



AALBORG UNIVERSITET
STUDENTERRAPPORT

Prædiktiv model til kapacitetsudnyttelse

P5 Semestersprojekt - Efteråret 2016

Gruppe 5405

4. Semester

School of Medicine and Health

Sundhedsteknologi

Fredrik Bajers Vej 7A

9220 Aalborg

Tema:

Klinisk teknologi

Projektperiode:

P5, Efteråret 2016

Projektgruppe:

5405

Medvirkende:

Linette Helena Poulsen

Maria Kaalund Kroustrup

Nirusha Jeevanadan

Rolf Oberlin Hansen

Sageevan Sayananthan

Sebastian Munk

Synopsis:



Vejleder:

Hovedvejleder: Sten Rasmussen

Bivejleder: Pia B. Elberg

Kliniske Vejleder: Christian Kruse.

Sider: XX

Appendikser: XX

Afsluttet:

Offentliggørelse af rapportens indhold, med kildeangivelse, må kun ske efter aftale med forfatterne.

Indholdsfortegnelse

1	Indledning	1
1.1	Initierende problemstilling	1
2	Problemanalyse	3
2.1	Kapacitet	3
2.1.1	Ortopædkirurgisk afdeling	4
2.2	Ubalance i kapacitetsudnyttelse	6
2.2.1	Personalesikkerhed	6
2.2.2	Patientsikkerhed	6
2.3	Belægningsgrad på ortopædkirurgisk afdeling	7
2.3.1	Problemformulering	9
3	Problemløsning	10
3.0.1	Forudsigelse af indlæggelsesvarigheden	10
3.0.2	Databehandlingsmodel	11

Indledning 1

Flere danske hospitalsafdelinger oplever i perioder at have flere patienter end der er kapacitet til. Dette medfører, at der sker en ubalance i kapacitetsudnyttelsen, da der forekommer mangel på sengepladser, personale og rum.[**Company2013**] I budgetfordelingen for Aalborg Universitetshospital i år 2017 indgår det, at ventetiden på en operation, for elektive patienter, skal reduceres fra 57 dage til 50 dage[**Budget2016**]. Dette forventes at medføre, at det daglige antal elektive patienter, der indlægges, vil skabe en reduktion i antallet af ledige sengepladser til akutte patienter. Derudover forventes det, at procentdelen af danskere over 65 år vil stige fra 29 % til 34 % og dermed også antallet af fremtidige patienter[**RegionNord2016**]. En stigning i antallet af patienter vil i takt med kortere ventetid på behandling skabe en udfordring ift. planlægning af de elektive indlæggelser. Planlægning af indlæggelser har indflydelse på personalets arbejdsdag ift. hvor mange patienter de skal varetage. Dertil kan sygeplejersker i nogle perioder opleve at skulle varetage ekstra patienter. Hertil mener hver anden regionalt ansat sygeplejerske på tværs af regionerne, at den travle arbejdsdag påvirker patientsikkerheden[**Kjeldsen2015**]. Et studie påviser, at ved blot én ekstra indlagt patient i 30 dage pr. sygeplejerske øges mortalitetsraten for patienten med 7 %[**Aiken2002**]. Foruden personalets øgede risiko for at begå fejl ift. behandlingen af patienter forekommer der ligeledes kapacitetsmangel, som medfører, at patienter overføres til uhensigtsmæssige områder som f.eks. gangarealer og fyldte stuer[**Madsen2014**]. Dette kan forårsage, at patienter såvel som pårørende oplever et skærpet privatliv[**Heidmann2014**].

På ortopædkirurgisk afdeling på Aalborg Universitetshospital opleves ligeledes en ubalance i kapacitetsudnyttelse. Over en 18 måneders periode forekommer der hhv. en belægning over og under 100 %[**SDS2015**]. Dette betyder, at ressourcerne ikke udnyttes optimalt, hvortil afdelingen eks. oplever perioder med mangel på personale og perioder med for meget personale ift. indlagte patienter.

