruiken waarvoor ze bestemd zijn.
20. *Op welke manier wordt de pipeline onderhouden en geoptimaliseerd om ervoor te zorgen dat deze efficiënt en betrouwbaar blijft functioneren?*
Consequent blijven invoeren. En niet gebruikte velden sluiten / dichtzetten.

Verantwoorden kwaliteit (techniek)

Gekeken naar standaard controles zoals:

- Data kwaliteit controles: Consistentie
- Snijd het sowieso hout?
- Dubbelingen
- Hoe vaak komt een specifieke ID voor

21. *Hoe wordt de kwaliteit van de data analyse modellen gemeten en beoordeeld?*
Op basis van je kwaliteitsnormen en fingerspitzengefuhl gevoel. Op basis van ervaringen en eerdere kennis kan men dit beoordelen.
22. *Op welke manier worden de resultaten van de modellen getoetst aan de gestelde kwaliteitsnormen?*
Binnen de organisatie zal er bepaalde kwaliteitsnorm neergelegd worden.
Business of IT neer te leggen? - Gebruik van applicaties.
23. *Welke rapportage- en communicatiemechanismen zijn er om verantwoording af te leggen over de kwaliteit van de modellen aan belanghebbenden?*
Master data tools inzetten om dit toe te passen.
MDM tools zoals specifieke rules en anders gebruiken we het niet.
Handhaven en toepassen.
Inrichting en toepassing. Aangepast op de organisatie van het bedrijf.
24. *Welke methoden worden toegepast om de prestaties en efficiëntie van de data analyse modellen te monitoren en te optimaliseren?*
Kwam echt op de mensen neer.
Normalisering van je data: ene voert vrij tekstveld in, ander alleen keuze voor 5 producten.
25. *Hoe worden technische uitdagingen, zoals gegevenskwaliteit, schaalbaarheid en integratie, aangepakt in het ontwikkelingsproces?*
Heel veel komt nu terug op gebied van AI. Prescriptieve vlak komt het nu terug - voorspelling maken op basis van welke aandelen wel of niet goed gaan. Traders hebben dit gezien en gebruiken.
Bijvoorbeeld met simuleren van algoritmes en door ontwikkelen.

A5.2.1 Interview 4

Openingsvraag: Hoe zie jij de houdbaarheid van data analyse modellen in relatie tot kwaliteit en houdbaarheid binnen het bedrijf?

De wens van de business verandert enorm. Dus de houdbaarheid ligt enorm waarop dit wordt gebruikt. De data die je er in stopt, maar vooral de wens van de markt en wat er tussentijds beschikbaar wordt gesteld.

Continue afstemming operatie (mens)

1. *Hoe wordt de afstemming tussen de operationele processen en de data analyse modellen gewaarborgd?*

Heel belangrijk onderdeel is de uitrol:

- Hoe uitgerold en hoe gecommuniceerd?
- Niet constant contact maar laagdrempelig contact- bijv. met een help knop
- Niet in de situatie komen dat het niet gebruikt wordt

Gebruik met bijvoorbeeld een chatfunctie. Heel duidelijk communiceren met gebruikers en het aantal clicks en views delen.

2. *Op welke manier wordt er gezorgd voor een voortdurende synchronisatie tussen de operationele gegevens en de analyse modellen? - hoe zorg je dat relevant blijft?*
Blijven onderzoeken. Je kan prima een trigger zetten op de verandering en frequentie waarop gekeken wordt, anticiperen met de eindgebruiker of er iets ontbreekt.
3. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen actueel blijven en blijven aansluiten bij de veranderende operationele omgeving?*
Monitoring op gebruik, trigger events.
De markt in de gaten houden en bepaalde trends zien. Onderzoeken hoe je het beter kan maken. Ligt aan het doel en eindgebruiker van het model.
4. *Welke uitdagingen komen experts tegen bij het begrijpen en toepassen van geavanceerde data analysetechnieken? - succesvolle toepassing gebruikt?*
Tool gemaakt voor farmaceuten - activiteiten die artsen gebruiken, verenigingen, journals etc. Ranking maken van de artsen - dan kan er een database gemaakt.
Specifieke manier van ranken van artsen zoals opkomende (rising stars), scores en banen.

Was één relevant onderwerp die niet klopte, dus vooral focus gelegd op de grote geheel waardoor nadruk minder op de details. Hierdoor uitrol vriendelijker en belangrijker voor de adoptie. Designs om bijvoorbeeld eerst met overzicht te beginnen en daarna de details in te duiden.

Taken en verantwoordelijkheden (mens)

5. *Hoe worden de taken en verantwoordelijkheden gedefinieerd en verdeeld binnen het ontwikkel- en onderhoudsproces van de data analyse modellen?*
Taak van het data team moet zijn om het voor de eindgebruiker duidelijk en laagdrempelig toe te passen. Dichter bij de personen houden en kijken naar de user experience. Makkelijker om voor team aan te geven hoe iets wel of niet werkt.
Eindgebruiker moet taak op zich nemen: Zonder klantvraag duidelijk uitspreken en begeleiden.
6. *Op welke manier worden eventuele overlappings of hiaten in taken en verantwoordelijkheden geïdentificeerd en aangepakt?*
Bij het ontwikkelen moet duidelijk al doel worden uiteengezet en uitgekomen. Zodat essentie van het model duidelijk kan worden toegepast en op kan worden teruggevallen.

7. *Welke mechanismen zijn er om ervoor te zorgen dat de betrokken stakeholder hun taken en verantwoordelijkheden op een consistente en effectieve manier uitvoeren?*
Ligt heel erg aan het type bedrijf waar je in zet. Bijvoorbeeld informatiemangers die communicatie verzorgen, en hoe moet je hier mee omgaan.
Klant gerichte dashboard zijn compleet anders en wordt veel specifiek ingevuld. Relevantie verschilt enorm per bedrijf en product.
8. *Wat zijn de belangrijkste vaardigheden en competenties die nodig zijn voor het ontwikkelen en onderhouden van data analyse modellen?*
Ja. Bij ontwikkelen en maken vooral skills bij de product owner vanuit klantperspectief en daardoor later minder houdbaar. Als je niet door ogen kijkt van degene die gebruikt heb je andere softskills nodig voor de houdbaarheid. Dit heeft ook met ervaring te maken van degene die het ontwikkeld dus afhankelijk van type bedrijf, en het model.

