



AALBORG UNIVERSITET
STUDENTERRAPPORT

Prædiktiv model til kapacitetsudnyttelse

P5 Semestersprojekt - Efteråret 2016

Gruppe 5405

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	7
2.3	Belægningsgrad på ortopædkirurgisk afdeling	8
2.3.1	Problemformulering	10
3	Problemløsning	11
3.1	Forudsigelse af indlæggelsesvarighed	11
3.1.1	Præoperativt	11
3.1.2	Postoperativt	12
3.2	Prædiktiv model	12
3.2.1	Præprocessering	13
3.2.2	Forudsigelse før operation *BEDRE OVERSKRIFT*	13
3.2.3	Indsamling af data	14
3.2.4	Eksklusionskriterier for data	14
3.2.5	Opdatering af model	15
3.3	Machine learning	15

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 reducere 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

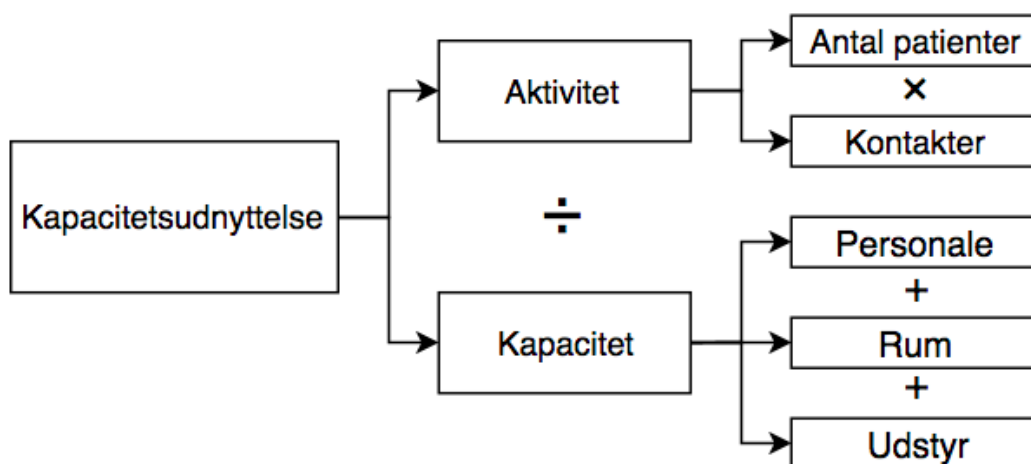
problemer på afdelingen?

Problemanalyse 2

2.1 Kapacitet

Kapacitetudnyttelse 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**]

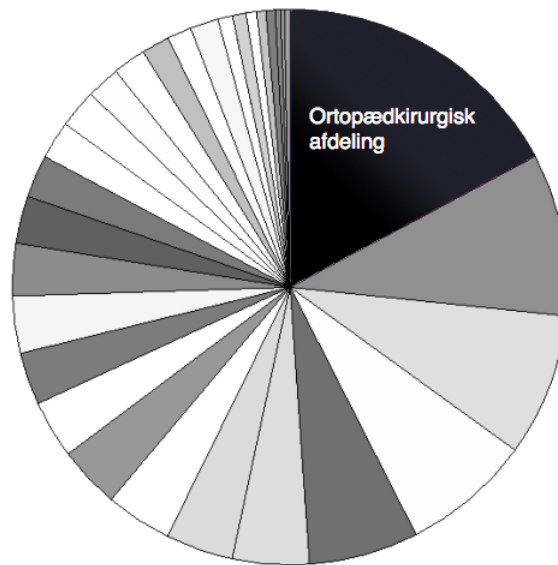
Belægning er defineret ud fra antallet af patienter, der er normeret til på en afdeling[**Heidmann2014**]. Når en 100 % belægning opnås, svarer dette til, at de disponible sengepladser på en afdeling er taget i brug. Ved en belægning på over 100 % betyder det, at der er flere patienter end afdelingen 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 Sundhedsstyrrelsen 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