1.1 Initierende problemstilling

På baggrund af ovenstående opstilles følgende initierende problemstilling: *Hvordan påvirkes ortopædkirurgisk afdeling på Aalborg Universitetshospital af ændringerne vedrørende kapacitetsudnyttelse og hvor udbredte er belægningsrelaterede problemer*

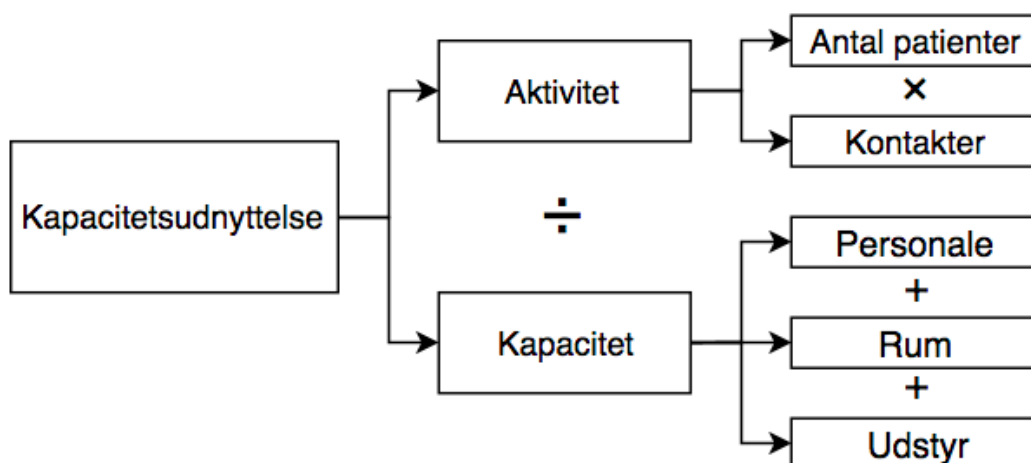
på afdelingen?

Problemanalyse 2

2.1 Kapacitet

Kapacitetsudnyttelse betegner forholdet mellem aktivitet og kapacitet. Aktivitet omhandler patient og kontakt, herunder består kontakt af forundersøgelse, behandling og kontrol. Kapacitet omfatter antallet af personale, udstyr og rum, hvor personalet består af læger, sygeplejersker og sekretærer. Udstyret beskriver antallet af maskiner på en afdeling og antallet af rum beskriver opbevarelsen af udstyret. Den samlede kapacitetsudnyttelse er defineret ud fra, at der produceres mest muligt for de investerede ressourcer.[Company2013]

Den samlede kapacitetsudnyttelse



Figur 2.1: Den samlede kapacitetsudnyttelse som er defineret ved forholdet mellem aktivitet og kapacitet. Aktivitet omfatter antallet af patienter samt kontakter og kapacitet omfatter personale, rum og udstyr.[Company2013]

Ud fra figur 2.1 fremgår det, at kapacitetsudnyttelse er forholdet mellem aktivitet og kapacitet. Dertil ses aktivitet som antal patienter multipliceret med kontakter. Kapaciteten udgør personale, rum og udstyr lagt sammen. Antallet af patienter,

der repræsenterer en del af aktivitet beskriver ligeledes belægning på hospitalets afdelinger.[**Company2013**]

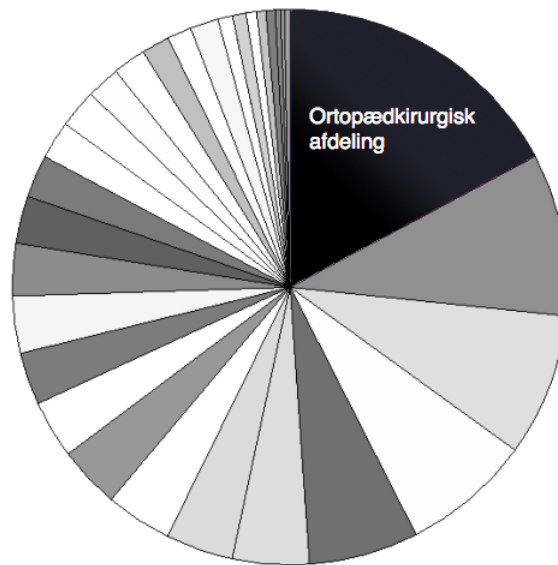
Belægning er defineret ud fra antallet af patienter, der er normeret til på en afdeling[**Heidmann2014**]. Når en 100 % belægning opnås, svarer dette til, at de disponible sengepladser på en afdeling er taget i brug. Ved en belægning på over 100 % betyder det, at der er flere patienter end afdelingen er normeret til, hvilket vil sige, at afdelingen yder mere end der er kapacitet til. Ud fra figur 2.1 vil dette betyde, at der ikke er ligevægt mellem aktivitet og kapacitet, hvilket i dette tilfælde vil forårsage kapacitetsmangel på afdelingen. Det kan derfor være nødvendigt, at personalet skal varetage flere patienter samt arbejdsopgaver, det kan ligeledes være nødvendigt at tilkalde ekstra personale for at opnå en balance i kapacitetsudnyttelsen. Hvis der derimod er en belægning på under 100 % er der omvendt færre patienter end afdelingen er normeret til. Dette betyder, at der er flere sengepladser end patienter, hvilket ligeledes fører til en ubalance i kapacitetsudnyttelsen. I denne situation er der mere personale end nødvendigt til at varetage de enkelte patienter, hvilket betyder, at der ikke er fuld udnyttelse af personalets arbejdskraft.[**Pauly1986**]