Flexibiliteit (proces)

9. *Hoe wordt flexibiliteit in de aanpak en methodologie geïntegreerd om te kunnen omgaan met veranderende eisen en behoeften?*
Vaak te veel flexibiliteit. In het begin vooral té flexibel geweest waardoor de houdbaarheid slecht is geworden.
Hoe meer flexibel, hoe groter kans op niet houdbaarheid. Vooral standaardisatie.
Je kan flexibeler zijn, maar elke stakeholder kan bepaalde voorwaarde meegeven. Minder bijzaken, meer hoofdzaken. Zo weinig mogelijk flexibel.
10. *Op welke manier kunnen de data analyse modellen gemakkelijk worden aangepast en afgestemd op nieuwe gegevensbronnen of veranderende bedrijfsomstandigheden?*
Veranderende omstandigheden niet direct gekoppeld aan flexibiliteit. Kleine aanpassingen maken e.d. kunnen dan wel later worden toegepast. Eerder continuïteit waarborgen. Dus vooral gestandaardiseerde werkwijze bepaalde mate van flexibiliteit.
Gegevensbron veranderd en markt wijzigt, is essentieel.
11. *Welke maatregelen worden genomen om ervoor te zorgen dat de modellen flexibel genoeg zijn om te kunnen evolueren en aanpassingen te ondergaan?*
Vooral gekeken naar de impact. Afhankelijk van de maatregelen die je neemt op basis van de flexibiliteit.
12. *Op welke manier worden risico's en knelpunten geïdentificeerd en aangepakt bij een proceswijziging?*
Doel duidelijk en afgekaderd projectopzet. Zo minimalistisch mogelijk te werk gaan en measures gebruiken. Minimaal aantal test werk, zodat je iets zo eenvoudig mogelijk maken. Bij voorkeur iets meer tijd in ontwikkeling dan extra controles in te bouwen. Zo min mogelijk varianten van de gebruiker en hoe een dashboard is ontwikkeld en de eindgebruiker bij betrekken in de werkwijze.

Iteratief proces (proces)

13. *Hoe wordt het iteratieve proces van ontwikkeling, evaluatie en verbetering van de data analyse modellen gestructureerd en beheerd?*
Duidelijk hebben wat de lineage is: Vooral snel de impact kunnen bepalen en maatregelen treffen. Plan opstellen om de impact en te zorgen hoe dit soepel verloopt. Anticiperen op de verandering via een bepaald proces - opschrijven en vastleggen hoe met de verandering. Dit is niet eenduidig te definiëren. Of wachten tot fout gaat er repareren, maar dat zal niet altijd prettig zijn.

14. *Welke methoden of benaderingen worden toegepast om ervoor te zorgen dat er voortdurende feedback en optimalisatie plaatsvindt tijdens het proces?*

Alles zo genormaliseerd en klein mogelijk houden.

Bijv. bij verschillende plekken neergezet - dat het verkeerd neer kwam - chaos te boven komen: De bron en lineage makkelijk kunnen vinden.

15. *Welke criteria worden gebruikt om te beoordelen of een data analyse model geschikt is voor implementatie en gebruik?*

Vanuit deployment gaat vooral om: Wie heeft waar toegang toe, wat zijn de voordelen, hoe bereikbaar is het en normaliseren van de tabellen zo veel mogelijk weg wordt gehaald bij de gebruiker.

Discussie niet op de inhoud maar op doel een essentie van het model.

16. *Hoe wordt het ontwikkelings- en onderhoudsproces van data analyse modellen gestructureerd en beheerd?*

Zie hierboven: Prima developen, aan het eind van product kijken of het op meest efficiënte manier is gebeurd. Juiste measures en robuust mogelijk.

Technisch robuuste data pipeline (techniek)

17. *Welke technologische hulpmiddelen en platforms worden gebruikt bij het ontwikkelen en implementeren van data analyse modellen?*

Als je iets bouwt moet je altijd rekening houden met de invoerproces: Zorg voor validatie van de data Duidelijk begrip zijn waar de data vandaan komt en de risico's inzichtelijk gemaakt.

18. *Op welke manier wordt er gezorgd voor de robuustheid en betrouwbaarheid van de gebruikte technische infrastructuur, zoals de data pipeline?*

Bepaalde scenario's bij het invoer proces beter kunnen begrijpen. Dit ligt bij de product owner: Wat zijn de processen waarop we dit gaan bouwen en wat betekent dit voor de klant. Bijvoorbeeld per ongeluk wijziging kan leiden tot niet functioneren van het model.

19. *Welke mechanismen zijn er om de kwaliteit, consistentie en integriteit van de data binnen de pipeline te waarborgen?*

In principe check je altijd op missing values, outliers en dergelijke. Heel veel systemen kan je preparatie variant waar je de data kwaliteit. Dit zal in je ETL gebouwd moeten worden op gestandaardiseerde manier zoals data types e.d.

Wat laat je precies zien?

Duidelijk met gebruiker de consequenties doorlopen als bijvoorbeeld data kwaliteit slecht is e.d.

20. *Op welke manier wordt de pipeline onderhouden en geoptimaliseerd om ervoor te zorgen dat deze efficiënt en betrouwbaar blijft functioneren?*

Methoden altijd overleggen maar niet tot in hoog detail bespreken met de gebruiker.

Verantwoorden kwaliteit (techniek)

21. *Hoe wordt de kwaliteit van de data analyse modellen gemeten en beoordeeld?*

Je moet je afvragen: Klopt qua data is dit mogelijk?

Als het niet relevant is, hoe ga je dit dan doen?

Dit zal met bepaald proces moeten worden afgevangen. 1x in x aantal maanden bijvoorbeeld de relevantie van alle modellen en measures doorlopen. Een manier vinden waarop de categorieën die je gebruikt voor je data nog toepasbaar is: Is het nog relevant?

22. *Op welke manier worden de resultaten van de modellen getoetst aan de gestelde kwaliteitsnormen?*

Wat verwacht je van een gebruik in het dashboard? - norm zetten en hoe veranderd dat op den duur

Als gebruik minder wordt, maar vooral kijken naar de eigenschappen van het dashboard en welke vragen moet deze beantwoorden?