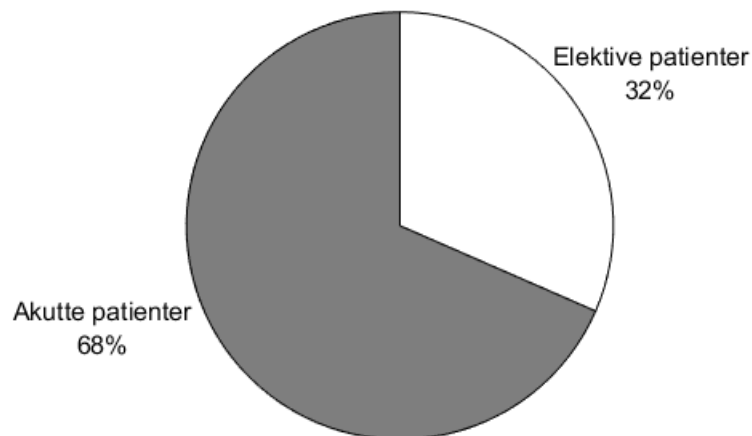
Patientindlæggelse

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] En fordeling af de elektive og akutte patienter fremgår af figur 2.3.

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

Akutte og elektive patienter på Aalborg Universitetshospital



Figur 2.3: *Elektive og akutte patienter på ortopædkirurgisk afdeling på Aalborg Universitetshospital over en tre måneders periode fra august til november år 2014.*

Af figur 2.3 ses det, at fordelingen mellem elektive og akutte patienter ikke er ligeligt fordelt på ortopædkirurgisk afdeling. Der ses over en tre måneders periode i år 2014, at de elektive patienter udgør 32 % og de akutte udgør 68 % af de samtlige patienter.

Elektive patienter indlægges i tidsrummet XX-XX og udskrives i tidsrummet XX-XX. Udskrivelsen af akutte patienter foregår i samme tidsrum. På afdelingen planlægges 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

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

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 arbejdesopgaver[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].

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]

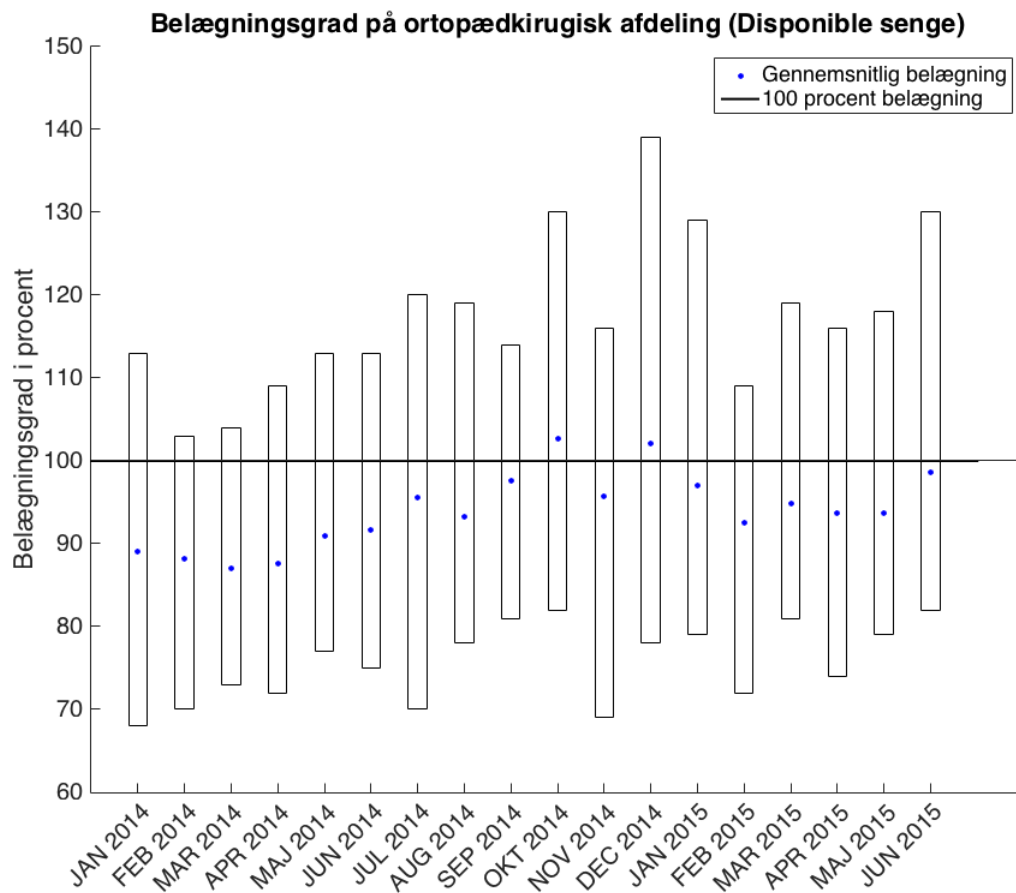
³FiXme Note: Spørgsmål til sygeplejersker: Hvordan prioriteres pauser under overbelægning?

