Clickstreams personificeret: En undersøgelse af Web Usage Mining til persona generering.

AALBORG UNIVERSITET
STUDENTERRAPPORT

. **Institut for Engineering and Tech**
Aalborg Universitet
Produkt- og Designpsykologi
Frederik Bajers vej 7A
9220 Aalborg Ø

Titel:
Clickstreams personificeret: En undersøgelse af Web Usage Mining til persona generering.

ECTS:
30 ECTS

Semester:
10. semester

Semester tema:
Kandidatspeciale

Projektperiode:
Forårssemester 2025

Projektgruppe:
PDP10 - 1083

Projektmedlemmer:
Anders Martinhoff Knudsen

Vejleder:
Christian Sejer Pedersen

Antal sider: 117

Bilag: 15, **Appendix:** 7

Abstract:

As the world changes, so must the persona litterature by considering the rising interrest in big data. This project aims to contribute to this, by an exploratory study of the use of Web Uage Mining based personas in an existing product design setting. To do this, the method is compared to a traditional persona generating process by executing both processes on the same user group. Results indicate a need to improve the current Web Usage Mining method to achieve af fully automatic persona generation. As for the traditional persona generating procedure, a need for a complete guide that takes contextualized persona template design into consideration is argured. To explore both methods relevance, resulting personas are used under same design task by product team members. Results support the continued relevance of traditional personas, and show Web Usage Mining based personas as relevant supplementary data, but not as a suitable alternative, as lack of personal information and effective data presentation is problematic. Future studies should aim at improving Web Usage Mining persona automation, and further explore relevant methods for including user frustrations and effective data visualisation techniques.

1 | Forord

Dette kandidatspeciale er udarbejdet i samarbejde med virksomheden Dataproces A/S, hvor ressourcer og sparring er samlet gennem løbende kontakt med medlemmer af MARC-teamet.

Jeg vil give en stor tak

Til hele MARC-teamet for at tro på projektet. Særlig tak til Jesper Justesen for den viden og tillid du har betroet mig, og Emma Jensen for den enorme tid og tålmodighed du har brugt på løbende sparring. Uden jer ville projektet ikke have været muligt. Jeg vil også takke Christian Sejer Pedersen for løbende vejledning i tider med usikkerhed, samt de kommunal ansatte, der har afsat tid til at deltage i interviews.

Formalia

Kildehenvisninger og litteraturlisten er lavet efter APA-styleguide, og kildehenvisning vises ved forfatrendes efternavn, samt årstal for udgivelse. I tilfælde af henvisning til hjemmesider, vil henvisning efter bedste evne referere forfatteren og udgivelsesår. Hvis dette ikke er fundet, henvises til ophaveren af rettighederne til hjemmesiden og dato for sidst tilgåelse af hjemmeside.

Bilag består af rådata, analyseværktøjer og resultater fra rapportens undersøgelser. Bilagene består af en zip-fil, der indeholder en fil med materiale for hver bilag. Det anbefales at gennemgå relevante bilag sammen med gennemlæsning af rapport, da dette vil supplere rapportens udsagn med dokumentation.

Appendix består af enten delelementer af analyser, supplerende dokumentation eller yderligere resultater, der er fjernet fra rapportens kapitler, da disse kun er dømt relevante for yderligere gennemlæsning.

Figurer og tabeller i rapporten er nøje valgt for at bidrage til forståelsen af teksten, og bør derfor studeres i sammenhæng med læsning af den tilhørende tekst.

Indholdsfortegnelse

2 | Introduktion

Internettet er en fantastisk opfindelse, der har givet enhver mulighed for at dele viden og services til hele verden igennem websites. For at sikre at disse websites også er brugbare, i stedet for at lede til frustrationer, er designet af brugerfladen vigtig. En måde at sikre sig en brugbar brugerflade, er at brugeren inddrages i designprocessen. (Bano og Zowghi 2013, Abelein og Paech 2015, Gunatilake et al. 2024-09).

Brugerinddragelse spænder over et bredt område, og der mangler et systematisk framework for at kategorisere brugerinddragelse som en metode (Zhang og Dong 2016, Wallisch og Paetzold 2020). På baggrund af dette har Zhang og Dong 2016 forsøgt at opsummere fire overordnede måder at inddrage brugeren på i litteraturen: (1) **Designere repræsenterer brugeren**. Her har designere en forståelse af brugergruppen, og bruger denne viden til at træffe beslutninger. (2) **Designere konsulterer brugerne**. Denne konsultation kan ses som et videnskabeligt eksperiment, hvor designere er videnskabspersoner, og brugere er et emne som skal undersøges. (3) **Brugere deltager i designprocessen**. Her bliver brugere interessenter i produktet, og skal indgå aktivt i beslutninger. (4) **Brugere som designere**. Her står brugerne selv for at modificere produktet, så det opfylder de ønskede krav.

Inden for design af websites bliver der gjort særligt meget brug af **Designere konsulterer brugerne** som tilgang til at inddrage brugeren (Zhang og Dong 2016). Dette kan eksempelvis gøres ved at udvikle en persona af brugeren der kan både hjælpe designere, udviklere, investorer og marketing til at træffe beslutninger (Salminen, Wenyun Guan et al. 2022).
I L. Nielsen 2019d defineres en persona som en beskrivelse af en fiktiv bruger, som er baseret på relevant information fra virkelige brugere og får læseren til at tro, at denne person kan være virkelig. Der findes ikke en standardiseret tilgang til at udvikle personaer, og tilgangen kan variere alt efter det bagvedliggende mål med personaen og den tilgængelige tid og ressourcer til at udvikle den. I en artikel af Floyd et al. 2008 beskrives fem forskellige tilgange til at udvikle personaer som følgende;

- **Cooperian personaer**. Denne tilgang tager udgangspunkt i den originale forfatter af begrebet persona (Cooper 1999). Her bliver brugerne interviewet, og på baggrund af en analyse af deres udtalelser, bliver personaer udviklet for hver type af brugere, som opstår i dataen. Personaer består oftest af kvalitativ data, og er anbefalet af Nielson Norman Group som den bedste tilgang til de fleste situationer, grundet en balance mellem pålidelighed og krav til brug af ressourcer (Laubheimer 2020).
- **Pruitt-Grudin personaer**. Denne type persona er langt mere data-dreven end Cooperian personaer, da der her udnyttes både kvalitativ og kvantitativ data til at generere personaer, og flere dataindsamlingsmetoder bliver ofte udnyttet til at samle et præcist indtryk af brugerne. Selvom denne tilgang giver grundlag for virkelighedsnære personaer, sættes der høje krav til den tid og ressourcer som skal sættes af til projektet.
- **Sinha personaer**. Her bliver der oftest kun gjort brug af kvantitativ data om brugerne, heraf oftest surveys. Statistiske metoder anvendes på dataen til at adskille brugertyper fra hinanden.
- **Ad hoc personaer**. Denne type persona opstår på baggrund af eksisterende viden om brugeren, og kræver derfor ikke yderligere dataindsamling.
- **User archetypes**. Ved denne tilgang beskrives personaer ikke som individer, men som grupper af personer.

I L. Nielsen 2019c beskrives Cooperian personaer yderligere som en mål-orienteret tilgang, da der oftest fokuseres på specifikke spørgsmål angående mål med brugen af produketet. Pruitt-Grudin personaer beskrives omvendt som en rolle-baseret tilgang. Målet er det samme som ved den mål-orienterede tilgang, men Pruitt-Grudin personaer baseres på en kritik af cooperian personaer, som menes at være for afhængig af brugskontekster, og ikke udnytter en bredere viden om brugeren.

Personaer bliver brugt til at udvikle produkter indenfor et bredt spænd af områder (Salminen, Wenyun Guan et al. 2022), og indholdet af en persona er stærkt afhængig af den kontekst, som personaen skal bruges i. Eksempelvis fandt en litteraturanalyse af Salminen, Guan, Nielsen et al. 2020, at personaer kan bestå af i alt 4 til 14 kategorier af informationer, og at mængden af information i personaer kan variere med op til 300%.

Personaer bliver oftest lavet ved brug af kvalitativ data (Brickey et al. 2012, Wang et al. 2025). Men ved denne tilgang opstår der begrænsninger for, hvor præcist brugeren kan blive repræsenteret, da det kræver flere ressourcer at nå en større andel af brugergruppen (Jansen et al. 2022). I en undersøgelse af Billestrup, Stage, Bruun et al. 2014 fandt de ydermere, at virksomheder ikke altid udvikler personaer ud fra litteratur, men oftest ved teknikker, der lånes af hinanden. Derudover blev erfaringer og kendskab til brugeren brugt i flere virksomheder som fundament for at udvikle personaer, i stedet for at udvikle personaer på baggrund af nyt data om brugeren. I dette lys er personaer tidligere blevet kritiseret for at være for "fiktive"og ikke videnskabelige nok til at have en relevant forbindelse til brugeren, idet resultaterne ikke kan reproduceres (L. Nielsen 2019c).

I modsætning til kvalitativ data, har en kvantitativ tilgang potentiale til at nå en langt større mængde af brugere på færre bekostninger (Jansen et al. 2022). Dette gøres ofte ved at sende surveys ud til brugere eller ved at indsamle web log data (Salminen, Guan, Jung et al. 2020).

Her vil sidstnævnte metode tilhøre området Web Usage Mining, som karakteriseres ved undersøgelse og analyse af data, som brugere genererer ved brugen af hjemmesider (Matias 2023). Særligt dette område oplever en stigende grad af opmærksomhed, grundet et øget interesse i big data (Taşgetiren og Aktas 2022a). Web Usage Mining er brugt til at optimere hjemmeside-baseret produkter inden for et bredt spænd af emner som eksempelvis usability af websites (Pellizon et al. 2017), brugeroplevelsen af websites (Palomino et al. 2021) og opdatering af produktkrav (Garcia og Paiva 2018).

Hertil er der et fortsat behov for fokus på eksperimentering og validering af eksisterende forskning indenfor udvikling af kvantitative personar, grundet en manglende viden om kontekstuelle begrænsninger (Salminen et al. 2025). Eksempelvis mangler der litteratur inden for undersøgelser, der sammenligner forskellige tilgange til at udvikle en persona under samme kontekst. (Jansen et al. 2022). Dette kan bidrage til forståelsen af forskellige typer personaer i forskellige kontekster, samt forståelsen af effekten af personaer generelt.

Dertil kan Web Usage Mining også være et relevant emne at undersøge indenfor generering af personaer, da clickstreams både kan være let tilgængeligt, og kan give et data-drevet indsigt i brugerens adfærd. I et studie af Zhang et al. 2016 er der tidligere gjort brug af denne tilgang til at udvikle fem personaer for hjemmesiden Platfora i samarbejde med konsulentfirmaet *Cooper*. Disse personaer blev udviklet ved at identificere workflows og derefter beskrive typiske brugere, på baggrund af disse workflows. Da de blev sammenlignet med eksisterende personaer for hjemmesiden, muliggjorde det også en verificering af fire af dem, mens to af dem viste tegn på et behov for opdatering. Dette studie viser dermed relevans for Web Usage Mining som en alternativ tilgang personagenerering.

Dog er der begrænsningerne ved tilgangen, da det alene på baggrund af clickstreams ikke er muligt at inddrage informationer som typisk vil indgå i en persona, som eksempelvis demografi, personlighed, arbejdsrelaterede problemstillinger Nielsen et al. 2015 eller bagvedliggende intentioner ved brugerens adfærd (Jansen et al. 2022). Derudover er begrænsningen af adgang til personlig data, der ellers vil give designere mulighed for at sympatisere med brugeren, en udfordring, da denne egenskab ved personaer tidligere har været beskrevet som værende essentiel (Cooper 1999).

Med det sagt kan fordelene ved data-orienterede personaer på baggrund af Web Usage Analytics ikke ignoreres, da det har potentiale til at inddrage brugeren med datadrevet adfærdsmønstre, der belyser en reel interaktion med produktet, og kan være et alternativ til andre metoder til personagenerering i en verden, som fortsat centreres omkring big data (Salminen et al. 2025).

2.0.1 Projektets forløb

Denne rapport vil have til formål at bidrage til litteraturen indenfor personaer. Dette vil gøres, ved at følge en kvalitativ, cooperiansk tilgang til persona generering i en given kontekst, og derefter en Web Usage Mining tilgang, som beskrevet i Zhang et al. 2016 til persona generering i samme kontekst. Her vil disse tilgange blive bedømt i forbindelse med den givne

kontekst, og derefter vil anvendeligheden af begge typer personaer diskuteres i samarbejde med en virksomhed. På baggrund af dette, opstilles følgende problemformulering;

Hvordan kan personaer inddrage brugere i designprocessen af et digitalt produkt?

For at undersøges denne problemformulering, vil følgende underspørgsmål besvares;

1. *Hvordan kan brugere repræsenteres ved hjælp af en kvalitativ persona genererings tilgang?*
2. *Hvordan kan brugere repræsenteres ved hjælp af en Web Usage Mining persona genererings tilgang?*
3. *Hvordan vil en kvalitativ baseret persona og en Web Usage Mining persona blive modtaget under en designproces af et digitalt produkt?*

På baggrund af disse underspørgsmål, er rapporten opdelt i tre kapitler, hvor kapitel 3 vil besvare underspørgsmål 1, kapitel 4 vil besvare underspørgsmål 2 og kapitel 5 vil besvare underspørgsmål 3.

2.0.2 Definering af begrebet persona

Begrebet Persona defineres oftest som en fiktiv person, baseret på rigtig data fra brugeren (L. Nielsen 2018), hvilket er gældende for personaer baseret på både kvantitativ og kvalitativ data. Kvantitativ funderet personaer er dog også anset som en "datadrevet alternativ"til en mere fiktiv, kvalitativ forfader (Jansen et al. 2022). Dette modstrider dog Alan Coopers beskrivelse af en persona, som lægger stor vægt på at bestå af rigtig data fra brugeren, med små fiktive tilføjelser som navn, billede og detaljer, der får personaen til at minde om et rigtigt menneske (Cooper et al. 2014, 85–87). Dette har ført litteratur til at kalde kvantitativ funderede personaer "data drevne personaer", hvilket kritiseres som værende en misvisende adskillelse (Salminen et al. 2025).

Definitionen af en persona som en fiktiv person, baseret på rigtig data fra brugeren, argumente-

res i denne forstand til at være gældende for både kvalitative og kvantiative baserede personaer.

Spørgsmålet er dog i denne forstand, hvad der skal til, før clickstreams ved brug af Web Usage Mining er tilstrækkelig personificeret, og der dermed kan være tale om en persona.

Clickstreams er tidligere nævnt i forbindelse med personaer til at evaluere repræsentativitet af valg af temaer (Thoma et al. 2009a) eller til at opdatere eksisterende persoaer (Zhang et al. 2016), men ikke til at spille en faktor for valg af indhold af personaer. I websikkerhed bliver clickstreams ofte brugt som et analytics værktøj til at identificere skadelige brugere (Kamatchi og Uma 2025), hvor visualisering af enkelte brugere er det endelige produkt (Nguyen et al. 2020). I L. Nielsen 2019d, s. 135 - 160 nævnes brug af clickstreams i personaer i forbindelse med automatisk personagenerering (eng. "Automatic Persona Generation"(APG)), som hertil typisk omhandler visualisering af brugen af sociale medier. En persona vil her eksempelvis bestå af et billede, personlig information som navn, køn og alder, interesser, citater som består af bedømmelser (Jung et al. 2018) og demografisk statistik (Association for Computing Machinery 2019), der opsamles fra brugernes profiler. APG eksisterer i denne forbindelse mellem online analytics og traditionelle, kvalitative personaer, og giver mulighed for effektivt at beskrive brugersegmenter (Salminen, S.-G. Jung et al. 2020). I Salminen et al. 2025 argumenteres der for, at online analytics adskiller sig fra personaer, ved at bestå af individuel brugerstatistik, mens personaer består af bruger ærketyper.

En Web Usage Mining persona vil i denne forbindelse placeres i højere grad mod online analytics end APG, da profil data som alder, navn, køn og interesser, der ellers virker personificerende (Nielsen et al. 2015), ikke vil være tilgængelige. En persona vil derfor i denne forstand bestå af adfærdsmæssige beskrivelser, som er tilgængelige igennem Web Usage Mining, hvor adskillelsen af disse personaer vil bestå af en adskillelse af adfærd.

2.0.3 Adskillelse af processer

Da den kvalitative og kvantitative metode er komplementære (Jansen et al. 2022), kan der være chance for, at data, som behandles i udviklingen af den ene type persona, kan påvirke

udviklingen af den anden. I lyset af projektets fokus er dette ikke ønsket. For at undgå, at de to metoder påvirker hinanden, er den kvalitative funderet metode gennemført inden begyndelsen af den Web Usage Mining baseret persona.

2.1 Dataproces

For at udføre projektet, er der indgået et samarbejde med Dataproces A/S. Dataproces er en tværfaglig konsulentvirksomhed, der blandt andet tilbyder it-ydelser til kunder i den offentlige sektor, som består af SaaS løsninger der automatisere manuelle opgaver hos kommuner, der alle tilgås gennem en web browser (A/S 2024a). Løsningerne udvikles på baggrund af et specialiseret kendskab til juridiske krav om kommunal drift, og derfor eksistere der også en faglig afstand mellem de juridisk faglige ansatte og teknisk faglige ansatte. Her består den juridisk faglige af team ledere, hvor teknisk faglige består af udviklere. Krav til design og redesign fremføres derfor af team ledere på begrund af faglig og brugermæssig kendskab, hvorefter implementeringen gennemføres af udviklere.

Derudover kendetegnes Dataproces ved løbende kundekontakt, enten ved direkte kontakt med kunder, events, customer support eller løbende ERFA-møder, hvor produktteamet præsentere idéer for kunder og brugere, og giver disse mulighed for at kommentere produktet.

2.1.1 MARC teamet

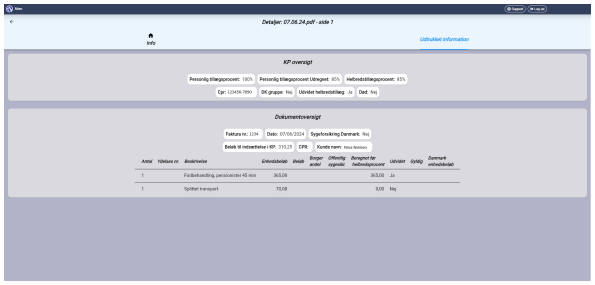
MARC er en af Dataproces's hovedprodukter og omhandler automatiseringsløsninger. MARC teamet består af en junior-udvikler, en senior-udvikler, en ansvarlige for customer support og undervisning, og en team leder. Videre udvikling af eksisterende MARC produkter foregår på baggrund af (1) konkrete ønsker om nye funktioner eller design ændringer fra brugere (2) dialog med kunder og (3) politisk og juridisk faglig viden indenfor området.

2.1.2 Case - MARC Helbredstillæg

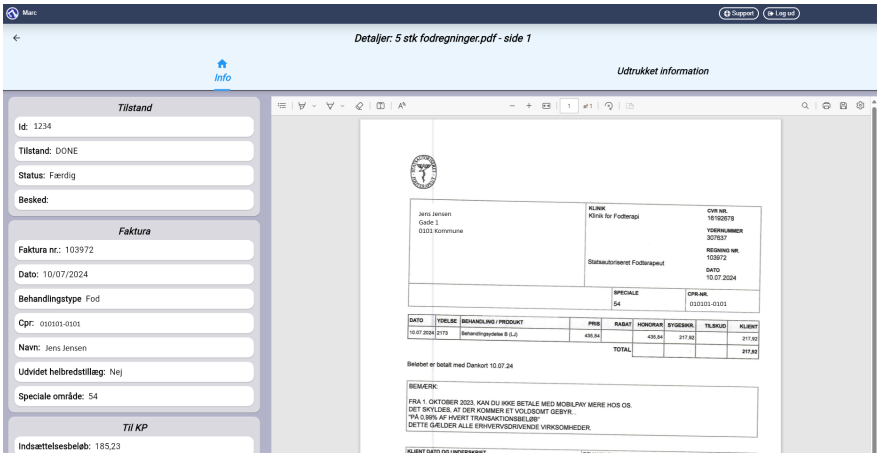
MARC Helbredstillæg automatisere indlæsningen af fysiske fakturaer, ved upload af kvitteringer til hjemmesiden, hvor en virtuel robot indsætter disse kvitteringer i et offentligt system ved navn Kommunernes Pensionssystem (A/S 2024b).

Efter en uformel dialig MARC teamet, er der givet udtryk for en manglende delt brugerrepræsentation, der besværliggøre sortering og rangering af indkomne ønsker. På baggrund af dette, viser MARC Helbredstillæg sig som en relevant case at bruge i forbindelse med undersøgelser af persona generering, hvor brugerinddragelse gennem personaer skal hjælpe teamet prioritere bruger ønsker. Nedenstående figur viser et overblik over MARC Helbredstillæg som inkluderer front page, dokument oversigt samt et eksempel på et dokument.

(A) Marc Helbredstillæg Front Page



(B) Dokument Oversigt

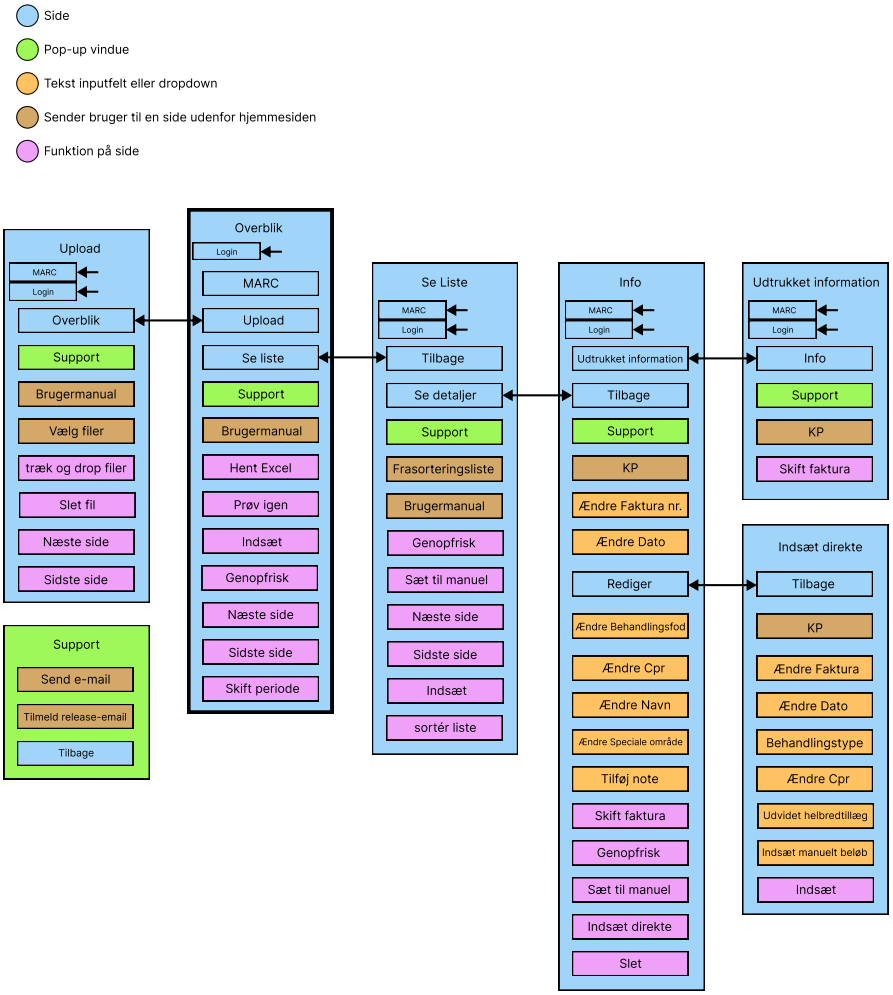


(C) Vis Dokument

FIGUR 2.1: Samlet oversigt af forskellige dokumentvisninger. Billeder er taget fra Dataproces.dk

For at forstå opbygningen af MARC Helbredstillæg er der tegnet et sitemap, som er en visuel

repræsentation af en hjemmesides organisering (J. Nielsen 2008) og kan bruges til at danne overblik over en hjemmeside (Tankala 2023). Dette sitemap kan ses på figur 2.2



FIGUR 2.2: Denne figur viser et sitemap over MARC Helbredstillæg, som fokuserer på de interaktive elementer på hjemmesiden. Sitemappen består af blå kasser, som alle repræsenterer en side på hjemmesiden. Hver blå kasse består yderligere af et antal af mindre kasser, der hver repræsenterer et element, som kan klikkes på. Disse kasser er farveinddelt alt efter, hvad der sker når elementet trykkes på; grøn symboliserer en pop-up, orange symboliserer et tekst inputfelt, pink repræsenterer en funktion på siden, brun symboliserer et element, der leder brugeren ud af hjemmesiden, når den trykkes på og blå symboliserer en knap, der fører videre til en anden side. Ved hver blå firkant er der en pil, der viser, hvilken side denne knap fører til. Hvis det er muligt at vende tilbage til en side, der er længere end én side væk, vises dette med en blå firkant under sidens navn samt en kort pil. Hovedsiden, som brugeren efter log-in bliver videreført til, er fremhævet med en fed kant. Rækkefølgen af klikbare elementer er sorteret efter type og repræsenterer ikke rækkefølgen, som elementerne bliver vist på siden.

3 | Kvalitativ persona

Persona blev introduceret i bogen "The Inmates Are Running The Asylum"i 1999 og var et modsvar til at bruge billedet af en elastisk end-user, der kunne bøjes og formes til udviklerens egne behov (Cooper 1999, s. 127). For at undgå dette, skulle brugeren beskrives af en persona, der består af en detaljeret beskrivelse af en bruger af produktet, med tilhørende billede. Personaens billede og beskrivelse skulle ikke nødvendigvis være af en virkelig person, men blot være baseret på data fra interviews af rigtige brugere (Cooper 1999, s. 128).
Sidenhen er der udgivet flere bøger fra forskellige forfattere, der bruger erfaring fra industrien til at give deres bud på en trinvis proces for at generere personaer. Der er udvalgt tre bøger som litteratur til valgte tilgang i denne undersøgelse, på baggrund af, at disse er hovedværker indenfor persona generering (L. Nielsen 2018).

Herunder er bogen "About face"af Alan copper et al. (Cooper et al. 2014) valgt som primær litteratur, da denne bog er den seneste udgave af den oprindelige forfatter af termet "persona". Supplerende værker vil bestå af "The essential persona lifecycle"af John Priutt og Tamara Aldin (Adlin og Pruitt 2010), samt "Personas - User Focused Design"af Lene Nielsen (L. Nielsen 2019d). Selvom disse tre tilgange varierer med henblik på holdning til personaer og bruger forskellige termer for hver proces, der gennemgås, er der klare ligheder på tværs af processerne. De tre kilder kan derfor supplere hinanden med yderligere detaljer om uklare punkter i processen, og supplere med yderligere indsigter inden for emnet.

3.1 Initierende undersøgelse

En essentiel del af processen, for at udvikle personaer, er en initierende undersøgelse, hvor allerede kendte informationer om brugergruppen samles for at danne en initierende hypotese om, hvilke type brugere der undersøges - samt identificerede huller i kendskabet til brugeren (Adlin og Pruitt 2010, Kapitel 3). Derudover er det også relevant at samle viden om firmaets og relevante interessenters plan, mål og vinkel for produktet, da dette kan informere om

rammer og begrænsninger for brugen af personaer (Cooper et al. 2014, s. 39). Her fremstilles følgende spørgsmål for en initierende undersøgelse (Cooper et al. 2014, s. 37 - 40);

- Hvad er produktet?
- Hvem bruger det?
- Hvad har brugerne mest brug for?
- Hvilke brugere og kunder er de mest vigtige for firmaet?
- Hvilke udfordringer møder de produktansvarlige og forretningen i den nære fremtid?
- Hvem er de største konkurrenter og hvorfor?
- Hvilket syn har forskellige interessenter og produktansvarlige på brugeren?
- Hvilket budget og ressourcer er afsat til projektet?
- Hvilke tekniske begrænsninger og muligheder eksisterer der?
- Hvilke hovedmål har firmaet med produktet?

For at finde svar på disse spørgsmål, anbefales det at gennemgå eksisterende litteratur og dokumentation om produktet, hvor manglende viden derefter kan opfanges gennem interviews med interessenter og produktansvarlige (Cooper et al. 2014, s. 39).