Kwaliteit van de data analyse modellen zijn te meten op basis van: Hoe past de eindgebruiker het toe voor zijn of haar gebruik?

23. *Welke rapportage- en communicatiemechanismen zijn er om verantwoording af te leggen over de kwaliteit van de modellen aan belanghebbenden?*

Automatisch een survey bijvoorbeeld uitsturen. De productowner zal dan frequent door moeten lopen of hetgeen dat gebruikt moet worden nog relevant is of niet. Zo laagdrempelig mogelijke communicatie.

Kiezen voor een communicatie model die niet mens intensief is: Zo makkelijk mogelijk. Een link met vraag te stellen over bijvoorbeeld bepaald model zoals gebruik van een chatbot. A-synchrone communicatie, niet direct op het persoon.

24. *Welke methoden worden toegepast om de prestaties en efficiëntie van de data analyse modellen te monitoren en te optimaliseren?*

Zie hierboven, eerder benoemd:

- Monitoren op gebruik
- Communicatie en
- Bij goede uitrol zal dit aanzienlijk schelen in adopties
- Toegankelijkheid bepalen
- Awareness vergroten

25. *Hoe worden technische uitdagingen, zoals gegevenskwaliteit, schaalbaarheid en integratie, aangepakt in het ontwikkelingsproces?*

Schaalbaarheid is één van de belangrijkste punten: Vooral bijdragen aan de doelstelling.

Bij voorkeur één dashboard die bijdraagt dan allerlei bijzaken - zo minimalistisch en eenvoudig mogelijk. Documentatie goed houden en rekening houden met dat iemand anders door kan ontwikkelen.

Nog iets toe te voegen aan de afsluiting?

Heel belangrijk onderwerp: Deployment is al lastig genoeg. Data analisten aannemen is niet alles, het is een product beheren en het is dus belangrijk om bijvoorbeeld vertrouwen op te bouwen. Als te vaak men is teleurgesteld is het lastig houdbaar om toe te passen.

A5.3 Overzicht gebruikte coderingen

Hieronder staat weergegeven hoe de coderingen gemaakt zijn voor de interviews die zijn afgenomen. Zie ook 2.5.3 Coderingschema. De samengevatte interviews zijn allereerst gecodeerd volgens de initiële codes, zoals hieronder weergegeven. En daarna zijn de specifieke codes toegepast.

- Beperkende factor duidelijk en inzichtelijk
- Business verantwoordelijk voor juiste gebruik
- Continu bijstellen targets door veranderende omgeving
- Continu communiceren met gebruikers
- Continu verbeteren
- Continue afstemming met eigenaren
- Correctheid en volledigheid
- Duidelijk onderhoudsproces
- Eenvoudig in onderhoud
- Essentie model toepassen
- Gebruikers stimuleren en triggeren
- Goede opmerking
- Indien niet relevant dan archiveren
- Keep it simpel
- Scenario verwerking
- Toepassen maatstaaf voor kwaliteit
- Veranderende omstandigheden

Kernbegrip (raamwerk)	Dimensie (thema)	Indicator - kort	Bron – verdiepende theorie
<i>Mens</i>	Taken en verantwoordelijkheden	Duidelijk rol medewerker beheer	Saltz, J. S., & Grady, N. W. 2017
	Taken en verantwoordelijkheden	Duidelijke proces taken en verantwoordelijkheden	Davenport, T. H. 2006
	Taken en verantwoordelijkheden	Gebruikerstoepassingen duidelijk toegepast	Mark Staples, Liming Zhu, John Grundy 2016
	Taken en verantwoordelijkheden	Versimpelde manier van weergeven	Ciampi, F., Demi, S., Magrini, A., Marzi, G., & Papa, A. 2021
	Continue afstemming operatie	Sturen op data kwaliteit en impact	Yeoh, W., & Koronios, A. 2015
	Continue afstemming operatie	Gebruikersperspectief duidelijk	Kutzias, D., Dukino, C., & Kett, H. 2021
	Continue afstemming operatie	Actuele informatie behoefte afgestemd	Luo, Z., Yeung, S. C. J., Zhang, M., Zheng, K., Zhu, L., Chen, G., Fan, F., Lin, Q., Ngiam, K. Y., & Ooi, B. C. 2020
	Continue afstemming operatie	Gebruik wordt gecontroleerd	Mark Staples, Liming Zhu, John Grundy 2016
	Continue afstemming operatie	Business geeft feedback	Mark Staples, Liming Zhu, John Grundy 2016
	Continue afstemming operatie	Regelmatige afstemming data (analyse) eigenaren	Victor Chang & Mohamed Abdel-Basset, Muthu Ramachandran 2018
	Continue afstemming operatie	Afstemming doorontwikkeling	Jianyu Zhao, Xi Xi, Lulu Zhang, Ching-Hsien Hsu, Priyan Malarvizhi Kumar 2022
	Continue afstemming operatie	Verbeteren met nieuwe technieken	Ciampi, F., Demi, S., Magrini, A., Marzi, G., & Papa, A. 2023
<i>Proces</i>	Felxibiliteit	Hergebruik eerdere werkzaamheden	Brodsky, A., & Luo, J. 2015
	Felxibiliteit	Regelematige afstemming beslissers en data (analyse) eigenaren	Victor Chang & Mohamed Abdel-Basset, Muthu Ramachandran 2018
	Felxibiliteit	Toepassen Efficiënte Data Analytics	Jianyu Zhao, Xi Xi, Lulu Zhang, Ching-Hsien Hsu, Priyan Malarvizhi Kumar 2022
	Iteratief proces	Continu verbeteren data analyse modellen	Yeoh, W., & Koronios, A. 2015
	Iteratief proces	Proces continu verbeterd	Kutzias, D., Dukino, C., & Kett, H. 2021
<i>Techniek</i>	Technisch robuuste data pipeline	Data lineage duidelijk inzichtelijk	Carlie Idoine, Peter Krensky, Erick Brethenoux, Jim Hare, Svetlana Sicular, Shubhangi Vashisth 2018
	Technisch robuuste data pipeline	Meest optimale data pipeline wordt geambieerd	Luo, Z., Yeung, S. C. J., Zhang, M., Zheng, K., Zhu, L., Chen, G., Fan, F., Lin, Q., Ngiam, K. Y., & Ooi, B. C. 2020
	Technisch robuuste data pipeline	Gebruik eenvoudige codes	Brodsky, A., & Luo, J. 2015
	Verantwoorden kwaliteit	Consistentie gebruik codes	Victor Chang & Mohamed Abdel-Basset, Muthu Ramachandran 2018
	Verantwoorden kwaliteit	Continue optimalisatie business en techniek	William Yeoh & Andy Koronios 2015
	Verantwoorden kwaliteit	Integriteit getest en beoordeeld	William Yeoh & Andy Koronios 2015
	Verantwoorden kwaliteit	Continue verbinding techniek en business	Yeoh, W., & Koronios, A. 2015
	Verantwoorden kwaliteit	Betrouwbaarheid en validiteit gewaarborgd	Chang 2018
	Verantwoorden kwaliteit	Toegankelijkheid ingeregeld in organisatie	Ransbotham 2015
	Verantwoorden kwaliteit	Duidelijke inzichten helpen bij besluitvorming	Ransbotham, S., Kiron, D., & Prentice, P. K. 2015
Verantwoorden kwaliteit	Kwaliteit van de oplossing afgewogen tegenover complexiteit	Luo, Z., Yeung, S. C. J., Zhang, M., Zheng, K., Zhu, L., Chen, G., Fan, F., Lin, Q., Ngiam, K. Y., & Ooi, B. C. 2020	

