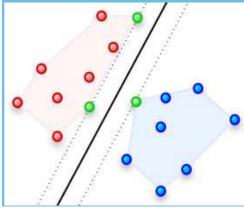
Klassifikation af Årsag til Utilsigtede Hændelser

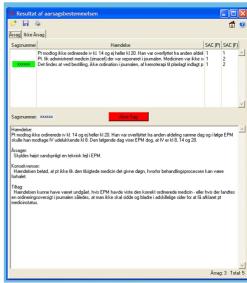
-baseret på Statistisk Natural Language Processing

Hændelse 1: En sygeplejerske i nattevagt kontakter forvagten telefonisk med henblik på ordination af smertestillende til en 68-årig indlagt kvinde. Lægen ordinerer 10 mg petidin i.v. Først da sygeplejersken har givet patienten medicinen, opdager hun, at der i EPM står, at patienten har cave morfin. Sygeplejersken undrer sig over ikke at være blevet advaret, da hun dispenserede petidin i EPM.

Hændelse 2: En indlagt 55-årig kvinde har registreret CAVE metronidazoltabletter iEPM. Under indlæggelsen bliver der ordineret metronidazol i.v. EPM advarer hverken ved ordination, dispensering eller administration.

Hændelse 3: En sygeplejerske opdager "tilfældigt", at et barn med cave aminoglycosid er registreret i EPM med "ingen cave". Cave-oplysningen iEPM er fire måneder gammel (fra tidligere indlæggelse) og er ikke blevet opdateret, da man i mellemtiden blev







Marie Juul Hansen & Nana Østergaard Rasmussen, Gruppe 1087a Sundhedsteknologi, Aalborg Universitet Afgangs-speciale Forår 2008

Aalborg Universitet

Institut for Sundhedsvidenskab og Teknologi

Titel:

Klassifikation af Årsag til Utilsigtede Hændelser - baseret på Statistisk Natural Language Processing

Tema:

Sundhedsteknologi

Projektgruppe:

Gruppe 1087a

Gruppemedlemmer:

Marie Juul Hansen Nana Østergaard Rasmussen

Vejledere:

Stig Kjær Andersen Christian Nøhr

Semester:

Speciale 10. Semester

Antal kopier: 5

Antal sider: 119

Synopsis:

I 2007 blev der i Danmark indrapporteret 23.521 utilsigtede hændelser. Formålet med rapportering af utilsigtede hændelser er, at få udarbejdet retningslinier for hvorledes disse forebygges, hvorved der er potentiale for at redde menneskeliv. Rapporterne indrapporteres i fritekst, hvilket på nuværende tidspunkt kræver gennemlæsning af en risikomanager for, at kortlægge hvad der er årsag til den utilsigtede hændelse. I dette projekt er det undersøgt, om de utilsigtede hændelser automatisk kan årsagsklassificeres ved brug af statistisk natural language processing. Dette er gjort ved at udvikle et system som 'proof-ofconcept'. Systemet består af en brugergrænseflade og en klassifikationsmodel. Klassifikationsmodellen er trænet og testet på i alt 132 utilsigtede hændelsesrapporter, der alle indeholder søgeordet 'EPM', og har til formål at klassificere om EPM er årsag til den utilsigtede hændelse eller ej. Dette er gjort ved konstruktion af en model baseret på en kombination af a priori viden fra eksperter og statistisk viden.

Resultat af klassifikationen viser et F-mål på 0,946 for de rapporter, der klassificeres som ikke årsag og et F-mål på 0,667 for de rapporter, der klassificeres som årsag. Disse resultater kan forbedres ved træning af modellen på flere utilsigtet hændelses rapporter. Det konkluderes, at statistisk natural language processing kan anvendes til klassifikation af utilsigtede hændelser, hvis der er integreret a priori viden i forhold til det konkrete søgeord.

Aalborg University

Department of Health Science and Technology

Title:

Classifying the Reason for Adverse Events - based on Statistical Natural Language Processing

Subject:

Biomedical Engeneering and Informatics

Project Group:

Group 1087a

Group Members:

Marie Juul Hansen Nana Østergaard Rasmussen

Supervisors:

Stig Kjær Andersen Christian Nøhr

Semester:

Master Thesis 10^{th} Semester

Publications: 5

Pages: 119

Abstract:

In year 2007, 23.521 adverse events were reported in Denmark. The purpose of reporting adverse events is to create guidelines for how to prevent future events, and hence save lives. The reports are written in natural language. It is time consuming for the risk manager to read through all of the reports carefully, to locate the reason for the adverse event.

This project examines if it is possible to classify the adverse events, based on the reason for the event by using statistical natural language processing. This is done by developing a system as a proof of concept. The system consists of a user interface and a classification model. The classification model is trained and tested on 132 adverse event reports, all containing the keyword 'EPM', where the purpose of the model is to classify whether EPM is the reason for the adverse event or not. This is done by creating a model based on a combination of a prior knowledge from domain experts and statistical knowledge.

The classification results in an F-measure at 0.946 for the reports, where EPM is not the reason, and an Fmeasure at 0.667 for the reports, where EPM is the reason for the adverse event. These results can be improved by training the model on more adverse event reports.

It can be concluded that statistical natural language processing can be used in classifying adverse events if a prior knowledge is included in relation to the specific keyword.

Forord

Denne rapport er udarbejdet af projektgruppe 1087a på 10. semester på Institut for Sundhedsvidenskab og Teknologi indenfor specialiseringen Medicinsk Informatik på Aalborg Universitet. Rapporten er udarbejdet i perioden fra d. 1. februar til d. 2. juni 2008, under temaet Sundhedsteknologi.

Rapporten henvender sig til vejledere, medstuderende og andre med interesse i Natural Lnaguage Processing, Utilsigtede hændelser og klassifikation af disse. Desuden henvender rapporten sig til Enhed for Patientsikkerhed ved Region Hovedstaden, idet projektet er udarbejdet i samarbejde med ansatte ved Enhed for Patientsikkerhed. I forbindelse med dette
samarbejde ønsker vi at rette en stor tak til Læge og Risikomanager Jonas Egebart og Brian Bjørn ved Enhed for Patientsikkerhed for et godt samarbejde samt bidrag af utilsigtede
hændelsesrapporter. Ligeledes takkes de for deres deltagelse i forbindelse med fastlæggelse
af problemstillingen samt brugervenlighedstest af det udviklede system. Yderligere ønsker
projektgruppen at takke Center for Sprogteknologi ved Københavns Universitet, herunder
Softwareudvikler Bart Jongejan, for lån af Natural Language Processing værktøjer.

Marie Juul Hansen	Nana Østergaard Rasmussen

Læsevejledning

Rapporten er inddelt i 13 kapitler, hvor projektgruppen anbefaler, at rapporten læses samlet og fortløbende. Til rapporten hører en appendiksdel indeholdende 5 kapitler. Disse kapitler indeholder supplerende oplysninger, som understøtter rapportens indhold. Der vil igennem rapporten blive refereret til disse appendikskapitler på steder, hvor det kunne være relevant med uddybende oplysninger. Ligeledes er der til rapporten vedlagt et bilag i form af Region Hovedstadens rapporteringsskema til utilsigtede hændelser.

Appendiks A indeholder en beskrivelse af hvorledes rapporter med utilsigtet hændelser (UTH) risikovurderes. Appendiks B giver en introduktion til NLP, og hvilke sprogniveauer der kan arbejdes med indenfor NLP. Appendiks C præsenterer resultaterne af standardmetoden til klassifikation af dokumenter. Appendiks D viser hvilke features, der anvendes til klassifikation af UTH-rapporterne samt et udpluk af de dertilhørende featurevektorer. Endeligt indeholder appendiks E et referat af brugertesten. Referat af afholdte interviews er at finde på følgende hjemmeside: $www.hst.aau.dk/\tilde{0}8gr1087a$.

Rapporten er baseret på eksisterende litteratur, interviews og information erhvervet igennem Internettet. Disse kilder henvises til efter Harward-metoden med betegnelsen [Efternavn, Årstal]. Når en henvisning er placeret inden et punktum i en sætning, refererer henvisningen til den pågældende sætning. Henvisninger efter et punktum refererer til hele foregående afsnit. Da rapporterne med utilsigtede hændelser kan indeholde patientfølsomme oplysninger, er sagsnummeret for den enkelte hændelse erstattet med et andet identifikationsnummer, således at det ikke er muligt at spore den enkelte hændelse. Ligeledes er det af hensyn til rapporternes fortrolighed valgt ikke at vedlægge kildekoden til systemet.

Indhold

1	\mathbf{Ind}	ledning	3
2	Ana	alyse af UTH-Rapporteringsforløb	5
	2.1	Rapportering af UTH i Danmark	5
	2.2	Rapporteringsskemaets Opbygning	6
	2.3	Sagsbehandling af UTH-Rapporter	9
	2.4	Problemformulering	11
3	Me	\mathbf{tode}	13
4	NL	P-System	17
	4.1	Kravsspecifikation	17
	4.2	Systemets Funktionaliteter	19
5	Ana	alyse af UTH-Rapporter	23
	5.1	Beskrivelse af UTH-Rapporter	23
	5.2	Karakteristika for UTH-rapporter	25
6	Ud	vikling af EPM Klassifikationsmodel	29
	6.1	NLP-Præprocessering	30
	6.2	Featureudvælgelse	31
	6.3	Klassifikationsalgoritmer	36
7	Ud	vikling af Brugergrænseflade	43
	7.1	Brugervenligt Design	44
	7.2	Systemets Brugergrænseflader	46
8	Inte	egration af NLP-System	51

INDHOLD

	8.1	Integration af Model og Brugergrænseflade	52
	8.2	Eksekvering af Systemet	52
	8.3	Systemets Plads i Rapporteringsforløbet	53
9	Gen	erel Klassifikationsmodel	55
J	9.1	A Priori Mønstre	55
	9.2	Statistiske Mønstre	56
	5.2	Suddistribute rightster.	90
10	Test	og Resultater	59
	10.1	Test af EPM Klassifikationsmodel	60
	10.2	Systemtest	64
	10.3	$Bruger ven lighed stest \ \ldots \ \ldots$	69
11	Disk	xussion —	79
	Dist	A CONTRACTOR OF THE CONTRACTOR	
12	Kon	klusion	85
13	Pers	spektivering	87
${f Li}$	tter	atur	91
Α.	pper	ndiks	94
	PP°		0 1
A	Risi	kovurdering af Utilsigtede Hændelser	95
В	Nat	ural Language Processing	99
_			
С	Bag	of Words Algoritme	105
D	\mathbf{Udt}	rukket Features	109
${f E}$	Tær	nke Højt Test	113
		-	
Bi	lag		117
	-~-8		
A	\mathbf{Reg}	ion Hovedstadens Rapporteringsskema	117

Kapitel 1

Indledning

Aktiviteter i sundhedsvæsenet, der involverer arbejdet med patienter indebærer en risiko for patientskade. Dette skyldes, at patientpleje er baseret på mennesker, og i takt med stigende kompleksitet og tværfaglighed i sundhedsvæsenet kan det ikke undgås, at mennesker begår fejl. [Schiøler 2001] De fejl, patienter kommer ud for under deres ophold på hospitalet, kaldes utilsigtede hændelser (UTH). UTH defineres som [sikkerpatient.dk 2008]:

En ikke-tilstræbt begivenhed, der skader patienten eller indebærer risiko for skade som følge af sundhedsvæsenets handlinger eller mangel på samme.

Den første danske undersøgelse omhandlende UTH blev foretaget i 2001 og viste, at 9% af alle patienter, der udskrives, har været udsat for en UTH under deres ophold på sygehuset. En UTH kan resultere i forlænget indlæggelse, mén eller død forårsaget af fejl, eller komplikationer, som ikke har med deres tilgrundliggende sygdom at gøre. [patientsikkerhed.dk 2008a] Udenlandske undersøgelser [Brennan et al. 1991, Hayward & Hofer 2001] oversat til danske forhold viser, at der årligt kan undgås 140 dødsfald, som alle er direkte forårsaget af UTH [Schiøler et al. 2001]. I 2007 blev der rapporteret 23.521 UTH til Dansk Patientsikkerhedsdatabase (DPSD) [dpsd.dk 2008], hvoraf en undersøgelse viser, at 4 ud af 10 utilsigtede hændelser kunne have været forebygget svarende til 9.408 hændelser [patientsikkerhed.dk 2008a].

