Aalborg Universitet

Biomedical Engineering and Informatics

Gruppe 20gr10411

Fænotypeinddeling og prædiktion af ældre borgeres longitudinale hjemmeplejebehov

Forfattere:
Mathilde Fischer Gadensgaard
Steffen Jensen

AAU-vejleder:
Alex Skovsbo Jørgensen, MSc

2. Juni 2020

AALBORG UNIVERSITET
STUDENTERRAPPORT

AALBORG UNIVERSITET
STUDENTERRAPPORT

Health Science and Technology
Biomedical Engineering & Informatics
Fredrik Bajers Vej 7A
9220 Aalborg Øst
http://www.hst.aau.dk

Titel:
Fænotypeinddeling og prædiktion af ældre borgeres hjemmeplejebehov

Projekt:
Kandidatspeciale

Projektperiode:
Februar 2020 - Juni 2020

Projektgruppe:
20gr10411

Gruppemedlemmer:
Mathilde Fischer Gadensgaard
Steffen Jensen

AAU-vejleder:
Alex Skovsbo Jørgensen, MSc

Sider: 105
Afsluttet: 02-06-2020

Synopsis:

Baggrund:
Den kommunale ældrepleje har behov for effektivisering i kraft af, at skulle varetage en stigende opgavemængde, og samtidig begrænse udgifterne hertil. En udfordring i denne forbindelse er et tiltagende differentieret ældrebillede med forskellige plejebehov og udvikling heraf. Det undersøges deraf om data fra elektroniske omsorgsjournaler kan anvendes i en data-drevet tilgang til udledning og prædiktion af plejebehovsfænotyper.

Metode:
Dette studie udviklede og implementerede et system bestående af to centrale dele. Første del omfattede en sammenligning af k-means clustering og hierarkisk agglomorativ clustering til udledning af plejebehovsfænotyper ud fra den longitudinale udvikling i ældre borgeres plejebehov i løbet af et år. Anden del blev udgjort af en klassifikationsmodel til prædiktion af borgerens plejebehovsfænotype ud fra stamdata og viden om deres første visitation. I denne forbindelse blev en Random Forest og et Neural Network sammenlignet.

Resultater:
Det udviklede *proof of concept*-system udledte seks unikke plejebehovsfænotyper ved brug af k-means clustering. I den efterfølgende klassifikation af borgeres plejebehovsfænotype, opnåede systemet en maksimal ROC AUC på 0.77 og PR AP på 0.43 med den implementerede Random Forest.

Konklusion:
Data fra elektronisk omsorgsjournal-systemer kan anvendes som potentielt grundlag for en data-drevet tilgang til inddeling af fænotyper, gennem k-means clustering. Desuden er brug af en Random Forest-klassifikationsmodel appliceret på data fra borgerens første visitation en mulig metode til estimering af det fremtidige plejebehov.

Rapportens indhold er frit tilgængeligt, men offentliggørelse (med kildeangivelse) må kun ske efter aftale med forfatterne.

Forord

Denne projektrapport er udarbejdet som et kandidatspeciale på civilingeniøruddannelsen, Sundhedsteknologi på Aalborg Universitet i foråret 2020. Der rettes stor tak til Faaborg-Midtfyn Kommune for deltagelse i interview samt for at stille data tilrådighed. Endvidere takkes KMD for facilitering af kontakt til Faaborg-Midtyn Kommune og for fagkyndig sparring gennem Flemming Holbæk Lundager og Jeppe Svendsen. Afslutningsvis, siges et tak til vejleder Alex Skovsbo Jørgensen for kompetent vejledning gennem hele projektforløbet.

Mathilde F. Gadensgaard Steffen Jensen

Læsevejledning
Dette specialeprojekt er opbygget med en indledende analyse af problemfeltet, hvorudfra problemformuleringen opstilles. Herefter beskrives i kapitel 4, projektets metode for udvikling og implementering af det tilhørende *proof-of-concept*-system. Desuden indeholder dette kapitel grundlæggende teori for de anvendte *machine learning*-modeller. Det efterfølgende kapitel beskriver resultater for hyperparameteroptimering i relation til de implementerede klassifikationsmodeller. Afslutningsvis præsenteres de opnåede resultater, der desuden behandles i syntesen, udgjort af diskussion og konklusion.

Kildehenvisningen følger Vancouver-modellen, hvor referencen anføres med et tal omsluttet af klammer, eksempelvis [1]. Nummerering af kilderne foregår fortløbende som de forekommer i teksten.

For at etablere konsistens mellem projektets figurer, der er udarbejdet med engelsk talstandard, og den skrevne tekst anvendes engelsk-formaterede tal med punktum-separator frem for komma.

Der anvendes en række akronymer og forkortelser, hvortil der først i projektrapporten findes en liste med angivelser af de fulde betegnelser. Første gang en forkortelse anvendes, er den fulde betegnelse inkluderet.

Abstract

Background:
The municipality-based eldercare is in the need of improving the efficiency on account of having to manage an increasing amount of tasks while limiting the expenses hereof. In this regard, the increasingly diverse group of elderly citizens poses a challenge. From this, it is investigated whether data from electronic home care records may be utilized in a data-driven application to extract and predict phenotypes of care requirements.

Methods and Materials:
The present study developed and implemented a system consisting of two central parts. The first part is made up of a comparison of k-means clustering and hierarchical agglomerative clustering for the task of extracting care requirement phenotypes from the longitudinal trajectory of elderly citizens in the duration of a year. The second part constituted a classifier for prediction of the care requirement phenotype of a citizen, based on demographic data and information regarding the first visitation. In this regard, a Random Forest and a Neural Network were compared.

Results:
The developed *proof of concept*-system, deduced six unique care requirement phenotypes using k-means clustering. In the subsequent classification of the care requirement phenotype, the system achieved a maximum ROC AUC of 0.77 and PR AP of 0.43 with the implemented Random Forest.

Conclusion:
Data from electronic home care records may be utilized in a data-driven application for extracting phenotypes, using k-means clustering. Furthermore, the use of a Random Forest classifier applied on data from the first visitation of the citizen, constitute a potential method for estimation of the future care requirement

Akronymer

AdaGrad	Adaptive Gradient Algorithm
Adam	Adaptive Moment Estimation
AP	Average Precision
ARIMA	Autoregressive Integrated Moving Average
AUC	Area Under the Curve
BUM	Bestiller-Udfører-Modtager
CART	Classification and Regression Tree
EOJ	Elektronisk Omsorgsjournal
EPJ	Elektronisk Patientjournal
FMK	Faaborg-Midtfyn Kommune
FS	Fælles sprog
HAC	Hierarkisk Agglomorativ Clustering
IQR	Interquartile Range
KL	Kommunernes Landsforening
KMC	K-Means Clustering
LSTM	Long Short-Term Memory
MLP	Multi-layer Perceptron
NN	Neural Network
PCA	Principal Components Analysis
PR	Precision-Recall
ReLU	Rectified Linear Unit
RF	Random Forest
RMSD	Root Mean Squared Deviations
RMSprop	Root Mean Squared Propagation
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SSD	Sum of Squared Deviations
t-SNE	t-Distributed Stochastic Neighbor Embedding

Indholdsfortegnelse

Kapitel 1 Indledning **1**

Kapitel 2 Problemanalyse **2**
2.1 Den decentrale offentlige sektor under pres . . . 2
2.1.1 Udvikling i demografi og helbred hos den ældre befolkning . . . 3
2.2 Nuværende arbejdsprocesser på ældre- og omsorgsområdet . . . 4
2.2.1 Visitation af social- og sundhedsydelser . . . 5
2.2.2 Styring og kvalitetsudvikling på ældre- og omsorgsområdet . . . 7
2.3 Fælles sprog III . . . 8
2.3.1 Fælles Sprog III tilstande . . . 9
2.3.2 Fælles Sprog III indsatskatalog . . . 9
2.4 Udnyttelse af den registrerede ældre- og omsorgsdata . . . 10
2.5 Case: Hjemmehjælp i Faaborg-Midtfyn Kommune . . . 11
2.6 Lignende studier . . . 14

Kapitel 3 Problemformulering **16**
3.1 Problemafgrænsning . . . 17

Kapitel 4 Metode **18**
4.1 Studieoverblik . . . 18
4.2 Data . . . 18
4.2.1 Borgerdemografi . . . 20
4.2.2 Datasegmentering og præprocessering . . . 20
4.3 Clustering af longitudinal udvikling i plejebehov . . . 21
4.3.1 Præ-clustering analyse . . . 23
4.3.2 Hierarkisk clustering . . . 26
4.3.3 K-means clustering . . . 28
4.3.4 Komparativ analyse og bestemmelse af antal clusters . . . 29
4.4 Klassifikationsmodeller til prædiktion af plejebehovsfænotype . . . 34
4.4.1 Bayesian Optimization . . . 34
4.4.2 Random Forest . . . 35
4.4.3 Neuralt netværk . . . 38
4.5 Evaluering af cluster-udledte fænotyper . . . 43
4.6 Evaluering af klassifikationsmodeller . . . 44
4.6.1 Receiver operating characteristic . . . 44
4.6.2 Precision og Recall . . . 47
4.7 Evaluering af vigtigste EOJ-features . . . 47
4.7.1 Permutation importance . . . 47
4.7.2 Drop column importance . . . 48

Kapitel 5 Hyperparameteroptimering **49**

5.1 Random Forest 49
5.2 Neuralt netværk 49

Kapitel 6 Resultater **52**
6.1 Clustering af longitudinal udvikling i leverede hjemmehjælp 52
6.1.1 Clustering performance 52
6.1.2 Evaluering af typeforløb som approksimation af individuelle plejebehovsforløb 56
6.1.3 Homogenitet i de udledte clusters og anvendelighed som fænotyper . 59
6.2 Klassifikation af plejebehovsfænotype 62
6.3 Feature importance 65

Kapitel 7 Syntese **69**
7.1 Diskussion 69
7.2 Konklusion 72

Litteratur **73**

Appendiks A Struktureret Litteratursøgning **77**

Appendiks B Interview m. Faaborg-Midtfyn Kommune

Appendiks C Feature signifikans-maps

Appendiks D Feature Importance

Indledning 1

Den kommunale varetagelse af ældre- og omsorgsområdet er under et flerfoldigt pres, idet der stilles krav til begrænsning af udgiftstilvæks, alt i mens mængden af opgaver og kompleksiteten heraf er tiltagende. Derudover er det en målsætning, at kvaliteten af de udbudte serviceydelser øges. Foruden denne eksisterende ressourceknaphed, foreskrives det at anddelen af ældre i befolkningen vil stige markant de kommende år, hvilket vil lægge yderligere pres på den kommunale ældrepleje.

Den tiltagende opgavemængde i ældreplejen er et resultat af, at de seneste år ligeledes har været præget af en tiltagen i antallet af ældre og kronikere. Samtidig forekommer et ønske om, at holde borgeren længst muligt i eget hjem, hvorved opgavebyrden af komplekse pleje- og behandlingsopgaver, der foretages i borgerens hjem, er tiltagende og omskiftelig.

For at kunne imødekomme krav om begrænsning af udgiftstilvækst uden, at serviceniveauet og kvaliteten af ydelser vil lider herunder, er der dermed behov for redskaber til effektiv prioritering, kvalitetsudvikling og forebyggelse.

En mulig metode for tilvejebringelse af sådanne redskaber er anvendelse af en data-drevet tilgang. Gennem systemer for elektroniske omsorgsjournaler, genereres en stor mængde data for ældre- og omsorgsområdet. Derudover har mange kommuner nyligt implementeret seneste version af dokumentationsstandarden Fælles Sprog, FSIII. Formålet med denne standard er at styrke det eksisterende databillede gennem strukturering og klassificering af data til anvendelse i ledelsesmæssige styrings- og prioriteringsopgaver. Herudfra kan den registrerede data i ældreplejen potentielt udnyttes i data-drevne redskaber, som efterspørges til bearbejdning af den store mængde information.

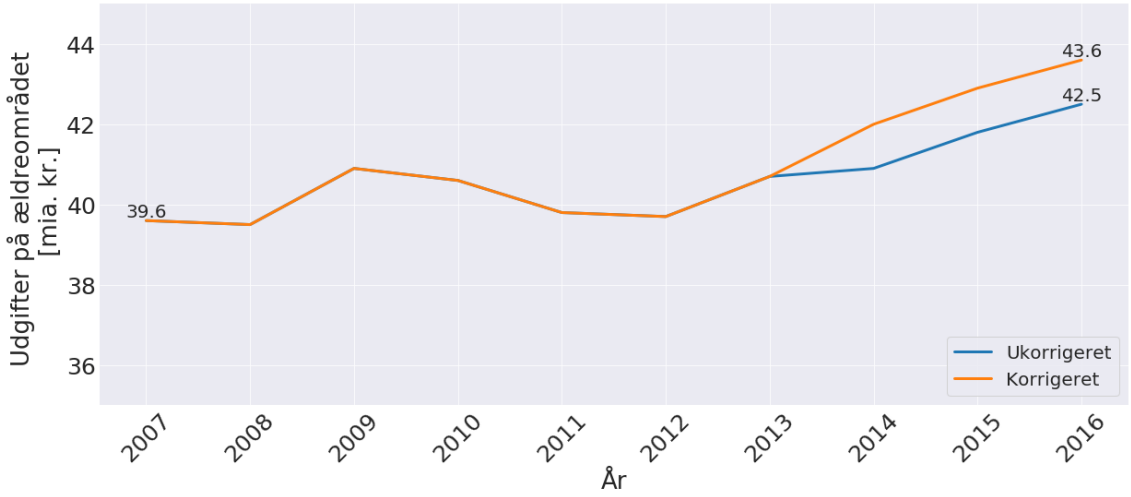
På baggrund af dette problemfelt undersøges det i nærværende studie, indledningsvist, hvilke udfordringer der er forbundet med, at udnytte ældre- og omsorgsdata fra elektroniske omsorgsjournaler til data-drevne beslutnigsstøttende effektiviseringsredskaber. Med afsæt i dette vidensgrundlag udvikles dernæst et proof of concept værktøj som løsningsforslag til de identificerede udfordringer.

Problemanalyse 2

Dette kapitel har til formål at identificere udfordringer forbundet med udvikling af data-drevne redskaber til prioritering og effektivisering af ældre- og omsorgsområdet.

2.1 Den decentrale offentlige sektor under pres

Pleje og behandling i hjemmet er en stadig større og nødvendig del af borgernes møde med sundhedsvæsnet. Dette følger af kortere liggetider på sygehusene, et voksende antal af ældre og kronikere samt en generel socialpolitisk holdning om, at borgeren skal være en aktiv del af eget behandlingsforløb, og deraf være længst muligt i eget hjem. Som følge heraf forekommer en kraftig stigning i aktiviteten i hjemmesygepleje, både i forhold til antal af borgere og frekvensen på besøg [1]. Endvidere beskrives en *opgaveglidning*, som betegner at specialiserede sygeplejeopgaver flyttes fra sekundær til primærsektoren, hvorved opgavebyrden af komplekse pleje- og behandlingsopgaver, der foretages i borgerens eget hjem, er tiltagende og omskiftelig [1, 2]. Et eksempel herpå er opgaver forbundet med ikke-medicinsk hjerterehabilitering, der i nogle regioner er uddelegeret til kommunen [3]. På trods af stigning i arbejdsbelastning i hjemmesygeplejen, er antallet af hjemmesygeplejersker forblevet konstant [4]. Denne udvikling er afspejlet i den internationale litteratur, der fremhæver, hvordan sundhedssektoren skal imødekomme en stigende mængde plejebehov hos borgerne, og samtidig opnå øget effekt af de udbudte plejeindsatser og behandlinger under reducerede økonomiske rammer [5].



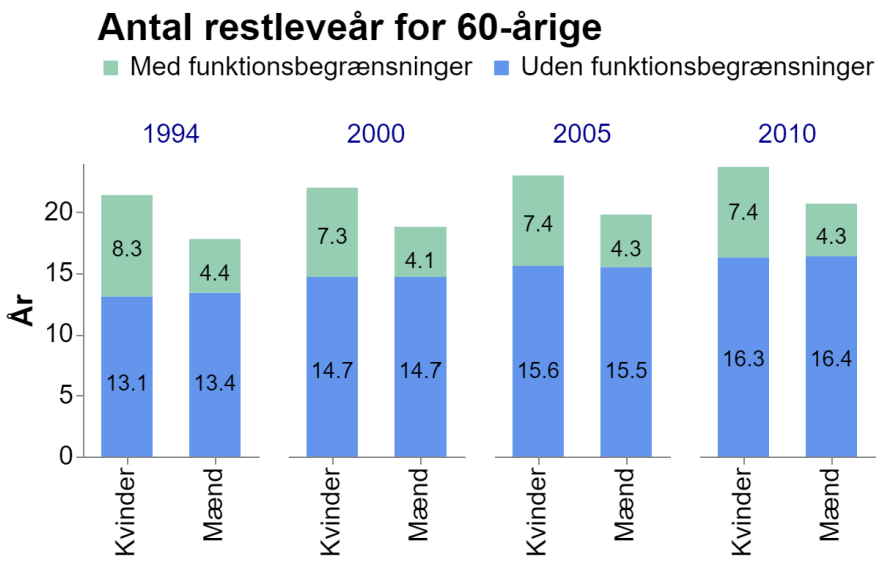
Figur 2.1. Kommunernes udgifter til ældreområdet, korrigeret for finanspolitisk løft (orange) og ukorrigeret (blå). Modificeret fra [6].

I Danmark er de stigende udgifter til ældreområdet blevet prioriteret finanspolitisk de seneste år, men som det fremgår af figur 2.1 hvor udgifterne korrigeres for dette løft, er der endog en stigning med omkring fire milliarder kroner fra 2007 til 2016 [6]. Herfor er der et behov for at undersøge, hvordan de tilgængelige ressourcer i primærsektoren udnyttes og prioriteres optimalt, og dermed udløser mest mulig velfærd [7]. Samtidig beskrives det i en rapport fra Hjemmehjælpskomissionen, at den kommunale indsats på ældre- om omsorgsområdet skal tage højde for et tiltagende, differentieret ældrebillede, i forhold til behovet for pleje og støtte, særligt forsaget af en stigning i antallet af borgere med demens og kroniske lidelser, samt en tendens til stigende ulighed i sundhed. [7]. Dette billede understøttes i en artikel af Rosella et al., der beskriver en tendens til, at de 5% dyreste borgere udgør omkring 50% af de samlede udgifter i sundhedsvæsnet i Canada. Her gælder det, at særligt borgere med flere samtidige helbredsnedsættende tilstande står for størstedelen af det samlede udgiftsbillede [8]. Såfremt denne tendens er afspejlet på ældre- og omsorgsområdet i Danmark, er der altså et stort potentiale i at forhindre uhensigtsmæssig udvikling i plejebehovet for borgere, hvor det kunne være undgået. Herfor bør muligheder for at udvikle målrettede indgreb og præventive tiltag undersøges. Eksempelvis kunne indsatsforløb målrettes flerfoldige plejebehov, hvor pleje- og behandlingsindsatser traditionelt er udviklet til individuelle helbredstilstande. En sådan strategi vil kræve analytiske værktøjer til udredning af det heterogene ældrebillede og dermed bedre indsigt i den enkelte borgers fremtidige plejebehov, samt hvorledes indsatser påvirker udviklingen af dette behov.

2.1.1 Udvikling i demografi og helbred hos den ældre befolkning

Danmarks Statistiks befolkningsprognose foreskriver, at den demografiske sammensætning i Danmark i fremtiden vil udgøres af en markant større andel ældre, mens den yngre befolkningsgruppe mindskes [9]. Den ældre befolkning defineres ofte ud fra folkepensionsalderen, hvilket vil sige, at personer over 66 år inkluderes i ældregruppen. Størstedelen af borgere, der modtager hjemmehjælp er over 75 år, hvilket i 2013 udgjorde 406.000 personer. Dette antal vil, ifølge Danmarks Statistik, være mere end fordoblet i 2050, idet denne gruppe er estimeret til omkring 909.000 personer. Denne udvikling kan ikke direkte sidestilles med, at arbejdsbyrden i hjemmeplejen vil øges tilsvarende, da der er indikationer for, at funktionsevnen ved bestemte alderstrin vokser, i takt med at levetiden forlænges. Der er altså sket en helbredsmæssig udvikling hos den ældre befolkning, hvor antallet af leveår uden funktionsbegrænsninger er steget i takt med stigningen i den generelle levealder. Dette er illustreret på figur 2.2. Den længere levetid uden funktionsbegrænsninger betyder, at tidspunktet for, hvornår der opstår et behov for hjemmehjælp er udskudt. For at opretholde og forbedre den positive udvikling i antallet af leveår uden funktionsnedsættelse, er det relevant at identificere, og rette fokus mod de grupper, der ikke følger samme fremgang som gennemsnittet. [7].

I kontrast til den positive udvikling i antallet af leveår uden funktionsbegrænsninger, opleves en stigning i forekomsten af kroniske sygdomme og demens for de højere alderstrin. I 2010 levede 50% af 65-74-årige og 66% af borgere over 75 år med en kronisk sygdom. Til sammenligning var andelen af 45-54-årige med en kronisk sygdom blot 15%. Denne aldersbetingede stigning i kronikkere betyder, at en del af ældregruppen kan have komplekse plejebehov, der stiller større krav til hjemmeplejen. Dette gælder ligeledes



Figur 2.2. Antallet af restleveår, med og uden funktionsbegrænsning, for 60-årige. Modificeret fra [7].

demenssyge borgere, der allerede udgør omkring 25% af hjemmehjælpsmodtagere, og generelt anslås at udgøre cirka en femtedel af borgere i 85-års alderen [7]. Foruden stigning i kroniske sygdomme, falder borgernes selvvurderede helbred i takt med alderen. Dette er evident i en undersøgelse af *sundhed og sygelighed i Danmark*, udarbejdet af Statens Institut for Folkesundhed i 2005. Her vurderede 70,6% af 65-74-årige deres helbred som "godt" eller "virkelig godt", mens det for 75-84-årige var 62,8% [10]. Alt i alt er den nuværende, og fremtidige ældregruppe præget af heterogenitet i forhold til selvvurderet helbred og kroniske sygdomme. Denne diversitet betyder, at det bør overvejes om ældregruppen kan inddeles i mere homogene fænotyper med fællestræk i forhold til deres sygdomsprofil, demografiske oplysninger, plejebehov eller andre karakteristika. Sådanne fænotyper kan være et muligt redskab til identificering af hvilke borgere, der vil udvikle et højt plejebehov. Desuden muliggøres tilpasning af plejeforløb for de forskellige typer borgere, med henblik på at øge effektivt og kvalitet i hjemmeplejen for den enkelte borger.

2.2 Nuværende arbejdsprocesser på ældre- og omsorgsområdet

I forbindelse med sundheds- og omsorgsindsatser, der udføres i borgernes eget hjem, skelnes der mellem hjemmepleje og hjemmesygepleje, hvor førstnævnte har til formål at støtte borgere med nedsat fysisk eller psykisk funktionsevne. Hjemmeplejeindsatser leveres efter Servicelovens §83, og indebærer hjælp til personlig pleje, praktiske opgaver og madservice [11]. Herigennem søges at gøre borgeren mest mulig selvhjulpen samt at øge livskvaliteten. Serviceloven vedtages i Folketinget som en rammelov, hvorved det er op til den enkelte kommunalbestyrelse at fastlægge serviceniveauet for at udfylde disse rammer. Hjemmesygepleje leveres i henhold til Sundhedslovens §138, der pålægger sundhedsvæsnet at yde sundhedsfremme gennem forebyggelse og behandling af sygdomme, lidelser og funktionsbegrænsninger [3]. Dermed bør sundhedsvæsnet, ifølge sundhedsloven, tilrettelægge ydelser med henblik på forebyggelse af uhensigtsmæssige

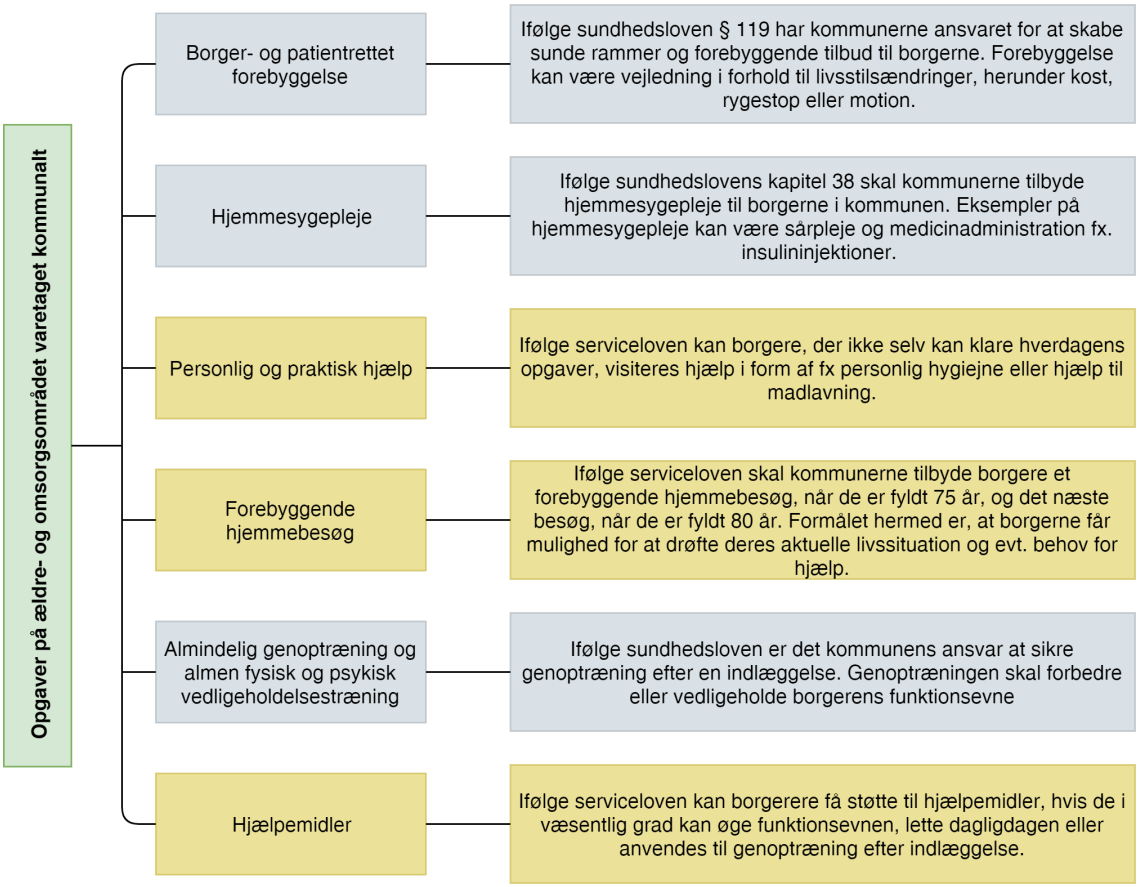
patientforløb. Sundhedsloven beskriver også at *"kommunalbestyrelsen er ansvarlig for, at der ydes vederlagsfri hjemmesygepleje efter lægehenvisning til personer med ophold i kommunen"* [12]. Her adskiller sundhedsloven sig fra serviceloven, idet en borger har *krav* på hjælp til et sundhedsmæssigt problem, såfremt der foreligger en sundhedsfaglig henvisning. For at modtage hjemmepleje i henhold til serviceloven, skal borgeren derimod ansøge herom, hvorefter denne ansøgning vurderes af en kommunal visitator [3]. De forskellige opgaver, der varetages kommunalt på ældre- og omsorgsområdet samt de bagvedliggende love, er opstillet i figur 2.3 Opgaverne i hjemmeplejen udføres oftest af social- og sundhedspersonale, mens opgaverne i hjemmesygeplejen kan udføres af personale som sygeplejersker, fysioterapeuter, social- og sundhedsassistenter og andre. Samtidig forekommer der et overlap mellem hjemmeplejen og hjemmesygeplejen, idet nogle opgaver fra hjemmesygeplejen uddelegeres til hjemmeplejen, og dermed konverteres fra sundheds- til serviceloven [4]. Opgavefordelingen mellem faggrupper organiseres i forhold til, hvordan kommunen tilrettelægger deres opgave i at yde vederlagsfri hjemmesygepleje [1, 3]. Det er forskelligt, hvordan kommunerne organiserer ældre- og omsorgsområdet; hvor nogle kommuner har en ordning med integration mellem hjemmeplejen og hjemmesygeplejen, mens det for andre er selvstændige enheder. Argumentet for en integreret ordningen er muligheden for sparring og koordinering af opgaver mellem sygeplejersker og øvrigt plejepersonale. Omvendt giver en selvstændig sygeplejeenhed mulighed for vidensdeling mellem sygeplejersker [2]. Der er dermed mange muligheder, når det gælder organisering af det nære sundhedsvæsnet, i form af hjemmepleje og hjemmesygepleje; alle med hver deres fordele og ulemper. Fælles er dog at ydelserne, som borgeren skal have leveret, gennemgår en visitationsprocess inden de tildeles.

2.2.1 Visitation af social- og sundhedsydelser

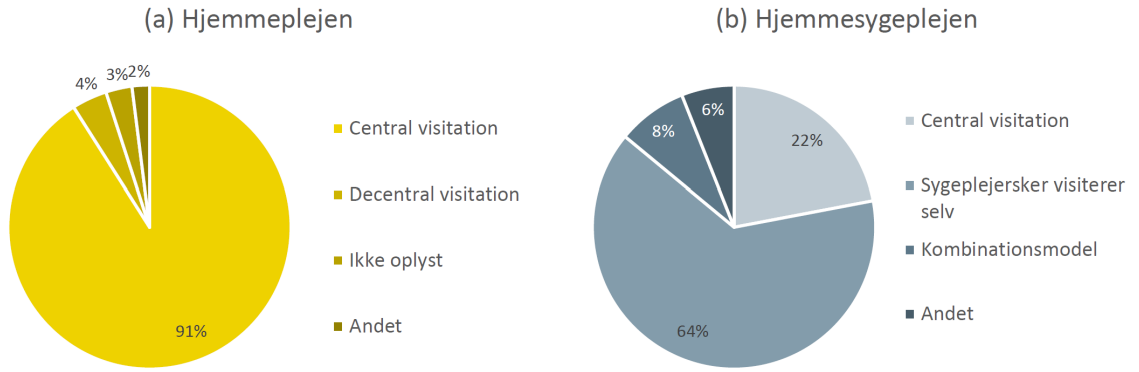
Som beskrevet, kræver udbydelsen af hjemmepleje, at de relevante ydelser visiteres på baggrund af en anmodning fra borgeren. Hjemmesygepleje visiteres derimod på baggrund af en sundhedsfaglig henvisning. Ydermere kan en hjemmesygeplejeopgave uddelegeres til hjemmeplejen, ved at opgaven sendes til visitationen, hvor den bevilliges som en hjemmeplejeydelse [4].

For hjemmesygeplejeindsatser kan visitationsformen inddeles i tre overordnede kategorier: Central visitation, selvvisiterende sygeplejersker og central visitation med decentral udmøntning. Fordelingen af kommuner med de forskellige visitationsformer, opgjort for 45 kommuner i en undersøgelse af Kommunernes Landsforening (KL), kan ses på **figur 2.4 (b)**.

Central visitation indebærer ofte, at der visiteres efter et fast indsatskatalog, hvortil frekvens og tid for hvert enkelt besøg skal besluttes. Fordelen herved er, at der skabes et central overblik og deraf ensartethed i visitationen, samt den faglige linje for alle enheder i kommunen. For selvvisiterende sygeplejersker visiteres indsatserne direkte af den udførende sygeplejerske. Herved udnyttes sygeplejerskernes indsigt i praksis. En ulempe ved denne visitationsform er, at der kan opstå forskellige faglige linjer mellem visitatorerne, og dermed risiko for et uensartet serviceniveau for sammenlignelige borgere [2]. Central visitation med decentral udmøntning beskriver en visitationsmodel, der kombinerer central visitation og selvvisiterende sygeplejersker. Formålet er at sikre overblikket hos den centrale



Figur 2.3. Overordnede opgaver varetaget af primærsektoren på ældre- og omsorgsområdet. Opgaver markeret med gråt, markerer at ydelserne visiteres på baggrund af sundhedsloven, mens det for de gule markeringer er serviceloven. Modificeret fra [3].



Figur 2.4. Fordeling af visitationsformerne i kommunerne for henholdsvis (a) hjemmeplejen [n= 261] og (b) hjemmesygeplejen [n= 45]. Modificeret fra [2] og [13].

visitationsenhed, mens sygeplejerskernes faglige viden udnyttes i selve udmøntningen. Denne model kan organiseres på flere måder, men et typisk eksempel er, at den centrale enhed fastsætter målsætninger, udstikker retningslinjer for ydelser samt følger op på effekten af behandlingen. Beslutningen omkring den specifikke visitation ligger hos den enkelte sygeplejerske, så behandlingen baseres på praksis. En ulempe ved denne model er, at den decentrale visitation kan resultere i uensartethed i serviceniveauet mellem de visiterende sygeplejersker, grundet subjektivitet [2]. Uanset den valgte visitationsform, inducerer de visiterende aktører altså en subjektivitet, hvilket indikerer et behov for redskaber til objektiv visitation og tilrettelæggelse af hjemmesygeplejeforløb.

Visitationen af hjemmepleje bygger på loven om frit valg, der foreskriver at borgeren har valgfrihed mellem leverandører af personlig og praktisk hjælp. Dette organiseres efter en Bestiller-Udfører-Modtager (BUM)-model, hvor myndighed og leverandør er to separate enheder. Denne giver mulighed for, at private leverandører kan konkurrere med kommunale og giver mere klarhed omkring ansvarsfordeling og afsatte ressourcer. Denne model har desuden medvirket til, at mange kommuner har valgt en centraliseret visitationsform på hjemmeplejeområdet [13]. Fordelingen af central, decentral og andre visitationsformer i hjemmeplejen hos landets kommuner i 2004 er illustreret på **figur 2.4 (a)**. Såfremt det gør sig gældende at central hjemmeplejevisitation, som for hjemmesygeplejevisitationen, inducerer subjektivitet i visitationen af indsatser, er objektiv beslutningsstøtte ligeledes relevant på dette område.

2.2.2 Styring og kvalitetsudvikling på ældre- og omsorgsområdet

Som beskrevet i afsnit 2.1 er den kommunale varetagelse af ældre- og omsorgsområdet udfordret af et tveægget pres, i form af krav til at begrænse udgifter, men samtidig varetage en stigende opgavemængde samt øge kvaliteten af de udbudte servicetilbud. For at muliggøre kvalitetsudvikling i de kommunale servicetilbud, skal der altså skabes økonomisk råderum, hvilket forudsættes af effektivisering. Som videns- og beslutningsgrundlag i effektiviseringsarbejdet, er det essentielt at identificere, hvordan de tilgængelige ressourcer udnyttes og prioriteres optimalt [14]. Her skal der tages højde for den igangværende udvikling på ældreområdet, der betyder at antallet af ældre stiger og at variationen i behovet for hjælp og pleje blandt den ældre befolkning, ligeledes er stigende. I denne forbindelse forventes det, at der vil være en andel ressourcestærke ældre med et minimalt behov for hjælp, en andel af ældre, hvis behov kan løses gennem hjælp til selvhjælp, og endelig en andel borgere med omfattende plejebehov. For at muliggøre forebyggelse og målrette indsatser, peges der på vigtigheden af at implementere værktøjer til tidlig opsporing af ændringer i borgernes plejebehov [7]. Der er dog stor variation i hvilken information kommunerne har til rådighed som grundlag for prioritering og styring på ældreområdet. De fleste kommuner har nøgletal tilgængelige omkring forbrug og budget, mens faglige resultater varierer i tilgængelighed og anvendelsesgrad. For at kunne effektivisere opgavevaretagelsen og målrette behandlinger mod netop de borgere, der vil få udbytte heraf, er det en forudsætning, at der følges op på effekt og kvalitet af de leverede ydelser. Sådan information kan efterfølgende anvendes i analyser af, hvad der er karakteristisk for borgere, der har gavn af en given indsats eller forløb. Dette kunne eksempelvis være specifikke egenskaber i forhold til borgerens helbredstilstand, sygdomsprofil, stamoplysninger eller andet. En sådan data-drevet tilgang til prioritering

og kvalitetsudvikling er også fremhævet i det fælles nationale kvalitetsprogram 2015-2018 [15], der beskriver hvordan *"Systematisk brug af relevante sundhedsdata om kvalitet, aktivitet og økonomi skal motivere og drive kvalitetsforbedringer hos sundhedspersonale, administratorer og ledelse"*. Det seneste årti er der arbejdet med at sikre en høj grad af dokumentation gennem detaljerede retningslinjer for dokumentationspraksis og understøttelse heraf gennem digitale løsninger, såsom Elektronisk Omsorgsjournal (EOJ)-systemer [7]. På trods af dette oplever mange kommuner stadig udfordringer på dataområdet [2]. Hidtil har dokumentationen på ældreområdet været fokuseret mod administrative og økonomiske formål, hvorfor der hovedsageligt er registreret data omkring tidsforbrug for de leverede ydelser. Dette skyldes til dels, at kommunernes afregning med leverandørende er baseret på timeprisberegninger ud fra det registrerede tidsforbrug. Dermed har der ikke været fokus på dokumentation af det faglige aspekt i indsatserne. Dette omfatter, blandt andet, borgerens tilstand, målsætning for forløbet og udfaldet heraf [7]. For at kunne vurdere hvornår en indsats har haft en positiv indvirkning på borgerens tilstand, kræver det en ensartet måde at vurdere effekten af en indsats og om målene hermed er opnåede. For at muliggøre sammenligning af et effektmål, kan det være nødvendigt at afgrænse til bestemte indsatser, da målet kan være forskelligt alt efter indsatsens natur og formål. Herfor er det en udfordring at opstille en model eller et mål, hvorudfra effekten af indsatser og forløb beskrives på en meningsfuld måde [2].

2.3 Fælles sprog III

KL lancerede i 1996 det første Fælles sprog (FS), der havde til opgave at skabe en overordnet begrebsramme til beskrivelsen af funktionsevnen for indvisiterede borgere i kommunerne, samt hvilke indsatser borgeren får. FS skulle på den måde skabe mulighed for udtræk af ensartede og sammenlignelige oplysninger om kommunernes borgere. Målet var større politisk gennemsigtighed med ressourceforbrug og serviceniveau til beslutningsstøtte i ressourceprioriteringer. [16]

De efterfølgende år steg fokusset på borgerinddragelse, retssikkerhed og kommunernes visitering samt dokumentationen heraf. Visitationerne blev centraliseret og specialiseret i takt med, at kommunerne skulle løse opgaver med stigende kompleksitet, der blev flyttet fra sygehusene til den kommunale hjemmepleje. Sammen med den teknologiske udvikling og udbredelsen af EOJ-systemer, steg forventningerne til, hvilke opgaver FS skulle kunne håndtere. Herfor lancerede KL i 2003 FSII. Målet blev "borgeren i centrum", med fokus på gennemsigtighed i forhold til borgerens behandling og pleje samt kvalitetssikring af den enkeltes borgers givne indsatser. [17]

I 2016 implementerede de første kommuner den seneste version af FS, FSIII. Den nyeste version dækker foruden hjemmepleje også rehabilitering og træning, ud fra serviceloven, samt hjemmesygeplejen, jævnfør sundhedsloven. Målet er interoperabilitet mellem sektorerne og øget borgertilfredshed i kraft af sammenhængende plejeforløb. Derudover er særligt de ledelsesmæssige styrings- og prioriteringsopgaver i fokus, for at danne grundlag for mere effektiv og lønsom arbejdstilrettelæggelse gennem struktureret og klassificeret data. [18]

2.3.1 Fælles Sprog III tilstande

FSIII tager udgangspunkt i en borgers tilstand på visiteringstidspunktet. Tilstandene opdeles i *funktionstilstande*, der relaterer sig til serviceloven og *helbredstilstande*, der relaterer sig til sundhedsloven.

Kategorien **funktionstilstand** indeholder 30 funktionstilstande, der grupperes i fem undergrupper, som alle beskriver en potentiel funktionsbegrænsning hos borgeren. Et eksempel på en undergruppe er *Egenomsorg*, hvortil en funktionstilstand er eksempelvis *Fødeindtagelse*. Dertil knyttes en tilstandspræcisering, der skal indeholde en niveauvurdering af tilstanden. Niveauet beskriver borgens selvstændighed i relation til funktionsbegrænsningen inddelt i niveauerne vist i tabel 2.1. Disse informationer vurderes, for den enkelte borger, af den udredende myndighed; det vil sige den visiterende aktør.

Tabel 2.1. Tilstandsniveauer i Fælles sprog III. Redigeret fra [18].

Niveau	Beskrivelse
0	Ingen eller ubetydelige begrænsninger
1	Lette begrænsninger
2	Moderate begrænsninger
3	Svære begrænsninger
4	Totale begrænsninger
9	Ikke relevant

Ydermere kan funktionstilstande suppleres yderligere med fritekst om vurderinger, årsager eller faglig notater til tilstanden.

Helbredstilstandene indeholder 43 tilstande der grupperes i 12 undergrupper, baseret på de sygeplejefaglige problemområder. Et eksempel på en undergruppe er *Udskillelse af affaldsstoffer*, hvortil en tilstand er *Problemer med mave og tarm*. Til helbredstilstanden skal der medfølge en vurdering af, hvordan denne kommer til udtryk, og desuden danner grundlag for borgerens fremtidige forventede tilstand. Som for funktionstilstande kan tilstanden suppleres med niveauer, årsager og faglige notater. [18]

Til alle tilstande opstilles en forventet tilstand. For funktionstilstande indebærer dette et forventet tilstandsniveau, mens det for helbredstilstande er struktureret som en fritekstvurdering af den forventede tilstand. Denne fritekst anbefales i FSIII, at indeholde en estimering af hvorvidt tilstanden forventes at "forsvinde", "mindskes" eller "forblive uændret". [18]

2.3.2 Fælles Sprog III indsatskatalog

Af FSIII - Metodehåndbogen fremgår det, at FSIII blandt andet, *"omhandler dokumentation af de visiterede/ordinerede indsatser og levering heraf"*. En indsats er en overordnet beskrivelse af den ydelse, der skal leveres hos borgeren. Til de forskellige tilstande kan der bevilliges en eller flere indsatser til borgeren, ud fra det udstedte indsatskatalog. Dette består af 31 indsatser til funktionstilstandene og 43 til helbredstilstandene. De forskellige indsatser er opgjort i KL's *FSIII Indsatskatalog (Sundhedslov)* [19] og *FSIII Indsatskatalog*

(Servicelov) [20]. Ifølge FSIII skal en indsats som minimum være knyttet til én tilstand og den tilhørende paragraf i sundheds- eller serviceloven. Et eksempel på en tilstand og en tilhørende indsats kan ses for en funktionstilstand i tabel 2.2, og for en helbredstilstand i tabel 2.3. En indsats kan suppleres med yderligere handlingsanvisninger på lokalt niveau, eksempelvis *hjælp til indtagelse af mad og drikke*. [18]

Tabel 2.2. Eksempel på en funktionstilstand og en relateret indsats fra Fælles sprog III indsatskataloget, dækket af serviceloven [18].

Funktionstilstand		Indsats
Egenomsorg	→	**§ 83 stk. 1, nr. 1 Personlig hjælp og pleje**
- Fødeindtagelse		- Ernæring

Tabel 2.3. Eksempel på en helbredstilstand og en relateret indsats fra Fælles sprog III indsatskataloget, dækket af sundhedsloven [18].