Det anses herved vigtigt, at der er balance mellem aktivitet og kapacitet, således de investerede ressourcer udnyttes optimalt. Det ønskes derfor at opnå en kapacitetsudnyttelse på 100 %. Ud fra dette vil der fremover undersøges betydningen af kapacitetsudnyttelse på ortopædkirurgisk afdeling på Aalborg Universitetshospital.

2.1.1 Ortopædkirurgisk afdeling

Kapacitetsudnyttelse afhænger af det budget som hver afdeling har til rådighed. Dette budget udregner Sundhedsstyrelsen ud fra diagnoserelaterede grupper (DRG). DRG anvendes til at analysere omkostninger og aktivitet på et hospital.[**DRG2016**] Ortopædkirurgisk afdeling har et budget på 700.872.744 kr, som svarer til 17,2 % af det samlede budget for alle afdelinger på Aalborg Universitetshospital. Det samlede DRG for afdelingerne på Aalborg Universitetshospital er illustreret af figur 2.2.[**Rasmussen2016**] Størstedelen af budgettet anvendes til personale- og patientudgifter, som svarer til hhv. 60 % og 32 %. Det resterende budget anvendes til bygninger, it, apparatur, inventar samt drift og service[**Noegletal2016**].

Fordeling af DRG budget for afdelingerne på Aalborg Universitetshospital



Figur 2.2: Fordeling af DRG for samtlige afdelinger på Aalborg Universitetshospital. Det fremgår, at ortopædkirurgisk afdeling har en større andel end de resterende afdelinger.[Rasmussen2016]

Personalearbejde

På ortopædkirurgisk afdeling på Aalborg Universitetshospital arbejder personalet i gennemsnit 37 timer om ugen[Danske2015]. Vagterne kan variere fra XX til XX timer, hvoraf det både kan være nat- og dagvagter. Der er indlagt betalte pauser, hvilket betyder, at personalet skal være til rådighed under pausen. Pauserne er opdelt i XX om dagen. Afdelingen er delt op i XX vagthold og har vagtskifte hver XX time. Personalet varetager XX patienter om dagen.¹

Patientindlæggelse

Som beskrevet i afsnit 2.1 ønskes en 100 % kapacitetsudnyttelse, dertil ønskes ligeledes en belægning på 100 %. For at opfylde dette skal der være ligevægt mellem antallet af sengepladser og patientindlæggelser. På ortopædkirurgisk afdeling har de XX sengepladser til rådighed, som er fordelt på XX afsnit.

Ortopædkirurgisk afdeling modtager både elektive samt akutte patienter. Elektive patienter omfatter både indlagte og ambulante patienter. Ved pludselig forværret tilstand kan elektive patienter skifte status fra elektiv til akut. Akutte patienter defineres som personer, der er henvist til hospitalet efter en akut opstået tilstand. Sammenlignes der med de resterende afdelinger på Aalborg Universitetshospital, har ortopædkirurgisk afdeling flest elektive indlæggelser.[RegionNord2016] Elektive patienter indlægges i tidsrummet XX-XX og udskrives i tidsrummet XX-XX. Udskrivelsen af akutte patienter foregår i samme tidsrum. På afdelingen planlægges

¹FiXme Note: Vi mangler informationer for at kunne skrive dette færdigt.

elektive patienter med forbehold for, at der er uforudsigelige indlæggelser af akutte patienter pr. XX. Herunder planlægges XX elektive patienter, således at der er plads til XX akutte patienter.²