Rapportering af UTH blev gjort lovpligtig ved Lov om Patientsikkerhed, der trådte i kraft d. 1. januar 2004 [retsinformation.dk 2007]. Loven om patientsikkerhed har til formål at forbedre patientsikkerheden i sundhedsvæsenet. Dansk Selskab for Patientsikkerhed definerer patientsikkerhed som [patientsikkerhed.dk 2008b]:

Sikkerheden for patienter mod skade og risiko for skade som følge af sundhedsvæsenets indsats og ydelser eller mangel på samme.

Patientsikkerhed er en forudsætning for en høj kvalitet i sundhedsvæsnet, hvor ét af midlerne til at sikre patienterne er rapportering, analyse og læring af UTH.

Følgende UTH er obligatoriske at indberette [Region Hovedstaden 2008];

- 1. Klinisk betydende hændelser opstået i forbindelse med medicinering.
- 2. Klinisk betydende hændelser opstået i forbindelse med kirurgiske og invasive indgreb.
- 3. Andre hændelser af alvorlig karakter, for eksempel når hændelsen resulterer i betydelig øget udredning eller behandlingsintensitet.

Overordnet kan UTH inddeles i skadevoldende og ikke-skadevoldende hændelser. Dernæst kan de inddeles i egentlige hændelser og nærhændelser, hvor en nærhændelse er en fejl, der aktivt bliver afværget før handlingen. Herudover er der yderligere en inddeling i henholdsvis komplikationer og fejl, hvor komplikationer defineres som en optræden af nye sygdomsprocesser, der er relateret til en tilstedeværende sygdom, og fejl defineres som en mangelfuld gennemførelse af en plan eller valg af forkert plan. [Sygehusfællesskab 2005] Sammenhængen imellem disse begreber fremgår af figur 1.1, hvoraf det ses, at det er de forebyggelige hændelser, der ønskes rapporteret.

En præcis viden om forekomsten og karakteren af UTH i sundhedsvæsenet er afgørende

Utilsigtede hændelser					
Skadevoldende u	hændelse				
I	Egentlige hændelse	er	Nær-hæ	endelser	
Komplikationer		Fe	ejl		
Ikke forebyggelige	Forebyggelige	Forebyggelige Ingen skade da Patientten var robust	Forebyggelige Blev forhindret ved et tilfælde	Forebygget På grund af velfungerende sikkerhedskultur	
En patient, der er fuldt orienteret, oplyser at kunne täle penicillin. Patienten får en injektion med dette lægemiddel og udvikler anafylaktisk chok.	En patient, der har oplyst ikke at kunne tåle penicillin, får en injektion med detle lægemiddel og udvikler anafylaktisk chok.	En patient, der har oplyst ikke at kunne tåle penicillin, får en injektion med dette lægemiddel. Det medfører ingen udtal allergisk reaktion.	Patienten har oplyst ikke at kunne tåle penicillin. Lige før lægemidlet er på vej til at blive njiceret, siger en pårørende, der netop er kommet til stede, at patienten ikke tåler penicillin, hvorfor hændelsen afværges.	En patient har oplyst ikke at kunne täle penicillin, og dette er registreret i det elektroniske medicinordinationssy stem, idet lægen alligevel ordinerer penicillin, fremkommer en advarsel, hvorfor ordinationen ændres.	
L	Rapporteringspligtige hændelser				

Figur 1.1: Oversigt over begreber i relation til utilsigtede hændelser, og hvilke der skal indrapporteres. [Region Hovedstaden 2008]

for forebyggelse af hændelserne. [Lipczak & Schiøler 2001] Denne viden opnås kun ved, at UTH bliver rapporteret, hvilket sikres da patientsikkerhedsloven gør det lovpligtigt for sundhedspersoner at rapportere både hændelser, de selv har været involveret i, og hændelser, de har observeret. Herved kan der opnås viden om i hvilke situationer, der sker UTH i sygehusvæsenet og dermed højnes muligheden for forebyggelse, idet der kan uddrages læring og iværksættes forebyggende initiativer. [Region Hovedstaden 2008] For at kunne øge patientsikkerheden er det dermed vigtigt, at rapporterne bliver anvendt korrekt. Dette leder frem til følgende initierende problemstilling:

Hvordan anvendes UTH-rapporter med henblik på at øge patientsikkerheden?

Kapitel 2

Analyse af UTH-Rapporteringsforløb

I dette kapitel analyseres hvorledes rapporteringsskemaet til indrapportering af utilsigtede hændelser er opbygget, samt hvorledes sagsbehandlingen af rapporterne finder sted.

Formål med UTH-rapporterne er at få at udarbejdet retningslinier med henblik på at forebygge UTH. Anvendelse af informationer til udarbejdelse af retningslinier kaldes også sekundær brug af klinisk information, hvor primær brug ville indebære en form for realtidshandling på baggrund af informationerne, hvilket ikke er tilfældet for UTH-rapporter.

For at opfylde formålet med UTH-rapporterne er det nødvendigt at analysere rapporterne og dermed opnå viden om, hvilken fejl der er sket. For at en lignende hændelse skal kunne forebygges, er det nødvendigt at få identificeret tre aspekter i forhold til hændelsen 1) hvad hændelsen indebar, 2) årsagen til hændelsen og 3) konsekvenserne af hændelsen. Når disse tre aspekter er identificeret og analyseret kan den opnåede viden anvendes til sekundær brug. Den sekundære brug inkluderer planlægningsmæssige formål og kvalitetsarbejdet samt klinisk- og epidemiologisk forskning [Safran et al. 2007]. Den sekundære brug udmunder i direkte respons til rapportøren af hændelsen baseret på den enkelte rapport eller udarbejdelse af retningslinier baseret på flere rapporter.

Analysen af hændelserne kan være en tidskrævende og langsommelig proces, afhængigt af måden hvorpå rapporterne er struktureret og hvordan informationerne i rapporterne er repræsenteret. For at optimere måden, hvorpå rapporterne behandles og dermed øge patientsikkerheden, er viden om strukturen af rapporterne nødvendig.

2.1 Rapportering af UTH i Danmark

UTH opstået i forbindelse med patientens behandling eller ophold på sygehuset skal ifølge dansk lovgivning rapporteres til den region, hvori hændelsen fandt sted. Rapporteringen skal fortages umiddelbart efter hændelsen er sket og senest syv dage herefter. I tilfælde af, at sundhedspersonen først senere bliver klar over, at en UTH har fundet sted, skal rapporteringen ske hurtigst muligt. [retsinformation.dk 2007] De sundhedspersoner, der rapporterer hændelserne, kan vælge at være anonyme, men hvis de vælger ikke at være det, kan de som følge af de rapporterede fejl ikke blive pålagt nogen form for sanktioner. [Hansen $2005\,b$] De offentlige sygehuse skal sende rapporterne videre til den patientsikkerhedsansvarlige i den

region, de tilhører, og ved de private sygehuse skal rapporteringen ske til en bestemt person på sygehuset. Efter rapporterne er sendt til regionerne, skal de gennemlæses og følge en standard for videre rapportering til Sundhedsstyrelsen. Rapporteringen til regionen skal udføres enten elektronisk via rapporteringssystemets hjemmeside www.dpsd.dk eller ved udfyldelse og indsendelse af en tilsvarende blanket udarbejdet af Sundhedsstyrelsen. [retsinformation.dk 2007]

Da patientsikkerhedsloven trådte i kraft, var der allerede, på daværende tidspunkt, tiltag for rapportering i enkelte amter. Disse amter, nu regioner, arbejder videre med deres egne rapporteringsskemaer. [sikkerpatient.dk 2008] Det adskiller sig ikke fra, at alle regioner og privat sygehuse skal rapportere hændelserne på samme vis til Sundhedsstyrelsen. Det tidligere H:S er blandt de sygehuse, som over længere tid har beskæftiget sig med rapportering af UTH, og derfor har deres eget rapporteringssystem. Dette tiltag er nu overført til Region Hovedstaden, hvilket har ført til en handlingsplan 'Patientsikkerhed i Region Hovedstaden' [Region Hovedstaden 2008]. Region Hovedstaden anvender både elektronisk og manuel rapportering af hændelserne, fælles er at de indeholder de samme informationer. Typen af rapporteringsskema afhænger af om det pågældende sygehus tidligere har tilhørt H:S. Hvis det har det har de deres eget skema, hvorimod hvis sygehuset ikke kommer fra det H:S anvendes rapporteringsskemaet på www.dpsd.dk [Egebart & Bjørn 2008]. I resten af denne rapport tages udgangspunkt i UTH-rapporterne og arbejdsgangene, som de forløber på det tidligere H:S Sygehus Hvidovre Hospital i Region Hovedstaden.

2.2 Rapporteringsskemaets Opbygning

Selve rapporteringsskemaet er delt op i otte dele 1) Hvornår skete hændelsen 2) Hvor skete hændelsen 3) Kontaktoplysninger 4) Hændelsesbeskrivelse 5) Årsager til hændelsen 6) Konsekvens af hændelsen 7) Forslag til tiltag og 8) Patientoplysninger. Rapportskemaets opbygning ses på skemaet på figur 2.1 og 2.2.

Rapporteringsformen er semistruktureret, idet fire af delene er struktureret med klart definerede felter, hvorimod de resterende fire dele er fritekstfelter. Alle disse felter skal udfyldes når rapporten indrapporteres. Dette er en fremstillingsform, som også anvendes ved for eksempel U.S Food and Drug Administration [fda.gov 2007] til rapporteringer af utilsigtede hændelser. I Region Hovedstaden anvendes samme struktur som i rapporteringsskemaet fra DPSD. Indholdet i rapporten er følgende:

Hvornår skete det? Denne del er struktureret med angivelse af dato, måned og år samt cirka klokkeslæt. Under afsnittet står der: 'Angiv dato og tidspunkt for hændelsen. Hvis hændelsen strækker sig over længere tid, angiv da starttidspunkt. Hvis du ikke er sikker på tidspunktet, angiv da dit bedste skøn eller tidspunktet hvor hændelsen blev opdaget.'.

Hvor skete det? Denne del er struktureret med angivelse af felterne; region, sygehus og afdeling. Yderligere er det muligt i et fritekstfelt at angive supplerende oplysninger om afdelingen. Under punktet står der: 'Angiv hospital og afdeling, hvor hændelsen fandt sted. Hvis flere steder er involverede, angiv da stedet hvor hændelsen begyndte.'.

Hvem er du? Her kan angives, navn, mailadresse, telefonnummer og rolle. Under punktet

står der: 'Skriv dit navn og evt. e-mail-adresse eller telefonnummer herunder. Hvis du vælger ikke at oplyse dit navn, kan det være vanskeligt at bruge din rapport, fordi det så ikke vil være muligt at indhente yderligere oplysninger om hændelsen. Din rapport bliver behandlet fortroligt.'.

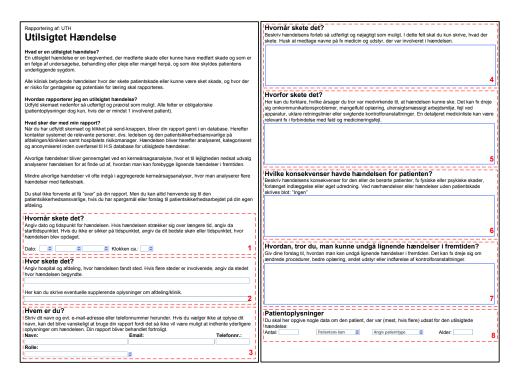
- Hvad skete der? Er et fritekstfelt, hvori der står; 'Beskriv hændelsens forløb så udførligt og nøjagtigt som muligt. I dette felt skal du kun skrive, hvad der skete. Husk at medtage navne på fx medicin og udstyr, der var involveret i hændelsen.'. Bortset fra denne tekst er der ingen retningslinjer for, hvad dette felt skal indeholde. Det kan være alt fra en enkelt sætning til en flere sider lang beskrivelse af hændelsen, afhængigt af den sundhedsperson, der rapporterer hændelsen.
- Hvorfor skete det? Er ligeledes et fritekstfelt uden begrænsninger, blot at der minimum skal skrives et tegn. Feltets formål er at angive, hvad der kunne være årsag til den utilsigtede hændelse. Under punktet står der: 'Her kan du forklare, hvilke årsager du tror, var medvirkende til, at hændelsen kunne ske. Det kan dreje sig om kommunikationsproblemer, mangelfuld oplæring, uhensigtsmæssigt arbejdsmiljø, fejl ved apparatur, uklare retningslinier eller svigtende kontrolforanstaltninger. En detaljeret medicinliste kan være relevant fx i forbindelse med fald eller medicineringsfejl.'.
- Hvilke konsekvenser havde hændelsen for patienten? I dette fritekstfelt står der; 'Beskriv hændelsens konsekvenser for den eller de berørte patienter, fx fysiske eller psykiske skader, forlænget indlæggelse eller øget udredning. Ved nærhændelser eller hændelser uden patientskade skrives blot: Ingen'. Minimum ét tegn skal være til stede for at kunne indsende skemaet.
- Hvordan, tror du, man kunne undgå lignende hændelser i fremtiden? Det sidste af fritekstfelterne er ikke obligatorisk at udfylde, der står; 'Giv dine forslag til, hvordan man kan undgå lignende hændelser i fremtiden. Det kan fx dreje sig om ændrede procedurer, bedre oplæring, andet udstyr eller indførelse af kontrolforanstaltninger.'.
- Patientoplysninger Består af strukturerede felter med patientens alder, køn og type. Patienttypen er kategoriseret i fem kategorier, hvor en af dem skal afkrydses; indlagt, dagpatient, ambulant, skadestue, ingen patient involveret og ved ikke. Ligeledes skal antallet af patienter angives. Under punktet står der: 'Du skal her opgive nogle data om den patient, der var (mest, hvis flere) udsat for den utilsigtede hændelse.'.