Helbredstilstand		Indsats
Udskillelse af affaldsstoffer	→	**§ 138 Sygepleje.**
- Problemer med mave og tarm		- Stomipleje

2.4 Udnyttelse af den registrerede ældre- og omsorgsdata

I 2019 fremlagde KL en analyse af, hvordan kommunerne anvender den registrerede data på ældre-/omsorgsområdet. I analysen rapporterede de kommunale medarbejdere, at de i høj grad anvender data til kvalitetsudvikling og planlægning, men i mindre grad til styring og prioriteringsformål. Modsat fremgik det, at data på et politisk og direktivt niveau, i høj grad blev brugt til styring og planlægning, men i mindre end 30% af kommunerne til kvalitetsudvikling. Kommunerne angav i samme rapport hvilke begrænsninger, de oplevede ved dataanvendelsen. Her fremgik det blandt andet, at over 40% af kommunerne oplever at *"Data er vanskelige at udtrække fra systemer o.l."*. Desuden rapporterer over 30%, at *"Udtræk og bearbejdning af data kræver specialiserede kompetencer"* samt at *"Der mangler ressourcer til at udtrække og behandle data"*. Analysen konkluderer, at registrering og udtræk af data fra fagsystemerne i kommunerne besværliggøres af manglende interoperabilitet med lokale løsninger til eksempelvis indsamling af borgerrapporterede oplysninger. Dog opleves der færre begrænsninger i de kommuner, hvor data anvendes meget, og hvor man har inkorporeret tekniske løsninger til at arbejde med data. Det fremgår dog ikke tydeligt af KL's analyse, hvilke tekniske løsninger, der er tale om. Ydermere fremgår det af analysen, at KL anser det for en succesfaktor, at kommunens medarbejdere oplever et tydeligt formål med anvendelsen af data til løsning af kerneopgaver. Det forventes derved, at kvaliteten og udnyttelsen af den data, der registreres kan øges. [21]

Gennem resultaterne fra den beskrevne analyse, belyses et behov for værktøjer, der kan støtte kommunerne i at gøre brug af den data, der registreres i EOJ-systemerne. FSIII indførte en fælles registreringspraksis på tværs af kommunerne, som kan skabe datagrundlag til tværsektorielle værktøjer, og potentielt føre til bedre brug af det obligatorisk registrerede data på ældre og -omsorgsområdet. Det bør deraf undersøges om denne standardiserede datamængde kan anvendes til understøttelse af styring og

prioritering gennem automatiserede, data-drevne værktøjer. Eksempelvis gennem FSIII indsatser, i relation til tilstandende og forbedringer i niveauerne heraf over tid.

2.5 Case: Hjemmehjælp i Faaborg-Midtfyn Kommune

For at identificere hvilke typer redskaber til prioritering, styring og kvalitetsudvikling, der er relevante og realistiske for hjemmehjælpsområdet i praksis, er der foretaget et interview med Faaborg-Midtfyn Kommune (FMK). Dette interview er udført som et semi-struktureret gruppeinterview med deltagelse af souschef, teknologikonsulent og ledelsesinformationskonsulent i forvaltningen for Sundhed og Omsorg i FMK. Desuden deltog en konsulent fra KMD. Den teoretiske baggrund og en komplet transskription af interviewet findes i appendiks B.

Gennem dette interview bekræftes KL's konklusion, beskrevet i afsnit 2.4, om at registrering og udtræk af data fra fagsystemerne i kommunerne ikke er tilstrækkelig. For FMK gælder det, at implementeringen af FSIII ikke har været strømlinet og standardiseret nok fra begyndelsen, hvilket har ført til forskellige fortolkninger og tilpasninger mellem kommunerne. Dette beskrives i følgende citat.

"...man arbejder meget ud fra 'hvordan plejer vi at gøre, og hvordan matcher det op imod FSIII, og hvordan synes vi så, det skal passe ind i vores kontekst'. Så derfor oplever vi ofte, at vi ikke kan sparre og dele opsætninger på tværs af kommuner, fordi vi har vores egen FSIII-opsætning. Vi oplever også, at det er rigtig, rigtig svært for medarbejderne at dokumentere efter det, og det kommer jo til syne i datakvaliteten også."
Kathrine Skov Kastrup, Teknologikonsulent, FMK.

Det beskrives altså, hvordan muligheden for generalisering på tværs af kommuner lider under forskellige og individuelle implementeringer af FSIII. Ligeledes indikeres det, at medarbejdernes adaption til dokumentationspraksis efter FSIII-standarden er en udfordring. Herved forurenes datakvaliteten af medarbejdernes forskellige fortolkninger af standarden. Dette kan eksempelvis være en registrering af samme type oplysninger under forskellige strukturerede datafelter, da meget af FSIII stadig er baseret på fritekst.

"Altså vi oplever, at selvom man er under samme hat; måske træning, så har man forskellig dokumentationspraksis fra lokation til lokation."
Kathrine Skov Kastrup, Teknologikonsulent, FMK.

Foruden manglende entydighed i FSIII-standarden opleves det, at medarbejderne i praksis tilsidesætter den strukturerede og klassificerede registreringsform til fordel for fritekstnotater. Dette kan være en indikation på, at den begrebsramme FSIII udbyder, ikke er tilstrækkelig til at repræsentere praksis. Konsekvensen heraf er, at information lagres i et format uden mulighed for kvantificering og databearbejdning.

"Når de så begynder at lave notater, der ligger i nogle rubrikker eller felter - eller hvad det hedder teknisk inde i systemet - hvor vi ikke kan trække ud, så har vi jo fuldstændig anarki på området."

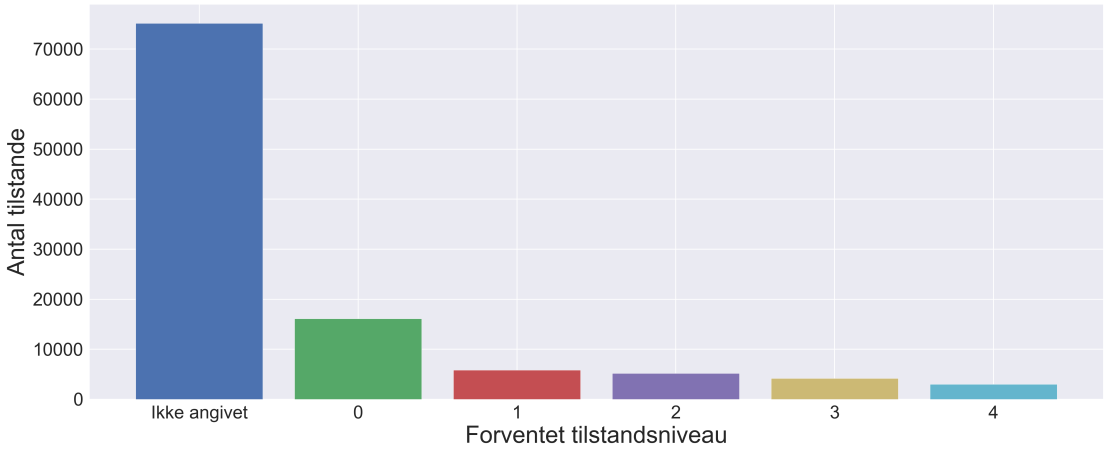
Allan Christiansen, Souschef for Sundhed og Omsorg, FMK.

Endvidere belyser det udførte interview mangel på effektmål og strukturering af målsætninger som en problemstilling, der hæmmer udnyttelsen af data på hjemmehjælpsområdet. Dette er evident i følgende citat.

"...det er visitationen, der bestiller en ydelse hos en leverandør; altså det kan både være private og kommunale leverandører. Dér bliver der sat nogle mål op, og hvis man så ikke kan måle progression hen mod et mål, så ved man jo aldrig, hvor langt man er nået. Så vi ved ikke, hvordan vi går fra A til B og så videre. Det er i hvert fald noget af det, jeg synes der halter i vores setup i forhold til at være styret efter en data-dreven organisation, fordi de data vi har, er jo ikke valide nok"

Allan Christiansen, Souschef for Sundhed og Omsorg, FMK.

Det ses altså, at der i forbindelse med målsætning mangler effektmål til at guide processen. Målsætning skal, ifølge FSIII, ske gennem dokumentation af forventet funktions- eller helbredstilstand, hvilket skal vurderes ved visitation af indsatsen, som beskrevet i afsnit 2.3. Undersøges FMKs data med fokus på FSIII vurderinger af forventede tilstande, opstillet i **figur 2.5**, bliver det tydeligt, at der i størstedelen af tilfældene, ikke er registreret en forventet tilstand.



Figur 2.5. Fordeling af antal indsatser i forhold til det registrerede forventede tilstandniveau. Data er udtrukket fra FMKs fagsystem, KMD Nexus.

Dette stemmer overens med udtalelsen om, at den effekt-relaterede data ikke er valid nok som grundlag for organisatoriske beslutninger. Dermed indeholder den registrerede data på hjemmeplejeområdet i FMK ikke effektmål, der direkte kan repræsentere den enkelte borgers effekt af en indsats. Dette understøttes yderligere af konsulenten for ledelsesinformation, der påpeger, at manglen på effektmål hæmmer muligheden for efterfølgende kategorisering af borgerne.

"Så man kan sige, det med at fastsætte nogle mål for 'hvad er det for en kategori den her borger tilhører', dér er vi slet ikke i dag. Og man kan sige, det med tilstande, dét er slet ikke noget vi er nået til at validere, så vi bruger det ikke som et aktivt arbejdsredskab, i

dag, til at se 'jamen, hvor er det vi forventer borgeren er om et halvt år og om et år?'. Der bliver sat nogle tilstande på, men det er slet ikke noget, vi trækker ud og bruger som data."

Jeppe Rosager Lund Pedersen, Konsulent for ledelsesinformation, FMK.

Samtidig fremhæves det i interviewet, at den data der er basis for afregning, det vil sige aktivitetsdata, forventes at være valide.

"Det vi primært har koncentreret os om, det er aktivitetsdaten: Hvor mange timer leverer vi, hvor mange borgere har vi inde og i den dur. Så det er på et meget lavt datamæssigt niveau stadigvæk." og *"Jamen altså inden for hjemmeplejeområdet, som jeg antager at I kigger på, i forhold til de leverede timer, der er vi på et validt niveau. Altså vi kan godt stole på data."*

Jeppe Rosager Lund Pedersen, Konsulent for ledelsesinformation, FMK.

For FMK er der altså data tilgængelig for aktiviteten omkring den enkelte borger, hvilket kan anskues som et mål for plejebehovet. Det kan overvejes, om denne data kan anvendes til udvikling af redskaber til effektiv prioritering. Behovet for denne type beslutningsstøtte er indikeret i følgende citat.

"Hele den ledelsesmæssige prioritering, i det jeg kalder en data-drevet organisation, det har ledelsen ikke mulighed for at gøre, og det mangler vi jo mega meget. Og så når vi er ressourceknappe, og det er alle kommunerne - så det er interessant for alle det her - hvordan kan man så, som afdelingsleder, prioritere sin tid i en visitation, der har for få visitatorer, og der ikke kan nå deres arbejde. Men hvad nu hvis de nåede det rigtige? og samtidig lod det rigtige, altså det, der ikke kunne gøres noget ved, ligge?"

Allan Christiansen, Souschef for Sundhed og Omsorg, FMK.

Lignende beskrives det af konsulenten for ledelsesinformation.

"Ja, eller hvis man kunne lave sådan nogle grupperinger. Lave nogle kategorier og sige: "Jamen de her borgere, med den her karakteristik af hjælp eller hvad I nu finder frem til af andre karakteristika, det er dem I skal foretage opfølgning på om et år eller om et halvt år. Fordi det er dét data viser, at det er ofte dér man vil se en forbedring i deres funktionsniveau og dermed vil I kunne opnå noget ved at foretage en revisitering"

Jeppe Rosager Lund Pedersen, Konsulent for ledelsesinformation, FMK.

Der forekommer dermed en efterspørgsel på prioriteringsredskaber, mere specifikt klassificeringsredskaber, så borgerne kan inddeles i fænotyper med lignende behov. Herudfra vil det være muligt at undersøge, blandt andet, målrettede indsatser og tidsintervaller for opfølgning. For at konstruere fænotyper med homogenitet i forhold til plejebehov samt muliggøre tidlig opsporing af ændringer heraf, bør der tages højde for udviklingen i plejebehovet over tid, og hvorvidt der findes generelle borgerkarakteristika, som er betydende for en given udvikling. Da der ikke er valid data for borgernes

tilstandsniveau, forventede tilstandsniveau og dermed heller ikke effekt, må udvikling af data-drevne prioriteringsredskaber baseres på aktivitetsdata.

2.6 Lignende studier

Inden for det sundhedsvidenskabelige domæne har de seneste år budt på studier, der fokuserer på at udvinde værdifuld information fra den store datamængde, der findes i kraft af systemer for Elektronisk Patientjournal (EPJ). Gennem data-mining-metoder som machine learning er der dokumenteret lovende resultater med at identificere mønstre i fremtidige medicinske tilstande eller diagnoser [22–25]. Disse nye metoder er dog ikke umiddelbart udbredt til applikationer på ældre- og omsorgsområdet. Dette på trods af, at den data der findes i EOJ-systemer bør kunne sammenlignes med data fra EPJ-systemer, i kraft af den longitudinale, strukturerede og heterogene karakter af dataen fra begge kilder. Herfor kan litteratur, der arbejder med data-mining-metoder på data fra EPJ-systemer, eller andre kilder til longitudinal sundhedsdata, anses som et vidensgrundlag for udvikling af tilsvarende værktøjer til gavn for den kommunale ældrepleje.

En af de udfordringer der er i forbindelse med data-mining af data fra EPJ-systemer, er håndtering af den temporale dimension, der grundet forskellig frekvens af dataregistrering resulterer i heterogenitet mellem patienters data. En udbredt måde at omgå denne udfordring er at aggregere dataen i antal tilfælde og frekvens. Herved mistes den information, der kan være indeholdt i den tidslige udvikling [22]. I et studie af Last et al. håndteres den temporale dimension ved at modellere hver patients månedlige medicinforbrug i 35 måneder som en 35-dimensionel vektor, der er input i en k-means clustering algoritme [26]. En anden tilgang fremsættes i et studie af Zhang et al., hvor frekvensen af udvalgte diagnosekoder suppleres med en "transition matrix", der beskriver tiden mellem forekomsten af to diagnoser, så et diagnosepar, der forekommer tæt sammen i tid anses for mere korrelerede end hvis der er lang tid imellem [22]. Endvidere, kan temporal data håndteres ved modeller, der er bygget til formålet. Et eksempel herpå er et Recurrent Neural Network (RNN), som blandt andet anvendes i et studie af Baghaei et al. [25], til prædikteringen af sepsis. Der er altså forskellige måder at håndtere temporal data i data-mining, alt efter hvad målet med modellen er.

Litteraturen har en række eksempler på studier, der forsøger at prædiktere eller klassificere på baggrund af temporal data. I et studie af Esteban et al. fremsættes en Temporal Latent Embedding model baseret på Markov embeddings og multilayer perceptron med Bernoulli likelihood som cost function, til prædikteringen af forløbet af hospitalsbesøg for nyresvigtspatienter. Integreringen af baggrundsdata for patienten samt hospitalshistorikken forbedrede signifikant modellen [24]. Endvidere fremsætter Forkan et. al. en Hidden Markov Model til tidlig prædiktion af anormale kliniske events ud fra longitudinale målinger af vitalparametre fra en række sensorer. Denne model opnår en accuracy på 97.7 %, men er dog begrænset til at kunne prædiktere for det næste 30-minutters tidsvindue. Foruden traditionelle machine learning modeller har flere studier undersøgt anvendelsen af *deep learning*. I denne sammenhæng har RNNs vundet indpas i de seneste år. Som beskrevet anvender Baghaei et al., et RNN, i bidirektionel udgave, som trænes med multivariat, temporal data med henblik på at prædiktere sepsis. For denne model opnås en maksimal accuracy på 75,5% [25]. Et studie af Kaushik et al.

benytter ligeledes et RNN, men i en variant med Long Short-Term Memory (LSTM). Denne type model overgik en traditionel time series forecasting model, Autoregressive Integrated Moving Average (ARIMA), i målet om at prædiktere patienters fremtidige forbrug af smertestillende medicin. Hvor RNN opnåede en Root Mean Squared Deviations (RMSD) på 13.693 mod 16.855 for ARIMA [27].

Generelt vinder en tankegang frem, om at sundhedsdata kan udnyttes til udvikling af avancerede, data-drevne værktøjer - ofte med et prædiktivt sigte. Appliceres dette perspektiv på hjemmehjælpsområdet, foreligger der er potentiale for effektivisering ved at kunne prædiktere, hvilken type plejebehov en borger vil få. Ud fra en sådan prædiktion, vil det være muligt at udvikle målrettede indsatser baseret på forskellige fænotyper af plejebehov.

Problemformulering 3

På trods af finanspolitisk prioritering de seneste år, er kommunernes udgifter til ældreområdet steget med omkring fire milliarder kroner fra 2007 til 2016 [6]. Denne stigning i udgiftsniveau reflekterer en tendens om, at det nære sundhedsvæsen skal imødekomme en stigende mængde plejebehov hos borgerne og samtidig opnå øget effekt af de udbudte plejeindsatser og behandlinger, uden yderligere økonomisk råderum [5]. Herfor er der behov for effektivisering af hjemmeplejen, hvilket forudsætter identificering af, hvordan de tilgængelige ressourcer udnyttes og prioriteres optimalt [14]. Her skal der tages højde for, den igangværende udvikling på ældreområdet, der betyder at antallet af ældre borgere stiger, samtidig med at variationen i behovet for hjemmepleje, og udviklingen heraf, er tiltagende blandt den ældre befolkning [7]. Såfremt dette differentierede ældrebillede kan udredes og borgere kategoriseres i fænotyper med tilsvarende udvikling i plejebehov, vil det være muligt at arbejde med målretning af indsatser samt prioritering af opfølgninger. Dette kan dermed være en metode til effektivisering, hvor de visiterede hjemmeplejeydelser tilpasses ud fra typen af borgere, der herved har større sandsynlighed for at opleve effekt af plejeforløbet. Desuden muliggøres tidlig opsporing af ændringer i borgernes plejebehov ud fra den givne fænotype, hvilket betragtes som en betingelse for forebyggelse og målretning af hjemmeplejeindsatser [7].

Generelt vidner litteraturen om et stort potentiale i at udnytte sundhedsdata til udvikling af avancerede, data-drevne værktøjer; særligt med et prædiktivt sigte. Appliceres dette perspektiv på hjemmehjælpsområdet, foreligger der er potentiale for, at anvende denne tilgang i udvikling af et værktøj til effektivisering ved prædiktion af, hvilken type plejebehov en borger vil få. Data på hjemmeplejeområdet genereres i stadig sigende kvantitet gennem EOJ-systemerne. Denne data er, til en vis grad, standardiseret efter dokumentationsstandarden FSIII. Undersøgelser af praksis viser dog, at der endnu ikke registreres mål for effekten af de udførte indsatser, idet dokumentationspraksis hovedsageligt er fokuseret mod administrative og økonomiske formål, og deraf hovedsageligt tidsforbrug for de leverede ydelser. Herfor kan identificeringen af fænotyper ikke guides af effektmål, men må anvende aktivitetsdata som et indirekte mål for plejebehov. Dermed kan det longitudinale forløb af plejebehovet inddrages, så fænotypeinddelingen af borgerne baseres på, hvorledes plejebehovet ændres som funktion af den første visitation. Hermed muligøres tidlig identificering af borgeres hjemmeplejefænotype, så forebyggende tiltag kan iværksættes. Disse muligheder og overvejelser leder frem til følgende problemformulering:

Hvordan kan data fra EOJ-systemer anvendes til udarbejdelse af et værktøj til identificering af fænotypeforløb for borgeres fremtidige, longitudinale hjemmeplejebehov samt prædiktion heraf, ud fra første visitation.

3.1 Problemafgrænsning

Den definerede problemformulering søges undersøgt gennem udviklingen af et *proof of concept* system til data-drevet fænotypeinddeling af hjemmeplejebehov samt klassifikation af en borgers fænotype, ud fra data registreret ved første visitation. På baggrund af dette formål udvikles en model der først og fremmest modellerer en række typiske plejeforløb gennem *clustering*, som efterfølgende kan udgøre klasse-labels i en *classifier*. Datagrundlaget for den implementerede clustering-algoritme er antallet af leverede minutters hjemmehjælp, da denne type aktivitetsdata anvendes i afregning for FMK og dermed har højest validitet. Dette er udledt af interviewet i afsnit 2.5.

For at etablere bedst mulige betingelser for dette proof of concept, afgrænses fokusområdet for at opnå et mere homogent datasæt. Derfor afgrænses målgruppen til modtagere af hjemmehjælp i form af praktisk hjælp og personlig pleje, hvorved varians induceret af eksempelvis hjemmeplejeforløb til handicap og misbrug elimineres. Desuden inkluderes kun ældre borgere, der i dette studie, defineres som borgere over 60 år. Afgrænsning er taget i anvendelse gennem begrænsning af datasættet, der beskrives nærmere i afsnit 4.2.

Metode 4

4.1 Studieoverblik

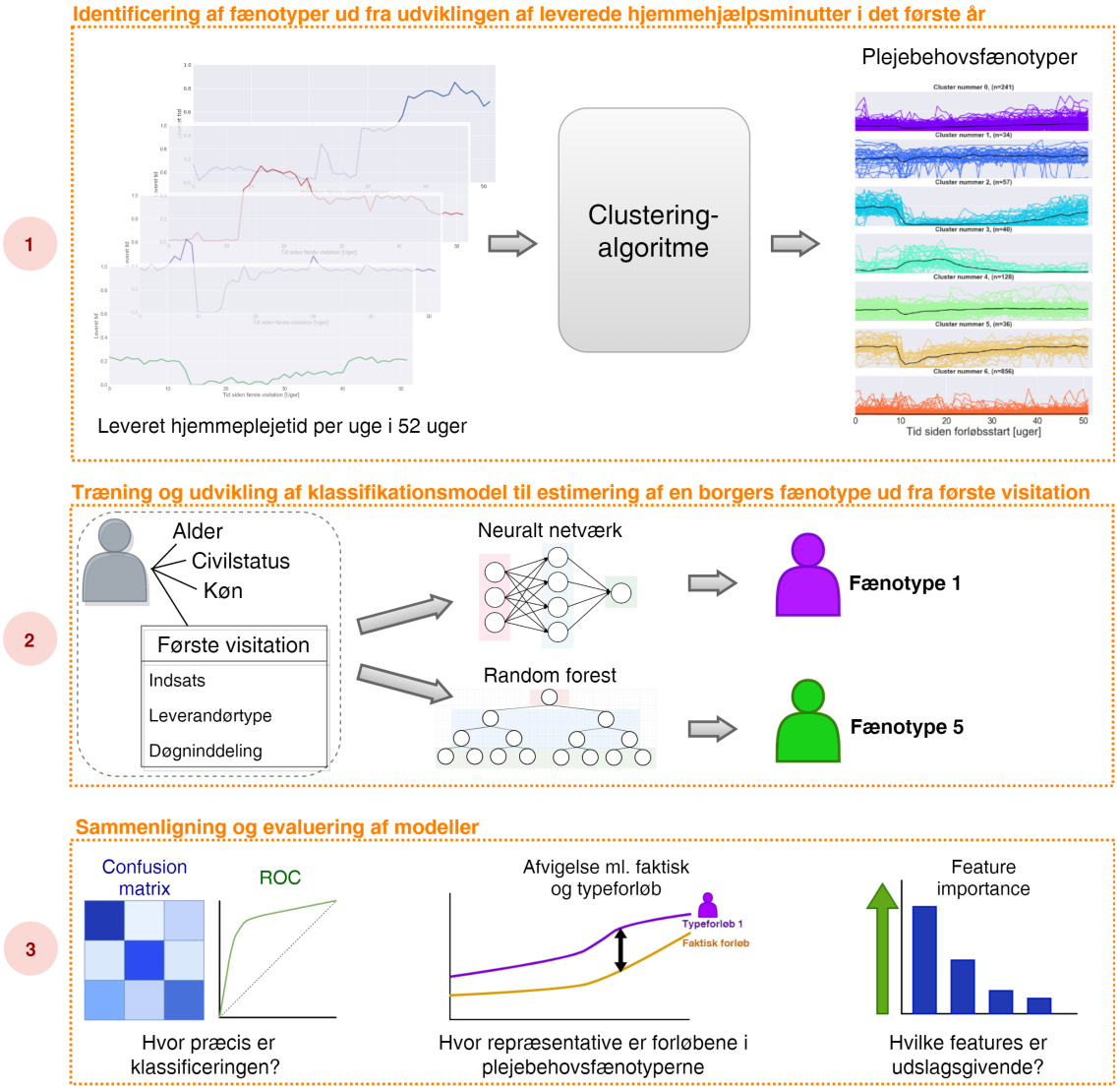
Som beskrevet i problemformuleringen i kapitel 3, er målet med dette studie at undersøge, hvorvidt data fra EOJ-systemer kan bidrage med værdifuld information til forklaring af det differentierede ældrebillede i relation til udviklingen i hjemmeplejebehov. Denne hypotese søges efterprøvet gennem udvikling af et data-drevet, prædiktivt fænotypeinddelings- og klassifikationsværktøj. De forskellige komponenter af denne udviklingsproces er illustreret i **figur 4.1**. Implementeringen af det beskrevne system er foretaget i Python med biblioteker som Keras, Scikit-learn og tslearn.

Først og fremmest udledes en række fænotyper for borgernes plejebehov. Dette realiseres gennem en *clustering*-algoritme, hvor input er hver borgers tidslige udvikling i antal minutters leveret hjemmehjælp. Dette er målt per uge i 52 uger. Gennem denne metode grupperes borgerne med lignende forløb i forhold til leveret pleje. For hver fænotype udregnes det gennemsnitlige forløb, som beskriver det generelle forløb i plejebehov for de respektive borgere. Dette betegnes som fænotypens *typeforløb*.

Fænotyperne, der er identificeret for hver borger gennem clustering, anvendes dernæst som tilstræbt resultat i træningen af to typer klassifikationsmodeller. Input til disse modeller er en række features om borgerens første visitation samt stamdata. Der trænes både en Random Forest (RF) og et Neural Network (NN) for at sammenligne modeller af forskellig kompleksitet. Det undersøges herved om det, med disse modeller, er muligt at klassificere en borger i den respektive plejebehovsfænotype og derigennem estimere forløbet i plejebehovet i det kommende år. Det endelige skridt i studiet indebærer sammenligning af performance for den implementerede RF- og NN-klassifikationsmodel, gennem confusion-matricer, Precision-Recall (PR)- og Receiver Operating Characteristic (ROC)-kurver. Derudover undersøges det, hvorvidt de udledte typeforløb, synes at være repræsentative for de individuelle forløb. Herfor udregnes differencen mellem hver borgers individuelle forløb af leverede minutter og typeforløbet for den respektive fænotype. Desuden undersøges det hvilke features, der er udslagsgivende for at en borger grupperes i en bestemt fænotype, gennem *feature importance*.

4.2 Data

Dataen anvendt i dette studie er stillet til rådighed af FMK. Dette gælder data fra hjemmeplejeområdet, som er genereret gennem EOJ-systemet KMD Nexus. Som beskrevet i afsnit 3.1, afgrænses nærværende *proof of concept* til, at omhandle hjemmepleje i form af praktisk hjælp og personlig pleje. Desuden er borgere med en alder på under 60 år



Figur 4.1. Oversigt for studiets tre dele, gennemgået i udviklingen af et data-drevet, prædiktivt fænotypeinddelings- og klassifikationsværktøj. Første del indebærer *clustering* af forløbet af leveret hjemmeplejetid over det første år, for hver borger. Disse fænotyper anvendes som klasselabels i anden del, der træner to forskellige klassifikationsmodeller på data fra borgerens første visitation. Endelig indeholder del tre en sammenligning af klassifikationsmodellerne, samt eksplorativ validering gennem *feature importance* og difference mellem de individuelle forløb af plejebehov og typeforløbet for den pågældende fænotype.

ekskluderet, for at fokusere datasættet på ældre borgere. For hver borger, består datasættet af to delmængder: (1) longitudinalt aktivitetsdata i form af antallet af leverede minutters hjemmehjælp pr. uge og (2) stamdata om borgeren og række variable om dennes indsatser, registreret ved første visitation. Et eksempel på sidstnævnte er opstillet i **tabel 4.2**, for en tilfældig borger. I alt består datasættet af 1334 borgere, hvilket opsplittes i et trænings- og testdatasæt af henholdsvis, 1067 og 267 borgere.

4.2.1 Borgerdemografi

Demografien i datasættet af borgere fra FMK, anvendt i nærværende projekt, er opstillet i **tabel 4.1**.

Tabel 4.1. Fordeling af demografiske egenskaber for borgerne i trænings- (n=1067) og testdatasættet (n=267).

	$n_{traening} = 1067(\%)$	$n_{test} = 267(\%)$
Alder (år ± standardafvigelse)	81.33 ± 9.14	81.19 ± 9.40
Antal indsatser ± standardafvigelse	2.23 ± 2.06	2.12 ± 1.96
Køn		
Kvinde	704 (65.98)	168 (62.92)
Mand	363 (34.02)	99 (37.08)
Civilstatus		
Gift	253 (23.71)	66 (24.72)
Ugift	100 (9.37)	27 (10.11)
Fraskilt	152 (14.24)	35 (13.11)
Samboende	1 (0.09)	0 (0)
Enke/-mand	561 (52.57)	139 (52.06)

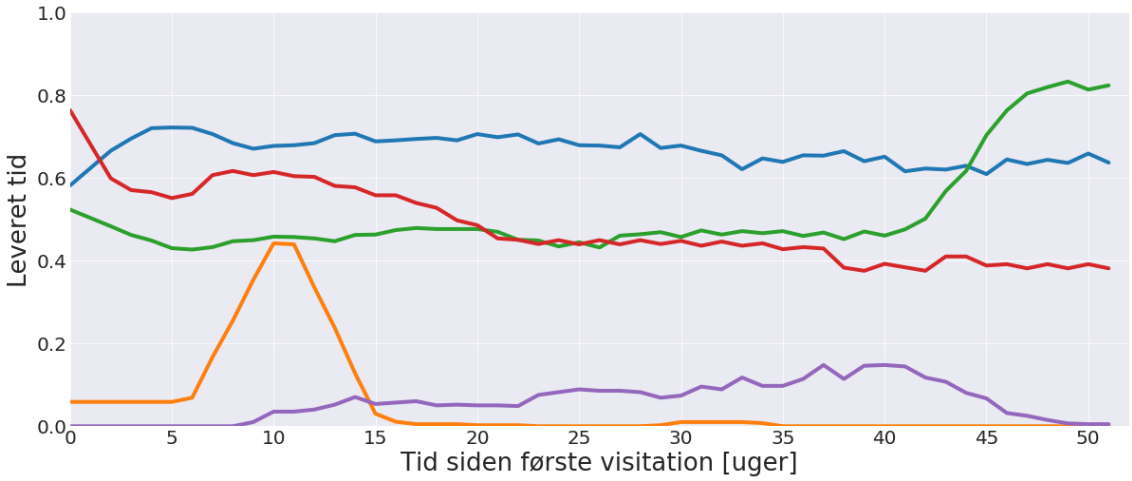
4.2.2 Datasegmentering og præprocessering

Da dataentiteterne for hver borger i det tilgængelige datasæt ikke er ensartede, er der foretaget en række segmenterings- og præprocesseringsoperationer, for at rense datasættet. Formået hermed er at etablere optimale betingelser for udvikling af en prædiktiv model, er dert. Først og fremmest er data renset for borgere, der ikke har data for et år, samt borgere hvor den longitudinale aktivitetsdata har værdien nul for alle datapunkter. Herudover er data præprocesseret forskelligt for de to delmængder, grundet deres forskelligartede karakter. Første delmængde består af tidsserier, mens anden delmænge er featurevektorer. Dette er beskrevet i de følgende afsnit.

Longitudinalt aktivitetsdata til fænotypeinddeling

Jævnfør det udførte interview med FMK i afsnit 2.5, registreres der endnu ikke data for effekten af leverede indsatser i hjemmeplejen. Samtidig blev det beskrevet, at aktivitetsdataen, der er udgangspunkt for afregning, kan betragtes som værende valide. Herfor anvendes aktivitetsdata i form af antallet af leverede minutters hjemmehjælp som et pseudomål for borgernes plejebehov. Det er ikke forventeligt, at den leverede tid ændrer sig signifikant fra dag til dag, hvorfor dataen er udtrukket per uge i et år, det vil sige i alt 52 uger. Det forventes at et års data er tilstrækkeligt til at udtrykke en tendens i udviklingen af borgerens plejebehov. Da det hovedsagligt er tendensen i forløbet der er af interesse, udglattes hvert forløb med et Savitzky–Golay filter, for at mindske betydningen af *outliers* og uregelmæssige udsving i forløbene. Savitzky–Golay filteret applicerer en foldning med et lav-ordens polynomium, fittet ud fra least squares. Dette gøres løbende i et vindue af datapunkter i forløbet [28]. Det anvendte filter har en vindueslængde på fem datapunkter, og der anvendes et lineært førsteordens polynomium, hvilket tillader interpolering af grænseværdier. Desuden er den longitudinale

data efterfølgende normaliseret til værdier mellem 0 og 1, i forhold til det globale maksimum i datasættet, for at sikre sammenlignelighed mellem clustering-modellerne gennem udligning af feature range. Som eksempel på denne type data, er den visualiseret for fem tilfældige borger i **figur 4.2**.



Figur 4.2. Eksempel på longitudinalt aktivitetsdata, i form af antal leverede minutters hjemmehjælp, for fem tilfældige borgere. Dette data er præprocesseret gennem trunkering til 52 uger, udglatning og normalisering i forhold til det globale maksimum i datasættet.

Stam- og visitationsdata fra først visiterede indsatser til brug i klassifikationsmodel

Anden delmængde af data anvendes som input-features til den implementerede klassifikationsmodel. Disse features udgøres af en række variable om hver visisteret indsats og borgeren, registreret ved borgerens første visitation. Et eksempel på en tilfældig borgers visiterede indsatser, ved første visitation er illustreret i **tabel 4.2**. Da alle features, på nær alder, bestilt og planlagt tid, er kategoriske variable, skal disse konverteres til numeriske værdier, for at kunne anvendes i det implementerede neurale netværk. Herfor konverteres de kategoriske variable til *one-hot*-vektorer, hvor hver mulig kategori indsættes som kolonne, og variablens værdi markeres med 1 i den pågældende kolonne, mens de øvrige er nul. Featurevektorerne for borgerne har forskellige størrelser, da der kan være visiteret flere indsatser til en borger ved første visitation. For at sikre at datamængden omkring den første visitation har samme dimensionalitet for alle borgere, komprimeres data for hver borger til en én-dimensionel vektor af samme længde. Dette gøres vednat summere de numeriske variable (Bestilt tid og Planlagt tid), samt one-hot variablerne. Desuden kombineres disse med stamdata-variablerne, i form af alder, køn og civil status.

4.3 Clustering af longitudinal udvikling i plejebehov

Til identificering af fænotyper, for forløbet af hjemmeplejebehov, appliceres en clusteringalgoritme på det longitudinale datasæt over leverede minutters hjemmepleje i borgerens første år med hjemmepleje. Bagrunden for at anvende den tidslige udvikling af leveret hjemmeplejetid er antagelsen om, at denne indeholder information om, hvorvidt de første visitationer borgeren tildeles er tilstrækkelige for at fastholde eller nedbringe

	Indsats 1	Indsats 2
Bestilt tid (minutter)	115.5	10.5
Døgninddeling	Ukendt	Ukendt
Alder	94	94
Køn	Mand	Mand
Civilstatus	Enke/-mand	Enke/-mand
Paragrafgruppe	Praktisk hjælp	Praktisk hjælp
Indsats (Niveau 01)	Servicelov	Servicelov
Indsats (Niveau 02)	FSIII	FSIII
Indsats (Niveau 03)	4.0 Praktisk hjælp i hjemmet §83.2	4.0 Praktisk hjælp i hjemmet §83.2
Indsats (Niveau 04)	4.2 Rengøring	4.2 Rengøring
Leverandørtype	Frit valg	Frit valg
Indsats arbejdsgang	Kalenderbestilt arbejdsgang	Kalenderbestilt arbejdsgang
Planlagt tid (minutter)	110	10
Besøgstype	Besøg	Besøg

Tabel 4.2. Eksempel på data for en tilfældig borger i den delmængde af data, der indeholder stamdata om borgeren og variable om dennes indsatser, registreret ved første visitation. Den inkluderede borger har fået visiteret to indsatser ved første visition, beskrevet i hver deres kolonne.

plejebehovet for den pågældende borger. Data for ét år udgøres af summen af leverede hjemmepleje-minutter per uge, hvorved hver borgers plejebehov repræsenteres som en 52-dimensionel vektor. På baggrund af denne højdimensionalitet, anses en clustering-algoritme som et optimalt værktøj til identificering af fænotyper. Den teoretiske baggrund og implementering for den anvendte clustering-algoritme er beskrevet i det følgende.

Inden for mønstergenkendelse og klassificering af data er cluster-analyse en ofte anvendt metode. Clusteranalyse inddeler datasættet i grupper, eller *clusters*, der naturligt forekommer i den undersøgte data. Blandt de forskellige metoder til clustering anses metoder, der ikke trænes op mod et kendt target, som de mest stabile og objektive metoder. Disse typer metoder betegnes som *unsupervised learning*. Herved udledes mønstre i dataen uden at inkludere forhåndsviden, også kaldet *a priori viden*, omkring grupperinger i dataen. Dermed er de features, der anvendes til at beskrive inputtet afgørende for, hvilke clusters der identificeres. Endvidere kan resultatet af en unsupervised clustering ikke anses som sand eller falsk, men bør bedømmes ud fra anvendeligheden af de fundne klasser. [29, Kap. 1].

Til clustering i dette projekt sammenlignes metoderne K-Means Clustering (KMC) og Hierarkisk Agglomorativ Clustering (HAC). Denne sammenligning foretages, da clustering performance for disse modeller generelt er applikationsafhængig. Netop disse metoder er udvalgt på baggrund af projektets datamæssige vilkår, da de kan håndtere et mindre datasæt og longitudinal, og deraf multidimensionel data. Clustering-metoderne og implementeringen heraf beskrives i de hertil dedikerede afsnit 4.3.3 og 4.3.2.

Cluster-inddeling baseres typisk på et similaritets- eller dissimilaritetsmål, der beskriver ligheden mellem to observationer. Denne type mål samles under betegnelsen *proximity*. Proximity afhænger særligt af data-skalen (kategorisk, kontinuert eller binært) for

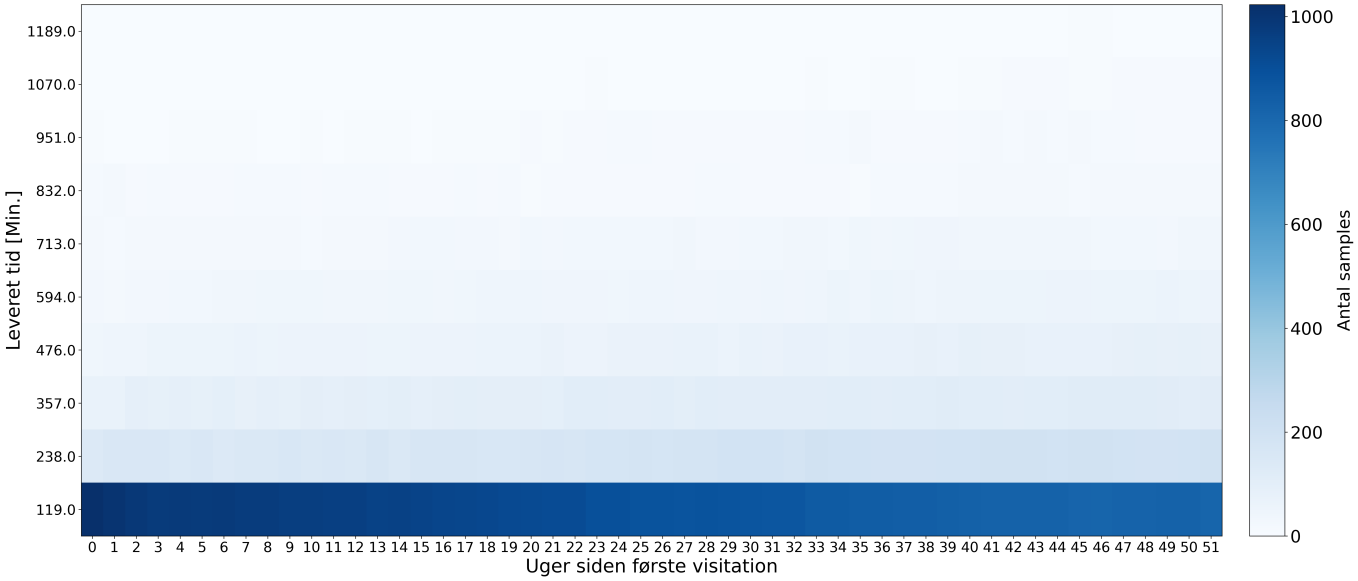
de undersøgte features. For kontinuert data repræsenteres proximity typisk med et distancemål, der kvantificerer afstanden mellem to observationer i feature space. Dermed er formålet med clustering at danne clusters med lavest mulig intra-cluster afstand og højest muligt inter-cluster afstand. Det mest populære distancemål er den euklidiske distance givet ved formel (4.1) [29, Kap. 3]

$$d_{ij} = \left[\sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right]^{\frac{1}{2}} \tag{4.1}$$

Hvor x er værdien af den k'te feature i det p-dimensionelle rum for individerne i og j [29, Kap. 3]. Euklidisk distance anvendes som distancemål i nærværende projekt, da dette er et anerkendt og velprøvet enhed, med intuitiv fortolkning i forhold til fysisk afstand.

4.3.1 Præ-clustering analyse

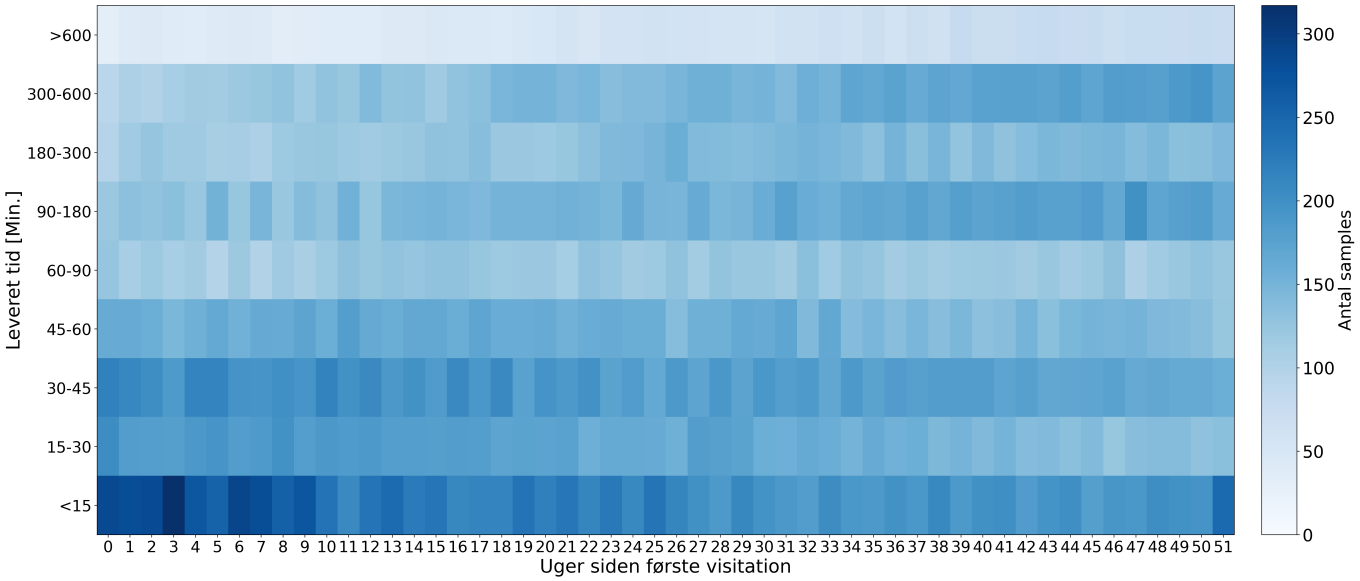
For at undersøge hvorvidt der forkommer tydelige grupperinger i data, indebærer første skridt i clustering typisk en visuel inspektion af det *feature space*, datasættet udspænder. Hertil anvendes eksempelvis histogrammer, scatter-plots eller dimensionalitetsreducerende metoder, såsom Principal Components Analysis (PCA) eller t-Distributed Stochastic Neighbor Embedding (t-SNE). Mangel på synlige grupperinger i dataen ved visuel inspektion betyder endog ikke, at grupperinger i dataen kan udelukkes. I nogle tilfælde vil visualisering af multivariat data, fremstille unimodale feature-distributioner, på trods af påviste grupper [29, Kap. 2]. Forud for clustering er det anvendte datasæt med longitudinalt aktivitetsdata analyseret gennem et *heat-map*, som illustrerer observationerne for hver relativ uge fordelt på 10 bins. Dette er inkluderet i **figur 4.3**.



Figur 4.3. Heatmap over det longitudinale antal leverede hjemmepleje-minutter.

Heraf ses det, at dataen primært var koncentreret omkring lave værdier af antal leverede hjemmeplejeminutter. Dette mønster går igen for alle 52 uger. For at udligne denne ubalance, er en alternativ udgave af den longitudinale aktivitetsdata beregnet. Her blev 10

intervaller eller trin, for datapunkternes værdi opstillet, ud fra et ønske om at opnå samme antal observationer i hvert trin. Overvægten af lave dataværdier resulterede i to trin med samme grænseværdi, hvorfor disse blev reduceret til én. Det endelige antal trin var dermed ni. Grænserne for disse trin blev tilpasset for at give intuitiv mening i relation til antal leverede minutter, eksempelvis blev den udregnede grænse på 55 minutter rundet op til 60 minutter. Et heatmap for det alternative, trininddelte datasæt er opstillet i **figur 4.4**.