2.2 Ubalance i kapacitetsudnyttelse

Ved kapacitetsmangel på ortopædkirurgisk afdeling forekommer en omstrukturering af personalets arbejdsopgaver, som sikre patientens behov, opretholdelse af kvalitet og udnyttelse af kompetencer. Dette er med henblik på at opnå en balance mellem de ressourcer og de krav, der stilles i den pågældende situation.[Bjerg2016]

2.2.1 Personalesikkerhed

I tilfælde af kapacitetsmangel er der udarbejdet en arbejdstilrettelæggelse af Region Nordjylland for personalet på ortopædkirurgisk afdeling. Ved kapacitetsmangel påtager lederen, eller dennes stedfortræder, ansvaret for at finde en løsning på dette problem. Dette kan betyde, at det afgående vagthold skal blive indtil en midlertidig løsning er fundet eller en tidligere indkaldelse af det næste vagthold. I nogle tilfælde kan det være nødvendigt at låne ressourcer fra andre afsnit eller indkalde personale fra vikarbureauet. Derudover undersøges det, hvorvidt behandlingen af elektive patienter kan aflyses.[Bjerg2016]

Ved overarbejde må en arbejdsuge for en sygeplejerske, ifølge arbejdstidsaftalen indgået med Dansk Sygeplejeråd, ikke overstige 48 timer[Danske2015].³ Hvis sundhedspersonalet er nødsaget til at arbejde længere end den normale arbejdstid, viser dette sig at have en negativ indvirkning på personalets arbejdsopgaver[Dinges2004]. Overarbejde kan resultere i en presset arbejdsdag og dermed en forringet kvalitet af behandlingen[Kjeldsen2015]. Dertil mener hver anden regionalt ansat sygeplejerske på tværs af regionerne, at den travle arbejdsdag påvirker patienternes sikkerhed[Kjeldsen2015].

2.2.2 Patientsikkerhed

Under perioder med kapacitetsmangel er det ofte nødvendigt at overflytte patienter til andre afdelinger, gangarealer eller fyldte stuer, herved er det ofte patienter, der snart udskrives, der overflyttes.⁴ Overflytningen kan belaste både fysiske og psykiske forhold for patienter såvel som pårørende[Heidmann2014]. Herunder kan skærpet privatliv forekomme hos patienter, der er flyttet til gangarealer og fyldte stuer[Madsen2014].

²FiXme Note: Vi mangler informationer for at kunne skrive dette færdigt + tilføjelse af Sebastians grafer (elektive/akutte) + (indlæggelse/udskrivelse)

³FiXme Note: Spørgsmål til sygeplejersker: Hvordan prioriteres pauser under overbelægning?

⁴FiXme Note: Sygeplejersker: Vi vil gerne høre om der prioriteres i forhold til hvilke patienter der flyttes. Er der en bestemt afdeling i flytter patienterne over på eventuelt en afdeling der ligner ortopædkirurgisk?