Figurerne 2.1 og 2.2 viser skemaet i papirformat bestående af to A4 ark. Rapporteringsskemaet, som det ser ud i papirformat, er ligeledes at finde i bilag A.

Den online udfyldelse af skemaet består af et skærmbillede, som skal udfyldes, på samme vis om papirskemaet, men i og med at det er elektronisk er der mulighed for at vælge datoen fra en kalender, og tidspunktet vælges fra en dropdown-menu.

2.2.1 Fritekst i Rapporteringsskema

Fritekstbeskrivelser er den tungeste måde at repræsentere medicinsk information på i forhold til senere hen at kunne anvende de informationer, der er i teksten. Samtidigt tillader det en hurtig måde at dokumentere hændelser på. Ved fritekstrapportering er der dog risiko for,



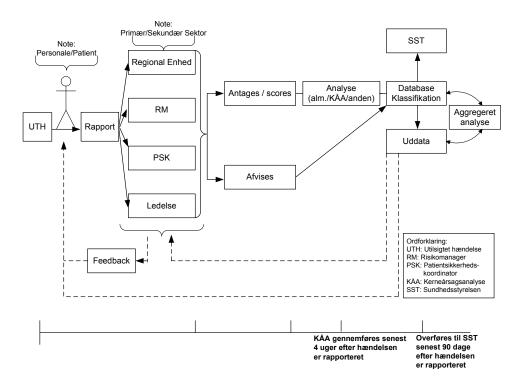
Figur 2.1: Side et i rapporteringsskemaet, hvor 1, 2 og 3 markerer strukturerede felter.

Figur 2.2: Side to i rapporteringsskemaet, hvor 4, 5, 6 og 7 indikerer fritekstfelter og 8 er et struktureret felt.

at der kan opstå inkonsistens mellem de forskellige felter. Hvorimod hvis hændelserne skal rapporteres i struktureret form, vil der være risiko for udeladelsesfejl, således at sundhedspersonen står tilbage med en manglende fuldendthed af rapporteringen. [Johnson et al. 2008] Som tidligere nævnt rapporteres UTH i semistrukturerede skemaer, Information kan trækkes direkte ud af de felter, hvor data er struktureret. Det vil sige, at data, i dette tilfælde ord, bliver til information idet feltet, hvor der i forvejen er defineret hvad feltet skal indeholde, bliver udfyldt. Det udfyldte ord er således i en kontekst, hvor det bidrager med information. For eksempel hvis teksten ud for feltet er 'Region' og der indtastes Region Hovedstaden, er der opnået information om, i hvilken region hændelsen er sket. Det er derefter tilgængeligt, at opnå statistisk viden om i hvilke regioner der forekommer flest UTH.

Derimod er tilgangen til at opnå viden ved at udtrække information mere besværet i de felter, der er lavet til fritekstfelter. For eksempel i fritekstfeltet med teksten 'Hændelsesbeskrivelse' giver hvert enkelt ord i friteksten ikke information i sig selv, men det kræver en sammenhængende forståelse af ordene, for at udtrække information.

Alternativet til fritekstfeltet er en struktureret formular, hvor der angives med enkelte ord eller afkrydsning, hvad der er årsag til hændelsen. Problemet ved denne strukturerede form er, at der kan forekomme situationer, hvor afkrydsning eller enkelt ord ikke er tilstrækkeligt beskrivende for hændelsen. Hvis den strukturerede form laves meget detaljeret for at undgå denne slags udeladelsesfejl risikeres det at blive tidskrævende at udfylde. [Johnson et al. 2008] Det betyder, at det er en balancegang, at udvælge hvornår det skal være struktureret, og hvornår det skal være anvendelse af fritekst. Region Hovedstaden har tidligere anvendt mere strukturerede rapporteringsskemaer, men er med tiden gået bort fra dette, da erfaringen viste, at det var for tilfældigt hvordan afkrydsningen fandt sted. Ofte blev de ikke udfyldt



Figur 2.3: Oversigt over sagsgangen når en UTH-rapport indsendes. [Region Hovedstaden 2008]

eller ligefrem udfyldt forkert. [Egebart & Bjørn 2008]. Med den valgte struktur i Region Hovedstadens UTH-rapporter er det nødvendigt, at have en klar oversigt hvordan rapporterne, og dermed hvordan friteksten håndteres.

2.3 Sagsbehandling af UTH-Rapporter

For at kunne drage nytte af UTH-rapporterne med henblik på at forbedre patientsikkerheden, gennemgår hver UTH-rapport en sagsbehandling. Efter at der er blevet rapporteret en UTH på et sygehus i Region Hovedstaden distribueres denne automatisk ud til 4 sagsbehandlere i sundhedssektoren via e-mail; 1) den regionale enhed som for Region Hovedstaden er Enhed for Patientsikkerhed, 2) risikomanageren tilknyttet det pågældende sygehus, 3) den patientsikkerhedskoordinatoren på sygehuset og 4) ledelsen på afdelingen på afdelingen hvor UTH fandt sted. Region Hovedstadens illustration af dette forløb ses på figur 2.3.

Herefter påbegyndes den egentlige sagsbehandling, der er baseret på gennemlæsning samt analyse af hændelserne. Først skal det besluttes, hvorvidt rapporten kan godkendes som en UTH eller ej. Dette skal besluttes senest tre hverdage efter, at hændelsen er rapporteret. Hvem der besidder denne opgave aftales lokalt imellem risikomanageren, patientsikkerhedskoordinatoren og afdelingsledelsen. Godkendes hændelsen, er alle tre sagsbehandlere involveret i sagsbehandlingen, hvor de har mulighed for at kommentere hændelsen. Risikomanageren har desuden rettighed til at involvere én eller flere eksterne eksperter inklusiv rapportøren, som kan bidrage yderligere til sagsbehandlingen. Efter endt sagsbehandling skal denne godkendes af afdelingsledelsen, hvorefter UTH scores med en Safety Assessment Code (SAC) med

en værdi fra 1-3. Scoringen er baseret på den faktuelle eller potentielle patientskade, kombineret med risikoen for gentagelse, hvor scoren på 3 symboliserer den største risiko, og er nærmere beskrevet i appendiks A. SAC-vurderingen udføres af enten risikomanageren eller patientsikkerhedskoordinatoren. Hændelsen skal vurderes senest 5 hverdage efter, at hændelsen er rapporteret. [Vingtoft & Pedersen 2007] Der kan være store variationer i denne scoring, da den lægger til grund for den enkelte persons vurdering, og det dermed er svært at opnå konsistens, da denne person ikke kan undgå at være farvet af situationen, på det sygehus vedkommende arbejder på. Hvis det er et stort sygehus, ses der flere UTH, hvorfor en person derfor kan være tilbøjelig til at vurdere risikoen for gentagelse højere end en person på et mindre sygehus med få hændelser. [Egebart & Bjørn 2008]

Endeligt kan det på baggrund af sagsbehandlingen overvejes, hvilken analyse der skal foretages af UTH, hvor valget står imellem kerneårsagsanalyse, aggregeret analyse eller en lokal analyse. Analysen foretages for at få identificeret årsagen til UTH for dermed at undersøge, om den kunne have været forhindret. Analysen skal være foretaget senest to uger efter, at hændelsen er rapporteret. [Vingtoft & Pedersen 2007]

Ved faktuel og potentiel score på 3 gennemføres en kerneårsagsanalyse. Ved faktuel/potentiel score på 2 og medicineringsfejl med potentiel score på 3 gennemføres en aggregeret kerneårsagsanalyse. Resten af hændelserne indgår i en lokal analyse. [Sygehusfællesskab 2005]

Rapporterne skal klassificeres manuelt af risikomanageren i henhold til ni hændelseskategorier, som foreskrevet af Sundhedsstyrelsen. Kategorierne er i forhold til hændelse i forbindelse med [dpsd.dk 2008]:

- 1. Medicinering
- 2. Kirurgiske eller invasive indgreb
- 3. Fald
- 4. Selvmord og selvmordsforsøg
- 5. Anæstesiprocedurer
- 6. Forveksling eller fejlkommunikation
- 7. Kontinuitetsbrud
- 8. Hjertestop eller uventet dødsfald
- 9. Øvrige hændelser

Klassifikationerne angives ved koder, og ligeledes bliver der på de tidligere H:S sygehuse også tilknyttet ATC-koder (Anatomical Therapeutic Chemical) til alle nævnte lægemidler, således at det også er muligt at udføre statistik på disse. Igen er det risikomanageren, der står for dette. Der kan være stor variation i hvorledes rapporterne bliver klassificeret alt efter hvilken risikomanager, der udfører opgaven. [Egebart & Bjørn 2008]

Når rapporten er indsendt af rapportøren får vedkommende en kvittering på dette. Herefter får personen besked om hvorvidt rapporten er godkendt eller afvist som UTH. Hvis den godkendes, tilbagesendes der efter endt sagsbehandling et resume af denne til rapportøren. [Vingtoft & Pedersen 2007]

Enhed for Patientsikkerhed kan nu på baggrund af UTH-rapporterne udforme generelle forebyggende initiativer og videreformidling af relevante resultater regionalt. [Sygehusfællesskab 2005, Mortensen & Dunker 2004 Enhed for Patientsikkerhed anvender to metoder til at uddrage læring af rapporterne. Den første metode indebærer at opnå brugbar information i rapporterne, når der opstår et ønske om at opnå mere viden om en specifik hændelsestype eller årsag hertil. På nuværende tidspunkt findes denne information ved manuelt at søge på et ord i friteksten og efterfølgende læse sammenhængen, hvori søgeordet fremkommer i rapporten. Det kan for eksempel være søgeordet 'venflon' hvis der ønskes, at undersøge om en venflon er årsag til den UTH. Denne tilgang er hurtig, men producerer også langt flere resultater end ønsket, og det kan blive tidskrævende at frasortere de falsk positive resultater manuelt. Det skyldes, at 'venflon' kan fremkomme i mange andre sammenhænge end som den direkte årsag til hændelsen. Det kræver derfor, at store mængder tekst skal gennemlæses ofte af flere omgange, for at finde frem til de steder hvor det var en venflon, som var årsag til hændelsen. Beslutningen om hvorvidt en hændelse skyldes en venflon eller ej beror dermed på en subjektiv vurdering. Nøgleordssøgningen bliver vderligere besværet af muligheden for stavefejl. Når information er blevet udtrækket kan der gives feedback til rapportøren i forhold til om hændelsen kunne have været undgået. Ved udtrækning af information fra flere rapporter kan der udarbejdes generelle retningslinier til brug i hele Region Hovedstaden typisk i form af temarapporter. [Egebart & Bjørn 2008]

Den anden metode består i, at Enhed for Patientsikkerhed ugentligt afholder møder, hvor én person fra enheden forinden har læst den forrige uges hændelser. Ud fra dette fremlægges de episoder, hvor personen mener, at der er potentiale for regional læring og enheden tager herefter stilling til hvordan det bør formidles. [Egebart & Bjørn 2008] Efter rapporterne er færdigbehandlet i Enhed for Patientsikkerhed sendes de kommenterede rapporter videre til Sundhedsstyrelsen i anonymiseret form senest 90 dage efter hændelsen er rapporteret [Vingtoft & Pedersen 2007]. Sundhedsstyrelsen udgiver, med udgangspunkt i rapporterne, forskellige publikationer og informationsmateriale omkring hændelserne for derved at formidle den læring, der skabes på baggrund af rapporterne. [sikkerpatient.dk 2008] Såfremt denne viden og læring bliver formidlet korrekt er der potentiale for at øge patientsikkerheden.