Figur 4.4. Heatmap over det longitudinale antal leverede hjemmepleje-minutter.

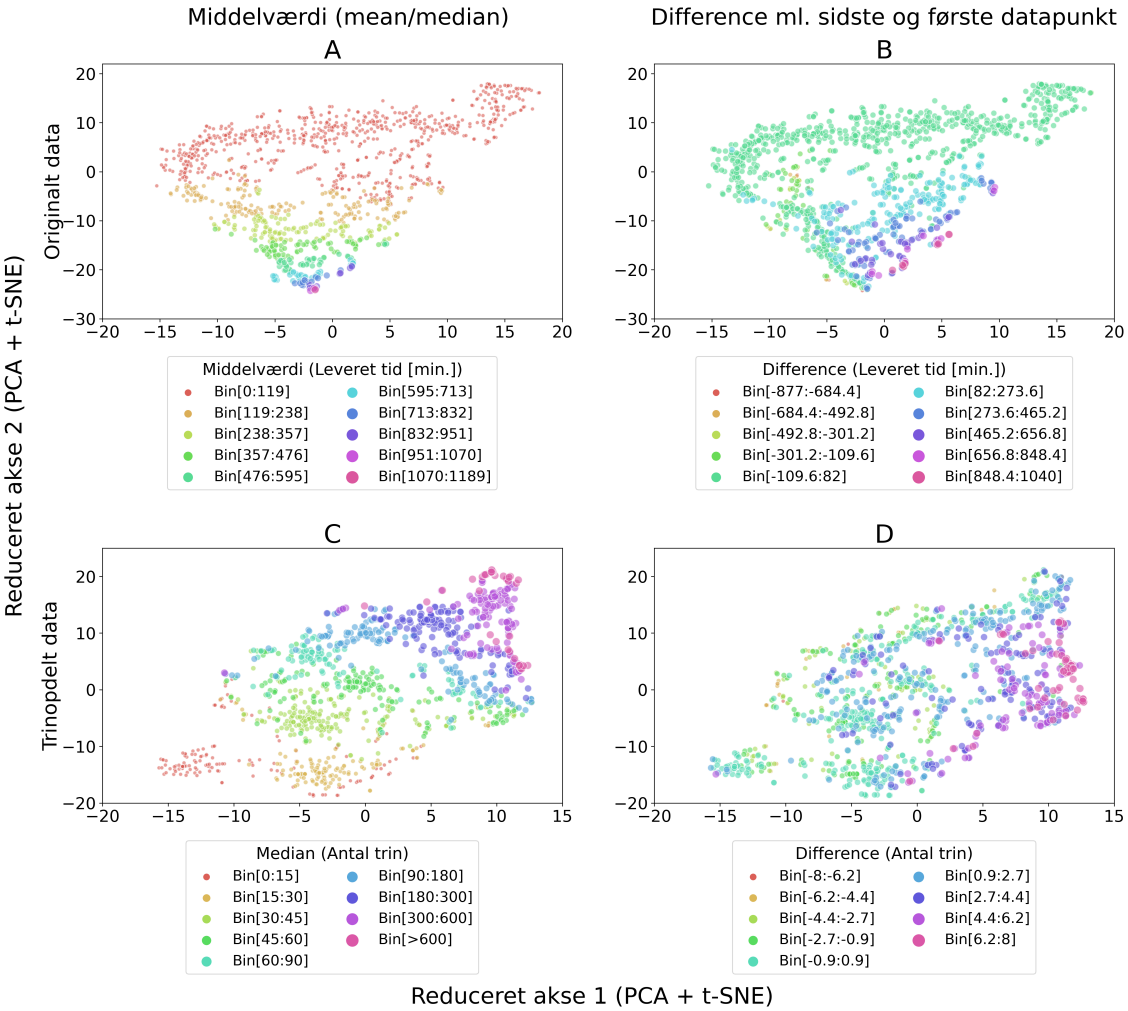
Den trininddelte data giver dermed en mere balanceret fordeling af observationer ved at tilføre mere finmasket niveauopdeling for de lave værdier, mens niveauerne for det høje område af spektret er inddelt i grove niveauer. Af **figur 4.4** ses nu også en stigning i leveret tid i takt med antal uger siden første visitation. Hvor en høj koncentration af observationer i de første uger primært findes i trinnet "<15", mens der, for de sidste uger, ses en øget koncentration i trinnene "90-180"og "300-600".

Ydermere undersøges den longitudinale data gennem en PCA samt efterfølgende t-SNE, for at undersøge hvordan borgerne fordeler sig i feature space. Gennem PCA, transformeres dataen til et lineært rum med lavere dimensionalitet. Målet for denne transformation er at bevare så meget som muligt af datasættets variation, og at observationer der ikke ligner hinanden placeres langt fra hinanden i det reducerede feature space. PCA fungerer ved, at finde en lineær transformation til det orthonormale rum som dataens eigenvektorer udspænder. Heri er variansen i dataen repræsenteret ved eigenværdierne. De features der beskriver mest af variansen, kan nu udvælges ved at sortere eigenværdierne fra højest til lavest og anvende en tærskelværdi for et specifikt antal eigenværdier eller hvor meget af variansen, der ønskes repræsenteret [30].

Efterfølgende appliceres t-SNE på det PCA-reducerede feature space, for at effektuere en yderligere reduktion ned til to dimensioner og dermed muliggøre visualisering i et scatter-plot. I t-SNE er formålet at placere lignende observationer tæt på hinanden i det reducerede feature space. Dette opnås ved at opstille en betinget sandsynlighedsfordeling for samhørsforholdet mellem hvert sæt observationer i feature space. Denne estimeres

ved at modellere afstandende mellem et *point-of-interest* og alle de øvrige punkter som en normalfordelingskurve centreret i det pågældende point of interest. Dermed vil sandsynligheden for at et givent punkt har en nærtliggende observation som nabo, være større end for et punkt længere væk. Det reducerede feature space opstilles nu i henhold til disse sandsynlighedsfordelinger, så punkter med høj sandsynlighed for at være naboer placeres tæt sammen [31].

Det reducerede feature space, beregnet gennem PCA og t-SNE er illustreret i **figur 4.5**, for både det originale longitudinale datasæt og det trininddelte. For at undersøge hvorvidt borgerne er naturligt grupperet i relation til plejebehovet, er de tildelt en farve efter henholdsvis gennemsnitsværdien af forløbet eller differencen mellem det sidste og første datapunkt. Disse aggregerede værdier inddeles i 10 bins af samme størrelse og farves i henhold hertil.

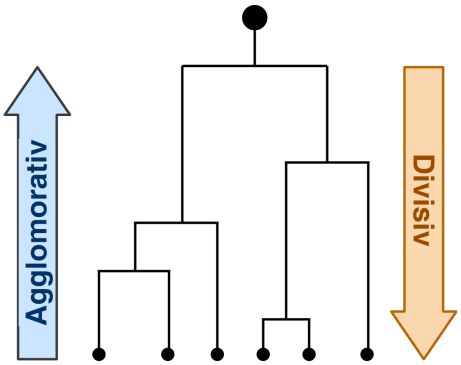


Figur 4.5. Kombineret Principal Component Analysis og t-Distributed Stochastic Neighbor Embedding for 52-dimensionel, longitudinal aktivitetsdata, i form af antal leverede hjemmeplejeminutter. Hver borger er repræsenteret ved et datapunkt i det reducerede feature space. Desuden er hver borger tildelt en farve, ved at aggregere borgerens forløb til henholdsvis en middelværdi (venstre kolonne) og differencen mellem sidste og første datapunkt (højre kolonne). Denne analyse er appliceret på det orginale datasæt (øverste række) og på et trininddelt datasæt (nederste række).

Her ses det, at borgere med tilsvarende gennemsnitsværdi for antal leverede hjemmehjælpsminutter over de 52 uger i datasættet er organiseret i forholdsvis afgrænsede områder. Dette er forventeligt eftersom det reducerede feature space er en repræsentation af de 52 datapunkter på uge-niveau. Da to lignende forløb vil producere et lignende gennemsnit, vil de dermed også ligge tæt i det reducerede feature space. Samtidig ses det, særligt for det originale datasæt i **figur 4.5 (B)**, at der også er en gruppering i det reducerede featurespace i relation til ændringen i antal leverede hjemmehjælpsminutter fra sidste datapunkt til første. Ved sammenligning af **figur 4.5 (A)** og **(B)**, ses en indikation på, at borgere der oplever en lav difference mellem sidste og første datapunkt, (*Bin[-0.09:0.07]*), på **(B)** hovedsageligt tilhører gruppen med en lav gennemsnitsværdi, (*Bin[0:119]*), på **(A)**. Tilsvarende ses det ved sammenligning af **figur 4.5 (C)** og **(D)**, at gruppen med højst ændring i plejebehov, (*Bin[0.69:0.89]*), på **(D)** primært findes i området med høje gennemsnitsværdier, (*Bin[300:600]* og *Bin[>600]*) på **(C)**. Alt i alt viser de reducerede feaure spaces, at der findes grupperinger i det longitudinale datasæt, både i originalt og trininddelt format, med relation til det gennemsnitlige plejebehov og udviklingen heraf. Dermed er der en indikation på, at borgere der ender ud med en specifik forskel fra udgangspunktet, vil gennemgå et lignede forløb i mellemtiden.

4.3.2 Hierarkisk clustering

Den mest fundamentale tilgang til clustering er hierarkiske metoder, der er let tilgængelige og alsidige. Hierarkisk clustering foregår sekventielt, enten agglomorativt eller divisivt. En agglomorativ hierarkisk clustering tager udgangspunkt i, at alle observationer udgør hver deres cluster, og fusioneres derefter sekventielt til større og større clusters. Modsat starter en divisiv hierarkisk clustering med én samlet cluster, der gradvist splittes i finere og finere clusters. Disse typer hierarkisk clustering er illustreret i relation til et dendrogram i **figur 4.6**. Hvorledes henholdsvis fusion eller split af clusters skal foretages, bestemmes ud fra maksimering af similaritetsmålet for de nye clusters. En ulempe ved de hierarkiske metoder er, at de kan opleve problemer ved at identificere ikke-sfæriske clusters samt ved håndtering af store mængder data på grund af tidsforbrug. [29, Kap. 4]



Figur 4.6. Agglomorativ og divisiv clustering. Førstnævnte tager udgangspunkt i, at hver observation udgør en individuel cluster, hvorefter disse fusioneres sekventielt til større clusters. Divisiv clustering anser alle observationer som én enkelt cluster, der splittes i mindre clusters. Modificeret fra [30, Kap. 10].

For den implementerede hierarkiske clustering-algoritme anvendes agglomorativ clustering og et similaritetsmål baseret på euklidisk distance, i henhold til *Wards method* [32].

Denne metode indebærer, at der for hvert sekventielt step samles clusters på baggrund af opdatering af en kriteriefunktion, der beskriver variansen i clusterne og dermed ønskes minimeret. Variansen i en cluster evalueres udfra et distancemål i det 52-dimensionelle feature space udspændt af den longitudinale data. Den implementerede kriteriefunktion er summen af den kvadrerede euklidiske distance mellem punkterne i hver cluster. Det reducerede matematiske udtryk for dette kaldes Sum of Squared Deviations (SSD). Denne beregning, fremsat i formel (4.2), foretages hver gang to clusters fusioneres.

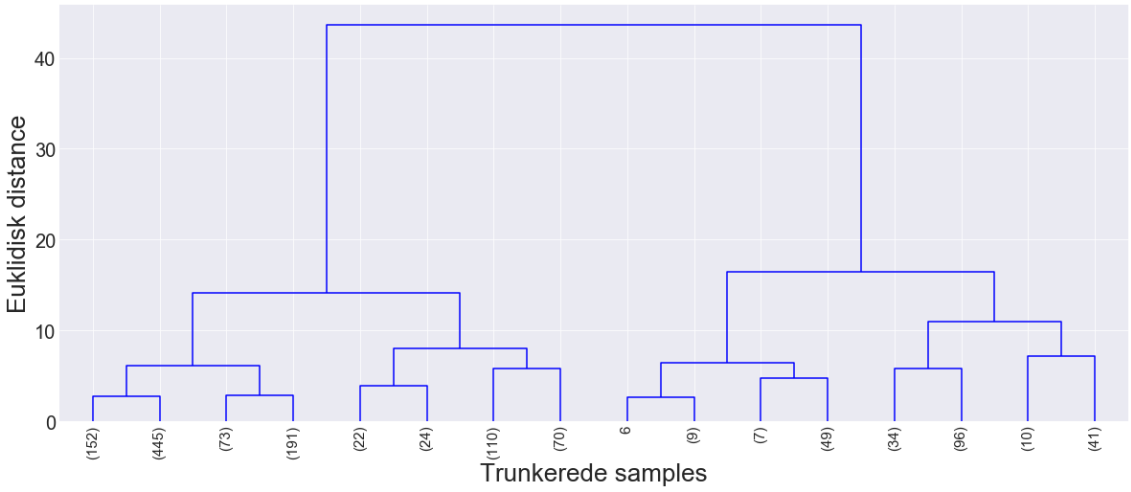
$$E = \sum_{m=1}^{g} E_m \tag{4.2}$$

hvor

$$E_m = \sum_{l=1}^{n_m} \sum_{k=1}^{p_k} (x_{ml,k} - \overline{x}_{m,k})^2 \tag{4.3}$$

$\overline{x}_{m,k}$ repræsenterer gennemsnittet af cluster m ($m = 1...g$) for variabel k ($k = 1...p_k$). $x_{ml,k}$ er værdien for variabel k for datapunkt l ($l = 1...n_m$) i cluster m. [32] [29, Kap. 4]

Den resulterende hierarkiske clustering kan inspiceres ved hjælp af et dendrogram-plot, der visualiserer den agglomorative fusion af clusters for hvert sekventielt step [29, Kap. 4]. Et dendrogram viser den hierarkiske clustering-proces, ved at plotte den hierarkiske forgrening i antal clusters sammen med similaritetsmålet på y-aksen. For det implementerede tilfælde, udgøres y-aksen af dissimilaritet, i form af euklidisk distance. Dermed vil et større spring på y-aksen markere, at det nye split nedbringer dissimilariteten meget. Et split på dendrogrammet repræsenterer, at den forudgående cluster, markeret ved en enkelt gren, opdeles i to nye clusters. Dendrogrammet for den implementerede HAC er fremsat i **figur 4.7**.



Figur 4.7. Dendrogram for den implementerede Hierarkisk Agglomorat Clustering (HAC) med Wards metode og euklidisk distance.Grupperne er trunkerede ved over fire splits for overskulig visualisering.

Det inkluderede dendrogram er trunkeret efter fire split i samme initierende gren for at lette visualiseringen. Af dendrogrammet ses det, at der ved fire splits i hver gren er en majoritetsgruppe på 33.4% ($n = 445$) af det totale antal borgere ($N = 1334$). Ud fra dendrogrammet kan en grænseværdi bestemmes, hvorved modellen inddeler data i et antal clusters mellem 1 og N, afhængig af antallet af clusters i det bestemte niveau. I dette studie evalueres de resulterende clusters, for en række forskellige grænseværdier, forud for beslutning om et endeligt antal clusters. Dette beskrives nærmere i afsnit 4.3.4.

4.3.3 K-means clustering

Til sammenligning af HAC undersøges og implementeres en KMC, der inddeler dataen i et forudbestemt antal, k, clusters. Overordnet fungerer KMC ved at antage k forskellige datapunkter i feature space, som værende centrum i de k clusters, der ønskes identificeret. De resterende punkter klassificeres dernæst med et cluster-tilhørsforhold, ud fra hvilket af de k centrum-punkter de ligger tættes på i feature space. Hver gang en ny observation tilføjes en cluster, beregnes et nyt centrum ud fra gennemsnittet af alle observationer i den pågældende cluster. Alle datapunkters cluster-tilhørsforhold evalueres nu i forhold til de nye centre. Processen gentages indtil intra-cluster distancen til centrum konvergerer, så det nye centrum ikke er signifikant forskelligt fra det forrige. Dermed fungerer k-means algoritmen som en optimeringsprocess, hvor målet er at maksimimere similariteten i hver cluster. Typisk anvendes summen af den kvadrede euklidiske distance, også kaldet inerti eller *within-cluster sum of squares*, som similaritetsmål [29, Kap. 5] [30, Kap. 10]. Dette similaritetsmål er implementeret i nærværende projekt.

Longitudinalt data, der er en-dimensionelt over tid, håndteres i KMC ved at behandle det som en n-dimensionel featurevektor, hvor hvert tids-datapunkt anskues som en individuel feature [29, Kap. 7]. For datasættet anvendt i dette projekt betyder dette, at antallet af leverede minutters hjemmehjælp for hver uge, vil udgøre features i en 52-dimensionel vektor. Hver borgers plejebehov kan dermed repræsenteres som et punkt i et 52-dimensionelt feature space udspændt af denne feature-vektor.

Ligesom for HAC er det en svaghed for KMC, at identificere ikke-sfæriske clusters. En yderligere svaghed ved KMC er, at de resulterende clusters og kvaliteten heraf er sensitiv overfor hvilke datapunkter, der vælges som de initierende centre. Traditionelt vælges disse tilfældigt, hvilket kan give suboptimale clusters, hvis et af de initierende centre eksempelvis er et yderpunkt eller hvis to initierende centre ligger tæt på hinanden. Herfor implementeres KMC algoritmen som *k-means++*, for at opnå en klogere initalisering. K-means++ er en metode, hvor kun ét af de initierente centrum-punkter vælges tilfældigt, hvorefter det næste centrum-punkt vælges ud fra at opnå den maksimale kvadrerede afstand til det tilfældigt valgte centrum. Dette fortsættes indtil alle k centrum-punkter er valgt. Det sikres dermed, at de initierende centrum-punkter ligger langt fra hinanden og risikoen for, at to initierende centrum-punkter tilhører samme cluster nedsættes. [33].

Endnu en svaghed ved KMC er, at antallet af clusters, modellen vil finde, er en hyperparameter, der skal bestemmes på forhånd. Dette kan bestemmes ud fra empiri eller ved udregningen af objektive mål, kaldet *cluster metrics*. For applikationen i nærværende projekt, hvor antallet af clusters i de longitudinale udviklinger i plejebehov ikke er givet, bestemmes antallet af clusters ud fra denne type mål. Dette er beskrevet i det følgende

afsnit.

4.3.4 Komparativ analyse og bestemmelse af antal clusters

Til sammenligning af de implementerede modeller, HAC og KMC, samt til bestemmelse af det optimale antal clusters, kan en række cluster metrics beregnes. Disse giver et kvantificerbart udtryk for kvaliteten af de resulterende clusters ud fra inter-cluster separation og intra-cluster homogenitet.

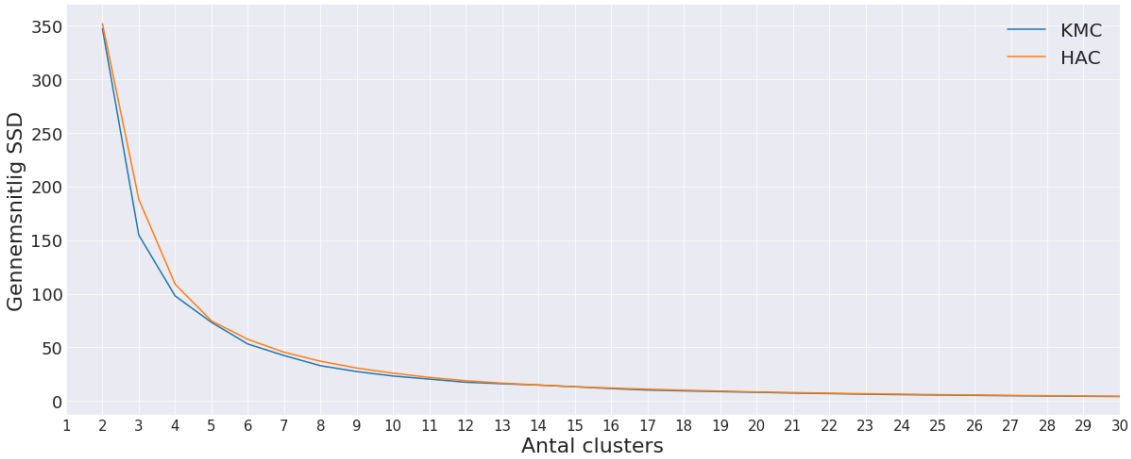
Der findes en lang række cluster metrics til bestemmelse af antallet af clusters, men fælles for dem alle er, at de ikke nødvendigvis giver meningsfyldte resultater i sig selv. Dette skyldes at der lægges vægt på forskellige aspekter af clustering-opgaven ud fra de bagvedliggende partionsmål. Herfor gennemgås resultaterne for flere af de mest populære mål. Dette inkluderer: Elbow, silhouette-score og Davies-Bouldin indeks. Disse er beskrevet i de følgende afsnit og understøtter beslutningen om det endelige antal clusters.

Det tages desuden i betragtning, at kvaliteten af en clustering-algoritme bør evalueres i forhold til den pågældende applikation. For nærværende projekt skal de resulterende clusters anvendes til beskrivelse af en borgers typiske hjemmeplejebehov ud fra det gennemsnitlige forløb i den pågældende borgers tildelte cluster. Dermed er det et succeskriterie for clustering-algoritmen, at dataen inddeles fint nok til at kunne adskille forskellige tendenser i udviklingen af plejebehovet. Herudfra følger det, at det gennemsnitlige forløb for en cluster vil være repræsentativt for de individuelle medlemmer i den pågældende cluster. Ydermere vurderes det, at et højt antal clusters ikke vil være tilstrækkeligt unikke og dermed vanskelige at fortolke i forhold til anvendelse som plejebehovs-fænotyper. Ud fra disse succeskriterier evalueres et antal clusters, k, mellem to og 30.

Elbow metoden:
Elbow-metoden indebærer, at et similaritetsmål evalueres som funktion af antallet af clusters, k. Similaritetsmålet vil forbedres i takt med, at antallet af clusters øges som følge af at variansen i de eksisterende clusters nedbrydes til flere, mere kompakte clusters. Det optimale antal clusters forekommer, når ændringen i similaritetsmålet nærmer sig konvergens. Dette punkt beskrives som kurvens 'albue', og efter dette punkt vil en yderligere forøgelse i antallet af clusters ikke bidrage til væsentlige forbedringer i similaritetsmålet. En svaghed ved elbow-metoden er, at vurdering af den præcise placering af kurvens albue er underlagt subjektivitet hos observatøren. [34].

For de implementerede clustering-modeller anvendes elbow-metoden med gennemsnitligt SSD fra de enkelte clusterpunkter til cluster-centeret, som tidligere er udtrykt i ligning (4.3). Desuden aggregeres til et gennemsnit over alle clusters. Dette metric er valgt, da det baserer sig på den euklidiske distance som begge modeller også er trænet på baggrund af. Den resulterende elbow-kurve for antal clusters i intervallet, $k = [2, ..., 30]$ er opstillet i **figur 4.8**.

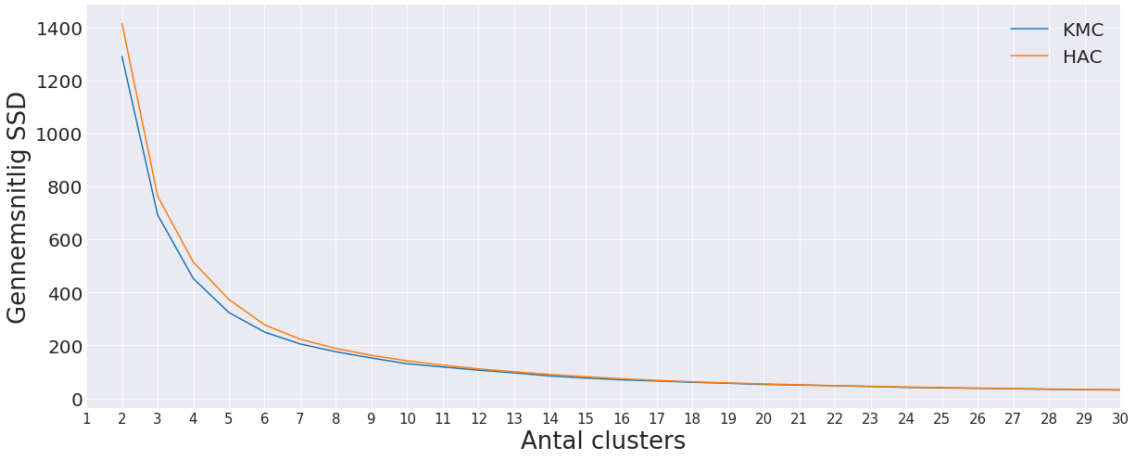
Af figuren ses det, at den gennemsnitlige SSD til cluster-centrum, falder støt indtil omkring fire clusters for både HAC og KMC. Herefter vurderes albuen at forløbe indtil omkring 10 clusters, hvorefter kurven synes at konvergere. Det vurderes herudfra, at et interval mellem fire og 10 er mulige kandidater til antallet af clusters. Herefter vil en forøgelse i antallet af



Figur 4.8. Udvikling af den gennemsnitlige Sum of Squared Deviations (SSD) som funktion af et stigende antal cluster for Hierarkisk Agglomorat Clustering (HAC) og k-means clustering (KMC). Denne type evaluering betegnes *elbow*-metode.

clusters ikke bidrage med et betydeligt fald i distancemålet.

Elbow-kurven med det trininddelte data, opstillet i **figur 4.8**, viser samme tendens. Her ses dog en højere gennemsnitlig SSD og senere konvergens med en albue, ved omkring fem til 11 clusters for begge modeller. Den højere gennemsnitlige SSD er forventelig, da inddelingen i trin vil resultere i, at nogle punkter ligger længere fra centrum og dermed får højere vægt grundet kvadreringen i udregningen af den gennemsnitlige SSD.



Figur 4.9. Udvikling af den gennemsnitlige Sum of Squared Deviations (SSD) som funktion af et stigende antal cluster for Hierarkisk Agglomorat Clustering (HAC) og K-means Clustering (KMC) på trinindelt data. Denne type evaluering betegnes *elbow*-metoden.

Silhouette-score:
For at understøtte den subjektive indikation på antal k udledt ved elbow-metoden, evalueres også silhouette-scoren. Denne sammenholder inter-cluster seperation med intra-cluster kompakthed. Dermed er denne score et udtryk for, hvor godt hvert datapunkt passer i deres respektive cluster, og samtidig hvor adskilt datapunktet er fra nærmeste cluster. Scoren beregnes for hvert datapunkt, i, ud fra ligning (4.4),

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{4.4}$$

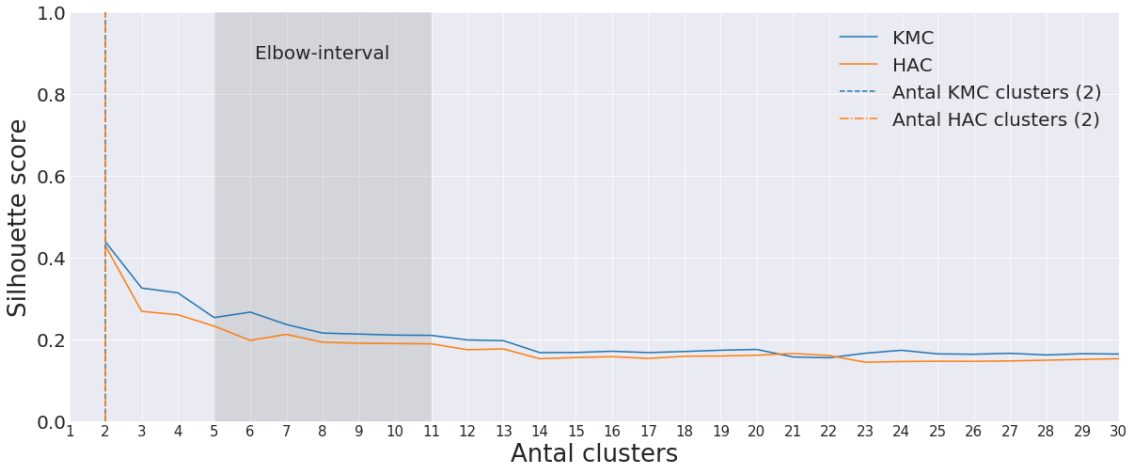
hvor $a(i)$ er den gennemsnitlige dissimilaritet mellem punktet og andre i samme cluster, og $b(i)$ er den gennemsnitlige dissimilaritet til datapunkterne i den nærmeste cluster, som punktet ikke tilhører. [35]. Dette resulterer i en værdi på en skala mellem -1 og +1, hvor høje værdier er udtryk for høj clustering-performance. En gennemsnitlig silhouette-score på over 0.5 indikerer et favorabelt cluster-resultat i forhold til intra-cluster similaritet og inter-cluster dissimilaritet. En score under 0.2 indikerer derimod fravær af cluster-strukturer med det pågældende antal clusters [29, Kap. 5]. Den gennemsnitlige silhouette-score over alle datapunkter med $k = [2, ..., 30]$ for de to implementerede modeller er vist på det originale data i **figur 4.10**.

Figur 4.10. Udvikling af silhouette-scoren ved et tiltagende antal clusters i Hierarkisk Agglomorat Clustering (HAC) og K-means Clustering (KMC). Antallet af clusters, der resulterer i den maksimale silhouette-score er markeret med en vertikal linje, da dette indikerer optimal clustering-performance.

Den resulterende silhouette-score indikerer ikke en tydelig fordel i at øge antallet af clusters fra $k = 2$, der giver den højeste score for begge modeller, $s_{HAC} = 0.660$ og $s_{KMC} = 0.668$. Der er dog indikation på tilstedeværelse af cluster-strukturer, i kraft af en silhouette-score over 0.2 med alle testede antal clusters for både HAC og KMC. I elbow-intervallet kan der for HAC ses en stagnering i en silhouette-score på omkring $s_{HAC} = 0.5$ fra $k = 4$ til $k = 8$. Da en score på 0.5 er et udtryk for rentabel clustering, kan dette interval anses for mulige kandidater til antallet af clusters. For KMC ses ikke samme indikation på et specifikt antal clusters, indenfor intervallet fundet ved elbow-metoden.

For det trininddelte datasæt, er den gennemsnitlige silhouette-score generelt lavere. Her opnås den højeste silhouette-score ved $k = 2$ med $s_{HAC} = 0.429$ og $s_{KMC} = 0.439$. Dette fremgår af **figur 4.11**. Stagneringen er i det trinopdelte data ikke evident i elbow-intervallet, men begge forløb viser nogle interessante udsving heri. Særligt ved $k = 6$ for KMC og $k = 7$ for HAC, hvor scoren stiger.

Davies-Bouldin indeks:
Afslutningsvis undersøges de implementerede modeller med Davies-Bouldin indekset, da



Figur 4.11. Udvikling af silhouette-scoren, på trininddelt data, ved et tiltagende antal clusters i Hierarkisk Agglomorat Clustering (HAC) og K-means Clustering (KMC). Antallet af clusters, der resulterer i den maksimale silhouette-score er markeret med en vertikal linje, da dette indikerer optimal clustering-performance.

dette mål, ligesom silhouette-scoren, minimerer subjektivitet i bestemmelse af k. Hvor silhouette-scoren fokuserer på alle datapunkters tilhørsforhold til de respektive clusters, baseres Davies-Bouldin indekset på minimering af ratioen mellem intra- og inter-cluster similaritet på de værste scenarier. Denne ratio, R, beskrevet for to clusters, i og j, er defineret ved formel (4.5).

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{4.5}$$

hvor S er intra-cluster similariten og M er inter-cluster dissimilariteten, det vil sige separationen mellem de to clusters. Disse parametre beregnes ved et distancemål som typisk er euklidisk distance. Dermed er intra-cluster similariteten givet ved distancerne mellem cluster-centrum og hvert datapunkt i den pågældende cluster. Tilsvarende er inter-cluster similariteten givet ved distancen til cluster-centrum i hver af de andre clusters.

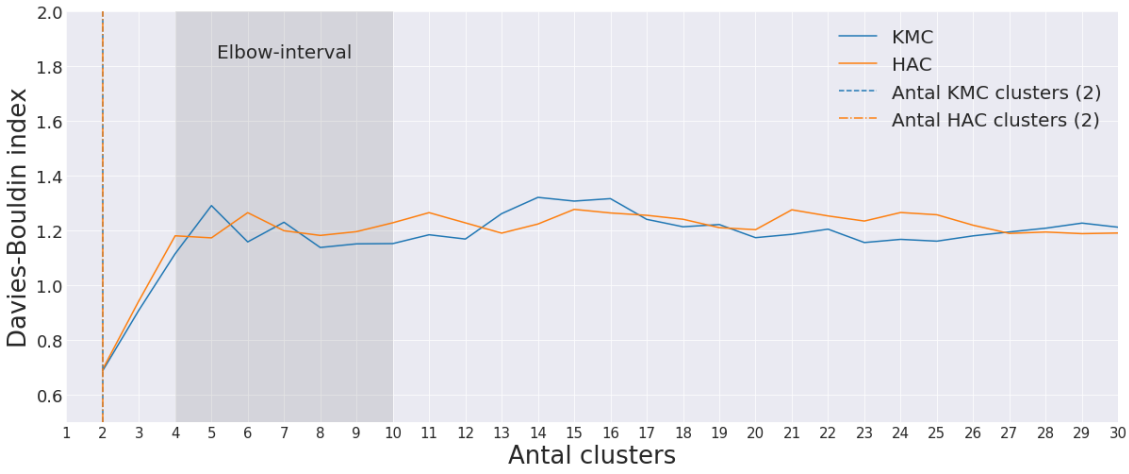
Davies-Bouldin indekset tager udgangspunkt i det værste cluster-scenario for de evaluerede clusters. Dette er defineret som det datapunkt, hvor separationen til resterende clusters er lavest (lav M_{ij}) og afstanden til eget cluster-centrum er højest (høj $S_i + S_j$). Dermed vil ratioen, R_{ij}, være høj, hvilket kan indikere at den udførte clustering er uhensigtsmæssig. Dette værste scenario, D, er udtrykt ved formel (4.6).

$$D_i = \max_{j \neq i} R_{ij} \tag{4.6}$$

Davies-Bouldin indekset er givet ved gennemsnittet af de værste scenarier for k clusters, som udtrykt ved ligning (4.7). [36]

$$DB = \frac{1}{k} \sum_{i=1}^{k} D_i \tag{4.7}$$

De to modellers Davies-Bouldin indeks med originalt data, for $k = [2, ..., 30]$, kan ses på **figur 4.12**.



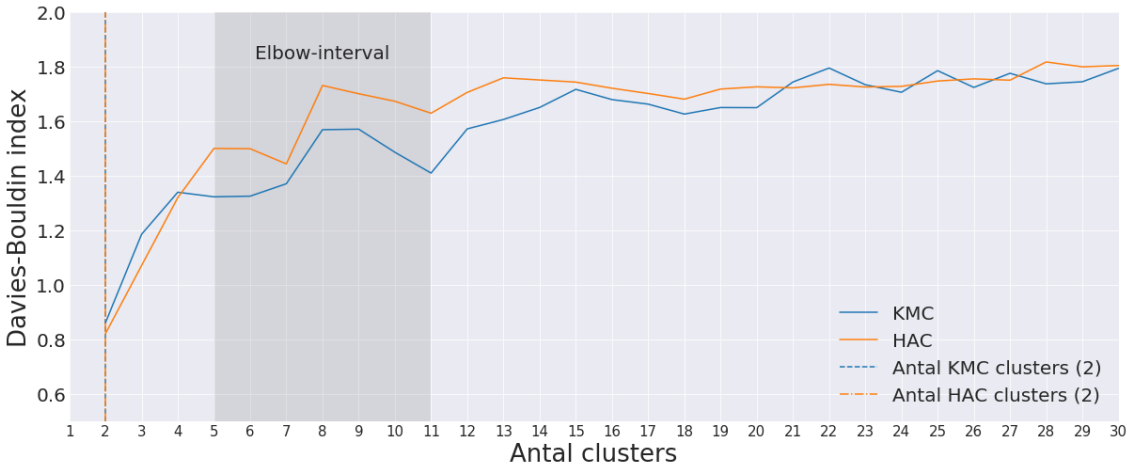
Figur 4.12. Udvikling af Davies-Bouldin indekset ved et stigende antal clusters i Hierarkisk Agglomorat Clustering (HAC) og K-means Clustering (KMC). Antallet af clusters, der resulterer i det minimale Davies-Bouldin indeks er markeret med en vertikal linje, da dette indikerer optimal clustering-performance.

Davies-Bouldin indekset indikerer, ligesom silhouette-scoren, ikke en fordel i at splitte datasættet op i flere end to clusters, med $DB_{HAC} = 0.694$ og $DB_{KMC} = 0.688$ ved $k = 2$. Desuden observeres ikke en betydelig forskel i performance mellem de to modeller. Indenfor elbow-intervallet ses et forholdsvist stabilt niveau, dog med et par fald i Davies-Bouldin indekset ved $k = 6$ og $k = 8$ for KMC, der kunne indikere bedre clustering-betingelser herfor. For HAC ses der, ud over en mindre stigning i indeksværdien ved $k = 6$, ingen indikationer for et optimalt antal clusters indenfor elbow-intervallet.

Det ses for det trinindelte data på **figur 4.13**, at indekset generelt er højere. Den laveste værdi for Davies-Bouldin indekset opnås ved $k = 2$ med $DB_{HAC} = 0.822$ og $DB_{KMC} = 0.862$. I elbow-intervallet, forekommer der for KMC en mindre stagnering indtil $k = 6$, hvor kurven igen stiger. Dermed indikeres det, at et antal clusters i dette interval har bedre clustering-betingelser, end de højere værdier ved $k > 6$. For HAC observeres et fald i Davies-Bouldin indekset ved $k = 7$, hvilket dermed er det optimale antal clusters indenfor elbow-initervallet.

Resulterende antal clusters

Gennem evaluering af de beskrevne cluster metrics er der ikke fundet entydig evidens for det optimale antal af clusters. Silhouette-scoren og Davies-Bouldin indekset peger begge på $k = 2$ som det optimale antal clusters for begge de implementerede modeller. Dette gælder både for det originale og det trininddelte datasæt. Elbow-metoden indikerede derimod, et antal mellem cirka fire og 10 clusters for begge modeller med det originale datasæt. Indefor dette interval peger silhoutte-scoren og Davies-Bouldin indekset, med det originale data, på op til syv clusters, hvoraf Davies-Bouldin indekset for KMC indikerer at seks clusters er det optimale antal. For det trininddelte datasæt vurderedes elbow-intervallet at være mellem cirka fem og 11. Heri peger både silhouette-scoren og Davies-Bouldin indekset på syv clusters for HAC. For KMC indikerer Davies-Bouldin indekset fem eller



Figur 4.13. Udvikling af Davies-Bouldin indekset, på trininddelt data, ved et stigende antal clusters i Hierarkisk Agglomorat Clustering (HAC) og K-means Clustering (KMC). Antallet af clusters, der resulterer i det minimale Davies-Bouldin indeks er markeret med en vertikal linje, da dette indikerer optimal clustering-performance.

seks clusters, hvortil silhouette-scoren indikerer seks clusters. Da de undersøgte cluster metrics ikke giver en tydelig indikation på ét specifikt antal clusters, baseres valget af k på de kvalitative vurderinger af udsving i tendenserne for de anvendte cluster metrics. I denne forbindelse betragtes intervallet hvori albuen vurderes tilstede sammenholdt med et antal k, der vurderes realistisk i relation til projektets applikation. Disse betingelser opfyldes ved seks clusters for begge KMC-modeller og HAC-modellen på det originale data og syv clusters med den trininddelte HAC, hvorfor disse værdier for k anvendes i det videre system.

4.4 Klassifikationsmodeller til prædiktion af plejebehovsfænotype

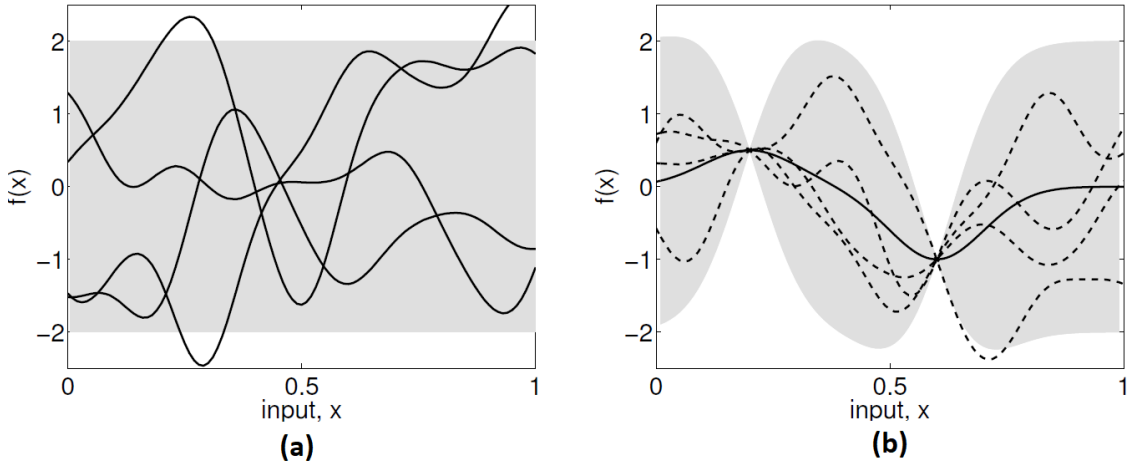
For at muliggøre prædiktion af en borgers fremtidige plejebehov designes og implementeres to typer klassifikationsmodeller, en Random Forest (RF) og et Neural Network (NN), til sammenligning af forskellige modelkompleksiteters udnyttelse af dataen. Formålet hermed er, at klassificere borgeren i en fænotype ud fra første visitation. Herefter kan typeforløbet for den pågældende fænotype anvendes som approksimation for den fremtidige udvikling i hjemmeplejebehov for borgeren.

4.4.1 Bayesian Optimization

I forbindelse med design af klassifikationsmodeller som RF og NN, skal en række hyperparametre bestemmes. For at identificere den kombination af hyperparametre der resulterer i maksimal model-performance kan en optimeringsalgoritme anvendes. En sådan metode udgør et alternativ til en udtømmende undersøgelse af alle kombinationer for de mulige hyperparameterværdier, kaldet *grid search*.

I nærværende studie optimeres klassifikationsmodellernes hyperparametre gennem *Bayesian Optimization*, der er en statistisk optimeringsmetode som forsøger at maksimere

en funktion med så få iterationer af hyperparameterkombinationer som muligt. Her antages det, at kriteriefunktionen der skal optimeres kan samples fra en *Gaussian process*. Sidstnævnte er en posterior sandsynlighedsfordeling af mulige funktioner, der kan beskrive kriterifunktionen. Efterhånden som antallet af reelle observationer vokser, gennem test af et sæt hyperparametre, forbedres den posteriore sandsynlighedsfordeling. Dette er illustreret på **figur 4.14**, hvor **(a)** er den initierende posteriore sandsynlighedsfordeling og **(b)** viser hvordan denne ændres efter to punkter på kurven er evalueret. Herved bliver algoritmen mere sikker på hvilken region af *parameter space*, der skal afsøges for at finde den optimale kombination af hyperparametre. Dermed anvendes alle forrige evalueringer til at prædiktere det næste gæt på et sæt af optimale hyperparametre [37].