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?

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

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 2.4 ses belægningsgraden fra år 2014 til 2015 på ortopædkirurgisk afdeling.[SDS2015]

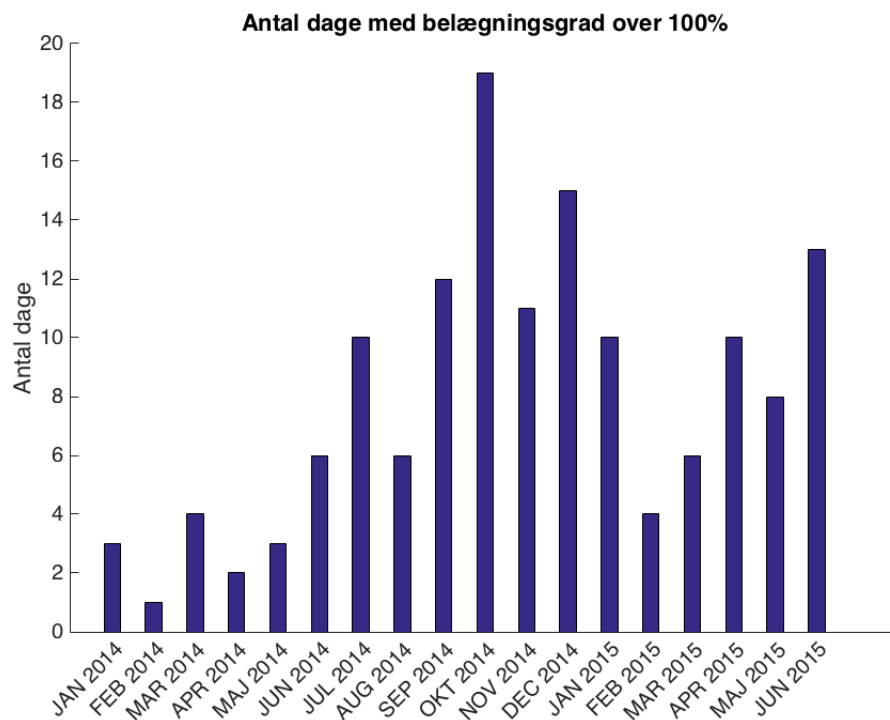


Figur 2.4: Belægningsgraden på ortopædkirurgisk afdelingen på Aalborg Universitetshospital målt over 18 måneder fra år 2014 til 2015. Søjlerne viser belægning ift. 100 %, hvortil maksimal og minimum belægning ligeledes illustreres. De blå punkter viser den gennemsnitlige belægning for hver måned.[SDS2015]

Det fremgår af figur 2.4, 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 2.4 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 2.5 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 2.4. [SDS2015]



Figur 2.5: 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 2.5, at der i oktober måned år 2014 opleves en belægning på over 100 % i 19 dage, sammenlignes dette med oktober måned på figur 2.4 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 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 blot én af faktorerene under aktivitet eller kapacitet[Bames2015]. Det kan herunder forsøges at effektivere planlægningen af patienter og dertil forsøge at estimere indlæggelsesvarigheden for de indlagte patienter. En planlægning kan resultere i strukturering af personalets arbejdsopgaver og bedre udnyttelse af disponible sengepladser. Planlægningen kan dog kun forekomme ved elektive patienter, da akutte patienters indlæggelse ikke kan forudsiges.