For at gøre dette, er der ingået kontakt med marketing afdelingen hos Dataproces. Herfra er der indsamlet salgsmateriale fra flere kilder bestående af informationssider fra Dataproces.dk, artikler udgivet hos ritzau med Dataproces som forfatter, salgsmatriale som bliver brugt til messer, heraf bestående af to Roll-ups, en folder og en video og slutteligt undervisningsmateriale om MARC produkterne, bestående af en powerpoint.

Her udføres en dokumentanalyse (Brinkmann og Tanggaard 2025, s. 163 - 164). For at gøre dette anvendes der en tematisk analyse ved brug af af en struktureret kodning (eng. "structured coding"). Denne metode adskiller sig fra en induktiv kodning af passager, ved at opbygge en kodebog på baggrund af et prædefineret fokus eller liste af spørgsmål (Guest et al. 2012, s. 55 - 56). For at bygge en kodebog, vil hvert spørgsmål være en kode, og hver kode vil bestå af fire tekstelementer; kort definition, fuld definition, hvornår koden skal tages i brug og hvornår koden ikke skal bruges (Guest et al. 2012, s. 53 - 61). Resulterende kogebog kan ses i appendix I

3.1.1 Første kodning af tekstmateriale

For at gennemgå de opsamlede dokumenter, er teksten segmenteret fra eventuelle billeder, grafer, grafisk opsætning og videoafspildning. Hertil er der prioriteret tekst, som omhandler MARC-produktet som en helhed og MARC helbredstilæg. Dertil er eventuel information om andre produkter indenfor MARC frasorteret, for at understøtte en mere målrettet analyse af MARC Helbredstillæg. Analysen er foregået ved at tilføje en farve til hver kode og markere dele af teksten, der efter kodebogen tilhører den givne kode. Resulterende kodning af dokument efter første gennemlæsning kan ses i bilag 3.

3.1.2 Anden kodning af tekstmateriale

PRODUCT_DESC og USER_NEEDS viste sig at være besværlige at skelne mellem, da specifikke dele af produktet antagelsesvist er blevet udvalgt til at blive beskrevet for at imødekomme brugerens behov. Eksempelvis ved tekststykket "*Desuden har MARC automatiseret oprettelse af journalnotat i KP, hvormed I vil kunne fremsøge informationer om sagerne senere hen.*"kan det tolkes både som en beskrivelse af produktet og et forsøg på at imødekomme et brugerbehov. For at kode teksten i disse situationer opdeles PRODUCT_DESC og USER_NEEDS i "hvad kan produktet gøre"og "hvad kan produktet gøre for dig". *Desuden har MARC automatiseret oprettelse af journalnotat i KP* kodes som PRODUCT_DESC og *I vil kunne fremsøge informationer om sagerne senere hen.* kodes som USER_NEEDS. Derudover er PROD_USERS ændret til at bestå af målgruppe i stedet for brugere, da det ikke gøres klart i de indsamlede dokumenter om, hvorvidt de nævnte modtagere er brugere eller kunder. Resulterende kodning af dokument efter anden gennemlæsning kan ses i bilag 2. I alt 8 koder er anvendt under dokumentanalysen, da der ikke blev fundet tekst som opfyldte kravene til kodning af koderne STAKEHOLDER_VIEW og COMPETITION. Den kodede tekst er samlet i hver sit dokument og analyseret med henhold til teamer, som kan ses i bilag 1.

Begrænsninger ved analysen

Det kan forekomme, at de analyserede dokumenter enten er forældet eller misfortolket, da forskellige dokumenter kan være lavet til forskellige målgrupper og skrevet med forskellige

niveauer af ekspertise. Resultatet af analysen skal derfor anses som en række antagelser der skal valideres.

3.1.3 Validering af antagelser

For at validere dokumentanalysen, tilsendes et dokument med resultaterne fra dokumentanalysen gennem mail til tre hovedansvarlige for MARC Helbredstillæg; en IT konsulent, der håndterer customer support, en seniorudvikler, der har bygget produktet samt team-lederen for MARC. Hertil er følgende spørgsmål stillet; (1) *Hvilke påstande fra analysen er rigtige, og hvilke påstande er forkerte?*, og (2) *Hvilke vigtige og relevante ting og pointer jeg ikke har fået med eller misset i hver påstand?*. Den fulde mail kan findes i bilag 4. Der er indsamlet besvarelse på mail fra en seniorudvikler, en IT konsulent, der håndtere customer support, samt et opkald fra team-lederen af MARC, hvoraf dataindsamlingen fra sidstnævnte samtale består af noter. For at tilrette resultaterne fra dokumentanalysen, er disse svar gennemlæst og samlet til en ny tilrettet tekst, som efterfølgende er læst igennem for validering i samarbejde med IT konsulenten der håndterer customer support. Besvarelser fra disse tre personer kan ses i bilag 5 og en tilrettet resultat af den initierende undersøgelse kan ses i appendix III;

3.2 Hypotesedannelse

Ifølge (Cooper et al. 2014, s. 46) består næste skridt af at danne en hypotese af brugeren, på baggrund af den indsamlede viden. Denne hypotese skal hjælpe med at besvare, hvordan brugerne og brugernes adfærd adskiller sig fra hinanden, og kan dermed hjælpe med at målrette interviews med brugerne. Derudover kan en hypotese om brugeren også bruges til at identificere, hvilke brugere der er vigtigst at fokusere på i de senere skridt i processen, samt sikre, at der er enighed om billedet af brugeren blandt alle parter (Adlin og Pruitt 2010, Kapitel 4). Hertil anbefales det at identificere fire områder ved den potentielle bruger; brugerroller, adfærd og demografiske variabler, domæne ekspertise versus teknisk ekspertise og brugerens miljø, som hver er en relevant område at adskille brugere fra hinanden (Cooper et al. 2014, s. 47 - 48).

3.2.1 Brugerroller

Da MARC Helbredstillæg bruges i en B2B (eng. "Business to Business") forretningskontekst, er brugerrolle og stilling forbundet (Cooper et al. 2014, s. 47). I den forbindelse eksisterer der en samlet brugerrolle for alle brugere af MARC Helbredstillæg, som er kommunelle sagsbehandlere, der sidder med helbredstillæg.

3.2.2 Adfærd og demografiske variabler

Denne måde at adskille brugere er mest brugbar indenfor produkter med private kunder (Cooper et al. 2014, s. 48). Da det antages at MARC Helbredstillæg befinder sig indenfor kategorien B2B, vil denne måde at differensiere brugere på ikke yderligere overvejes.

3.2.3 Domæne ekspertise versus teknisk ekspertise

Ifølge dokumentanalysen fra kapitel 3.1 kan brugere af MARC Helbredstillæg variere bredt i forhold til domæne ekspertise, alt efter faglig baggrund og erfaring indenfor helbredstillæg; *[...] der er tale om en bred gruppe af forskellige fagligheder. Sagsbehandlere varierer meget fagligt.*. Dette kan føre til et behov for yderligere oplæring og evalueringsmøder af brugen af løsningen, adgang til customer support og standardisering af arbejdsgange, der hjælper onboarding af nye medarbejdere; *De har brug for standardiserede arbejdsprocesser, understøttet af et korrekt datafundament frem for at følge den enkelte medarbejder, der støtter onboarding af nye medlemmer, da det er svært at rekruttere.* Dog er det erfaret, at kommuner takker nej til evalueringsmøder, men at dette kan skyldes, at spørgsmål bliver besvaret gennem customer support; *Med henhold til evalueringsmøder opleves der, at kommunerne sommetider takker nej til dette tilbud, fordi de ikke har behov herfor, f.eks. fordi de får deres spørgsmål besvaret igennem supporten.*. Brugerne har antageligtvis en mindre teknisk ekspertise, da et af de identificeret brugerbehov er en løsning, der er nem at bruge; *På brugerniveau har brugerne brug for et nemt system til automatisering af fakturaer til refusion af helbredstillæg [...].*

3.2.4 Brugerens miljø

Alle brugere af MARC Helbredstillæg er ansatte i danske kommuner, hvor disse kommuner kan variere i både størrelse og økonomiske rammer. Erfaring fra dokumentanalysen har vist, at Dataproces tidligere har differensieret mellem tre forskellige størrelser af kommuner; *Dataproces skelner i udmeldinger mellem 3 kommunestørrelseskategorier; de 50 mindste kommuner benævnes kommuner, de 38 midterste benævnes større kommuner og de 10 største benævnes top-10 kommuner.*. Dog er der ikke givet udtryk for, at en opdeling af kommuner på baggrund af størrelse er relevant, da alle kommuner tillægges samme fokus; *Dataproces skelner på sin vis ikke i, om man vil have store eller små kommuner. De vil gerne have løsningen ud til alle kommuner.*.

3.3 Resulterende hypotese

På baggrund af ovenstående analyse af tre områder angående brugerne, antages det at brugerne adskiller sig i højest grad ved domæne ekspertise og faglig baggrund. Brugerens behov følger heraf også deres baggrund, ved at brugere med en svækket faglig baggrund har brug for mere hjælp. Derudover er de opgaver - og dermed adfærd, afhængig af, hvilke opgaver som bliver stillet en sagsbehandler i borgerservice, som på nuværende tidspunkt endnu ikke er belyst. Kommuner kan variere i størrelse og økonomi, og det er ikke kendt hvilken betydning dette spiller for brugen af produktet. Identificerede opgaver består hovedsageligt af upload af håndskrevne fakturaer, samt manuel indsættelse af de faktureaer, der fejler. Andre opgaver kan bestå af download og visning af de indsatte fakturaer.

3.4 Begrænsning af hypotesedannelse

For at danne en optimal hypotese om brugeren, er det anbefalet at indgå i tæt samarbejde med de produktansvarlige, og opretholde en proces der sikre feedback fra relevante interresenter gennem hele processen, for at undgå misforståelser eller indsamling af forkert information (Adlin og Pruitt 2010, kapitel 4). Det er også anbefalet at bruge interviews af produktansvarlige som tilgang til at undersøge konteksten af produktet. Grundet en begrænsning af tid har

processen forløbet med en mindre grad af direkte kontakt til produkt teamet, hvor validering af antagelser om brugerne er foregået gennem mailkorrespodence. Dette har begrænset valideringsprocessen, og dermed forhøjer chancen for at bygge personaerne på et forkert udgangspunkt. For at komme udenom dette, anbefaler Adlin og Pruitt 2010, Kapitel 4 eksempeltvis at danne ad-hoc personas, før der indgås kontakt med brugerne, da dette kan give interresenter et tydeligt udgangspunkt for at afkræfte eller bekræfte antagelser om brugeren. Grundet begrænsning af tid har dette dog ikke været muligt. Andre områder som bruger problemer, brugerfejl og klager ikke blevet undersøgt. Disse informationer kunne bidrage til at rette fokus i en bestemt retning under interviews af brugere, og kunne undersøges gennem eksempeltvis customer support.

3.5 Interview af brugere

I dette afsnit gennemgås handlingsplanen baseret på guiden af Cooper et al. 2014, samt udførelsen og analysen af brugerinterviews.

3.5.1 Ressourcer til interview

For at validere eller afkræfte hver identificeret variabel i hypotesen, anbefales det iføge Cooper et al. 2014, s. 49 at blive undersøgt over fire til seks interviews, hvoraf disse interviews skal vare i op til en time Cooper et al. 2014, s. 50. For at få mest ud af interviews, kan personer blive valgt til interview gennem på baggrund af persona hypotesen (Cooper et al. 2014, s. 35). Dette kan sikre, at personer, som adskiller sig på relevante parametre, bliver undersøgt. Dette har dog ikke været muligt at gøre, da tid med brugere af MARC Helbredstillæg har været en begrænset ressource. Derfor er der i stedet anvendt opportunity sampling, hvor seks udvalgte kommuner er kontaktet. Gennem dette har tre deltagere vendt tilbage med et ønske om at blive interviewet, med en aftale om 30 minutter til interviewet. Her vil interviewet foregå over et online møde ved brug af Microsoft Teams.

3.5.2 Definering af plan

For at vælge en bestemt metode og tilgang til interview af brugere, er det ifølge L. Nielsen 2019a først vigtigt at bestemme målet med interviewet; hvilke type mennesker, som skal bruge personaen efterfølgende og hvad de skal bruge den skal spille en rolle de beslutninger der skal træffes.

For at imødekomme modtagere, bestående af både udviklere, team ledere og customer support ansvarlige, kan spørgsmål angående konkredte problemer med user interface eller misforståelser være relevante, da disse situationer kan oversættes til konkredte beslutninger om design.

I forbindelse med at generere personaer, er et typisk mål med interviews at belyse, hvordan brugere adskiller sig fra hinanden (L. Nielsen 2019a). For at gøre dette med den tilgængelige mængde af ressourcer, er det valgt at hovedsageligt fokusere på brugerens domæne ekspertise.

Disse spørgsmål skal hovedsageligt være rettet emner der omhandler brugeren, i modsætning til produktet (Cooper et al. 2014, s. 42). Relevante emner kan bestå af **brugskontekst**, **domæneviden**, **nuværende opgaver og handligner med produkter**, **mål og motivation til at bruge produktet**, **brugerens mentale model af produktet** samt **problemer og frustrationer over produktet** (Cooper et al. 2014, s. 42 - 43).

For at udforske, hvordan brugerne adskiller sig fra hinanden, og på baggrund af antagelsen om, at brugere er forskellige på baggrund af domæne ekspertise, vil **domæneviden** være hovedemnet for interviewet. For at gøre dette, er emnet brudt ned til yderligere underemner; (1) teknisk ekspertise og interresse, (2) faglig ekspertise og interresse samt (3) praktisk erfaring med MARC og helbredstillæg. Derefter vil der blive fokuseret på **problemer og frustrationer over produktet** for at imødekomme brugskonteksten af personaen i en design proces, der omhandler redesign. Slutteligt vil der blive suppleret med spørgsmål angående andre faktorer, der kan spille ind i **brugskontekst**, som antal af kollegaer, der håndtere samme opgaver, og vidensdeling.

3.5.3 Valg af interview metode

I L. Nielsen 2019a skelnes der mellem struktureret, semi-struktureret og ustruktureret interview, alt efter hvor målrettet interviewet skal være.

Da der er afsat en begrænset mængde tid til hver interview, er der valgt at gøre brug af struktureret interviews, da dette vil sikre, at de vigtigste aspekter af hver person indsamles. 15 spørgsmål blev herpå formuleret, som kan ses i bilag 6. En pilottest af spørgsmålene blev herefter gennemført i samarbejde med en privatperson, som er ansat hos Port of Aalborg. For at gøre dette, blev "sagsbehandler"udskifter med Port of Aalborg, og MARC Helbredstillæg blev udskiftet med et system ved navn Get Organized. Her blev der identificeret spørgsmålene *Hvad er dine styrker og dine svagheder som sagsbehandler?* og *Hvad elsker du og hvad hader du ved at være sagsbehandler?* for værende for personlige, og derfor ubehagelige at besvare. De tilrettet spørgsmål efter pilottest kan ses i bilag 6. Som validering af spørgsmål, er de tilsendt den ansvarlige for MARC Helbredstillæg customer support. Tilbagesvaret førte ikke til ændringer i spørgmål, men yderligere ændringer i rækkefølge. De resulterende spørgsmål kan ses i tabel 3.1.

Type	Spørgsmål
Domæneviden	Hvad er din faglige baggrund?
Domæneviden	Hvor lang tid har du arbejdet som sagsbehandler?
Domæneviden	Hvad kan du bedst lide og hvad kan du mindst lide at være sagsbehandler?
Domæneviden	Hvordan er det for dig at arbejde med helbredstillæg?
Domæneviden	Hvordan vil du klassificere din viden og interesse indenfor helbredstillæg?
Domæneviden	Hvor længe har du brugt MARC Helbredstillæg?
Domæneviden	Hvilke opgaver løser du med MARC Helbredstillæg?
Domæneviden	Hvad er dine faglige styrker og dine svagheder som bruger af MARC Helbredstillæg?
Problemer	Hvilke ting er du glad for ved MARC Helbredstillæg?
Problemer	Hvilke ting frustrerer dig ved MARC Helbredstillæg?
Problemer	Bliver du nogle gang forvirret når du bruger MARC Helbredstillæg?
Problemer	Hvad betyder det at have MARC Helbredstillæg for din arbejdsdag?
Domæneviden	Hvor komfortabel er du med at lære nye digitale værktøjer?
Domæneviden	Når du skal lære et nyt system at kende, hvad tænker du så?
Brugskontekst	Samarbejder eller sparer med andre kollegaer angående helbredstillæg?
Brugskontekst	Hvordan er du blevet oplært i at håndtere helbredstillæg og sagsbehandling?

Tabel 3.1: Denne tabel viser de resulterende spørgsmål til interview af brugere af MARC Helbredstillæg, samt den tilhørende type af spørgsmål.

3.5.4 Iterering af spørgsmål og tilgang

For at udnytte erfaring fra tidligere interviews til at forbedre kommende interviews, kan valget af metode ses som en iterativ proces, hvor ubesvarede spørgsmål, erfaringer og undringer kan forme tilgangen og valg af fokus af interviews (Cooper et al. 2014, s. 56). Ved første interviews erfares, at de valgte spørgsmål leder til detaljerede svar, men at svar overlapper spørgsmål. Særligt spørgsmålet *Bliver du nogle gang forvirret når du bruger MARC Helbredstillæg?* blev anset som en gentagelse af tidligere samtaler. For at håndtere dette, er det valgt at skifte metoden til en mere semistruktureret tilgang, da dette bedre vil kunne håndtere situationer, hvor samtalen springer mellem spørgsmål. Derudover var spøgsmålet *Hvad er dine faglige*

styrker og dine svagheder som bruger af MARC Helbredstillæg? svært at besvare for personen, og er derfor fjernet. Slutteligt var spørgsmålet *Hvad kan du bedst lide og hvad kan du mindst lide at være sagsbehandler?* problematisk, da der indeholder to spørgsmål i samme formulering, og der blev kun besvaret et af spørgsmålene. For at undgå dette, er spørgsmålet splittet i to. Slutteligt er spørgsmålet *Hvad gør en god dag god som sagsbehandler?*, for at samle mere information om brugerens mål i løbet af hverdagen (Cooper et al. 2014, s. 52).

3.5.5 Behandling af interview

Efter interviewet er optagelsen transkriberet ved hjælp af et CLAAUDIA Transcriber, som er tilgået gennem Aalborg Universitet. Resulterende transkribering er efterfølgende behandlet for linjeskift og læsbarhed gennem et python script (se bilag 7) og gennemgået sammen med optagelsen for at rette eventuelle fejl og mangler. For at reducere større tekstsegmenter i transkriberingen, er der anvendt meningskondensering, da dette vil sikre en bibeholdelse af meningsindhold gennem fortolkning (Brinkmann og Tanggaard 2025, s. 57). Resulterende dokumenter kan findes i bilag 7.

L. Nielsen 2019b beskriver Formålet med at behandle den indsamlede data fra interviews, som værende skabelse af mening, når der både kigges på det individuelle interview, og på tværs af interviews. Dette kan gøres enten ved at strukturere en tematisk analyse (L. Nielsen 2019b) eller ved blot at gennemgå dataen for noter og markeringer på en ustruktureret vis i mindre nuanceret og komplekse domæner (Cooper et al. 2014, s. 56). Grundet mængden af indsamlet data for hver interview, samt kompleksiteten ved domænet, er der valgt en struktureret, tematitisk analytisk tilgang.

3.5.6 Diskussion af valg af interview spørgsmål

Persona hypotesen har kun været delvis behjælpelig til at udvinde spørgsmål til interviews, da det har hjulpet med retning af spørgsmål, men ikke med konkrete emner der skal spørges ind til. Dette kan skyldes, at den initierende undersøgelse ikke har været grundig nok. Det anbefales ikke at bruge et fast sæt af spørgsmål til interviews, da dette kan fremmedgøre (eng. "alienate") personen og potentielt misse ukendt data (Cooper et al. 2014, s. 52). Dog

bliver en struktureret interview tilgang nævnt som en valid tilgang af L. Nielsen 2019b, men nævnt uden yderligere uddybelse af anvendelse.

Det er uklart hvor relevante de valgte spørgsmål har været for at danne mål-orienterede personaer, da mål-orienterede spørgsmål ikke bliver klart defineret, andet end eksempler på mulige spørgsmål (Cooper et al. 2014, s. 52). Det er også uklart, hvor relevant mål-orienterede spørgsmål er for slut-brugeren af personaerne, da produktet bruges i en job kontekst og brugerne derfor er motiveret til at bruge produktet, da det er en del af deres daglige arbejde.

3.6 Generering af personaer

I Cooper et al. 2014, s. 81 - 92 formuleres en 8-trin proces for at udvikle personaer ud fra interview data om brugere, som i dette projekt er valgt som tilgang. Dette afsnit vil gennemgå disse trin, samt hvordan de er fulgt.

3.6.1 1. groupér interviews efter rolle

dette trin består i at groupére interviews på baggrund af den rolle, som personen indebærer. Dette gøres for at kunne sammenligne personer med samme brugskontekst, og dermed fokusere på personlige forskelle, i stedet for forskelle i brug af produktet.

Person 1 vedrøre bevillingen af helbredstillæg; *Som sagt er det selve bevillingen af helbredstillæg jeg sidder med.*. Person 2 og 3 arbejder primært med kørsel af kvitteringer i MARC Helbredstillæg, hvor person 2 både uploader kvitteringer og tjekker K.P; *Derfra uplaoder jeg så kvitteringerne til MARC. så efterhånden, som den frasorterer, så går jeg listen igennem og bearbejder det, hvad der mangler. Her slår jeg op i K.P. og finder de nødvendige informationer og indsætter dem direkte og kører marc igen.*, mens person 3 blot uploader kvitteringerne; *Jeg står for hele processen med MARC. Mine kollegaer er dem der tjekker op i K.P. og kontakter borgerne.*. Herpå opstår to mulige grouperinger; (1) groupere person 1 for sig selv på baggrund af håndtering af hele bevillingen, og groupere person 2 og 3 sammen på baggrund af håndtering af af MARC Helbredstillæg, eller (2) groupere alle personer for sig selv, da

person 2 slår op i K.P, mens person 3 ikke gør. Grundet den lave antal af data, og at alle personerne adskiller sig på baggrund af ansvarsområder er det besluttet at groupere alle tre personer hver for sig. Groupering kan ses i følgende tabel.

Personer	Rolle
Person 1	Fuld Bevilling af Helbredstillæg
Person 2	Upload og kørsel af kvitteringer i MARC Helbredstillæg, samt slår op i K.P.
Person 3	Upload og kørsel af kvitteringer i MARC Helbredstillæg.

3.6.2 2. identificér adfærdsmæssige variabler

Her identificeres adfærdsmæssige variabler for hver rolle. Adfærdsmæssige variable referere i denne forstand til faktorer, der påvirker brugerens adfærd direkte eller indirekte, og danner grundlaget for at definere personaerne. Hertil anbefales det at fokusere på fem emner;

- **Aktiviteter** - Hvad brugeren gør
- **Holdninger** - Hvordan brugeren forholder sig til domænet og teknologien
- **baggrund** - uddannelse og oplæring; ævne til at lære
- **Motivation** - Hvorfor brugeren er engageret i produktet og domænet
- **Evner** - faglige og tekniske evner relateret til produktet og domænet

For at gøre dette, er interviews gennemgået en struktureret kodning på baggrund af de fem emner, og disse koder er sammenlignet på tværs af interviews for at finde gennemgående variabler angående adfærd. Disse variabler er valgt på baggrund af, at de skal (1) være nævnt på tværs af alle interviews og (2) beskrive en adfærd, eller en variabel der påvirker adfærd. Grundet den begrænset mængde data, og til trods for at alle personer er grouperet i hver sin gruppe i tidligere trin, er alle interviews samlet i samme gruppe for dette trin. Kodning samt sammenligning af interviews kan findes i bilag 9, og de identificeret adfærdsmæssige variabler kan ses i tabel 3.2.

Emne	Variabel	Beskrivelse
Aktiviteter	Opgavemangfoldighed	Hvor mange skiftende opgaver, som brugeren bliver udsat for
Aktiviteter	Sparring	Hvor ofte brugeren sparer med kollegaer
Aktiviteter	Håndtere jævnlige problemer	Hvor ofte brugeren håndtere uventede problemer i MARC Helbredstillæg
Aktiviteter	Mængde af ansvar	Ansvarsområdet indenfor arbejdet med MARC Helbredstillæg
Holdninger	Holdning til at lave fejl	Hvordan brugeren forholder sig til at begå fejl med IT-systemer
Holdninger	Forhold til nye systemer	Hvordan brugeren forholder sig til at lære nye systemer
Holdninger	Holdning til MARC Helbredstillæg	Hvilken holdning brugeren har angående MARC Helbredstillæg
Holdninger	Holdning til forskelligartede opgaver	Hvordan brugeren forholder sig til at have mange forskelligartede opgaver
Holdninger	Gode og dårlige dage	Hvor oftest har brugeren dårlige dage
Baggrund	Erfaring hos kommunen	Hvor længe har brugeren arbejdet hos kommunen
Baggrund	Evne til at lære nye it systemer	Hvor gode evner har brugeren til at sætte sig ind i nye it systemer
Baggrund	Uddannelse	Hvor lang en uddannelse har brugeren
Evner	Erfaring med MARC Helbredstillæg	Hvor meget erfaring har brugeren med MARCHelbredstillæg

Tabel 3.2: Denne tabel viser de identificeredet adfærdsmæssige variabler.

Ved nogle af variablerne er der ikke komplette udtalelser om området, som eksempeltvis ved "dårlige dage"udtaler person 2 sig ikke om dårlige dage, men blot hvad der gør en god dag god på det seneste; *Ja, altså god er jo lige pt. For mig i hvert fald at være sammen med mine kolleger, at jeg har mulighed for stadigvæk at bidrage med noget, selvom jeg kun er på de der 12 timer om ugen.*. Her bliver der i stedet lavet en bedømmelse på baggrund af ordvalg, som eksempeltvis brugen af "lige pt."som hentyder at der også dårlige dage, som ikke bliver snakket om, i modsætning til person 3, som nævner at denne bruger ikke har dårlige dage.

3.6.3 3. forbind interviews til adfærdsmæssige variabler

For at undersøge mønstre i brugernes adfærd på baggrund af disse variabler, består dette trin i at forbinde interviewet bruger med alle variabler, og identificér steder, hvor flere personer enten følger hinanden eller lægger ovenpå hinanden. Hertil er den præcise placering af hver adfærdsmæssig variabel ikke vigtigt, da det er forskellen på variablerne på tværs af personerne, som er relevant.

For at gøre dette, er teksterne genlæst, som er ledt til de identificerede adfærdsmæssige variabler, og tolkes, hvordan disse personer vil ligge i forhold til hinanden på en skala. Her er skalaen udviklet på baggrund af et eksempel fra Cooper et al. 2014, s. 84.

FIGUR 3.1: Denne figur viser, hvordan de tre personer er bedømt til at ligge i forhold til hinanden, i forbindelse med de identificerede adfærdsmæssige variabler.

3.6.4 4. identificér signifikante adfærdsmæssige variabler

For at kunne udlede personaer ud fra figur 3.1, anbefales det at finde et mønster på seks til otte adfærdsmæssige variabler, hvor samme brugere ligger tæt. Dette mønster betegnes som et signifikant adfærdsmæssig variabel, og kan bruges til at beskrive en persona (Cooper et al. 2014, s. 85), I forbindelse med dette, argumenteres person 2 og 3 for at følge en signifikant adfærdsmæssig variabel, da disse to brugere kan ses at ligge tæt på otte forskellige adfærdsmæssige variabler, som vist på figur 3.2.

FIGUR 3.2: Denne figur viser, hvordan de tre personer er bedømt til at ligge i forhold til hinanden, i forbindelse med de identificerede adfærdsmæssige variabler, hvor person 2 og person 3 ses at ligge tæt ved flere variabler.

3.6.5 5. syntetisere karakteristikker og definér mål

Dette skridt består af første forsøg på at danne en persona, ved at gennemgå noter for de personer, som danner en fælles signifikant adfærdsmæssig variabel, og derfra samle informationer, hvor otte emner burde inddrages for hvert mønster (Cooper et al. 2014, s. 85);

- Adfærden i sig selv (aktiviteter og motivation bag dem)
- Brugsomgivelserne (miljøet hvor adfærden forekommer)
- Frustrationer og smertepunkter i forbindelse med nuværende løsninger
- Demografiske træk forbundet med adfærden
- Færdigheder, erfaring eller evner relateret til adfærden
- Holdninger og følelser forbundet med adfærden
- Relevante interaktioner med andre personer, produkter eller tjenester
- Alternative eller konkurrerende måder at gøre det samme på – især analoge metoder

For at gøre dette, er de kodet segmenter af interview teksten fra afsnit 3.6.4 genanvendt til at udtrække udtalelser fra brugerne til hver detalje. Ved person 2 og 3 - hereftere refereret til som persona 1, er udtalelserne samlet til en fælles udtalelse, der bedst muligt dækker fælles træk fra begge interviews. Person 1 - herefter refereret til som persona 2, bliver brugt til at generere en persona for sig selv. Detaljer for hver persona er i dette trin beskrevet i punktform. Resulterende personaer, samt tilhørende udtalelser som begrundelse af tekstvalg for første iteration, kan ses i bilag 10.