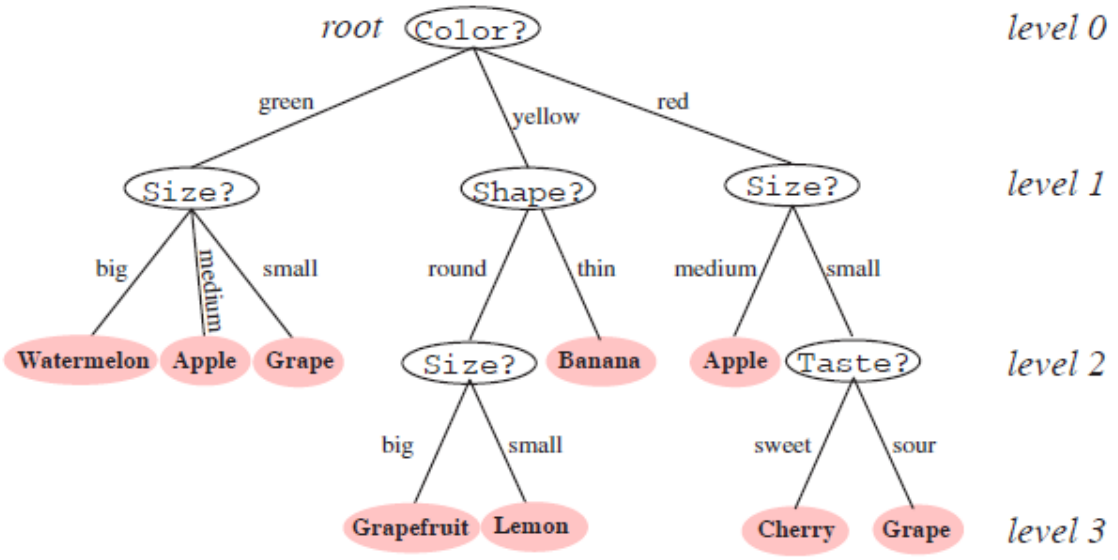


Figur 4.14. En *Gaussian process* for funktionen $f(x)$ med (a) forudgående distribution af sandsynlige funktioner og (b) ændringen i den posteriore sandsynlighedsfordeling efter to datapunkter er evalueret. [38]

Den implementerede Bayesian hyperparameteroptimering tog udgangspunkt i et GitHub-repo af Fernando M. F. Nogueira [39]. Her vælges et antal gange algoritmen skal foretage en initierende tilfældig evaluering af et sæt hyperparametre. Disse anvendes som udgangspunkt for at opstille den posteriore sandsynlighedsfordeling. Ud over disse tilfældige evalueringer kan der opsættes specifikke hyperparameterkombinationer kaldet probes, der ønskes evalueret som en del af opstilling af den posteriore sandsynlighedsfordeling. Desuden specificeres antallet af iterationer til evaluering af sandsynlighedsbegrundede hyperparameterkombinationer. De specifikke hyperparametre der skal optimeres er beskrevet for hver af modellerne i de følgende afsnit og resultaterne herfor præsenteres i kapitel 5

4.4.2 Random Forest

En RF-klassifikationsmodel er opbygget som en samling af individuelt trænede klassifikationsmodeller kaldet *Decision trees*. Hvert decision tree giver et klassifikations-output i forhold til, hvilken klasse inputtet tilhører, hvoraf flertallet bestemmer den endelige klassifikation i RF-modellen. Decision trees er en type non-lineære modeller, baseret på en serie af regelbaserede betingelser omkring inputtet, hvor den næste betingelse, der opstilles er baseret på udfaldet på den gældende. Denne sekvens af betingelser opstilles som et retningsbestemt træ af beslutninger, eller *decisions*, som illustreret i **figur 4.15**.



Figur 4.15. Eksempel på et *decision tree* til klassificering ud fra en række beslutninger som evalueres på baggrund inputtets egenskaber [30]

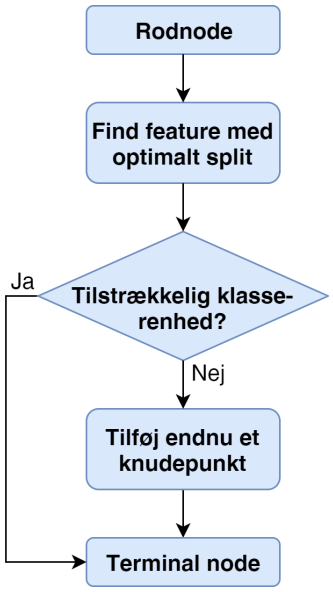
Det øverste knudepunkt, kaldet rodnoden, er forbundet med et hierarki af efterfølgende knudepunkter gennem en række forgreninger. For hvert knudepunkt afgør en betingelse omkring en af inputtets egenskaber hvilken forgrening, der skal følges videre til næste niveau i træet. For at øge den beregningsmæssige effektivitet anvendes typisk binære beslutninger. Betingelsesudfaldene i et knudepunkt skal derfor være entydige og udtømmende, så et bestemt input altid vil følge samme vej gennem træet, og altid vil nå til de terminale knudepunkter, der markerer hvilken klasse, det pågældende input tilhører. For et ideelt decision tree indeholder hvert terminalt knudepunkt en undergruppe af observationer med samme klasselabel. Klasserenheden i et knudepunkt kan beskrives af et mål kaldet, *Gini impurity*. Dette mål beskriver sandsynligheden for, at en tilfældigt observation ville klassificeres forkert, hvis klassificeret ud fra distributionen af observationer i det pågældende knudepunkt. Gini impurity er givet ved formel (4.8).

$$I_G(n) = 1 - \sum_{i=1}^{J} (p_i)^2 \tag{4.8}$$

Hvor n betegner et knudepunkt, J er alle klasser og p_i repræsenterer brøkdelen af klasseobservationer ud af det samlede antal observationer i knudepunktet. Gini impurity anvendes ofte under træning af decision trees som mål for, hvilket datasplit i et knudepunkt, der vil resultere i mest heterogene grupper.

For at træne et decision tree til at opnå bedst mulig klassifikation, og dermed mest mulig klasserenhed i de terminale knudepunkter anvendes en rekursiv, såkaldt *greedy* optimering. Her evalueres alle features med henblik på at splitte træningsdata op, så datasplittet resulterer i mere rene samplegrupperinger. Derfor identificeres hvilken feature, og beslutning herom, der minimerer heterogeniteten i de resulterende grupper mest muligt. Da hver undergruppe ofte vil udgøres af en blanding af klasselabels, bestemmes det ud fra

et stop-kriterie, om undergruppen er acceptabel, eller om der skal foretages endnu et split ud fra en anden feature. Denne proces kaldes Classification and Regression Tree (CART) og er illustreret på **figur 4.16**.



Figur 4.16. Rekursiv, *greedy* optimeringsprocess, ved navn CART, anvendt til træning af et decision tree. Modificeret fra [30].

Da inputtets vej til klassifikation styres af en række afgørelser baseret på inputtets features, kan decision trees håndtere nominel og ordinal data. Dette er datatyper, der beskriver inputtet med kategorier frem for numeriske værdier. Desuden er decision trees forholdsvis transparente og selvforklarende i forhold til den klassificering der foretages, da klassifikationsprocessen kan summeres op til en sekvens af logiske regler.[30].

En RF-klassifier er som beskrevet en samling af et antal individuelle descision trees, som anvendes som én enhed. Klassificeringen foregår ved at inputtet propageres gennem alle decision trees i RF-modellen, hvorefter klassen bestemmes ud fra flertallet. Fordelen herved er, at flertallet udligner individuelle træers fejl. Forudsætningerne for at en RF-model kan fungere er, at der er et mønster i data, som gør modellerne i stand til at give mere kvalificerede klassificeringer end ved tilfældige gæt. Ydermere skal der være lav korrelation mellem klassifikationerne fra de individuelle træer. For at sikre opfyldelse af sidstnævnte kriterie, anvendes en metode kaldet *Bootstrap Aggregation*. Dette indebærer, at der til træning af hvert individuelt træ udvælges tilfældige observationer fra datasættet, så en delmængde udelades. Desuden dublikeres en andel af observationerne, så der oversamples til samme størrelse datasæt som det originale. Herved sikres det, at der trænes forskellige træer. Ydermere fremtvinges variation mellem træerne i RF-modellen ved at begrænse feature-sættet, når de individuelle træer skal foretage et split. Som beskrevet overvejes alle features i beregning af det optimale split for et traditionelt decision tree. For RF-træer estimeres det optimale split i stedet ud fra en tilfældig delmængde af features i det originale feature-sæt, hvilket bidrager til at opnå ukorrelerede træer. Endvidere nedbringes risikoen for over-fitting, forsaget af greedy optimering med hele feature-sættet [40].

Der er forskellige måder at implementere en multiklasse RF på, men et studie af Adnan og Islam [41] peger på en One-vs-Rest-arkitektur som den mest lovende i en række

klassificeringsproblemer, hvorfor denne arkitektur er implementeret i det nærværende studie. Denne metode tager udgangspunkt i, at lave et binært replika af target-værdierne for hver klasse med hver deres individuelle dertilhørende decision trees.

I implementeringen af RF-modellen, er der en række hyperparametre, der skal bestemmes. Disse omfatter følgende:

- Antallet af individuelle decision trees.
- Antallet af features der skal i betragtning når et nyt split skal identificeres.
- Den maksimale dybde på et decision tree.
- Det maksimale antal terminale noder.
- Det mindste antal observationer der kræves i en node, for at denne kan splittes.
- Det maksimale antal features der skal evalueres for at opstille det optimale split.
- Kriteriefunktionen til kvantificering af klasserenhed.

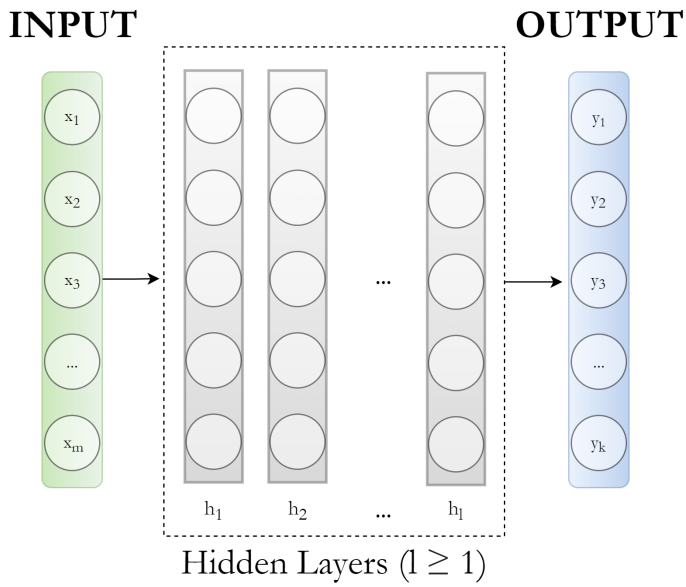
Hensigten med indstilling af hyperparametrene er at designe en model med bedst mulige klassificeringer, uden at modellen blot har memoreret træningsdataen, og dermed ikke er generaliserbar til ny data. Til at finde den kombination af hyperparametre, der resulterede i bedst modelperformance, anvendtes optimeringsalgoritmen, Bayesian optimization, beskrevet i afsnit 4.4.1. Det er vigtigt at overveje, at optimering i forhold til træningsdataen kan resultere i en over-fittet model. For at nedbringe denne risiko anvendes *k-fold cross-validation*, hvor træningsdataen inddeles i k undergrupper. Modellen trænes iterativt med alle undergrupper undtagen én, som i stedet anvendes til validering. Dette fortsættes, indtil alle undergrupper har været anvendt som valideringsdata. I hyperparameteroptimeringen testes en ny kombination af hyperparameterværdier for hvert sæt k-fold cross-validations. Herved kan modelperformance ved en række kombinationer sammenlignes og den bedste udvælges. I træning af den implementerede model anvendtes 5-fold cross-validation med evaluering på gennemsnitlig f1-score på klasseniveau som mål for modelperformance. f1-scoren anvendes, da denne håndterer eventuel ubalanceret fordeling af observationer i klasserne, hvilket er forventeligt i dette studie grundet den høje koncentration af lave værdier beskrevet i præ-clustering analysen i afsnit 4.3.1. f1-scoren er beskrevet nærmere i afsnit 4.6.2.

4.4.3 Neuralt netværk

Til klassifikationen af borgernes første visitation, er der ydermere implementeret et Neural Network (NN). Grundlæggende er et NN en lag-inddelt samling af parametriserede knudepunkter, hvorigennem en non-lineær mapning fra en input-featurevektor til det ønskede klassifikations-output kan læres. Gennem denne potentielt non-lineær modellering, tillader denne type model en øget kompleksitet, der i mange applikationer øger performance. Arkitekturen af et NN udgøres af tre forskellige lagtyper: input-, hidden og output-lag. Hvor inputtet tager form efter den featurevektor der anvendes som repræsentation for input-observationen. Dernæst følger et antal hidden-lag der indeholder et antal kunstige neuroner. Output-laget indeholder et antal neuroner svarende til antallet af klasser. Output-signalet fra neuronerne i dette lag er en diskriminantfunktion, hvoraf den med højest værdi afgør output-klassen [30].

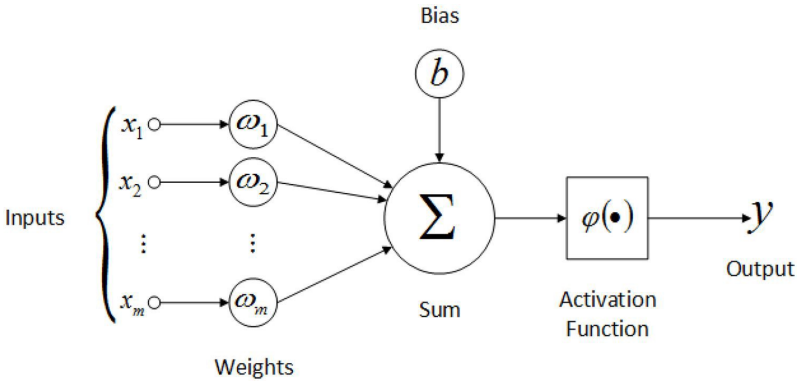
Den mest basale feedforward arkitektur, kaldet Multi-layer Perceptron (MLP), er valgt som udgangspunkt i dette studie for, at bevare tolkbarhed af de resulterende klassificeringer og

da de anvendte input features ikke indeholder signaldata. Som navnet *multilayer* antyder, er der mindst tre lag i netværket. *Perceptron* beskriver, at lagene består af en række kunstige neuroner, der også kaldes perceptroner. Et MLP-netværk har foruden et input- og output-lag, altid mindst ét hidden-lag. Et eksempel på en MLP-arkitektur er illustreret på **figur 4.17**.



Figur 4.17. Den basale opbygning af et Multilayer Perceptron (MLP) netværk, med et input-lag bestående af m features, mindst ét hidden-lag og et output-lag med k klasser.

Grundstenen i et NN, er de kunstige neuroner, der processerer flere inputs og producerer ét output. Disse er forbundet lagene imellem. I et såkaldt "fully connected" NN, er alle neuroner i hvert lag forbundet med hinanden, hvorved værdien for hver enkelt neuron gives videre til alle neuroner i næste lag. Den grundlæggende opbygning af en neuron er vist i **figur 4.18**.



Figur 4.18. Den basale opbygning af en kunstig neuron. Modificeret fra [42]

Input-værdierne til en neuron vægtes med vægtparametre, ω, og summeres til en netto aktivering. Desuden justeres denne sum med et bias, b. Gennem disse vægte er det muligt, at diskriminere mellem vigtigheden af neuronens inputs, x_m, under træning af modellen. Outputtet af neuronen er en non-lineær funktion af inputtet kaldet aktiveringsfunktionen, $\phi(\cdot)$ [30, Kap. 6]. Antallet af lag og neuroner i hvert lag er en hyperparametre, der skal

bestemmes i relation til applikationen. I dette studies optimeres disse gennem bayesian optimization.

Træning af det neurale netværk med backpropagation

Ved træning af NN-modellen forståes bestemmelse af vægtparametrene i hver neuron. Denne træning foregår gennem *supervised learning*, det vil sige ud fra træningsdaten og det tilhørende ønskede output. Præcis hvordan denne træning forløber bestemmes af netværkets *optimizer*. Her er den mest udbredte og simple metode klassisk *backpropagation*.

Ved backpropagation udnyttes afvigelsen, eller *loss*, $L(y, y')$, mellem den prædikterede klasse, y', og den sande klasse, y. Dette loss propageres tilbage gennem lagene i netværket, hvor vægte og bias opdateres i henhold til retning og størrelse for gradienten af loss. Formålet hermed er at opdatere vægtene og bias i neuronerne til værdier der resulterer i en lavere loss og dermed outputs tættere på de sande klasser. Denne opdatering laves ud fra formel 4.9.

$$\Delta w_{ij}^{l} = -\alpha \frac{\partial L(y, y')}{\partial w_{ij}^{l}} \tag{4.9}$$

For at beregne størrelsen af opdateringen af en given vægt, Δw, for node j i lag h_l fra input node i, anvendes kædereglen til udledning af den negative gradient. Denne regel baseres på at prædiktionen, y', er en funktion af alle parametrene, givet at parametrene hver især er afhængige af parametrene fra det tidligere lag. Herfra er gradienten givet ved den førsteordens afledte af loss, $\frac{\partial L(y,y')}{\partial w_{ij}^{l}}$, med respekt til den aktuelle vægt w_{ij}^{l}. Herved kan loss propageres tilbage gennem hele netværket, indtil en gradient for hver parameter er beregnet og vægte og bias justeret i forhold hertil, i et forsøg på at identificere det globale minimum i loss-funktionen. [43]. Dette gentages for alle observationer i træningssættet for at opnå et netværk med optimale parametre til at repræsentere variationen i datasættet. Dette kan gøres på tværs af hele træningssættet eller i *batches* afhængigt af applikationen.

For at sikre at outliers i dataen ikke får negativ indflydelse for optimeringen af netværket, kan parameteroptimeringen reguleres gennem *learning rates* og *learning decays*. Learning rate, α, er den relative tilladte ændring i vægtene og learning decay et udtryk for iterativt at sænke α gennem træningen for at øge generaliserbarheden i systemet. [30, Kap. 6]

Oftest har gradient descent, i sig selv, en risiko for at finde suboptimale lokale minimum. Metoder, såsom Adaptive Gradient Algorithm (AdaGrad), Root Mean Squared Propagation (RMSprop) eller Adaptive Moment Estimation (Adam) forsøger at løse dette problem gennem anvendelse af en adaptiv learning rate, der tilpasses ud fra størrelse og/eller retningen af de forudgående gradienter, kaldet momentum [44]. Momentum tilføjes backpropagation-udtrykket som vist i ligning 4.10.

$$\Delta w_{ij}^{l}(t) = -\alpha \frac{\partial L(y, y')}{\partial w_{ij}^{l}(t)} + \mu \Delta w_{ij}^{l}(t-1) \tag{4.10}$$

Her tilføjes ændringen i vægten ved forrige iteration, $\Delta w_{ij}^{l}(t-1)$, justeret af momentumkoefficienten, μ. Ofte anvendes RMSprop eller Adam i mange NN, hvorfor disse vil

blive sammenholdt i optimeringen af det nærværende system, med forskellige indledende learning rates.

Loss-funktion
For dette projekt anvendes borgerens fænotype, tildelt gennem clustering, som den sande klasse. I et multiklasse-opsæt som dette, hvor der tillades ét korrekt output, men systemet har mulighed for at prædiktere alle klasserne, er de facto standarden for at beregne loss *categorical cross entropy*. Denne loss-funktion er en sammenligning af distributionen for hver klasse i output-laget med distribution af den sande klasse, i form af en one-hot-vektor. Jo mere de ligner hinanden, desto lavere bliver fejlen. Dette beregnes ved formel (4.11). [30, Kap. 6].

$$L(y, y') = -\sum_{i=1}^{k}(y_i \times \log(y'_i)) \tag{4.11}$$

Hvor $L(y, y')$ betegner loss mellem logaritmen af den prædikterede klasse, $log(y')$ og den sande klasse y, summeret for hver kategori, k. Kategorisk cross-entropy tager ikke højde for, at en skæv klassefordeling kan resultere i, at majoritetsklassen får mere indflydelse på loss. Da datasættet anvendt i dette projekt indebærer en potentiel ubalance af antallet af borgere i hver klasse, blev en f1-relateret loss implementeret frem for kategorisk cross-entropy. Et problemet med anvendelse af f1-scoren som loss-funktion er, at den ikke er differentiabel og dermed ikke direkte implementerbar i backpropagation. Dette er løst ved en modificering af f1-score, kaldet "soft f1", så den baserer sig på en kontinuerlig summering af likelihood værdierne over hvert batch. Dette gøres ved at anvende sandsynlighederne for sande positive ($y' \cdot y$), falske negative ($y' \cdot (1 - y)$) og falske positive ($(1 - y') \cdot y$) uden threshold til udregningen af f1. [45] Udregningen af f1-scoren er yderligere beskrevet i afsnit 4.6.2

Træningprotokol
I forbindelse med træningprotokollen for et NN, findes der en række parametre, der skal justeres. Som udgangspunkt bør en procentdel af dataen i træningssættet tilsidesættes til løbende validering af modellen. Denne validering kan bruges til bestemmelse af antallet af epochs eller som led i et stopkriterie på træningen. Antallet af epochs er antallet af gange som modellen træner på hele træningssættet, opdelt i et antal delmængder kaldet *batches*. Dette bør sættes så loss bliver så lav som mulig, uden at systemet overfitter til træningsdataen. Overfitting vil sige at validerings-loss stiger selvom trænings-loss falder ved et øget antal epochs. For at undgå dette kan der sættes et stopkriterie, som eksempelvis kan være at stoppe træningen hvis validerings-loss ikke mindskes. Derved stoppes træningen ved den model, der giver de mest generaliserbare resultater, og ergo den mindst overfittede model. Da det ofte er beregningsmæssigts krævende, at evaluere på det fulde træningssæt, kan der anvendes batches i træningsprotokollen. Disse består, som beskrevet, af en delmængde af træninggsættet, hvorudfra gradienten til backpropatation beregnes og vægtene dermed opdateres. Størrelsen på denne delmængde, kaldet *batch size*, bør være stor nok til at sikre, at alle klasser er repræsenteret og at variansen i sættet er tilsvarende det fulde træningssæt. [30, Kap. 6]. En del af dette er allerede sikret i nærværende studie, ved at balancere træningsdataen. Yderligere er batch size sat til

antallet af clusters multipliceret med en faktor, a, der er optimeret som hyperparameter ved implementeringen.

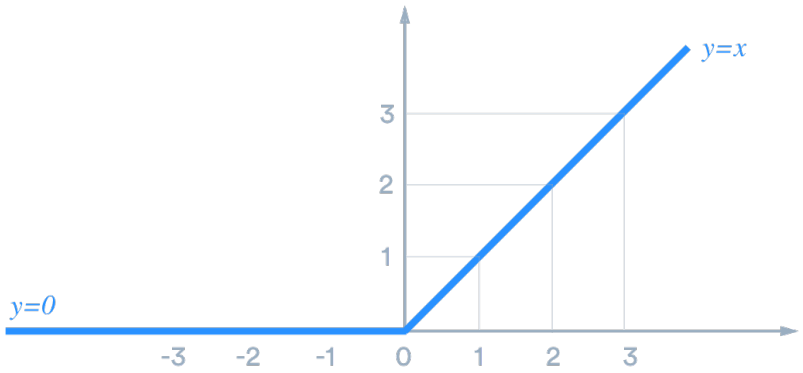
Valg af aktiveringsfunktion

Aktiveringsfunktionen bestemmer, som beskrevet, outputtet af en kunstig neuron og er dermed afgørende for om neuronens inputs betragtes som afgørende for den pågældende klassifikation. Herfor har aktiveringsfunktionen stor betydning for modellens stabilitet og hvorvidt modellens performance konvergerer under træning.

En aktiveringsfunktion skal opfylde at den er tæt på lineær for små netto aktiveringer og tiltagende non-lineær for større netto aktiveringer. Da en lav aktivering er ensbetydende med at den pågældende neuron ikke indeholder signifikant information, vil disse blive dæmpet under træning med en sådan funktion. Bagrunden herfor, er at træning af et NN typisk foregår ved 'gradient descend'-baserede metoder, der finder den minimale loss gennem differentiering af loss-kurven. Desuden skal aktiveringsfunktionen være monoton, kontinuert og differentiabel for at reducere risikoen for lokale minima under optimering.

For det implementerede NN, anvendes Rectified Linear Unit (ReLU)-funktionen, der opfylder de beskrevne kriterier for en aktiveringsfunktion og samtidig er beregningsmæssig effektiv. Denne effektivitet er et resultat af at negative input-værdier ignoreres, mens de positive input bibeholdes [46]. Dette er udtrykt ved ligning 4.12 og illustreret på **figur 4.19**.

$$f(x) = \max(0, x) \tag{4.12}$$



Figur 4.19. Aktiveringsfunktionen Rectified Linear Unit (ReLU). Modificeret fra [46].

ReLU-funktionen har dermed en uendelig øvre grænse, hvilket er beregningsmæssigt belastende. For at imødekomme denne problematik, er *batch*-normalisering et hyppigt anvendt værktøj. Denne tilføjelse normaliserer inputtet i et lag, hvilket desuden mindsker netværkets træningstid. [47]. ReLU-funktionen tager ikke højde for en problematik kaldet "forsvindende gradienter", hvor gradienten bliver forsvindende lille og vægtene deraf ikke bliver opdateret. Dette er et generelt problem i mange aktiveringsfunktioner, men ReLU har til trods vundet popularitet, da dens performance matcher lignende funktioner, såsom sigmoid, men er mindre beregningsmæssig belastende for systemet [48].

I multiklassetilfælde, hvilket klassificeringen af plejebehovsfænotyper er, bør aktiveringsfunktionen i sidste lag være en *softmax*-funktion. For denne aktiveringsfunktion er outputtet skaleret mellem nul og et, hvorved det kan behandles som en sandsynlighed for hver klasse, k, i forhold til hvilken klasse input-observationen tilhører. Den samlede sum af sandsynligheder for klasserne er 1. Denne sandsynlighed er udregnet fra det logistiske output ved *softmax*-funktion i ligning 4.13.

$$f(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{k} e^{y_j}} \tag{4.13}$$

Hvor e^{y_i} er den naturlige eksponentialfunktion til det givne input, y_i, som normaliseres ved division med summen af e til de resterende inputs, e^{y_j} i det pågældende lag. Dermed er den klasse, den respektive input-observation med størst sandsynlighed tilhører, givet ved output-neuronen med højeste output. Desuden er værdien for denne output-neuron et udtryk for konfidensen af prædiktionen. [49].

Dropout

Aktiveringsfunktionen i et NN kan desuden reguleres gennem *drop out*, der er en rate af tilladte aktive noder i et lag ved hvert gennemløb i træningen. Formålet med at regulere aktiveringerne er, at øge systemets generaliserbarhed og mindske overfitting ved at tvinge systemet til at lære fleksibilitet udenom de deaktiverede noder. [50]. En række drop outrater er undersøgt i implementeringen.

4.5 Evaluering af cluster-udledte fænotyper

For at undersøge validiteten af de clusters, der udledes fra de implementerede clusteringalgoritmer, beskrevet i afsnit 4.3, evalueres de indledningsvis i forhold til clusteringperformance. I denne forbindelse beregnes først Interquartile Range (IQR), der er et nonparametrisk mål for spredning i forhold til medianen. Dette mål beregnes som afstanden mellem 25. og 75. percentil af dataen. IQR beskriver dermed hvor tæt koncentreret de midterste 50 % af dataen er [51, Kap. 4]. Desuden opstilles de udledte clusters i et silhouette-plot, der visualiserer hver cluster repræsenteret ved en såkaldt silhuet. Denne er baseret på silhouette-scoren, beskrevet i afsnit 4.3.4, og viser antallet af observationer, der ligger tydeligt inden for den pågældende cluster og antallet der ligger mellem clusteres og dermed har et mere usikkert tilhørsforhold til clusteren. Dermed viser vidden på hver cluster i plottet den relative kvalitet af clusteren, mens højden viser andelen af observationer i hver cluster ud af det samlede antal [35]. Denne visualisering giver desuden mulighed for sammenligning af cluster-performance mellem de to implementerede modeller, KMC og HAC, samt de resulterende clusters for det originale datasæt og det trininddelte datasæt. Endvidere sammenlignes de udledte clusters i et scatter-plot, hvor observationerne visualiseres i et reduceret feature space, beregnet gennem en kombination af PCA og t-SNE. Disse metoder er tidligere beskrevet i afsnit 4.3.1. I dette plot tildeles hver observation en farve i relation til det cluster-tilhørsforhold de er tildelt. Herved muliggøres undersøgelse af seperation mellem clusters samt forskelle i hvordan de to modeller har placeret decision boundaries til adskillelse af clusteres.

Endvidere undersøges de cluster-udledte typeforløb i relation til deres repræsentation af borgerens faktiske forløb. I denne forbindelse udregnes RMSD, givet ved formel (4.14), mellem hvert datapunkt i en borgers faktiske forløb og typeforløber fra borgerens tildelte cluster.

$$RMSD = \sqrt{\frac{\sum_t^T (y'_t - y_t)^2}{T}} \tag{4.14}$$

Hvor y er borgerens faktiske forløb, y' er det tildelte typeforløb, t er et enkelt datapunkt i forløbet og T det samlede antal datapunkter i borgerens forløb. RMSD er valgt i denne evaluering, da applicering af kvadratrod udligner kvadreringen på afvigelserne og derved øger tolkbarheden i relation til faktisk distance.

For at evaluere brugbarheden af de udledte clusters som plejebehovsfænotyper, analyseres borgersammensætningen i hver klasse. Denne undersøgelse underbygges med statiske sammenligninger af featureværdierne mellem clusters. I denne forbindelse foretages for hver feature, enten en ensidet variansanalyse (ANOVA) eller den ikke-parametriske ækvivalent, Kruskall-Walis, afhængigt af om featureværdierne er normalfordelte. Normalfordelingen testes gennem en Shapiro-Wilk test. Såfremt det findes i variansanalysen, at der er statistisk signifikant forskel i en feature mellem nogle af clusters, undersøges det ved en post hoc test, hvilke clusters der giver anledning til denne forskel. Tukeys test anvendes post hoc, da denne kan håndtere grupper af forskellig størrelse samt korrigerer for den øgede sandsynlighed, der er for at opnå et falsk statistisk signifikant resultat på grund af udførelsen af flere sammenligninger.

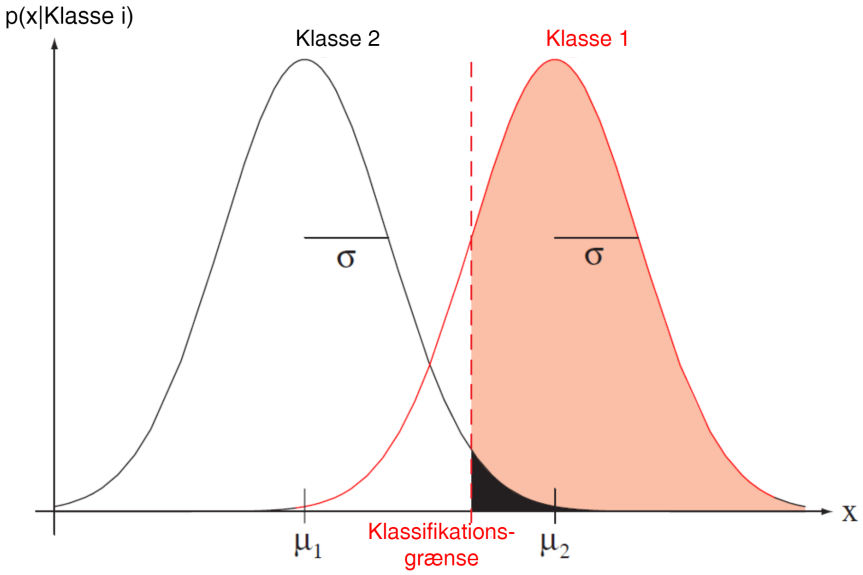
4.6 Evaluering af klassifikationsmodeller

I forbindelse med validering af machine learning algoritmer, er litteraturen skiftet væk fra hovedsageligt at evaluere på baggrund af nøjagtigheden eller *accuracy*, der udelukkende beskriver andelen af rigtigt klassificerede observationer ud af det totale datasæt. I stedet anbefales det at anvende mere informative mål, som ROC- og PR-kurver [52]. Disse metoder er beskrevet i de følgende afsnit.

4.6.1 Receiver operating characteristic

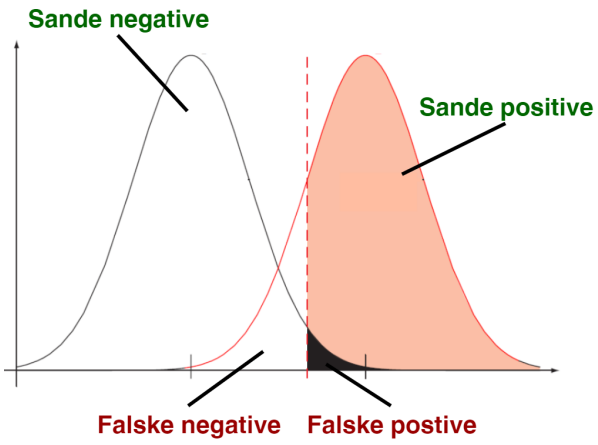
Til evaluering af klassifikationsperformance anvendes ROC-kurven. Dette er et mål for, hvor godt en supervised klassifikationsmodel kan diskriminere mellem klasserne ved brug af forskellige klassifikationsgrænser. Denne metode tager udgangspunkt i et binært klassifikationsproblem, hvorved modellens opgave er, at opstille en grænse, der bedst muligt adskiller sandsynlighedsfordelingerne for de to klasser. Denne situation er illustreret på **figur 4.20**, hvor to klassers sandsynlighedsfordelinger er plottet som en funktion af observationen, x. Modellen klassificerer en ny observation ud fra, om den ligger over eller under den identificerede grænse [30].

For en binær klassification med en positiv og negativ klasse kan udfaldet af klassifikationen inddeles i fire kategorier. *Sand positiv* betegner, at en observation klassificeret som positiv, korrekt tilhøre denne klasse. *Sand negativ* beskriver, at en observation klassificeret



Figur 4.20. Sandsynlighedsfordelinger, p(x|Klasse i), for to klasser i et binært klassifikationsproblem, hvor x er en observation i datasættet. Gennem en klassifikationsmodel defineres en klassifikationsgrænse (*decision boundary*), der søger at adskille sandsynlighedsfordelingerne mest muligt. Modificeret fra [30].

som negativ, korrekt er klassificeret til ikke at tilhøre den positive klasse. Begge disse grupper er dermed korrekte klassifikationer. De to resterende kategorier beskriver forkerte klassifikationer, hvor *falsk positiv* indebærer, at en observation klassificeret som positiv, reelt tilhører negativ. *Falsk negativ* betegner, at en observation der faktisk tilhører den positive klasse, klassificeres som værende negativ. **Figur 4.22** illustrerer hvordan kategorierne manifesterer sig i de to sandsynlighedsfordelinger for et binært klassifikationsproblem. Desuden kan kategorierne for udfaldet opstilles i en tabel kaldet en *confusion matrix*, som vist i **figur 4.21**.



Figur 4.21. Kategorierne for udfaldet af en binær klassifikation illustrerer for sandsynlighedsfordelingerne for de to klasser ved en tilfældig klassifikationsgrænse. Modificeret fra [30].

		Den sande klasse	
		Klasse 1	Klasse 2
Den prædikterede klasse	Klasse 1	Sand positiv	Falsk positiv
	Klasse 2	Falsk negativ	Sand negativ

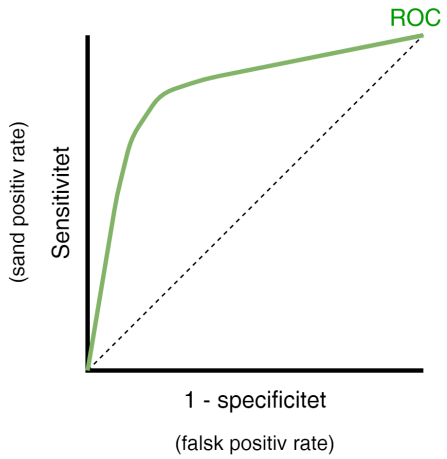
Figur 4.22. Confusion matrix for binært klassifikationsproblem. Modificeret fra [30].

Ud fra confusion matrix for en klassifikationsmodel kan sensitivitet og specificitet udregnes. Hvor sensitiviteten, også kaldet sand positiv rate, beskriver en models evne til korrekt at klassificere de observationer, der tilhører den positive klasse. Udregningen heraf er opstillet i formel (4.15). Omvendt repræsenterer specificiteten modellens evne til at klassificere de observationer, der tilhører den negative klasse. Udtrykket herfor er givet i formel (4.16).

$$\text{Sensitivitet} = \frac{\text{sande positive}}{\text{sande positive} + \text{falske negative}} \tag{4.15}$$

$$\text{Specificitet} = \frac{\text{sande negative}}{\text{sande negative} + \text{falske positive}} \tag{4.16}$$

ROC-kurven er den resulterende kurve, når sensitiviteten plottes mod (1 - specificiteten), ved en række forskellige klassifikationsgrænseværdier. Hermed er det muligt at visualisere alle mulige klassifikations-grænseværdier på én gang. Et eksempel på en ROC-kurve er vist på **figur 4.23**. Fra ROC-kurven beregnes Area Under the Curve (AUC), der er arealet under ROC-kurven. Hvis klassifikationsmodellen opnår høj separation mellem klasserne vil AUC være tæt på 1, mens en dårlig separation af klasserne vil resultere i en AUC på omkring 0.5, hvilket udtrykker at klassifikationen er tilfældig. ROC-kurven og AUC er et udbredt og nyttigt mål, der dog risikere at negligere performance af minioritetsklasser ved svært ubalancerede klasser. Dette scenario kan bedre evalueres gennem PR der beskrives i afsnit 4.6.2 [52].



Figur 4.23. Eksempel på en Receiver operating characteristic kurve. Modificeret fra [30].

Som beskrevet er ROC-kurven tilegnet binære klassifikationsproblemer. Herfor må den modificeres for at kunne appliceres i en multiklasse-situation. I dette tilfældes anvendes en *One-vs.-Rest*-tilgang, hvor en ROC-kurve opstilles for hver klasse, som ved et binært klassifikationsproblem. Her anskues alle andre klasser end den aktuelle klasse, som én samlet, komplimentær klasse og der udregnes dermed en specifik ROC-kurve for hver klasse. De klasse-specifikke ROC-kurver, og dermed AUC-værdier, kan aggregeres til et samlet mål for klassifikationsmodellens performance. Dette kan enten gøres på makro- eller mikro-niveau. Makro-aggregering foregår ved at beregne gennemsnittet af AUC-værdierne for hver klasse-specifik ROC-kurve. Ved mikro-aggregering anvendes i stedet et gennemsnit af sand-positiv- og falsk-positiv-raten til beregning af én samlet AUC-værdi. Mikro-aggregering kan

dog skjule dårlig performance af visse klasser da en eventuel majoritetsklasse vil have større vægt i denne type gennemsnittet.

4.6.2 Precision og Recall

En PR-kurve er et alternativ til ROC-kurven, som giver et mere informativt indblik i performance for en klassifikationsmodel, når det anvendte datasæt er ublanceret i forhold til klasserne. Dette skyldes at PR baseres på præcisionen frem for specificiteten, der grundet inklusionen af de sande negative påvirkes af ubalancen mellem klasserne.

På samme måde som ROC-kurven tager en PR-kurve udgangspunkt i confusion matrix for klassifikationen. Recall er givet ved sensitiviteten, mens Precision er et udtryk for hvor mange af de observationer, der klassificeres som værende en bestemt klasse, der faktisk tilhører denne klasse. Dette er beregnet ved formel (4.17). En høj værdi for både precision og recall indikerer, at klassifikationsmodellen genererer akkurate resultater samt identificerer størstedelen af de observationer, der tilhører den pågældende klasse. En PR-kurve opnås ved at plotte *recall* på x-aksen og *precision* på y-aksen.

$$\text{Precision} = \frac{\text{sande positive}}{\text{sande positive} + \text{falske positive}} \tag{4.17}$$

Desuden kan en samlet score, kaldet f1-scoren beregnes ud fra precision og recall. Denne er givet ved formel (4.18) og udgør et harmonisk gennemsnit af de to mål.

$$\text{f1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precison} + \text{Recall}} \tag{4.18}$$

Forskellen mellem PR og ROC er fremtrædende, når antallet af negative observationer er betydeligt større end positive [52]. Dette er tilfældet i en multiklasse-klassifikationsmodel, da denne, som beskrevet i forbindelse med ROC, må evalueres ved at imitere en binær klassifikation for hver klasse. Herved anses hver klasse iterativt som den positive klasse, mens de øvrige klasser samlet udgør den negative klasse, hvorfor den negative klasse vil være større end den positive.

4.7 Evaluering af vigtigste EOJ-features

Som en del af evalueringen af klassifikationsmodellerne undersøges det, hvilke features der er udslagsgivende for, at en borger tilskrives en bestemt klasse. Dette mål for hver features indflydelse på klassifikationen kaldes *feature importance* og evalueres for både den implementerede RF og NN. Til undersøgelse heraf anvendes de to metoder beskrevet i de følgende afsnit.

4.7.1 Permutation importance

Den første metode, der anvendes til udregning af feature importance, er *Permutation importance*. Her evalueres indflydelsen af hver feature ved at blande værdierne herfor vilkårligt blandt borgerne. Herefter trænes en ny model med dette modificerede datasæt. Den pågældende feature's vigtighed defineres som forskellen i klassifikationsperformance for

denne nye model og den originale model. Som mål for klassifikationsperformance anvendes f1-scoren, beskrevet i afsnit 4.6.2, da denne er et samlet mål for Precision og Recall samt tager højde for et ubalanceret antal klasser. Dette gentages for alle features. Herved kan feature importance undersøges uden af negligere interaktion mellem features [53]. Denne metode anvendes, da den er applicérbar på alle typer modeller [53] og dermed giver et sammenligneligt resultat for de to implementerede modeller i dette projekt.

4.7.2 Drop column importance

En yderligere tilgang til undersøgelse af feature importance er gennem metoden *Drop Column*. Her undersøges vigtigheden af en feature ved at træne en model uden den pågældende feature, og sammenligne klassifikationsperformance med den originale model med alle features. Dette gentages for alle features. Denne metode viser desuden, om nogen features har negativ indflydelse på klassifikationsperformance, målt ved f1-scoren.

Hyperparameteroptimering 5

Dette kapitel beskriver resultaterne af den udførte Bayesian optimization, beskrevet i afsnit 4.4.1, til bestemmelse af hyperparametre for hver af de implementerede klassifikationsmodeller, RF og NN. For begge modeller anvendes 20 initierende tilfældige gæt på et hyperparameteropsæt og 100 iterationer af sandsynlighedsbegrundede hyperparameterevalueringer.

5.1 Random Forest

De hyperparameterværdier, der blev identificeret gennem optimering af den implementerede RF-model er opstillet i **tabel 5.1**. Som beskrevet i afsnit 4.4.2 baseres optimeringen på f1-scoren som mål for model-performance, hvorfor denne anvendes som kriteriefunktion. Søgeintervallerne for hyperparameterne er sat i henhold til indledende empiriske undersøgelser og så de er brede nok til at inkludere realistiske værdier.

Tabel 5.1. Hyperparametre anvendt i den implementerede Random Forest-model, med henholdsvis det originale og det trininddelte data. Værdierne herfor er identificeret gennem bayesiansk optimering i de angivne søgeintervaller.

Hyperparameter	**Søgeinterval**	**Originalt**	**Trininddelt**
Antal decision trees	[10;1500]	1000	1430
Maksimale dybde	[3;500]	148	246
Maksimale antal terminalnoder	[7;300]	150	208
Mindste antal observationer for split	[2;20]	2	2
Maksimale antal features til split	[10%;99%]	99%	30%
Kvantificeringskriterie	[Gini imp., Entropy]	Gini imp.	Gini imp.
Cross-validation f1-score		0.8990	0.8228

Gennem træning med disse parametre opnås RF-modeller, med henholdsvis 263 og 341 knudepunkter i gennemsnit per individuelt decision tree, på originalt og trinindelt data. Ydermere har modellerne en gennemsnitlig maksimal dybde af de individuelle træer på henholdsvis 18 med det original datasæt og 19 med det trinindelte data.

5.2 Neuralt netværk

For NN-modelles behandles arkitekturen, i form af antal af lag og noder, som hyperparamtre og optimeres sammen med dropout rate, learning rate og batch size. Derudover sammenholdes de to optimizers, Adam og RMSprop, for bedste resultater. Optimeringen tager udgangspunkt i f1-scoren på det fulde træningssæt, som mål

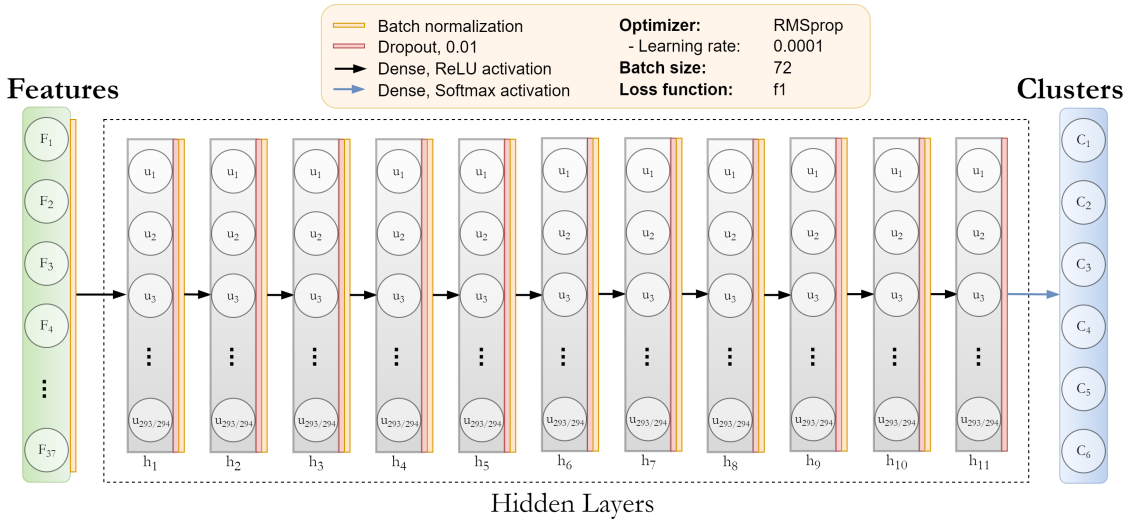
for modelperformance. Optimeringen forløber med tre indledende empiriske probes af parametre, identificeret ved opsætning af netværkdesignet. Hvert testet netværk trænes indtil f1-core på et tilfældigt 20 % udsnit af træningsdataen ikke mindskes over 50 epochs, hvorfra det bedste netværk gemmes til evaluering.