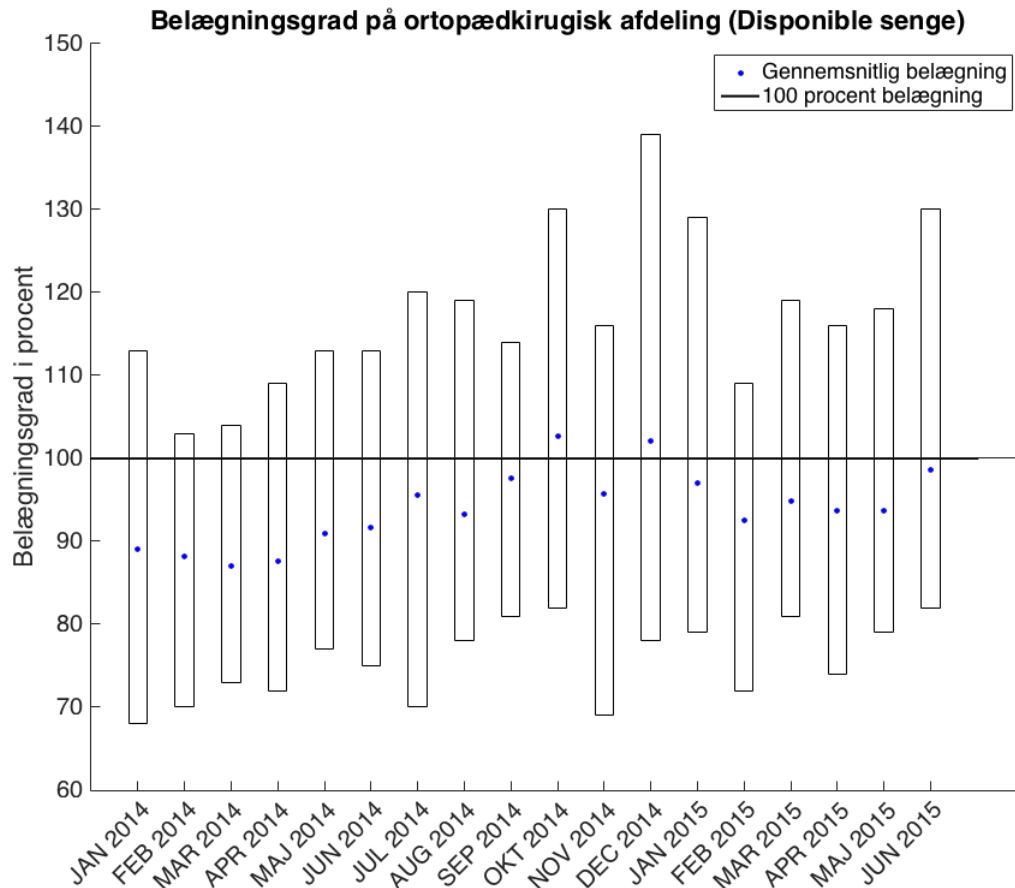
Som nævnt i afsnit 2.2.1 forringes kvaliteten af behandlingen ved overarbejde, dertil ses det ligeledes, at mortalitetsraten øges med 1,2 % ved en overskridelse af belægningen med 10 %, ifølge et dansk studie fra år 2014[Madsen2014]. Hertil understreges det, at der kan være ukendte parametre, der påvirker mortalitetsraten, og det nødvendigvis ikke er belægning, der er den primære årsag til en øget mortalitet. For at undgå forringet kvalitet af behandling forsøges det at få patienterne udskrevet tidligere, således et ønske om balance mellem aktivitet og kapacitet opnås.

Der tilkaldes en brandvagt til afdelingen, hvis en belægning over 100 % har fundet sted i over 4 timer for således at sikre patienterne ved evakuering under brand. En brandvagt kan højest overvåge to afdelinger på samme etage, hvorfor det kan være nødvendigt, at der indkaldes flere. Det er afdelingens ansvar at afvikle overbelægningen hurtigst muligt ved at udskrive patienter eller overflytte patienter til andre afdelinger.⁵ Hver gang der tilkaldes en brandvagt faktureres dette af Aalborg Universitetshospital.[Beredskab2016]

2.3 Belægningsgrad på ortopædkirurgisk afdeling

På ortopædkirurgisk afdeling på Aalborg Universitetshospital ses en varierende belægningsgrad for hver måned. Som tidligere nævnt i afsnit 2.1 ønskes en fuld kapacitetsudnyttelse, hvoraf alle sengepladser ønskes at være i brug. Belægningsgraden er antallet af de anvendte disponible senge. På figur 2.3 ses belægningsgraden fra år 2014 til 2015 på ortopædkirurgisk afdeling.[SDS2015]