FIGUUR 9 - GEBRUIKTE CODES VOOR CODEREN INTERVIEWS

A6 Case studies

A6.0.1 Casestudy protocol

Voor dit onderzoek is er gekozen om multiple embedded cases toe te passen. Er is gekozen voor dit type case-onderzoek om het fenomeen duidelijk te kunnen onderzoeken op basis van de beschikbare resources. In dit onderzoek wordt er niet veel diepgang gecreëerd, maar worden de case vooral onderzocht in het gedrag op het fenomeen, met als doel om bij te dragen aan de onderzoeksdoelstelling. Gezien de beperkte tijd en middelen is dit de beste uitkomst om cases te toetsen. De uitvraag naar specifieke eigenschappen van de cases is gebeurt aan de hand van een aanvullend interview dat is afgenomen bij een extra expert. Hierbij is gevraagd om een korte samenvatting te geven waar het in theoretisch kader omschreven fenomeen zich heeft voorgedaan bij de case organisatie.

De cases zijn geselecteerd aan de hand van een vooraf gesteld format waarbij gezorgd wordt dat alle benodigde gegevens bevat. De vier cases zullen getoetst worden aan de volgende eigenschappen (of embedded units), die aaneensluiten met de thematische analyse:

- **Menselijke aspect**
Focus op de rollen, vaardigheden en interacties tussen de stakeholders, gebruikers en ontwikkelaars. Vooral gekeken naar bruikbaarheid en effectiviteit.
- **Procesgericht aspecten**
Analyseren van ontwikkeling- en onderhoudsprocessen van de modellen. Zijn er bijvoorbeeld bepaalde interacties gebruikt, governance structuren of live cycles waar men rekening houdt.
- **Technische aspecten**
Hoe zijn de technologische infrastructuren opgebouwd voor het behoud van kwaliteit en levenshoudbaarheid. Omschrijving van effectieve maatregelen die zijn getroffen.
- **Cross-functionele aspecten**
Richt zich op vlak van communicatie en samenwerking tussen diverse stakeholders. Vooral gekeken naar wat de geleverde bijdrage is en welke compromissen worden gesloten.

In deze cases zullen de volgende eigenschappen verder worden omschreven:

- Algemene omschrijving
- Succes van het model
- Embedded units en verbanden
- Theoretische en praktische bijdrage onderzoek

De omschrijven van de cases worden hieronder verder uiteengezet. Aan het einde volgt er nog een algehele samenvatting, analyse en beschouwing van de resultaten.

A6.0.2 Geselecteerde cases

Er zijn in totaal vier cases geselecteerd, hieronder de omschrijving:

1. Huisvestingsanalyse – tijdelijke analyse, veranderende units
2. Planningsanalyse – bepalen vulgraad planning voor nieuwe aan te nemen werken
3. QHSE dashboard – rapportage controle op nalevering veiligheid checklisten
4. Productie- en contract monitoring model – inzicht in naleving contractafspraken en productie

A6.1 Case 1 | Huisvesting analyse

Huisvesting is een belangrijke opgave voor de case-organisatie vanwege de behoefte aan landelijke dekking van haar diensten. Strategisch is het essentieel om elk kwartaal de kosten en verdeling van de vestigingen te evalueren in relatie tot personeelscapaciteit en -bezetting. Voorheen was de personeelsbezetting op de vestigingen onduidelijk, wat het maken van strategische beslissingen belemmerde. Het ontwikkelde model biedt gedetailleerd inzicht in de personeelssamenstelling en -dynamiek, wat cruciaal is voor het effectief beheren en plannen van huisvestingsbehoeften, vooral in het licht van diverse bouwprojecten met grote of variërende personeelsbezetting. Zie ter illustratie Figuur 10 - Overzichtsk kaart met daarin gehuisveste medewerkers, project- en kantoorlocaties.

Succes model

Het succes van het model is te danken aan verschillende factoren. Het biedt een duidelijk en eenvoudig inzicht, maakt gebruik van eenvoudige berekeningen en maakt het mogelijk om relatief overzichtelijk te zien wat er gebeurt. Dit alles faciliteert het nemen van duidelijke beslissingen en stelt de organisatie in staat om verschillende scenario's gemakkelijk te toetsen. Belangrijkste onderschatte punten zijn de dimensie en/of stamgegevens die men continu moet bijwerken. De personeelsbezetting wordt bijgehouden aan de hand van apparaten die aan personen zijn gekoppeld, maar de administratie bleek niet altijd up-to-date. Het is een uitdaging om berekeningen te maken voor personeel dat op één dag op meerdere vestigingen werkt. Bovendien is het een eenvoudige berekening die alleen rekening houdt met personeel dat fysiek aanwezig is; personeel dat niet inlogt op de wifi van de vestiging wordt niet meegenomen in de analyse. Tot slot, een grote update van de databronsystemen heeft het dashboard ernstig verstoord, ondanks eerder gemaakte afspraken. Dit leidde tot een periode waarin de analyse niet bruikbaar was. Het model waardevolle inzichten voor de organisatie, maar het vereist voortdurende aandacht en aanpassing om effectief te blijven in het ondersteunen van huisvestingsbeslissingen.