3.1 Forudsigelse af indlæggelsesvarighed

*** INDLEDNING *** * Hvorfor opdeles parametre i før og efter operation?

3.1.1 Præoperativt

Dette afsnit baseres primært ud fra interviewet på ortopædkirurgisk afdeling på Aalborg universitetshospital.

Hvilke parametre før en operation kan have indflydelse på varigheden af indlæggelsen? Er der nogle parametre som lægerne vægter højere end andre i estimeringen?

Derudover vil vi via. litteratur, grafer og statistiske analyser understøtte ovenstående. Vi har på nuværende tidspunkt fundet frem til, at operationstypen, patienttypen (akut/elektiv), alder, kombination af parametre og komorbiditeter har

indflydelse. Hertil har vi også fundet ud af parametre som det ikke er påvist har en indflydelse på indlæggelsesvarigheden.

I dette afsnit vil vi gerne analysere på, hvorledes faktorer såsom ledige sengepladser og kirurger kan have indflydelse på patienters indlæggelsesvarighed. Ved begrænset sengepladser og kirurger vil patienters operationer eventuelt udskydes.

3.1.2 Postoperativt

Den planlagte indlæggelsesvarighed for patienter kan ændre sig efter en operation, hvis der opstår komplikationer både løbende samt postoperativt. Hertil kan de opståede komplikationer tilføjes den prædiktive model, således det er muligt at estimere indlæggelsesvarigheden igen. Det ses, at patienter med forlænget indlæggelsesvarighed oftest er ældre, har flere komorbiditeter samt været igennem uforventede komplikationer.[**Krell2014**] Herudover ses en sammenhæng mellem længere operationstid og flere indlæggelsesdage.[**Kjeldsen2015b**]

Komplikationer under og efter operation (NY OVERSKRIFT)

3.2 Prædiktiv model

Til at estimere indlæggelsesvarigheden for patienter kan en prædiktiv model anvendes. Prædiktiv modellering er en model der udarbejdes med henblik på at forudsige hændelser. Denne model skal gøre det muligt at forstå og kvantificere nøjagtigheden af forudsigelsen ift. fremtidig data.[**Kuhn2013**] Kvantificeringen sker på baggrund af algoritmer.

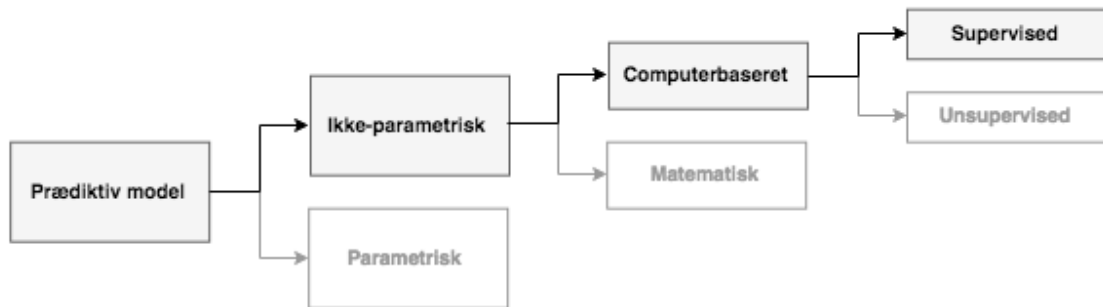
Inden for sundhedssektoren er det muligt at prædiktere forskellige former for hændelser og forløb. Dette kan eksempelvis være en forudsigelse om, hvorvidt en patient, indlagt med hjertestop, har risiko for endnu et hjertestop, hvoraf vurderingen f.eks. baseres på demografi, livsstil samt kliniske målinger[**Hastie2008**].