⁵FiXme Note: Har skal indsættes hvis de har et samarbejde med en anden afdeling

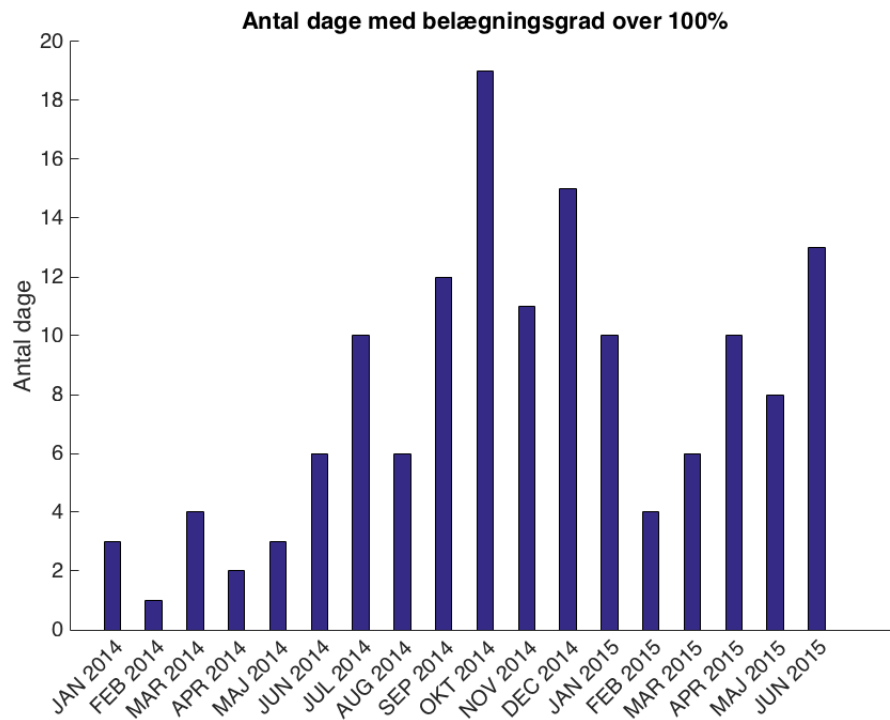


Figur 2.3: Belægningsgraden på ortopædkirurgisk afdelingen på Aalborg Universitetshospital målt over 18 måneder fra år 2014 til 2015. Søjlerne viser belægning ift. 100 %, hvortil maksimal og minimum belægning ligeledes illustreres. De blå punkter viser den gennemsnitlige belægning for hver måned.[SDS2015]

Det fremgår af figur 2.3, at ortopædkirurgisk afdeling oplever en belægning hhv. over og under den ønskede belægning på 100 %. I december måned år 2014 ses en maksimal belægning på 139 % og en minimums belægning på 78 %. Maksimums belægning kan indikere, at der er flere indlagte patienter end afdelingen er disponeret til, herved har afdelingen oplevet kapacitetsmangel. Minimums belægning kan indikere, at der ikke har været tilstrækkelige elektive patienter i perioder, hvilket ligeledes medfører ubalance i kapacitetsudnyttelsen. Af figur 2.3 er den gennemsnitlige belægning pr. måned hyppigst under 100 %. Oktober og december måned i år 2014 oplever dog en gennemsnitlig belægning over 100 %. Den gennemsnitlige belægning ses varierende mellem 90 og 100 % for de resterende måneder, hvilket kan indikere, at afdelingen oplever kapacitetsmangel i kortvarige perioder.[SDS2015] Det fremgår ikke af den anvendte data, hvorvidt belægningen opleves i timer eller flere døgn. Dertil skal der tages forbehold for, at det ikke er angivet om det er elektive eller akutte patienter, der udgør en belægning over 100 %.[SDS2015]

For at underbygge belægningsgraden yderligere, illustrerer figur 2.4 antal dage pr. måned med en belægningsgrad på over 100 %. Denne graf er udarbejdet ud fra

ortopædkirurgisk afdeling over de samme 18 måneder som figur 2.3. [SDS2015]



Figur 2.4: *Belægningsgrad over 100 % målt i antal dage over 18 måneder fra år 2014 til juni 2015 for ortopædkirurgisk afdeling på Aalborg Universitetshospital.* [SDS2015]

Det fremgår af figur 2.4, at der i oktober måned år 2014 opleves en belægning på over 100 % i 19 dage, sammenlignes dette med oktober måned på figur 2.3 ses en belægning på 130 %. Der ses ligeledes en sammenhæng mellem de resterende måneder for de to grafer. Ud fra den anvendte data fremgår det ikke, hvor mange patienter, der udgør en belægningsgrad over 100 %, samt hvor længe de enkelte patienter er indlagt på afdelingen. Da belægningsgraden og antal dage kan variere for hver måned anses 18 måneder ikke som værende repræsentativ, for at kunne vurdere problemets omfang. Ud fra belægningsgraden kan det dog tyde på, at en effektivisering af planlægningen af patienter på ortopædkirurgisk afdelingen vil kunne medføre en balance i kapacitetsudnyttelsen.

2.3.1 Problemformulering

Hvordan kan indlæggelsesvarigheden for patienter på ortopædkirurgisk afdeling på Aalborg Universitetshospital forudsiges med henblik på at opretholde en kapacitetsudnyttelse på 100 %?