2.4 Problemformulering

UTH rapporteres med det formål at opnå viden om, i hvilke situationer hændelserne forekommer, hvad årsagen var og hvad konsekvenserne blev. Denne information skal benyttes til at forebygge kommende lignende hændelser ved at iværksætte retningslinjer og forebyggende initiativer både lokalt, regionalt og nationalt. Den lokale anvendelse af UTH-rapporterne er den direkte opfølgning af hændelsen til den sundhedsperson, som rapporterede hændelsen. Regionalt kan analyser af flere rapporter anvendes til at udstede generelle retningslinjer, hvilket varetages af Enhed for Patientsikkerhed i Region Hovedstaden. Nationalt er det Sundhedsstyrelsen, som kan danne et overblik af behovet for nationale tiltag.

Når der skal laves tiltag og retningslinjer på regionalt plan i form af temarapporter genanvendes informationerne i rapporterne på nuværende tidspunkt ved brug af nøgleordssøgninger for dermed at opnå viden om bestemte situationer eller hændelser. Nøgleordssøgningerne er efterfulgt af manuel gennemlæsning af de rapporter, der matchede søgeordet. Dette er en

meget tidskrævende proces, idet nøgleordssøgninger resulterer i en stor mængde af falsk positive resultater, da søgeordet kan indgå andre sammenhænge i teksten. En automatisering af denne proces vil kunne effektivisere, spare tid, og sikre en klassifikation efter de samme parametre hver gang. Desuden vil kvaliteten af analysearbejdet højnes, idet søgningen efter årsagen til UTH bliver ensrettet. Processen kan dog ikke automatiseres fuldt ud, da en søgning vil forudsætte en endelig vurdering af en ekspert i form af gennemlæsning af de udvalgte rapporter. [Egebart & Bjørn 2008]

Med udgangspunkt i en problemstilling omhandlende automatisk at kunne opnå information fra fritekst, kan *Text Mining* [Feldman & Sanger 2006] og *Natural Language Processing* (NLP) nævnes [Johnson *et al.* 2008]. Begge områder dækker bredt, men ingen af dem er endnu anvendt i forbindelse med udtræk af danske UTH-rapporter.

Text Mining inkluderer blandt andet kategorisering af tekst, tekstgruppering, koncept / enhedsudtræk og dokumentopsummering. [Weiss et al. 2005]

Metoden til at udføre NLP kan deles op i henholdsvis symbolsk og empirisk tilgang. Den symbolske tilgang omhandler integration af viden om sprog baseret på regler. [Dale et al. 2000] Empirisk NLP bygger på tekstkorpus, en stor mængde data, kombineret med statistisk processering. [Somers 2004]

En kombination af outputtet fra et NLP-system med standardteknikker sammenholdt med tekstkategorisering kan præstere et bedre resultat end de to fremgangsmåder adskilt [Matthews 2006].

En succesfuld anvendelse af disse fremgangsmåder på friteksten i danske UTH-rapporter vil øge effektivisering af fremstillingen af retningslinjer. Dermed vil der være potentiale for at forbedre patientsikkerheden, idet tilgængeligheden af friteksten i rapporterne til sekundær brug øges. [Egebart & Bjørn 2008]

2.4.1 Afgrænsning

Det er valgt at afgrænse dette projekt til kun at omhandle UTH-rapporter, hvor Elektronisk Patient Medicinerings (EPM) modulet indgår i hændelsesbeskrivelsen, da det er denne data, der er til rådighed og ligeledes var den seneste nøgleordssøgning som Enhed for Patientsikkerhed havde foretaget ved projektets begyndelse. EPM er navnet på Region Hovedstadens medicinmodul i den elektroniske patientjournal. Denne afgrænsning forhindre ikke muligheden for at vurdere konceptet som helhed, idet det antages, at hvis det kan lade sig gøre at identificere om EPM er årsag til en UTH eller ej, så er det også muligt at udvide til andre søgeord. EPM kan være nævnt i rapporterne, fordi det er blevet anvendt i forbindelse med medicineringsprocessen eller fordi EPM var årsag til den utilsigtede hændelse. I dette tilfælde vil en nøgleordssøgning på EPM give falsk positive resultater i de tilfælde, hvor EPM blot blev anvendt, hvis formålet var at finde de UTH, hvor EPM var årsagen. Dette leder frem til følgende konkrete problemstilling:

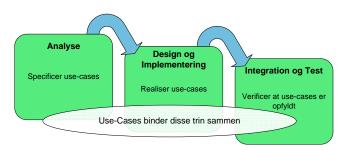
Hvordan kan et NLP-baseret system, ved hjælp af klassifikation, identificere om EPM er årsag til en UTH?

Kapitel 3

Metode

I dette kapitel redegøres for, hvilken metode der er anvendt til at undersøge om NLP kan anvendes til at klassificere årsagen til utilsigtet hændelser. Dette er sket gennem analyse, design og udvikling af et NLP-baseret system til klassifikation af utilsigtet hændelser.

Udviklingsmetoden for løsningsforslaget er baseret på to traditionelle metoder til softwareudvikling; Struktureret Programudvikling (SPU) [Biering-Sørensen et al. 2002] og Unified Modeling Language (UML) [Eriksson et al. 2004]. Ved at kombinere disse to metoder opnås en udviklingsmetode, der kaldes SPU-UML. Denne metode bygger på den traditionelle vandfaldsmodel, som indeholder syv forskellige faser; Kravspecifikation, Design, Implementering, Integration, Afprøvning og fejlfinding, Installation og Vedligeholdelse. [Winston 1970] I denne



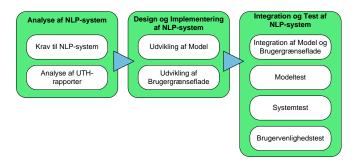
Figur 3.1: På figuren ses SPU-UML metoden, hvor use-cases indgår i alle trin af udviklingen under udarbejdelse af et løsningsforslag.

rapport er det udført ved, først at have analysefasen hvor det specificeres hvilke krav, der er til systemet. Herefter følger design og implementeringsfasen, hvor kravene implementeres. Når alle dele af systemet er lavet, sættes disse sammen i integrationsfasen, hvorefter systemet er klar til at blive testet. Vandfaldsmodellen er et special tilfælde af de iterative modeller, idet der kun udføres én enkelt iteration af de enkelte trin [Hansen 2005a]. Det siges også, at faserne er diskrete, hvilket vil sige, at én fase afsluttes, førend en ny begyndes [Biering-Sørensen et al. 2002].

SPU-UML metoden er en use-casedrevet udviklingsproces. Det vil sige, at udviklingen i løsningsforslaget styres ud fra use-cases, der er identificeret i forbindelse med kravspecifikationen. Use-cases beskriver de funktionelle krav til systemet. De bliver implementeret for at sikre, at alle funktionaliteter er realiseret i systemet, for derefter at give mulighed for at verificere og teste systemet på baggrund af de opstillede use-cases. Da use-cases udgør de funktionelle krav, vil de have indflydelse på alle faser i vandfaldsmodellen, hvilket fremgår af

figur 3.1. [Hansen 2005a]

I dette projekt kan de tre faser på figur 3.1 omsættes til følgende tre trin: 1) Analyse af NLP-system, 2) Design og Implementering af NLP-system og 3) Integration og Test af NLP-system, som vist på figur 3.2.



Figur 3.2: På figuren ses hvilke trin, der indgår i udarbejdelse af et løsningsforslag til besvarelse af problemformuleringen.

Analyse of NLP-system

Det har ikke tidligere været forsøgt at klassificere danske UTH-rapporter, på baggrund af årsagen, ved brug af NLP-teknikker, hvorfor denne rapport vil undersøge, hvorvidt det er muligt som et 'proof-of-concept'. For at kunne udføre disse undersøgelser er valgt at konstruere et NLP-system, således en fremtidig bruger vil have mulighed for at deltage i vurderingen af klassifikationen.

For at kunne designe et NLP-system, der kan afgøre hvorvidt EPM er årsag til en utilsigtet hændelse eller ej, er det nødvendigt at opnå en samlet forståelse af, i hvilken kontekst systemet skal anvendes. Dette gøres ved, at udarbejde en kravspecifikation i samarbejde med brugeren af systemet, hvilket fører til en række krav til systemets funktionalitet. Brugeren af systemet er risikomanagere ved Enhed for Patientsikkerhed i Region Hovedstaden. Samarbejdet med Enhed for Patientsikkerhed er foregået igennem kontakt til to personer: Brian Bjørn og Jonas Egebart, der begge er risikomanager og læger. Deres krav til systemet er herefter blevet omsat til use-cases, for dermed at kunne anvende den use-case drevne udviklingsmetode. Udgangspunktet for at kunne designe et system er at inddrage data i form at UTH-rapporter. UTH-rapporterne er valgt ud fra den specifikke problemstilling, hvor der er fortaget en søgning i Region Hovedstadens database med utilsigtede hændelser, hvor alle der indeholdt ordet 'EPM' i 3. kvartal 2007 er udvalgt svarende til 131 stk. Rapporterne er randomiserede til enten et trænings- eller testsæt med en fordeling på to tredjedele til træningssættet og en tredjedel til testsættet. Kun rapporter indeholdende tekst i alle fire fritekstfelter er inkluderet i sættene. UTH-træningsrapporterne er analyseret for at afgøre hvilke karakteristika, der kendetegner UTH-rapporter, hvor EPM er årsag til den utilsigtede hændelse og UTH-rapporter, hvor det ikke er.

Design og Implementering af NLP-system

Design og implementering af NLP-systemet indebærer, at kravene til systemet omsættes til en teknisk løsning. Den tekniske løsning er inddelt i 2 dele: udvikling af en model, der kan k-

lassificere UTH-rapporter i forhold til om EPM er årsag til hændelsen eller ej, samt udvikling af brugergrænseflade, der grafisk kan repræsentere resultaterne af klassifikationen. Dette er de to komponenter, der tilsammen udgør den tekniske løsning. Metoden til at udvikle denne klassifikationsmodel bygger på mønstergenkendelsesprincipper. Modellen konstrueres ud fra viden om UTH-rapporter opnået igennem analyse og NLP-processering. Designet til brugergrænsefladen er udarbejdet over tre omgange. Først er Enhed for Patientsikkerhed blevet præsenteret for et mock-up, hvorefter brugergrænsefladen er blevet implementeret. System er testet i forhold til brugervenlighedsmetoder og efterfølgende forbedret i henhold til resultatet af disse.

Integration og Test af NLP-system

Ligesom udviklingen af systemet er integration af systemet foretaget i programmeringssproget Python. Efter integration af systemet eskal systemet testes for at validere om løsningsforslaget løser problemstillingen. Testen består af tre dele; en test af selve klassifikationsmodellen, test af det samlede system samt en brugervenlighedstest. Det er valgt at inddrage alle tre tests, da de alle bidrager med noget forskelligt. Test af modellen gør det muligt at vurdere resultatet af klassifikationen af UTH-testrapporter. Det samlede system testes, for at undersøge hvorvidt de funktionelle krav til systemet i form af use-cases er blevet opfyldt. Endeligt foretages der en brugervenlighedstest med de to føromtalte brugere fra Enhed for Patientsikkerhed, for at få identificeret hvorvidt brugeren forstår og kan anvende systemet, eller om der er områder der bør ændres.

I de næste tre kapitler vil de ovenstående tre dele af metoden blive gennemgået, med henblik på at designe og implementere et NLP-system.

Kapitel 4

NLP-System

Ud fra den foregående analyse af UTH-rapporteringsforløbet er der identificeret et behov for et system, der kan øge tilgængeligheden af hvorvidt Region Hovedstadens medicinmodul EPM er årsag til den utilsigtede hændelse eller ej. Efterfølgende vil disse behov blive analyseret for at kunne designe et NLP-system.

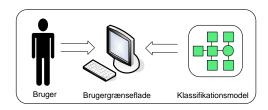
Et godt design er karakteriseret ved, at det er let at forstå, implementere, teste, vedligeholde og samtidigt har en høj performance. Én metode til at strukturere designet af et system på er ved inddragelse af aspekter fra modelleringssproget Unified Modelling Language (UML). [Biering-Sørensen et al. 2002] I denne rapport er UML inddraget ved at anvende use-cases til at beskrive hvilke funktionaliteter, der skal indgå i designet af systemet.