Embedded units & verbanden

Menselijke aspect

Bij de ontwikkeling van dit model is zeer integer gehandeld, vooral omdat de monitoring van de personeelsbezetting tot argwaan kan leiden. Er is effectief samengewerkt met verschillende disciplines, waaronder IT, Business Intelligence, huisvesting en de directie. Deze samenwerking heeft uiteindelijk zijn vruchten afgeworpen, wat bijdraagt aan de betrouwbaarheid en acceptatie van het model binnen de organisatie.

Procesgericht aspecten

Er zijn significante uitdagingen ervaren bij het actueel houden van de context en de data. Ondanks de gemaakte afspraken en de uitdagingen met frequentie van updates, bleek het zeer waardevol om gezamenlijk en regelmatig, specifiek per kwartaal, het model te evalueren. Dit bevorderde de verdere ontwikkeling en fijnafstemming van het model.

Technische aspecten

De technische implementatie omvatte een beperkt aantal alerts, wat soms resulteerde in vertraagde detectie van fouten in het model. Handmatige controles van de data pipelines waren effectief om grote fouten te voorkomen. Door het model eenvoudig en overzichtelijk te houden, was het onderhoud efficiënt en beheersbaar.

Cross-functionele aspecten

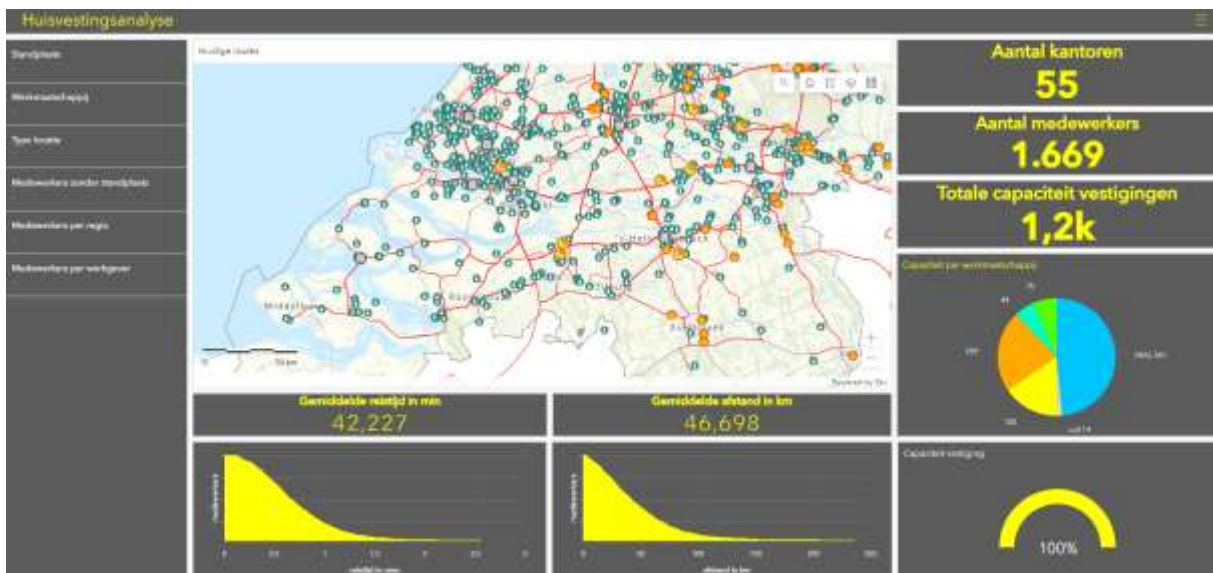
Het is duidelijk geworden dat technische ondersteuning essentieel is. Goede afspraken vormen de basis voor het correct onderhouden van het data analyse model. Belangrijk is ook de selectie van de juiste personen die betrokken zijn bij de ontwikkeling en het onderhoud van het model. De inclusie van diverse perspectieven en expertise bevordert niet alleen de functionaliteit van het model, maar ook de brede acceptatie binnen de organisatie.

Theoretische en praktische bijdrage onderzoek

De belangrijkste lessen uit het onderzoek naar de houdbaarheid van data analyse modellen in de huisvestingscase zijn:

1. Flexibiliteit en aanpasbaarheid zijn essentieel vanwege de veranderlijke bedrijfsomgeving
2. De nauwkeurigheid en actualiteit van data zijn cruciaal voor betrouwbare modelresultaten
3. Een holistische benadering die menselijke, proces- en technologische aspecten integreert, is noodzakelijk voor succes
4. Het model moet uitzonderingen en anomalieën kunnen verwerken
5. Robuustheid tegen systeemupdates is belangrijk.
6. Gebruiksvriendelijkheid en toegankelijkheid zijn essentieel voor effectieve besluitvorming.
7. De lange termijn houdbaarheid vereist continue evaluatie en aanpassing van het model.

Deze inzichten zijn fundamenteel voor het waarborgen van de effectiviteit en duurzaamheid van data analyse modellen. Belangrijk is dat het maken van duidelijke afspraken en deze in de continuïteit van het werken te waarborgen al een flinke uitdaging blijkt.



FIGUUR 10 - OVERZICHTSKAART MET DAARIN GEHUISVESTE MEDEWERKERS, PROJECT- EN KANTOORLOCATIES

A6.2 – Case 2 | Planningsanalyse

De case organisatie opereert landelijk met regionale in te zetten ploegen om projecten buiten uit te voeren. Daarbij is het van belang dat overzicht behouden wordt en een voorlopige planning in het systeem wordt bijhouden voor verdeling van de werken en ervoor te zorgen dat de ploegen continuïteit behouden in de werkzaamheden. De organisatie heeft daarbij belang om zo veel mogelijk met eigen personeel te werken en waar nodig tijdig eventuele subcontractors in te schakelen. De doorlooptijd van de projecten is erg variërend, waardoor het aantal uit te voeren werk in tonnen wordt uitgedrukt en uitgespreid wordt over de beschikbare ploegen. Hiermee kan ook gelijk rekening gehouden worden bij nieuwe aanbestedingen. Ook zit er enkele correcties die worden berekend om de werkelijke uitvoer van de planning wordt behaalt. Zie Figuur 11 - Gebruikte visualisatie in de planningsanalyse ter illustratie.