Problemløsning 3

På ortopædkirurgisk afdeling på Aalborg Universitetshospital ønskes der, på baggrund af afsnit 2.1, en 100 % kapacitetsudnyttelse med henblik på at udnytte de tilgængelige ressourcer optimalt. Kapacitetsudnyttelse er forholdet mellem aktivitet og kapacitet, hvoraf antal patienter er en betydende faktor ift. aktivitet. Det fremgår af afsnit 2.3, at belægningsgraden på ortopædkirurgisk afdelingen er varierende for hver måned. Ved en belægning over 100 % vil afdelingen opleve kapacitetsmangel, hvilket vil medføre, at personalet skal yde mere end afdelingen har kapacitet til. Derimod vil afdelingen ved en belægning under 100 % forårsage, at der ikke er fuld udnyttelse af personalets arbejdskraft. Derved oplever ortopædkirurgisk afdeling en ubalance i kapacitetsudnyttelsen. For at opnå en kapacitetsudnyttelse på 100 %, skal en ligevægt mellem aktivitet og kapacitet forekomme. Denne ligevægt kan tilnærmes ved at justere på en af faktorerne under aktivitet og kapacitet[Bames2015]. Det kan herunder forsøges at effektivere planlægningen af patienter og dertil forsøge at estimere indlæggelsesvarigheden for de indlagte patienter. På baggrund af dette opstilles følgende hypotese:

Indlæggelsesvarigheden for patienter kan benyttes til at effektivisere kapacitetsudnyttelsen.

3.0.1 Forudsigelse af indlæggelsesvarigheden

For at forbedre kapacitetsudnyttelsen på ortopædkirurgisk afdelingen undersøges det, hvorvidt indlæggelsesvarigheden kan forudsiges. Aalborg Universitetshospital har i et tidligere projekt indsamlet data fra 970 hospitalsindlæggelser fra ortopædkirurgisk afdeling. Datasættet er indsamlet gennem digitale patientjournaler i perioden 1. august til 31. oktober år 2014. Disse data er fordelt på 78 forskellige parametre. Flere af parametrene for patienterne er ikke ukendte, hvorfor præprocessing af datasættet anses nødvendigt. Ud fra parametrene fra datasættet vil indlæggelsesvarigheden forsøges estimeres.

Prædiktiv model

Til at estimere indlæggelsesvarigheden for patienter kan en prædiktiv model anvendes. Prædiktiv modellering definerer en proces, hvor en model udarbejdes med henblik på at forudsige hændelser. Denne model skal gøre det muligt at forstå og kvantificere nøjagtigheden af modellens forudsigelser ift. fremtidig data.[**Kuhn2013**] Inden for sundhedssektoren er det muligt at prædiktere forskellige former for hændelser og forløb ved anvendelse af disse modeller. Dette kan eksempelvis være en forudsigelse om, hvorvidt en patient, indlagt med hjertestop, har risiko for endnu et hjertestop, hvoraf vurderingen f.eks. baseres på demografi, kost samt kliniske målinger[**Hastie2008**].

Prædiktiv modellering kan opdeles i de to kategorier parametrisk og ikke-parametrisk. Forskellen på disse to kategorier ses overordnet ved om antallet af parametre kan varieres. Da flere parametrene i datasættet ikke er ukendte benyttes der ikke-parametrisk modellering.[**Sheskin2000**]

En ikke-parametrisk model kan opbygges som en matematisk model eller en computerbaseret databehandlingsmodel (på engelsk: computational models). Matematiske modeller er mindre optimale at avende ved ikke-parametriske datasæt¹. Af denne årsag fokuserer projektet fremadrettet på computerbaseret databehandling, da denne egner sig bedst til det tilgængelige ikke-parametriske datasæt². [**Sheskin2000**]

3.0.2 Databehandlingsmodel

Computerbaserede databehandlingsmodeller varierer fra matematiske modeller, da databehandlingsmodellerne kan være vanskelige at opstille ligninger for. I stedet anvendes ofte simulering til at bestemme en forudsigelse. Denne metode kaldes ligeledes en "black box"prædiktiv model, da modellen ikke giver et indblik i, hvordan den kommer fra input til output. Det er ikke muligt at få et indblik, da der ikke anvendes en entydig formel eller matematisk model, men derimod en model, der kan tage højde for manglende data samt data uden lineær sammenhæng. Dermed vil det være relevant at anvende databehandlingsmodeller ud fra den tilgængelige data for dette projekt, da der ikke nødvendigvis er en direkte lineær sammenhæng mellem data og indlæggelsestid. Dertil er ikke alle parametre registrerede for alle individer i datasættet. [**Kuhn2013**]