Prædiktiv modellering kan opdeles i de to kategorier: parametrisk og ikke-parametrisk. Parametrisk anvendes, når samtlige parametre er kendte, hvorimod ikke-parametriske benyttes, hvis én eller flere er ukendte.[**Sheskin2000**]

Den prædiktive model kan både anvende matematiske- og computerbaserede modeller. Matematiske modeller er en ligningsbaseret model, der forudsiger på baggrund af ændring i input. Herunder anvendes ofte regression, hvor der tages udgangspunkt i lineær sammenhæng. Computerbaserede modeller kræver ofte en simuleringsteknik til forudsigelse.[**MathWorks2016**]

Som tidligere nævnt sker kvantificering ud fra algoritmer. For at kunne udarbejde en algoritme kræves et træningssæt[**DIKU2010**]. Et træningssæt kan både være supervised eller unsupervised. Supervised learning er når indholdet af datasamples har til formål at forudsige en hændelse på baggrund af den kendte input-output relation[**Brownlee2013**]. Modsat er unsupervised learning, når indholdet af datasamples ikke har til formål at prædiktere en hændelse, men derimod finde en

sammenhæng mellem data[**Brownlee2013**, **Kuhn2013**]. På baggrund af ovenstående er figur 3.1 udarbejdet. Denne illustrerer de valg, der burde tages ift. prædiktive modeller.



Figur 3.1: Valg ift. prædiktiv modellering. De markerede felter illustrerer beslutningstagen.

Det fremgår af figur 3.1, hvilke modeller der bør anvendes ud fra datasættet fra ortopædkirurgisk afdeling. Da flere af parametrene i datasættet ikke fremgår bør der anvendes ikke-parametrisk modellering. Datasættet består af flere parametre, hvilket medfører, at det ikke ses hensigtsmæssigt at anvende ligningsbaseret modellering. Hertil kræves en simuleringsteknik, der anvendes under computerbaseret modellering. Der afgrænses til supervised learning, da datasættet indeholder input-output relation.

3.2.1 Præprocessering

Da flere parametre i datasættet fra ortopædkirurgisk afdeling på Aalborg Universitetshospital mangler, anses det nødvendigt at foretage præprocessering. Præprocessering foregår manuelt før en prædiktiv modellering kan foretages. Der findes flere metoder, der kan anvendes for at kompensere for de manglende parametre. Kompenseringen kan forekomme ved kassering af værdier, tilegne manglende værdier, reducering af værdier og imputering. Imputering opdeles i tre underkategorier: Prædiktiv værdi imputation, Distribution-baserede imputation og Unik-værdi imputation. Prædiktiv værdi imputation erstatter den manglende værdi med estimate-rede værdier. Distribution-baserede imputation vægter værdien af manglende data mindre end det resterende, således disse får en mindre betydning under generering. Unik-værdi imputation erstatter den manglede værdi med en vilkårlig værdi fra samme parameter.[**Saar2007**]

3.2.2 Forudsigelse før operation *BEDRE OVERSKRIFT*

Formålet med at konstruere en prædiktiv model er at estimere indlæggelsesvarigheden for patienter. Modellen er et redskab til sundhedspersonalet, der skal vurdere

indlæggelsesvarigheden når en operation planlægges og gennemføres. Den prædiktive model bør udarbejdes, da kapacitetsproblemer kan medføre en ubalance mellem kapacitet og aktivitet, som nævnt i afsnit 2.1. Brugen af modellen, indsamlingen samt indtastningen af data bør ikke være mere tidskrævende end personalets nuværende arbejdsopgaver. Ved en mere tidskrævende arbejdsprocedure vil der være mindre tid til de nuværende arbejdsopgaver, hvorfor ønsket om effektivisering af kapacitetsudnyttelse ikke vil blive afhjulpet.

**** NY INDLEDNING **** (OVERSTÅENDE KAN MÅSKE FLYTTES OP TIL PRÆDIKTIV MODEL) * Skal indeholde overskrifter for hele afsnittet

3.2.3 Indsamling af data

Den prædiktiv model skal designes specifikt til ortopædkirurgisk afdeling på Aalborg Universitetshospital, da der er bestemte parametre som f.eks. operationer, procedurer, udstyr, personale og medicinering. For at sikre, at indsamling af data er tilstrækkelig og relevant skal der opstilles stramme retningslinjer for, hvordan disse indsamles og noteres. På baggrund af afgrænsningen til supervised learning i afsnit 3.2, anses det nødvendigt, at hver indsamlede parameter kan have indflydelse for den enkelte patients indlæggelsesforløb. Disse overvejelser øger sandsynligheden for at opnå en brugbar prædiktiv model.