3.6.6 6. tjek for overflødighed og fuldendthed

Dette trin består af at gennemgå de genererede personaer for manglende data, samt undersøge, om personaernes tekst adskiller sig fra hinanden på en meningsfuld måde. Dette gøres for at iterere over personaerne og sikre en synlig adskillelse og sammenlignlighed. For at gøre dette, bedømmes sammenlignligheden af personaerne, ved at undersøge, hvor godt teksterne beskriver samme emner, og i hvilken grad personaerne adskiller sig fra hinanden under hvert emne. Dette gøres, da det giver mulighed for at fjerne tekst, som er ens mellem personaerne og dermed overflødelig, samt belyse emner der mangler data.

Efter gennemlæsning er der tilføjet to ekstra punkter til persona 2; *jeg lære bedst nye systemer når jeg bliver ordenligt sat ind i dem og kan få hjælp* og *nogle gange er der nogle fejlkoder som jeg bliver forvirret over.* Derudover er *jeg har arbejdet hos kommunen i ca. 6 år* og *når jeg får nyt teknologi tænker jeg "let's do it" og så prøver jeg mig frem* tilføjet i persona 1. Disse fire tekster er tilføjet for at gøre personaerne mere sammenlignlige. Derudover er *bevilling af helbredstillæg* tilføjet til første punkt i persona 2 for præcisere ansvarsområdet. resulterende personaer for anden iteration kan ses i bilag 10.

3.6.7 7. designér personatyper

I dette tring laves en prioritering af personaerne, hvor en primær persona vælges, ved at bedømme, hvilke krav der kan opfyldes uden at modarbejde kravene fra de resterende personaer. Herved anbefales det også at undgå beslutning på baggrund af at repræsentere så mange brugere som muligt, da dette vil føre til de mest dominerende kravspecifikationer,

og ikke nødvendigvis de kravspecifikationer opfylder flest behov (Cooper et al. 2014, s. 80 - 82).

For at vælge en primær persona, gennemgås dataen tihørende persona 1 og 2 for fællestræk. Herpå kan det ses, at disse personaer udtrykker behov der ikke vedrøre hinanden. Persona 2 ønsker mere gennemsigtighed i systemerne *jeg vil gerne have gennemsigtige systemer så jeg kan se hvad der sker og hvad der er sket*, ønsker at udvikle sig *[...] motivere mig, for jeg føler at jeg udvikler mig fagligt og personmæssigt* og bryder sig ikke om at begå fejl *jeg er bange for at lave fejl, men fejl sker og det er også okay* mens persona 1 ønsker først og fremmest at blive færdig med sine opgaver *jeg bliver også motiverede ved at se mit fremskridt i arbejdet og at jeg bliver færdig med mine opgaver* og udtrykker ikke bekymringer for at begå fejl *Jeg er ikke bange for at lave fejl, for jeg har altid en livslinje hvis det går galt*. Herpå ses der ikke en tydelig primær persona, der også opfylder behov af andre personaer.

For at styrke forudsætningerne med at træffe den bedste beslutningen, gennemgås personaernes tilhørende data, hvor særligt komplementære udtalelser mellem person 1 og 2 undersøges, da disse ifølge figur 3.1 har mere tilfælles end person 1 og 3. Eksempeltvis udtrykker person 2 også et ønske om at forståelse af det tekniske bag MARC Helbredstillæg *Jeg er ikke bange for IT, tværtimod, så giver det mig en fed fornemmelse når jeg forstår hvad der sker*, og en glæde ved opgaver, der giver anledning til fordybelse *Jeg kan lide at man kan fordybe sig og få noget for hånden, hvor man kan se fremskridt i sit arbejde*. Dette kan tyde på, at persona 2 også kan dække mere usynlige ensligheder mellem de to personaer.

Dog kan der også argumenteres for, at persona 1 er mere relevant som primære persona, da denne er baseret på brugere med drift af produktet som hovedansvar, og er derfor mere relevant i forbindelse af et redesign, hvorimod persona 2 indeholder i højere grad informationer om hele arbejdet med helbredstillæg, og argumenteres derfor for at mere relevant i forbindelse med udvidelse af produktet.

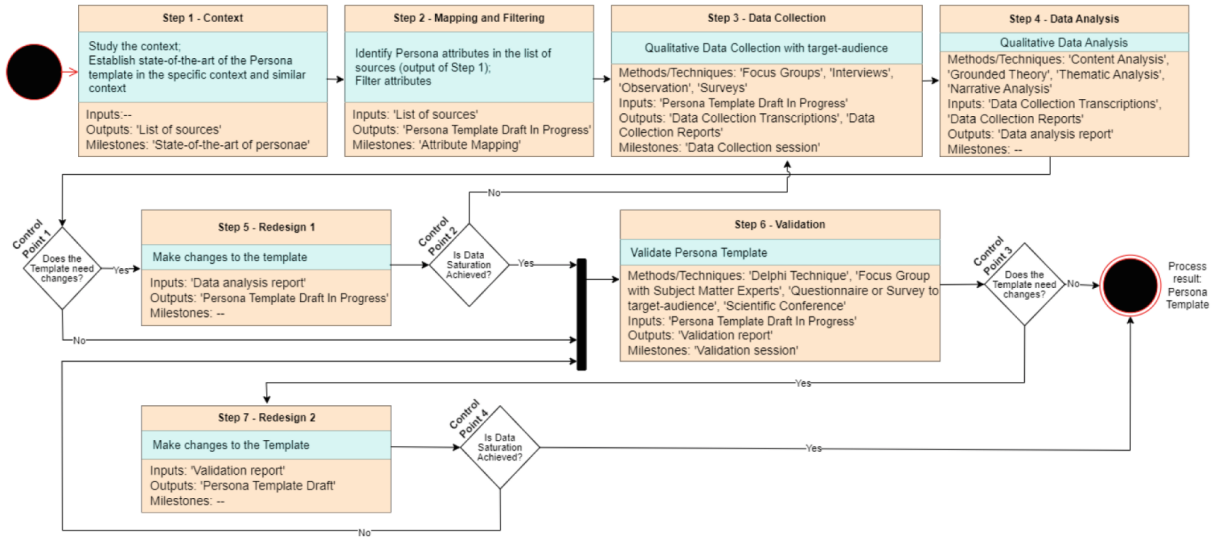
På baggrund af gennemgåelse af personaernes tilhørende data, er det valgt at vælge persona 2 som den primære persona, hvoraf persona 1 er valgt som sekundær. Dog er der ikke et

tydeligt argument for valg af primær persona, da begge personaer vedrøre forskellige behov.

3.6.8 8. udvid beskrivelse af attributter og adfærd

Dette trin består af at strukturere personaen og udvælge tekststykker, som skal med i personaen, og strukturere denne tekst, samt tilhørende billede og navn gennem en persona skabelon. Her anbefales det at lave en sammenfattende tekst for hver persona, ud fra de tidligere valgte tekststykker.

Valg af skabelon skal dog være baseret på konteksten af den givne persona, og ikke på litteratur alene, da dette kan føre til et overflødigt design af skabelonen (Couto et al. 2025). I forlængelse med dette præsentere Couto et al. 2025 en iterativ design proces, der vil sikre et kontekst-drevet beslutningsgrundlag, som kan ses på figur 3.3.



FIGUR 3.3: Denne figur viser Method for Creating Persona Templates (MCPT) metoden, præsenteret i Couto et al. 2025, som kan bruges til at beslutte design af persona skabelon på baggrund af kontekst.

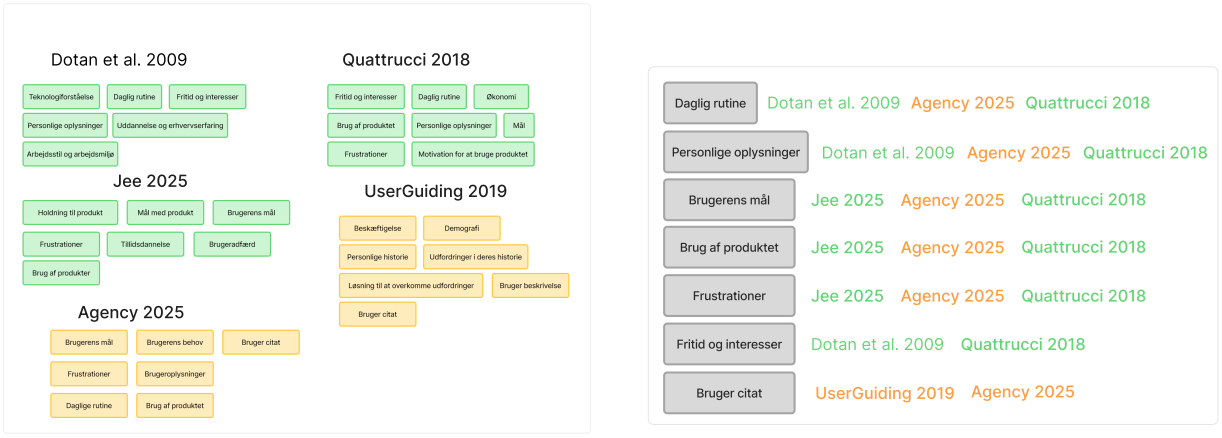
Valg af struktur af tekst

Det er til dags dato ikke blevet fundet tidligere artikler eller litteratur om personaer i samme kontekst omhandlende et hjemmesidebaseret SaaS B2B. Derfor undersøges personaer i kontekst af redesign og business-to-business hver for sig for at udvide mulig litteratur.

Undersøgelse af personaer i forbindelse med personaer i en redesign kontekst er ledt til fund af tre artikler: Dotan et al. 2009 undersøger brugen af fire kvalitative personaer i et redesign af et informationsprogram, Quattrucci 2018 har genereret tre personaer til et redesign af Public Transport Victoria App og Jee 2025 præsentere en liste af anbefalede informationer til en persona angående redesign af hjemmesider.

Angående personaer i en B2B kontekst er der fundet to artikler: UserGuiding 2019, præsentere en liste af informationer, som anbefales indraget i personaer og Agency 2025 præsentere en fulendt skabelon for en B2B persona med eksempel tekst.

For at samle et overblik over de forskellige anvendte emner, samt sammenligne emner der går på tværs af litteratur, er disse emner analyseret og samlet i figur 3.4;



FIGUR 3.4: Denne figur viser to billeder. Billedet til venstre viser et overblik over de anvendte emner i den fundne litteratur. Sammenligning af emner på tværs af litteratur kan ses på billede til højre.

Udfra figur 3.4 ses det, at Jee 2025, Agency 2025 og Quattrucci 2018 deler tre emner, mens Agency 2025 og Quattrucci 2018 deler fem emner. For at samle så bred en persona som muligt i denne første iteration, er alle syv emner valgt til struktur af persona skabelon.

Valg af billede

Ifølge en litteratur analyse af Salminen et al. 2024 er der på nuværende tidspunkt ikke konsensus om brugen af billede i en persona, men at tilvalg af billede heller ikke er en triviel

beslutning, da dette fremstiller nye problematikker om korrekt valg af billede med potentiel steoretypiske træk som alder, køn, miljø og ansigtsudtryk, som alle kan påvirke perceptionen og brugen af personaen. Dette omfatter brugen af generativ A.I., som eksempeltivs chatGPT 4.0, til at skræddersyge persona billeder, der kan tilpasse specifikke kontekster, og dermed styrke relevansen af billedet. Et studie af Salminen, S.-g. Jung et al. 2022 undersøgte brugen af A.I. genereret billeder sammenlignet med billeder af rigtige personer, og fandt, at de genererede billeder ikke havde en signifikant negativ effekt på oplevelsen af personaerne.

På baggrund af dette vil der blive forsøgt genereret persona billeder med brug af en ChatGPT 4o-billedgenereringsmodel af Naif Alotaibi, ved navn "image generator".

Opsætning af skabelon

Ud fra de tidligere nævnt litteratur, blev opsætningen præsenteret i Quattrucci 2018 og Agency 2025. Begge har billedet af personen i venstre side af opsætningen, hvor personlige informationer står enten i under billedet, eller i nederste del af billedet, hvor teksten er vist på højre side af opsætningen, og opdelt i overskrifter. Teksten er enten delvist opdelt i prosatekst og punkter Quattrucci 2018 eller udenlukkende punkter Agency 2025. Resulterende opsætning med tilhørene tekst inhold og billede kan ses i appendix IV;

evaluering af skabelon og tekstvalgt

For at evaluere opsætning af personaer, er de gennemgået sammen med en medarbejder hos MARC som sidder med MARC Helbredstillægs customer support. Gennemgåelsen fokuserede på relevans af tekstvalg, valg af billede og opsætning i forhold til, hvor virkelighedsnære personaerne virker til at være. Feedback var overvejende positiv, men med information under emnerne Daglig rutine og Interesse, som personlig udvikling, interresse og samarbejde med kollegaer, blev disse nævnt som mindre interresante informationer i kontekst af redesign. Anitas person var for ung til at virke overbevisende, og skal derfor ændres til ar være ældre, og Jonas skal ændres til at være en kvinde, da meget få mænd opleves at arbejde i denne position. Relevante elementer bestod af billede, frustrationer, personlige mål, personlige informationer og citater. Ved spørgsmål til den vigtigste persona, blev Anita nævnt som den

vigtigste.

På baggrund af denne feedback vil Jonas ændres til en kvinde og Anitas alder vil ændres. Teksten under emnerne Daglig rutine og Interresse vil ikke ændres, da der ikke er tilstrækkelig data til et alternativ. Endelige personaer kan ses i bilag 15.

3.6.9 Validering af repræsentativitet

Kvalitative personaers repræsentativitet er svært at validere, da tilgangen bygger på et smalt datagrundlag, og er derfor begrænset i både repræsentativitet og reproducérbarhed (Salminen et al. 2025). Herpå findes der flere måder at forholde sig til validiteten af personaer i litteraturen. I Cooper et al. 2014, s. 95 indgår valideringen af personaer naturligt ved at følge den præsenterede proces, da processen dermed sikrer, at de resulterende personaer er repræsentative. Albrechtsen et al. 2016 argumentere for at inddrage brugeren i hele processen af persona generereing og dermed sikre, at design og det indraget information forbliver repræsentativt. Zhang et al. 2016 brugte domæne eksperter til at validere deres fund af personaer, samt sammenlignede clickstream data med deres eksisterende personaer for at validere repræsentativiteten af begge typer af personaer. Thoma et al. 2009b argumentere for brugen af kvantitativ data gennem faktor analyse af surveys til at validere repræsentativiteten af kvantitativ data.

I (Adlin og Pruitt 2010) mener John Priutt at kvalitativ data alene ikke er nok, og at en udenlukkende kvalitativ tilgang derfor er metodisk begrænset, hvor en løsning er, at inddrage kvantitativ data, som eksempeltvis questionairs. Alan Cooper argumentere derimod for, at denne kritik stammer fra en misforståelse af mål-orienterede personaer, da han mener at de bliver fejlagtigt sammenlignet med marketssegmenter, som opdeler brugere på baggrund af demografiske og psykografiske (eng. "psychographic") faktorer, hvor personaer derimod forsøger at adskille personer på baggrund af adfærd med produktet, og repræsentere derfor ikke altid konkrete brugergrupper (Cooper et al. 2014, s. 95).

3.7 Diskussion

Ifølge Cooper et al. 2014, s. 50 kræves der mellem fire og seks interviews for at kunne bekræfte eller afkræfte repræsentativiteten af hver variabel identificeret i persona hypotesen. Med to identificeret relevante variabler (domæne ekspertise og brugeren miljø), kræver denne undersøgelse dermed mindst 8 interviews. Med tre udførte interviews kan der derfor ikke siges, om brugerne variere i højere grad på et andet område. Dog stemmer den løbende dialog og feedback med MARC-teamet overens med den indsamlet data, hvilket tyder på, at domæneviden har været til en vis grad relevant at adskille personerne på. Grundet en top-down tilgang til valg af spørgsmål, samt valg af en semi-struktureret interview, er der dog chance for at data indsamlingen har været påvirket af et bias, og den indsamlet data er formet af det data, der er ledt efter. Repræsentativiteten af disse personaer er derfor også diskutable.

Bedre repræsentativitet kunne opnås ved probability sampling, som et alternativ til opportunity sampling, samt mere åbne spørgsmål, der i mindre grad er afhængig af en hypotese. Ifølge Cooper et al. 2014, s. 49 - 52 kan dette opnås ved spørgsmål, der formes løbende, på baggrund af tidligere interviews, samt interviews der varer op til en time.

Hertil anses kvalitativ undersøgelse ydermere ikke blot som interviews, men som en etnografisk undersøgelse, der kan gøre brug af mere fordybende (eng. "immersive") aktiviteter, som observering af brug af produkt, rollespil og "show and tell-lege. Disse aktiviteter kunne yderlig støtte dataindsamlingen.

Brugen af den trinvise guide af Cooper et al. 2014 har hovedsageligt været fyldestgørende, da hvert trin har ført til næste trin på en logisk og gennemsigtig facon. Dog bliver der ikke taget højde for en kontekst-drevet udvikling af persona skabelonen. Denne del af guiden mangler overordnet set detaljer, hvilket er problematisk, da denne del kan betegnes som en kritisk del af processen ved generering af personaer (Couto et al. 2025). En senere udgivelse kunne med fordel opdatere udvikling af persona skabelonen således, at der tages højde for den givne

undersøgte kontekst, gennem en iterativ proces.

3.8 Konklussion

For at undersøge, hvordan brugere kan repræsenteres ved hjælp af en kvalitativ persona genererings tilgang, er der gjort brug af en mål-orienteret persona tilgang i form af en trinvis guide, hvor tre brugere er interviewet i forhold til deres domæneviden, frustrationer med produktet og samarbejde med kollegaer. Resultatet består af to personaer, der er evalueret i samarbejde med en repræsentant fra MARC-teamet. Antal af interviews og valg af brugere begrænser repræsentativiteten, men valgte guide har hovedsageligt vist sig fyldestgørende, med en mangel på detaljer for udvikling af en kontekst-drevet persona skabelon.

4 | Web Usage Mining Persona

Web Usage Mining defineres af Ibrahim og Obaid 2021 som indsamling af data, der kan informere om brugeres adfærdsmønstre og færden på hjemmesider. Dette er en underkategori til begrebet Web Mining (Ibrahim og Obaid 2021). Derudover dækker Web Mining over *Web Content Mining*, der referere til indsamling af hjemmesiders data, som kan være alt fra filmanmeldelser til billeder og videoer (kumar 2017), og *Web Structure Mining*, der referere til indsamling af opbygningen af hjemmesider og generelle strukturelle faktorer ved internettet (Tyagi et al. 2017). En typisk tilgang til at udføre Web Usage Mining består af fire stadier (Ibrahim og Obaid 2021). For at indsamle og analysere Web Usage data til at generere personaer, er det derfor relevant at følge disse fire trin.

4.1 1. Dataindsamling

Første trin består af at indsamle relevant data fra aktive brugere, og metoden til indsamling af data variere, efter hvilken type data der indsamles. **Web server log files** kan bestå af forespørgsler på ressourcer fra serveren (Access Log), informationer om brugerens browser og enhed (Agent Log), fejlbeskeder ved behandling af forespørgsler (Error Log) og informationer om sider, som brugeren tidligere har besøgt (Referer Log) (Varnagar et al. 2013, Ali Mohammed et al. 2024). Eksempelvis har et studie af Rawira et al. 2023 anvendt Server Log Files til at angive brugeres navigering på www.mea.gov.in ved at logge GET forespørgelser af sideskift på hjemmesiden, tidsstempel for forespørgslen samt tidligere navigeret side. Et alternativ til Web Server Log Files er brug af **Proxy Server Log Files**. Lignende informationer kan tilgås ved disse log filer. Dog kan det være svært at identificere den enkelte brugere med Proxy Server Logs, da dette typisk dokumenterer trafikken af brugere til flere servere af gangen (Varnagar et al. 2013). Grundet denne begrænsning er denne metode til indsamling ikke yderligere overvejet. Den tredje mulighed inden for dataindsamling er ved brug af **Client Side Log Files**. Her logges dataen på brugerens egen enhed. Dette muliggøre indsamling af eksempelvis scrolls, kliks og musebevægelser, da logning har adgang til den kode, som

brugeren eksekvere i browseren (Varnagar et al. 2013). Lognings placering medfører dog også, at dataen skal overføres fra brugeren computer, hvilket besværligører dataindsamlingen (Ali Mohammed et al. 2024).

4.1.1 Undersøgelse af mulig data til indsamling

For at vælge den korrekte dataindsamlingsmetode, undersøges mulige data for indsamling, på baggrund af relevans for beskrivelse af brugeradfærd. I Nguyen et al. 2020 klassificeres Web Usage Mining data i tre forskellige kategorier på baggrund af den processering, som dataen skal gennemgås, før at det kan bruges til at karakterisere aspekter af brugeren.

4.1.1.1 Inherent Features

Denne type indebærer data, som eksplicit modtages fra en brugeres session (Nguyen et al. 2020). Inherent Features (IF) består af fakta om brugeren, der ikke kræver yderligere processering for at udlede information om brugeren. Antallet af tilgængelige IF kan variere alt efter systemopsætning og software. Eksempeltvis returnerer "Combined Log Format"fra apache 2.4's web server log IF bestående af ip-adresse, tidligere besøgt side, bruger-id fra et autentificeret log-in, samt information om browser, enhedstype og operativ system (Apache 2025). Hertil kan javascript-biblioteker som Fingerprint.JS bruges til at samle browser indstillinger som skærmopløsning, font, sprogindstillinger og timezone, der kan bruges til at vurdere brugerens sprog, visningsindstillinger, land og type af enhed (Mudassar et al. 2023-3-7).

4.1.1.2 Derived Features

Denne type data er afledt af en mindre kompleks processering af informationer om brugeren, og består af brugerens handlinger, hvor eksempler kan bestå af sessionens længde, samt hvor aktiv en bruger er på baggrund af antallet af sessioner (Nguyen et al. 2020). Ydermere kan analyse af musebevægelser også anvendes til at antage læsning af tekst på en side (Kirsh 2022), samt lave antagelser om brugerens opmærksomhed generelt (Kirsh og Joy 2020). Handlinger og workflows kan beregnes gennem clickstreams bestående af enkelte kliks i kombination med en timestamp og id for sessionen og brugeren (Zhang et al. 2016) og kan kombineres med

clip-board data for mulige yderligere indsigter i adfærd (Kirsh 2020).

4.1.1.3 Latent Features

Denne type data afledes af mere komplekse processeringer. Det kan indebære udførslen af high-level opgaver på en hjemmeside og tildele en brugertype til en bruger på baggrund af gentagende adfærdsmønstre (Zhang et al. 2016), kategorisere atypisk adfærd (Nguyen et al. 2020), og opfange "usability smells"ved sammenligning af en tilsigtet brug af en hjemmeside, med brugerens reelle brug hjemmesiden (Ribeiro et al. 2019).

Nedenunder ses en tabel over de forskellige værdier som kan hentes gennem Web Server Log Files og Client Side Log Files.

Kategori	Værdi	Variabler	Indsamling
Inherent	Netværk	IP adresse	WSLF
Inherent	Tidligere besøgt side	`log_referer` (WSLF), `document.referer` (CSLF)	WSLF, CSLF
Inherent	Bruger id	`frank (%u)` (WSLF), unikt ID gemt i browseren (CSLF)	WSLF, CSLF
Inherent	Session start	`(%t)` (WSLF), timestamp (CSLF)	WSLF, CSLF
Inherent	Agent	Enhed, operativsystem og browser	WSLF, CSLF
Inherent	Browser indstillinger	Font, timezone, sprog og skærmopløsning	CSLF
Derived	Session varighed	Første anmodning minus sidste anmodning (WSLF), indlæsning af første side minus lukning af browser (CSLF)	WSLF, CSLF
Derived	Aktivhed	Antal sessioner	WSLF, CSLF
Derived	Tekst læsning	Museposition, timestamp, session ID, bruger ID	CSLF
Derived	Opmærksomhed	Museposition, session ID, bruger ID	CSLF
Derived	Handlinger/workflow	Klikket element, scrollet element, timestamp, clipboard, session ID, bruger ID, tekst læsning, opmærksomhed	CSLF
Latent	Abnormal adfærd	Handlinger/workflow, handlinger/workflow fra andre brugere	CSLF
Latent	Brugertype	Handlinger/workflow	CSLF
Latent	High-level opgaver	Handlinger/workflow	CSLF
Latent	Usability smells	Handlinger/workflow, tærskel	CSLF

Tabel 4.1: Denne tabel viser en liste over de forskellige identificerede typer af data, som er mulige at udtrække fra forskellige typer af logfiler. Kolonnen "kategori"viser dataens type, "værdi"viser dataens relevante information om brugeren, "variabler"viser det data, der kan bruges til at udlede den relevante information om brugeren og "indsamling"viser de dataindsamlingsmetoderm der kan gøres brug af, for at tilgå dataen. Her står CSLF for Server Side Log Files, og WSLF står for Web Server Log Files.

4.1.2 Valg af data til indsamling

For at kunne koble data og adfærd til en bestemt bruger, skal en given bruger kunne identificeres. Her kan der anvendes bruger-id, da et bruger-id gives ved log-in og er derfor unikt på tværs af enheder, browsers og faner. Alternative metoder kunne eksempelvis gøre brug af et device fingerprint gennem javascript, bestående af data om brugerens browser informationer, eller ip-adresse og user agent gennem Web Server Log Files (Mudassar et al. 2023-3-7). Da MARC Helbredstillæg tilgås gennem en log-in portal, kan bruger-id derfor bruges til at identificere brugen og tilkoble attributter uden yderligere processering.

Browser indstillinger og enhedstype kan informere om brugerens opsætning af MARC Helbredstillæg. Derudover kan Derived Features som session varighed og antal af sessioner bruges til at forstå, hvor ofte og hvor længe MARC Helbredstillæg bruges ad gangen. Latent features som brugertype og high-level opgaver kan bruges til at kategorisere brugerne i forskellige kategorier alt efter typer af opgaver, som der løses, samt hvilke opgaver der oftest løses af disse typer af brugere (Zhang et al. 2016). På baggrund af dette, er disse valgt til indsamling.

For at kunne udtrække disse værdier, skal der gøres brug af data angående brugerens klikket element og timestamp for klikket element. Yderligere informationer, som kan understøtte identificering af handlinger og workflows kan indebære scrolls på siden og indsamles derfor også. Derudover ville det være optimalt at indsamle timestamp for log-ud for klar identificering af varighed af sessioner. Grundet den tekniske opsætning af MARC Helbredstillæg har disse informationer dog ikke været mulige at indsamle. Det ville også være relevant at indsamle data om musebevægelser for yderligere at undersøge tekstlæsning og opmærksomhed i løbet af en session. Grundet en begrænsning i tid og ressourcer er dette data dog blevet fravalgt. Data angående brugerens clipboard kan bestå af personfølsom data som cpr-numre, og anvendes derfor ikke i dette projekt. Andre Latent Features som abnormal adfærd og usability smells kan potentielt bruges til yderligere at kategorisere forskellige adfærdsmønstre, men grundet mangel på tid, er disse ikke anvendt. En liste over de valgte informationer, der trækkes fra brugeren, kan ses herunder;

- Bruger-id
- Timestamp for log-in
- Klikket element
- Timestamp for klikket element
- Enhedstype
- Operativsystem
- Sprogindstillinger
- Fontindstillinger
- Scrolls
- Skærmopløsning

4.1.3 Valg af metode til data indsamling

For at kunne indsamle de valgte data om brugeren, er det nødvendigt at indsamle data på baggrund af Client Side Log Files (CSLF), da klikkede elementer og scrolls kræver adgang til brugerens enhed (Ali Mohammed et al. 2024). Derudover kan alle yderlige informationer om brugerens browserindstillinger og enhed samles gennem CSLF.