4.1 Kravsspecifikation

Ud fra første møde med Enhed for Patientsikkerhed ved Region Hovedstaden er der udarbejdet en beskrivelse af hvilket krav, der er til systemet [Egebart & Bjørn 2008]. Brugeren af systemet vil være en person tilknyttet Enhed for Patientsikkerhed ved Region Hovedstaden. Formålet med systemet er at foretage en søgning i henhold til en specifik årsag for dermed at få afklaret hvilke utilsigtede hændelser, der skyldes en specifik årsag. Dette kaldes en årsagsbestemmelse. Årsagsbestemmelsen skal klassificeres ved brug af NLP, og når der arbejdes med NLP-fritekst, er der tale om to overordnede metoder [Dale et al. 2000]:

- 1. Opnå information
- 2. Udtrække information

Forskellen på opnå information og udtrække information i denne sammenhæng, er om det ønskes, at vide hvilket sted i teksten søgeordet kan identificeres som årsag, eller om det ønskes at opnå samlet information for UTH-rapporten. I rapportens problemstilling er formålet at klassificere UTH-rapporten, i henhold til, om EPM er årsag eller ej. Det betyder, at der ønskes at opnå information, idet der ikke er behov for at vide det konkrete søgeord. Dette kan ses som en dokumentklassifikation, med det formål at reducere antallet af relevante rapporter og derved lette arbejdsgangen i forbindelse med sekundær brug af UTH-rapporterne. Brugeren vil stadig skulle have mulighed for at ændre en given klassifikation efter at systemet har fortaget en klassifikation. Scenariet, hvor en bruger interagerer med systemet og den underliggende klassifikation via en brugergrænseflade, er skitseret på figur 4.1.



Figur 4.1: Brugeren skal interagere med en klassifikationsmodel, via en brugergrænseflade.

Brugeren af systemet specificerer, i form af et søgeord, årsagen til den utilsigtede hændelse. Idet udviklingen af systemet er afgrænset til kun at omhandle EPM som årsag, kan søgeordet kun være dette. Ligeledes skal brugeren angive i hvilken fil søgningen ønskes foretaget. Systemet skal nu udvælge de UTH-rapporter, hvori det specificerede søgeord indgår. Herefter skal de udvalgte rapporter klassificeres i henhold til de rapporter, hvor søgeordet EPM er årsagen og de rapporter, hvor søgeordet EPM indgår af andre grunde. Klassifikationen skal foretages ved brug af NLP teknikker. Denne klassifikation skal efterfølgende præsenteres på en brugergrænseflade.

Input til systemet bliver dermed en fil med UTH-rapporter samt et søgeord, som specificerer årsagen. Resultatet af klassifikationen er to klasser med henholdsvis rapporter, hvor søgeordet var årsag til hændelsen, også kaldet den positive klasse og rapporter hvor søgeordet ikke var årsag til hændelsen, også kaldet den negative klasse.

Beskrivelsen af systemet kan omsættes til en række krav til systemet. En god og brugbar kravspecifikation er kendetegnet ved at kravene er: korrekte, entydige, fuldstændige, konsistente, verificerbare, modificerbare og sporbare. Kravspecifikationen er korrekt, hvis alle kravene i den er korrekte, hvilket vil sige, at alle kravene, er mulige at opfylde. Specifikationen bliver entydig, idet hvert krav kun har én enkelt betydning, og fuldstændig hvis alle funktionaliteter og dertilhørende krav er identificeret og beskrevet. En konsistent kravspecifikation er kendetegnet ved, at der ikke er nogle af kravene, der er i konflikt med hinanden. Hvis det er muligt, at afgøre hvorvidt et krav er opfyldt eller ej, siges kravspecifikationen at være verificerbar, og at den er modificerbar er kendetegnet ved, at det er let at ændre i kravene senere hen. Endeligt bliver en kravspecifikation sporbar, såfremt det er muligt, at spore hvorfra kravene stammer. [Biering-Sørensen et al. 2002] Ud fra disse retningslinier for en kravspecifikation og beskrivelsen af hvad systemets funktionalitet skal være, er der opstillet følgende krav til systemet:

- 1. Systemet skal klassificere UTH-rapporter til én af to mulige klasser; enten hvor søgeord er årsag til hændelsen, eller hvor søgeordet blot indgår.
- 2. Systemet skal give brugeren mulighed for at angive en kildefil indeholde UTH-rapporter.
- 3. Systemets input skal være kompatibelt med databasen hvorfra UTH-rapporterne skal hentes.
- 4. Systemet skal give brugeren mulighed for at angive et søgeord, der specificerer den

årsag, der skal klassificeres efter med mulighed for flere stavekombinationer af årsagen.

- 5. Systemet skal klassificere alle positive UTH-rapporter korrekt.
- 6. Systemet skal give mulighed for at se både de positive og de negative klassificerede UTH-rapporter i en oversigt, samt den individuelle rapport.
- 7. Systemet skal give brugeren mulighed for at korrigere klassifikationerne ved at flytte en rapport fra en positiv til en negativ klasse og omvendt såfremt en klassificering er forkert.
- 8. Systemet skal kunne sortere klassifikationerne i henhold til SAC-score og sagsnummer.
- 9. Systemet skal kunne gemme en klassifikation.
- 10. Systemet skal kunne vise og udskrive de klassificerede UTH-rapporter, enten på papir eller til en pdf fil.

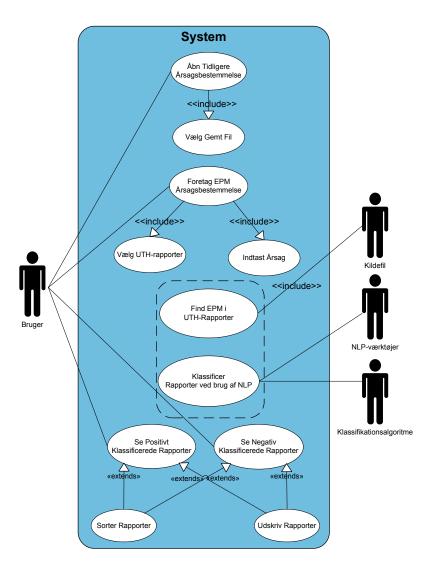
De ti krav skal omsættes til hvorledes udvikling af et system skal forløbe, og de giver efter endt udvikling mulighed for at følge op på om brugernes krav er opfyldt til systemet.

4.2 Systemets Funktionaliteter

Selve kravene til systemets funktionalitet kan omsættes til en grafisk repræsentation af det ønskede system ved hjælp af et use-case diagram, se figur 4.2. Et use-casediagram viser systems funktionalitet repræsenteret ved aktører og deres krav, hvoraf linjerne viser, at aktøren giver en værdi til use-casen. Hver use-case beskriver en del af systems udformning, idet den svarer til et specifikt krav til systemet. På figuren anvendes to UML-notationer; Include og Extend. Include angiver, at den pågældende use-case er inkluderet i den ovenstående use-case, hvilket vil sige, at udførslen af use-casen er obligatorisk for, at den ovenstående use-case kan fuldføres. Extend betyder, at use-casen kan forløbe, men ikke er en nødvendighed for, at den ovenstående use-case kan fuldføres. [Eriksson et al. 2004] Aktørerne til systemet er; En bruger, en kildefil, NLP-værktøjer fra Center for Sprogteknologi (CST) ved Københavns Universitet og en klassifikationsalgoritme. En aktør er nogen eller noget udenfor systemet, der enten bruger systemet eller er brugt af systemet [Eriksson et al. 2004]. Brugeren er en aktiv aktør, idet vedkommende skal interagere med systemet i de tilknyttede use-cases for, at funktionaliteterne bliver udført. Derimod er kildefilen, NLP-værktøjerne og klassifikationsalgoritme alle passive aktører, idet systemet anvender aktørerne for at opnå funktionaliteterne i de pågældende use-cases. I det følgende vil indholdet i de enkelte use-cases blive gennemgået nærmere.

Åbn Tidligere Årsagsbestemmelse Denne use-case repræsenterer, at brugeren af systemet har mulighed for, at åbne en tidligere foretaget søgning, såfremt der allerede er blevet gemt en.

Vælg Gemt Fil For at der kan åbnes en tidligere årsagsbestemmelse, skal stien til den gemte fil angives. Dette kan enten gøres ved at browse efter filens placering, eller ved at indtaste stien direkte.



Figur 4.2: Use-casediagrammet viser systemets funktionalitet illustreret ved use-cases og dertilhørende aktører. Use-casene kan inddeles i to grupper. Én omhandlende funktionaliteter i forhold til klassifikationsmodellen, hvilke på figuren er omgivet af en stiplet boks. De resterende use-cases er alle funktionaliteter i forbindelse med en brugergrænseflade.

- Foretag EPM Årsagsbestemmelse Denne use-case repræsenterer, at brugeren af systemet kan foretage en ny årsagsbestemmelse.
- Vælg Kildefil For at der kan foretages en ny årsagsbestemmelse skal der vælges en kildefil med de rapporter, der ønskes klassificeret. Dette gøres ved at browse sig frem til den korrekte sti for filens placering eller selv at indtaste stien.
- Indtast Årsag For at foretage en ny årsagsbestemmelse er det ligeledes nødvendigt at årsagen, som der skal klassificeres efter, angives. Da dette er et 'proof-of-concept' er det muligt at angive én type årsag dog med forskellige stavekombinationer. Den mulige årsag er som tidligere nævnt EPM.
- Find EPM Rapporter i Kildefil For at reducere antallet af rapporter, der skal klassificeres, udvælges kun de rapporter i kildefilen, hvori søgeordet indgår.
- Klassificer Rapporter ved brug af NLP Hver enkelt rapport, der indeholder søgeordet skal klassificeres i én af to grupper: enten gruppen hvor søgekriteriet er årsagen til den utilsigtede hændelse eller den anden gruppe, hvor søgekriteriet blot indgår i teksten uden at være årsag til UTH.
- Se Positivt Klassificerede Rapporter Efter at rapporterne er blevet klassificeret skal de positivt klassificerede rapporter vises for brugeren af systemet.
- Se Negativt Klassificerede Rapporter Brugeren af systemet har mulighed for at se de rapporter, hvor EPM ikke var årsag til den utilsigtede hændelse.
- **Sorter Rapporter** Brugeren af systemet har mulighed for at filtrere rapporterne i henhold til sagsnummer og SAC-scorer.
- Udskriv Rapporter Brugeren af systemet har mulighed for at udskrive alle de klassificerede rapporter samt få vist udskrift i form af en pdf fil.

De beskrevne use-cases kan inddeles i to grupper. Én der omhandler funktionaliteter i forhold til udvikling af den model systemet skal bygges op om, og én der omhandler funktionaliteter forbundet med den interaktion med systemet i form af en brugergrænseflade. De modelbaserede use-cases er: Find EPM Rapporter i Kildefil, og Klassificer Rapporter ved brug af NLP, se den stiplede boks på figur 4.2, hvorimod de resterende 9 use-cases alle er brugergrænsefladebaserede.

Ud fra disse funktionaliteter er det dermed muligt at påbegynde en egentlige udvikling af systemet, for at undersøge om NLP kan bruges til klassifikation af årsagen til UTH-rapporter.

Kapitel 5

Analyse af UTH-Rapporter

For at kunne udvikle en model til klassifikation af UTH-rapporter, er det nødvendigt at inddrage data, som i dette tilfælde er på forhånd klassificerede UTH-rapporter. For at have et beslutningsgrundlag for, hvorledes systemet skal udvikles, vil der igennem dette kapitel blive opnået kendskab til indholdet i UTH-rapporterne.

Det er vigtigt, at udvælgelsen af UTH-rapporter afspejler i formålet med klassifikationen. Da der i dette projekt ønskes at klassificere, hvorvidt EPM er årsag til den utilsigtede hændelse eller ej, indeholder de tilgængelige UTH-rapporter alle ordet 'EPM' eller 'EMP', da dette er de to måder, hvorpå der bliver refereret til medicinmodulet i UTH-rapporterne. Fremover vil kun ordet 'EPM' blive nævnt, men det dækker over begge notationer.

5.1 Beskrivelse af UTH-Rapporter

UTH-rapporterne stammer alle fra Region Hovedstadens database med UTH-rapporter fra 3. kvartal 2007 svarende til 132 rapporter i alt.

Friteksten i de fire fritekstfelter, hændelsesbeskrivelse, årsager til hændelsen, konsekvens af hændelsen og forslag til tiltag i rapporterne, tidligere omtalt i kapitel 2, er forinden blevet anonymiseret af Enhed for Patientsikkerhed. Derudover er hver rapport blevet SAC-scoret i henhold til den potentielle og faktuelle risiko for lignende hændelser.