3.2.4 Eksklusionskriterier for data

For at udarbejde en prædiktiv model opstilles kriterier ift. dataindsamling ud fra formålet med modellen. Da nogle modeller er sensitive for strømlinet data er det vigtigt at bestemme, hvorvidt der skal opstilles eksklusionskriterier til dataindsamlingen eller om en forudbestemt prædiktiv model skal anvendes.[**Kuhn2013**] For at systemet kan sammenholde parametre er det nødvendigt, at data er indskrevet efter faste retningslinjer. Hvis der ikke opstilles faste retningslinjer for indskrevet data, kan dette have indflydelse på estimeringen af indlæggelsesvarigheden.[**Kuhn2013**]

Kategorisering af data

Ved at kategorisere data, anses det derfor muligt at lave en prædiktiv model med en mindre mængde data. Denne model er dog ikke så specifik som en model uden kategoriseret data, da en mulig variation i parametrene kan reduceres.[**Rowan2007**] Derfor bør det overvejes, hvor stor datamængden skal være for at konstruere modellens træningssæt. Ved ikke at kategorisere data, bliver modellen mere specifik ift. hver enkelt patient, men kræver dertil også en større database for at lave en funktionel model. Dette kan f.eks. inkludere meget specifikke kirurgiske indgreb eller sjældne komorbiditeter. Et eksempel på kategorisering af data kan være alder, hvor denne kan inddeles i aldersgrupper. En gruppe kan eksempelvis være 20-29 år hvilket kræver en mindre database, end aldersinddelingen 20-25 år, der udspænder et mindre interval.[**Rowan2007**]

3.2.5 Opdatering af model

En vigtig del af en prædiktiv model er, at denne kan tilpasse ændringer i parametres vægtning løbende ved ny data.[Kuhn2013] Ved ændring i medicinering eller procedure af kirurgiske indgreb kan prædiktering af det gældende indgreb give misvisende estimeringer af indlæggelsesvarigheden. Derfor kan modellen ved transitioner mellem procedureændringer blive invalideret og dermed skal den prædiktive model have inkorporeret mere data før denne kan forudsige indlæggelsesvarigheden. Hvis et datapunkt ikke kan kategoriseres i systemet, kan det være nødsaget at ekskludere denne data fra indskrivelse i databasen.

3.3 Machine learning

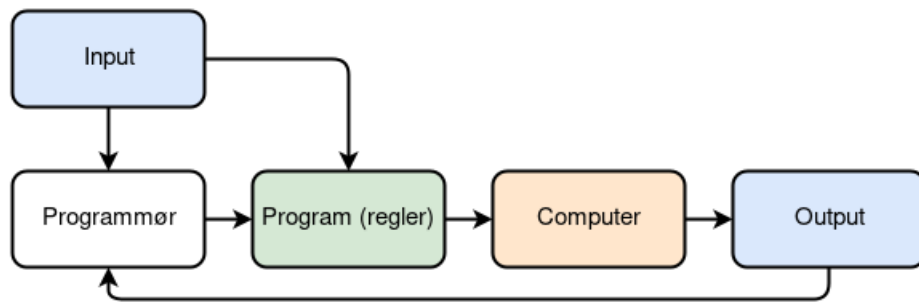
**** DET KAN VI MEGA MEGET BRUGE! **** Machine learning er en underkategori af kunstig intelligens og er en metode, der kan anvendes til at gøre en computer i stand til at genkende mønstre i data. Derved er det muligt at bearbejde datasæt af en størrelse, der er umulige for mennesker at behandle. Det bruges uset mange steder og har mange formål, som f.eks. søgemaskiner, forudsigelse af aktiekurser og ansigtsgenkendelse.[DIKU2010]¹

En konventionel computeralgoritme arbejder efter et input og et regelsæt, der afgør, hvordan dette input skal behandles. Dette resulterer i et output, begrænset af programmørens evne til at opstille regler. *Ligegyldigt*