Succes van het model

Het succes van het data analyse model in het asfaltprojectbeheer is duidelijk merkbaar in verschillende aspecten. Een van de meest significante voordelen is dat er geen data van verschillende vestigingen opgehaald hoeft te worden. In plaats daarvan zorgt een reeks van live bijgewerkte ETL-processen voor een dynamische en actuele datastroom. Dit heeft geresulteerd in een toegenomen vertrouwen tussen de verschillende regio's, wat weer leidt tot meer communicatie onderling. Het proces dat is ingezet om te komen tot zuivere data, heeft een sleutelrol gespeeld in de evolutie naar een professioneel, data-gedreven bedrijf.

Dit model heeft ook gezorgd voor een sterkere discipline in de regionale administraties. Een accurate administratie werd essentieel, omdat zonder deze discipline geen helder inzicht mogelijk was in de bedrijfsvoering. Echter, het project kende ook belangrijke leerpunten. Zo bleek het bijhouden van seizoenspatronen een complexe taak en werden wijzigingen of verschoven werken onvoldoende bijgehouden. Het ontbreken van een centrale 'spelbewaker' of eigenaar die de administratie consequent bijhield, was een knelpunt.

Financiële aannames die op basis van dit model werden gedaan, droegen een hoger risico dan aanvankelijk gedacht, wat leidde tot te optimistische prognoses en aannames. Bovendien berustten veel berekeningen op aannames. Tot slot was de communicatie en instructie over de werking van het model beperkt, wat de effectiviteit en het begrip van het systeem onder de gebruikers beïnvloedde.

Embedded units & verbanden

Menselijke Dimensie

De invoering van het model heeft positief bijgedragen aan het vertrouwen en de communicatie tussen verschillende regio's. Door de noodzaak van accurate gegevensinvoer, werden medewerkers aangespoord tot nauwkeuriger werk. Echter, het project ondervond uitdagingen in het ontbreken van een centrale verantwoordelijke (of 'spelbewaker'), wat leidde tot inconsistenties in de administratie. De beperkte communicatie en instructie over de werking van het model benadrukt de noodzaak voor meer betrokkenheid en training van de betrokken personen.

Procesgerichte Dimensie

Het model bevorderde de professionalisering van het bedrijf naar een data-gedreven aanpak. De automatisering van dataverzameling via ETL-processen en de focus op zuivere data hebben de efficiëntie van de processen verhoogd. Toch bleken er uitdagingen te zijn in het bijhouden van

seizoenspatronen en wijzigingen in de planning, wat wijst op ruimte voor verbetering in het procesmanagement.

Technische Robuustheid

Technisch gezien zorgde de live data-updating voor een constante stroom van actuele informatie, wat een sterke technische basis toont. De afhankelijkheid van het model op aannames en de daaraan verbonden risico's in financiële prognoses wijzen echter op een behoefte aan verbetering in de algoritmische nauwkeurigheid en validiteit van de gebruikte modellen. De uitdaging ligt in het versterken van de technische robuustheid om meer betrouwbare en risicobewuste output te genereren.

Cross-functionele aspecten

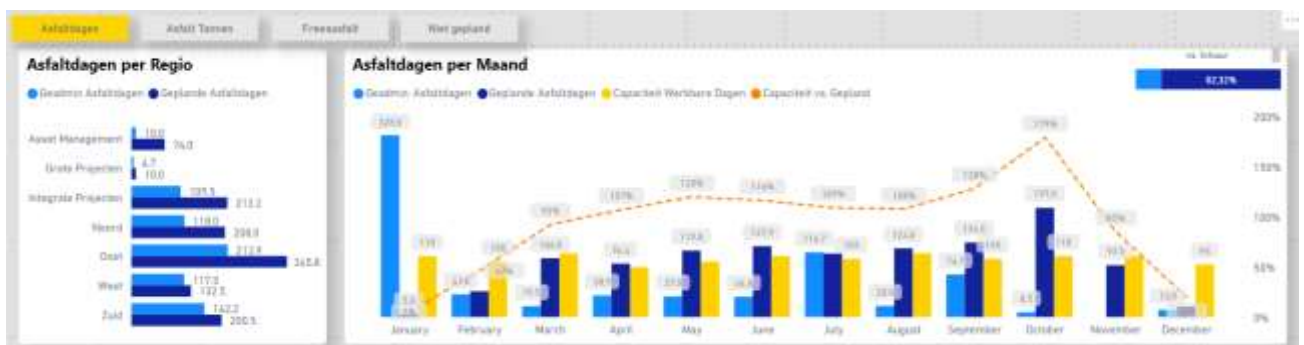
In het algemeen illustreert deze analyse dat het model sterke punten heeft in elk van de drie dimensies, maar ook dat er ruimte is voor verbetering, met name in het versterken van de menselijke betrokkenheid, het verfijnen van processen, en het verhogen van de technische nauwkeurigheid en betrouwbaarheid.

Theoretische en praktische bijdrage onderzoek

De belangrijkste lessen uit het onderzoek naar de houdbaarheid van data analyse modellen in de huisvestingscase zijn:

1. Belang van duidelijke communicatie en betrokkenheid van alle stakeholders.
2. Noodzaak van duidelijke rolverdeling en verantwoordelijkheden binnen projecten.
3. Het belang van een flexibele en iteratieve benadering in procesmanagement.
4. De cruciale rol van technische infrastructuur en data-integriteit.
5. Continu monitoren en waarborgen van de kwaliteit van data en modellen.
6. Bewustzijn van de levenscyclus en houdbaarheid van data analyse modellen.
7. Aanpassingsvermogen van modellen aan veranderende omstandigheden en inzichten.
8. Integratie van kwaliteitscontroles en risicobeheer in het ontwikkelingsproces van modellen.

Deze punten benadrukken het belang van een holistische benadering, waarbij zowel de menselijke, procesmatige als technische aspecten gelijktijdig worden aangepakt voor effectieve data analyse en modelbeheer.