4.1.4 Indsamling af data

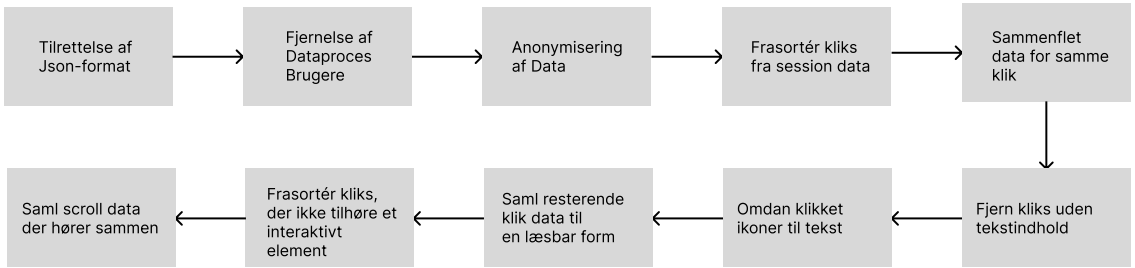
For at logge data om brugeren, er Inherent Features samlet ved start af sessionen. For at logge data om brugerens adfærd, er MARC Helbredstillæg modificeret til at logge brugerens kliks med et tilhørende timestamp. For at identificere et klikket element, logges elementets nærmeste tekstindhold, som i tilfælde af klik på knapper enten vil være knappens indholdte tekst eller ikon. Dertil er yderligere tekst i nærheden af det klikket element indsamlet til at danne en unik identifikation af hver klikket element. For at indsamle log dataen, er MARC Helbredstillæg forbundet til en intern server hos Dataproces, hvor hver log sendes øjeblikkeligt efter logning. Eksempel på format for logning kan ses i appendix V.

4.1.5 Periode for indsamling af data

Efter Dialog med Dataproces angående optimal tidsperiode til dataindsamling, erfares det, at MARC Helbredstillæg er opdelt i perioder, hvor en ny periode starter mellem d. 15 og d. 22 i måneden og slutter d. 16 til d. 21 måneden afhængig af måneden - for april måned er perioden d. 20 til d. 17 maj. I denne periode vil der typisk opstå mest aktivitet i starten af perioden, hvor aktiviteten derefter falder. Dog anbefaler Dataproces kunderne af MARC Helbredstillæg, at gennemføre kvitteringer løbende. Grundet mangel på tid er der ikke indsamlet data over en fuld periode, men forsøgt at følge perioden på bedst mulig vis. Dataindsamling fandt hertil sted over 11 hverdage fra d. 15 april til d. 5 maj, hvor der er indsamlet 30704 datapunkter.

4.2 2. Præprocessering

En afgørende del af behandlingen af Web Usage data, er præprocessering, som består af filtrering af utilsigtet data (Mittal et al. 2022). Utilsigtet data vil i denne forbindelse bestå af misklik, da alle klikkede tekstelementer vil blive gemt. For at gøre dette er der genereret en liste af klikbare elementer. Denne liste er genereret ved manuelt at trykke på alle klikbare elementer og gemme resultatet i en fil, der så bliver brugt til at frasortere alle kliks, der ikke eksistere i filen. Derudover frafiltreres aktivitet fra dataproces brugere, samt aktivitet tidligere end d. 15 maj kl 9:00, som er opstået grundet testudførelser. Endeligt samles datapunkter, som har supplerende data sendt seperat. Filer, samt kode kan ses i bilag 8.

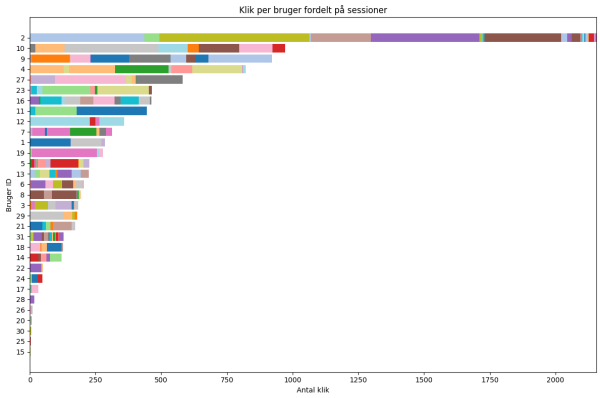


FIGUR 4.1: Denne figur viser en trinvis proces for præprocessering af den indsamlet data.

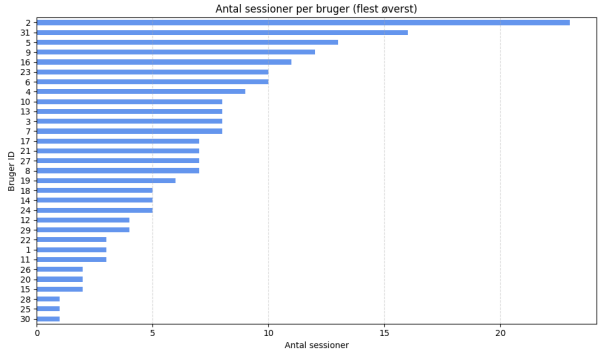
Frasortering af dataproces brugere, samt data før d. 15 maj kl 9:00 resulterede i fjernelse af 7026 datapunkter, og samling af data fjernede 1526 rækker, hvilket resultere i 21678 resterende rækker. Heraf fjernes 2203 grundet misklik og 7637 under samling af scroll elementer. Endelig liste består af **11838 elementer**, hvor **9743 er kliks** og **2095 er scrolls**. Dataen indeholder i alt **211 sessioner** af **31 brugere**, fordelt over **10 kommuner** som vist i nedenstående tabel;

Kommune ID	1	5	9	2	7	6	10	4	3	8
Antal brugere	5	5	4	3	3	3	3	2	2	1

En undersøgelse af brugers aktivhed, på baggrund af antal kliks i løbet af indsamlingsperioden, og antal sessioner, kan det ses der på figur 4.2 at enkelte kommuner står for en stor del af kliks, og at antallet af kliks variere mellem sessioner.



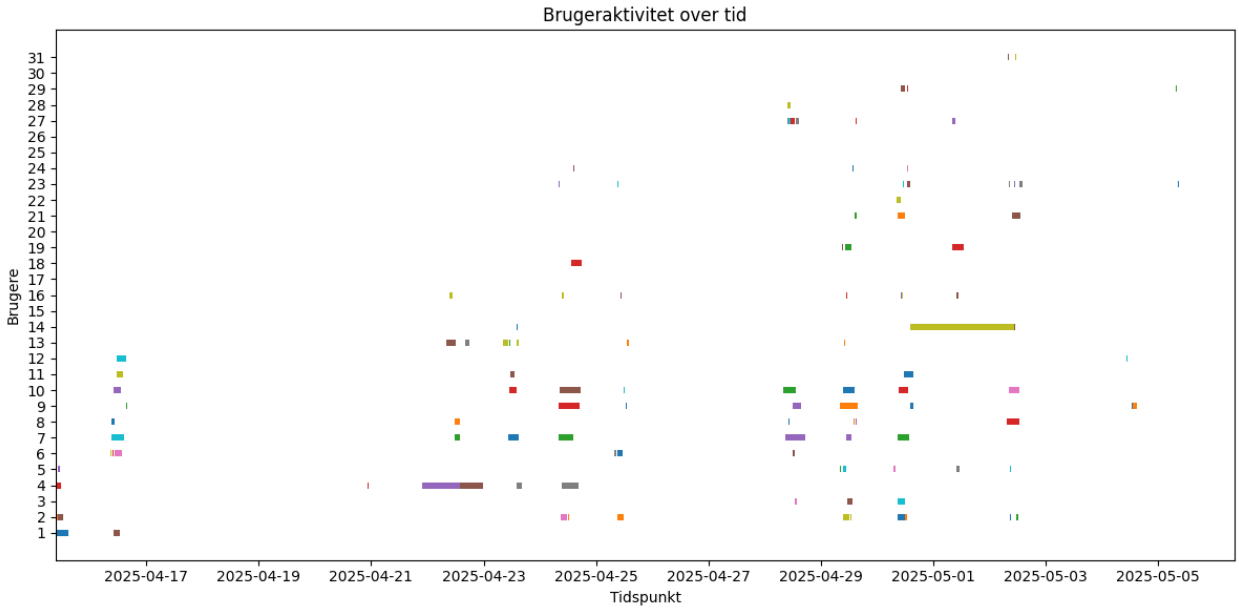
(A) Denne figur viser antal kliks per bruger, fordelt på sessioner, hvor skift i session kan ses i skift af farve.



(B) Denne figur viser antallet af sessioner per bruger.

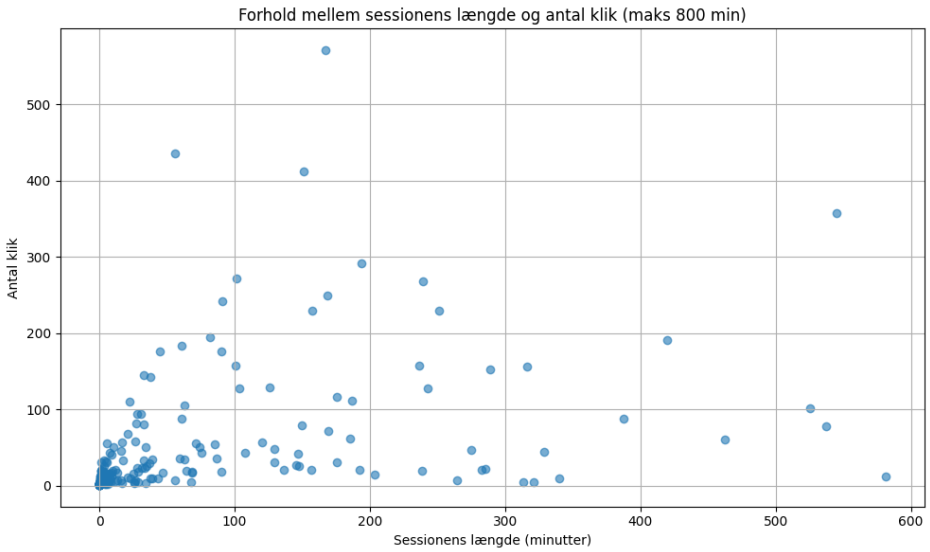
FIGUR 4.2: Denne figur består af to grafer, hvor (A) ammenligning af brugeres klikaktivitet og sessionsfordeling

Når der kigges på brugeraktivitet over tid for hver bruger på figur 4.3, på baggrund af session længde, ses der også tydelige forskelle i varigheder af sessioner.



FIGUR 4.3: Denne figur viser brugeraktiviteten for hver bruger, over hele dataindsamlingsperioden.

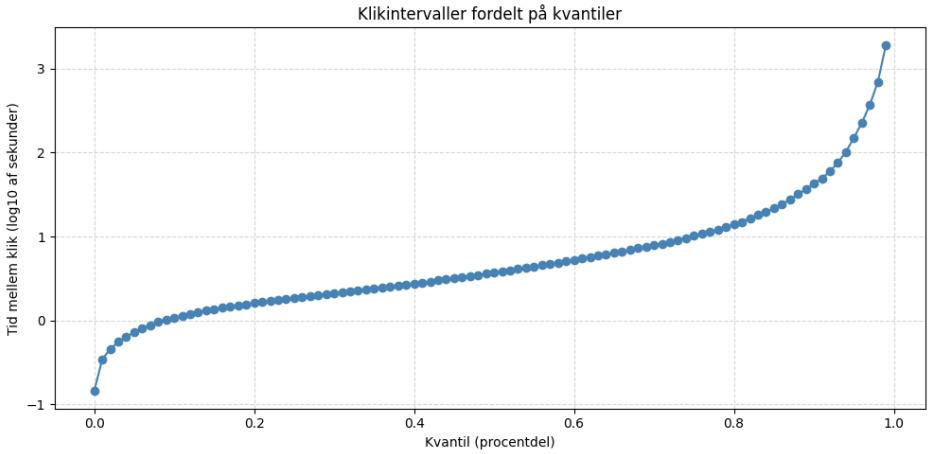
En sammenligning af antal klik per session og sessioners længde, set på figur 4.4 viser dog ikke tegn et tydeligt forhold mellem antal kliks og sessionslængde.



FIGUR 4.4: denne figur viser et punktdiagram, hvor hver punkt er en session, og placeringen af punktet er afhængig af forholdet mellem sessionens længde (set på x-aksen) og antal kliks (set på y-aksen).

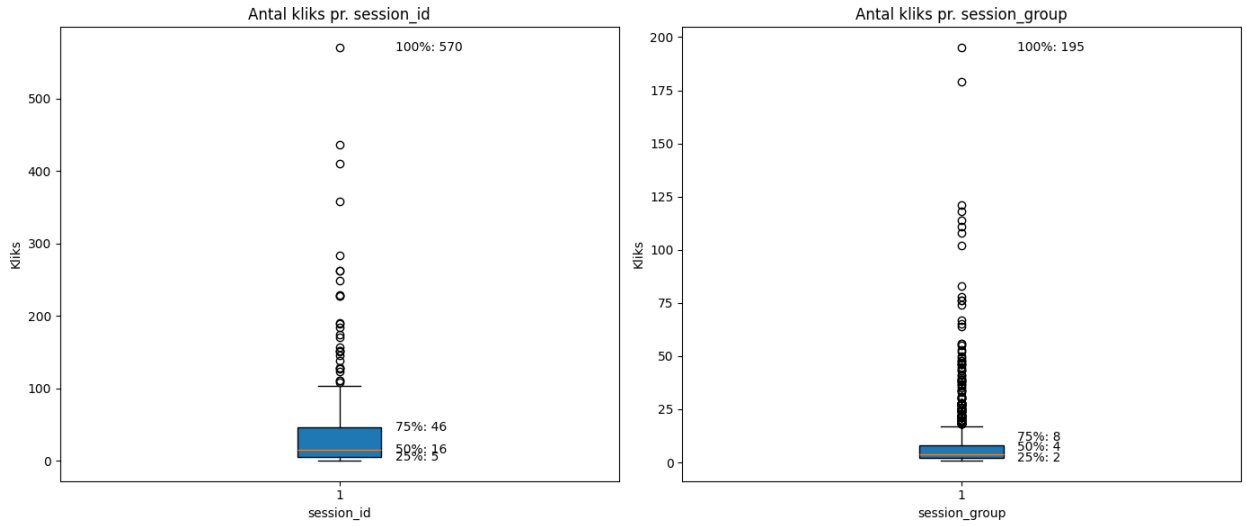
En undersøgelse af fordelingen af tidsmæssig afstand mellem kliks på baggrund af figur 4.5 viser en delvist typsik logaritmisk fordeling, dog med atypiske ender, hvor den første

kan indikere, at enten bestemte opgaver eller bestemte adfærdsmønstre adskiller sig ved at der kræves mange kliks på kort tid. Den anden ende kan indikere, at der sker et skift fra handlinger, til en hviletilstand, hvor sessionen er åben, men brugeren er inaktiv. Dette sker cirka ved percentilen 0.9, som dækker over alle kliks op til 42,7354 sekunders mellemrum.



FIGUR 4.5: Denne figur viser klikintervaller mellem kliks, hvor x-aksen er kvantiler, og y-aksen er en logaritmisk skala af sekunder.

Dette kan tyde på at brugeres adfærd enten varierer mellem hinanden alt efter, hvor hurtigt de navigere på MARC Helbredstillæg, eller at forskellige opgaver kan kræve forskellige mængder af tid mellem handlinger. Det tyder også på, at lange sessioner kan forklares ved perioder med inaktivitet, og at sessioner dermed kan indeholde flere unikke clickstreams. Dette er også set i i artiklen af Zhang et al. 2016, hvor nogle sessioner varede flere dage, og clickstreams derfor blev begrænset til at vare 24 timer. I dette datasæt er der kun en enkel tilfælde af en session, der varer mere end 24 timer. For at adskille clickstreams fra sessioner, er clickstreams derfor opdelt på baggrund af en maksimal tidsmæssig længde mellem kliks på 42 sekunder. Dette resultere i **1197 clickstreams**, hvor medianen af antal kliks per clickstream falder fra 16 til 4, som kan ses på figur 4.6.



FIGUR 4.6: Denne figur viser to boksplot. Boksplottet til venstre viser antal kliks per session, før sessioner og clickstreams er splittet. Boksplottet til højre viser antal kliks per session efter sessioner og clickstreams er splittet.

4.3 3. Opdagelse af mønstre

Dette trin består af at finde mønstre i dataen, der senere kan analyseres for indsigter om brugeren. En måde at finde mønstre, kan være ved klyngedannelse (eng. "clustering") af ensartede brugere, sider eller clickstreams (Ibrahim og Obaid 2021). Dette projekt vil fokusere på sidstnævnte, da ensartede clickstreams kan samles til et overordnet adfærdsmønster (eng. "workflow"), og dermed også forbinde en bruger til udførdslen af en bestemt type opgave, hvilket er relevant for at danne Web Usage Mining personaer (Zhang et al. 2016).

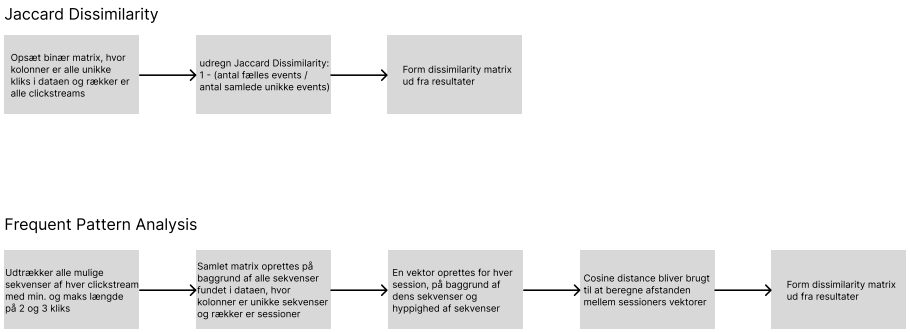
4.3.1 Klyngedannesle af clickstreams

Dette kan gøres ved at danne afstansmatrix på baggrund af forskelle i clickstreams (Zhang et al. 2016). I Zhang et al. 2016 bliver der brugt et Jaccard Dissimilarity for hver parvis sammenligning af hver clickstream (HAMERS et al. 1989). Denne metode adskiller clickstreams på baggrund af forskellen mellem antallet af unikke clicks mellem to clickstreams.
Alternative metoder til at adskille clickstreams, er tidligere belyst gennem en kortlægning af eksisterende litteratur på området af Taşgetiren og Aktas 2022b. Her er clickstreams

blandt andet blevet adskilt fra hinanden på baggrund af sandsynligheden for overgangen af et klik til et andet (Markov Chain Model), antallet af mønstre eller sekvenser, som to clickstreams har til fælles (Frequent Pattern Analysis), sekvensbaseret sammenligning af clickstreams (Pattern2Vec) og analyse og sammenligning af semantisk mening af sekvenser af ord dannet ud fra klikkede elementer (word2vec/node2vec). Herunder har embedding metoder (som eksempeltvis word2vec, node2vec og pattern2Vec) set en stigende opmærksomhed i litteraturen (Taşgetiren og Aktas 2022b). Særligt pattern2vec viser potentiale for at være en effektiv tilgang til at adskille clickstreams semantisk, på baggrund af hele sekvensen af kliks, men kræver dog også store mængder af data (Olmezogullari og Aktas 2022). Grundet mængden af data i kontekst af denne rapport, er embedding metoder derfor fravalgt.

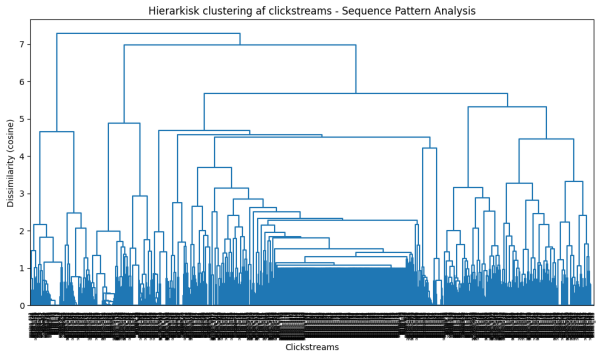
Frequent Pattern Analysis (FPA) nedbryder clickstream i mindre mønstre og sammenligner, hvilke mønstre der optræder flere gange på tværs af clickstreams (Rashidi 2014, Kapitel 5). Hertil tages der ikke hensyn til, hvilken rækkefølge mønstre optræder i. Eksempelvis vil a,b,c,d og c,d,a,b begge have mønsteret a,b og c,d tilfælles. Derfor kan den være særlig brugbar i tilfælde, hvor en clickstream indeholder flere overlappende arbejdsgange som kan være uafhængige af hinanden, men deler mange af de samme kliks. Dog er FPA mindre effektiv over for større og mere komplekse datasæt, hvor antallet af redundante mønstre stiger med kompleksitet (Jamshed et al. 2024). For at undgå dette, kan der bruges modificered FPA som eksempeltvis non-gapping FPA og closed FPA, der kan håndtere støj ved forskellige mønsterreduceringsteknikker (Wu et al. 2020). Dette kræver dog længere clickstream sekvenser, da mindre og unikke sekvenser vil frasorteres.

For at undersøge den bedste tilgang til at adskille clickstreams, vil der derfor blive forsøgt en umodificeret FPA. Resultat sammenlignes med en Jaccard Dissimilarity tilgang, da denne tilgang tidligere er brugt i forbindelse med klyngedannelse af clickstreams til persona generering (Zhang et al. 2016).

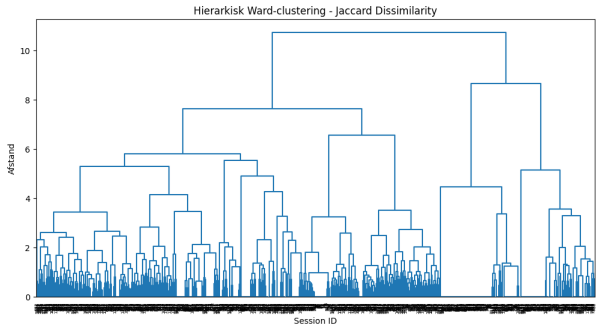


FIGUR 4.7: Denne figur viser anvendelsen af Jaccard Dissimilarity og Frequent Pattern Analysis til at adskille clickstreams fra hinanden.

Figur 4.7 viser udførslen af de to metoder, og den tilhørende kode kan findes i bilag 8. Klyngeopdeling blev dannet ved brug af Ward Hierarchical Clustering algoritme Zhang et al. 2016. Resultater ses på nedenstående figurer.

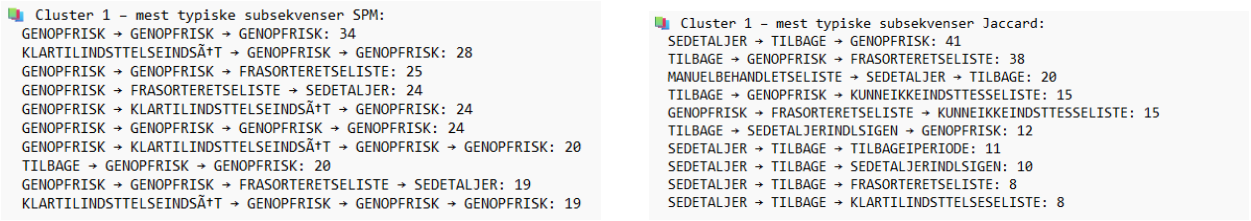


FIGUR 4.8: Denne figur viser en Ward Hierarchical Clustering af clickstream dataen, på baggrund af en Frequent Pattern Analysis adskillelse af clickstreams.



FIGUR 4.9: Denne figur viser en Ward Hierarchical Clustering af clickstream dataen, på baggrund af en Jaccard Dissimilarity af clickstreams.

Sammenligningen viser, at der opstår tydelig støj i FPA tilgangen sammenlignet med Jaccard Dissimilarity, da sekvenser, som består af gentagende klik, genererer en overrepræsentation af subsekvenser af disse gentagne klik, og bekræfter kritikken af brugen af en klassisk FPA tilgang til kompleks data som værende problematisk (Jamshed et al. 2024). En sammenligning af sekvenser fra samme klynge kan ses på figur 4.15 På baggrund af dette, er det valgt at bruge Jaccard Dissimilarity.



FIGUR 4.10: Denne figur viser de 10 mest hyppige sekvenser i klynge 1 ved en Frequent Pattern Analysis adskillelse (venstre) og en Jaccard Dissimilarity adskillelse (højre).

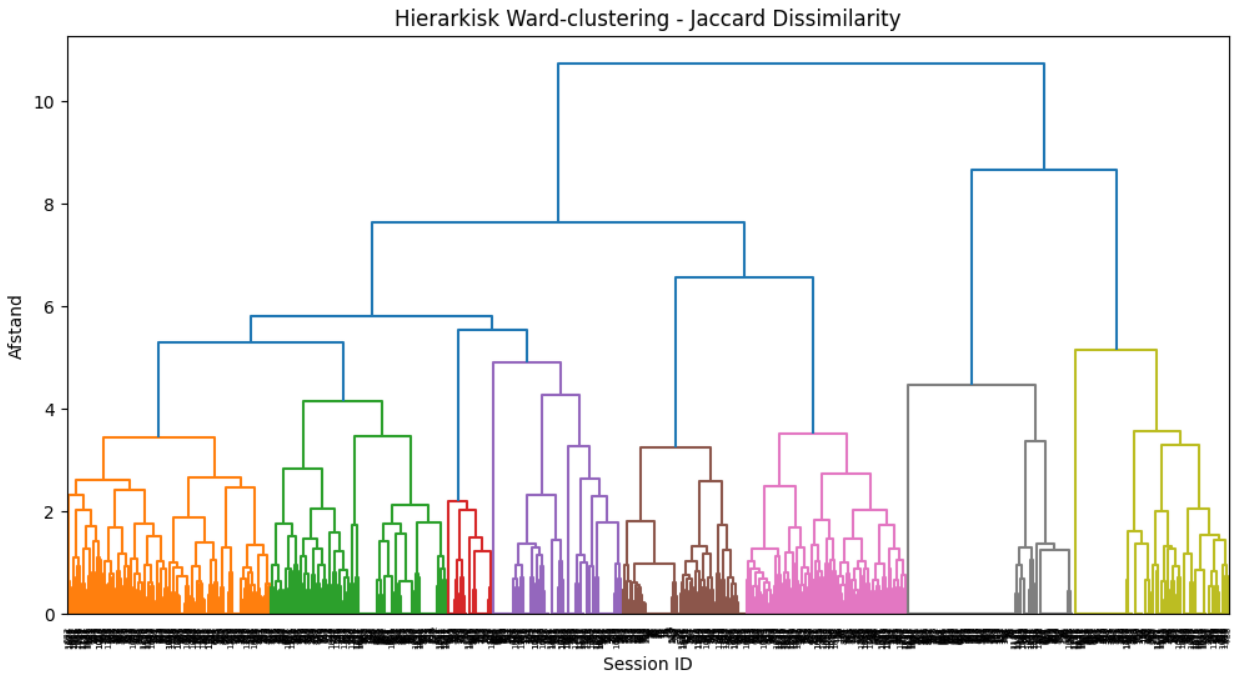
4.4 4. Analyse af mønstre

Analysen af de fundne mønstre er sidste del af Web Usage Mining procesen. Her vil analyse af mønstre bestå af at (1) analysere workflows ud fra clickstream klynger (2) vælge et antal af personaer og (3) udvikle personaer.

4.4.1 Analyse af clickstream klynger

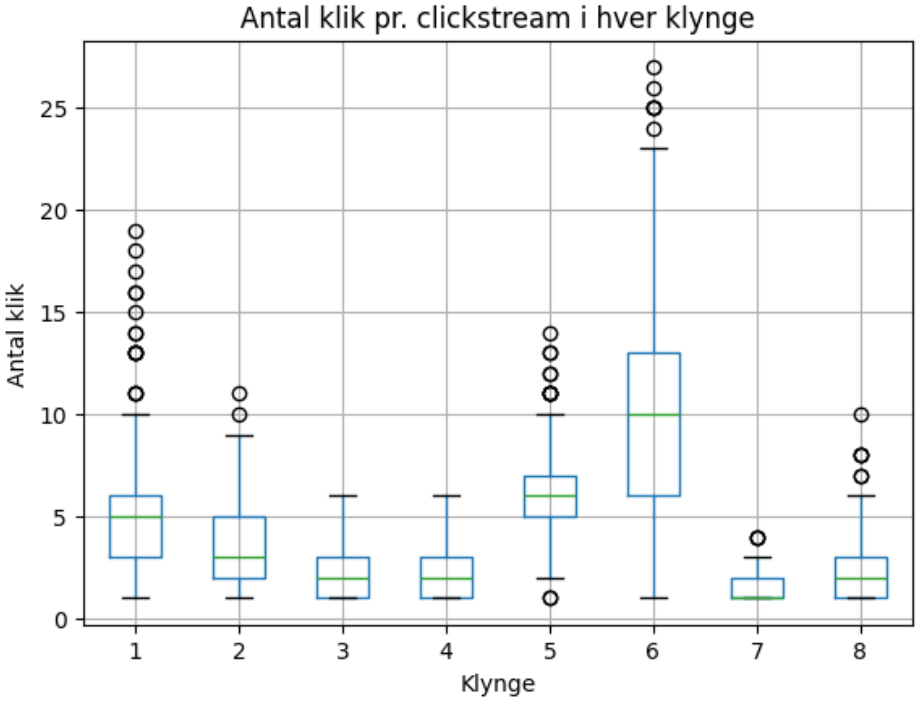
For at analysere workflows ud fra clickstream klynger, anvender Zhang et al. 2016 en eksplorativ tilgang, hvor en bestemt tærskel for grupering af klynger vælges. Typiske kliksekvenser gennemgås derefter for hver klynge, i samarbejde med en ekspertgruppe, hvor en korrekt groupering af klynger, samt beskrivelse af tilhørende workflow herpå vælges.