Supervised eller unsupervised learning

For at opstille en anvendelig algoritme er det nødvendigt at have et træningssæt. [**DIKU2010**]. Et træningssæt tager udgangspunkt i en datamængde, der enten har kendte eller ukendte labels. Systemets træningssæt er supervised, hvis input-output relationen er kendt, således det vides, hvilken parameter outputtet af systemet angives i. [**Brownlee2013**] Herudover findes ligeledes unsupervised learning, hvor indholdet af datasamples ikke har til formål at prædiktere en hændelse, men derimod

¹FiXme Note: HVORFOR? Dette skal uddybes

²FiXme Note: HVORFOR? Dette skal uddybes

finde en sammenhæng mellem data.[Brownlee2013, Kuhn2013] Formålet med den prædiktive model i denne rapport er at forudsige indlæggelsesvarigheden. Da indlæggelsesvarigheden er en del af det tildelte datasæt, er det hensigtsmæssigt at anvende modeller, der er baseret på supervised learning. Ved anvendelse af supervised learning er det muligt at frasortere parametre, der ikke har en statistisk indvirkning på det prædefinerede output parameter. Det er dermed muligt at udarbejde en model, der ekskluderer unødvendig data, da noget data kan have en negativ indvirkning på prædiktionen af indlæggelsesvarigheden. Dermed arbejdes der fremadrettet med modeller, der tager udgangspunkt i supervised learning.

Generering af manglende data

Ved manglende data i et datasæt findes der forskellige tilgange der kan kompensere for dette. Med udgangspunkt i [Saar2007] vurderes det, hvordan data med manglende indhold i enkelte parametre behandles. Der tages udgangspunkt i imputering, da det ikke er muligt at genskabe målingerne fra det originale datasæt. Ved at anvende Distribution-based Imputation (DBI), er det med manglende værdier muligt at vægte datasæt med manglende parametre, således at disse har en mindre betydning end de resterende målinger i datasættet under genereringen af et træningssæt. Årsagen til at de manglende parametre ikke udregnes vha. machine learning ud fra de resterende data, er, at de ikke nødvendigvis giver det rigtige udfald. F.eks. kan komorbiditeter være svære at bestemme, da datasættet ikke indeholder samtlige mulige komorbiditeter der findes. Dertil er det muligt at fjerne parametre fra hele træningssættet hvis hoveddelen af de registrerede patienter ikke har fået målt parameteren. Denne reduced feature model har flere fordele og ulemper som diskuteres i afsnit 3.0.2.

Reducering af parametre

Det kan være en fordel at fjerne enkelte parametre, hvis de er problematiske eller afviger fra de resterende data. Færre prædiktorer mindsker databehandlingstiden og kompleksiteten. Herudover kan kvaliteten af en prædiktiv model mindskes pga. problematiske variabler, der ikke indeholder information tilsvarende de resterende målepunkter.. Derved øges stabiliteten ved at fjerne disse problematiske variabler. Det kan eksempelvis skabe besvær ved lineær regression, da variablerne vil kunne forårsage fejl. Her vil data, der ikke indeholder nødvendig information kunne fjernes uden problemer.[Kuhn2013]

Generelt er det godt at undgå data med prædiktorer, der har høj sammenhæng, da de reelt er mere komplekse end den information de kan give. Disse vil ligeledes kunne være besværlige ved lineære regression og skabe ustabile modeller. Hvis to prædiktorer stort set er ens, betyder det at de indeholder samme information. For at kunne fjerne parametre kræves det, at antallet af unikke datapunkter skal være lavt ift. antallet af datasamples. Det vil ikke forringe modellens kvalitet ved at fjerne en af disse, men derimod øge hastigheden for prædiktionen og bruge mindre lagrings-

plads til systemet. Derfor kan en parameter som f.eks. BMI fjernes fra datasættet, hvis vægt og højde også indgår som parametre.[**Kuhn2013**]