De hyperparameterværdier der blev identificeret gennem optimering af det implementerede NN er opstillet i **tabel 5.2**.

Tabel 5.2. Hyperparametre anvendt i det implementerede neurale netværk, med henholdsvis det originale og det trininddelte data. Værdierne herfor er identificeret gennem bayesian optimering i de angivne søgeintervaller.

Hyperparameter	**Søgeinterval**	**Baseline**	**Trininddelt**
Antal hidden-lag	[1;11]	11	11
Antal hidden units	[8;512]	293	294
Batch size	Antal clusters * [5;12]	72	72
Dropout rate	[0.01;0.5]	0.01	0.01
Learning rate	[0.0001;0.1]	0.0001	0.0001
Optimizer	[Adam, RMSprop]	RMSprop	RMSprop
f1-score		0.8554	0.7337
Epoch ved patience 50		45	90

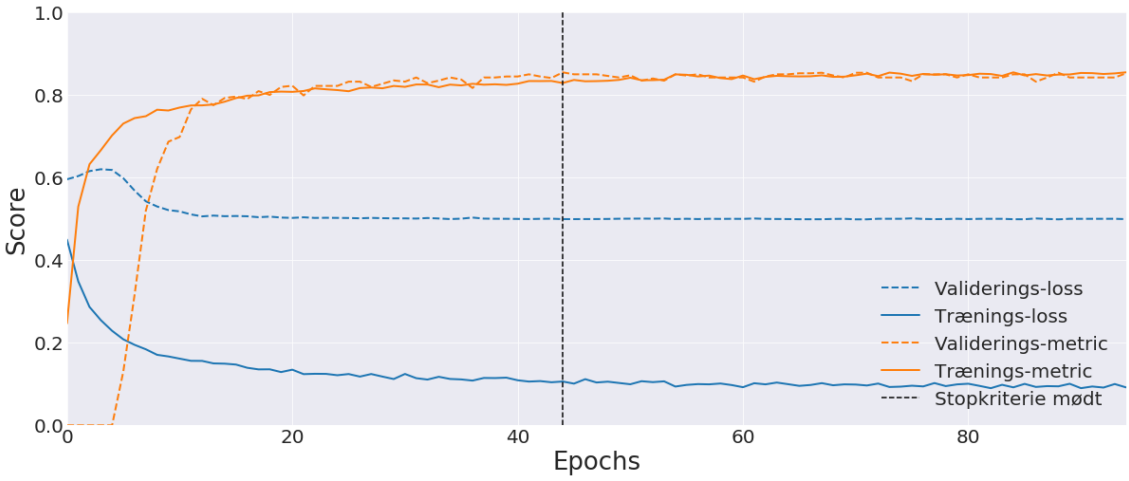
Optimeringen resulterede i en modelarkitektur som fremgår af **figur 5.1**. Modellen vil, i kontekst af det gældende datasæt, fremover blive refereret til som NN. I hvert lag er der implementeret batch-normalisering og et dropout på 1 %, dog med to undtagelser. Der appliceres ikke dropout inden første hidden-lag, for at undgå at inducere en fiktiv "missing data". Desuden er der ikke tilføjet batch-normalisering inden output-laget, da det manipulerer med softmax-funktionen og skaber forkerte cluster-sandsynligheder.



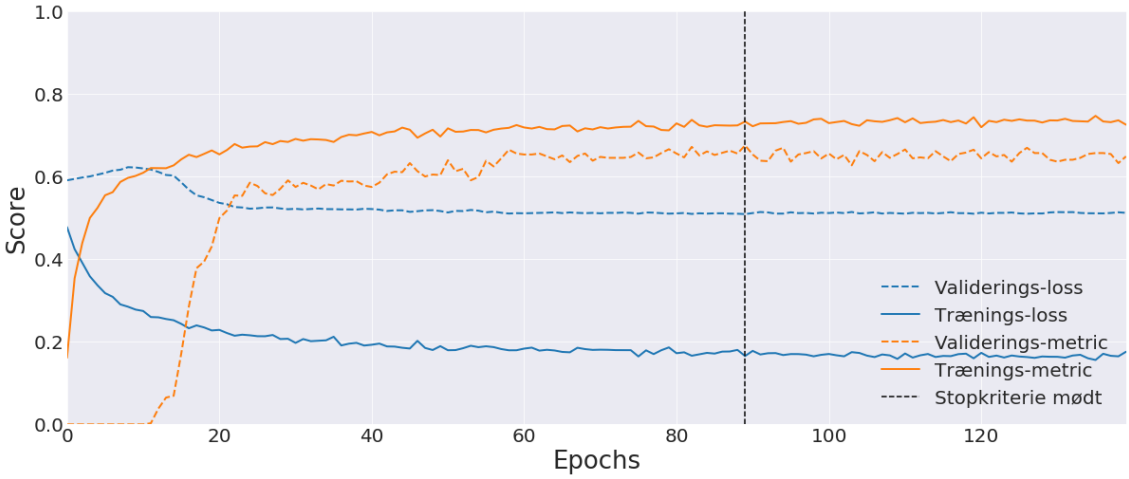
Figur 5.1. Arkitektur for Neuralt Netværk (NN) til cluster klassificering, optimeret til 293 hidden units, med det originale data, og 294 hidden units, med trininddelt data, i hvert hidden layer.

Modellerne blev konfigureret til at stoppe træningen ved konvergering af f1 i over 50

epochs. Trænings- og validerings-loss samt f1-score, for hvert epoch, kan ses på **figur 5.2** med originalt data og på **figur 5.3** med trininddelt data. Begge modeller ser ud til at være stoppet ved konvergerende metrics. Loss er tydeligt højere på det trininddelte data og f1-lavere, men derudover har begge modeller lignende forløb. Dog er der et større mellemrum mellem trænings- og validerings-f1 på det trininddelte data der også kan indikere at modellen overfitter mere end på det originale data.



Figur 5.2. Loss (soft-f1) og f1-score gennem træning af modellen optimeret på originalt data.



Figur 5.3. Loss (soft-f1) og f1-score gennem træning af modellen optimeret på trininddelt data.

Resultater 6

I de følgende afsnit beskrives resultaterne opnået for de forskellige dele fra projektoversigten, præsenteret i afsnit 4.1.

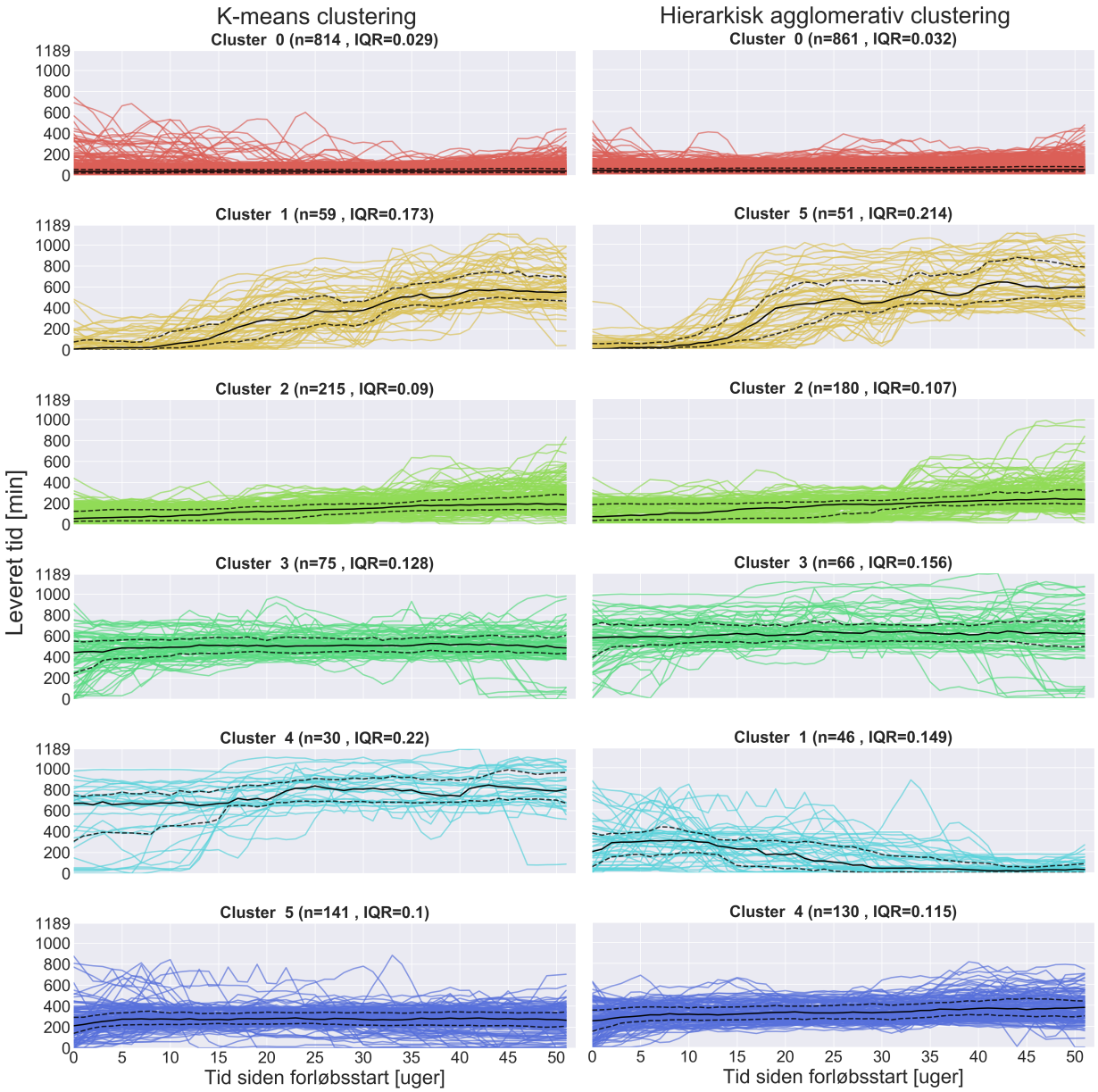
6.1 Clustering af longitudinal udvikling i leverede hjemmehjælp

Første del af dette projekt var at identificere fænotyper for udviklingen i ældre borgeres plejebehov. Dette opnås gennem clustering af forløb af antal leverede hjemmehjælpsminutter per uge i 52 uger. Desuden undersøges clustering af disse forløb, når værdierne herfor trininddeles i ni trin, for at udligne ubalancen mellem antallet af datapunkter, med høje og lave værdier i leverede hjemmehjælpsminutter. Til applicering af clustering er en HAC og en KMC implementeret, som beskrevet i henholdsvis afsnit 4.3.2 og afsnit 4.3.3. Jævnfør evaluering af en række cluster metrics, som beskrevet i afsnit 4.3.4, er HAC og KMC implementeret med $k = 6$, til identificering af seks clusters for det originale datasæt. For det trininddelte datasæt er HAC implementeret med $k = 7$, mens KMC har $k = 6$. Først og fremmest evalueres og sammenlignes performance for de implementerede clusteringmodeller i relation til det originale og trininddelte datasæt. Dernæst undersøges, hvorvidt de udledte clusters kan anvendes som fænotyper for udviklingen i hjemmeplejebehov og dermed repræsentere de individuelle borgere i hver cluster.

6.1.1 Clustering performance

De resulterende clusters for forløbene af plejebehovet i det originale datasæt er visualiseret i **figur 6.1** sammen med det respektive gennemsnitlige typeforløb.

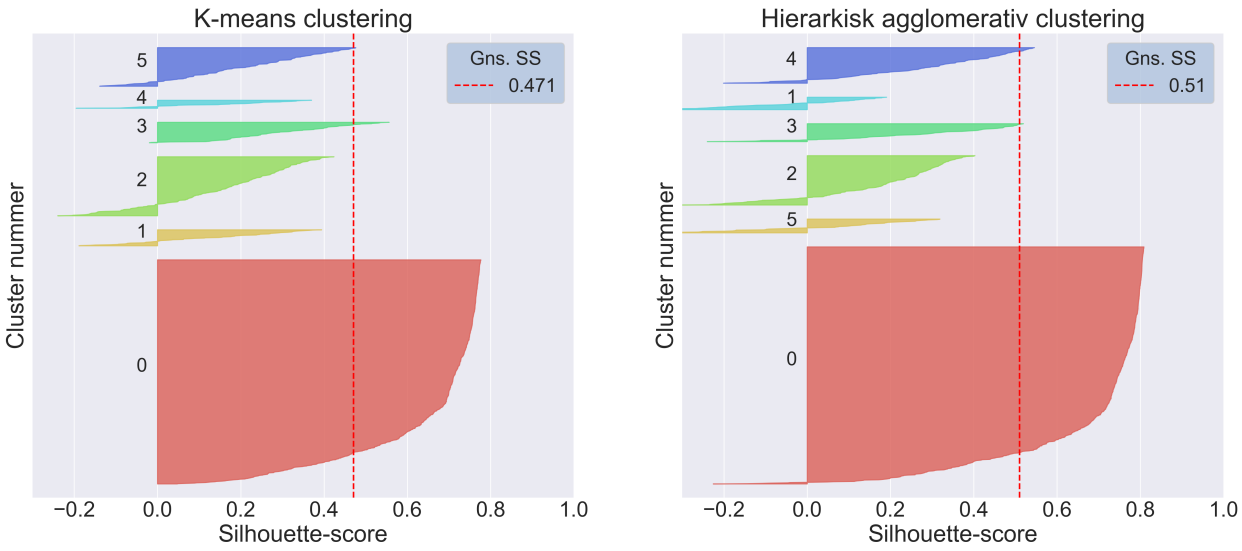
Her ses det, at begge clustering-modeller danner en majoritetsgruppe (cluster 0) med 61.02% (n=814) og 64.54% (n=861), af det samlede antal observationer (N=1334), for henholdsvis KMC og HAC. Samtidig udgør de mindste clusters 2.25% (n=30) for KMC, cluster 4, og 3.45% (n=46) for HAC, cluster 1. Generelt synes de to modeller at udlede sammenlignelige clusters, hvor hver cluster har en tilsvarende udgave i den konkurrerende model. **Figur 6.1** er organiseret, så hver cluster fra KMC er sidestillet med den mest lignende fra HAC. Heraf ses det, at der for cluster 4 fra KMC og cluster 1 fra HAC, ikke findes en ækvivalent cluster i den modsatte model, hvilket skaber variation mellem de resulterende typeforløb for de to modeller. Den unikke cluster i KMC består af de mest plejetunge borgere, der resultere i et højt typeforløb med et gennemsnit på omkring 750 leverede minutter. Modsat består den i HAC af borgere der starter med omkring 200 minutters leveret pleje men falder støt over det første år. For de ensartede cluster-par opnår



Figur 6.1. De seks clusters af forløb i antal leverede hjemmehjælpsminutter over 52 uger. Hver cluster er visualiseret med alle medlemmer og det respektive gennemsnitsforløb (sort). Disse er identificeret med henholdsvis den implementerede k-means og hierarkisk agglomorativ clustering.

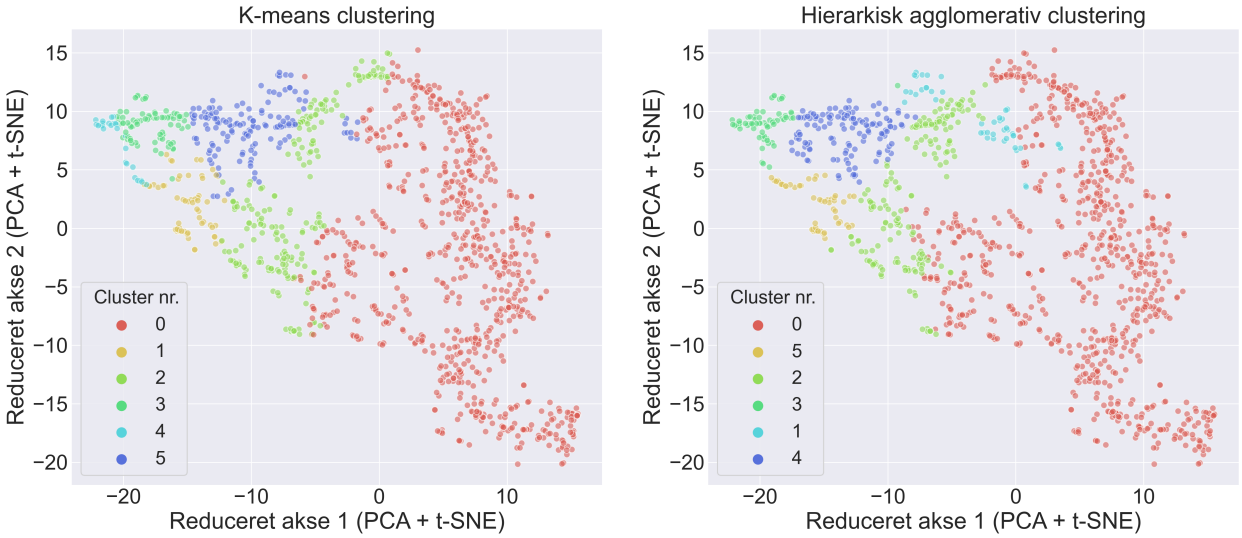
KMC en højere IQR. For yderligere undersøgelse af intra-cluster similaritet sammenholdt med inter-cluster dissimilaritet, er der opstillet et silhouette-plot i **figur 6.2**.

HAC-modellen opnår en højere gennemsnitlig silhouette-score på 0.51, mod KMC's 0.471. Derudover afslører silhouette-graferne, at de mindre clusters generelt har lavere silhouette-score end majoritetsgruppen og dermed trækker den samlede score ned. Samtidig udviser alle clusters fra HAC flere og lavere negative værdier end KMC, hvilket indikerer at der for HAC er flere observationer, der sandsynligvis er clustret forkert end for KMC.



Figur 6.2. Silhouette-plot for henholdsvis k-means og hierarkisk agglomorativ clusting appliceret det originale datasæt. Gns. SS betegner den gennemsnitlige silhouette-score over alle clusters. Denne er markeret med en stiplet vertikal linje.

For at undersøge hvordan de identificerede clusters fordeler sig i feature space, er alle observationer visualiseret i et reduceret features space med deres respektive cluster-label i **figur 6.3**.



Figur 6.3. Observationer clusteret med henholdsvis k-means og hierarkisk agglomorativ clustering. Disse er visualiseret i et reduceret feature space beregnet gennem Principal Component Analysis og t-Distributed Stochastic Neighbor Embedding.

Dette reducerede feature space er opnået gennem applicering af PCA og t-SNE, som beskrevet i afsnit 4.3.1. Udledninger fra dette plot skal ses i lyset af at feature space er reduceret fra 52 dimensioner til to, hvilket betyder at repræsentationen ikke er fuldstændig og observationer derfor kan være placeret anderledes i forhold til hinanden. Det reducerede feature space afdækker, at der er små forskelle i, hvor modellerne placerer *decision*

boundaries i forbindelse med at bestemme hvilke observationer, der tilskrives de forskellige clusters. Eksempelvis ses det for cluster 5 i KMC og cluster 4 i HAC, der er visuelt tilsvarende, at grænsen mellem disse grupper og den tilstødende cluster 3 er placeret forskelligt. Dette plot afslører desuden baggrunden for den markante forskel mellem cluster 4 fra KMC og cluster 1 fra HAC, idet der er tale om to forskellige områder af feature space. Det ses desuden at cluster 1 for HAC er udgjort at to separate clusters, mens den tilsvarende cluster 4 for KMC består af en samlet cluster. Dette er kontrast til disse clusters respektive IQR, hvor cluster 1 fra HAC indikeres at være mere kompakt $IQR_{cluster1HAC} = 0.149$ mod $IQR_{cluster4KMC} = 0.22$. Samtidig afslører silhouette-plottet dog at observationerne i cluster 1 for HAC har et svagere tilhørsforhold end observationer i cluster 4 for KMC.

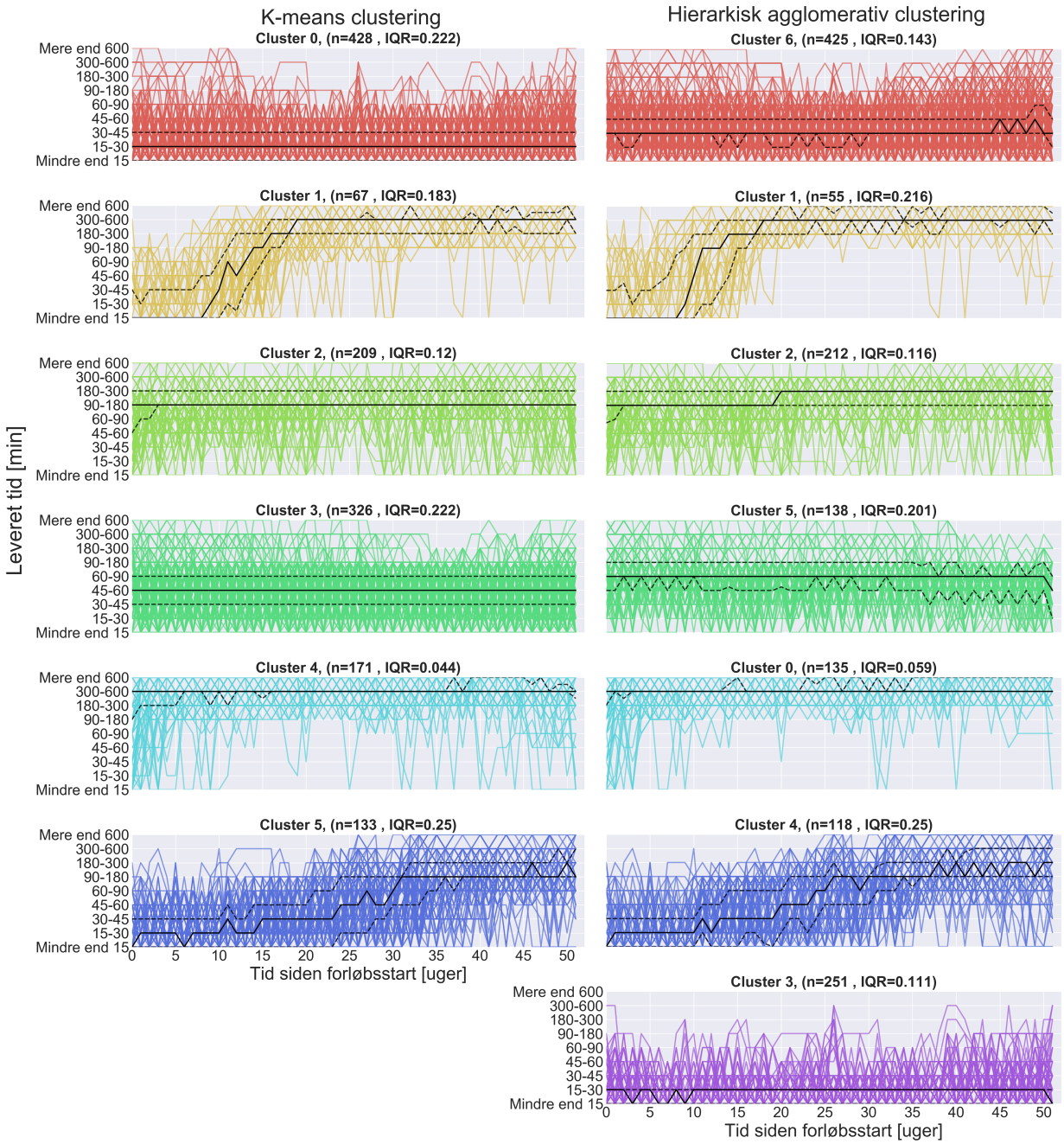
På samme måde evalueres de identificerede clusters for det trininddelte datasæt. De sammenlignelige clusters fra hver af de implementerede modeller er visualiseret sideordnet i **figur 6.4**.

På trods af at HAC er sat til finde syv clusters, mens KMC finder seks, gælder det ligeledes for det trininddelte datasæt, at de to implementerede clustering-modeller danner lignende clusters. Det ses desuden, at den yderligere cluster i HAC, cluster 3, udviser et lavt. stabilt niveau der er sammenligneligt med cluster 0 for KMC. Ligesom for det originale datasæt, ses en tendens til at danne en majoritetsgruppe, hvor det for KMC er 32.08% (n=428) af det totale antal observationer og 31.86% (n=425) for HAC. Minoritetsgruppen er 5.02%(n=67) for KMC og 4.12% (n=55) for HAC. De udledte clusters er dog generelt bedre balanceret i relation til antallet af medlemmer for det trininddelte datasæt. Samtidig ses en generel højere IQR. Dette tilskrives den bedre balancering af dataværdierne i det totale range, som beskrevet i afsnit 4.3.1. Denne balancering er endvidere tydelig på silhouette-plottet i **figur 6.5**. Her ses det, hvordan den yderligere cluster for HAC resulterer i en betydelig andel negative værdier. Ud fra mængden af negative værdier i silhouette-scorene har grupperne i HAC generelt flere observationer, der sandsynligvis er clusteret forkert.

For det trininddelte data, er det reducerede feature space og fordelingen af clusters heri visualiseret i **figur 6.6**. Her ses det, at de sidestillede cluster 0, i KMC, og cluster 6, i HAC, blot har et mindre område med overlap, og derudover findes på forskellige områder af feature space. Derimod udgør den ekstra cluster 3 for HAC, et betydelige område af cluster 0 fra KMC.

De beregnede cluster metrics for de to implementerede modeller ved applicering på henholdsvis det originale og trininddelte datasæt er opstillet i **tabel 6.1**. Her ses det, at KMC opnår bedst performance af de to modeller på flest af de inkluderede cluster metrics med begge datasæt. Af de to datasæt scorer KMC derudover bedst på det originale.

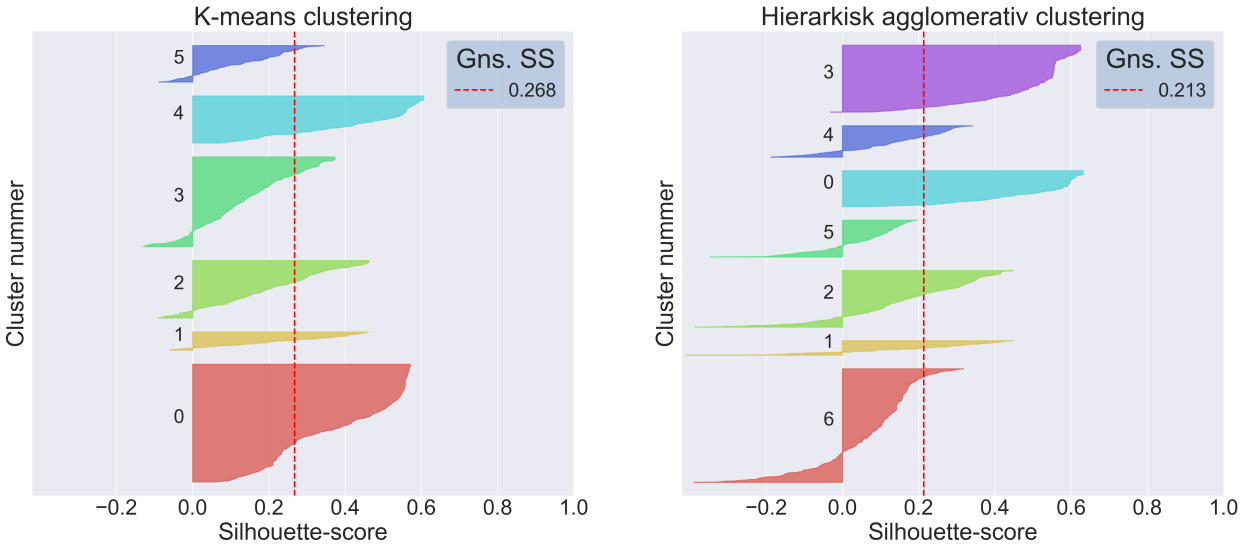
Alt i alt, outperformer KMC svagt HAC, på det originale datasæt, baseret på de beregnede cluster metris samt silhouette-plottet, der viser færre observationer med usikkerhed i tilhørsforholdet for de individuelle clusters i KMC. Endvidere er majoritets-gruppen for HAC størrer, hvilket er en udfordring for den videre anvendelse i klassifikationsmodellerne og dermed ikke ønskværdigt. Ligeledes gælder det for det trininddelte datasæt, at KMC outperformer HAC, målt på både cluster metrics og silhouette-score. Den yderligere cluster udledt i HAC, øger dermed ikke clustering performace.



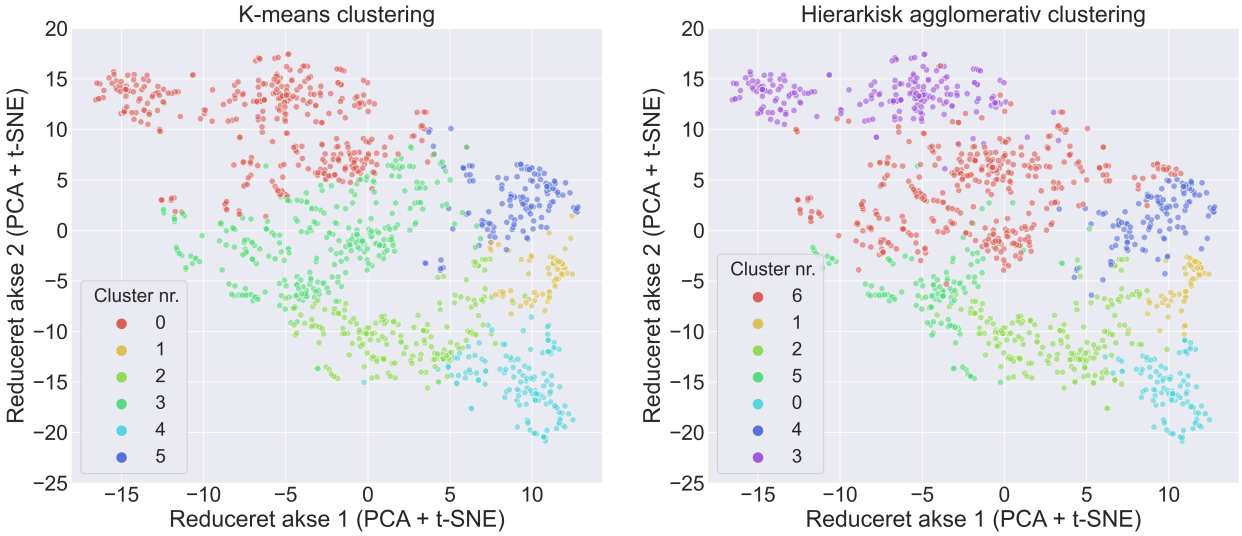
Figur 6.4. Clusters af trininddelte forløb i antal leverede hjemmehjælpsminutter over 52 uger. Hver cluster er visualiseret med alle medlemmer og det respektive gennemsnitsforløb (sort). Disse er identificeret med henholdsvis den implementerede k-means og hierarkisk agglomorativ clustering.

6.1.2 Evaluering af typeforløb som approksimation af individuelle plejebehovsforløb

For hver cluster beregnes et typeforløb herfor som medianen af datapunkterne for hver uge over alle 52 uger. Medianen anvendes da datasættets overvægt af lave værdier, som beskrevet i afsnit 4.3.1, bevirker at normalfordeling ikke kan antages. Typeforløb er visualiseret i **figur 6.7** for det originale datasæt og **figur 6.8** for det trininddelte data.



Figur 6.5. Silhouette-plot for henholdsvis k-means og hierarkisk agglomorativ clustering appliceret det trininddelte datasæt. Gns. SS betegner den gennemsnitlige silhouette-score over alle clusters. Denne er markeret med en stiplet vertikal linje.

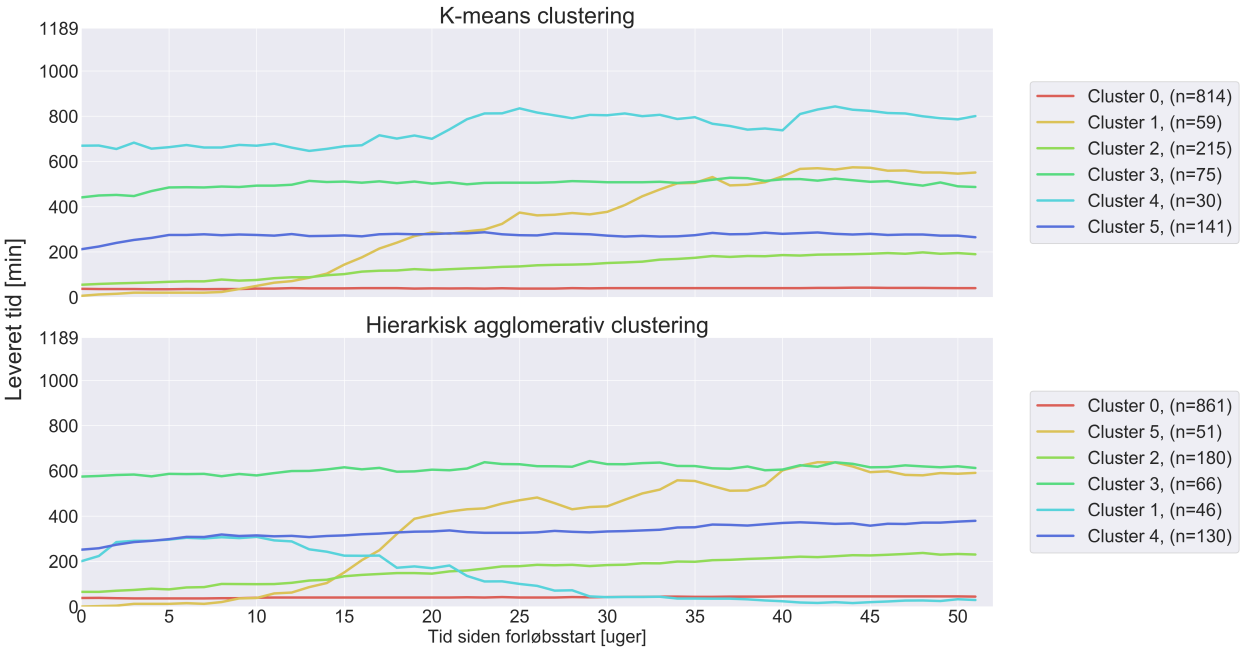


Figur 6.6. Observationer clusteret med henholdsvis k-means og hierarkisk agglomorativ clustering for det trininddelte datasæt. Disse er visualiseret i et reduceret feature space beregnet gennem Principal Component Analysis og t-Distributed Stochastic Neighbor Embedding

Herudfra ses det, at begge metoder var i stand til, at udlede unikke forløb i dataen. Det ses desuden at KMC har det højeste forløb på det originale data, men at ingen af de to modeller udleder et typeforløb i den øverste tredjedel af data-range. Ligeledes er der ingen af de to modeller, der på det trinopdelte data, har typeforløb i øverste trin (Mere end 600). Generelt udleder begge modeller en række stabile typeforløb uden markant ændring i plejebehov. Dette gælder for begge datasæt. Disse stabile typeforløb differentieres blot på niveauet af plejebehov. Desuden identificerer de to modeller, for det originale datasæt, et typeforløb (cluster 1 for KMC og cluster 5 for HAC) med udgangspunkt i et lavt plejebehov der efter

Tabel 6.1. Cluster metrics for clusteranalysen med hierarkisk agglomorativ (HAC) og k-means clustering (KMC).

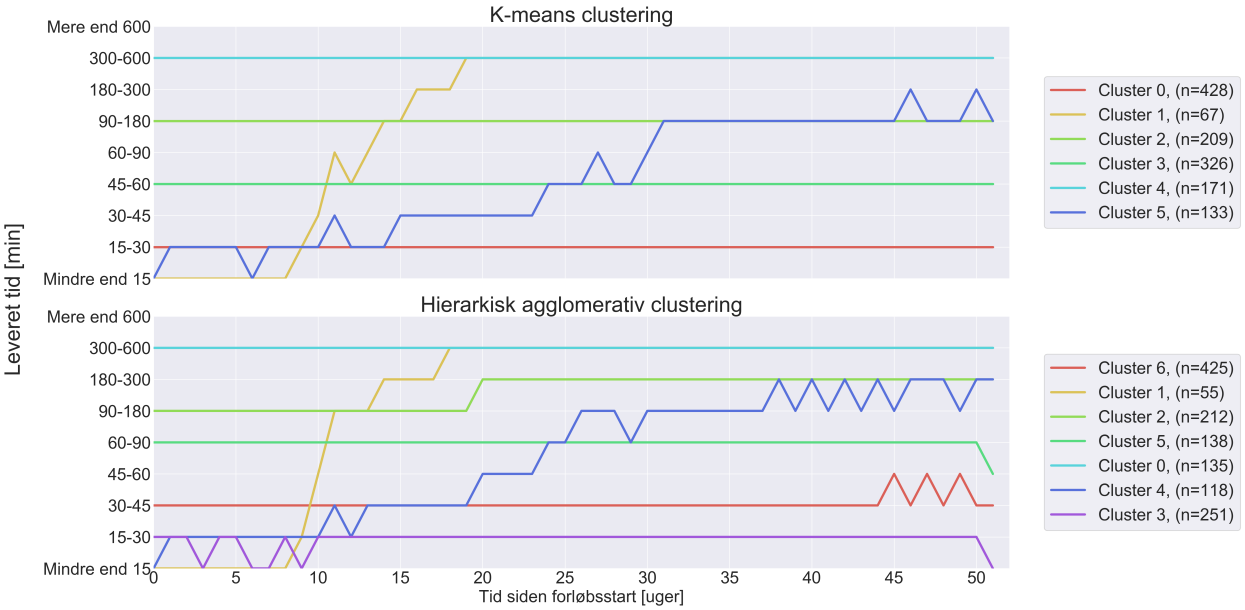
	HAC		**KMC**	
	Originalt	Trininddelt	Originalt	Trininddelt
Antal clusters (k)	6	7	6	6
Summeret IQR	0.773	1.095	0.74	1.041
Elbow-metode	57.509	223.354	51.282	250.502
Silhouette-score	0.510	0.213	0.471	0.268
Davies-Bouldin index	1.266	1.445	1.159	1.327



Figur 6.7. Typeforløb for antal leverede hjemmehjælpsminutter. Disse er beregnet som gennemsnitsforløbet af de resulterende clusters fra den implementerede k-means og hierarkisk agglomorative clustering-algoritme.

10-15 uger stiger støt til et ugentligt niveau i leveret hjemmehjælp omkring 600 minutter. En lignende tendens ses for det trininddelte datasæt, hvor begge modeller udleder to ikke-stabile typeforløb med forskellig grad af stigning i antal af leverede hjemmeplejeminutter, efter cirka 10 uger.

For at evaluere hvor tilstrækkeligt typeforløbene formår, at repræsentere udviklingen i den enkelte borgers plejebehov, er en række repræsentative eksempler opstillet i **figur 6.9**. Eksemplet for det originale datasæt viser, at der for borgerne med højest afvigelse fra typeforløbet for både KMC og HAC forekommer en stejl stigning i leveret tid for eksemplet i KMC og et brat fald for eksemplet i HAC. Disse forløbskarakteristika er ikke repræsenteret ved nogen af de udledte typeforløb og deraf heller ikke det tildelte. Modsats ses det at eksemplerne med lavest afvigelse fra typeforløbet, har samme niveau i plejebehov og følger den stabile udvikling i plejebehov portrætteret ved de tildelte typeforløb.



Figur 6.8. Typeforløb for trininddelt antal leverede hjemmehjælpsminutter. Disse er beregnet som gennemsnitsforløbet af de resulterende clusters fra den implementerede k-means og hierarkisk agglomorative clustering-algoritme.

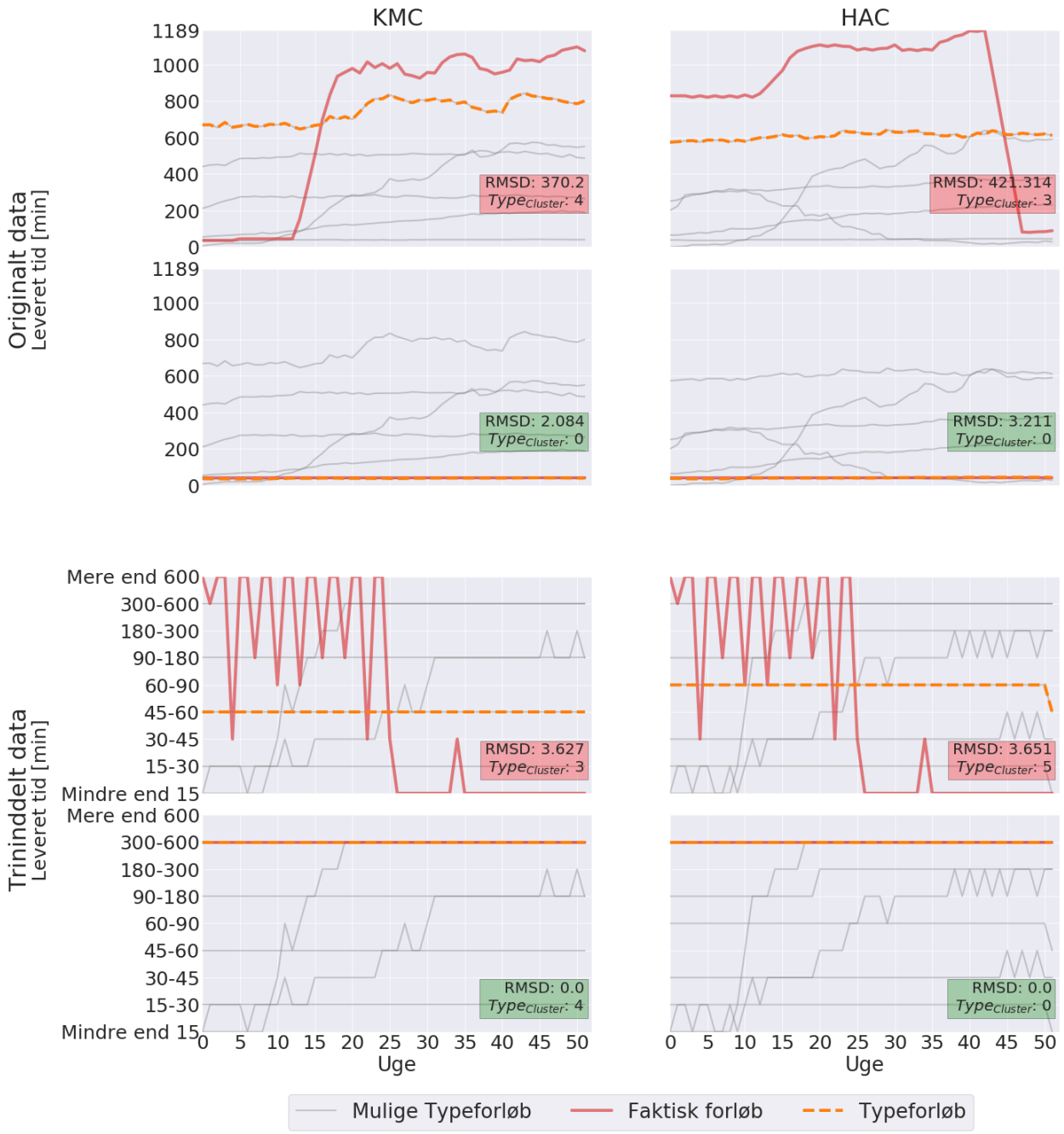
For eksemplet med højest afvigelse fra typeforløbet i det trininddelte datasæt forekommer borgerens faktiske forløb irregulært i forhold til de mulige typeforløb, hvorved ingen af de udledte typeforløb vil kunne repræsentere forløbet. Af de inkluderede eksempler med laveste RMSD fra de tildelte typeforløb, ses det at diskretiseringen i det trininddelte data resulterer i en fuldstændig repræsentation af det faktiske forløb.