I dette projekt vil der blive forsøgt samme tilgang. For at gøre dette, vil en FPA tilgang anvendes til at udtrække 20 hyppigste sekvenser af hver klynge, med en makimal længde på 5 og en minimal længde på 3. Disse sekvenser vil gennemgåes for et samlet tema og sammenlignes med de resterende klynger. Gennem dette er det erfaret, at otte klynger er den maksimale opdeling, inden klyngerne begynder at miste mening. Denne opdeling kan ses på figur 4.11.



FIGUR 4.11: Denne figur viser en Ward Hierarchical Clustering af clickstream dataen, ved hjælp af Jaccard Dissimilarity, hvor otte klynger er farvet i forskellige farver. X-aksen består af clickstreams (ikke sesisoner), mens y-aksen viser afstandem mellem clickstreams, på baggrund af e jaccard Dissimilarity.

Figur 4.12, viser forskellen på længder af clickstreams for hver klynge.



FIGUR 4.12: Denne figur viser et boksplot over antallet af kliks i hver clickstream, i hver klynge.

Det overordnede tema for hver klynge, samt repræsentativ farve i figur 4.11 kan ses i følgende tabel.

Klynge	Beskrivelse	Antal clickstreams	Klyngefarve
1	Åbne lister og indsæt igen	210	Orange
2	Åbne lister uden at lave ændringer	181	Grøn
3	Tilbage/frem i periode og genopfrisk	47	Rød
4	Udenlukkende håndtering af frasorteringslisten	134	Lilla
5	Opret note og send til manuel	128	Brun
6	Ret kvitteringer og indsæt	165	Lyserød
7	Vent og genopfrisk	175	Grå
8	Upload filer	160	Gul

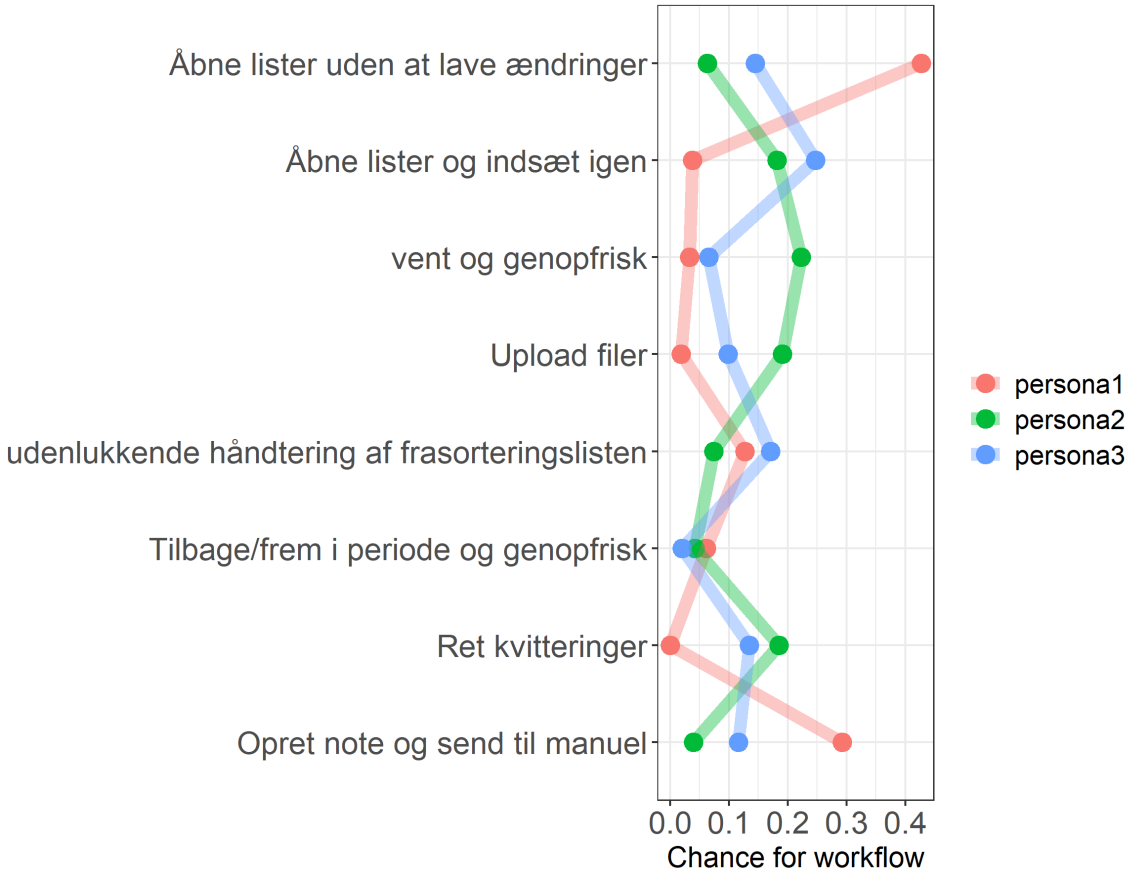
For at undersøge validiteten af denne adskillelse af workflows, er den customer support ansvarlige for MARC Helbredstillæg kontaktet og præsenteret resultaterne. Her bliver der gjort udtryk for en valid opdeling af opgaver, samt at personen typisk oplever tre typer brugere; en bruger til at uploade kvitteringer og en anden bruger til at behandle uploadede kvitteringer. Andre kommuner har en bruger, som både står for upload og håndtering af kvitteringer.

4.4.2 Valg af antal af personaer

For at bestemme antallet af personaer, er der i Zhang et al. 2016 brugt antallet af tidligere udviklet unikke personaer til at guide beslutningen. Det er i dette tilfælde ikke muligt, men tidligere nævnt samtale med Dataprces i afsnit 4.4 tyder på et antal af personaer på 3, hvor hver persona har en tydelig, dominerende workflow, som består af enten upload af filer, gennemgåelse af kvittering eller begge dele. For at undersøge dette, følges samme tilgang som i Zhang et al. 2016 til at udvinde personaer fra workflows ved hjælp af en Multinomial Mixture Model, estimeret via Expectation-Maximization (McLachlan et al. 2000). Dette er gjort ved brug af en R-pakke ved navn mixtools:

```
count_mat = create_countmat(
  data$Newuser,
  data$cluster_id,
  length(unique(data$Newuser)),
  ncluster
)
set.seed(1)  # sikre reproducerbarhed
mixfit = multmixEM(count_mat, k = 3)
personas = apply(mixfit$posterior, 1, which.max)
```

Den fulde kode kan ses i bilag 8. En eksplorativ undersøgelse viser, at flere end 3 personaer leder til enten arbitrære opdelinger, eller personaer der følger samme fordeling af chance for workflows. På baggrund af dette, samt tidligere dialog med Dataproces, er antallet af personaer valgt som værende 3. Figur 4.13 viser chancen for, at en bestemt persona udfører en bestemt type workflow.



FIGUR 4.13: Denne figur viser et profilplot over dannelse af tre personaer, hvor x-aksen er chancen for den bestemte persona at udføre en af de otte workflows, som ses på y-aksen. Chancen for udførdsel af alle otte workflows giver 1 for hver persona.

Figur 4.13 viser et profilplot over chancen for at udføre en bestemt workflow for hver persona. Her ses det, at persona 2 uploader hyppigst filer af de tre personaer, samt er den persona, der oftest retter og indsætter kvitteringer og genopfrisker repetetivt. Persona 1 uploader derimod sjældent filer, men navigere ofte uden at starte robotten og opretter ofte noter på kvitteringer og sender dem til "manuel". Persona 3 viser sig som en mere generel opdeling af workflows, med højeste chance for navigering og start af robot igen, samt udenlukkende håndtering af frasorteringslisten.

4.4.3 Metode til formidling af data

Clickstreams er tidligere nævnt i forbindelse med personaer til at evaluere repræsentativitet af valg af temaer Thoma et al. 2009a eller til at opdatere eksisterende persoaer Zhang et al. 2016,

men ikke til at spille en afgørende faktor i valg af indhold af personaer. Indenfor websikkerhed bliver clickstreams dog ofte brugt som et analytics værktøj til at identificere skadelige brugere Kamatchi og Uma 2025, hvor visualisering af user profiles spiller en betydelig rolle for effektiv brug (Nguyen et al. 2020). Derfor kan en visualisering af data, i modsætning til at beskrive data i form at tekst, være en relevant metode til at formidle Web Usage Mining personaer.

4.4.4 Valg af data til persona

For at træffe et valg om relevante emner, anvendes samme identificerede emner, som beskrevet i afsnit 3.6.8. Herfra er det valgt at fokusere på brug af produktet. Personlige oplysninger som navn og billede kan syntetiseres for at give personaen en personlighed, hvor tilvalg af billede grundes en positiv modtagelse af dette ved kvalitativ persona generering. Frustrationer kan potentielt yderligere inddrages gennem mere komplekse top-down tilgange som usability smells Ribeiro et al. 2019, men grundet en mangel på tid, er dette ikke medtaget.

Da generering af clickstreams er en direkte effekt af brugen af MARC Helbredstillæg, vil emnet omhandlende brug af produktet bestå af formidling af clickstreams gennem workflows. Formålet ved formidling af workflows, vil være at formidle personaens handlinger på en reduceret måde, for at gøre dataen nem at huske, uden at reducere dataen så meget, at den bliver redundant for læseren. Derudover vil Inherent data supplere hver persona, bestående af data om skærmstørrelse, brug af browser og operativsystem. Slutteligt vil repræsentativiteten af hver persona vises, ved at vise personaens tilhørende brugere og antal af sessioner.

4.4.5 Formidling af workflows

Formidling af sekvens data (i det her tilfælde sekvenser af workflows) er et kendt problem indenfor sekvens data analyse, da mere data oftest fører til højere kompleksitet (Munzner og Maguire 2015 - 2015, s. 14 - 16). Dog eksisterer der på nuværende tidspunkt et stort spænd af forskellige visualiseringsmetoder. Guo et al. 2022 har gennemgået en meta-analyse af artikler, som vedrører visuel analyse teknikker af sekvens data før 2020, bestående af i alt 104 unikke metoder.

Antallet af valideringer af visualiseringsmetoder er dog sparsomme. Zinat et al. 2023-06 har sammenlignet tre af disse metoder, fundet i Guo et al. 2022; sentenTree, CoreFlow og SequenceSynopsis. Disse er valgt på baggrund af brugbarhed af automatisk visualisering af data på tværs af domæner og sammenlignet på baggrund af anormalitetsdetektion (eng. "Analysis Task"), groupering (eng. "clustering") og typiske mønstre (eng. "common patterns"). Hertil fandt de, at Sequence Synposis scorede højest med henblik på opgaver indenfor disse tre områder. Nguyen et al. 2020 og Xie et al. 2021 har brugt en ekspert gruppe af analytikere til at validere brugbarhed og relevans af deres design af datavisualisering. Grundet mangel på tilgængelighed af detaljer om metoderne, er disse dog fravalgt.

4.4.6 Design af første iteration

For at finde en relevant metode til at visualisere brugen af MARC Helbredstillæg, er de resterende fremstillede 104 metoder i Guo et al. 2022 gennemgået, ved brug af http://eventvis.idvxlab.com/ (sidst tilgået 13-05-2025 kl. 09:21). Her bliver hver metode bedømt på baggrund af (1) visning af sekvenser med et lavt antal af mulige workflows, (2) visning af mange gentagende workflows og (3) visning af høj varians af længder af sekvenser af workflows.

Her er det valgt at fokusere på Visual Cluster Exploration ved brug af et Self Organizing Map (VCE-SOM) (Wei et al. 2012), da denne metode er enkel at implementer, og kan visualisere mange gentagne handlinger gennem enkelte modifikationer. Dette er gjort ved at udregne et Self-Organizing Map for hver session, bestående af en liste af workflows, ved brug af first-order Markov Chains. Derefter plottes hvert element på en graf, hvor de mest repræsentative sekvenser fylder mest i plottet. Visuel overlap håndteres ved søge et tomt felt i en spiral-lignende mønster ud fra det initierende punkt. Modifikationer fra den originale algoritme af (Wei et al. 2012) består heraf i, at gentagne sekvenser opsummeres med et tal, og dermed forkorter lange sekvenser med hyppige gentagelser, der ellers vil besværliggøre visualisering.

Grundet variansen i dataen, har det ikke været muligt at implementere størrelsesforskelle af sekvenser på baggrund af repræsentativitet. For at visualisere repræsentative mønstre i dataen, er der i stedet lavet en supplerende figur, som består af en medoid-baseret repræsentation af det workflow, der oftest opstår alene, eller de workflows, der ofte opstår sammen. Grundet den høje varians i dataen, er dette er gjort ved brug af Jaccard Dissimilarity. Dette medføre, at den medoid-baseret "typiske sekvens"ikke består af en bestemt rækkefølge af workflows, men af workflows der oftest opstår sammen. Brugte kode kan findes i appendix II og sammenfattet kode kan ses i bilag 8. De fundne medoid sekvenser for hver persona kan ses i tabel 4.2;

Persona	**Medoid Sekvens**	**Gns. Jaccard Dissimilarity**
Persona 3	[2.0, 1.0, 5.0, 1.0, 5.0, 4.0, 5.0, 6.0, 5.0, 2.0, 1.0]	0.61
Persona 2	[8.0, 7.0, 1.0]	0.62
Persona 1	[2.0]	0.47

Tabel 4.2: Denne tabel viser medoid sekvenser af workflows og tilhørende Jaccard Dissimilarity score for hver persona. Workflows repræsenteres med tal fra 1.0 til 8.0.

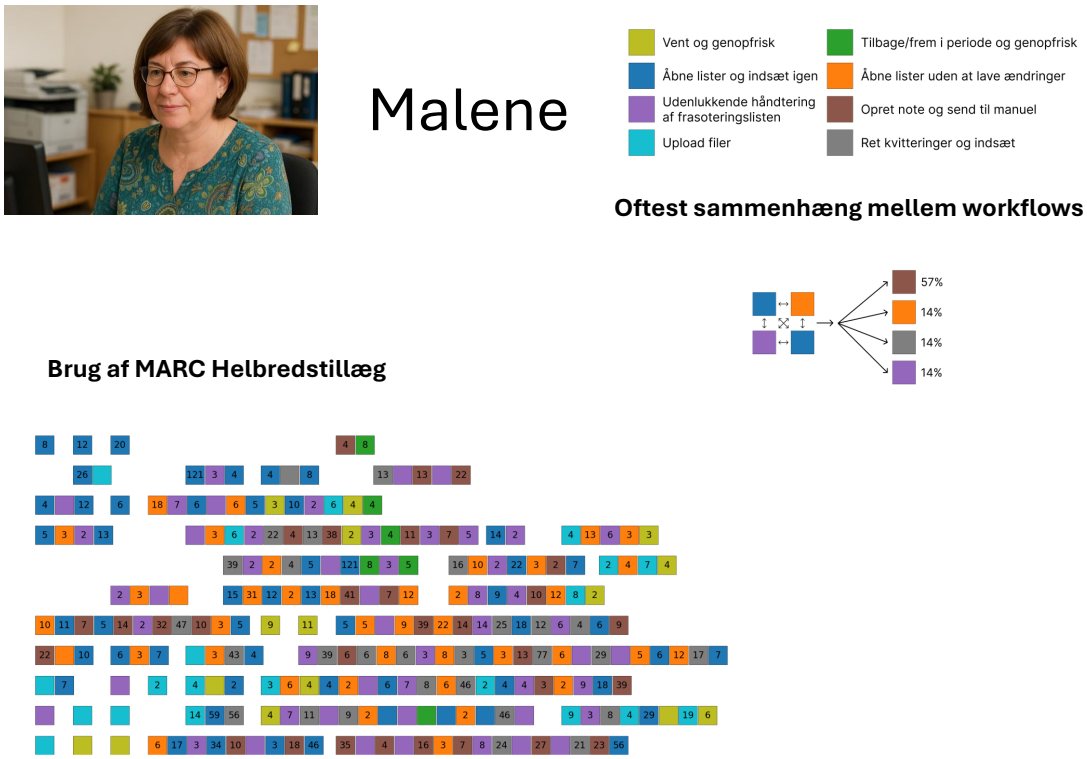
For at gøre persona 3 nemmere at visualisere, er længden på medoid sekvensen forkortet til fire workflows, hvilket giver en workflowsne [1.0, 2.0, 4.0, 1.0] med en gennemsnitlig Jaccard Distance på 0.63.

Derudover er chancen for næste workflow udregnet, ved at gennemgå alle forekomster i dataen, hvor disse workflows opstår i samme sekvens, og tælle forekomsterne for næste workflow. Resultatet af dette kan ses i tabel 4.3.

Kontekst	**Næste Element**	**Support**
[2.0]	4.0	18
	5.0	7
	3.0	4
	1.0	2
	7.0	1
[8.0, 7.0, 1.0]	7.0	3
	4.0	3
	6.0	3
	5.0	1
	3.0	1
[1.0, 2.0, 4.0, 1.0]	5.0	4
	2.0	1
	6.0	1
	4.0	1

TABEL 4.3: Denne tabel viser chancen for næste element, på baggrund af antal forekomster, hvor kolonnen "kontekst"indeholder Jaccard Dissimilarity medoid forekomsterne, kolonnen "næste element"viser de forskellige workflows, som er efterfulgt medoid og kolonnen "support"viser antallet af forekomster af disse workflows.

Et eksempel på den resulterende visualisering kan ses på figur 4.14



FIGUR 4.14: Denne figur viser et billede af persona 3, og indeholder (1) et billede, (2) navn, (3) visualisering af et plottet SOM af de sekvenser af workflows der tilhører persona 3 (venstre, nederst) og (4) visualisering af medoid baseret på en Jaccard Dissimilarity sammenligning.

4.4.7 Evaluering af iteration

For at undersøge relevans af brug af en modificeret (VCE-SOM) til visualisering, er resultatet, som vist på figur 4.14, blev gennemgået sammen med den ansvarlige for customer support for MARC Helbredstillæg. Her er der fokuseret på feedback om læsbarhed, samt relevans for valg af datavisualisering. Feedback bestod af følgende fire emner:

Beskrivelse af workflows

Feedback bestod hovedsageligt af kritik af beskrivelsen af workflows, som værende for uspecifikke, og at forklaringen af workflows er af høj interresse, da disse kan indeholde informationer om mere konkredt brug af MARC Helbredstillæg, som har høj relevans for den nuværende diskussion om MARC Helbredstillægs design. Herunder blev der omtalt forholdet mellem "indlæs igen"og "indsæt direkte", som er to måder at starte behandlingen af en kvittering,

men hvor "indsæt direkte"forårsager problemer for brugeren.

Relevans af SOM visualisering

Der blev gjort udtryk for relevansen af SOM visualiseringen, på baggrund af at forstå rækkefølgen af handlinger, som enten en bekræftelse eller afkræftelse af den nuværende forståelse af brugerens adfærd. Dog blev rækkeopdelingen af sessioner misforstået som værende tilhørende samme bruger, hvor disse rækkeopdelinger i virkeligheden ikke har nogen betydning for hverken brugere eller rækkefølgen for sessionernes forekomst. Løsning kunne bestå i at undgå at placere sessioner på en synlig række. Grundet tid har dette dog ikke været muligt at implementere. Dog blev der gjort udtryk for et ønske om yderlige overbliksdannelse over adfærd, eksempeltvis ved en figur, der viser hyppigheden af workflows over alle sessioner.

Relevans af medoid visualisering

Medoid visualiseringen blev omtalt som en konstruktiv supplement til SOM visualiseringen, hvor personen fokuserede på selve medoid samlingen, og ikke frekvenserne for efterfølgende workflows.

Yderligere kommentarer

Personen udtrykte et behøv for supplerende tekst til hver graf, der kunne opsummere de mest relevante karaktertræk for personaen, da disse blev anset som værende besværlige at analysere på baggrund af SOM-visualiseringen alene. Personen viste en overvejende positiv respons til iterationen, og mente at den præsenterede data kunne have en bredere relevans for forståelsen af MARC Helbredstillægs brugere, uden at kunne beskrive en nærmere relevans for brugeren i en konkret design kontekst.

4.4.8 Design af anden iteration

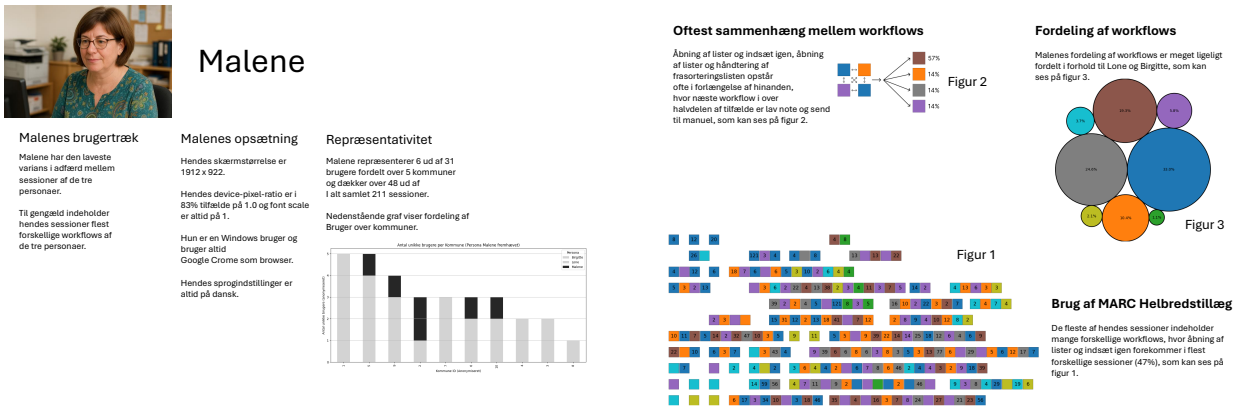
For at imødekomme ønsket om en yderligere specificeret beskrivelse af hver workflow, er hver workflow gennemgået ved at undersøge frekvensen af kliks indenfor indlæsning/indsættelse af kvitteringer og brug af lister, da disse er efter gennemgang af første iteration anvist som værende særlig interresante. Derudover er der tilføjet en kort beskrivelse og første, samt tredje

kvartil for antal kliks og længde af hver workflow.

Personaen splittes yderligere i tre sider, hvor side 1 består af (1) et billede, (2) opsummering af personaens brugertræk (3) opsætning, bestående af Inherent Features som typisk skærmstørrelse, typisk device-to-pixel ratio, valg af browser, computertype og browser indstillinger som font og sproginstillinger (4) repræsentativitet af personaen, bestående af antallet af brugere som indgår i personaen, antallet af sessioner, samt hvilke kommuner disser brugere er fra.

Side 2 består af SOM visualisering og medoid visualisering med sammenhængende tekst, der beskriver mest synlige træk fra visualiseringerne, samt en tredje figur, bestående af en Packed Circle Plot, hvor hver cirkels radius er defineret ud fra forekomsten af hvert workflow i hele personaens datasæt.

Side 3 består af en uddybende beskrivelse af hver workflow, bestående af en overskrift, kort beskrivelse, forekonster af relevante kliks i forbindelse med workflowet (som eksempeltvis hyppigheden af brugen af "indsæt direkte"og "indlæs igen") og slutteligt statestik for antal klik for workflowet og workflow varighed. Anden iteration af alle tre personaer kan ses i bilag 11. Eksempel på side 1 og side 2 for anden iteration af persona 3 kan ses på figur 4.15. Side 3 kan ses i appendix VI.



FIGUR 4.15: Denne figur viser side 1 og side 2 af anden iteration af persona 3, hvor side 1 ses til venstre og side 2 ses til højre.

4.4.9 Evaluering af iteration 2

Grundet tid har det ikke været muligt at gennemgå iteration 2 med repræsentative fra teamet bag MARC Helbredstillæg for yderligere tilrettelser.

4.5 Diskussion

Dette afsnit vil diskutere processen vedrørende generering af Web Usage Mining personaer, og vil berøre valg af dataindsamlingsmetode, identificering af workflows, analyse af klynger, muligheder ved brug af interaktive personaer og valg af visualiseringsmetode.

4.5.1 Brug af en CSLF tilgang til dataindsamling

Brugen af Client Side Log Files (CSLF) har vist sig effektiv til at indsamle data for brugerens interaktioner med MARC Helbredstillæg. Det har her været muligt at indsamle størstedelen af brugerens kliks på interaktive elementer på hjemmesiden. Dog har denne tilgang også vist sig bekostelig at implementere, da front-end frameworks, i dette tilfælde Flutter, er begrænset i redskaber til bruger-tracking. I modsætning til WSLF, har CSLF også krævet opsætning infrastruktur, bestående af en server til modtagelse af sendte logs, der yderligere

sætter krav til brugen af ressourcer. Dette bekræfter kritikken af CSLF som værende ressourcekrævende (Ibrahim og Obaid 2021). Dog har CSLF været i dette tilfæde påkrævet for at generere Web Usage Mining personaer på baggrund af tilgangen præsenteret af Zhang et al. 2016, da kliks på elementer, der ikke ikke sender en forespørgsel til server-siden, kræver adgang til klient-siden af programmet. Dette er dog afhængig af programmets opbygning, og programmer, der er opbygget med henblik på bruger-tracking kan have potentiale til at nedskære ressourcer til tracking betydeligt, ved enten af påtvinge et krav om forespørgsel ved hver brugerinteraktion og dermed muliggøre en WSLF tilgang til at indsamle alt interaktion, eller opbygge programmet med henblik på bruger-tracking til at starte med.

CSLF har dog fortsat en potentiel relevandt indsamlingsmetode til at undersøge brugeradfærd, da metrikker som scrolls og musebevægelser kan have potentiale til at danne en detaljeret betegnelse af brugerens adfærd, og yderligere specificere workflows. Dette potentiale har dog desværre ikke været muligt at undersøge i denne rapport grundet begrænset tid.

4.5.2 Identificering af workflows

I Zhang et al. 2016 består identificering af workflows, af klyngedannesle af clickstreams og en efterfølgende analyse af klynger.

Klyngedannelse

Siden udgivelsen af Zhang et al. 2016 er der kommet mere fokus på embedding metoder til at analysere big data, hvor eksempeltvis pattern2vec er designet specifikt til at adskille sekvens data, som clickstreams, og viser evne til at opdele data på et mere dynamisk grundlag end lineære algoritmer (Olmezogullari og Aktas 2022). Dette kunne potentielt overkomme problematikker observeret i denne artikel, som håndtering af stor varians af clickstream længder, og lige vægtning mellem interresante og redundante mønstre, hvor vægtningen er ukendt fra start, grundet en buttom-up tilgang. Dog kræver embedding metoder typisk mere data end modeller som Jaccard Dissimilarity og FPA (Reyhani Hamedani og Kim 2021). I tilfælde af systemer med få brugere, som MARC Helbredstillæg, vil mere data kræve at indsamle data over en længere periode.

Brugeradfærd kan dog ændre sig over tid, og gammel data kan derfor miste repræsentativitet. Men en løsning kunne hertil være online neurale modeller for klyngedannesle, der opdatere opdeling af workflows løbende, samt glemmer gammel data (Yu og Webb 2019).

En alternativ fremgangsmåde kunne også bestå af en top-down klyngedannelse, ved at samle sekvenser af kliks i meningsfulde handlinger på baggrund af tilsigtet brug, og dermed sikre, at "overgangsklik"som i det her tilfælde eksempeltvis er "tilbage"og "gem", ikke er vægtet lige så højt, som "indsæt direkte"og "indlæs igen". En top-down tilgang til analyse af clickstreams bruges allerede til at automatisere usability testing (Harms et al. 2014) og indikering af potentielle brugsproblemer (Ribeiro et al. 2019). Top-down klyngedannelse kan dog potentielt lede til at negletiere ukendte brugsmønstre og brugeradfærd, og er derfor også biased mod at finde den adfærd, som allerede er kendt (Electrical og Electronics Engineers 2019-6). Valg af metode til klyngedannelse skal derfor tage højde for både mængden af data, variansen af clickstream længder og eksisterende kendskab til brugerens adfærd.