Mængden af tekst i rapporter er varierende, da det er forskelligt, hvor detaljeret rapportøren beskriver den enkelte utilsigtede hændelse. Gennemsnitligt er der 97 ord i hver rapport med et minimum på 14 ord og maksimum på 372 ord.

5.1.1 Inkonsistens i UTH-Rapporterne

Teksten i fritekstfelterne kan bære præg af stavefejl, slåfejl, ufuldendte sætninger, diverse forkortelser og grammatiske fejl, hvilket al sammen er med til at forringe kvaliteten af UTH-rapporterne som data. Anvendelse af NLP på inkonsistente UTH-rapporter får betydning for hvilke resultater, der kommer ud heraf. Det skyldes, at et af de grundlæggende elementer i NLP er tegnsætning, og når der anvendes ukorrekte forkortelser og manglende tegnsætninger som for eksempel i nedenstående sag, hvor EPM ikke er årsag til hændelsen, kan det skabe problemer:

Sagsnummer: xxxxxx1

Hændelse: PT.s medicin dispenseret i EPM og ikke administ.

Årsag: ved ikke

Konsekvens: ingenhuske at administr.

Tiltag: medicin i EPM

Ud af de 16 ord, som indgår i denne hændelse, er der tre ureglementerede forkortelser: 'PT.s', 'administ.' og 'administr.' samt et manglende ophold i mellem 'ingenhuske'. Årsagen til inkonsistens i rapporterne kan skyldes, at det er lovpligtigt at rapportere UTH-rapporter, og der nationalt ønskes så mange som muligt, hvilket betyder, at der ikke stilles krav til rapportøren om hvordan udfyldelse af fritekstfelterne skal ske, og at en stavekontrol måske ville skabe besvær omkring indrapporteringen [Egebart & Bjørn 2008]. Ligeledes er det heller ikke stringent, at de fire fritekstfelter med henholdsvis hændelsesbeskrivelsen, årsager, konsekvenser og tiltag indeholder den dertilhørende tekst, idet nogle rapportører kommer til at beskrive, for eksempel, både hændelse og årsag i hændelsesbeskrivelsesfeltet, hvilket yderligere komplicerer datastrukturen. Ligeledes er der enkelte rapporter, hvor der ikke er tekst i alle de fire fritekstfelter, selvom dette er det eneste krav Region Hovedstaden stiller til rapportøren i forhold til indholdet i fritekstfelterne. I dette projekt er det valgt kun at arbejde med rapporter, der har indhold i alle fire fritekstfelter, hvorfor én rapport er blevet frasorteret, hvorved antallet af tilgængelige UTH-rapporter reduceres til 131.

De 134 rapporter er inddelt i et trænings- og testsæt. Rapporterne er automatisk randomiseret til en af disse sæt med en fordeling på to tredjedele til træningssættet og en tredjedel til testsættet. At denne tildeling er sket automatisk bevirker, at det ikke er muligt at påvirke resultatet af test af systemet, idet testsættet først inddrages efter udvikling af systemet er fuldendt.

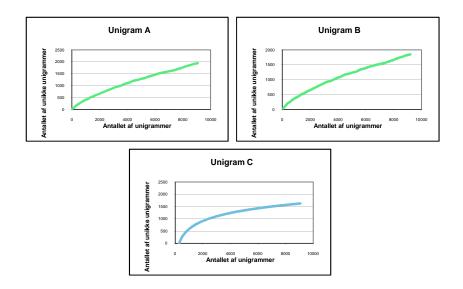
Klasserne for træningsrapporterne er manuelt blevet identificeret af eksperter i forbindelse med udarbejdelsen af dette projekt. Ud fra denne identifikation ses, at der eksisterer flere rapporter, hvor EPM blot indgår, men ikke er årsagen, frem for rapporter hvor EPM er årsag til den utilsigtede hændelse, hvilket er vigtigt at være opmærksom på under udarbejdelse af en model, der kan klassificere, hvorvidt EPM er årsag til hændelsen.

5.1.2 Antallet af UTH-Rapporter

Kravet til de UTH-rapporter, der anvendes i projektet er, at det er repræsentativt for alle UTH-rapporter indeholdende søgeordet 'EPM'. For at undersøge hvorvidt mængden af træningsrapporter er tilstrækkeligt til udvikling af systemet, kan der foretages en analyse af ordforekomster i træningssættet.

Hvis træningssættet, bestående af de 88 UTH-rapporter, er repræsentativt for alle UTH-rapporter indeholdende EPM, skal forekomsten af nye ord stagnere i takt med et stigende antal af rapporter, som vist på figur 5.1 C. Dette kan undersøges ved at fortage en n-gram analyse, der er en statistisk metode til at fastslå, hvor mange nye unikke ordobservationer, der detekteres i et givent ordkorpus. Hvis n-gram analysen er baseret på en sekvens med n=1, kaldes dette en unigramanalyse. Er antallet af ord ikke endeligt, vil det ikke være muligt at dække hele domænet i udviklingen af algoritmen. [Dale $et\ al.\ 2000$]

Unigramanalysen er udført på UTH-træningsrapporterne. Diverse tegn såsom punktum og



Figur 5.1: Unigramanalyser af træningsrapporter, hvor antallet af unikke ordforekomster ses i forhold til det samlede antal af ord. A) Unigramanlyse på træningsrapporter, B) Unigramanalyse på træningsrapporter rettet for stavefejl og forkortningsfejl, C) Eksempel på unigramanalyse hvor UTH-rapporterne er repræsentative for domænet.

komma er forinden blevet frasorteret, da disse ikke har indflydelse på typerne af ord. Ligeledes er stammen fundet for hvert enkelt ord, da ordets gradbøjning ikke har betydning for unigrammet. Resultatet af analysen illustreres ved grafen på figur 5.1 A, hvor antallet af unikke unigrammer er plottet i forhold til antallet af unigrammer. Her ses det, at antallet af unikke ord ikke er endeligt, hvilket betyder, at der kan argumenteres for, at 88 rapporter til træning ikke er tilstrækkeligt til udvikling af en model, der skal kunne bruges på alle UTH-rapporter. Resultatet skyldes, at ordvalget varierer i henhold til rapportøren og de føromtalte faktorer såsom stavefejl, der er med til at forringe datakvaliteten. Ligeledes indeholder fritekstfelterne en stor mængde medicinske ord, som kan have indflydelse på det værktøj, der anvendes til at finde stammen af hvert ord, da dette værktøj ikke er trænet på medicinske tekster. For at se hvilke indflydelse stavefejl og forkortningsfejl har på unigrammet, er alle træningsrapporter kørt igennem en stavekontrol samt forkortelser er blevet udskrevet, hvorefter unigrammet er lavet igen. Resultatet af dette ses på figur 5.1 B, hvor det ses at antallet af unikke unigrammer er reduceret, men at det endnu ikke er tilstrækkeligt til at de kan siges at være repræsentative for hele domænet. Stavekontrollen, der er anvendt er ikke udviklet til medicinske tekster. Ved at indføre en medicinsk stavekontrol på rapporterne, vil det være muligt at ensrette ordene i fritekstfelterne yderligere, hvilket vil have en indvirkning på unigrammet.

5.2 Karakteristika for UTH-rapporter

Generelt er der ikke stor variation i typerne af ord, der fremkommer i de to klasser, idet rapporterne fra de to klasser begge omhandler utilsigtede hændelser, hvor ordet EPM indgår. Hvis træningssættet var bestående af alle tænkelige UTH-rapporter er det muligt, at der

ville være større variation imellem rapporter indeholdende EPM og alle andre rapporter. Det er derfor vigtigt at få identificeret de forskelle, der måtte være, med henblik på at kunne klassificere UTH-rapporter i henhold til hvorvidt EPM er årsagen eller ej.

På foranledning af dette projekt har eksperter identificeret de ord, fraser eller sætninger i træningsrapporterne, der ifølge dem, afgør hvorvidt årsagen skyldes EPM eller ej i hver enkelt rapport. Dette er gjort med det formål at få identificeret karakteristika for henholdsvis den positive og negative klasser. Af disse markeringer ses, at de afgørende fraser og sætninger ofte indeholder ordet 'EPM', hvorfor det ved identifikation af karakteristika er relevant at se på den kontekst, ordet fremkommer i. Ud fra markeringerne er følgende karakteristika blevet identificeret for henholdsvis den positive og negative klasse.

5.2.1 Årsagskarakteristika

I de rapporter, hvor EPM er årsag til den utilsigtede hændelse, skyldes hændelsen, at funktionen af EPM ikke har fungeret korrekt på det pågældende tidspunkt. Dette kan for eksempel være på grund af, at systemet er nede eller at der er manglende eller misvisende funktionalitet i systemets struktur. Derfor vil det ofte være konkrete tekniske ord, der direkte relaterer til systemet, som kendetegner disse rapporter. Dette kan for eksempel være 'faneblad' eller 'advarselsboks'. Ud fra eksperternes identifikation er der udvalgt en række substantiver, der kendetegner denne situation. Ligeledes kan der identificeres nogle karakteristiske verbumfraser, såsom 'advarer ikke' og 'systemet er nede'. Et eksempel her på er:

Sagsnummer: xxxxxx2

Hændelse: Medicin til pt med meningitis automatisk seponeret i EPM kl 10.59. Opdaget tilfældigt.

Årsag: Ligner EPM systemfejl, men årsag ikke klarlagt.

Konsekvens: Ingen, da fejlen opdages tilfældigt. DER ER DOG TALE OM

POTENTIELT LIVSTRUENDE FEJL.

Tiltag: ?

I denne sag ses det, at frasen 'EPM systemfejl' indikerer, at der kan være tale om en hændelse med EPM som årsag. Rapportøren selv, er i tvivl om EPM er årsagen, hvilket det efterfølgende er blevet vurderet af eksperterne til at være.

5.2.2 Ikke Årsagskarakteristika

I de rapporter, hvor EPM ikke er årsag til den utilsigtede hændelse, er EPM ofte nævnt uden egentlig tilknytning til selve årsagen til hændelsen. Grunden til at EPM nævnes i disse sager, kan ses som led i beskrivelse af årsagen til den utilsigtede hændelse, hvor årsagerne overordnet omhandler medicineringsfejl eller arbejdsgange hvor ordet EPM har været nævnt i forbindelse med. Et eksempel på sådan en rapport er:

Sagsnummer: xxxxxx3

Hændelse: I EPM var dele af medicinen til patienten ordineret son selvmedicinering. Dette var dog en fejl, da patient ikke havde medbragt

egen medicin.

Årsag: Lægen må have ordineret forkert

Konsekvens: Kunne have resulteret i, at patienten kun fik dele af sin medicin. Dele af medicinen, der var ordineret som selvmedicinering var bl.a furix og magnyl.

Tiltag: Lægen må dobbelttjekke ordinationen

I denne rapport ses det, at hændelsen ikke direkte skyldes EPM, men at lægen har ordineret forkert i EPM, hvorfor eksperterne har klassificeret denne rapport som negativ.

Da UTH-rapporterne er skrevet i et naturligt sprog vil være flere variationer af det samme ord for eksempel i forhold til tid og antal, hvorfor det bliver vanskeligt at identificere karakteristika blot ved direkte nøgleordssøgninger. Det er derfor nødvendigt at få skabt en form for mønster for det naturlige sprog, der er uafhængig af tid og antal for derved at kunne genkende karakteristika og dermed gøre det muligt at adskille rapporter, hvor EPM er årsag til den utilsigtede hændelse, og hvor EPM ikke er årsag.

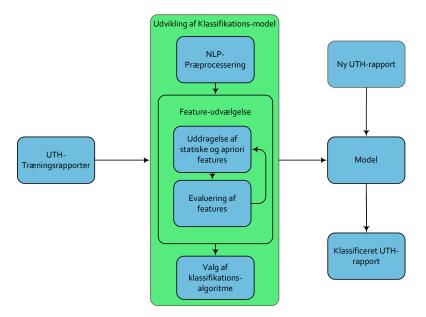
Ud fra analyse, beskrivelse og karakteristika for UTH-rapporterne er der dannet et grundlag for at kunne udforme en model til klassifikation af UTH-rapporterne i henhold til om EPM er årsag eller ej.

Kapitel 6

Udvikling af EPM Klassifikationsmodel

For at udvikle et samlet system, der opfylder alle Enhed for Patientsikkerheds krav, er det nødvendigt at udvikle en model, således at rapporterne kan klassificeres. Modellen skal gøre det muligt at udføre de to use-cases Foretag EPM Årsagsbestemmelse og Find EPM rapporter i Kildefil, som blev fremsat under krav til systemet.