Den samlede afvigelse i RMSD mellem de faktiske forløb og typeforløbene er visualiseret ved et boksplot for de to implementerede modeller i **figur 6.10**. KMC scorer en anelse bedre end HAC med 38.4 mod 40.6 minutter, men opnår ikke statistisk signifikant forskel i RMSD. Derudover ses det, at det trininddelte datasæt opnår en højere relativ afvigelse. For RMSD i modellerne er mellem 28 og 75 observationer outliers, resulterende i 2-6 % af den samlede datamængde.

For at opsummere opnår KMC-modellen generelt højre performance end HAC. Dette omfatter både intra-cluster IQR, clustering metrics, silhouette-plots og RMSD mellem borgernes faktisk forløb og det tildelte typeforløb. Med afsæt i disse fund anvendes de udledte clusters, og dermed fænotyper, fra KMC i studiets videre dele.

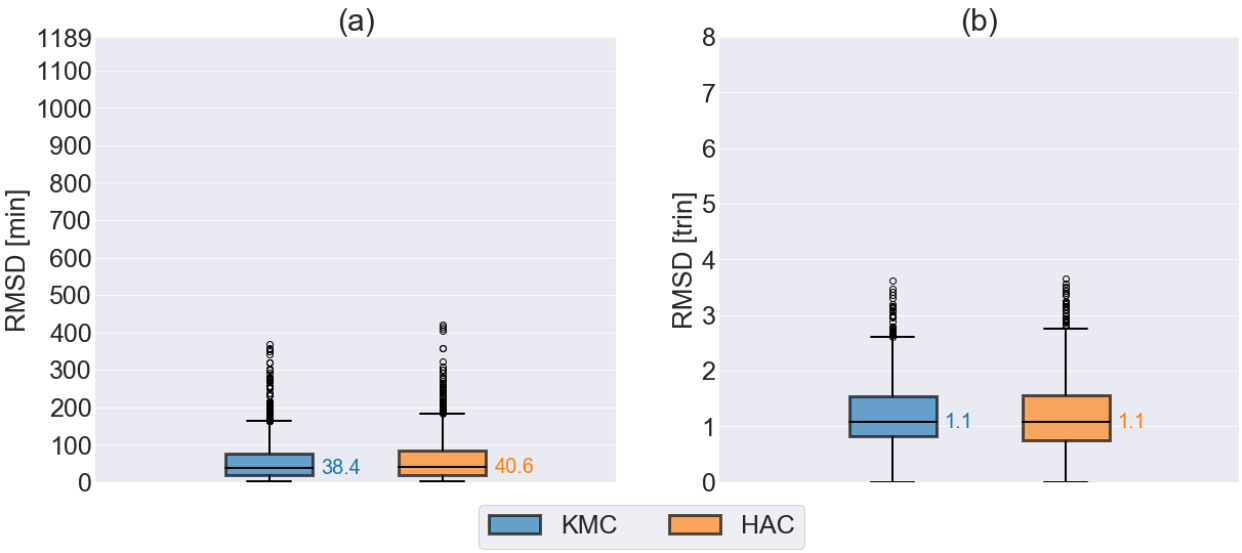
6.1.3 Homogenitet i de udledte clusters og anvendelighed som fænotyper

For at evaluere hvorvidt de identificerede fænotyper fra KMC udgør homogene grupper, blev det undersøget, hvad der karakteriserer borgerne i hver cluster. I denne forbindelse blev den statistiske signifikans af forskellen mellem distributionen af hver feature beregnet, som beskrevet i afsnit 4.5. Denne undersøgelse viste at der for 25 af de i alt 38 features, var statistisk signifikant forskel mellem minimum ét cluster-par. Appendiks C indeholder en oversigt med signiffikans-maps over alle cluster-par for features med signifikant forskel.

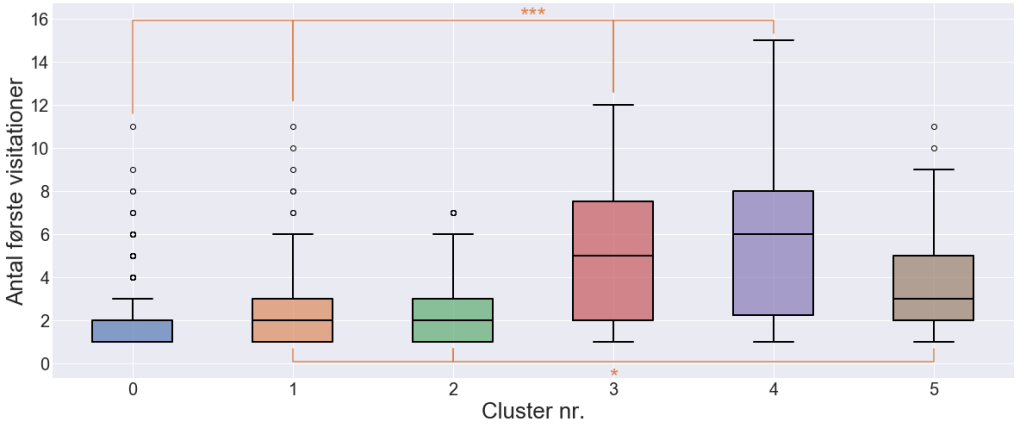


Figur 6.9. Repræsentative eksempler på borgerens faktiske forløb sammen med typeforløbet udledt fra de to implementerede clustering-modeller. Forløb med røde mærkater for RMSD tilhører borgere med mest afvigelse fra typeforløbet, mens forløb med grønne mærkater er borgere, hvis faktiske forløb ligger tættest på typeforløbet.

Undersøgelsen blev fokuseret på cluster 1, da typeforløbet for denne cluster adskiller sig fra de øvrige ved at portrættere en stigning i plejebehov. En feature der for denne cluster var statistisk signifikant forskellige fra de øvrige clusters, var blandt andre *Antal første visitationer*. Denne feature indeholder antallet af forskellige visitationer tildelt borgeren ved første visitation. Distributionerne af featureværdien for denne er illustreret gennem boksplots på **figur 6.11** for alle clusters sammen med den statistiske styrke af forskellen mellem dem.



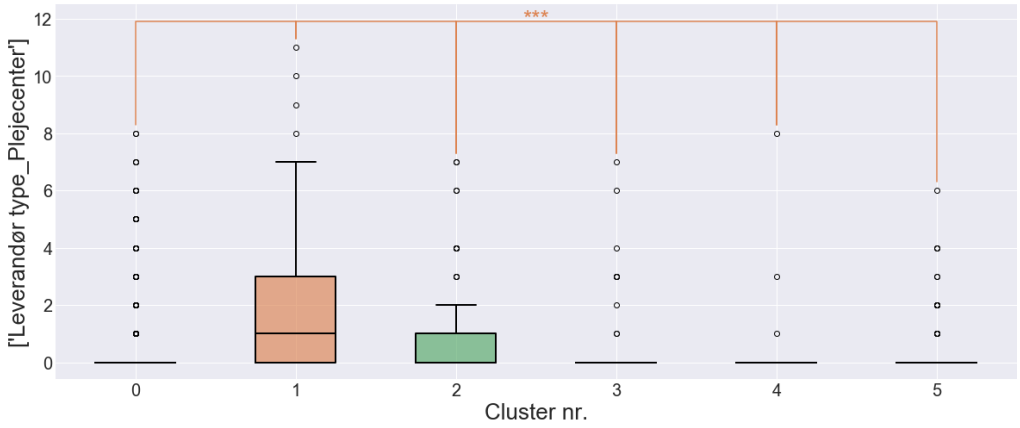
Figur 6.10. Boksplots over Root Mean Squared Deviation (RMSD) mellem borgernes faktiske forløb og typeforløbet for den tildelte cluster. For det originale datasæt (a) og det trininddelte (b). Y-aksen er afgrænset ved den maksimale mulige afvigelse for at muliggøre sammenligning mellem datasættene.



Figur 6.11. Boxplot for cluster over fordelingen af værdier for featuren *Antal første visitationer*. Styrken for den statistiske signifikans af forskellen mellem distributioner er udelukkende angivet for cluster 1 (orange).

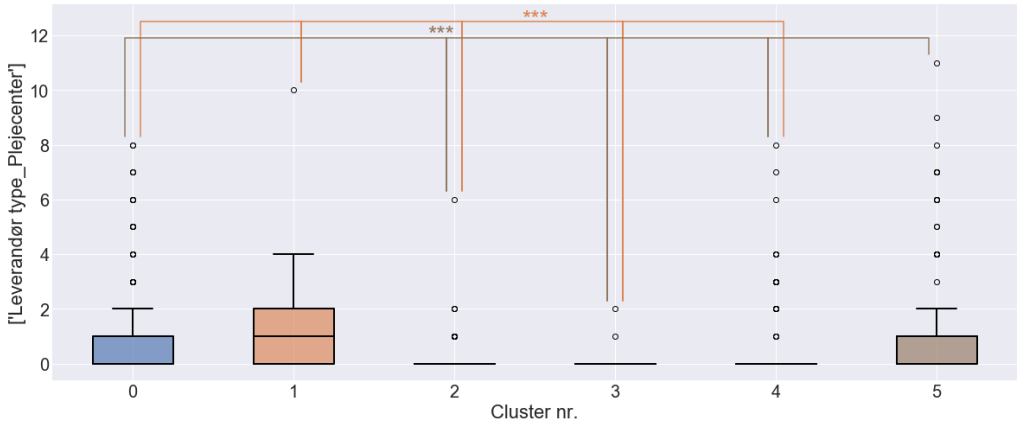
Her ses det at borgerne i cluster 1 gennemsnitligt har 2 initierende visitationer, jævnfør medianen. Samtidig ligner denne distribution cluster 2, hvortil der blot er svag statistisk signifikans af forskellen (p=0.0489). Herudfra synes det initierenede antal visitationer ikke at være unikt for cluster 1. Endvidere gælder det for featuren *Leverandør type_ Plejecenter*, at distributionen var statistisk signifikant forskellig fra de øvrige clusters. Denne distribution er, i **figur 6.12**, ligeledes opstillet i boksplot med angivelse af den statistiske styrke af forskellen til de øvrige clusters.

Her ses en markant forskel mellem distributionen for cluster 1, hvilket er underbygget af stærk statistisk signifikans i forskellen. Borgerne i cluster 1 adskiller sig dermed ved et højere niveau af featureværdien for *Leverandør type_ Plejecenter*.



Figur 6.12. Boxplot for cluster over fordelingen af værdier for featuren *Leverandør type_ Plejecenter*. Styrken for den statistiske signifikans af forskellen mellem distributioner er udelukkende angivet for cluster 1 (orange).

Ligeledes blev clusters udledt af det trindinddelte datasæt undersøgt for homogenitet. Her var statistisk signifikant forskel mellem minimum ét cluster-par for 29 af de i alt 38 features. Signifikans-maps for disse features findes i appendiks C. Da cluster 1 og 5 udviste en ikke-stabil udvikling i typeforløbet var disse fokus for videre undersøgelse. Som for det originale datasæt, var featuren *Leverandør type_ Plejecenter* statistisk signifikant forskellig mellem disse clusters og de øvrige. Dette er illustreret gennem boksplots i **figur 6.13**.



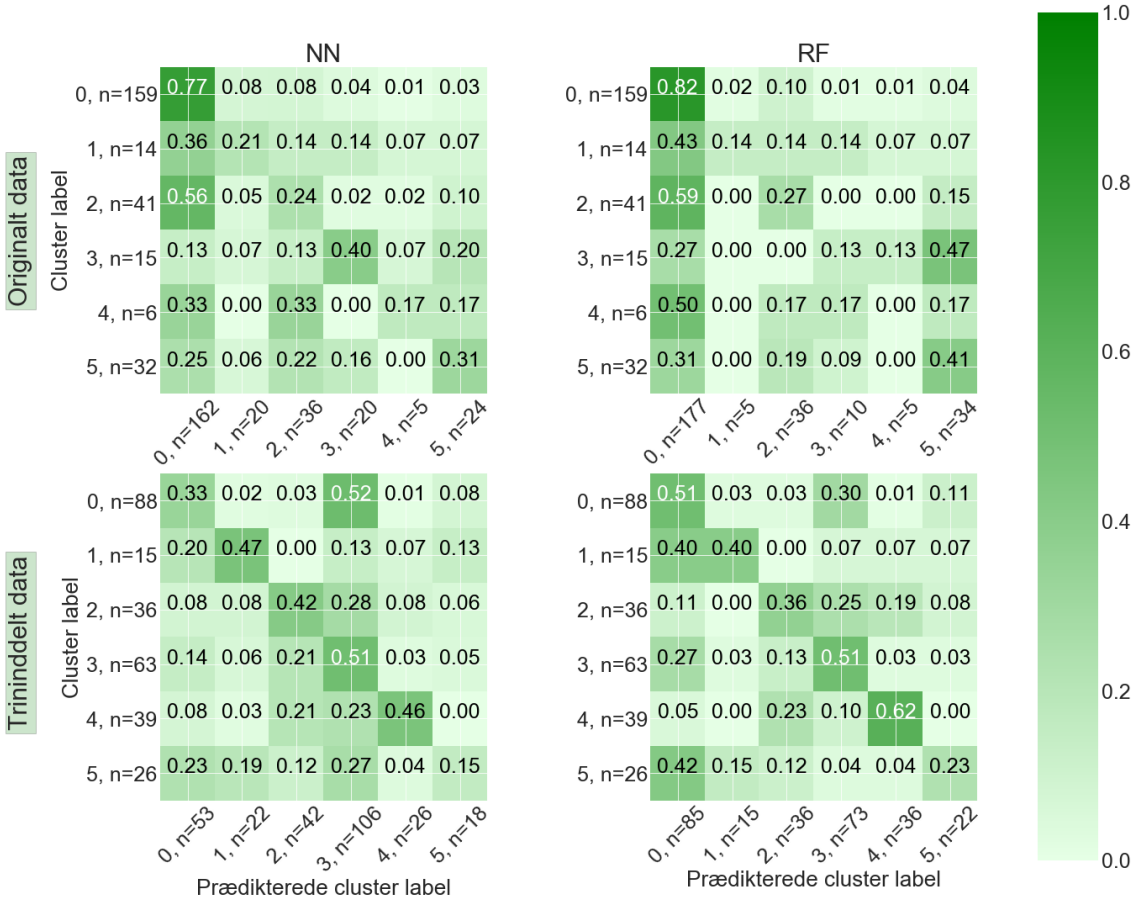
Figur 6.13. Boxplot for cluster over fordelingen af værdier for featuren *Leverandør type_ Plejecenter* for clusters udledt af de trininddelte datasæt. Styrken for den statistiske signifikans af forskellen mellem distributioner er udelukkende angivet for cluster 1 (orange) og cluster 5 (brun).

6.2 Klassifikation af plejebehovsfænotype

For at evaluere klassifikations-performance af den implementerede RF- og NN-model, beskrevet i afsnit 5.1 og 5.2, blev der foretaget prædiktioner af plejebehovsfænotype på baggrund af data for første visitation, for borgerne i testdatasættet.

Modellernes klassifikationer er illustreret ved confusion matricen i **figur 6.14**. De to modeller opnår, på det originale data, en højere gennemsnitlig accuracy på $Acc_{NN} = 0.569$ og $Acc_{RF} = 0.592$ mod $Acc_{NN_trin} = 0.393$ og $Acc_{RF_trin} = 0.472$ på det trininddelte

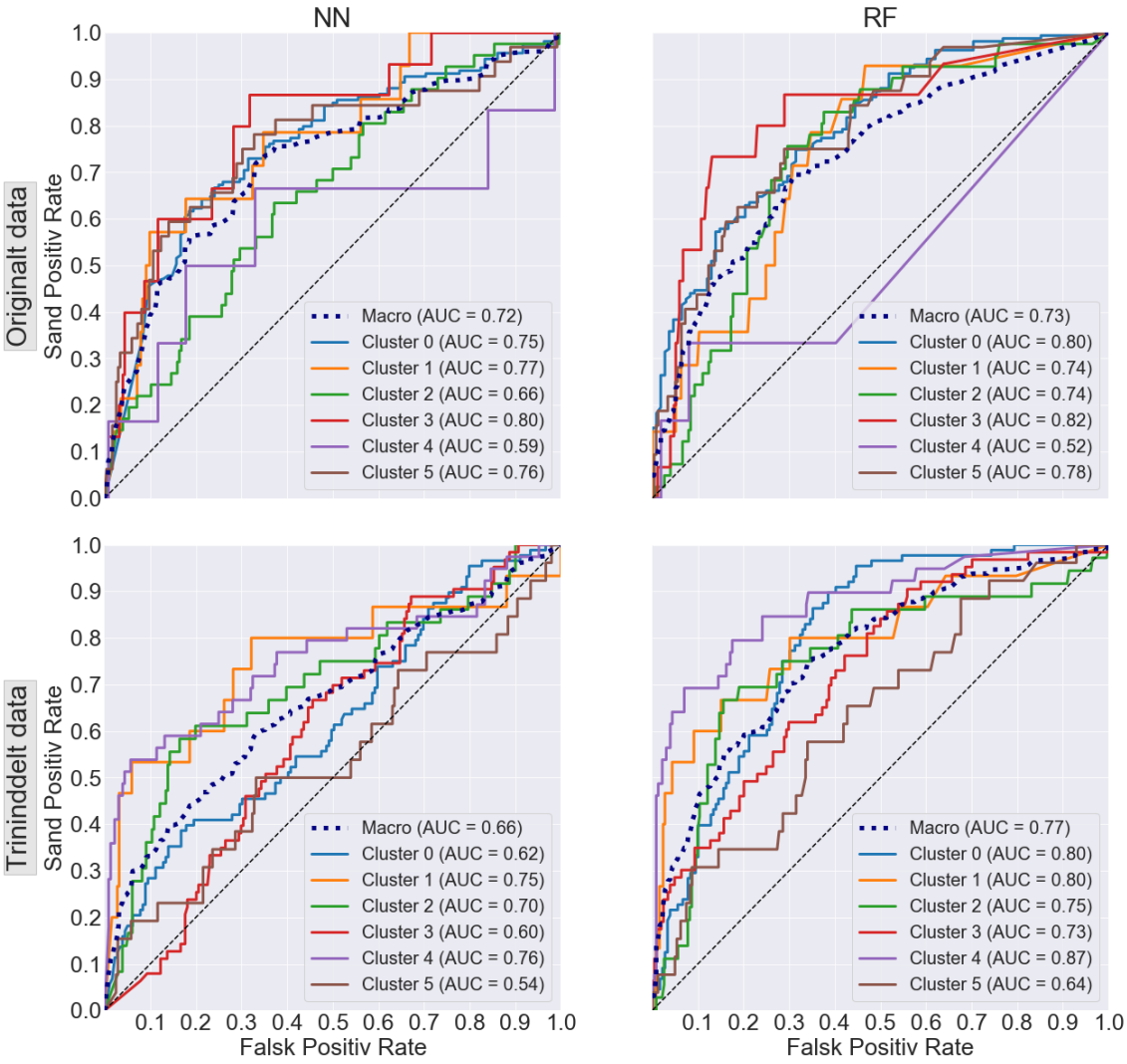
data. Som det fremgår af **figur 6.14**, er der dog tydeligere diagonaler ved det trininddelte data for begge modeller. Dette kan tilskrives den bedre accuracy på majoritets-clusteren (cluster 0) i den originale data og den mere balancerede fordeling af observationer i clusterne for det trininddelte data. Størrelsen på testsættet, resulterer dog i at confusion matricen skildrer et skævt billede af performance på det originale og trininddelte data.



Figur 6.14. Confusion matrice for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data. Hver række er normaliseret med antallet af observationer i hvert cluster label.

ROC-kurverne på **figur 6.15** viser ikke samme tendens som confusion matricerne i **figur 6.14**. Alle clusters i de samme modeller scorer sammenligneligt med undtagelse af minoritets-clusteren i det originale data (cluster 4, $n = 6$), der tydeligt bærer præg af manglende observationer. Det ses også at RF på trininddelt scorer højere ($AUC = 0.77$) end på det originale data ($AUC = 0.73$), og desuden højest af alle. Modsat scorer NN på det trininddelte lavere ($AUC = 0.66$) end på originalt data ($AUC = 0.72$) og dermed lavest. Ingen clustre i det trininddelte data bærer præg af manglende observationer på ROC-kurverne.

PR-kurven i **figur 6.16** tydeliggøre problematikken i størrelsen på testsættet og forskellen i antal af observationer mellem clusters. Kun de største clusters, særligt cluster 0 i det originale data, opnår en f1-score omkring 0.8. Gennemsnitligt opnår RF på det trininddelte data en AUC, der er klassens Average Precision (AP), på 0.43, hvilket er højere end for de resterende modeller med AP på $0.32 - 0.34$. Dermed er det den eneste model, der er tæt

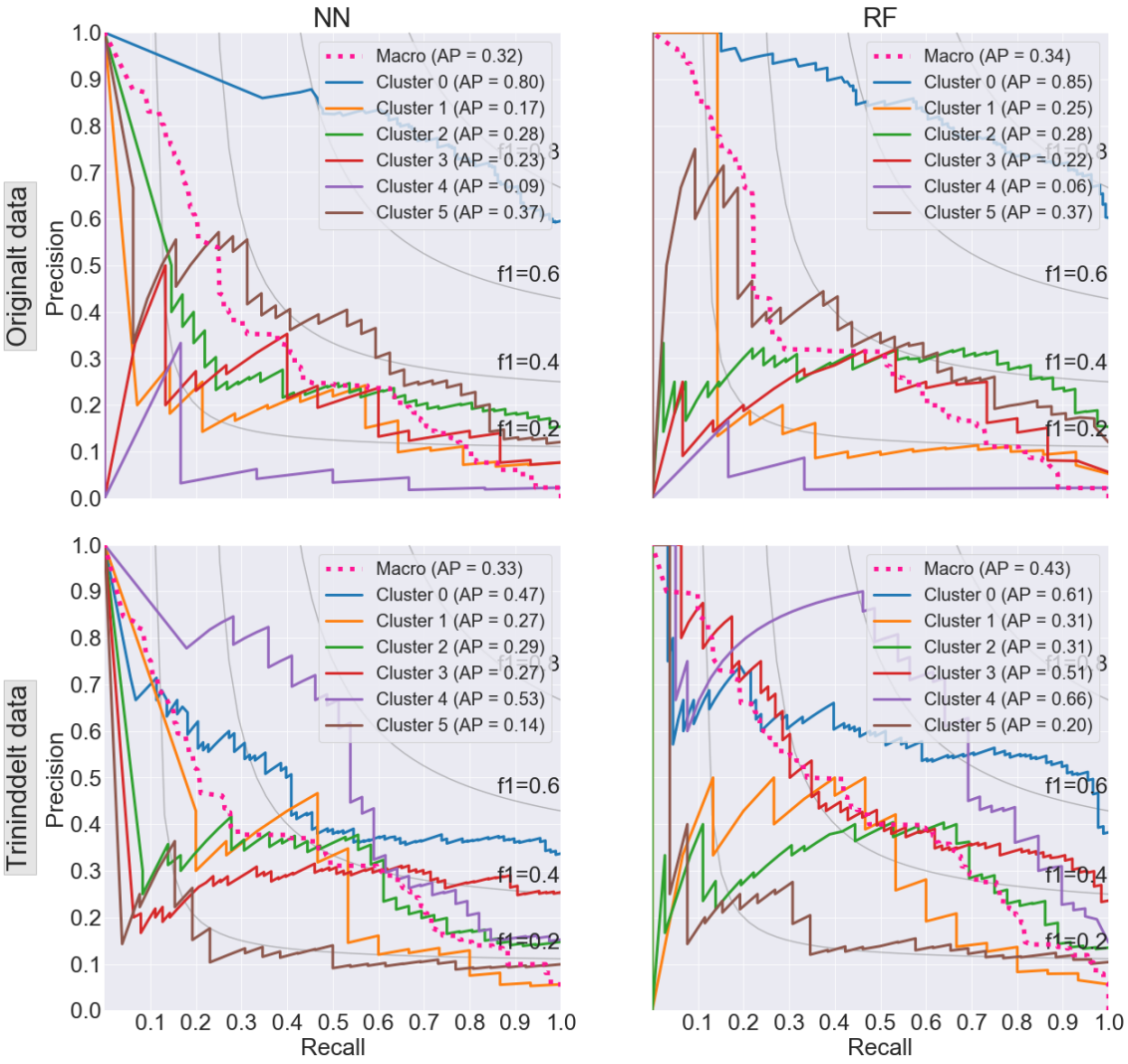


Figur 6.15. Receiver Operating Characteristic (ROC) kurver af hver cluster for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data. Alle kurver er ledsaget af Area Under the Curve (AUC). Hvert plot viser desuden macro-gennemsnitskurven (stiplede, blå linje).

på at være bedre end en tilfældig klassificering (AP>0.5) på samme datasæt, men bærer tydeligt præg af størrelsen på de mindste clusters som de fremgik af **figur 6.14** (cluster 1: n=15, cluster 2: n=36 og cluster 5: n=26).

Repræsentative eksempler med høj og lav RMSD mellem typeforløbet for den prædikterede cluster og borgerens faktiske forløb for begge modeller, med både originalt og trininddelt data, er vist på **figur 6.17**. Det er særligt værd at bemærke på **figur 6.17**, at prædiktionen for tilfældene med lavest RMSD er typeforløbe længest muligt fra det faktiske forløb. Dette gælder dog ikke for RF på det trinindelte datasæt, der prædikterer et forløb hvor de første cirka 10 uger, stemmer overens med det faktiske forløb.

Shapiro-Wilk-testen viste ikke-normalfordelte RMSD-fordelinger for begge modeller med både originalt og trininddelt data. Medianer og kvartiler for RMSD er illustreret i boxplottene i **figur 6.18**. Mann–Whitney–Wilcoxons non-parametriske test viste ingen forskel mellem grupperne ($p_{original} = 0.147$ og $p_{trininddelt} = 0.399$ mod $\alpha = 0.05$). RF



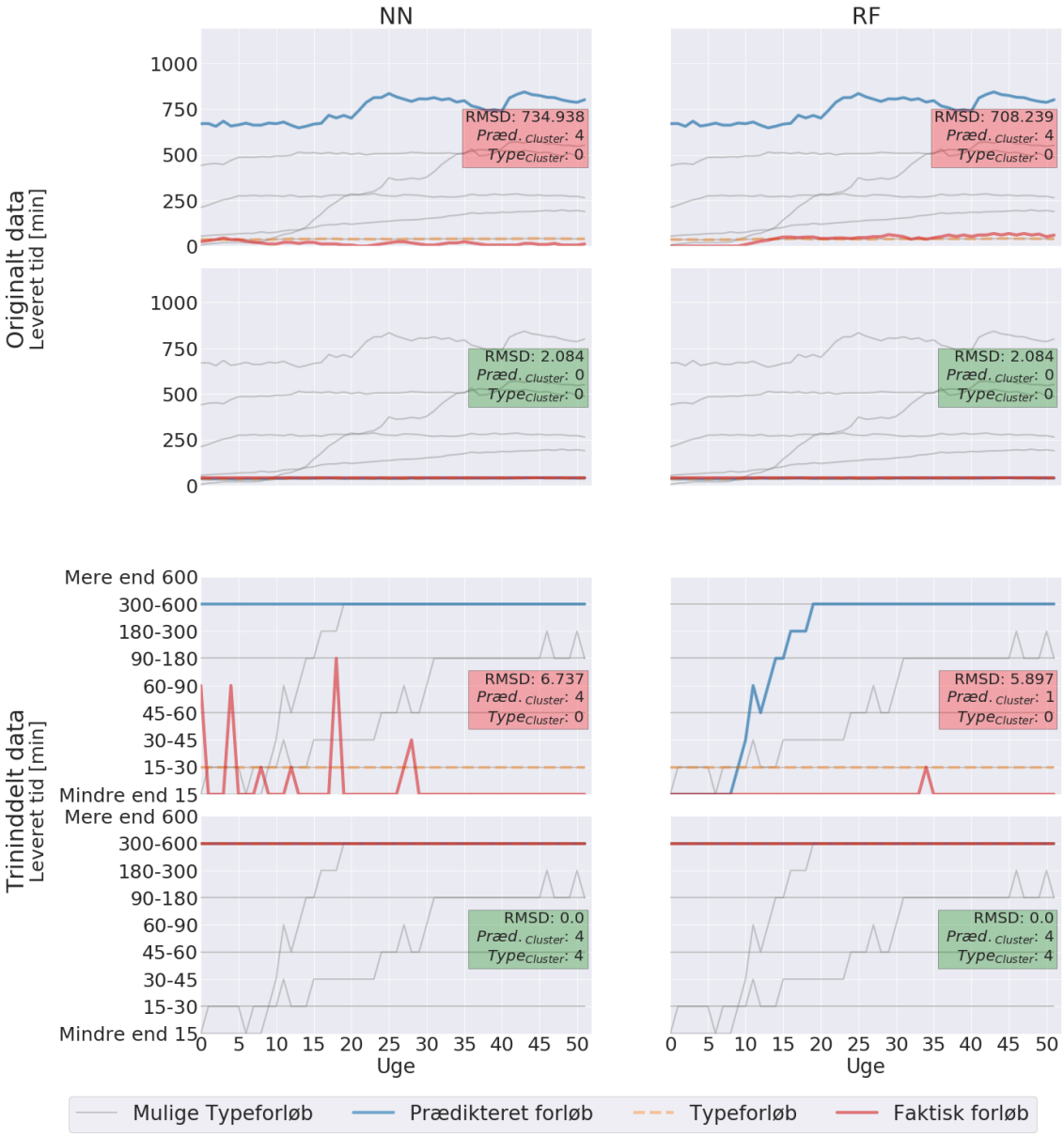
Figur 6.16. Precision-Recall kurver af hver cluster for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data. Alle kurver er ledsaget af Average Precision (AP). Hvert plot viser desuden macro-gennemsnitskurven (stiplede, pink linje).

resulterer i en lavere RMSD median score (57.7 min) end NN (70.6 min) på det originale data, men samtidig 20 outliers mod ni outliers i NN.

6.3 Feature importance

Feature importance evalueres ud fra det oprindelige træningssæt og giver derfor ikke indblik i generaliserbarheden af modellerne eller om de relevante features har lignende indflydelse for testsættet. Feature importance for alle de inkluderede features kan ses i appendiks D.

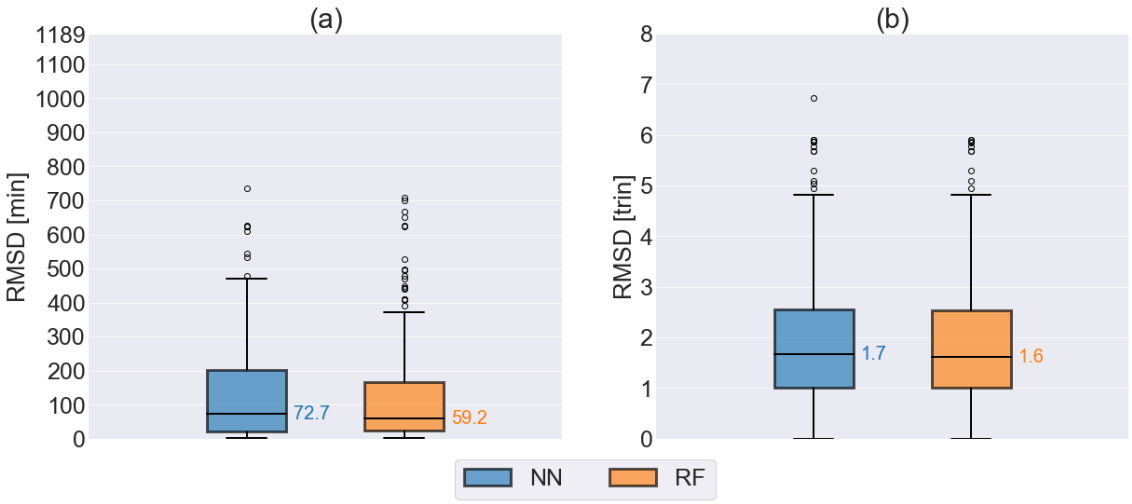
Som det fremgår fra permutation importance af de fem højest og lavest scorende features på **figur 6.19** er det særligt *Alder*, *Planlagt tid (minutter)* og *Bestilt tid (minutter)*, der har størst indflydelse på klassificeringen. Modsat viser figuren også at eksempelvis *Civil status_samboende* ($n = 1$), *Besøgstype_Akut besøg* ($n = 0$) og *Indsats (Niveau 04)_4.5 Hovedrengøring* ($n = 1$) har ingen eller lille indflydelse på klassificeringen. Dette



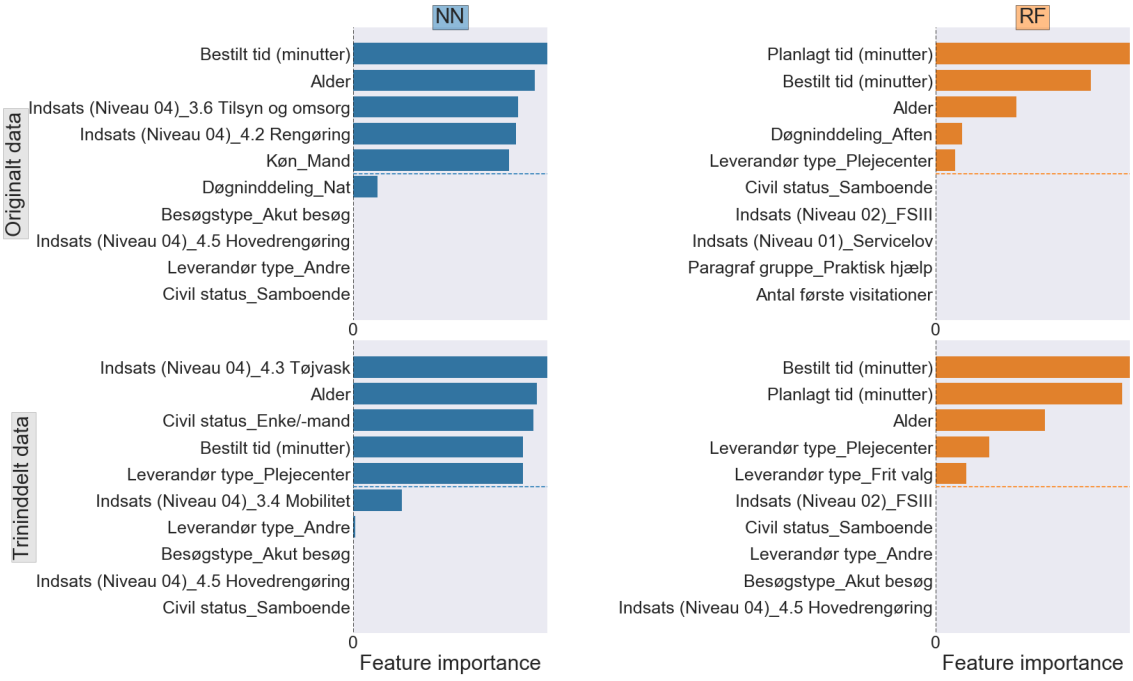
Figur 6.17. Tilfælde af laveste (rød) og højeste (grøn) Root Mean Squared Deviations (RMSD) for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data.

er forventeligt, da der er mindre end to borgere med hver af disse features i det inkluderede testsæt, hvilket ikke er tilstrækkeligt til at vise et mønster eller en sammenhæng.

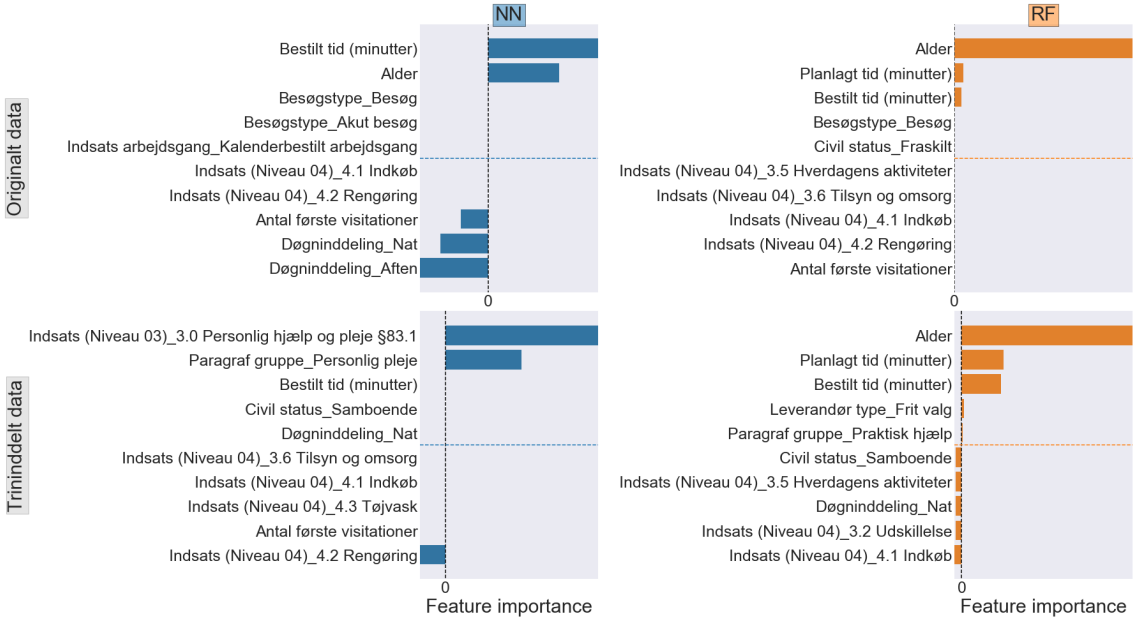
Drop column importance for de fem højest og lavest scorende features på **figur 6.20**, viser ligeledes at *Alder*, *Planlagt tid (minutter)* og *Bestilt tid (minutter)* er vigtige enkeltstående features for klassificeringen i begge RF-modeller og NN på det originale data. Kun NN på det trininddelte data viser ingen vigtighed i disse features og generelt ikke meget indflydelse fra de forskellige features. Ikke overraskende er der en sammenhæng mellem hjemmeplejetiden visiteret i første visitation og den prædikterede cluster, eftersom at de fleste clusters var stabile. Ligeledes synes det naturligt at alder er sammenhængende med hvor meget hjemmepleje en borger får.



Figur 6.18. Median Root Mean Squared Deviations(RMSD) for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) i antal minutter på originalt data (a) og i antal trin på trininddelt data (b).



Figur 6.19. De fem højeste og laveste permutation feature importances for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data. 'Feature importance' er relativ til differencen mellem højest og lavest scorende feature for hver model.



Figur 6.20. De fem højeste og laveste drop column feature importances for henholdsvis Neuralt Netværk (NN) og Random Forest (RF) på originalt og trininddelt data. 'Feature importance' er relativ til differencen mellem højest og lavest scorende feature for hver model.

Syntese 7

7.1 Diskussion

Baggrunden for nærværende studie var, at muliggøre målretning og prioritering af hjemmeplejeydelser gennem udnyttelse af EOJ-data til prædiktion af den fremtidige udvikling i hjemmeplejebehov for ældre borgere. Dette blev opnået gennem udvikling og implementering af en clustering-algoritme til identificering af plejebehovsfænotyper og deraf en række typeforløb for plejebehovet i hver fænotype. Desuden blev en klassifikationsmodel implementeret til estimering af hvilken fænotype borgeren tilhører, ud fra stamdata om borgeren og information om indsatserne visiteret ved første visitation. Typeforløbet for den pågældende fænotype blev dernæst anvendt som en approksimation på borgerens forventede udvikling i plejebehov det følgende år. Dermed er det videnskabelige bidrag af dette studie en ny data-drevet tilgang til estimering af hjemmeplejemodtageres forestående udvikling i plejebehov. Herved skabes et potentiale grundlag for målretning af plejeindsatser samt tidlig opsporing af borgere med potentiale for at gennemgå en uhensigtsmæssig udvikling i plejebehovet, ud fra deres prædikterede plejebehovsforløb.

Det implementerede proof of concept system formåede gennem KMC, at udlede seks plejebehovsfænotyper med unikke typeforløb. Disse forløb inkluderede fem omtrent konstante forløb, men på forskellige niveauer i plejebehov. Denne mangel på ændring i forløbet, kan anses som et udtryk for et hensigtsmæssigt plejeforløb hvor borgerens plejebehov fastholdes. Derudover identificeredes en fænotype med et lavt udgangspunkt i plejebehov, men en markant stigning efter omkring 10-15 uger. Denne udvikling kan dermed indikere et uhensigtsmæssigt behandlet plejebehov. Signifikant for denne fænotype var, at borgerne fik visiteret indsatser fra et plejecenter. Dette kan være en forklaring på stigningen i plejebehovet, idet det synes naturligt at en borgers plejebehov forventes at stige, hvis der visiteres ydelser på et plejecenter. Det er desuden muligt at denne type borgere omfatter kronikere eller demente, der som beskrevet er en stigende del af hjemmeplejemodtagere. For dette tilfælde kan en mulig forklaring på forløbets udvikling være borgere, der får visiteret enkeltstående ydelser eller forløb på specialiserede plejecentre, hvorefter deres tilstand bliver tiltagende plejekrævende.

Grundet ubalance i fordelingen af observationer mellem fænotyperne, blev dataen inddelt i brugerdefinerede trin. Dette resulterede i bedre balancering af data i fænotyperne, men med højere relativ RMSD mellem de faktiske forløb for borgerne i en fænotype og de respektive typeforløb. Denne trininddeling kan potentielt forbedre klassificeringen af borgere på bekostning af, at den grovere opløsning af forløbet vil have mere usikkerhed mellem de prædikterede typeforløb og det faktiske forløb, grundet bredden af trinene. Ydermere giver

trininddeling mere tolkbarhed i plejebehovsniveauet, da det anskues i relation til de øvrige intervaller på en plejebehovsskala.

For klassifikation af borgernes fænotyper ud fra features fra første visitation, var den maksimale opnåede ROC AUC 0.77 og PR AP på 0.43 givet ved den implementerede RF-model med trininddelt data. I denne forbindelse var alder, leverandørtypen og startniveauet for antal bestilte minutters hjemmehjælp udslagsgivende for klassifikationerne. Det er forventeligt at alder vil have indflydelse på hjemmeplejebehovet, mens leverandørtypen forekommer at være vigtig i relation til, om der er tale om et plejecenter eller en frit-valgs-leverandør. Desuden kan vigtigheden af det initierende niveau i visiteret pleje, forklares ud fra flertallet af stabile plejebehovsniveauer, da det initierende niveau dermed, i mange tilfælde, vil være svarende til typeforløbets niveau.

Ved gennemgang af litteraturen for afdækning af sammenlignelige metoder til estimering af hjemmeplejemodtageres kommende plejebehov blev det fundet, at ingen andre studier udnytter clustering til identificering af fænotyper for borgere med lignende tidslige udviklinger i hjemmeplejebehov. Derimod er det en udbredt metode at inddele borgere efter tærskelværdier anvendt på enkeltstående mål eller aggregerede værdier. Et eksempel herpå er Rosella et. al., der kategoriserer sundhedsmæssigt ressourcetunge borgere som værende i top 5 percentilet for den samlede udgift, ud fra et bestemt nedslag i tid [8]. Gennem aggregering mistes den information, der kan være indeholdt i den tidslige udvikling, hvilket kan være afgørende for at vurdere hvorvidt en borger har modtaget et optimalt plejeforløb. Dette begrundes med at negligering af udviklingen over tid ,vil betyde at borgere, der oplever en stigning i plejebehovet grupperes med borgere, der udviser et stabilt niveau. Ved anvendelse af den præsenterede metode, hvor den tidslige udvikling modelleres gennem KMC, overkommes denne udfordring, idet de resulterende fænotyper indeholder borgere med lignende longitudinale mønstre i plejebehov. En yderligere fordel herved er, at det lades være op til dataen at definere grupperingerne i dataen, hvorved hidtil ukendte, latente fænotyper potentielt kan udledes. Dette var tilfældet med den uhensigtsmæssige fænotype i dette studie, der ikke ville kunne udledes med simple tærskelværdier.

Ydermere indeholder litteraturen ikke eksempler på, at unsupervised forløbsclustering af borgernes udvikling i plejebehov anvendes som udgangspunkt for prædiktion af det fremtidige hjemmeplejebehov. Dette skyldes først og fremmest at machine learning modeller af denne art, ikke er udbredt til applikationer på ældre- og omsorgsområdet. Dette på trods af den store mængde data, der genereres i EOJ-systemer. Desuden anvendes aktuelle machine learning modeller på tilsvarende områder, typisk i situationer, hvor et effektmål eller kendte klasser guider klassifikationen. Da registrering af effektmål for hjemmeplejeindsatser endnu ikke er en integreret del af de eksisterende arbejdsprocesser, fremsætter proof of concept modellen i dette projekt, dermed en alternativ metode til identificering af klasser til brug i prædiktiv klassifikationssammenhæng i hjemmeplejeregi.