Hertil er Jaccard Dissimilarity ikke et optimalt valg i forbindelse med MARC Helbredstillæg, da metoden både negleciere rækkefølge af kliks, hvilket besværliggøre tolkning af adfærd, og bedømmer clickstreams forskel på clickstreams på baggrund af længde, hvilket adskiller data på en potentiel arbitrær grundlag.

Analyse af klynger

En essentiel del ved tilgangen, er at kunne omdanne clickstream klynger til en meningsfuld, generalisérbar adfærd (eng. "common workflow") på et datadrevet grundlag. Dog er der i (Zhang et al. 2016) en mangelende struktureret tilgang, som på nuværende tidspunkt består af at gennemgå typsike sekvenser og derfra analysere et generelt tema frem. Derudover er nuværende metode afhængig af domæneekspert inddragelse for validering, hvilke kan resultere i forkastelse af adfærd, som ellers kan understøttes af data, hvis adfærden ikke accepteres af eksperterne (Electrical og Electronics Engineers 2019-6). For at skabe en gennemgående, data-dreven tilgang, mangler der derfor at yderligere undersøge metoder til at analysere

og fortolke workflows ud fra klynger. Dette er afgørende at opnå, for at effektivisere valget af antal af klynger, og dermed muliggøre undersøgelsen flere opdelinger. Ved lav vaians af clickstreams i klyngeopdelingerne, kunne det være relevant at anvende en modificeret FPA eller Markow Chain til at syntetisere en typisk sekvens af kliks. Ved høj varians kunne det være relevant at anvende neurale embedding metoder som word2vec, som opfatter kliks som ord, og sekvenser som sætninger, der behandles semantisk (Electrical og Electronics Engineers 2020-12-10).

4.5.3 Statiske og interaktive personaer

Følge samtalen fra afsnit 4.4.7 bliver gennemsigtighed anset som en vigtig faktor for brugbarheden af Web Usage Mining personaer i kontekst af MARC Helbredstillæg. En måde at tilføje mere gennemsigtighed kunne være at indføre interaktive elementer i personaen, som eksempeltvis at åbne workflows og se bagvedlæggende kliks, eller funktionalitet til at spore forskelle ved brugere under samme persona. Interaktiv datavisualisering er tidligere brugt i web sikkerhed, hvor brugergrupper eksempeltvis kan bestå af flere bruger profiler (Nguyen et al. 2020).

4.5.4 Visualisering af workflows

På baggrund af litteratur indenfor anomalitetsdetektion (eng. "anlomali detection") antages det, at et successcriterie for Web Usage Mining personaer er at kunne visualisere brugeradfærd grafisk (Nguyen et al. 2020). Dog har det været uklart, hvilken form for visualisering der vil være optimal for at gøre dette. Ved evaluering af brugen af en modificeret VCE-SOM tilgang, har den tilhørende medoid visualisering ført til mest omtale, hvilket kan indikere, at denne har været effektiv til at formidle personaen, mens SOM-visualiseringen har været kompleks at fortolke for personen.

For at træffe et bedre grundlag for valg af visualiseringsmetode, vil det være relevant at undersøge flere typer af visualisering. Dette kunne gøres, i forbindelse med MARC Helbredstillæg, ved at inddrage MARC-teamet i designprocessen af personaen, på samme

vis som med Co-Designet Personaer (Albrechtsen et al. 2016), da persona skabelonen burde designes på baggrund af slutbrugeren (L. Nielsen 2019b).

4.6 Konklusion

For at undersøge, hvordan brugere kan repræsenteres ved hjælp af en Web Usage Mining persona genererings tilgang, er clickstreams indsamlet for 10 brugere over en periode på 11 hverdage og opdelt i otte klynger i samarbejde med MARC-teamet, hvor et typisk workflow er tolket ud fra hver klynge. Tilgangen har dog vist sig tidskrævende, grundet manglende effektiv bruger-tracking på klient-siden, samt manglende metode til automatisk generering af typisk workflow ud fra clickstream klynger. Analyse af workflows, samt dialog med MARC-teamet, har resulteret i tre personaer, hvor design af persona skabelon, bestående hovedsageligt af visualisering af workflows, er udviklet i samarbejde med en repræsentant fra MARC-teamet.

5 | Sammenligning af personaer

I Salminen et al. 2025 præsenteres en påstand om, at personificerede analytisk data er bedst til stratetisk beslutningstagen, mens individuel brugerstatestik er bedre til automatiske beslutningstagen. Dette er særligt relevandt i forbindelse med sammenligning af en kvalitativ persona generering og en persona generering på baggrund af Web Usage Mining, da sidstnævnte skal være i stand til at kunne hjælpe med stratetisk beslutningstagen, og ikke blot automatisk beslutningstagen. Dermed former denne påstand også rammerne for undersøgelser af brug af personaer, som et spørgsmål om, hvornår personaer er brugbare, og ikke om, hvor vidt personaer er brugbare eller ej. Med dette, medfølger også en stærk afhængighed af kontekst for brugen af personaer.

i Kerr og Kelly 2025 bliver brugen personaer undersøgt, ved at udføre workshops og derefter spørge ind til holdninger til deltagernes holdning til personaerne. På denne måde kunne deltagere forholde sig til personaer på baggrund af erfaring med anvendelse af dem.

For at deltagerne skal opnå samme erfaring med brug af de udviklere personaer, inden de skal forholde sig til dem, vil en lignende tilgang vil følges her. dette gøres ved at opstille brugssituationer med personaerne.

Anvendelse af personaer variere bredt, og afhænger af både kontekst og type af persona (Salminen, Wenyun Guan et al. 2022). Bødker et al. 2012 undersøgte brugen af personaer, ved at opsætte en design workshop med tre interresenter, og præsenterede citater i printet form. Madsen et al. 2010 undersøgte personaer i forbindelse med scenarier, ved at præsentere en start på en historie og derefter observere personaens rolle i udvikling af brugsscenarier. Salminen, S.-G. Jung et al. 2020 sammenligner analycs og AGP gennem en opgave, som er givet individuelle deltagere, der omhandler identificering af brugersegmenter.

Da nuværende problemstilling for MARC teamet er, at prioritere ønsker korrekt i forhold

til effekt, som beskrevet i afsnit 2.1, vil opgaven bestå i at sortere ønskerne på baggrund af størst positive effekt på bredeste målgruppe, og dermed også efter relevans. For at holde opgaver relevant til konteksten af situationen, er der indgået kontakt med customer support ansvarlige for MARC Helbredstillæg for kunders nuværende ønsker til ændringer. Her er der identificeret seks brugerønsker.

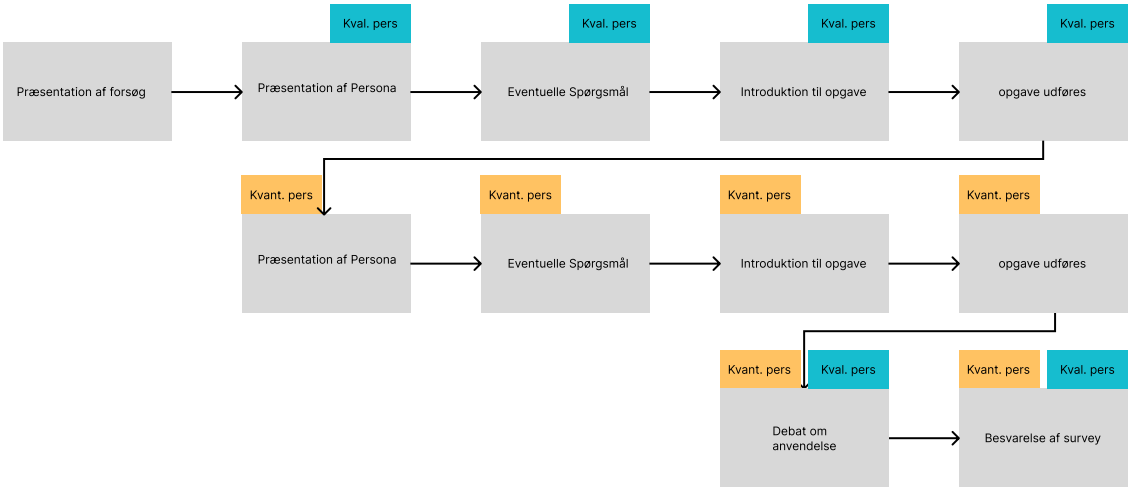
For at finde svar på, i hvilken forbindelse en kvalitativ og en Web Usage Mining baseret persona vil blive brugt af MARC-teamet, vil der efterfølgende gennemføres en åben debat, hvor alle fire medlemmer har mulighed for at tage stilling til muligheder og begrænsninger ved begge typer personaer.

For at eksekvere debatten, bliver tre spørgsmål præsenteret for teamet, følge konstruktet (eng. "construct") "vilje for anvendelse"(eng. "Willingness to use"), identificeret i artiklen af Salminen, Santos et al. 2020. Her vil konstruktet "vilje for anvendelse"dække det fundamentale spørgsmål stillet i Salminen et al. 2025 angående, hvordan personaer vil blive brugt. Dette gøres ved at oversætte konstruktets tre underspørgsmål til dansk, og omformulere dem til åbne spørgsmål, som vist i tabel 5.1.

Tabel 5.1: Denne tabel viser oversættelsen og omformuleringen af tre underspørgsmål relateret til konstruktet "vilje for anvendelse".

Spørgsmål (engelsk)	**Åbent spørgsmål (dansk)**
I would make use of this persona in my task.	Hvordan forestiller du dig, at du kunne bruge denne persona i forbindelse med MARC Helbredstillæg?
I can imagine ways to make use of the persona information in my task.	Hvilke måder kunne du forestille dig at bruge informationerne fra personaen i arbejdet med MARC Helbredstillæg?
This persona would improve my ability to make decisions about the customers it describes.	Kunne denne persona hjælpe dig med at træffe beslutninger om de kunder, den beskriver?

Efterfølgende vil de fire medlemmer besvare en survey, udviklet af Salminen, Santos et al. 2020, som vedrører relevante konstrukter, der er forbundet med en personas kvalitet; Overenstemmelse (eng. "consistency"), fuldkommenhed (eng. "completeness"), troværdighed (eng. "credibility"), klarhed (eng. "clarity"), ligesindethed (eng. "similarity"), sympatisk (eng. "likability") og empatisk (eng. "empathy"). Denne survey er oversat til dansk, hvor oversættelserne kan ses i appendix VII. Den endelige plan kan ses på figur 5.1.



FIGUR 5.1: Denne figur viser en gennemgang af mødet, samt hvilke personaer der er tilstede under hver del af mødet.

Besvarelse af surveys kan ses i appendix VIII, samt rådatan kan findes i bilag 13. Grundet mangel på tid er denne data dog ikke yderligere undersøgt.

5.0.1 Udførdsel af mødet

Alle medlemmer af MARC-teamet, bestående af en junior udvikler (JU), en senior udvikler (SU), en team leder (TL) og en ansvarlig for customer support (CS) har haft mulighed for at deltage i en workshop på to timer. Ingen har tidligere haft kendskab til personaer, pånær CS, som har lært om personaer i et kursus på universitetet. En lydoptagelse af mødet vil sikre fuld dokumentation af udtalelser. Præsentationen af mødet vil indeholde en kort forklaring af formålet med mødet, samt en gennemgang af mødet. Præsentation af personaer

vil foregå ved først at uddele en personatype i printet format på et A4 ark til hver deltager, da dette bidrager til en aktiv brug af personaerne (Bødker et al. 2012). Derefter gennemgås hver information på hver personaer, gennem en visning af hver persona i en powerpoint præsentation. Efterfølgende bliver deltagerne spurgt om de har spørgsmål til enten det de har set i forbindelse med personaerne eller til mødet. Anvendte Powerpoint kan findes i bilag 14.

Introduktion til opgaven med kvalitativ persona vil bestå af en beskrivelse af opgaven, som værende en sorteringsopgave, og gennemgang af hver brugerønske. Brugerønskerne er printet på hver sit stykke papir, og hver deltager har hver sin kopi af brugerønskerne. Dette gøres for at give deltagerne mulighed for både at arbejde med sorteringsopgaven enkeltvis, og sammen med hinanden. Udførslen af opgaven er fuldendt, når alle er enige om en sortering, og når rangeringen af brugerønskerne kan forklares ved argumentation.

Introduktion og udførsel af en opgave med Web Usage Mining persona vil forløbe ligeledes introduktion og udførdslen af opgaven med kvalitativ persona. Dog vil de kvalitative personaer fjernes fra deltagerne inden opgaven begynder. Dette skal minimere effekten af den kvalitative persona til sorteringensogaven med Web Usage Mining personaen.

Debat om anvendelse vil foregå ved at præsentere et udsagn, som præsenteret i tabel 5.1, og derefter observere udtalelser. Når samtalen stopper med at producere nyt indhold, vil næste udsagn præsenteres.

Efter alle tre udsagn, uddeles en survey til hver deltager med spørgsmålene fra tabel VII.1, hvor hver spørgsmål er forbundet et tekstfelt og en 9-trins likert skala.

5.0.2 Analyse af data

Grundet mangel på tid er der blevet valgt kun at analysere debatten. Grundet deltagelse af den Customer Support ansvarlige under debatten, der også har været en aktiv del af sparring om persona generering, er det relevant at undersøge hele mødet med MARC-teamet ved brug af en diskurs analyse, da personaer før har vist sig og blive brugt til at promovere personlige

agendaer (Salminen, Wenyun Guan et al. 2022). Diskurs analyse kan også belyse, hvordan personaerne bliver brugt i argumentationen af rangering af brugerønsker, i forbindelse med magtforholdet mellem de forskellige medlemmer af MARC-teamet (Wodak og Meyer 2016). Grundet mangel på tid har dette dog ikke været muligt. Debatten vil i stedet blive analyseret på baggrund af udtalelser om personaerne, ved brug af tematisk analyse.

Dette er gjort ved at transkribere lydoptagelse for debatten, og derefter kode teksten og lave temaer på baggrund af kodningen. Kodningen og temaer er efterfølgende gennemgået for tilrettelse og endeligt opsummeret på baggrund af temaer.

Tematisk analyse resulterede i syv emner, besåtende af Idéer til brug af statistik med MARC Helbredstillæg, Brug af kvantitativ persona, Brug af kvalitativ persona, Kritik af persona, Sætte ting i perspektiv, Sammenligning og eksisterende antagelser. Tematisk analyse kan findes i bilag 12.

Disse syv emner er derefter skrevet til en sammenfattet tekst, set i afsnit 5.1.

5.1 Resultat

Under brugen af WUM persona opstod der flere idéer til at bruge user analytics. Idéer bestod af (1) et mail system der automatisk hjalp nye bruger,ved at gøre dem opmærksomme på oversete funktioner;

TL: "*I løsningen kunne man godt sætte et eller andet smart op, som sætter automatisk en mail ind og siger, "hey, vi ser at du udenlukkende bruger indsæt direkte"*".

(2) automatisk opdagelse af brugerfejl;

CS: "*det kan man selvfølgelig ikke måle på den her type data, men hvis der pludselig bliver uploaded mange af dem ved en fejl, så indikerer det jo, at der er et behov*".

(3) brug af kliks til at indikere påpulære og upopulære funktioner;

CS: "*Hvis man også kunne måle på, at den ikke blev brugt, så kunne man jo også sige, okay, så er det ikke den, vi skal videreudvikle lige nu, eller så er det den, vi skal videreudvikle*".

(4) brug af tid som metrik til at måle brugeren;

JU: "*hvor lang tid der er inde på alle metrikker, som du kan måle, som er antallet af kliks, forskellige steder, og også det her med de sekunder, og hvordan ting tager, og hvor lang tid de har været brugt*".

(5) måling af effekt på brugens antal af kliks og varighed af workflows ved nye ændringer af MARC Helbredstillæg;

TL: "*At når du prøver at rette i noget, at vi så egentlig kan se, om vi kan både få antallet af kliks og varighed ned ved hjælp [...] når vi laver det her, så påvirker det sådan her*".

Disse idéer giver udtryk for interesse i muligheden for at automatisere beslutninger, ved brug user analytics, som at sende mail til kunder der bruger systemet forkert og opdage brugerfejl.

Brugen af WUM personaerne bestod delvist af at fortolke workflows. Her blev workflowet "vent og genopfrisk"fortolket til at brugeren ikke syntes robotten er hurtig nok; SU: "*Det eneste, jeg tolkede ved det, det var, at hun ikke syntes, den gjorde det hurtigt nok*". Der blev også kortvarigt tolket på sekvenser af workflows, hvor hyppigheden af en bestemt sekvens kan give ny viden om, hvor meget en bestemt opgave fylder for brugeren;

CS: "*det der med bare at rette, kvitteringer og indlæst igen, hvor det potentielt kommer i sådan et loop, ikke? Hvor tit er det at de gør det? Det lader jo til, at det er ret meget her*".

Workflows blev dog i de fleste tilfælde omtalt i forbindelse med metrikker, som antal kliks og varighed; CS: "*kunne jo også måske være, kunne vi spare dem for clicks?*", JU: "*Og det er også noget, man kan få ud af at se, hvis der er noget, de klikker rigtig meget på*". På baggrund af dette, er den observerede brug af WUM personaerne bestået af at tolke brugerens intentioner ud fra workflows, samt tolke hyppighed af bestemte brugeropgaver, som at uploade og rette kvitteringer, ud fra sekvenser af workflows. De tilhørende metrikker til workflows blev også brugt til at folholde sig kritisk overfor den bestemte workflow; TL: "*13-56 kliks, det er jo fuldstændig galt. Jeg ved ikke engang, hvordan I kan få 56 kliks til at være på én kvittering*".

Ved brug af kvalitative personaer blev der diskuteret løsninger på baggrund af personaernes holdninger til IT. Eksempeltvis nævnte SU, at personaen Britta er bange for at lave fejl, og det derfor kunne være smart med bedre fejlbeskeder;

SU: "*Jeg vil sige bedre fejlbeskeder [...] Var det ikke Britta, der var lidt nervøs for at lave fejl? [...] Der tænker jeg, at jo mere hjælp systemet giver, til hvorfor noget fejler, så gør det måske dem mindre nervøs.*".

Herunder blev mængden af tilgængelig data i MARC Helbredstillæg også diskuteret, i forbindelse med forskelle i personaernes behov;

CS: "*er det så personaer, som skal have præsenteret alt data, eller skal vi måske tænke, at det også kan forvirre dem? Og så skal vi afgrænse det lidt, så det kun er det vigtigste, de får*".

Et vigtigt element ved den kvalitative persona blev også nævnt som værende informationer, der kan bruges til bedre at møde kunder;

CS: "*den kvalitative har rigtig meget at sige i forhold til både, hvordan vi skal møde vores kunder og brugere i supporten, og i undervisning af implementeringer, men også i forhold til at videreudvikle det*".

Her består møde med kunden både af undervisning og support, men også af at hjælpe nye kunder i gang; TL: "*hvordan sørger vi for at tage bedst muligt imod dem, og hjælpe dem i gang med løsningen, baseret på, hvilken person de er*". Hertil blev der også nævnt en mere indirekte fordel ved at anvende kvalitative personaer, da interviews og kundekontakten hjælper styrke relationen til kunderne; TL "*Jeg har det påstået, at der kommer en større gevinst ud af det, ved at have snakket*".

I løbet af debatten blev en fordel ved den kvalitative persona nævnt flere gange af CS, som værende at kunne sætte indkomne brugerønsker i perspektiv, i forhold til de resterende brugere; CS: "*er kunne være så meget potentiale i, at man lige kaster den op, og så lige tænker, okay, hvordan harmonerer man med de her personaer*". Denne fordel blev nævnt i forbindelse med at effektivisere beslutningstagen; "*Så jeg tror, der var noget effektivisering i, at vi bruger persona, tanken, ideen*", og i forbindelse med at dele viden om brugere på tværs af team-medlemmer; CS: "*Det her er den måde også at eksplicitere den viden, der faktisk er*".

Både den kvalitative og WUM personaer blev nævnt i forbindelse med team-medlemmernes eksisterende antagelser om brugerne. Den kvalitative udfordrede deres antagelser om brugeren, da der er en antagelse om, at de fleste brugere er ligesom de brugere, der tilhører Gentofte Kommune; CS: "*Fordi, at jeg ville tænke, at Gensofte var det, vi skulle regne med*". Dertil nævner CS en styke i at få udfordret disse antagelser. JU nævner også at kvalitative personaer bekræfter formålet med MARC Helbredstillæg, da personaen Anitas ønsker på arbejdspladsen stemmer overens med det, MARC Helbredstillæg tilbyder;

JU: "*Så det endeligt kan automatiseres, skal automatiseres, det er jo næsten sådan essensen af det her produkt [...] Men så bliver man jo også bare, hvad skal man sige, bekræftet i det*".

Under debatten blev der både nævnt kritik af WUM personaerne og de kvalitative personaer. Kritik af kvalitative personaer bestod af manglende repræsentation af særligt problematiske kunder; TL: "*vi mangler lidt gentoftepersonaen [...] Det er nok den, vi vil få mest ud af at kunne hjælpe*"samt en mistillid til, om personaens udtalelser stemmer overens med brugernes

holdninger; TL "*spørgsmålet er jo egentlig så om der er overensstemmelse med hvad de siger, og hvordan de egentlig har det*".
Kritik af WUM personaerne bestod af, at personaerne ikke indeholder nok evidens;

SU: "*Altså jeg kan godt lide den her en lille smule (kvantitativ persona). Men problemet det er at, man ved jo faktisk ikke om de her ting her det er godt eller skidt. Og det kræver, evidens*",

samt risiko for at misfortolke personaer, og dermed træffe forkerte beslutninger;

SU: "*hvis man kun kigger på det her (WUM personas), så kan man godt risikere at lave fejl, hvis man ikke er klog*".

Ved sammenligning af WUM og kvalitativ persona, blev det også nævnt, at de kvalitative er bedre til at forklare brugeroplevelser;

TL: "*kvalitative data kan måske forklare deres oplevelse lidt bedre end det kvantitative. Fordi de glæder og frustrationer, som de har, kan vi nødvendigvis ikke se indplykket*".

Hertil blev det også givet udtryk for en fælles enighed om, at WUM og de kvalitative personer supplere hinanden, og at det for dem ikke er et spørgsmål om enten eller; JU: "*Jeg tror ikke noget, der hedder kun, når vi snakker data generelt. Der vil man jo altid gøre begge dele*", men at i denne forbindelse også er en enighed om, at de kvalitative personaer er mere fordelagtige end WUM personaerne; JU: "*Ja, vi vil jo ikke kunne undvære det kvalitative data*". Her bliver de kvalitative personaer nævnt som værende gode i forbindelse med et nyere produkt, med brugere der ikke minder om dem selv; TL: "*Det er jo stadig et nyt produkt. Det er jo stadig en ny gruppe af mennesker, som ikke minder om os. Og derfor tror jeg, den kvalitativ er bedst*", mens WUM personaerne er gode til at forbedre eksisterende kunders oplevelse; CS: "*Og den kvantitative er vel mere, når de så er kommet i gang med løsningen, hvordan kan vi så forbedre deres oplevelse*".

5.1.1 Sammenligning af resultater med litteraturen

Der opstod flere idéer til at bruge user analytics, ved omtalen af WUM personaerne, som eksempeltvis at automatisere feedback til brugerne om brugen af MARC Helbredstillæg. Dette støtter udsagnet af Salminen et al. 2025 om user analytics som værende mest passende til automatiserede beslutningstagen. Det bekræfter derudover også relevansen af automatiserede tiltage til at kunne fange brugerfejl eller utiltænkt brugeradfærd, som eksempeltvis ved brug af usability smells (Ribeiro et al. 2019), da dette også blev nævnt som en af idéerne til brug af user analytics.

Dårlige eller sparsommelige beskrivleser gør, at personaer ikke bliver brugt, men når personaer er succesfulde, hjælper de udviklere med at forstå brugere og deres behov (Billestrup, Stage, Nielsen et al. 2014). Der bliver set en tydelig forskel på WUM og den kvalitative persona i denne forstand, hvor den kvalitative persona blev omtalt som værende informativ, mens WUM personaerne blev kritiseret for at kræve for meget analyse af læseren, og kan føre til forkerte beslutninger ved mistolkning. Dette er en essentiel kritik af WUM personaerne, da hovedformålet er at kunne træffe kritiske beslutninger om produktet på samme vis som med andre typer personaer (Salminen et al. 2025), og viser, at det valgte WUM persona design mangler kritiske ændringer for at kunne opnå dette.

Et af disse kunne bestå af information om brugeren, der hjælper læseren med at forstå dem. Her bliver der nævnt informationer som glæder og frustrationer. Dette bekræfter også tidligere antagelser om disse persona informationer som værende vigtige i forbindelse med redesign (Cooper et al. 2014, Dotan et al. 2009).

Flere artikler har nævnt clickstream data som en supplement til en persona (Zhang et al. 2016, Salminen et al. 2025, L. Nielsen 2018). Ved debatten er der også observeret denne holdning til WUM personaerne som et supplement til de kvalitative personaer. Dette bekræfter også denne antagelse. Dog bliver der også nævnt, at de kvalitative personaer er de "essentielle personaer", hvor WUM personaerne er supplerende.

Der er observeret en generel positiv modtagelse af den kvalitative persona. Dette bekræfter også en fortsat relevans af kvalitative personaer i en verden, der i stigende grad gør brug af big data og user analytics (Salminen et al. 2025).

5.1.2 Personaer og kundeforhold

En overraskende kommentar om kvalitative personaer, har været dens indirekte effekt af at styrke kundekontakten. Hertil er brugerinterviews anset som en måde at skabe et forhold til brugerne, og ikke blot som en opgave, der skal gennemføres for at indhente data. Dette fremstiller også en essentiel fordel ved at forbigå brugerkontakt ved WUM personaer, som en ulempe. Hertil bliver der dog også nævnt, at WUM personaer kan supplere kvalitative personaer ved produkter med et stort antal brugere.

5.2 Konklussion

For at undersøge, Hvordan en kvalitativ baseret persona og en Web Usage Mining persona vil blive modtaget under en designproces af et digitalt produkt, er alle medlemmer af MARC-teamet inviteret til en debat om brugen af personaer i forbindelse med MARC Helbredstillæg. Her fortrækkes en kvalitativ persona, som begrundes med personaens læsbarhed, relevans af informationer og brugbarhed i relevante kontekster, som består af prioritering af brugerønsker, mødet med kunden og hjælp til opstart af nye kunder. WUM personaerne blev anset som et relevant supplement til kvalitative personaer, men brugskontekst bestod af automatisering af beslutningstagen, og ikke af kritisk beslutningstagen, som ellers adskiller personaer fra user analytics. Valgte visualiseringsmetode af WUM persona data blev derudover kritiseret for at kræve for meget analyse at anvende, hvilket indikere et behov for et mere opsummerende persona skabelon design.

Bibliografi

A/S, Dataproces. 2024a. „Organisationen: Professionalisme, innovation og samarbejde“. Hentet 3. marts 2025. https://www.dataproces.dk/organisationen.

———. 2024b. „Organisationen: Professionalisme, innovation og samarbejde“. Hentet 3. marts 2025. https://www.dataproces.dk/losninger/marc/marc-helbredstill%C3%A6g.

Abelein, Ulrike og Barbara Paech. 2015. „Understanding the Influence of User Participation and Involvement on System Success – a Systematic Mapping Study“ [**inlang**eng]. *Empirical software engineering : an international journal* (New York) 20 (1): 28–81. ISSN: 1382-3256.

Adlin, Tamara og John Pruitt. 2010. *The essential persona lifecycle : your guide to building and using personas* [**inlang**eng]. 1. udgave. Chantilly: Morgan Kaufmann/Elsevier. ISBN: 0123814189.

Agency, The WIN-LOSS. 2025. „Using Qualitative Interviews to Refine Personas“. Hentet 3. marts 2025. https://win-loss.agency/post/using-qualitative-interviews-to-refine-personas/.

Albrechtsen, Charlotte, Majbrit Pedersen, Nicholai Friis Pedersen og Tine Wirenfeldt Jensen. 2016. „Proposing Co-Design of Personas as a Method to Heighten Validity and Engage Users: A Case from Higher Education“ [**inlang**eng]. *International journal of sociotechnology and knowledge development* 8 (4): 55–67. ISSN: 1941-6253.