Det generelle formål med at udvikle en model er, at lave en ramme for, at kunne anvende logik og matematik, som kan evalueres uafhængigt og af samme årsag kan tilføres i andre situationer, end den er udviklet til. I problemstillingen, i dette projekt ønskes, at lave en model som på baggrund af UTH-træningsrapporterne kan klassificere, om et søgeord i et UTH-rapport hører til klassen 'årsag' eller klassen 'ikke årsag'. På figur 6.1 ses principperne i generering af



Figur 6.1: UTH-træningsrapporter anvendes til at skabe en klassifikationsmodel, hvorefter modellen kan anvendes til at klassificere hvilken klasse en ny UTH-rapport tilhører.

en model ved hjælp af featureudvælgelse og klassifikation af UTH-træningsrapporter. Der er

i første omgang taget udgangspunkt i at anvende et bestemt søgeord 'EPM' og ud fra dette udvikle en model. Senere kan det testes, om det kan overføres til hvilket som helst søgeord, hvilket vil blive nærmere beskrevet i kapitel 9. Selve udviklingen af modellen består af tre overordnede trin [Duda et al. 2001]:

- 1. Præprocessering af UTH-rapporterne, hvis formål er at ensarte disse.
- Featureudvælgelse, hvis formål er at opnå karakteristika om de positive og negative UTH-rapporter og udvælge de features som skelner bedst mellem den positive og den negative klasse.
- 3. Klassifikation af UTH-rapporterne, hvis formål er at få opdelt UTH-rapporterne i årsag og ikke årsag.

Når udviklingen af modellen er afsluttet er denne i stand til, at klassificere nye UTH-rapporter med EPM som søgeord. For at kunne lave en model, som er i stand til at klassificere korrekt kræver det, at de nye rapporter har samme karakter, som de rapporter modellen er bygget ud fra [Weiss et al. 2005].

Den første boks på figur 6.1, 'UTH-Træningsrapporter', er tidligere omtalt i kapitel 5, hvor de anvendte UTH-rapporter er beskrevet og kvaliteten af disse er undersøgt. I dette kapitel vil den næste boks 'Udvikling af Klassifikationsmodel' blive gennemgået. indeholdende tre trin; NLP-præprocessering, Featureudvælgelse og Valg af klassifikationsalgoritme.

6.1 NLP-Præprocessering

Formålet med præprocesseringen er at tilføre yderligere information til teksten. Den yderligere information består i at hvert ord får tilknyttet forskellige former for grammatisk information, hvilket kaldes *NLP-tagning*. Til *NLP-tagningen* anvendes der tre værktøjer udviklet af CST. Værktøjerne kan downloades og anvendes under GNU licensen, på nær nogle specielle danske lingvistiske resurser, som skal anvendes hvis værktøjerne skal benyttes på dansk. De danske resurser er frit tilgængelige til ikke-kommercielt brug. [cst.dk 2008]

NLP er en del af området indenfor 'kunstig intelligens' og refereres også til som 'beregneligt sprog'. NLP arbejder med analyse, forståelse og automatisk generering af naturligt sprog. Hvilket betyder, at NLP spænder over flere grene af videnskaben, da det både omhandler sprogvidenskab og computervidenskab. [Dale et al. 2000] For yderligere beskrivelse af NLP, se appendiks B. NLP kan anvendes på mange niveauer i teksten, fra det enkelte ords grammatik til sætningsniveau. Hvilket niveau, der anvendes, afhænger af hvad formålet er. Her er formålet, at klassificere hvilken af de to ønskede klasser en UTH-rapport tilhører.

I Region Hovedstaden ligger UTH-rapporterne i en Access database. I databasetabellen tilsvarer hver række en sag, og hver kolonne tilsvarer fritekstfelterne fra rapporteringsskemaet; Hændelsesbeskrivelse, Årsag, Konsekvens, Forslag til Tiltag, Potentiel SAC score og Faktuel SAC score. For at bevare den informationsgivende struktur under NLP-tagningen er første skridt i præprocesseringen at eksportere UTH-rapporterne til en tabulatorsepareret fil. Hvert ord tagges med NLP-værktøjerne; Tokeniser, Lemmatiser og Part-Of-Speech(POS) tagger:

Tokeniser opdeler teksten i tokens. Tokens kan bestå af ord, interpunktionstegn og andre tegn. Tokeniseringen sker ved adskillelse af tokens med et mellemrum. Den anvendte tokeniser fra CST samler ligeledes flerordforbindelser som for eksempel 'i forhold til' til en token. [cst.dk 2008]

Lemmatiser Lemmatiseren beregner grundformen af ord på baggrund af en række regler og en ordbog, der begge afspejler forholdet mellem ordformer og grundformer. Lemmatiseren tager ikke højde for ordets grammatiske type, for eksempel skelnes der ikke imellem om ordet høj, er et adjektiv eller et substantiv. For at kunne skelne mellem disse kræves information fra POS-taggeren. [cst.dk 2008]

POS-tagger POS-tags kan bruges i viderebehandling af teksten for eksempel til at udtrække bestemte ordklasser, eller ordklassemønstre. Et POS-tag kan for eksempel være 'N_INDEF_SING' til ordet 'journal', hvor POS'en betydet, at ordet er et substantiv i ubestemt ental. For at kunne kategorisere alle slags ord, til en POS, har projektet 'Penn Tree bank' [Marcus et al. 1993], lavet en liste med 38 forskellige POS-tags. Disse POS-tags er oversat fra engelsk til dansk af CST og anvendes i POS-taggeren. [cst.dk 2008] I appendiks B findes en oversigt over de danske POS-tags.

Sidste del i præprocesseringen består i at kombinere resultaterne fra de tre NLP-værktøjer, således, at hvert ord bliver repræsenteret ved: Ord, Grundform og POS, idet den oprindelige information derved er bevaret samtidigt med at der er tilføjet ny information, se figur 6.2 for et eksempel på denne fremstilling. Denne struktur gør det muligt at udtrække, både generelle features samt NLP-features. Herefter har præprocesseringen skabt et grundlag for at lave en featureudvælgelse, af både det oprindelige ord, grundform og POS.

Ord	Grundform	POS
Forvagten optager journal , herunder Patientens medicinanamnese .	forvagt optage journal , herunder patient medicinanamnese	NDEF_SING V_PRES V_INDEF_SIN TEGN ADV N_DEF_SING_GEN N_INDEF_SING

Figur 6.2: Præprocesseringen fører til et output, hvor første kolonne er det oprindelige ord, anden kolonne er det lemmatiserede ord og tredje kolonne er den grammatiske type POS.

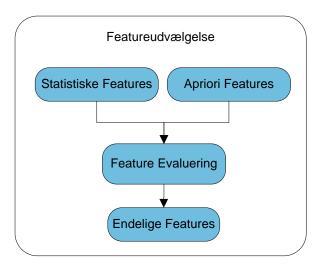
6.2 Featureudvælgelse

Featureudvælgelse sker på baggrund af mønstergenkendelse. Sammenhængen mellem at lave en god featureudvælgelse og en korrekt klassifikation er arbitrær. Hvis featureudvælgelsen er ideel, vil det betyde, at repræsentationen af UTH-rapporterne vil gøre klassifikationen triviel. Ligeledes vil en ideel klassifikationsalgoritme ikke kræve det store af featureudvælgelsen. [Weiss et al. 2005] Det er dog sjældent muligt at lave hverken featureudvælgelsen eller

klassifikationen triviel [Duda et al. 2001].

Som tidligere nævnt kan klassifikation af UTH-rapporter ses som en dokumentklassifikation. Indenfor dokumentklassifikation bliver features ofte udvalgt ved hjælp af algoritmen, Bag of Words (BOW) der er yderligere beskrevet i appendiks C. I BOW er en rapport repræsenteret ved features, som repræsenterer en uordnet samling af ord uden hensynstagen til hverken grammatik eller ordenes rækkefølge. Samlingen af ord er repræsenteret ud fra alle ordforekomster i UTH-træningsrapporterne. Ved afprøvning af denne algoritme for UTH-træningsrapporter, blev 1 ud af 17 positive rapporter klassificeret korrekt og alle på nær en af de negative rapporter blev klassificeret korrekt. Dette bekræfter, at en featureudvælgelse ved brug af BOW features ikke er tilstrækkelig til håndteringen af UTH-rapporter. Dette skyldes, at de ord, der indgår i henholdsvis de positive og negative rapporter, minder for meget om hinanden. Derfor bør vægten lægges på at finde features, som kan adskille de to klasser.

At finde mønstre som kan adskille de steder, hvor EPM er årsagen, og de steder hvor EPM ikke er årsagen til den utilsigtede hændelse, kan gøres på to måder. Mønstre kan blive identificeret på baggrund af a priori viden eller på baggrund af statistisk information udtrukket fra UTH-rapporterne [Duda et al. 2001]. De to metoder refereres også til som empirisk og symbolsk NLP tilgang [Dale et al. 2000]. I denne uddragelse af features, er det interessant at anvende NLP-features. Featureudvælgelsen er en proces, hvor tilgangen er, at uddrage mulige features og efterfølgende udvælge de features, der bidrager med tilstrækkelig information til klassifikationsalgoritmen. Denne fremgangsmåde er skitseret på figur 6.3.

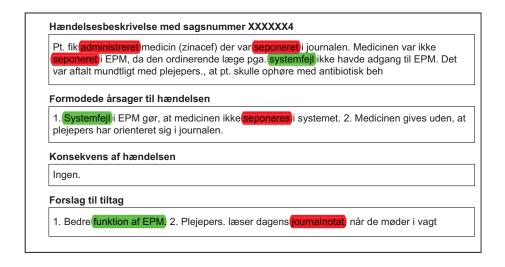


Figur 6.3: Featureudvælgelsesprocessen består af uddragelse af statiske og a priori features. Herefter evalueres features, for at ende op med de features, der skal anvendes til den endelige model.

6.2.1 Uddragelse af A Priori Features

A priori viden skal enten hentes fra litteraturen eller fra eksperter indenfor det givne domæne. For at opnå a priori viden om hvilke ord og fraser, der får eksperterne til at tage beslutningen om hvorvidt EPM er årsag eller ej, har eksperter systematisk gennemgået træningsrap-

porterne for at opnå denne viden. Eksperterne har læst de positive rapporter og de negative rapporter, og markeret hvilke ord og fraser, der får dem til at klassificere rapporterne, som de gør. Implicit i denne handling, ligger den viden de har om de danske sprog, som blandt andet inkluderer, at sætninger på dansk ikke kun består af ord men af led, som kæder ordene sammen. Et eksempel, er udvalgte fraser fra de positive træningsrapporter 'teknisk fejl', 'advare ikke', 'automatisk seponere', 'epm struktur', 'trykke forkert', 'taste forkert', 'ændring af epm' og 'funktion af epm'. Fraserne er efterfølgende blevet gjort mere generelle ved for eksempel at ændre frasen 'ændring af emp', til 'ændring PRÆP epm' hvor PRÆP står for præposition, således at det er uafhængigt af om der står 'i', 'på' eller 'af' foran ordet epm. Figur 6.4 viser et eksempel på en positiv UTH-rapport med markering de steder, hvor a priori



Figur 6.4: Et eksempel på en positiv UTH-rapport, hvor a priori ord, som karakteriserer positive UTH-rapporter, er markeret med grøn og a priori ord, der karakteriserer de negative UTH-rapporter, er markeret med rødt. Ligeledes er frasen 'funktionen af EPM' markeret.

features er tilstede. Røde markeringer indikerer karakteristika for den negative klasse, mens grønne markeringer indikerer karakteristika for den positive klasse.

På baggrund af de identificerede mønstre og den sproglige viden er der uddraget fire typer af a priori features; 1) Udvalgte ord 2) Udvalgte Fraser 3) Udvalgte verber med EPM i samme sætning og 4) Udvalgte substantiver med EPM i samme sætning, se tabel 6.1. Disse fire typer af features forholder sig både til de positive træningsrapporter og de negative træningsrapporter, hvilket fører til i alt otte individuelle features.

Repræsentation af alle a priori features med en featurevektorer for hver UTH-rapport kan ses i appendiks D.