Det kan diskuteres hvorvidt anvendelsen af antal leverede minutters hjemmepleje som et pseudomål for en borgers plejebehov, er en valid tilgang. Foruden borgerens plejebehov er den leverede hjemmeplejetid produktet af en række yderligere faktorer, herunder serviceniveauet fastlagt for den visiterende myndighed, frekvensen på opfølgninger og midlertidige hændelser i borgerens liv, der ikke er relateret til ændring i det reelle plejebehov. Det er ikke muligt for den implementerede clustering-algoritme at adskille disse

yderligere faktorer fra borgerens reelle plejebehov, idet der ikke er information tilgængelig herfor. Denne begrænsning kunne afhjælpes ved at danne fænotyper på baggrund af et mere akkurat mål for borgerens plejebehov, såsom FSIIIs tilstandsvurderinger, såfremt arbejdsprocesser på dette område var modne til at generere valid data. Alternativt kan der inkluderes yderligere information i clustering-algoritmen, for at give mere information relateret til plejebehovet. Eksempelvis kunne aktivitetsdataen kombineres med sundhedsdata fra EPJ-systemer. Et eksempel på relevant sundhedsinformation er features relateret til demens og andre kroniske sygdomme, der vides at have indflydelse på plejebehovet. Desuden ville performance af klassifikationen formentlig kunne øges ved at inkluderer features der mere direkte repræsenterer borgerens sundhed og selvstændighed.

En anden begrænsning ved det implementerede proof of concept, er datasættets størrelse og udformning. Den udførte clustering til identificering af fænotyper påvirkes betydeligt af, at datasættet indeholder en overvægt af lave værdier for leverede minutters hjemmehjælp. Dette resulterer i en majoritetsgruppe med forløb i den lave del af dataspændet, og dermed en ubalance i antallet af borgere i hver fænotype. Dette er til dels afhjulpet gennem trininddeling af datasættet med finere opløsning for de lave værdier og grovere opløsning for de højere værdier. Desuden forsøges ubalancen udlignet gennem oversampling i klassifikationsmodellen, men betyder endog, at de øvrige fænotyper er baseret på få observationer. For at styrke validiteten og generaliserbarheden af de udledte fænotyper kræves dermed et større datasæt. Størrelsen på datasættet er ligeledes en udfordring for de implementerede neurale netværk, da en stor datamænge kræves for at træne denne type modeller til at lære variansen i dataen. Særligt ses dette ved det trininddelte data hvor det neurale netværk lærer meget lidt fra de inkluderede features og generelt klassificerer tilnærmelsesvis tilfældigt. Ydermere betyder datasættets størrelse, at testsættet blot indeholder 267 observationer. Når dette kombineres med at majoritetsgruppen udgør henholdsvis 61.02% og 32.08% for det originale og trininddelte datasæt, kan der argumenteres for, at testsættet ikke er af tilstrækkelig størrelse til at opnå statistisk styrke og formå at repræsentere det virkelige databillede akkurat. Dermed kan størrelsen på testsættet inducere et skævt billede af klassifikations-performance.

I den anvendte Bayesian optimization er kriteriefunktionen f1-scoren udregnet på det fulde træningssæt. Dette betyder, at denne kan være kunstigt høj, hvis netværket har overfittet til træningsdataen. Det er dog taget højde for dette ved at blande træningssættet og applicere et nyt træning-/valideringssplit ved hver epoch. Alternativt kunne loss eller f1 anvendes fra den epoch som træningen for det testede netværk er stoppet på. Denne kan dog være baseret på et heldigt split i dataen der giver specielt gode resultater for lige det data. Dette kunne løses ved at anvende et gennemsnit over de fem omkringliggende epochs og på den måde imitere en five-fold cross validering, hvor der trænes og valideres på et nyt 80/20 split for hver targetværdi.

Bortset fra de beskrevne svagheder demonstreres en potentiel metode til udledning af plejebehovsfænotyper samt estimering af en hjemmeplejemodtagers fremtidige udvikling i plejebehov som funktion af første hjemmeplejevisitation. Den praktiske anvendelighed af modellen er særligt betinget af evnen til at prædiktere, om en borger forventes at følge en uhensigtsmæssig udvikling i plejebehovet. I denne forbindelse forventes det, at validiteten for det prædikterede forløb kan øges gennem inkludering af features, der

reflekterer borgerens tilstand, eksempelvis FSIII tilstandsvurderinger eller lignende. Dette vil dog kræve konsistent og ensartet registrering heraf.

7.2 Konklusion

På baggrund af de igangværende forandringer på ældreområdet, der resulterer i stigende variationen i behovet for hjemmepleje, og udviklingen heraf, er der afdækket et behov for et redskab til fænotypeinddeling af udviklingen i plejebehov. Rationalet bag dette postulat er, at et sådant redskab vil kunne bidrage til udredning af det differentierede billede af forskellige plejebehov blandt ældre. Hermed muliggøres potentielt målretning af indsatser samt forebyggelse af progression til et unødvendigt højt plejebehovsniveau. Dette kan potentielt være nyttigt i relation til den aktuelle udfordring i den kommunale hjemmepleje, hvor en stigende opgavemængde skal varetages på trods af krav om at begrænse udgifter. Ud fra dette identificerede behov blev følgende problemformulering opstillet:

Hvordan kan data fra EOJ-systemer anvendes til udarbejdelse af et værktøj til identificering af fænotypeforløb for borgeres fremtidige, longitudinale hjemmeplejebehov samt prædiktion heraf, ud fra første visitation.

Med afsæt i denne problemformulering er der implementeret et proof of concept system, der inddeler hjemmeplejemodtagerne i fænotyper på baggrund af forløbet i deres plejebehov. Desuden opnår det implementerede system at kunne klassificere borgere i deres respektive fænotype ud fra stamdata og viden om deres første visitation. På baggrund af denne klassifikation approksimeres borgerens fremtidige udvikling i plejebehov ved typeforløbet for den tildelte fænotype.

Særligt fremhæves RF-modellen på det trininddelte datasæt, der blev clusteret med KMC. Denne opsætningen gav den bedste ROC AUC på 0.77 og PR AP på 0.43. Dette på trods af, at de prædikterede trinindelte typeforløb afveg mere fra de faktiske forløb end i det originale datasæt, med $\text{RMSD}_{RF_trin} = 1.6$. Modellen identificerede desuden særligt alder og den initierende visiterede plejetid som vigtige features til klassifikationen. Yderligere fandt denne model at leverandørtypen var en udslagsgivende feature for klassificeringen.

Det kan i kraft af det udviklede system konkluderes, at data fra EOJ-systemer kan anvendes som potentielt grundlag for en data-drevet tilgang til inddeling af fænotyper, gennem KMC, ud fra kvantitativt hjemmeplejebehov. Desuden er brug af en RF- eller NN-klassifikationsmodel på baggrund af data fra borgerens første visitation en mulig metode til estimering af det fremtidige plejebehov, men bør undersøges yderligere med større datasæt.

Litteratur

[1] Vinge S, Kramhøft K, Bøgh A. Hjemmesygeplejen i Danmark. DSI Institut for Sundhedsvæsen; 2006.

[2] Kommunernes Landsforening. Styring af hjemmesygeplejen - Inspirationsnotat. Kommunernes Landsforening; 2014.

[3] Hagestrøm J, Rasmussen HB. Det sammenhængende borger- og patientforløb. 1st ed. 1. Munksgaard; 2017.

[4] Kommunernes Landsforening. Dokumentation og styring af hjemmesygepleje - Baggrundsnotat. Kommunernes Landsforening; 2010.

[5] Hu Z, Hao S, Jin B, Shin AY, Zhu C, Huang M, et al. Online prediction of health care utilization in the next six months based on electronic health record information: a cohort and validation study. Journal of medical Internet research. 2015;17(9):e219.

[6] Sundheds- og Ældreudvalget 2017-18. Redegørelse om ældreområdet. Sundheds- og Ældreministeriet; 2018.

[7] Hjemmehjælpskomissionen. Fremtidens hjemmehjælp - Ældres ressourcer i centrum for en sammenhængende indsats. Socialstyrelsen; 2013.

[8] Rosella LC, Kornas K, Yao Z, Manuel DG, Bornbaum C, Fransoo R, et al. Predicting high health care resource utilization in a single-payer public health care system: development and validation of the high resource user population risk tool. Medical care. 2018;56(10):e61.

[9] Danmarks Statistik. Markant flere ældre i fremtiden; 2018. Available from: `https://www.dst.dk/da/Statistik/nyt/NytHtml?cid=26827`.

[10] Christensen AI, Ekholm O, Davidsen M, Juel K. Sundhed og sygelighed i Danmark 2010 og udviklingen siden 1987. Statens Institut for Folkesundhed, Syddansk Universitet;.

[11] Danske Love. Serviceloven §83;. Acessed: 20.02.2020. Available from: `https://danskelove.dk/serviceloven/83`.

[12] Danske Love. Sundhedsloven §138;. Acessed: 20.02.2020. Available from: `https://danskelove.dk/serviceloven/83`.

[13] Ankestyrelsen. Frit valg i ældreplejen - erfaringer fra landet kommuner. Ankestyrelsens Undersøgelser; 2010.

[14] McKinsey & Company. Organisering og effektivitet på hjemmehælpsområdet - fokus på personlig pleje og praktisk hjælp. KL, Finansministeriet og Socialministeriet; 2006.

[15] Ministeriet for Sundhed og Forebyggelse. Nationalt kvalitetsprogram for sundhedsområdet 2015-2018. Ministeriet for Sundhed og Forebyggelse; 2015.

[16] Poulsen MV. Værd at vide om Fælles sprog – Kom videre med Fælles sprog; 2002.

[17] Poulsen MV. Fælles sprog II - Dokumentation af afgørelser på det kommunale ældre- og handicapområde; 2004.

[18] Fælles Sprog III Metodehåndbog; 2019.

[19] KL. Fælles Sprog III - Indsatskatalog (sundhedslov); 2019.

[20] KL. Fælles Sprog III - Indsatskatalog (service); 2019.

[21] Kommunernes Landsforening. Kortlægning af kommunernes anvendelse af sundheds- og ældredata. Kommunernes Landsforening; 2019.

[22] Zhang J, Xiong H, Huang Y, Wu H, Leach K, Barnes LE. M-SEQ: Early detection of anxiety and depression via temporal orders of diagnoses in electronic health data. In: 2015 IEEE International Conference on Big Data (Big Data). IEEE; 2015. p. 2569–2577.

[23] Sen C, Hartvigsen T, Kong X, Rundensteiner E. Learning Temporal Relevance in Longitudinal Medical Notes. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE; 2019. p. 2474–2483.

[24] Esteban C, Schmidt D, Krompaß D, Tresp V. Predicting Sequences of Clinical Events by using a Personalized Temporal Latent Embedding Model. In: 2015 International Conference on Healthcare Informatics. IEEE; 2015. p. 130–139.

[25] Baghaei KT, Rahimi S. Sepsis Prediction: An Attention-Based Interpretable Approach. In: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE; 2019. p. 1–6.

[26] Last M, Carel R, Barak D. Utilization of data-mining techniques for evaluation of patterns of asthma drugs use by ambulatory patients in a large health maintenance organization. In: Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007). IEEE; 2007. p. 169–174.

[27] Kaushik S, Choudhury A, Dasgupta N, Natarajan S, Pickett LA, Dutt V. Using LSTMs for Predicting Patient's Expenditure on Medications. In: 2017 International Conference on Machine Learning and Data Science (MLDS). IEEE; 2017. p. 120–127.

[28] Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry. 1964 Jul;36(8):1627–1639.

[29] Everitt BS, Landau S, Leese M, Stahl D. Cluster Analysis. 5th ed. Balding DJ, Cressie NAC, Fitzmaurice GM, Goldstein H, Molenberghs G, Scott DW, et al., editors. Probability and Statistics. John Wiley & Sons, Ltd; 2011.

[30] Duda RO, Hart PE, Stork DG. Pattern Classification. Wiley; 2001.

[31] Maaten Lvd, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008;9(Nov):2579–2605.

[32] Joe H Ward J. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association. 1963 Mar;58(301):236–244.

[33] Arthur D, Vassilvitskii S. k-means++: The Advantages of Careful Seeding. In: Bansal N, Pruhs K, Stein C, editors. SODA. SIAM; 2007. p. 1027–1035.

[34] Thorndike RL. Who Belongs In The Family? Psychometrika. 1953 Dec;18(4):267–276.

[35] Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics. 1987;20:53–65.

[36] Davies DL, Bouldin DW. A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1979 Apr;PAMI-1(2):224–227.

[37] Snoek J, Larochelle H, Adams RP. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. arXiv. 2012;1206.2944[stat.ML].

[38] Rasmussen CE, Williams CKI, Processes G, Press MIT, Jordan MI. Gaussian Processes for Machine Learning; 2006.

[39] Nogueira FMF. BayesianOptimization; 2014. GitHub.

[40] Breiman L. Random Forests. Machine Learning. 2001;45(1):5–32.

[41] Adnan N, Islam Z. One-Vs-All Binarization Technique in the Context of Random Forest. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2015;p. 385–390.

[42] Ahire JB. The Artificial Neural Networks Handbook: Part 4; 2018. Blog post.

[43] Goodfellow I, Bengio Y, Courville A. Variational Autoencoders. In: Deep learning book. MIT Press; 2016. p. 693–696.

[44] Bengio Y. Practical Recommendations for Gradient-Based Training of Deep Architectures. CoRR. 2012 Sep;arXiv:1206.5533(2).

[45] Maiza MA. The Unknown Benefits of using a Soft-F1 Loss in Classification Systems; 2019. Towards Data Science - A Medium publication sharing concepts, ideas, and codes.

[46] Clevert DA, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). arXiv preprint arXiv:151107289. 2015;.

[47] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. CoRR. 2015 Mar;arXiv:1502.03167(3).

[48] Ramachandran P, Zoph B, Le QV. Searching for Activation Functions. CoRR. 2017 Oct;arXiv:1710.05941.

[49] Bridle J. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. Neurocomputing NATO ASI Series (Series F: Computer and Systems Sciences). 1990;68.

[50] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overtting. Journal of Machine Learning Research. 2014;15:1929–1958.

[51] Zar JH. Biostatistical Analysis (5th Edition). Prentice-Hall, Inc.; 2007.

[52] Davis J, Goadrich M. The Relationship Between Precision-Recall and ROC Curves. In: Proceedings of the 23rd international conference on Machine learning; 2006. p. 233–240.

[53] Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10):1340–1347.

[54] Moher D, Liberati A, Tetzlaff J, Altman D. Items for Systematic Reviews and Meta-Analysis: The PRISMA statement; 2009.

[55] Malterud K. Qualitative research: Standards, challenges, and guidelines. The Lancet. 2001;358(9280):483,488.

Struktureret Litteratursøgning A

For at indhente et evidensgrundlag for løsningsmulighederne i relation til problemet, se afsnit 3, er der udført en struktureret litteratursøgning. Denne søgning havde til formål, at identificere studier, der beskæftiger sig med metoder til prædiktering af fremtidige outcomes baseret på longitudinal og multidimensional sundhedsdata. Relevante søgetermer blev identificeret ud fra den initierende problemstilling samt gennem en ustruktureret litteratursøgning. Disse søgetermer inddeltes efterfølgende i søgeblokkene: Home Care Data, Longitudinal, Machine Learning, Prediction, Care Demand. I hver blok blev synonymer og lignende begreber inkluderet. Disse inkluderedes i søgstrengen med OR-operatoren, mens hver blok blev sammenholdt med AND-operatoren. De identificerede søgeblokke og -termer kan ses i **tabel A.1**.

Tabel A.1. Søgeblokke og tilhørende søgetermer identificeret i forbindelse med den strukturerede litteratursøgning.

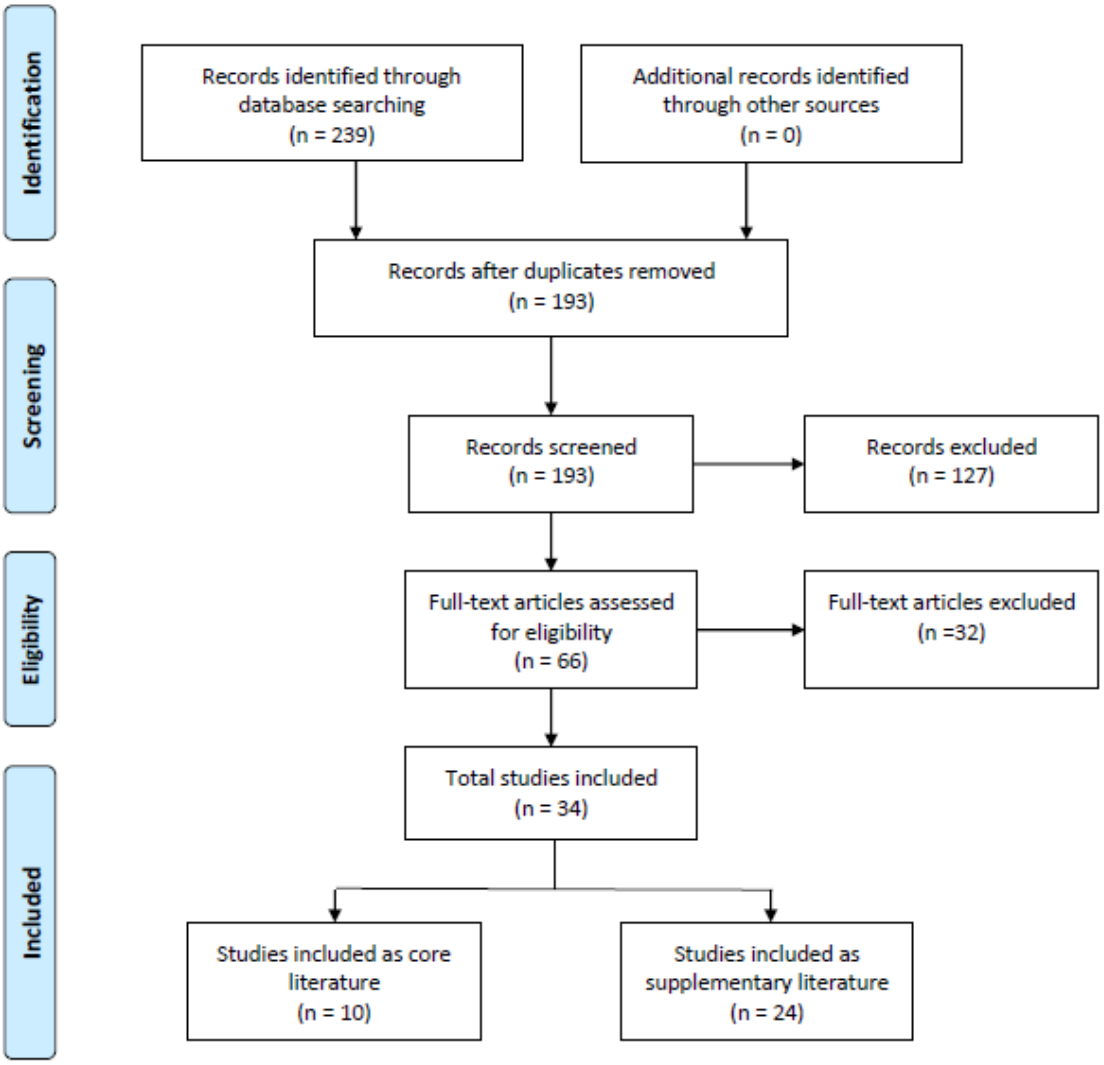
Home Care Data	**Longitudinal**	**Machine Learning**
health care data	temporal	deep learning
electronic health record	panel	artificial intelligence
electronic medical record	time series	AI
electronic care record		neural network
electronic social care record		cluster*
health data		classification
administrative health data		classifier
		algorithm*
		data-driven
		data mining
		regress*
Prediction	**Care Demand**	
predicting	health care utilization	
estimat*	home care utilization	
modelling	resource utilization	
forecasting	costly patients	
forecast	health profile	
identify*	care profile	
risk stratification	care trend	
	care trends	
	care sequence	
	care sequences	

Den resulterende søgestreng blev anvendt i følgende videnskabelige databaser: Pubmed,

Embase, Web of Science, Scopus og IEEE. Disse databaser er udvalgt, da de tilsammen udgør et bredt grundlag i tværfeltet mellem sundhedsvidenskab og ingeniørfag. I alt resulterede søgningen i 239 artikler, hvilket blev reduceret til 193 efter fjernelse af duplikater. Disse artikler blev screenet på titel og abstract, for hurtigt at frasortere de der opfyldte de opstillede eksklusionskriter. Artikler blev fjernet såfremt de ikke indeholdt et prædiktivt aspekt, eksempelvis retrospektive kohort-studier, der undersøgte korrelationen mellem demografiske faktorer og et givent outcome. Endvidere frasorteredes studier med fokus på modeller til 'event'-detektion ud fra retrospektiv, longitudinal data. Sidstnævnte eksklusionskriterie skyldes, at den type metoder oftest blot var baseret på at undersøge, hvorvidt en dataentitet afviger fra tidligere data. Endelig blev studier ekskluderet, hvis der var tale om en visuel analyse af dataen.

Ud fra de beskrevne eksklusionskriterier blev 127 referencer frasorteret i screeningsfasen, hvortil 32 af de resterende 66 artikler blev ekskluderet efter fuldtekst-læsning. Da relevansen af de tilbageværende 34 artikler var varierende, blev disse inddelt i kategorierne, kernelitteratur (10 referencer) og supplerende (24 referencer). Processen for udførelsen af den strukturerede litteratursøgning og antallet af referencer i de respektive faser er illustreret gennem PRISMA-diagrammet i figur A.1.

Den strukturerede litteratursøgning er suppleret med yderligere litteratur identificeret gennem referencelisten af de inkluderede studier. I tilfælde hvor et specifikt område måtte undersøges i dybden, blev litteratur indhentet gennem målrettede søgninger. Ydermere er litteratursøgningen suppleret med grå litteratur, såsom rapporter udgivet af offentlige organer.



Figur A.1. PRISMA-diagram over processen for, og antallet af referencer gennem, de forskellige stadier i den strukturerede litteratursøgning udført i forbindelse med nærværende projekt. Modificeret fra [54].

Interview m. Faaborg-Midtfyn Kommune

B

B.1 Formål

Formålet med det udførte interview var at afdække hvorledes en model trænet på data fra EOJ-systemet kan skabe mest mulig værdi for det praktiske arbejde på ældre- og omsorgsområdet. Dette har afsæt i, at det identificerede problemfelt peger på behov for redskaber til styring og prioritering i visitationen af omsorgs- og sygeplejeydelser, men uden at specificere sådanne værktøjer yderligere. Dette søges uddybet og konkretiseret gennem et interview med FMK, der foruden deltagelse i interview stiller ældre- og omsorgsdata til rådighed i projektet.

B.2 Metode

Nærværende studie gør brug af et semi-struktureret interview. Dette indebærer, at interviewet tager udgangspunkt i en interviewguide, men uden specifikke svarmuligheder formuleret på forhånd. Desuden tillades afvigelse fra interview-guiden, for at understøtte opfølgende spørgsmål. Dette metodevalg er foretaget på baggrund af, at et struktureret interview findes for rigidt, da interviewet skal fungere i en eksplorativ kontekst, som en del af problemanalysen til indsnævring af fokus for udvikling af et *proof-of-concept*-værktøj. En interview-guide findes nødvendig, for at sikre tilstrækkelig afdækning af emnet.

Forud for interviewets påbegyndelse er der indhentet tilladelse til optagelse af samtalen.

I evalueringen af resultaterne fra interviewet og validiteten heraf, holdes det for øje, at kvalitative metoder påvirkes af en række faktorer. Dette omfatter blandt andet subjektivitet hos respondenten, den effekt undersøgeren har på interviewet og fortolkningen resultaterne. Subjektivitet hos respondenten kan ikke undgås og vil med sikkerhed have indflydelse på resultaterne. Herfor skal valg af respondenter og spørgsmålene hertil, afpasses til de research-spørgsmål, der ønskes undersøgt. [55]. Valget af respondenter er baseret på at danne et tværfagligt felt med roller, der har indsigt i både den faglige praksis og anvendelsen af data og teknologi i organisationen.

B.3 Transskription

Interviewet blev foretaget virtuelt d. 16.04.2020 kl. 15.00 med en varighed på ca. 60 min. Mathilde Gadensgaard agerer interviewer på vegne af projektgruppen. Alle informanter har givet mundtlig tilladelse til optagelse og anvendelse af samtalen. Nedenfor følger information om informanterne samt en ordret og komplet transskription af interviewet.

Navn:	Allan Christiansen
Stilling:	Souschef for Sundhed og Omsorg, FMK
Køn:	Mand
Annotation i transkription:	A

Navn:	Kathrine Skov Kastrup
Stilling:	Teknologikonsulent, FMK
Køn:	Kvinde
Annotation i transkription:	K

Navn:	Jeppe Rosager Lund Pedersen
Stilling:	Konsulent for ledelsesinformation, FMK
Køn:	Mand
Annotation i transkription:	JP

Navn:	Jeppe Svendsen
Stilling:	Konsulent, KMD
Køn:	Mand
Annotation i transkription:	JS

I: *Hvis vi lige tager en runde her, hvor I måske lige præsenterer jer selv i forhold til jeres stilling, og hvad det er for nogle ansvarsområder I sidder med i den forbindelse.*

K: Allan starter.

A: Jeg starter, okay. Jeg hedder Allan og jeg er souschef i Sundhed og Omsorgsforvaltningen i Faaborg-Midtfyn Kommune. Det gælder alt, der hedder det traditionelle ældreområde; det vil sige kostområdet og plejehjem, hjemmepleje, sygepleje, træning, genoptræning, sundhedsområdet. Det er det primære dér, og så har vi hele det specialiserede socialområde og handicapområde med alt det traditionelle, det socialfaglige område og handicap og psykiatri.

K: Du sidder også med økonomi ikke Allan?

A: Nåå jojo, altså jeg kan godt fortælle, at vi har et budget på en milliard og et underskud på 35 millioner, hvis det er interessant. Vi forsøger at styre økonomien blandt andet sammen med Jeppe, der er informations.. øh koordinatior. Hvad kalder du dig Jeppe?

JP: Ja, jeg ved faktisk ikke, hvad titlen er.

I: *Okay, så det vil også sige det er dig, Allan, der er styrende for at lave prioriteringer af serviceniveau og så videre, i Faaborg-Midtfyn Kommune?*

A: Altså vi har jo et... det jeg nævnte før, det er sådan driftsniveau, og så har vi et center som, blandt andet, vores myndighed ligger i og det er én myndighed, der visiterer samtlige ydelser inden for det felt, som jeg nævnte før. Der er det egentlig vores myndighedschef der har hovedansvaret for at udarbejde kvalitetsstandarder og sådan nogle ting. Men typisk så har vi i det sekretariat, hvor jeg så er daglig leder af, der har vi nogle konsulenter, blandt andet der hvor Kathrine også sidder, og så er der også nogle lidt mere fag-faglige konsulenter, der har hænder på de udarbejdelser af kvalitetsstandarderne. Og så har jeg ligesom ansvaret over for vores øverste chef i Sundhed og Omsorg i at kvalitetsstandarderne bliver udarbejdet rigtigt, og at de bliver politisk behandlet og sådan nogle ting der. Så i sidste ende er det jo selvfølgelig et politisk serviceniveau vi har. Men altså, på den måde er jeg jo inde over kvalitetsstandarderne. Men sådan det praktiske arbejde, der ligger hos jo myndigheden, men stadigvæk hvor Kathrine og jeg, tilsammen, har et stor indblik i hvordan der arbejdes med det qua vores.. altså min rigtig lange erfaring og Kathrines lidt mindre, men stadigvæk indgående erfaring i det. For ikke at komme galt afsted, Kathrine.

K: Nej, det var pænt sagt.

A: Så vi er godt dækket ind i forhold til det arbejde der ligges der.

I: *Og Kathrine hvis du lige vil uddybe, sådan hvad det er præcis, du sidder og arbejder med?*

K: Ja, jeg sidder jo så i det sekretariat, der er i centeret Sundhed og Omsorg, og jeg er ansat som konsulent, og så er jeg også teamkoordinator for vores systemadministrator-team for Nexus-systemet. Så jeg arbejder meget med projektledelse, selvfølgelig, men har også faciliteret et samarbejde over imod ledelsesinformationskonsulenten, Jeppe, i forhold til, hvordan arbejder vi med analyser, og hvordan får vi måske oparbejdet lidt bedre dokumentation af vores omsorgssystem. Vi er i gang med at lancere et stort servicetjek af hele Nexus, for egentlig at få nogle bedre data og bruge data lidt mere intelligent end bare til antal og sådan af de lidt mere lineære ting. Så det sidder jeg egentlig og bruger, eller kommer til at bruge, rigtig meget tid på.

I: *Ja, når I snakker om at bruge data anderledes, hvad har i tankerne eller i støbeskeen der?*

K: Altså det er meget ny endnu, kan man sige, men det som Jeppe og jeg har talt om - og sammen med Allan, selvfølgelig - det er egentlig sådan at sige, alt efter hvilket niveau ledelse, hvad er det så for en information, der er vigtig at få, i forhold til at styre både økonomi, men også sådan strategisk, kvalitetsmæssigt, udviklingsmæssigt. En almindelig daglig leder af en hjemmeplejegruppe har måske ikke brug for at se så meget på effekt, men mere på, om de leverer det de skal og overholder deres tid og kvalitetsstandarder og alt sådan noget. Men når det så kommer højere op i ledelseslaget, op imod hvor Allan jo så sidder, så mere det her med at se på kvalitet af, "nå men vi giver et digi-rehab-foløb...- og måske holdt op imod noget virtuel skærmbesøg - "får de så generelt mindre hjælp end dem som får almindelig besøg og ikke træner?". Altså sådan noget. Nu er jeg ikke super, det er mere KMD-Jeppe der er god til sådan noget. Men altså hvor man kombinerer dataen i langt højere grad, end vi har haft tradition for, for at kunne se på effekten. Nu arbejder jeg også med velfærdsteknologi-projekter, og der kigger man jo meget på effekt, og det er ikke mit indtryk, at man har gjort det så meget i forhold til den service, vi leverer til borgerne på mange andre områder.

I: *Ja, fordi det er jo faktisk også noget af det, vi har kigget lidt på her i forbindelse med at se på, hvad det er for nogle nye redskaber man kan begynde at kigge hen imod. Det her*

med 'hvad skal man egentlig bruge som mål for effekt, som det er nu'. Så hvad arbejder i med nu som mål for effekt?

K: Jamen det kan være hvis Jeppe Ros lige når at præsentere sig, for han sidder med alle de der rapporter nu.

I: *Ja*

JP: Ja. Jamen jeg sidder i den centrale økonomiforvaltning på rådhuset i Ringe. Så jeg er ikke på samme lokation som Allan og Kathrine, men jeg arbejder med ledelsesinformationen for netop Sundhed og Omsorg, og det har jeg gjort siden maj sidste år og inden da, der var jeg almindelig økonomikonsulent for området. Så jeg har også... ja det var jeg i fire år inden da, så jeg har også en nogenlunde historik i forhold til området. Og har set det både fra den økonomiske vinkel, og nu ser jeg det mere fra den statistiske og datamæssige vinkel.

I: *Ja, er det også data fra Nexus du trækker ud der?*

JP: Det er det ja.

I: *Er det i samarbejde med Jeppe (KMD) der også?*

JP: Nej, det gør vi selv. Vi står selv for dataudtræk. Vi har købt den data-mart som KMD stiller til rådighed og sælger til kommunerne, hvis de har lyst, og det er så derfra, vi trækker vores data. Så det er ikke KMD, der trækker det for os, men de stiller den så til rådighed og pakker den. Og så har vi så tilkøbt data.

K: Må jeg tilføje noget? Sundhed og Omsorg bruger selv Analyseportalen; det er sådan vores primære, når vi får analyser. Men så snart der er noget, der er to debitioner på, så ryger det altid over til Jeppe (FMK), og så er det ham, der sidder og regner på det. Så vi har ligesom sådan en arbejdsgang i forhold til arbejdet med analyser.

I: *Okay, men det vil sige I sidder selv og laver udtrækkene fra Analyseportalen. Når I snakker om det her service-eftersyn med Nexus, er det så også i forhold til den data I fremstiller der?*

K: Yes. Vi har rigtig meget roderi - for mange døde borgere, ting der ikke er afsluttede og sådan noget, så det er også et håb om at fjerne alt det, der ligger og fylder og roder. Så vi får et bedre billede.

I: *Sådan lidt i den forbindelse, så er der jo også hele den her Fælles Sprog III dokumentationsstandard. Hvor man snakker lidt om, at der er forskellige fortolkninger heraf. Hvordan har I arbejdet med at indføre FSIII?*

K: Skal jeg sige noget Allan, eller vil du?

A: Du må godt, så kan jeg supplere.

K: Altså jeg var ikke med fra starten, da vi implementerede, men vi var første bølge fra KL, og vi var også første bølge for KMDs Nexus. Det er også derfor, vi skal have et service-tjek nu, fordi det var noget skrammel. Men man kan sige, at der blev den jo fortolket. Der blev uddannet de her instruktører og så fortolkede de jo metoden og tegnede den og så byggede vi den jo sådan helt selv, fordi KMDs tale dengang var jo, at vi kunne få det, ligesom vi ville have det. Det er jo det mange kommuner sådan lidt har taget med sig, som selvfølgelig er blevet vendt om nu, for vi alle sammen blevet klogere, men man arbejder meget ud fra, 'hvordan plejer vi at gøre, og hvordan matcher det op imod FSIII, og hvordan synes vi så, det skal passe ind i vores kontekst'. Så derfor oplever vi ofte, at vi ikke sådan kan sparre og dele opsætninger på tværs af kommuner, fordi vi har vores egen FSIII-opsætning. Vi oplever også, at det er rigtig, rigtig svært for medarbejderne at dokumentere efter det, og det kommer jo til syne i datakvaliteten også. Vil du supplere Allan?

A: Nej, altså det er jo det, og det rammer meget fint Kathrine.
K: Altså vi lever meget op til den business-case, der er lavet fra KL, at de penge man troede man ville spare på FSIII, det har man ikke kunne realisere. Det har givet mere dokumentationstid, og det har det også for os.
I: *og hele det her håb om, at man får mere strukturerede og ensartede data, er måske ikke noget I ville nikke genkendende til, eller hvordan?*
K: Nej, og det synes jeg var et rigtig fint billede på, da vi havde Kirsten fra KMD ude og observere i vores telefonvagt, fordi at selvom de havde samme funktioner og arbejdsbaggrunde, så dokumenterede de forskelligt, fordi det kan du jo gøre i et system. Så selvom FSIII er en metode, og man bør kende den, så er der jo mulighed for at gøre tingene på en anden måde og gå en anden vej rundt om.
I: *Okay, så hele den ekstra dokumentationstid der er kommet til i kraft af FSIII, er det fordi man skal ind og lave de her tilstandsvurderinger, eller hvor er det lige præcis den ekstra tid ligger, vil I mene?*
K: Ja, det er det jo nok, og så kan man sige, at alt der er svært, det tager jo også længere tid, hvis man ikke har en rutine i det. Og så har der været klager over mange klik i systemet og det er jo fordi der mange tilstande og man skal bevillige aftale med ydelser. Altså der er mange arbejdsgange i noget.
I: *Okay, og så snakkede vi lidt om det her med at det ikke rigtigt er blevet nemmere at sammenligne sig mellem kommuner. Hvad med inden for jeres egen organisation, har det giver mere struktur i dataen der?*
K: Altså vi oplever, at selvom man er under samme hat, måske træning, så har man forskellig dokumentationspraksis fra lokation til lokation. Så det vil ikke sige, det kan også være Jeppe kan se lidt et billede. Der er selvfølgelig også nogle steder, der er mere eksemplariske. Hjemmeplejen er sgu nok meget gode, men plejehjemmene de gør det også lidt forskelligt fra sted til sted. Så det er ikke det, jeg oplever.
A: Nå men så kan jeg i hvert fald tillade mig at sige noget, som også i forhold til jeres studier, når vi snakker om det. Fordi det er rigtig nok at målet også blandt andet var mere struktureret dokumentation, så vi kunne sammenligne os kommuner imellem og det var jo også KLs ambition for mange, eller en del, år siden da det her blev sat i gang. Der var jeg faktisk selv med i nogle arbejdsgrupper omkring det. Der hvor man kan sige, at vi er randt ind i de udfordringer, der er dels, lidt som du sagde Kathrine, at vi skulle selv bygge det og vi havde frie hænder, og vi var en af det der hed bølge-et-kommunerne dengang og så sammen med KMD. Nu er vi så blevet klogere på, at sådan skulle det ikke have været, men det vidste vi jo ikke og så må vi jo lave om på det. Det er jo den ene del af det, den anden del det er jo hele implementeringsfasen i sådan noget som det her, fordi der er intet system, uanset hvordan de er bygget eller hvilke systemer der er, der er bedre end den implementering af den enkelte kunde, altså i det her tilfælde Faaborg-Midtfyn Kommune, håndterer internt i forhold til nogle strukturer. Det vi i hvert fald er blevet klogere på, det er at det kræver virkeligt top-down ledelse. Det er bare ledelse og ledelse og ledelse. Alt det laissez faire med, at så er der 40 mennesker der skal diskutere om man kan klikke på det ene eller det andet, eller om den ene opløsning skal ligge det ene eller det andet sted - det er bare ikke noget der skal eksistere. Man skal købe et produkt, og det er egentlig der vi, langt hen ad vejen, som jeg ser det i hvert fald, er på vej hen med det service-tjek, at nu ensarter vi og så kommer det til at rulle, i et eller andet omfang, eller rigtig stort omfang, ud over hele organisationen. Det betyder så også, at der fra

mange af dem der har bygget en hel masse forskellige ting, ikke for at gå i detaljen, men ude i driften, er behov for, hvad hedder det, at sådan noget skrive-notater og alt muligt andet, hvor man ikke kan klikke ind i et felt og trække data ud på tværs, men alt muligt mærkeligt, der er bygget og sluger en fandens masse tid, undskyld mig udtrykket. Det er jo det vi ikke har været gode nok til og det er dét vi rydder op i nu. Så det er meget det organisatoriske setup og det synes jeg også, uden helt at gennemskue hvad der er jeres hovedpointe eller hvad I går efter ved interviewer i dag, men bare for at I får den del med også; at hvis der ikke er en stram styring, en stram organisation og dermed også en stram, og meget præcis, implementeringsprocess omkring det her, så bliver det ikke bedre. Så kommer vi ingen veje i at få gode data og valide data. Det tror jeg også, Jeppe Svendsen kan snakke mere omkring, når det er at vi snakker om alt det her. Det kræver disciplin, så det var mere for at åbne den del af ballet op. Det er for mig at se noget af det mest afgørende, fordi systemerne og teknikken, det har jeg sådan set ingen problem med, med den kendskab jeg har til det. Det skal nok virke. Det er simpelthen hvordan vi håndterer i organisationen, og når vi har ni plejehjem, der gør det på ni forskellige måder, så behøver man jo ikke være Einstien for at forstå at det er dårlig ledelse, eller i hvert fald ikke tilstrækkelig ledelse.

JS: Nej fordi Allan, der er jo næsten tre elementer her. Der er et system, der er en metode og der er en ledelse i forhold til at få folk til at gøre det samme på den samme måde. Altså gå i takt ikke. Og så siger du: "jamen det er egentlig ikke metoden, det er egentlig heller ikke systemet, det er mere organisation og ledelse i forhold til at få folk til at gå i takt".

K: Det er egentlig også det, vi lidt har tænkt ind i det samarbejde her med Jeppe (FMK), omkring også at de her ledere, altså sådan medarbejder-ledere ude i driften, at nogle af de analyser de også får, det bliver jo nogle for at se hvor gode, for eksempel, er de til at kvittere, hvor gode er de til at afsluttet nogle ting, hvor gode er de til at få lukket døde borgere. Sådan at de egentlig kan lede på det, der skal ledes på og have fokus på det, der skal være fokus på, hele tiden.

I: *Ja, sådan at man nærmest bruger dataen til at fortælle, hvor der er nogle uensartede..*

K: Ja, altså sådan en controlling af hele tiden at følge "er vi gode nok til de og de arbejdsgange?". Altså sidde og se bagom og tage nogle audits på og.. ja, se hvordan de bruger systemet.

A: Men det svarer jo lidt til, Jeppe Rosager så må du jo gerne lige byde ind på den her også, men det svarer jo lidt til, hvis vi sammenligner med, at vi har økonomisystemet, så ved vi jo godt, når vi sender penge ud, hvor vi konterer det henne og når vi modtager penge, så ved vi også, hvor vi konterer det, eller også om det er på en momsbærende konto eller ej og alle de ting der. Ligeså snart det er økonomi, så kan vi, langt hen ad vejen, godt finde ud af at opføre os nogenlunde fornuftigt i at bogføre rigtigt og at trække rigtige rapporter ud og alt muligt andet. Men ligeså snart vi kommer over i de her fagsystemer, så er der så mange forskellige fagprofessionelle, der et eller andet sted skal bruge det samme system, men har hver deres opfattelse af hvad der er godt.. nu siger jeg det bare.. ikke for borgeren, men for dem selv. Det vi er interesserede i, det er progression i borger-sagerne jo, Kathrine var lidt inde på effekten i starten. Giver det progression, når en medarbejder, der sidder inde i myndigheden på socialområdet bevilliger en ydelse ud fra paragraf 85 eller en af de andre paragraffer og stiller nogle mål op? Og så skal driften sådan set planlægge, hvordan de vil følge op på det. Når de så begynder at lave notater, der ligger i nogle rubrikker eller

felter, eller hvad det hedder teknisk inde i systemet, hvor vi ikke kan trække ud, så har vi jo fuldstændig anarki på området. Så ved vi jo ikke hvad der foregår i systemet, men det gør vi i økonomisystemet, der har vi ingen problemer. Så det handler meget om kultur, og det er tilbage til det jeg sagde før, rimelig stringent ledelse og forvaltning af, hvordan det er, vi arbejder med det her. Så det er i hvert fald nogle af de erfaringer, jeg synes jeg sidder tilbage med, og jeg synes vi skal blive bedre til.

I: *Ja, altså for man kan sige at systemet skulle jo gerne kunne understøtte de forskellige arbejdsgange, men i den forbindelse, så når du taler om mål for, hvordan der bliver visiteret og så videre, hvor bliver den information lagret?*

A: Ja, og der skal jeg jo starte i visitationen, for det er visitationen der bestiller en ydelse hos en leverandør, altså det kan både være private og kommunale leverandører. Der bliver der sat nogle mål op, og hvis man så ikke kan måle progression hen mod et mål, så ved man jo aldrig, hvor langt man er nået. Så vi ved ikke, hvordan vi går fra A til B og så videre, og det er i hvert fald noget af det, jeg synes der halter i vores setup i forhold til at være styret efter en data-dreven organisation, fordi de data vi har, de er jo ikke valide nok, altså det er Kathrine jo også inde på. Når vi har døde borgere liggende, der er levende i systemet, men de er borte og begravede ude i virkeligheden, så er fanden jo løs i laksegade, når vi skal begynde og planlægge og prioritere.

I: *Hvad er det så man bruger som mål på nuværende tidspunkt, fordi nu snakker vi lidt om at noget af det, man vil begynde at tænke ind, det var nogle effektmål. Men hvis man skal måle nu, hvor pleje-tung en borger er og så videre, hvad er det så man bruger?*

A: Der tror jeg mere at Kathrine og Jeppe skal på banen i forhold til de udtræk I sidder og laver og også udvikler på i øjeblikket. I kan lige sige noget mere om det.

JP: Ja, jeg kan godt prøve. Altså, vi har jo ikke en kategorisering, ligesom I ligger op til i jeres projektbeskrivelse, hvor man ligesom inddeler borgerne i forskellige typer og så siger det her er et segment af borgere og det her er et andet segment af borgere. Det har vi slet ikke i dag, så langt er vi slet, slet ikke nået med dataarbejdet her i kommunen. Altså jeg vil sige, vi er en meget ung organisation i forhold til at bruge data. Det har vi ikke gjort i ret lang tid, og det vi primært har brugt vores tid på, er at få styr på de her basisregistreringer, som Allan rigtigt omtaler. Så man kan sige, det med at fastsætte nogle mål for "hvad er det for en kategori den her borger tilhører"der er vi slet ikke i dag. Og man kan sige det med tilstande, det er slet ikke noget, vi er nået til at validere, så vi bruger det ikke som et aktivt arbejdsredskab, i dag, til at se "jamen hvor er det vi forventer borgeren er om et halvt år og om et år?". Der bliver sat nogle tilstande på, men det er slet ikke noget, vi trækker ud og bruger som data.