Ali Mohammed, Mohammed, Rula A. Hamid og Reem Razzaq AbdulHussein. 2024. „Data Collection and Preprocessing in Web Usage Mining: Implementation and Analysis“ [**inlang**ara ; eng]. *Iraqi journal for computers and informatics (Online)* 50 (2): 54–74. ISSN: 2313-190X.

Apache. 2025. „Log Files“. Hentet 3. marts 2025. https://httpd.apache.org/docs/2.4/logs.html.

Association for Computing Machinery, issuing body., author. 2019. „Design Issues in Automatically Generated Persona Profiles“. I *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval /,* 225–229. CHIIR '19: Conference on Human Information Interaction and Retrieval. New York,NY,United States : Association for Computing Machinery, ISBN: 1-4503-6025-4.

Bano, Muneera og Didar Zowghi. 2013. „User involvement in software development and system success: a systematic literature review“ [**inlang**eng]. I *ACM International Conference Proceeding Series,* 125–130. New York, NY, USA: ACM. ISBN: 1450318487.

Billestrup, Jane, Jan Stage, Anders Bruun, Lene Nielsen, Kira S. Nielsen, Cristian Bogdan, Regina Bernhaupt, Marco Winckler, Stefan Sauer og Peter Forbrig. 2014. „Creating and Using Personas in Software Development: Experiences from Practice“ [**inlang**eng]. I *Human-Centered Software Engineering,* 8742:251–258. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 3662448106.

Billestrup, Jane, Jan Stage, Lene Nielsen og Kira Storgaard Nielsen. 2014. „The use of personas in software development: Advantages, obstacles and experiences in practice“ [**inlang**English]. I *Proceedings IRIS36: August 11-14 2013 at Gran, Norway,* redigeret af Tone Bratteteig, Margunn Aanestad og Espen Skorve. The 36th Information Systems Research Conference in Scandinavia (IRIS 36) ; Conference date: 11-08-2013 Through 14-08-2013. Department of Informatics, University of Oslo.

Brickey, J., S. Walczak og T. Burgess. 2012. „Comparing Semi-Automated Clustering Methods for Persona Development“ [**inlang**eng]. *IEEE transactions on software engineering* (LOS ALAMITOS) 38 (3): 537–546. ISSN: 0098-5589.

Brinkmann, Svend. og Lene Tanggaard. 2025. *Kvalitative metoder : en grundbog* [**inlang**dan]. 4. udgave. Kbh: Hans Reitzels Forlag. ISBN: 9788702412413.

Bødker, Susanne, Ellen Christiansen, Tom Nyvang, Pär-Ola Zander, Jesper Simonsen, Yanki Lee, Heike Winschiers-Theophilus, Kim Halskov og Keld Bødker. 2012. „Personas, people and participation: challenges from the trenches of local government“ [**inlang**eng]. I *PDC*

2012 Embracing New Territories of Participation. proceedings of the 12th Participatory Design Conference, August 12-16, 2012, Roskilde University, Denmark / Volume 1, Research papers, 1:91–100. New York, NY, USA: ACM. ISBN: 9781450308465.

Cooper, A. 1999. *The inmates are running the asylum : why high-tech products drive us crazy and how to restore the sanity.* [**inlang**eng]. Indianapolis, Ind: Sams. ISBN: 0672316498.

Cooper, Alan, Robert Reimann, David Cronin, Christopher Noessel, Jason Csizmadi og Doug LeMoine. 2014. *About face: the essentials of interaction design, fourth edition* [**inlang**eng]. 4th ed. Wiley. ISBN: 1118766407.

Couto, Fábio, Mariana Curado Malta, Dylan D. Schmorrow, Hirohiko Mori, Masaaki Kurosu, Ayako Hashizume, Yumi Asahi og Cali M. Fidopiastis. 2025. „Contributions for the Development of Personae: Method for Creating Persona Templates (MCPT)“ [**inlang**eng]. I *HCI International 2024 – Late Breaking Papers,* 15374:3–22. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland. ISBN: 3031768027.

Dotan, Amir, Neil Maiden, Valentina Lichtner, Lola Germanovich, Tom Gross, Philippe Palanque, Raquel Oliveira Prates et al. 2009. „Designing with Only Four People in Mind? – A Case Study of Using Personas to Redesign a Work-Integrated Learning Support System“ [**inlang**eng]. I *Human-Computer Interaction - INTERACT 2009,* 5727:497–509. Lecture Notes in Computer Science 2. Germany: Springer Berlin / Heidelberg. ISBN: 9783642036576.

Electrical, Institute of og issuing body. Electronics Engineers author. 2019-6. „Hierarchical Clustering for Discrimination Discovery: A Top-Down Approach“. I *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) /,* 187–194. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). Piscataway, NJ : IEEE. ISBN: 9781728114897.

———. 2020-12-10. „Representation of Click-Stream DataSequences for Learning User Navigational Behavior by Using Embeddings“. I *2020 IEEE International Conference on Big Data (Big Data) : proceedings : Dec 10 - Dec 13, 2020, virtual event /,* 3173–3179. 2020

IEEE International Conference on Big Data (Big Data). Piscataway, New Jersey : IEEE, ISBN: 1-7281-6252-1.

Floyd, Ingbert R., M. Cameron Jones og Michael B. Twidale. 2008. „RESOLVING INCOMMENSURABLE DEBATES: A PRELIMINARY IDENTIFICATION OF PERSONA KINDS, ATTRIBUTES, AND CHARACTERISTICS“ [**inlang**eng]. *Artifact (London, England)* 2 (1): 12–26. ISSN: 1749-3463.

Garcia, Jorge Esparteiro og Ana C. R. Paiva. 2018. „Manage Software Requirements Specification Using Web Analytics Data“. I *Trends and Advances in Information Systems and Technologies,* redigeret af Álvaro Rocha, Hojjat Adeli, Luís Paulo Reis og Sandra Costanzo, 257–266. Cham: Springer International Publishing. ISBN: 978-3-319-77712-2.

Guest, Greg, Kathleen M MacQueen og Emily E Namey. 2012. *Applied Thematic Analysis* [**inlang**eng]. 1. udgave. Thousand Oaks: SAGE Publications, Incorporated. ISBN: 1412971675.

Gunatilake, Hashini, John Grundy, Rashina Hoda og Ingo Mueller. 2024-09. „The impact of human aspects on the interactions between software developers and end-users in software engineering: A systematic literature review“. *Information and software technology.* ([Amsterdam] :) 173. ISSN: 0950-5849.

Guo, Yi, Shunan Guo, Zhuochen Jin, Smiti Kaul, David Gotz og Nan Cao. 2022. „Survey on Visual Analysis of Event Sequence Data“ [**inlang**eng]. *IEEE transactions on visualization and computer graphics* (LOS ALAMITOS) 28 (12): 5091–5112. ISSN: 1077-2626.

HAMERS, L, Y HEMERYCK, G HERWEYERS, M JANSSEN, H KETERS, R ROUSSEAU og A VANHOUTTE. 1989. „SIMILARITY MEASURES IN SCIENTOMETRIC RESEARCH - THE JACCARD INDEX VERSUS SALTON COSINE FORMULA“ [**inlang**eng]. *Information processing management* (OXFORD) 25 (3): 315–318. ISSN: 0306-4573.

Harms, Patrick, Jens Grabowski, Cristian Bogdan, Regina Bernhaupt, Marco Winckler, Stefan Sauer og Peter Forbrig. 2014. „Usage-Based Automatic Detection of Usability Smells“

[**inlang**eng]. I *Human-Centered Software Engineering,* 8742:217–234. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 3662448106.

Ibrahim, Kareem K. og Ahmed J. Obaid. 2021. „Web Mining Techniques and Technologies: A Landscape View“ [**inlang**eng]. *Journal of physics. Conference series* (Bristol) 1879 (3): 32125–. ISSN: 1742-6588.

Jamshed, Aatif, Bhawna Mallick og Rajendra Kumar Bharti. 2024. „A Systematic Review on Sequential Pattern Mining-Types, Algorithms and Applications“ [**inlang**eng]. *Wireless personal communications* (New York) 138 (4): 2371–2405. ISSN: 0929-6212.

Jansen, Bernard J., Soon Gyo Jung, Lene Nielsen, Kathleen W. Guan og Joni Salminen. 2022. „How to Create Personas: Three Persona Creation Methodologies with Implications for Practical Employment“ [**inlang**eng]. *Pacific Asia journal of the Association for Information Systems* (Atlanta) 14 (3): 1–28. ISSN: 1943-7536.

Jee, Sarah. 2025. „Creating User Personas prior to a website redesign“. Hentet 3. marts 2025. https://getwiththebrand.co/creating-user-personas-prior-to-a-website-redesign/.

Jung, Soon-gyo, Joni Salminen, Haewoon Kwak, Jisun An og Bernard J. Jansen. 2018. „Automatic Persona Generation (APG): A Rationale and Demonstration“ [**inlang**eng]. I *Chiir '18: Conference on Human Information Interaction and Retrieval,* 2018-:321–324. New York, NY, USA: ACM. ISBN: 1450349250.

Kamatchi, K. og E. Uma. 2025. „Insights into user behavioral-based insider threat detection: systematic review“ [**inlang**eng]. *International journal of information security* (Berlin/Heidelberg) 24 (2): 88–. ISSN: 1615-5262.

Kerr, Jeremy og Nick Kelly. 2025. „Use of personas in co-designing learning experiences with teachers: An exploratory case study“ [**inlang**eng]. *International journal of technology and design education* (Dordrecht) 35 (1): 189–207. ISSN: 0957-7572.

Kirsh, Ilan. 2020. „What Web Users Copy to the Clipboard on a Website: A Case Study“. I *Proceedings of the 16th International Conference on Web Information Systems and*

Technologies, 303–312. 16th International Conference on Web Information Systems and Technologies. SCITEPRESS - Science / Technology Publications. ISBN: 9789897584787.

Kirsh, Ilan. 2022. „Virtual Finger-Point Reading Behaviors: A Case Study of Mouse Cursor Movements on a Website“ [**inlang**eng]. *Big data research* (AMSTERDAM) 29:100328–. ISSN: 2214-5796.

Kirsh, Ilan og Mike Joy. 2020. „Splitting the Web Analytics Atom: From Page Metrics and KPIs to Sub-Page Metrics and KPIs“ [**inlang**eng]. I *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics,* 33–43. New York, NY, USA: ACM. ISBN: 9781450375429.

kumar, Anurag. 2017. „A Study on Web Content Mining“ [**inlang**eng]. *International journal of engineering and computer science,* ISSN: 2319-7242.

Laubheimer, Page. 2020. „3 Persona Types: Lightweight, Qualitative, and Statistical“. Hentet 24. februar 2025. https://www.nngroup.com/articles/persona-types/.

Madsen, Sabine, Lene Nielsen, Rikke Orngreen, Torkil Clemmensen, Dinesh Katre og Pradeep Yammiyavar. 2010. „Exploring Persona-Scenarios - Using Storytelling to Create Design Ideas“ [**inlang**eng]. I *Human Work Interaction Design: Usability in Social, Cultural and Organizational Contexts,* bind AICT-316, 57–66. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 3642117619.

Matias, Sheila Marie Mobo. 2023. „A Comprehensive Summary on Category of Web Usage Mining“ [**inlang**eng]. I *ACM International Conference Proceeding Series,* 288–292. New York, NY, USA: ACM. ISBN: 9798400709227.

McLachlan, Geoffrey J., David Peel og Geoffrey J. McLachlan. 2000. *Finite mixture models* [**inlang**eng]. Wiley series in probability and statistics. Applied probability and statistics section. New York ; Wiley. ISBN: 0471006262.

Mittal, Ruchi, Varun Malik, Jaiteg Singh, Vikram Singh, Amit Mittal, Paulo J. S Gonçalves, Yashwant Singh, Pradeep Kumar Singh, Maheshkumar H Kolekar og Arpan Kumar Kar.

2022. „Web Usage Mining—Process, Tools and Practices“ [**inlang**eng]. I *Recent Innovations in Computing,* 832:449–457. Lecture Notes in Electrical Engineering. Singapore: Springer. ISBN: 9789811682476.

Mudassar, Muhammad, Muhammad Ali, Atif Ali, Zulqarnain Farid, Rabia Asif, Muhammad Hassaan Mehmood og Abdul Salam Mohammed. 2023-3-7. „An Analysis of Browser and Machine Fingerprinting Techniques“. I *2023 International Conference on Business Analytics for Technology and Security (ICBATS).* 2023 International Conference on Business Analytics for Technology and Security (ICBATS). IEEE. ISBN: 9798350335651.

Munzner, Tamara og Éamonn Maguire. 2015 - 2015. *Visualization analysis design* [**inlang**eng]. 1st edition. A K Peters visualization series. Boca Raton, Florida: CRC Press. ISBN: 9780429088902.

Nguyen, Phong H., Rafael Henkin, Siming Chen, Natalia Andrienko, Gennady Andrienko, Olivier Thonnard og Cagatay Turkay. 2020. „VASABI: Hierarchical User Profiles for Interactive Visual User Behaviour Analytics“ [**inlang**eng]. *IEEE transactions on visualization and computer graphics* (LOS ALAMITOS) 26 (1): 77–86. ISSN: 1077-2626.

Nielsen, Jakob. 2008. „Site Map Usability“. Hentet 3. marts 2025. https://www.nngroup.com/articles/site-map-usability/.

Nielsen, Lene. 2018. „Design personas - new ways, new contexts“ [**inlang**eng]. *Persona Studies* 4 (2): 1–4. ISSN: 2205-5258.

———. 2019a. „A Slice of the World“ [**inlang**eng]. I *Personas - User Focused Design,* 27–38. Human–Computer Interaction Series. United Kingdom: Springer London, Limited. ISBN: 9781447174264.

———. 2019b. „Finding Connections“ [**inlang**eng]. I *Personas - User Focused Design,* 39–53. Human–Computer Interaction Series. United Kingdom: Springer London, Limited. ISBN: 9781447174264.

Nielsen, Lene. 2019c. „Introduction: Stories About Users“ [**inlang**eng]. I *Personas - User Focused Design,* 1–25. Human–Computer Interaction Series. United Kingdom: Springer London, Limited. ISBN: 9781447174264.

———. 2019d. *Personas - User Focused Design* [**inlang**eng]. Second edition. Human–Computer Interaction Series. London: Springer Nature. ISBN: 1447174275.

Nielsen, Lene, Kira Storgaard Hansen, Jan Stage og Jane Billestrup. 2015. „A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations“ [**inlang**eng]. *International journal of sociotechnology and knowledge development* (Hershey) 7 (1): 45–61. ISSN: 1941-6253.

Olmezogullari, Erdi og Mehmet S. Aktas. 2022. „Pattern2Vec: Representation of clickstream data sequences for learning user navigational behavior“ [**inlang**eng]. *Concurrency and computation* (HOBOKEN) 34 (9). ISSN: 1532-0626.

Palomino, Fryda, Freddy Paz, Arturo Moquillaza, Marcelo M Soares, Aaron Marcus og Elizabeth Rosenzweig. 2021. „Web Analytics for User Experience: A Systematic Literature Review“ [**inlang**eng]. I *Design, User Experience, and Usability: UX Research and Design,* 12779:312–326. Lecture Notes in Computer Science. Switzerland: Springer International Publishing AG. ISBN: 9783030782207.

Pellizon, Lucas Henrique, Joelma Choma, Tiago Silva da Silva, Eduardo Guerra, Luciana Zaina, Osvaldo Gervasi, Alfredo Cuzzocrea et al. 2017. „Software Analytics for Web Usability: A Systematic Mapping“ [**inlang**eng]. I *Computational Science and Its Applications - ICCSA 2017,* 10409:246–261. Lecture Notes in Computer Science. Switzerland: Springer International Publishing AG. ISBN: 9783319624068.

Quattrucci, Tony. 2018. „Public Transport Victoria App Redesign: A UX Case Study“. Hentet 5. februar 2025. https://medium.com/@tony.quattrucci/public-transport-victoria-a-ux-case-study-de6ae49e8235.

Rashidi, Parisa. 2014. *Stream Sequence Mining for Human Activity Discovery.* ISBN: 9780123985323.

Rawira, Panunsiya, Vatcharaporn Esichaikul, Chutiporn Anutariya og Marcello M Bonsangue. 2023. „Web Usage Mining for Determining a Website's Usage Pattern: A Case Study of Government Website“ [**inlang**eng]. I *Data Science and Artificial Intelligence,* 1942:88–100. Communications in Computer and Information Science. Singapore: Springer. ISBN: 9789819979684.

Reyhani Hamedani, Masoud og Sang-Wook Kim. 2021. „On Investigating Both Effectiveness and Efficiency of Embedding Methods in Task of Similarity Computation of Nodes in Graphs“ [**inlang**eng]. *Applied sciences* (Basel) 11 (1): 162–. ISSN: 2076-3417.

Ribeiro, Rafael Fontinele, Matheus de Meneses Campanhã Souza, Pedro Almir Martins de Oliveira og Pedro de Alcântara dos Santos Neto. 2019. „Usability problems discovery based on the automatic detection of usability smells“ [**inlang**eng]. I *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing,* 147772:2328–2335. New York, NY, USA: ACM. ISBN: 9781450359337.

Salminen, Joni, Kathleen Guan, Soon-Gyo Jung, Shammur A. Chowdhury og Bernard J. Jansen. 2020. „A Literature Review of Quantitative Persona Creation“. I *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–14. CHI '20. Honolulu, HI, USA: Association for Computing Machinery. ISBN: 9781450367080. https://doi.org/10.1145/3313831.3376502. https://doi.org/10.1145/3313831.3376502.

Salminen, Joni, Kathleen Guan, Lene Nielsen, Soon-gyo Jung og Bernard J. Jansen. 2020. „A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles“. I *Human Interface and the Management of Information. Designing Information,* redigeret af Sakae Yamamoto og Hirohiko Mori, 125–144. Cham: Springer International Publishing. ISBN: 978-3-030-50020-7.

Salminen, Joni, Bernard J. Jansen, Jisun An, Haewoon Kwak og Soon-gyo Jung. 2025. „Are Personas Done? Evaluating Their Usefulness in the Age of Digital Analytics“ [**inlang**eng]. *Persona Studies* 4 (2): 47–65. ISSN: 2205-5258.

Salminen, Joni, Soon-Gyo Jung, Shammur Chowdhury, Sercan Sengün og Bernard J. Jansen. 2020. „Personas and Analytics: A Comparative User Study of Efficiency and Effectiveness for a User Identification Task“ [**inlang**eng]. I *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–13. New York, NY, USA: ACM. ISBN: 9781450367080.

Salminen, Joni, Soon-gyo Jung, Ahmed Mohamed Sayed Kamel, João M. Santos og Bernard J. Jansen. 2022. „Using artificially generated pictures in customer-facing systems: an evaluation study with data-driven personas“ [**inlang**eng]. *Behaviour information technology* (ABINGDON) 41 (5): 905–921. ISSN: 0144-929X.

Salminen, Joni, Joao M. Santos, Haewoon Kwak, Jisun An, Soon-gyo Jung og Bernard J. Jansen. 2020. „Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas“ [**inlang**eng]. *International journal of human-computer studies* 141:102437–. ISSN: 1071-5819.

Salminen, Joni, João M. Santos, Soon-gyo Jung og Bernard J. Jansen. 2024. „Picturing the fictitious person: An exploratory study on the effect of images on user perceptions of AI-generated personas“ [**inlang**eng]. *Computers in Human Behavior: Artificial Humans* 2 (1): 100052–. ISSN: 2949-8821.

Salminen, Joni, Kathleen Wenyun Guan, Soon-Gyo Jung og Bernard Jansen. 2022. „Use Cases for Design Personas: A Systematic Review and New Frontiers“. I *CHI '22 : Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems : April 30-May 5, 2022, New Orleans, LA, USA /.* CHI '22: CHI Conference on Human Factors in Computing Systems. New York,NY : Association for Computing Machinery, ISBN: 1-4503-9157-5.

Tankala, Samhita. 2023. „Information Architecture vs. Sitemaps: What's the Difference?“ Hentet 3. marts 2025. https://www.nngroup.com/articles/information-architecture-sitemaps/.

Taşgetiren, Nail og Mehmet S Aktas. 2022a. *Mining Web User Behavior: A Systematic Mapping Study,* 13377:667–683. Computational Science and Its Applications – ICCSA 2022 Workshops. Cham, Switzerland : Springer Nature Switzerland AG, ISBN: 3-031-10535-4.

———. 2022b. *Mining Web User Behavior: A Systematic Mapping Study,* 13377:667–683. Computational Science and Its Applications – ICCSA 2022 Workshops. Cham, Switzerland : Springer Nature Switzerland AG, ISBN: 3-031-10535-4.

Thoma, Volker, Bryn Williams, Tom Gross, Philippe Palanque, Raquel Oliveira Prates, Paula Kotzé, Jan Gulliksen, Lars Oestreicher og Marco Winckler. 2009a. „Developing and Validating Personas in e-Commerce: A Heuristic Approach“ [**inlang**eng]. I *Human-Computer Interaction - INTERACT 2009,* 5727:524–527. Lecture Notes in Computer Science 2. Germany: Springer Berlin / Heidelberg. ISBN: 9783642036576.

———. 2009b. „Developing and Validating Personas in e-Commerce: A Heuristic Approach“ [**inlang**eng]. I *Human-Computer Interaction - INTERACT 2009,* 5727:524–527. Lecture Notes in Computer Science 2. Germany: Springer Berlin / Heidelberg. ISBN: 9783642036576.

Tyagi, Neha, Santosh Kumar Gupta, Durgesh Kumar Mishra, Vasudha Bhatnagar og V. B Aggarwal. 2017. „Web Structure Mining Algorithms: A Survey“ [**inlang**eng]. I *Big Data Analytics,* 654:305–317. Advances in Intelligent Systems and Computing. Singapore: Springer. ISBN: 9811066191.

UserGuiding. 2019. „How to Create User Personas for Your B2B Product?“ Hentet 3. marts 2025. https://medium.com/mytake/how-to-create-user-personas-for-your-b2b-product-1272c97f65c5.

Varnagar, C. R., N. N. Madhak, T. M. Kodinariya og J. N. Rathod. 2013. „Web usage mining: A review on process, methods and techniques“ [**inlang**eng]. I *2013 International Conference on Information Communication and Embedded Systems (ICICES),* 40–46. IEEE. ISBN: 9781467357869.

Wallisch, A. og K. Paetzold. 2020. „METHODOLOGICAL FOUNDATIONS OF USER INVOLVEMENT RESEARCH: A CONTRIBUTION TO USER-CENTRED DESIGN THEORY“ [**inlang**eng]. *Proceedings of the Design Society: DESIGN Conference* 1:71–80. ISSN: 2633-7762.

Wang, Yi, Chetan Arora, Xiao Liu, Thuong Hoang, Vasudha Malhotra, Ben Cheng og John Grundy. 2025. „Who uses personas in requirements engineering: The practitioners' perspective“ [**inlang**eng]. *Information and software technology* 178:107609–. ISSN: 0950-5849.

Wei, Jishang, Zeqian Shen, N. Sundaresan og Kwan-Liu Ma. 2012. „Visual cluster exploration of web clickstream data“ [**inlang**eng]. I *2012 IEEE Conference on Visual Analytics Science and Technology,* 3–12. IEEE. ISBN: 1467347523.

Wodak, Ruth og Michael Meyer. 2016. *Methods of critical discourse studies.* [**inlang**eng]. 3. edition / edited by Ruth Wodak, Michael Meyer. Los Angeles, Calif: SAGE. ISBN: 9781446282410.

Wu, Youxi, Changrui Zhu, Yan Li, Lei Guo og Xindong Wu. 2020. „NetNCSP: Nonoverlapping closed sequential pattern mining“ [**inlang**eng]. *Knowledge-based systems* (AMSTERDAM) 196:105812–105812. ISSN: 0950-7051.

Xie, Xiao, Fan Du og Yingcai Wu. 2021. „A Visual Analytics Approach for Exploratory Causal Analysis: Exploration, Validation, and Applications“ [**inlang**eng]. *IEEE transactions on visualization and computer graphics* (LOS ALAMITOS) 27 (2): 1448–1458. ISSN: 1077-2626.

Yu, Hualong og Geoffrey I. Webb. 2019. „Adaptive online extreme learning machine by regulating forgetting factor by concept drift map“ [**inlang**eng]. *Neurocomputing (Amsterdam)* (AMSTERDAM) 343 (C): 141–153. ISSN: 0925-2312.

Zhang, Bin og Hua Dong. 2016. *User Involvement in Design: The Four Models.* Elektronisk udgave., 9754:141–152. Human Aspects of IT for the Aged Population. Design for Aging. Springer International Publishing, ISBN: 978-3-319-39942-3.

Zhang, Xiang, Hans-Frederick Brown og Anil Shankar. 2016. „Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry“. I *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems /,* 5350–5359. CHI'16: CHI Conference on Human Factors in Computing Systems. New York : ACM, ISBN: 9781450333627.

Zinat, Kazi Tasnim, Jinhua Yang, Arjun Gandhi, Nistha Mitra og Zhicheng Liu. 2023-06. „A Comparative Evaluation of Visual Summarization Techniques for Event Sequences“. *Computer graphics forum.* 42 (3): 173–185. ISSN: 0167-7055.

Appendix I | Kodebog dokumentanalyse

Spørgsmål	Kode-navn	Kort beskrivelse af hvad der kodes
Hvad er produktet?	PROD_DESC	**Kort beskrivelse:** Beskrivelser af produktet **Fuld beskrivelse**: Alt tekst, hvor produktets funktion, formål, features eller formelle beskrivelse nævnes **Hvornår det skal bruges:** Når funktionelle aspekter af produktet nævnes **Hvornår skal det ikke bruges:** Når meningsberørte beskrivelser af produktet nævnes **Kort beskrivelse:** Beskrivelse af brugere **Fuld beskrivelse**: Udsagn om faktiske brugere – deres roller, kontekst, typer, antal, eller adfærd.
Hvem bruger det?	PROD_USERS	**Hvornår det skal bruges:** Når tekst eller formulering af tekst kan beskrive hvem der bruger produktet **Hvornår det ikke skal bruges:** Beskrivelser af, hvordan produktet bruges eller hvad produktet bidrager med

Hvad har brugerne mest brug for?	USER_NEEDS	**Kort beskrivelse:** Beskrivelse af brugerbehov **Fuld beskrivelse**: Omtale af brugernes behov, pain points eller mål. **Hvornår det skal bruges:** Når aspekter af produktet opstår der har nytte for brugeren **Hvornår det ikke skal bruges:** Når brugeren beskrives og ved andre beskrivelser der ikke vedrør brugeren
Hvilke brugere og kunder er de mest vigtige for firmaet?	KEY_USERS	**Kort beskrivelse:** Fokus på særlige målgrupper. **Fuld beskrivelse**: Specifikke grupper prioriteres af firmaet. Herunder målgrupper, segmenter eller stakeholders. **Hvornår skal det bruges:** Når en specifik bruger eller målgruppe nævnes eller prioriteres Hvor skal det ikke bruges: Ved bredere beskrivelser af brugere eller brugergrupper

		Kort beskrivelse: Nævnte mangler eller begrænsninger.
Hvilke udfordringer møder de produktansvarlige og forretningen i den nære fremtid?	BUSINESS_ CHALLENGE	**Fuld beskrivelse**: fremtidige problemer, risikoer eller barrierer, især for forretning eller produktledelse. **Hvornår skal det bruges:** Når specifikke og nuværende uløste problemstillinger nævnes. **Hvornår skal det ikke bruges:** Når tidligere problemstilligner nævnes, som er blevet løst.
Hvem er de største konkurrenter og hvorfor?	COMPETITION	**Kort beskrivelse:** Nævnte konkurrenter **Fuld beskrivelse**: Konkurrenter nævnes, inkl. styrker/svagheder og differentiering. **Hvornår skal det bruges:** Når alternative produkter til MARC Helbredstillæg nævnes **Hvornår skal det ikke bruges:** Når lignende produkter nævner, men hvor produktets formål er forskellige

Hvilket syn har forskellige interessenter og produktansvarlige på brugeren?	STAKEHOLDER_VIEW	**Kort beskrivelse:** Forskelle i beskrivelse af brugeren. **Fuld beskrivelse**: Hvor synspunkter, meninger eller forståelser af brugeren diskuteres – især variation imellem stakeholders. **Hvornår skal det bruges:** Når interresenter snakker om brugeren og der opstår modstridende. **Hvornår skal det ikke bruges:** Når interresenter snakker om brugeren og beskrivelserne er komplementerende.