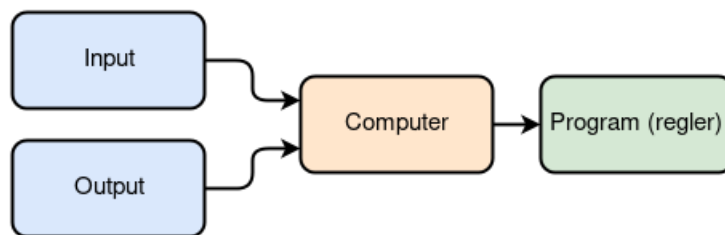
Machine learning arbejder derimod typisk efter et input, hvilket kaldes data, og et output, hvilket kaldes label. Det betyder at machine learning selv skaber reglerne for hvordan der kommer fra data til label. Når systemet har et solidt regelsæt kan det selv lave labels til ny lignende data. Forskellen er illustreret på figur 3.2. *FORSKELLEN ER IKKE VIGTIG! MEN VI KAN SAGTENS BRUGE DET OM ML! Dette gøres i en logisk proces kaldet induktiv inferens, hvilket er en metode, hvor computeren tilnærmer sandsynligheden for et givent output, ud fra en sammenligning med lignende eksempler. Der findes forskellige måder at gøre dette på, men fælles for alle undersøges det, hvorvidt den nye data ligner den eksisterende data.[DIKU2010]

¹FiXme Note: måske finde et andet ord for uset

Traditionel programmering



Machine learning



Figur 3.2: Det traditionelle programmeringsparadigme i forhold til machine learning *SKÆR LIDT*

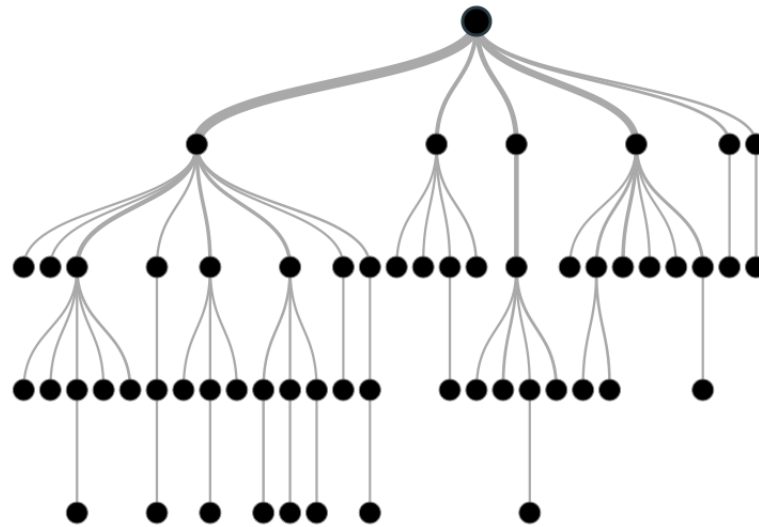
Figuren viser forskellen mellem den traditionelle tilgang til algoritmedesign i forhold til machine learning.

Når en algoritme designs til machine learning gøres dette på baggrund af et træningssæt. [DIKU2010] Træningssættet tager udgangspunkt i en datamængde, der enten har kendte eller ukendte labels. Hvis indholdet af de forskellige samples i datasættet har en kendt label er der tale om supervised learning, altså opsætning af en genkendelse hvor systemet guides i den ønskede retning. [Brownlee2013] Dertil findes også unsupervised learning hvor indholdet af data samples ikke har nogen label. Ved unsupervised learning undersøges matematiske sammenhænge og tendenser. [Brownlee2013] I denne rapport fokuseres der fremadrettet kun på supervised learning.