I: *Så når nu du skal se på det fra et økonomisk-perspektiv, så er det vel ud fra hvor meget tid der bliver brugt på den enkelte borger og så videre?*

JP: Lige nøjagtig. Det vi primært har koncentreret os om, det er aktivitetsdaten: Hvor mange timer leverer vi, hvor mange borgere har vi inde og i den dur. Så det er på et meget lavt datamæssigt niveau stadigvæk. Det er det.

I: *Hvad er din vurdering af hvor virkelighedssvarende det er så?*

JP: Jamen altså inden for hjemmeplejeområdet, som jeg antager at I kigger på, i forhold til de leverede timer, der er vi på et validt niveau. Altså vi kan godt stole på data. Der hvor vi har vores udfordringer, det er, som både Allan og Kathrine siger, det er, at vi har nogle visiterede timer som ligger på døde borgere, og vi har visitationer der ikke er foretaget revisiteringer på i rigtig lang tid. Så man kan sige, de visiterede timer, det er

ikke nødvendigvis et udtryk for borgerens, eller alles, plejebehov, aktuelt. Derfor skal I også gøre jer en overvejelse i forhold til, når I trækker data ud - vil I se på de visiterede eller de leverede timer?. Hvor jeg nok vil anbefale at I skeler mest til de leverede timer, også fordi, når jeg kigger ind i data nu her.. altså vores data, det kan I begynde at bruge fra 1.1.2018. Der har vi ligesom fået alle hjemmeplejegrupper med i data og fået det registreret i Nexus. Der kan jeg se, at når vi starter der 1.1.2018, der ligger vi på et rigtig højt niveau i visiterede timer, og det falder hen over tid, men de leverede timer de er faktisk forholdsvis stabile. Det siger mig, at der er et eller andet, der er sket i overgangen til Nexus fra det gamle system, som gør at man måske ikke har fået visiteret helt korrekt. Det vil jo give jer en bias I jeres analyse, hvis I løber an på de visiterede timer, fordi så vil I måske opleve, at borgerene vil blive forbedret, hvilket muligvis ikke er tilfældet.

I: *Ja, og vel eftersom at der afregnes på den leverede tid, så er det vel også der man vil sikre sig at det er i orden?*

JP: Lige nøjagtig.

I: *Okay.*

I: *Ja, men så har vi jo snakket lidt om det her med at begynde at bruge data til nogle lidt mere avancerede analyser, og en af de ting der begynder at komme meget frem, er det her med at man begynder at lave prædiktioner ud fra dataen omkring en borger, for eksempel i forhold til fremtidigt plejebehov og så videre. Så vil vi gerne høre, hvor I ser brugen af den her slags prædiktionsværktøjer i jeres organisation?*

A: Det ved jeg squ ikke lige. Hvad siger du Kathrine?

K: Jeg synes det er lidt svært. Nu må I sige hvis jeg skyder ved siden af, men når en borger søger om en plejehjemsplads, for eksempel, så mødes der sådan et visitationsudvalg, der sidder og snakker om borgeren og hvad de tænker og sådan noget. Der tænker jeg da, at hvis man kunne bruge sådan en form for analyse, at man måske havde et lidt bedre grundlag for det, eller at man ikke behøvede bruge så meget tid på det. Så man havde en fornemmelse af, "nå men vi ved formodentligt vil ti procent af de der komme til at skulle have plejehjemsplads ved os - hvordan ser det ud med pladserne. Øh ja.. sådan noget inden for det måske?

I: *Ja, som en slags styringsunderstøttelse, eller hvad skal man sige?*

K: Ja, altså hvis man tænker mere på tværs i kommunen, i forhold til de unge. Når de bliver 18, så kommer de over til os og ja om man kan sige der, bruge data til sådan at forudsige et eller andet.

JP: Hvis jeg må byde ind også?

I: *Ja.*

JP: Som jeg ser det i øjeblikket, så i forhold til at prædiktere fremtiden, jamen så tænker jeg ikke, det er så relevant i forhold til den økonomiske side, fordi der har vi et værktøj i dag, som faktisk er ganske udemærket. Vi har en demografimodel, som en af mine kollegaer i økonomiafdelingen sidder og udarbejder i forbindelse med budgetforhandlingerne. Så lige den store aggregerede udvikling i plejebehov, den tænker jeg ikke er så interessant. Der hvor jeg ser det er interessant, det er i forhold til opfølgning, altså den her revisitering af borgeren. Som det er lige nu, er man meget presset i myndighedsafdelingen, altså man har svært ved at finde ressourcer til at lave en opfølgning, og der synes jeg det kunne være super fedt, hvis man kunne et prædiktionsværktøj der siger: "Jamen, det er den her type af borgere vi skal gå efter, fordi det er dem data viser der ofte vil falde i plejebehov". I og med at vi ikke har ressourcerne til at foretage revisitering på alle borgere, så bliver vi

nødt til at være skarpe på, hvad er det for borgere vi revisiterer? og det skulle jo gerne helst være de borgere, som vi har det største potentiale for at sætte ned i tid.

I: *Så det er mere nede på det enkelte borger-niveau?*

JP: Ja, eller hvis man kunne lave sådan nogle grupperinger. Lave nogle kategorier og sige: "Jamen de her borgere, med den her karakteristik af hjælp, eller hvad I nu finder frem til af andre karakteristika, det er dem I skal foretage opfølgning på om et år eller om et halvt år. Fordi det er det data viser, at det er ofte der, at man vil se en forbedring i deres funktionsniveau, og dermed vil I kunne opnå noget ved at foretage en revisitering.

JS: Må jeg lige spørge om noget Jeppe? Fordi jeg tænker den demografimodel i laver, og nu gætter jeg bare, den tager vel udgangspunkt i et forventet plejebehov når du bliver en vis alder?

JP: Ja

JS: Og kan man sige er et lineært regnestykke på det. Det man vel også kunne bruge prædiktionsmodeller til, det var jo, sådan set, at få et lidt mere kvalificeret input på den og sige: "Det er ikke alle der er over 75, som gennemsnitligt koster x-antal timer". Men også måske noget mere se, hvad er det for nogle grupper af borgere vi potentielt står over for, måske også hvor de er henne og dermed også måske kunne begynde at gøre noget på den forebyggende bane?

JP: Det kunne man sagtens fordi så raffineret er modellen slet, slet ikke. Hvis jeg lige kort skal skitsere det, så er det, det her med at man har nogle alderskategorier, hvor man siger, for de borgere, der er i den her aldersgruppe, der ydes cirka så meget hjælp i gennemsnit per borger. Og så har vi en befolkningsprognose der siger om et år, om to, tre, fire år, der vil vi have så mange borgere i hver kategori og så er det bare at gange og dividere ud og så har vi en model.

A: Men Jeppe er der ikke indregnet noget 'sundhedsaldring' og sådan noget?

JP: Jo lige nøjagtig, det er der så også. Det går så lidt den anden vej, og siger at vi forventer at der er noget 'sund aldring' i det her, så en 80-årig i dag vil modtage mere hjælp, sammenlignet med en 80-årig om fire år.

A: Ja, men til gengæld så vil den 80-årige, når den så bliver 83, modtage meget hurtigere hjælp. Altså det er det der med, at man udskyder udgifterne. Når man lever sundere og så når at blive det ældre, så lige pludselig, sagt lidt groft, så får man hele pakken på en gang i stedet for at det bliver sådan gradueret op, kan man sige.

JP: Lige præcis. Altså vi vil altid nå derhen, at størstedelen af de gamle borgere har brug for et eller andet hjælp, en kort periode inden at de dør og det er ligegyldigt om de dør som 85-årige eller som 100-årige. Det har du fuldstændig ret i.

K: Men det kunne måske være rart at have sådan nogle forudsigelser i forhold til dem, vi giver forebyggende hjemmebesøg, og også dem, vi sender de der rehab-medarbejdere (rehabiliterings medarbejdere) ud til. Fordi, hvad er det for nogen vi bør fokusere på? eller hvor bør vi sætte ind? eller hvad er effekten af det? Altså vi ved jo ikke egentlig om dem, vi har der hedder terapeuter i visitationen, om det de gør virker. Forebyggende hjemmebesøg, det er også sådan noget man tit skærer af i mange kommuner, fordi man har svært ved lige at se... jamen de tager ud og snakker med dem, ja... kunne man få noget data og så bruge det til at fremskrive et eller andet.

I: *I forhold til hvornår man skal lave forebyggende hjemmebesøg og så videre, eller hvad?*

K: Ja, og hvad er det, der virker? og det bliver al sammen dyrere, når de bliver 80, men ham her kan vi se, at han bliver i hvert fald rigtig dyr. Altså sådan et eller andet der, det er bare en tanke.

A: Men noget af det, vi snakker om her, der er jo også noget som visitationen skulle bruge. Som I var inde på før, så hvis vi nu siger, at vi har 1.000 borgere og vi har x-antal visitatorer - hvis vi nu bare snakker 83-området (§83), så sidder de og bevilliger til det, men der er jo ikke en afdelingsleder, ikke mig bekendt i hvert fald, der sidder med nogle centrale udtræk på det og siger ud af den population på 1000 borgere her, så får borgerne så og så meget hjælp når de er i den alderskategori og hvor kan det bedst tænkes.. I forhold til at lave revisitation, hvordan skal en afdelingsleder prioritere medarbejdernes ressourcer? Der er det jo den enkelte medarbejder der nu har et eller andet, sådan som jeg er orienteret i hvert fald, så får de nogle data ud der siger: "Om otte måneder skal Fru. Jensen revisiteres"og ja, det når vi så ikke og så ringer vi og gør et eller andet. Hele den ledelsesmæssige prioritering i det jeg kalder en data-drevet organisation, det har ledelsen ikke mulighed for at gøre og det mangler vi jo mega meget. Og så er det jo når vi er ressourceknappe, og det er alle kommunerne, så det er interessant for alle det her, hvordan kan man så, som afdelingsleder prioritere sin tid i en visitation, der har for få visitatorer, og der ikke kan nå deres arbejde. Men hvad nu hvis de nået det rigtige? og samtidig lod det rigtige, altså det der ikke kunne gøres noget ved, ligge? Altså den prioritering den er der jo slet ikke, fordi overblikket ikke er der - og det synes jeg jo er katastrofalt.

JS: Ja nu laver i jo forebyggende hjemmebesøg, fordi folk har fødselsdag ikke?

A: Jaja.

JS: Nå jamen, det er jo kriteriet for, hvorfor I gør det. Og det kan man jo sige: "jamen var der nogle andre data-mæssige ting som gjorde, at det der det kunne... så ved jeg godt at der er også nogle ting som I skal gøre, men altså, var der nogle som vi skulle koncentrere os mere om end andre?

K: Ja både i forhold til kompetencer, altså "hvad er det for nogle kompetencer vi skal have om fem år?", "hvad skal vi skrue op og ned for?". Altså i forhold til at få rustet sig lidt til fremtiden, hele tiden. Nu snakkede vi med sådan et konsulenthus, den anden dag, der sagde at alle hjemmeplejere burde have en iPad... nåe ja men hvorfor? Hvad skulle vi bruge det til? Men hvis vi så kunne forudsige, at de kunne det og det til den tid ikke, så var man jo bedre rustet til fremtiden.

I: *Så har jeg lige et spørgsmål, i forbindelse med hele det her data-drevet organisation, fordi det vil man jo så gøre ud fra den data der ER tilgængelig, men hvor meget spiller økonomien og den prioritering man har på det gældende tidspunkt ind på den data? Hvis man vil gøre for at se på hvornår man, for eksempel, skal have et forebyggende hjemmebesøg, ud fra hvilke borgere der stiger i den tid, de skal have hjemmepleje og så videre, så er det vel også noget der har rod i økonomi?*

A: Jamen, jeg tror vi skal skelne imellem, de forebyggende hjemmebesøg.. altså er det ældrekonsulenterne, vi snakker om her? Det skal jeg lige være sikker på, altså i forhold til lovginingen, at de skal tilbydes de her forebyggende hjemmebesøg eller er det når de er kommet ind i systemet og så forebyggelse dér?

I: *Nej men jeg tænker faktisk ikke at det nødvendigvis er de forebyggende hjemmebesøg. Der er også, for eksempel, hvis vi snakker om det her med at se på, hvad det er for nogle borgere der stiger i tid. For de kan vel også stige i tid fordi man lige pludselig laver en anden prioritering?*

K: Altså tænker du på, om dem der sidder som myndighed, om de tænker økonomi ind i deres sagsbehandling?

I: *Ja, for eksempel.*
A: Altså helt officielt, så må de ikke det jo. Men de må gerne være økonomisk bevidste, fordi det er en faglig og saglig og individuel, konkret vurdering efter lovgivningen og der må du ikke skele til økonomien. Men det er jo ikke det samme som at vores visitation ikke må være effektiv jo. Hvis man ligger økonomien ind, så har vi jo sager i den danske kommunale verden, hvor det har kostet nogle topchefer jobbet.
K: Men man kan sige at hvis man bevilliger et hjælpemiddel så er det "bedst, billigst, anvendeligt", eller sådan et eller andet, så selvfølgelig har de det i øje også. Men som, Allan siger, så er det jo det andet som vægter. Men ledelsen gør det jo, kan man sige, fordi hvis man skal bevillige et hjælpemiddel over 50.000, tror jeg, så skal det godkendes ved en leder, og vil den der drøftelse om økonomi jo komme ind, fordi det jo så er lederens ansvar at holde budgetterne. Og også ved nogle særforanstaltninger, så er de også altid noget ledelse inde over i forhold til, når der skal ekstra økonomi.
A: Men der hvor man er begyndt at bruge økonomien, det er ikke på den enkelte borger eller på den enkelte medarbejder, det er jo inden for en afdeling. Det kunne være i myndigheden, men det kunne også være ude i hjemmeplejegruppen, og Jeppe det kan du jo snakke mere om i forhold til ledelsesinformation, hvor vores ledere, for eksempel, i hjemmeplejen, hvor der er en samlet økonomi på godt 100 millioner, eller der omkring, fordi de får penge efter de leverede timer. For en hjemmeplejegruppe, når tiden går op og ned, så kan man jo omregne det til kroner og ører - ligesåvel som vi kapitalisere vores sygefravær i kommunen, og ligepludselig går det op for en afdeling, at hvis man sænker sygefraværet med 0.5 så kan man faktisk lige pludselig være tre mere på arbejde om året. Og den tænkning må man jo gerne ligge ind i en visitation, at hvis vi var mere effektive med at følge op, og dermed kunne skabe en mere effektiv økonomi, så kunne man måske være to visitatorer mere, og så lige pludselig kommer man i den positive rulle. Det er jo sådan man kan bruge økonomien som en løftestang, men ikke på den enkelte borger direkte. Derfor er der en sammenhæng, i min optik, med det.
K: Men KMD kommer jo til at lancere et stykke værktøj, som vi har hørt det, så man kan sige: "Jeg har den her borger og jeg kan se tre mulige bevilliger, A, B, C, den dyreste det er A ikke - så der er den jo inde. Jeg ved godt det ikke er sådan direkte Allan, men det vokser jo lidt.
A: Jo jo.
I: *Ja, fordi det er jo det, at hvis man spørger en model så er der jo en risiko for, at den vil sige "jo mere hjælp, jo bedre", så derfor er det jo også en ting der er nødvendig, for at det skal kunne fungere i virkeligheden, kan man sige.*
K: Ja, man bruger meget sådan "mindst mulig indgriben"også, så vi starter ud med hold, eller virtuelt sammen, så det sidste trin på trappen det er egentlig, at der kommer et menneske ud til dig. Det er jo ikke direkte økonomi, men det er det jo alligevel, det er bare nogle andre ord på det - for det er selvfølgelig billigere at have dem på hold end individuelt.
I: *Så skifter jeg lige lidt emne her, men vi har været lidt inde på det her med, at man kan prøve at arbejde præventivt ved at målrette indsatser og sætte ind tidligt for at nedbringe plejebehov på lagt sigt. Hvordan arbejder I mere konkret med den slags tankegang? Eller om i gør?*
K: Vi har jo gjort lidt i forhold til det der digi-rehab-træning, de har ude i grupperne, hvor de godt må give det sådan lidt forebyggende, men det er jo ikke data på den måde, det er jo bare en vurdering af en hjemmepleje, en sosu.

I: *Ja, hvad går det ud på?*

K: Det går ud på, at hjemmeplejen kan se en borger, der begynder lige pludselig ikke at kunne samle noget op, eller har faldet, eller der har været et eller har fået et APV-hjælpemiddel. Der laver digi-rehab faktisk også et projekt med KL om en algoritme der kan sige, hvornår man skal tilbydes forebyggende træning. Men, de har ligesom nogle indikatorer de kan kigge på, så hvis der sker de her hændelser, så må de tilbyde borgeren et 12-ugers træningsprogram, to gange om ugen, når de kommer på besøg alligevel og yder hjemmehjælp. Så er det egentlig ikke tanken, at borgeren skal ned i tid, men bare skal kunne bibeholde sit funktionsniveau længere, end det måske var tilfældet, hvis man ikke havde trænet. Det er jo sådan noget der er rigtig svært at måle effekten på, men vi synes det ser ud til at virke.

A: Det du spørger ind til, det er hvordan arbejder vi med det forebyggende? Vi gør det vi skal i forhold til lovgivningen, med paragraf §83a. Når der kommer en henvendelse, så har vi nogle terapeuter, der i den ene del af vores visitation arbejder med et forløb på otte eller 12 uger, hvor de sætter nogle mål op. Hvis borgeren nu er kommet hjem med et brækket lårben, for nu at sige et eller andet, så er der både noget genoptræning, men der er også brug for lidt pleje, men det skulle jo så helst ikke være varig pleje. Der er målet at gøre borgeren fuldstændig selvhjulpen inden for den der periode. Efter den første visitation, så kommer anden visitation, og hvis man på det tidspunkt vurderer : "jamen vi kan ikke slippe hjælpen helt", så kommer der den formelle visitation ind over, og så vurderer man jo der hvor meget hjælp skal der så til i en given periode. Så på den måde er alle borgere jo inde over et 83a-forløb.. eller, altså ikke alle, men der hvor det giver mening. Det er det der med at vende borgerne i porten, ligesom man gør på sygehusene. Det fungerer jo - så kan man altid diskutere om det fungere godt eller skidt, men modellen er jo en del af vores visitations hverdag.

JP: Der ligger jo også forebyggelse i andre dele af Sundhed og Omsorg, end lige hjemmeplejen. Træningsområdet laver jo noget forebyggelse i forhold til borgere, der kommer ind på plejehjemmene og laver noget form for vedligeholdende træning.

I: *Hvad er det så for nogle værktøjer man bruger? For det jeg hører, er at det er vurderinger af hvor meget tid, der skal bruges og sådan. Hvad har man af hjælp til at foretage de her vurderinger?*

K: Jeg tror ikke, jeg forstår hvad du spørger om.

I: *For eksempel, det her med forebyggelse med den her træning, så skal der vurderes hvor lang tid, eller er det et fast forløb på de otte eller 12 uger?*

K: Altså det kommer an på, hvis det er sådan et 83a-forløb som Allan nævner, så er det jo en afgrænset periode med et rehabiliteringsforløb, men de får sådan en, som Jeppe siger, på plejehjemmene, 86 tror jeg den hedder, så er det ligeså længe...

A: Det er vedligeholdende.

K: Ja, så er det ligeså længe de trækker vejret, og det er jo fysioterapeutisk-vurdering.

A: Men vores kvalitetsstandarder de siger at i forhold til træning og genoptræning og sådan nogle ting, så er der et maks antal gange man kan komme til træning. Så skal der en visitation ind. Så på den måde har vi skærpet det meget her inden for det sidste halvandet-års-tid på alle vores kvalitetsstandarder. Der er blevet lidt tidsafgrænsede forløb der er i dem, og målet er jo at der bliver fulgt løbende op på dem. Det vil sige, når der starter en bevilling, så stopper der faktisk samtidig en bevilling også, og så er man nødt til at gå ind og revisitere dér. Så det er en anden stramning og det har jo ikke noget med

teknik at gøre, men der er et mere organisatorisk og administrativt tiltag, der er lavet. Tænkningen er, at det skal være med til at lave mere effektiv visitation, revisitation. Det gælder både hjemmeplejen og ude i træningsområdet, for der har vi set nogle mega, mega lange forløb uden at der er målt effekt. Der er det den enkelte medarbejder der bare, uden at skele til nogen kvalitetsstandarder har arbejdet med det. Det er der strammet op på nu, og det begynder også virke, kan vi se på forskellige parametre.

I: *Okay, så det er ud fra en tankegang om, at de ressourcer det tager at lave en revisitation, de er godt givet ud i forhold til at få stoppet forløb der ikke nødvendige?*

A: Ja det er ligesom at investere 100 kr. så skulle jeg have 110 kr. ugen efter jo. Så har vi lige pludselig til en medarbejder mere, eller også noget til kommunekassen.

I: *Men det er ens for alle borgere hvornår et forløb vil stoppe og der laves revisitation, eller er det ud fra nogle parametre på borgeren?*

A: Nej, det er beskrevet i vores kvalitetsstandarder. Dem vil I også kunne finde på vores hjemmeside. Og så er det op til den enkelte visitator, altså bevilligende visitator at følge op på det. Samtidig har vores personale ude i driften egenkontrolsforpligtigelse til at melde tilbage, fordi de er så den fremskudte visitation samtidig med at de er drift, og det øver vi os også lidt i kulturen jo. Det er det med, hvis der er for meget nedgang i ydelserne, så ved vi jo godt hvad det betyder, så betyder det, at der skal være færre medarbejdere. Så det er også en kultur-del og så er vi tilbage til noget af det jeg startede med at sige. Men det arbejder vi meget på at intensivere også blandt andet, Jeppe det kan du i hvert fald også snakke med omkring, egenkontrollen. i hjemmeplejen er den blevet meget mere synlig og meget mere italesat ude i hjemmeplejegrupperne, end den var tidligere. Vi kan jo se om det er på grund af visitation, centralt, eller om det er på grund af egenkontrol, at ydelserne revurderes.

JP: Ja, altså vi er begyndt at arbejde mere med det. Det jeg hører du spørger ind til, Mathilde, det er om man har et redskab, et form for regneark eller et skema, man bruger til at vurdere, hvor mange timer og lang en periode skal det her forløb være. Det korte svar er, at det har man ikke i dag. Det er den enkelte visitator, som Allan siger, eller den enkelte terapeut, der sidder og vurderer, ud fra egen faglige viden og erfaring, der siger: "Jamen du skal beviliges så mange timer her og de her indsatser". Så det at I vil komme med en model, man kan putte nogle tal ind i og så kommer den selv med en vurdering af, at den her borger skal have dét her forløb, med så mange timers ugentlig hjælp, det vil være noget helt, helt nyt for hele organisationen i Sundhed og Omsorg.

K: Man kan også sige, det her med straksbevillinger og hjælpemidler er jo noget der er kommet til i mange kommuner også, og det er måske første skridt til det. For det har været en stor omvæltning for mange af de her medarbejdere, at de ikke måtte køre ud til borgeren, men at de skulle lave en administrativ bevilling ud fra nogle krydser man satte i et skema.

I: *Ja, hvis du lige kunne uddybe "straksbevilling". Vil det sige at man på stedet...*

K: Ja så ringer man ind til vagttelefonen og siger at der er det og det med benene og sådan og sådan og så skal man egentlig kunne bevillige en rollator. For prisen på en rollator kontra sagsbehandling og tid. Der er det ligeså billigt at bevillige en rollator, på den måde - bare give den i telefonen og så kan de hente den eller få kørt den ud af et depot. Fordi ellers tager sådan nogle bevillinger, især hjælpemidler, rigtig lang tid, fordi man skal ud til borgerne og se hvad de gør og se nogle aktiviteter, så hjem og sagsbehandle på det og finde ud af: Hvad er serviceniveauet på det? Hvad siger lovgivningen? Er der

nogle principafgørelser i forhold til hvilket hjælpemiddel det skal være. Så der er tit en lang arbejdsproces, for hjælpemidler der måske ikke koster mere end 500 kr. I forhold til deling af data og brug af data, så sidder de jo i en myndighed hvor de både bevilliger hjælpemidler, hjemmepleje, personlig og praktisk hjælp, bostøtte, kørsel og alt sådan noget. Her har vi talt meget om, i hvert fald, det her med at bruge hinandens data. Det er man ikke så gode til, fordi man kan sige måske var der mange bevillinger der kunne ske ud fra, at man kunne se at ældrekonsulenten havde været ude og skrevet det og det, nogle tilstande og der var nogle andre, der havde været ude. Så man så havde at når de tilstande er sådan og sådan, så bevilliger vi, men det gør man ikke så meget brug af. Det er sådan noget jeg ved de arbejder hen imod.

I: *Ja, så der sådan at hvert område laver deres vurdering?*

K: Ja, de har hver deres tilgang. Altså man kan sige hvis man har en arbejdsprocesmodel som metode inden for sin faglighed, så vil man gå frem efter syv trin og så er det svært for sådan nogle fagligheder at bruge andres data på det, fordi der er en anden kultur. Det vil være det der skal vendes.

I: *Når en borger har ansøgt om hjemmehjælp og skal tale med visitationen, hvad er det så for noget data der bliver registreret i først møde?*

K: Det kan jeg ikke svare direkte på, men det kan vi finde ud af og sende til jer, hvis det er.

I: Okay, men vil I umiddelbart tænke at.. altså fordi den første vurdering, det er vel den man visiterer ud fra, så bliver der registreret data omkring borgerens tilstand, eller er det mere en vurdering der bliver lavet og ikke nødvendigvis dokumenteret så meget?

K: Altså man har jo journaliseringspligt, kan man sige, så man har jo en pligt, til at dokumentere det man høre borgeren fortælle, når de ringer. Og så når de er ude, så er der jo også nogle faste ting de spørger ud fra, i forhold til hvad lovgivningen siger, for at de kan give afslag eller bevillige. Men jeg kan ikke ned i detaljen, det ved jeg ikke om de andre kan?

A: Nej det ved jeg ikke i detaljer, men de har jo en guide, de går efter i forhold til hvordan de visiterer borgerne.

JP: Altså som jeg har forstået det, så mange af visiteringerne, det er jo skrivebordsvisitationer, hvor man får en beskrivelse ind af borgeren, eventuelt fra sygehuset eller anden instans. Så det jo ikke hver gang man er ude ved borgeren, inden at man bevilliger hjælp.

K: Ellers har de tit sendt nogle terapeuter i byen for dem. Man kan sige, hende der sidder i sambo-vagt, Allan hedder den det?

A: Ja.

K: Altså hun sidder jo og får alle de her udskrivelser fra sygehuset og så bevilliger hun jo hjælp ud til de borgere, ud fra hvad sygehuset har skrevet. Så der kører det jo bare på samlebånd, kan man sige.

I: Okay, det var egentlig sådan en nysgerrighed i forhold til om det første spor er selve visitationen eller om det også er alt det der ligger til grund derfor, men det lyder til at der er nogle faste ting, der ligesom kommer ind.

A: Jamen det er lidt det er med, som vi lige er inde på, at det er lidt forskellige måder borgerne kommer ind i visitationen på jo. Fra sygehuset, der er det ganske klart, at der kommer der en udskrivelse fra sygehusene via vores sambo, der, og så står der jo sådan, og sådan og sådan og så er det kommunen der skal levere et eller andet sted. Men hvis det bare er Fru. Jensen henne på gaden, hvor det bare er de pårørende der ringer og siger:

"Nu har min mor brug for rengøring, fordi nu kan hun ikke gøre rent mere", så er det jo en anden situation. Der har vi jo, i og med at vores visitation i hvert fald er presset, der er det sådan at hjemmeplejen kan jo godt gå ind og levere ydelsen, jeg tror det er de første måske 10 eller 14 dage, inden visitationen så reelt kommer ud. Så er der sådan en periode, at den bevilling de i virkeligheden kører på, er det driften der har bevilliget det. Og sådan er der forskellige variationer, alt efter hvad det er for nogle ydelser, vi snakker om i hjemmeplejen. Men ellers, der må jo ligge i Nexus, en eller anden skabelon de gennemgår når de visiterer borgerne, altså.

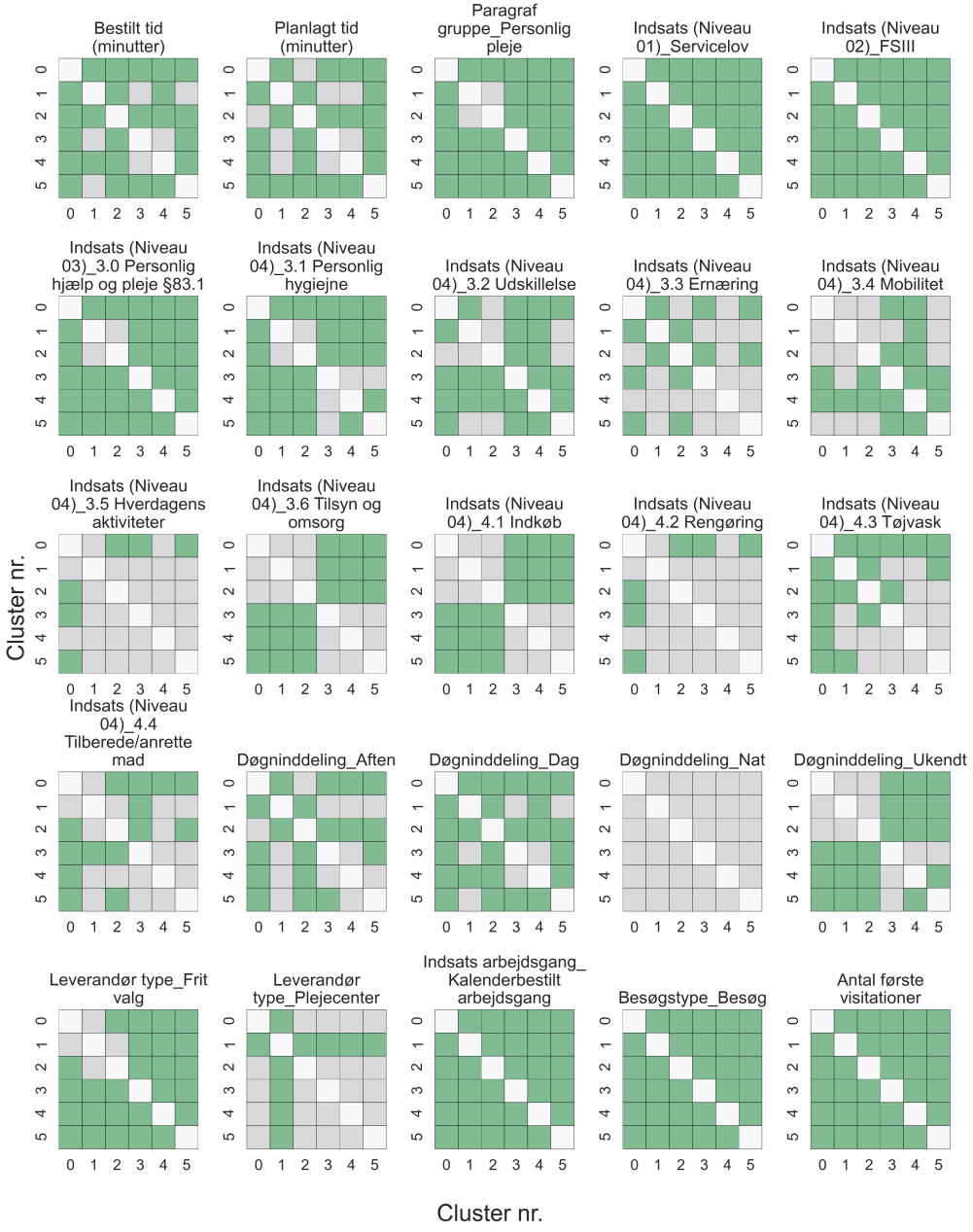
K: Ja og der er også noget inden for Fælles Sprog III de skal udfylde jo.

A: Jaja, det er jo alle de ting der.

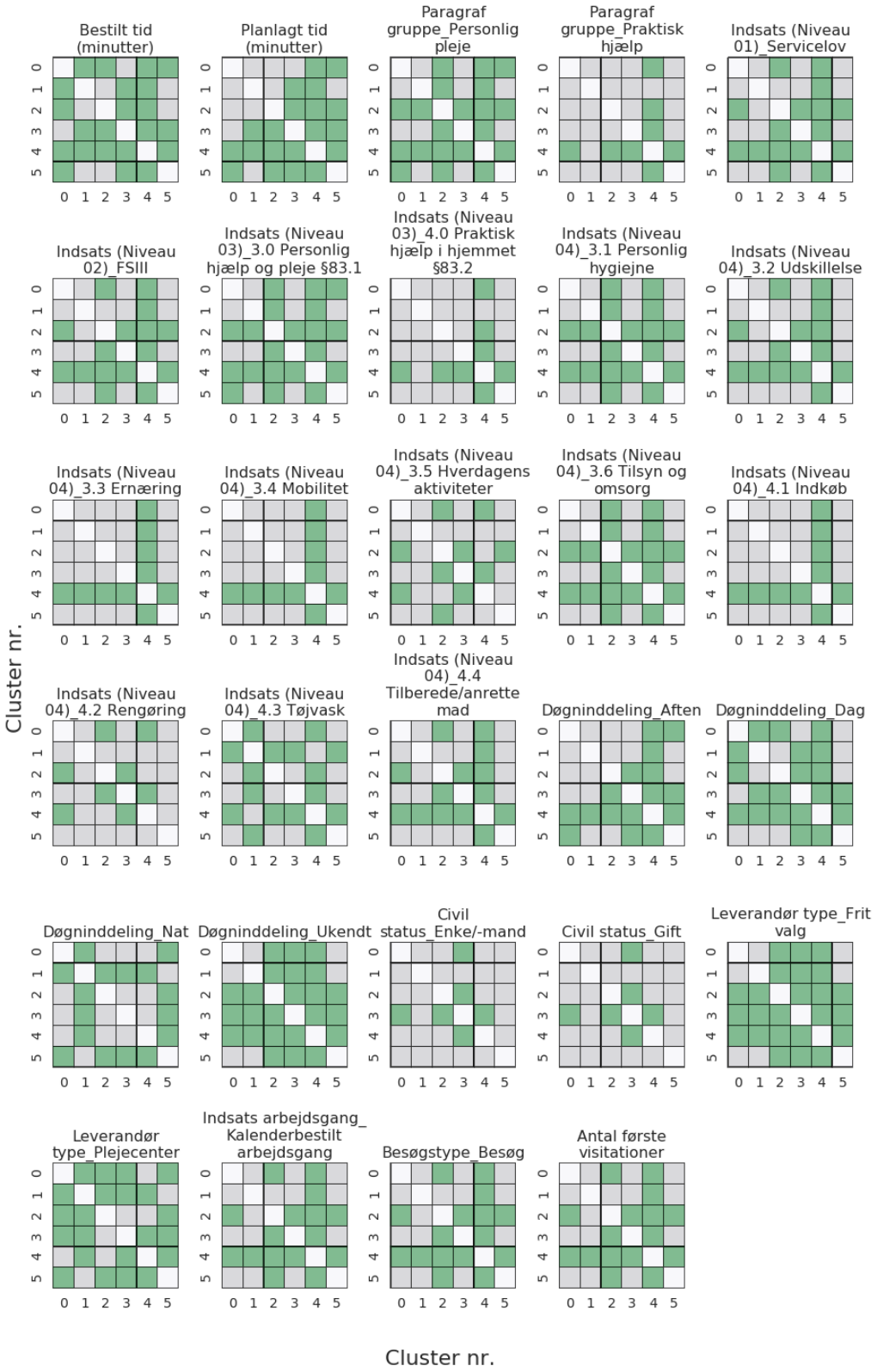
I: Jamen det er sådan set alt hvad vi vil vide.

Feature signifikans-maps C

Dette appendiks indeholder signifikans-maps, på originalt data i **figur C.1** og trininddelt i **figur C.2**, der beskriver hvorvidt en given feature i det anvendte datasæt, er statistisk signifikant forskellig mellem hvert cluster-par. Dette er, som beskrevet i afsnit 4.5, fundet gennem en Kruskall-Wallis og efterfølgende Tukey HSD test post hoc.



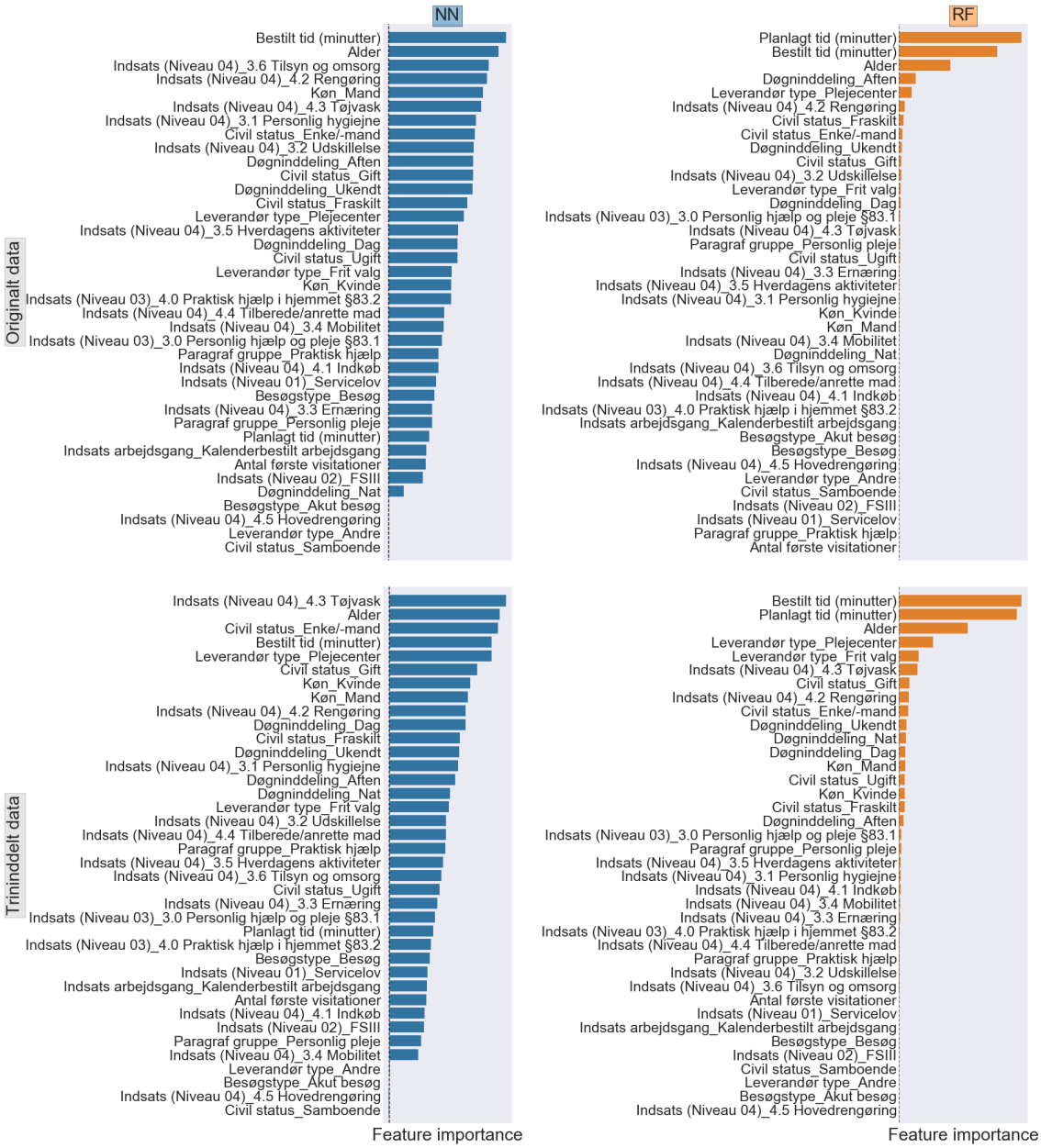
Figur C.1. Signifikans-maps for hver feature i det originale datasæt. Cluster-par, hvor den pågældende er statistisk signifikant forskellig er markeret med grøn.



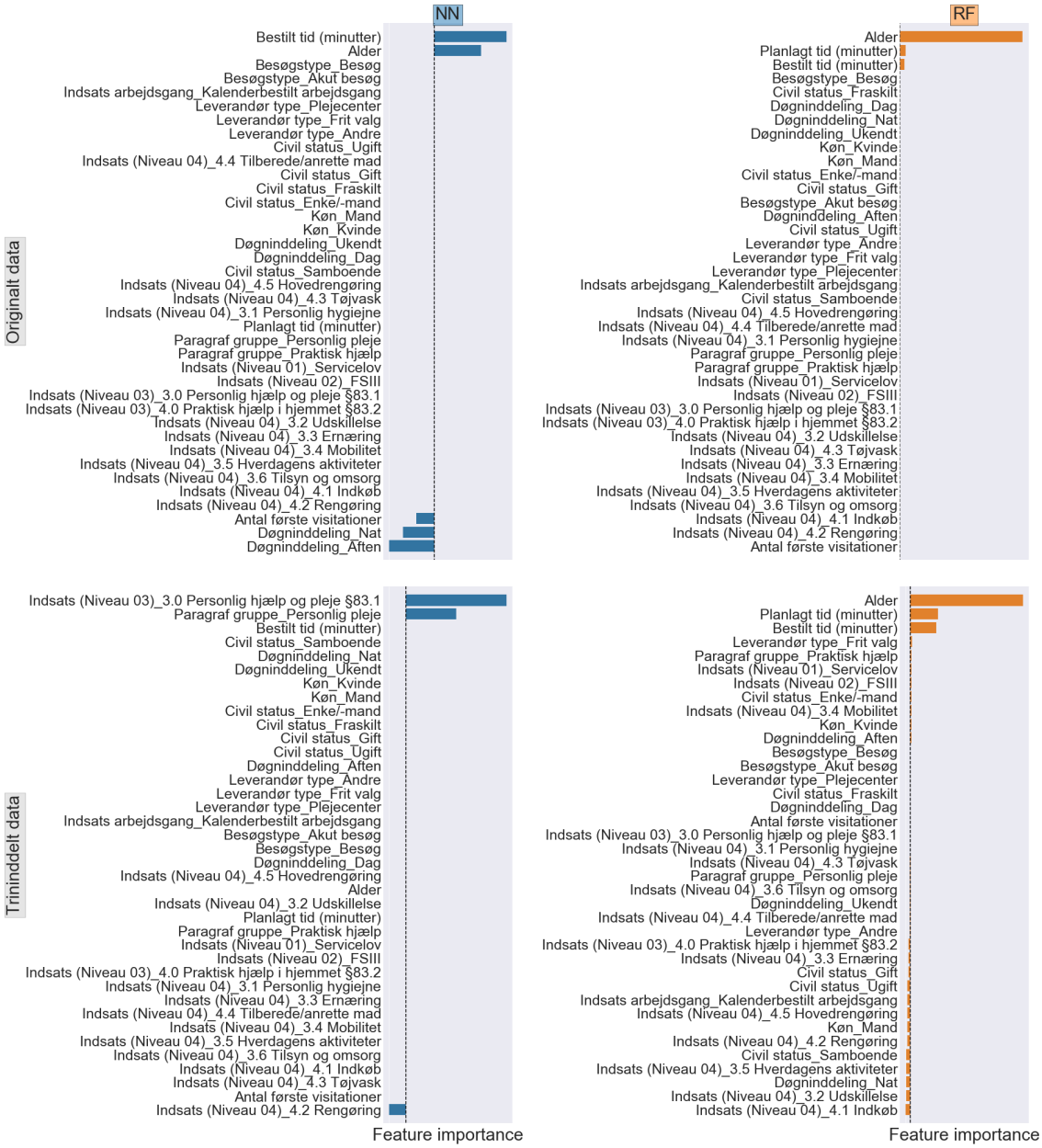
Figur C.2. Signifikans-maps for hver feature i det trininddelte datasæt. Cluster-par, hvor den pågældende er statistisk signifikant forskellig er markeret med grøn.

Feature Importance D

De følgende to figure, **figur D.1** og **figur D.2**, viser permutation og 'drop column' importance for alle inkluderede features.



Figur D.1. Feature importances over aller inkluderede features, for henholdsvis Neuralt Netværk (NN) og Random Forest (RF), på originalt og trininddelt data.



Figur D.2. Feature importances over aller inkluderede features, for henholdsvis Neuralt Netværk (NN) og Random Forest (RF), på originalt og trininddelt data.