Hvilket budget og ressourcer er afsat til projektet?	RESOURCES	**Kort beskrivelse:** Information om produktets ressourcer. **Fuld beskrivelse**: Information om penge, tid, mandskab, værktøjer og støtte. Både på nuværende tid og i fremtiden. **Hvornår skal det bruges:** Når der er information om projektets ressourcebrug eller plan for fremtidige ressourcer. **Hvornår skal det ikke bruges:** Når der snakkes om selve pruduktet uden kontekts om ressourcebrug.
Hvilke tekniske begrænsninger og muligheder eksisterer der?	TECH_ CONSTRAINTS	**Kort beskrivelse:** Tekniske rammer. **Fuld beskrivelse**: Både begrænsninger og muligheder i teknologi, systemer, integrationer. **Hvornår skal det bruges:** Når tekniske uløste problemer, begrænsninger eller fordele nævnes. **Hvornår skal det ikke bruges:** Ved bugs eller andre begrænsninger der skyldes andet end den valgte teknologi.

Hvilke hovedmål har firmaet med produktet?	PRODUCT_GOALS	**Kort beskrivelse:** Formål med produktet. **Fuld beskrivelse**: Strategiske, taktiske eller økonomiske mål eller ønsket effekt med produktet. **Hvornår skal det bruges:** Når formålet med produktet nævnes. **Hvornår skal det ikke bruges:** Når andet end afkast eller formålet nævnes.

Appendix II | Kode til at udregne næste sekvens

```
from collections import Counter

def predict_next_element_unordered_no_overlap(sequences, context):
    candidates = Counter()
    context_len = len(context)
    context_set = set(context)
    for seq in sequences:
        i = 0
        while i <= len(seq) - context_len:
            window = seq[i:i + context_len]
            if set(window) == context_set and len(set(window)) == len(
   context_set):
                if i + context_len < len(seq):
                    next_element = seq[i + context_len]
                    candidates[next_element] += 1
                i += context_len
            else:
                i += 1
    if not candidates:
        return None
    return candidates.most_common()

pattern1 = [2.0]
pattern2 = [8.0, 7.0, 1.0]
pattern3 = [1.0, 2.0, 4.0, 1.0]
next_predictions = predict_next_element_unordered_no_overlap(
   reduced_sequences, pattern1)
next_predictions2 = predict_next_element_unordered_no_overlap(
   reduced_sequences2, pattern2)
```

```
next_predictions3 = predict_next_element_unordered_no_overlap(
    reduced_sequences3, pattern3)

print(f"\n     N ste element predictions for kontekst {pattern1}:")
if next_predictions:
    for next_state, count in next_predictions:
        print(f"  Next: {next_state}, Support: {count}")
else:
    print("  Ingen n ste element fundet.")
```

Listing II.1: Python-funktion for ikke-overlappende forudsigelser

Appendix III | Resultat af initierende undersøgelse

Hvad er produktet? (PROD_DESC)

MARC Helbredstillæg er en RPA-løsning, der håndterer og udbetaler fakturaer angående helbredstillæg. Løsningen understøtter fodbehandling, udvidet fodbehandling, fysioterapi, kiropraktor, tandlæge og psykolog. Produktet bruger Azure Intelligence Document for at behandle fakturaerne/kvitteringer, både almindelige og håndskrevne fakturaer, efter de er uploaded, og fakturaerne kan efterfølgende både åbnes og downloades i løsningen. Dermed komplementerer systemet KOMBITs løsning, der behandler de elektroniske fakturaer. Systemet slår borgeren op og sørger for udbetaling i KP, hvor en fil med faktura også gemmes på sagen, og der desuden oprettes et journalnotat.

Hvem bruger det? (PROD_USERS)

Opgaven er typisk placeret i borgerservice i kommunen, hvor sagsbehandlere sidder med opgaven. Efter ibrugtagning planlægges opgaven og eventuelle andre relevante faggrupper inddrages. Hvad angår titel, så er der mange forskellige titler at finde i deres mailsignatur, og selvom opgaverne formentlig minder om hinanden i borgerservice på tværs af kommunerne, så tyder det på, at der er tale om en bred gruppe af forskellige fagligheder. Sagsbehandlere varierer fagligt. En del er hk kontoruddannet og kender ikke til det faglige i arbejdet med MARC Helbredstillæg.

Hvad har brugerne mest brug for?(USER_NEEDS)

På brugerniveau har brugerne brug for et nemt system til automatisering af fakturaer til refusion af helbredstillæg, der kan frigive tid og ressourcer for blandt andet sagsbehandlere, så de kan fokusere på mere komplekse sager. På organisationsniveau er der brug for understøttelse af fælles kommunal rammearkitektur og tværprofessionelt samarbejde samt vidensdeling på tværs af forvaltninger. De har brug for standardiserede arbejdsprocesser, understøttet af et

korrekt datafundament frem for at følge den enkelte medarbejder, der støtter onboarding af nye medlemmer, da det er svært at rekruttere. De har brug for sikring af fuld dokumentation og muligheden for fremsøgning af informationer om sager. De skal have et overblik over fakturaer og alle kommunaløkonomiske betalinger og refusioner.

Hvilke brugere og kunder er mest vigtige for firmaet? (KEY_USERS)

Dataproces er målrettet den offentlige sektor og deres administration. Det er ledere og administrative, der køber produktet, men det er sagsbehandlere, der bruger produktet. Økonomiafdelingen har tidligere været i fokus, men i MARC Helbredstillæg er det sagsbehandlere, der er i fokus. Dataproces skelner på sin vis ikke i, om man vil have store eller små kommuner. De vil gerne have løsningen ud til alle kommuner.

Hvilke udfordringer møder de produktansvarlige og forretningen i den nære fremtid? (BUSINESS_CHALLENGE)

Kvitteringer kan være meget forskellige og flere skal håndteres på hver deres måde. Kommunerne gør det derudover ikke på samme måde. Dette skaber problemer. Derfor kan der også opleves udfordringer ved implementering af nye kommuner, hvis fakturaerne, der afleveres ved kommunen, har et layout, som er nyt og/eller svært for robotten.

Hvem er de største konkurrenter og hvorfor? (COMPETITION)

MARC Helbredstillægs konkurenter består af kommunerne selv. Kommunerne kan, med de rette kompetencer, vælge at udvikle et lignende produkt til eget brug. Enkelte kommuner har derfor også benyttet denne mulighed. Et andet problem kan ligge i, hvis kviteringer digitaliseres. Hermed ville kommunerne ikke behøve den løsning som MARC Helbredtillæg udbyder. Dette vuderes der dog ikke på nuværende tidspunkt at være sandsynligt.

Hvilket budget og ressourcer er afsat til projektet? (RESOURCES)

MARC Helbredstillæg er blevet testet og trænet samt udviklet i tværfagligt samarbejde internt hos Dataproces og med danske kommuner. Der er afsat fri support, igangsættelse samt faglig og teknisk opstartsforløb via Microsoft teams. Derudover er der afsat evalueringsmøder

via teams, fysisk undervisning og ekstra undervisning via teams og invitation samt afholdelse af ERFA møder, som skal understøtte vidensdeling af brugere på tværs af kommuner. Med henhold til evalueringsmøder opleves der, at kommunerne sommetider takker nej til dette tilbud, fordi de ikke har behov herfor. Dette skyldes eksempelvis, at de får deres spørgsmål besvaret igennem supporten.

Hvilke tekniske begrænsninger og muligheder eksisterer der? (TECH_CONSTRAINTS)
Løsningen er simpel at sætte op, brugervenlig og skaber overblik over fakturaerne og tager ca. 65-75% af de både almindelige- og håndskrevne uploadede fakturaer uden manuel gennemgang. Løsningen er bygget på en opkobling til ky (kommunernes ydelsessystem) og kp (kommunens pensionssystem), hvilket gør løsningen ustabil, sammenlignet med en løsning bygget på en API. Fakturalæsningen er begrænset af, hvor godt Azure OCR-behandlingen går, og dermed er dette ude af Dataproces kontrol. Ved informationer, som MARC Heldbredtillæg ikke kan tyde, eller i tilfælde, hvor der stemples oveni teksten, skal medarbejderen selv registrere informationer. Typiske frasorteringsårsager er manglende helbredstillæg i KP, i forbindelse med fakturadato, ukendt layout, beløb er 0 eller linjebeløb mangler. På nuværende tidspunkt kan der ikke modtages Excel/csv fil med opdateringer af ydelsesnummer fra kunderne, og Sygeforsikring Danmark vil heller ikke udlevere disse lister. Derfor er det nødvendigt at gennemgå pdf dokumenter og hjemmesider for informationer om ændret ydelsesnummer, da de ændrer sig fra år til år.

Hvilke hovedmål har firmaet med produktet? (PRODUCT_GOALS)
MARC er designet til at være medarbejderens personlige assistent og optimering af den kommunale, digitale administration. Målet for produktet er effektivisering af administrative opgaver, automatisere de trivielle arbejdsgange, frigive tid til de komplekse opgaver og tillid til, at det fungerer, at det virker op til deadline, og at der ikke laves fejl.

Appendix IV | Første iteration af kvalitative personaer

Anita Storgaard

Om Anita
Anita er 31 år og bor sammen med sin kæreste i en lejlighed, samme by som hun arbejder.

Hun har været i borgerservice i 11 år, hvor hun har arbejdet med MARC Helbredstillæg i omkring 1 år.

"teknologi er jo fremtiden, så man bliver nød til at sætte sig ind i det"

"jeg er bange for at lave fejl, men fejl sker og det er også okay"

Brug af produktet

Anita og hendes kollegaer samarbejder om hele processen angående MARC Helbredstillæg, fra at uploade kvitteringer til at skrive til borgerne om diverse udfald. Herunder er MARC Helbredstillæg et hjælpemiddel til K.P., som er hendes hovedsystem til helbredstillæg.

Frustrationer

- overordnet tilfreds med MARC Helbredstillæg
- oplever det som et brugervenligt system.
- Bliver ind i mellem forvirret over visse fejlkoder.
- Skal trække fakturaer, der har fået afslag, ud af MARC for at kunne journalisere dem i K.P.

Personlige mål

- Bedre kunne takle de problematikker der kommer hendes vej
- udvikle sig personlig og fagligt

Daglig rutine

Anita arbejder med mange forskellige opgaver, hvor bevilling af helbredstillæg er en fast del. Hun samarbejder med kollegaer og sparrer om faglige problemstillinger. Når en borger mangler et helbredstillæg, igangsætter hun et grundigt detektivarbejde i kommunens IT-systemer og bruger udelukkelsesmetoden for at finde fejlen. Derfor er gennemsigtighed i systemerne vigtig for hende.

Interesse

Anita holder af alsidigheden i sine opgaver, da den giver hende både faglig og personlig udvikling. Hun trives bedst med opgaver, der involverer faglige og menneskelige problemstillinger – især i mødet med forskellige borgertyper. IT-arbejde, som med MARC Helbredstillæg, kan derimod føles ensformigt.

Når nye systemer introduceres, bliver hun usikker, fordi fejl er uundgåelige i starten – og det er hun ikke tryg ved. Men efter at have vænnet sig til tanken, går hun til opgaven med gåpåmod, især hvis hun får støtte i opstartsfasen.

FIGUR IV.1: Denne figur viser et billede af første iteration af persona 2

Jonas Bits

Om Jonas
Jonas er 43 år og bor med sin hustru lidt udenfor byen.

Han har arbejdet i kommunen som kontormedarbejder i 6 år, hvor han har arbejdet med MARC Helbredstillæg i omkring 7 måneder.

"Jeg er ikke bange for at lave fejl, for jeg har altid en livslinje hvis det går galt"

"når jeg får nyt teknologi tænker jeg "let's do it" og så prøver jeg mig frem"

Brug af produktet

Jonases brug af MARC Helbredstillæg består hovedsageligt af at uploade fysiske kvitteringer, starte robotten og håndtere frasorterede kvitteringer.

Frustrationer

- Rigtig glad for MARC Helbredstillæg
- MARC Helbredstillæg laver nogen gange uventede fejl, der kommer bag på ham
- Robotten frasortere 50 til 80 procent af kvitteringer, som han skal gennemgå manuelt

Personlige mål

- Gøre en forskel for pensionerede borgere
- Gennemføre alle kvitteringer hurtigt og uden fejl

Daglig rutine

Jonas arbejder som flexansat tre timer fire gange om ugen, og sidder med MARC Helbredstillæg, hvor kollegaer tager sig af og borgerkontakten. Han sparer i ny og næ med kollegaerne om faglige spørgsmål, men sidder hovedsageligt alene med sine opgaver, som består af en liste af kvitteringer som skal køres. Når listen er tom, er Jonas færdig med sit arbejde, ellers fortsætter han dagen efter.

Interesse

Jonas interessere sig ikke decideret for Helbredstillæg, men finder rigtig stor mening i, at hans arbejde gør en forskel for de pensionerede borgere. Derudover motivere det ham at kunne se fremskridt i sit arbejde og vide, at han er færdig når han er færdig.

Han har altid været god til IT, og er temmelig komfortabel med at lære nye digitale værktøjer. Han er ikke bange for at lave fejl, for der er altid en livslinje, hvis det går galt, så han prøver sig frem og syntes egentlig, at det er ret sjovt med nye IT systemer.

FIGUR IV.2: Denne figur viser et billede af første iteration af persona 1

Appendix V | Format for logning af data

Felt	Værdi
event	session_start
timestamp	3000-32-13T25:61:61.13
token	dlkqwelriqlwkerhqw.qwelkrjqwlierkqwer
sessionId	1234234-882e-41a1-324-342dfsdrf432
browser	Chrome
operatingSystem	Web
platform	Win32
userAgent	Mozilla/5.0 ...
language	da_DK
screenResolution	1920 x 1305
fontScale	1
devicePixelRatio	1
touchSupport	No

Tabel V.1: Denne tabel viser strukturen på et JSON-element, der indeholder inherent features om brugeren, vist med fiktiv data som eksempel

Felt	**Værdi**
event	klik
timestamp	3000-32-13T25:61:61.13
token	dlkqwelriqlwkerhqw.qwelkrjqwlierkqwer
sessionId	1234234-882e-41a1-324-342dfsdrf432
event	click
buttonText	? unicode: U+E192
rowData	? unicode: U+E192

Tabel V.2: Denne tabel viser strukturen på et JSON-element, der indeholder data om et klik, vist med fiktiv data som eksempel

Appendix VI | Beskrivelse af workflows

Åbne lister og indsæt igen

Typisk workflow
Se detaljer på kvitteringer og indsæt eller indlæs kvitteringen igen.

Mest sete lister:
Frasorteret: 6.39%
Kunne ikke indsættes: 5.35%
Manuel behandlet: 2.62%

Mest brugte igangsættelser:
Klar til indsættelse indsæt: 3.66%
Indlæs igen fra listen: 2.07%
Indlæs igen: 2.02%

Typisk kliks: 4 til 12
Typisk varighed: 12 til 77 sek

Ret kvitteringer

Typisk workflow
redigering af kvitteringer i frasorteringslisten, hvorefter de indsættes eller indlæses igen.

Metode til indlæsning:

Indsæt direkte: 7.13%
Indsæt: 7.08%
Indlæs igen: 2.23%

Typisk kliks: 13 til 56
Typisk varighed: 97 til 537 sek

Lav note og send til manuel

Typisk workflow
Se detaljer på en kvittering og skriv i note feltet, tryk tilbage og håndteret status ændres.

Mest besøgte lister:

Frasorteret: 1.49%
Manuel behandlet: 1.32%
Kunne ikke indsættes: 1.32%

Typisk kliks: 6 til 22
Typisk varighed: 42 til 230 sek

Vent og genopfrisk

Typisk workflow
Trykker genopfrisk efter handling og venter derefter.

Typisk handling før genopfrisk:

Klar til indsættelse indsæt: 12.40%
Frasorteret prøv igen: 6.07%
Kunne ikke indsættes prøv igen: 1.32%

Typisk kliks: 1 til 7
Typisk varighed: 1 til 506 sek

Åbne lister uden at lave ændringer

Typisk workflow
Se detaljer på en kvittering uden at redigere, indsætte eller indlæse kvitteringen igen.

Mest besøgte lister:

Manuel behandlet: 5%
Alle: 3.77%
Frasorteret: 0.79%

Mest udførte handlinger:

Næste side: 22%
Se detaljer: 15.16%
Tilbage i periode: 7.45%

Typisk kliks: 3 til 12
Typisk varighed: 10 til 128 sek

Udenlukkende håndtering af frasoteringslisten

Typisk workflow
Se detaljer på kvitteringer i frasorteringslisten uden at redigere, men med andre handligner.

Mest brugte handligner:

Gå til KP: 12%
Slet: 4.11%
Sæt til manuel: 2%
Genoptag behandling: 1.76%

Typisk kliks: 1 til 3
Typisk varighed: 1 til 24 sek

Upload filer

Typisk workflow
Naviger til upload og vælg filer eller drop filer i feltet til upload og derefter genopfrisk.

Metode til upload:

Vælg filer: 11.93%
Drag and drop: 18.44%

Typisk kliks: 3 til 5
Typisk varighed: 1 til 102 sek

Tilbage/frem i periode og genopfrisk

Typisk workflow
Tryk frem i periode og tilbage i periode efter hinanden, ofte i forlængelse med genopfrisk.

Typisk kliks: 2 til 5
Typisk varighed: 2 til 34 sek

FIGUR VI.1: Denne figur viser den uddybet beskrivelse af hver workflow.

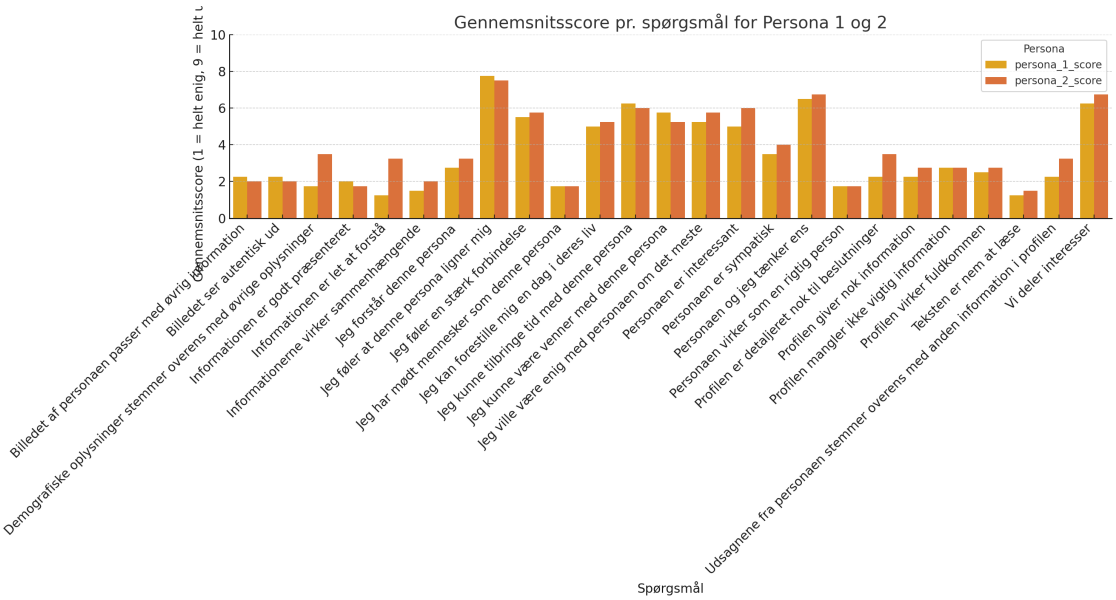
Appendix VII | Oversættelse af konstrukter

Konstrukt	Spørgsmål (engelsk)	Dansk oversættelse
Consistency1	The quotes of the persona match other information shown in the persona profile.	Udsagnene fra personaen stemmer overens med det andet information i personaens profil.
Consistency2	The picture of the persona matches other information shown in the persona profile.	Billedet af personaen passer med den øvrige information i profilen.
Consistency3	The persona information seems consistent.	Informationerne om personaen virker sammenhængende.
Consistency4	The persona's demographic information (age, gender, country) corresponds with other information shown.	Personaens demografiske oplysninger (alder og køn) stemmer overens med øvrige oplysninger.
Completeness1	The persona profile is detailed enough to make decisions about the customers it describes.	Personaen er tilstrækkeligt detaljeret til at træffe beslutninger om de beskrevne kunder.
Completeness2	The persona profile seems complete.	Personaens profil virker fuldkommen.
Completeness3	The persona profile provides enough information to understand the people it describes.	Persona profilen giver nok information til at forstå de mennesker, den beskriver.
Completeness4	The persona profile is not missing vital information.	Persona profilen mangler ikke vigtig information.

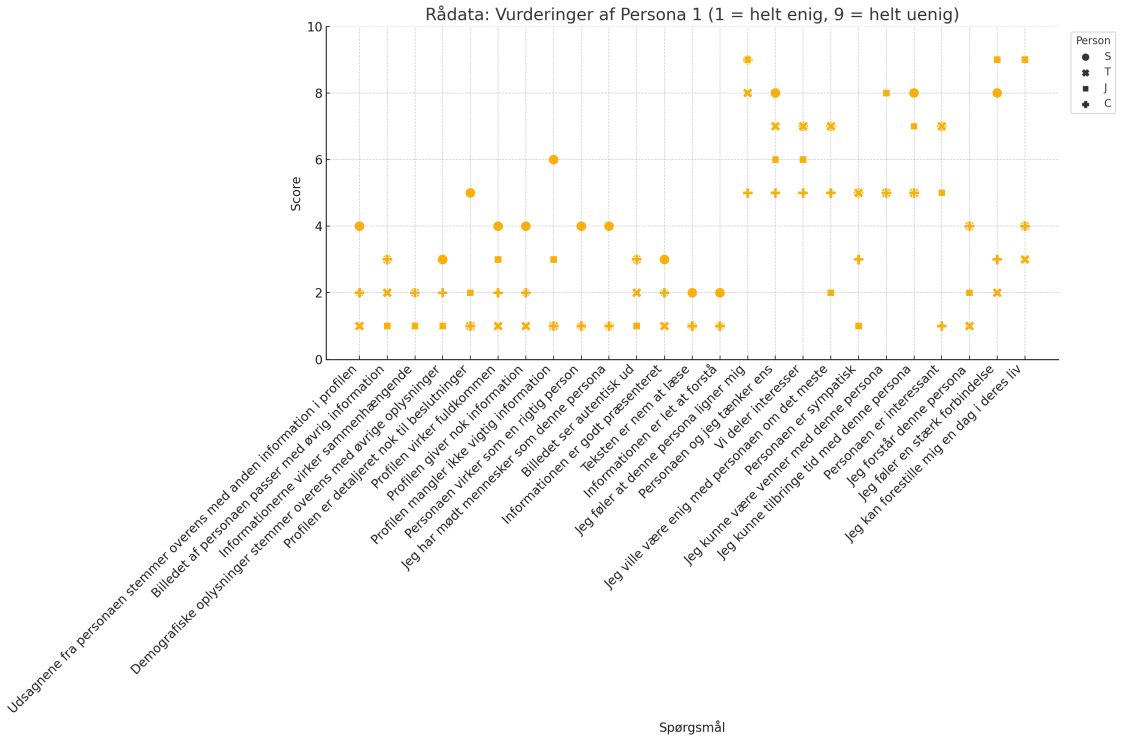
Konstrukt	**Spørgsmål (engelsk)**	**Dansk oversættelse**
Credibility1	The persona seems like a real person.	Personaen virker som en rigtig person.
Credibility2	I have met people like this persona.	Jeg har mødt mennesker som denne persona.
Credibility3	The picture of the persona looks authentic.	Billedet af personaen ser autentisk ud.
Clarity1	The information about the persona is well presented.	Informationen om personaen er godt præsenteret.
Clarity2	The text in the persona profile is clear enough to read.	Teksten i personaens profil er nem at læse.
Clarity3	The information in the persona profile is easy to understand.	Informationen i personaens profil er let at forstå.
Similarity1	This persona feels similar to me.	Jeg føler, at denne persona ligner mig.
Similarity2	The persona and I think alike.	Personaen og jeg tænker ens.
Similarity3	The persona and I share similar interests.	Personaen og jeg deler interesser.
Similarity4	I believe I would agree with this persona on most matters.	Jeg tror, jeg ville være enig med denne persona om de fleste ting.
Likability1	I find this persona likable.	Jeg synes, denne persona er sympatisk.
Likability2	I could be friends with this persona.	Jeg kunne godt være venner med denne persona.
Likability3	This persona feels like someone I could spend time with.	Denne persona virker som en, jeg kunne tilbringe tid sammen med.
Likability4	This persona is interesting.	Denne persona er interessant.
Empathy1	I feel like I understand this persona.	Jeg føler, at jeg forstår denne persona.

Konstrukt	**Spørgsmål (engelsk)**	**Dansk oversættelse**
Empathy2	I feel strong ties to this persona.	Jeg føler en stærk forbindelse til denne persona.
Empathy3	I can imagine a day in the life of this persona.	Jeg kan forestille mig en dag i denne personas liv.

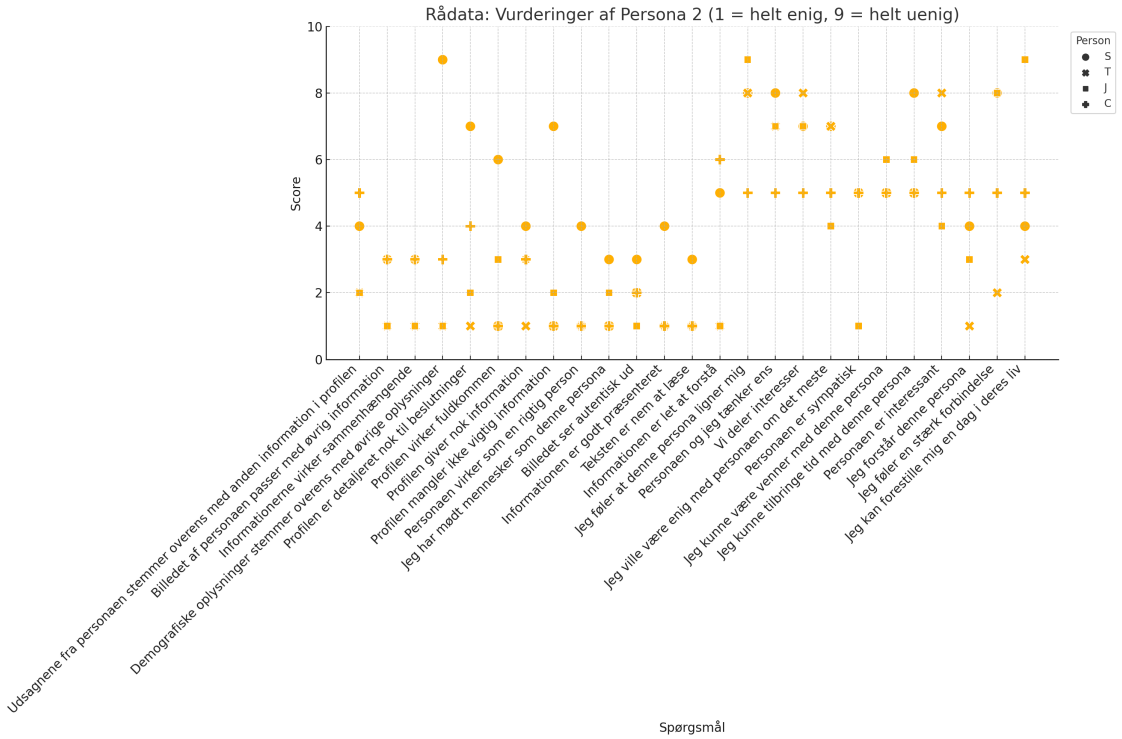
Appendix VIII | Resultat af survey om personaer



FIGUR VIII.1: Denne figur viser et pindediagram over den gennemsnitlige score, som deltagerne har givet hvert spørgsmål, hvor 1 betyder meget enig og 9 betyder meget uenig. Persona 1 referere til de kvalitative personaer, mens persona 2 referere til Web Usage Mining personaer.



FIGUR VIII.2: Denne figur viser et plot over den score, som deltagerne har givet hvert spørgsmål, hvor 1 betyder meget enig og 9 betyder meget uenig. Persona 1 referere til de kvalitative personaer. Hver person referere til et symbol, som kan ses i højre side af figurenm hvor S = senior udvikler, T = team leder, J = junior udvikler og C = ansvarlig for customer support.



FIGUR VIII.3: Denne figur viser et plot over den score, som deltagerne har givet hvert spørgsmål, hvor 1 betyder meget enig og 9 betyder meget uenig. Persona 2 referere til Web Usage Mining personaer. Hver person referere til et symbol, som kan ses i højre side af figurenm hvor S = senior udvikler, T = team leder, J = junior udvikler og C = ansvarlig for customer support.