FIGUUR 11 - GEBRUIKTE VISUALISATIE IN DE PLANNINGSANALYSE

A6.3 – Case 3 | QHSE dashboard

QHSE staat voor Quality, Healty, Safety en Environment. Veiligheid gaat binnen de case organisatie boven alles, en is in gros deel van de werk overleggen het eerste punt op de agenda. Om deze reden worden er frequent in diverse systemen de frequente en verplichte veiligheidsvoorschriften gemonitord. De buitendienst medewerkers en diens directe collega's moeten verplicht de meest recente veiligheidsvoorschriften van uit de organisatie en branche hebben gelezen. Hiervoor dient er overzicht te komen wat met gedaan heeft in het QHSE-dashboard. Het QHSE-dashboard is een innovatieve rapportage, ontwikkeld om de stand van deze veiligheidsvoorschriften binnen de case organisatie te monitoren. Sinds de introductie heeft dit significante impact gehad. Het combineert real-time data analyse met gebruiksvriendelijk interface om de prestaties weer te geven. De ontwikkeling en implementatie van het dashboard zijn in nauwe samenwerking tussen de maker, beheerder en de eindgebruikers tot stand gekomen. Zie ter illustratie Figuur 12 - Gebruikte visualisatie in het QHSE dashboard.

Succes van het model

Het QHSE-dashboard heeft zijn waarde bewezen door een toename in het gebruik en de positieve feedback van de gebruikers. Het succes is vooral merkbaar in hoe het dashboard heeft bijgedragen aan verbeterde besluitvormingsprocessen en een verhoogde transparantie in prestatiebeoordelingen binnen de veiligheid. De regelmatige updates en verbeteringen tonen de voortdurende inzet om het dashboard af te stemmen op de evoluerende behoeften van de organisatie. Belangrijkste leerpunten zijn dat het dashboard, is dat door het succes het belangrijk was het dashboard voldoende in stand te houden. Daarbij zijn onvoorziene zaken vaak aan de orde van de dag, waardoor het waarborgen van stabiliteit constant besproken moet worden. Dit vergt veel afstemming met gebruikers en de beheerders. Ook is gebleken dat testen zorgvuldig moet worden uitgevoerd, ook buiten de data eigenaren en bij verschillende testpersonen.

Embedded units & verbanden

Menselijke aspect

Het menselijke aspect van het dashboard wordt benadrukt doordat er veel gekeken is naar gebruikersinteractie en feedback. Dit is niet alleen terug te zien in de gebruiksvriendelijke interface, maar ook in de manier waarop het dashboard wordt aangepast en geëvolueerd op basis van directe gebruikersinput. Doordat er dagelijks zo'n 40 gebruikers hier naar kijken is de gebruikersfeedback essentieel geweest in het vormgeven van het dashboard.

Procesgericht aspecten

Het QHSE-dashboard ondersteunt een procesgerichte benadering door het bieden van gedetailleerde inzichten in workflow en operationele efficiëntie. Het stelt de organisatie in staat om processen continu te monitoren en te verbeteren, resulterend in een meer gestroomlijnde en effectieve werkomgeving. Door middel van de geregistreerde versie-updates is het mogelijk om de evolutie van het dashboard te volgen en te zien hoe het steeds meer is afgestemd op de specifieke procesbehoeften van de organisatie.

Technische aspecten

Het technische fundament van het QHSE-dashboard is robuust en flexibel, met geavanceerde data analysefuncties en maatwerkopties. De complexiteit van de gebruikte berekeningen en wijzigingen

getuigen van de technische diepgang van het dashboard, voornamelijk veroorzaakt door de complexe ontsluiting van verschillende data bronnen. Deze technische aspecten spelen een cruciale rol in de nauwkeurigheid en betrouwbaarheid van de gegenereerde inzichten, waardoor het een onmisbaar hulpmiddel is geworden voor data gestuurde besluitvorming.

Cross-functionele aspecten

Het QHSE-dashboard zorgt voor verbeterde samenwerking bij de buitendienst en directe collega's, door een gemeenschappelijk platform te bieden waarop verschillende afdelingen kunnen samenwerken. De integratie van gegevens uit diverse bedrijfsonderdelen maakt het mogelijk om een meer samenhangend beeld te krijgen van de organisatie als geheel. Dit bevordert synergiën tussen verschillende teams en verbetert de algehele organisatiecohesie.

Theoretische en praktische bijdrage onderzoek

Het QHSE-dashboard levert zowel theoretisch als praktisch een belangrijke bijdrage aan het begrip van data gedreven werken. Theoretisch biedt het inzichten in de rol van realtime data analyse in organisatiebeheer, terwijl het in de praktijk een bewezen voorbeeld is van hoe data-gedreven tools kunnen worden ingezet om operationele efficiëntie en besluitvorming te verbeteren. Het dashboard illustreert de balans die gezocht moet worden tussen technologie, menselijke interactie en bedrijfsprocessen, en onderstreept het belang van continue aanpassing en ontwikkeling in technologische hulpmiddelen binnen organisaties.

De belangrijkste lessen uit het onderzoek naar de houdbaarheid van data analyse modellen in de huisvestingscase zijn:

1. Door het hoge verbruik, vergt dit dashboard een gezonde mate van flexibiliteit
2. Gebruikers interactie is enorm belangrijk om goede feedback te ontvangen
3. Een versimpelde interactie van de
4. Adequaat handelen met gebruikers, heeft geleid tot hogere tevredenheid en betrokkenheid
5. Technische mankementen kunnen niet altijd worden voorzien
6. Creativiteit is erg belangrijk in het oplossen van complexe problemen
7. Belangrijk is om zorgvuldig te testen, ook buiten de bekende eigenaren en gebruikers
8. Zorg da het dashboard een aanland plek wordt waar complexe techniek vereenvoudigt wordt weer gegeven en interactie ontstaat, wat moet leiden tot goede constructieve discussies



FIGUUR 12 - GEBRUIKTE VISUALISATIE IN HET QHSE DASHBOARD

A6.4 – Case 4 | Productie- en contract monitorings model

Om productie bij te houden van een specifieke divisie, is er een monitoringstool gemaakt die de actuele voortgang en terugkoppeling voorziet en zo nodig allarmeert als contract afspraken niet worden gehaald. Het "Dienstleidingen" model is gericht op het inzichtelijk maken van contract afspraken, met een specifieke focus op het aansluiten van KPI's tussen uitvoering en opdrachtgever. Het model heeft als doel om een helder overzicht te bieden van de prestaties en naleving van contractuele afspraken. Zie ook Figuur 13 - Overzichten die inzicht bieden in behaalde productie- en contract monitoring.