*****DET HAR VI!*****

Ved supervised learning forsøges det at udarbejde en model, hvor data tilknyttes til labels og det bliver klassificeret. Det er vigtigt, at modellen afbildes så korrekt som muligt, for både træningssættet og de nye data. En af grundene til at denne proces er vigtig, er at den så kan generalisere, men samtidig kan blive stærkere og bedre rustet når der kommer ny data ind. Machine learning er i princippet en algoritme, som arbejder ud fra hypoteser og generalisering af data. Desto flere data den har at arbejde med igennem tiden desto bedre bliver træningssættet også, hvilket resulterer i lav risiko for at få forkerte labels ud. På denne måde bliver træningssættet bedre, og algoritmen bliver bedre til at rubicere nye data.[DIKU2010] Hvordan systemet træffer en beslutning kan variere mellem algoritmer, men overordnet kan det visualiseres som et beslutningstræ; et forsimplet, ensrettet flowchart som vist

på figur 3.4, hvor hvert punkt er en indsnævring mod den label der passer bedst på det pågældende data.[Barber2012]

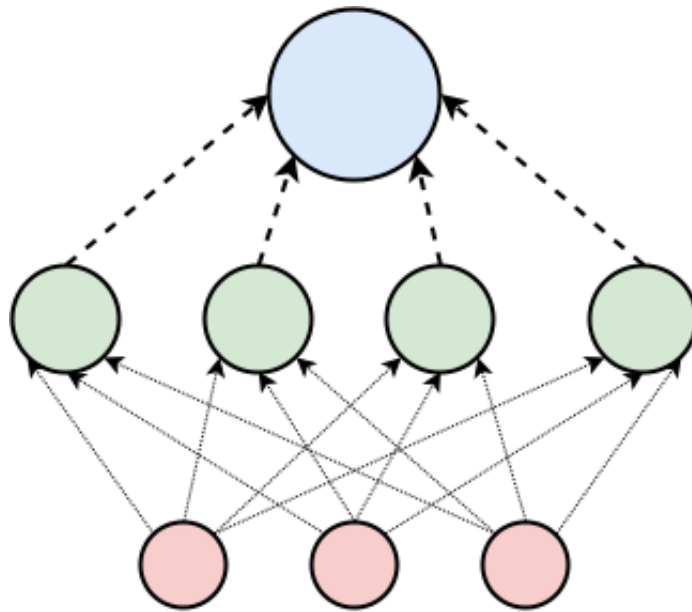


Figur 3.3: Visualisering af machine learning beslutningstræ

Figuren viser en forsimplet visualisering af et vilkårligt beslutningstræ fra en machine learning algoritme.[Saraswat2016]

Der findes flere metoder at udføre supervised learning på. En af de første, og stadig populære, er neurale netværk. Her tages der udgangspunkt i et netværk af celler, hvor disse celler får input hvor de i sidste ende summerer deres givne input til en neuron, hvilket kan ses på nedenstående figur. (skal lige have figuren ind) Fra at cellen får et input til der fås et output fra neuroner, forgår der en proces mellem de to stadier. Processen her kaldes hidden units, der er egentlig det fundamentale og kan næsten sammenlignes med hjernen. Dataene bliver klassificeret undervejs, af de enkelte celler, hvilket betyder at netværket minimeres for fejl, samt at generaliseringsevnen vil blive stærkere. Det gælder om for netværket, at få opbygget et struktur, så den kan forudsige de aspekter den bliver spurgt ind til. [DIKU2010] En af måderne netværket vil blive stærkere igennem tiden er, at den anvender begrebet feedforward netværk, der betyder at når netværket bliver anvendt bliver den trænet, samt også efter den er blevet anvendt. Dette skyldes blandt andet en mekanisme kaldt backpropagation. Netværket går ind og sammenligner det output den får ud sammenlignet med det der blev ønsket, og herefter kan der ud fra sammenlignings grundlaget gives feedback til de hidden units og output unitsene. ² Et eksempel på et neuralt netværk kan ses på figur ??.

²FiXme Note: Har lagt kilden op i mendeley og hedder david, shai shalev- kunne ikke lige finde det navn man skulle citerer til



Figur 3.4: Neuralt netværk med skjult lag

På billedet ses et neuralt netværk med skjult lag. Datainput sker i de røde neuroner, beregnes i det skjulte grønne lag, hvorefter det outputtes i den blå neuron.

Andre metoder kan også anvendes til at udføre supervised learning. Support Vector Machines (SVM) er en populær algoritme til særligt praktiske problemer. Boosting-algoritmer, som f.eks. AdaBoost er også populære og anvender et princip hvor mange simple klassifikationer samles i en flertalsafstemning, frem for ét beslutningstræ per datapoint.