Succes van het model

Het succes van het Dienstleidingen model, is door het geoperationaliseerde samenvatting van de belangrijkste relevante contractafspraken. Niet iedereen kent alle contractafspraken binnen de organisatie tot op de letter, maar moet hier wel naar handelen. De contractuele prestaties worden middels dit model getoetst om te zien of deze voldoen aan de KPI's. Zowel de nauwkeurigheid van de data als de tijdigheid van de informatievoorziening is daarbij essentieel, aangezien de financiële afrekening gebeurt aan de hand van deze prestaties. De doorontwikkeling van de prestatie-indicatoren is een belangrijk onderdeel aangezien de data stroom wijzigt, door verandering in de bondssystemen. Dit is uiteraard het effect van veranderende contractafspraken.

Embedded units & verbanden

Menselijke aspect

Het menselijke aspect van dit model is vooral zichtbaar in de manier waarop het team de KPI's heeft afgestemd op de werkelijke behoeften en verwachtingen van de stakeholders. Dit vereiste een diepgaand begrip van zowel de operationele processen als de verwachtingen van de opdrachtgevers. De letter van het contract wordt niet altijd even duidelijk gedefinieerd wat ertoe leidt dat er een duidelijke afspraken gemaakt moeten worden om een goede dynamiek te realiseren.

Procesgericht aspecten

Het model is erg afhankelijk van een procesgerichte aanpak, waarbij de nadruk ligt op het continu monitoren en verbeteren van de prestaties ten aanzien van contractuele afspraken. De continue doorontwikkeling van het model toont aan hoe het zich heeft aangepast aan veranderende omstandigheden en verbeterde inzichten in de procesefficiëntie. Belangrijk daarbij is dat alle partijen op dezelfde manier de letter van het contract interpreteren.

Technische aspecten

Vooraf de nauwkeurige en tijdige dataverzameling en -analyse maakte dit traject ingewikkeld. Het gebruik van geavanceerde meetmethoden en data analyse technieken, maakt het technisch lastig alle datastromen op elkaar aan te passen. De technieken stellen het project in staat om complexe datasets te transformeren in bruikbare inzichten, die essentieel zijn voor het monitoren van contractuele naleving en het behalen van KPI's.

Cross-functionele aspecten

Het Dienstleidingen model bevordert samenwerking door de integratie van verschillende personen bij de opdrachtnemer en -gever. Door deze samenwerking kunnen diverse perspectieven en expertise worden samengebracht om een meer holistisch beeld te krijgen van de contractuele

prestaties. Dit is essentieel voor het waarborgen van de integriteit en nauwkeurigheid van de contractbeoordelingen.

Theoretische en praktische bijdrage onderzoek

Het "Dienstleidingen" model biedt waardevolle inzichten in zowel de theoretische als praktische aspecten van het in stand houden van een succesvol data analyse model. Het benadrukt het belang van multidisciplinaire benadering van zowel opdrachtgever als -nemer. Dit project dient als een praktijkvoorbeeld van hoe organisaties effectief kunnen omgaan met complexe contractuele verplichtingen en hoe data-gedreven besluitvorming kan bijdragen aan het verbeteren van operationele prestaties om de gemaakte afspraken te monitoren volgens het contract.

De belangrijkste lessen uit het onderzoek naar de houdbaarheid van data analyse modellen in de huisvestingscase zijn:

1. De effectiviteit van het model hangt sterk af van de interpretatie van de uitkomst
2. Het aansluiten van KPI's op operationele gegevens is cruciaal voor de prestatiebeoordelingen
3. Het is essentieel aan te passen bij snel veranderende omstandigheden en nieuwe inzichten
4. Multidisciplinaire samenwerking is essentieel om overzicht te krijgen van de prestaties
5. Balans tussen mens, proces en techniek leidt tot betere resultaten, -inzichten en focus
6. Diepgaand begrip van het contract is nodig om het data analyse model te begrijpen
7. Belangrijk is om te begrijpen dat de bron moet worden aangepakt. Als hier begrip voor is, zijn alle stakeholders snel mee te krijgen



FIGUUR 13 - OVERZICHTEN DIE INZICHT BIEDEN IN BEHAALDE PRODUCTIE- EN CONTRACT MONITORING

A6.5 Belangrijkste conclusies cases

Voor het onderzoek naar de houdbaarheid van data analyse modellen en de factoren die hun kwaliteit en levensduur beïnvloeden, biedt dit case-onderzoek waardevolle inzichten. De kern van de studie was de interactie tussen de ontwikkeling en het onderhoud van deze modellen, gevoed door de expertise en ervaringen van een professional uit het werkveld. De geleerde lessen uit de vier cases bevestigen dat continue afstemming met de operationele praktijk en de heldere verdeling van taken en verantwoordelijkheden essentieel zijn voor het in stand houden van de bruikbaarheid en relevantie van data analyse modellen.

De flexibiliteit van het proces en de iteratieve aard van modelontwikkeling zijn cruciaal om de modellen te laten evolueren in lijn met veranderende bedrijfsdynamiek en data-inzichten. Dit vereist een robuuste technische infrastructuur die zorgt voor een betrouwbare datastroom en faciliteert in de noodzakelijke aanpassingen. Kwaliteitsborging blijkt een continu proces te zijn, waarbij technische validatie en onderhoud van modellen verzekeren dat ze niet alleen aan de huidige, maar ook aan toekomstige eisen voldoen.

Deze bevindingen tonen aan dat de houdbaarheid van data analyse modellen een specifiek fenomeen is, waarbij de menselijke interactie, procesflexibiliteit en technische precisie gezamenlijk bijdragen aan de duurzame waarde van de modellen binnen organisaties. Hiermee wordt de kloof tussen de theorie en praktijk van modelontwikkeling en -onderhoud aanzienlijk verkleind.