6.2.2 Uddragelse af Statistiske Features

Der er 12 typer af statistiske features, vist i tabel 6.2 fra nummer 5 til 17, der alle er udtrukket på baggrund af træningsrapporter. Disse er; 5) EPM i Hændelsesbeskrivelsen, 6) EPM i Årsagsbeskrivelsen, 7) EPM i Konsekvensbeskrivelsen, 8) EPM i Tiltagsbeskrivelsen, 9) Antal EPM, 10) Første sætningsnummer med EPM, 11) Første afsnit med EPM, 12) Potentiel SAC-

A priori Features	A priori Type Lukket gruppe	Featurevektor
1) Udvalgte Ord	Ord	[1, 0]
2) Udvalgte Fraser	Fraser	[1, 0]
3) Udvalgte Verber + EPM	Verber	[1, 0]
4) Udvalgte Substantiver + EPM	Substantiver	[1, 0]

Tabel 6.1: Featureudvælgelsen baseret på a priori viden. De fire featuretyper er uddraget for både de positive UTH-rapporter og de negative UTH-rapporter, hvilket betyder at der i alt er otte features. Repræsentationen [1, 0] angiver de mulige featureværdier. Det vil sige, at hvis et givent ord eller frase er tilstede i en UTH-rapport, angives '1', og '0' hvis det ikke er tilstede. En oversigt over de identificerede ord og fraser kan ses i appendiks på figur D.4 på side 111

score, 13) Faktuel SAC-score, 14) Substantiver, 15) Verber, 16) Tre ords fraser og 17) Tre ords POS, se tabel 6.2 for en samlet oversigt over de enkelte features.

I features nummer 5, 6, 7, 8 og 9 undersøges det, om der er forskel på hvor mange gange ordet 'EPM' fremkommer i de fire fritekstfelter. Forskellen i frekvensen er evalueret ved hjælp af udregning af RMS værdierne for både den negative og den positive klasse. Feature 5, 6 og 7 er valgt, fordi der er en tendens til, at ordet 'EPM' fremkommer oftere i de negative fritekstfelter for henholdsvis hændelsesbeskrivelse, årsag til hændelse og konsekvens af hændelse. Feature nummer 8 er for fritekstfeltet Tiltag, hvor tendensen er, at ordet 'EPM' forekommer oftere i de positive rapporter. Feature nummer 9 repræsenterer at, der samlet set er en tendens til, at ordet 'EPM' fremkommer flere gange i de positive rapporter end i de negative.

I feature nummer 10, 'Første sætningsnummer med EPM', er tendensen, at EPM i positive rapporter forekommer længere nede end i negative rapporter. I feature nummer 11, 'Første afsnit med EPM', er tendensen den samme som i den forrige feature.

Feature 12 og 13 repræsenterer henholdsvis den potentielle og faktuelle SAC-score, som hver rapporten får tildelt.

Verberne i henholdsvis de negative træningsrapporter og positive træningsrapporter er udskrevet og sammenlignet automatisk. De verber, som ikke er tilstede i de modsatte rapporter, er lavet til en ordbogsfeature. En ordbogsfeature er en liste af ord [Weiss et al. 2005], i dette tilfælde verber, hvor hvert enkelt ord i UTH-rapporten sammenlignes med om et af verberne fra ordbogsfeaturen er tilstede eller ej. Ved at lave denne type features er støj et problem på grund af de mange stavefejl, som fremkommer i UTH-rapporterne. Stavefejl bevirker, at et ord, der er stavet forkert kan blive tilføjet ordbogsfeaturen. For eksempel hvis ordet 'skal' er blevet til 'skla' ved en tastefejl i en positiv UTH-rapport, kommer den positive ordbogsfeature til at indeholde ordet 'skla'. Dette kan betragtes som støj, da der ikke er højere sandsynlighed for en negativ UTH-rapport frem for en positiv, ved forekomsten af dette ord. Ordbogfeatures bliver ligeledes konstrueret med substantiver for både den negative og positive klasse og udgør dermed feature nummer 14 og 15. Da der både er substantiver og verber for henholdsvis den positive og negative klasse, er der reelt fire features, selvom de i tabel 6.2 kun er vist som to features for overskuelighedens skyld.

Baseret på træningsrapporter er alle treordsfraser optalt, de fraser, der har en frekvens på

Statistiske Features	Featurevektor
5) EPM i Hændelsesbeskrivelsen	[1, 0]
6) EPM i Årsagsbeskrivelsen	[1, 0]
7) EPM i Konsekvensbeskrivelsen	[1, 0]
8) EPM i Tiltagsbeskrivelsen	[1, 0]
9) Antal EPM	Numerisk
10) Første sætningsnummer med EPM	Numerisk
11) Første afsnit med EPM	[1,2,3,4]
12) Potentiel SAC-score	[1,2,3]
13) Faktuel SAC-score	[1,2,3]
14) Substantiver	[1, 0]
15) Verber	[1, 0]
16) Treords fraser	[1, 0]
17) Treords POS	[1,0]

Tabel 6.2: Featureudvælgelsen af features baseret på statistisk information.

fire eller derover, i dette tilfælde 21, som kan ses i appendiks D, anvendes for at se om de er tilstede eller ikke tilstede i rapporten. Disse udgør feature 16. Samme princip er anvendt ved feature 17, for treords POS, der fremkommer i sammenhæng.

Uddragelsen af henholdsvis statistiske og a priori features er fortaget på baggrund af, at jo flere features der er til stede jo højere sandsynlighed er der for at diskriminere imellem klasserne. Denne antagelse ville være korrekt i tilfælde af, at en uendelig stor mængde træningsrapporter var tilstede. [Hall 1999] Men da dette ikke er tilfældet, bør der tages højde for eventuel fremkommen redundans og overtilpasning af UTH-rapporterne. Redundante features vil betyde unødvendig processering og dermed en øget processeringstid uden øget information. Mest kritisk er det, ved overtilpasning af UTH-rapporterne, hvor features er overtilpasset til træningsrapporterne, hvilket kan medføre, at testrapporterne vil blive fejlklassificeret på baggrund af en overtilpasset model [Duda et al. 2001]. For at undgå disse to problematikker er det nødvendig at evaluere de 17 typer identificerede features. Reelt er der 71 features, hvilket skyldes at nogle features findes for både positive klasse og den negative klasse, samt at hver tre ords frase er tilstede som en enkelt feature.

6.2.3 Evaluaring af Features

Algoritmer til featureevaluering udfører en søgning i featurerummet, hvor hver enkelt feature evalueres. Dette kræver en kombination af en featureevalueringsmetode og en søgealgoritme. Evalueringsmetoden vælges på baggrund af, at der ønskes at ende op med features, som

har lav korrelation klasserne imellem samtidig med, at der tages højde for den individuelle prediktive evne for hver feature, sammen med graden af redundans her i blandt.

Metoden til evaluering af features kan opdeles i to kategorier, Feature Ranking og Subset Selection, der begge er tilgængelige i softwareprogrammet Weka. Weka er et program beregnet til data mining og machine learning. Det er udviklet af University of Waikato i New Zealand og er en forkortelse for 'Waikato Environment for Knowledge Analysis', og er frit tilgængeligt på http://www.cs.waikato.ac.nz/ml/weka/.

Feature Ranking rangerer features ud fra mængden af information de bidrager med, hvorefter det er muligt at eliminere alle features, som ikke opnår en fastsat tærskelværdi. Subset Selection søger et sæt af features som bedst repræsenterer adskillelsen mellem klasserne. [Witten & Frank 2005]

Det vælges at anvende Feature Ranking metoden, hvor features udvælges efter evalueringsmetoden Corelationbased Feature Selection og søgemetoden GreedyStepwise. GreedyStepwise udfører en 'grådig' søgning forlæns og baglæns gennem featurerummet af featureundergrupper. Søgningen startes enten ved ingen features eller et arbitrært punkt i featurerummet. Metoden gennemløber alle features, og de der fører til et fald i evalueringen vil ikke blive tilføjet. Det endelige resultat for featureudvælgelsen består af seks features, som ses i tabel 6.3.

Udvalgte Features	Featurevektor
1) Udvalgte ord	[1,0]
2) Udvalgte Fraser	[1, 0]
5) EPM i Hændelsesbeskrivelsen	[1, 0]
8) EPM i Tiltagsbeskrivelsen	[1, 0]
12) Potentiel SAC-score	[1,2,3]
15a) Statistiske Positive Verber	[1,0]

Tabel 6.3: De seks features der er tilbage efter evalueringen, af de 17 features, som blev uddraget oprindeligt. 15a) betyder at det kun er for den positive klasse.

6.3 Klassifikationsalgoritmer

Efter udvælgelse af de seks features, som skal bruges til den endelige klassifikation, skal det besluttes hvilken form for klassifikationsalgoritme, der skal anvendes. Klassifikationsalgoritmer kan deles ind i to undergrupper; supervised og unsupervised klassifikation. Ved supervised læring, definerer en ekspert kategorierne på forhånd, hvorimod ved unsupervised læring former algoritmen naturlige grupper af mønstre, hvor der ikke tages højde for a priori viden om kategorierne. På nuværende tidspunkt er der allerede anvendt supervised information til udtræk af features, derfor skal valget af klassifikationsalgoritme findes inden for den supervised gruppe. [Duda et al. 2001] De tre oftest benyttede supervised klassifikationsalgoritmer indenfor tekstklassificering, som det ønskes at undersøge, er Bayes, K-nærmeste Nabo (KNN) og Suport Vector Machines (SVM). [Colas & Brazdil 2006] Klassifikationsalgoritmerne testes ved en 10-fold krydsvalidering, hvor princippet består i at opdele træningssættet i 10 dele,

hvor de ni af delene anvendes til træning og den sidste del til test. Processen forløber iterativt i ti delprocesser, hvor det tiende testsæt skifter plads for hver gang. [Witten & Frank 2005] Til at måle præstationen af klassifikationsalgoritmerne er målene; præcision, recall og F-mål anvendt, for både den negative og positive klasse. Disse mål er valgt eftersom de oftest anvendes inden for klassifikation af tekst [Buckland & Gey 1994]. Præcision er bestemt ved antallet af korrekt klassificerede i forhold til antallet af klassificerede. For den positive klasse er dette udtrykt ved:

$$Præcision = \frac{Sand\ Positiv}{Sand\ Positiv + Falsk\ Positiv}$$
(6.1)

Recall er et forhold som måler antallet af sande positive i forhold til antal positive

$$Recall = \frac{Sand\ Positiv}{Sand\ Positiv + Falsk\ Negativ}$$
(6.2)

F-mål er et vægtet harmonisk gennemsnit af præcision and recall og er givet ved:

$$F-\text{mål} = 2 \frac{\text{Præcision} \cdot \text{Recall}}{\text{Præcision} + \text{Recall}}$$
(6.3)

Bayes Klassifikationsalgoritme

Bayes klassifikationsalgoritme er baseret på antagelsen om, at beslutningsproblemet er stillet probabilistisk, og at alle relevante probabilistiske værdier er kendte, hvilket betyder at information om den rapport, der skal klassificeres, er givet inden klassifikationen. Den naturlige

SP Rate	ED Data				
	FP Rate	Præsision	Recall	F-Mål	Klasse
0.824	0.028	0.875	0.824	0.848	Positiv
0.972	0.176	0.958	0.972	0.965	Negativ
Forudsagt					

. 0. 0.	asage				
Posi	tiv	Ne	egativ		-
4	14	×	3	Positiv] 욹
×	2	\checkmark	69	Negativ	Š

Figur 6.5: Resultaterne for de seks udvalgte features klassificeret ved hjælp af Bayes klassifikationsalgoritme. SP står for sand positiv, FP står for falsk positiv, præcision, recall og F-mål er valideringsmål. Tallet markeret med et grønt √ indikerer, at rapporterne er klassificeret korrekt, hvorimod det røde kryds betyder fejlklassifikationer.

tilstandsform er beskrevet ved et antal stadier $(\omega_1, \omega_2, ..., \omega_n)$, hvor n tilsvarer antallet af klasser. I dette tilfælde er der kun to klasser 'årsag' eller 'ikke årsag', altså n=2. Det antages, at sandsynligheden for hvert stadie $P(\omega_n)$ er givet, såvel som for den betingede sandsynlighed $p(x|\omega_n)$. Den betingende sandsynlighed er givet ved [Duda et al. 2001]:

$$p(\omega_n, x) = P(\omega_n | x) p(x)$$

eller

$$p(\omega_n, x) = p(x|\omega_n)P(\omega_n)$$

Disse to formler leder frem til Bayes sætning:

$$P(\omega_n|x) = \frac{p(x|\omega_n)P(\omega_n)}{p(x)}$$

eller

$$posterior = \frac{sandsynlighed \cdot forudgående}{evidens}$$
 (6.4)

hvor beslutningsreglen er baseret på:

Beslut
$$\omega_1$$
 Hvis $P(\omega_1|x) > P(\omega_2|x)$; Ellers Beslut ω_2

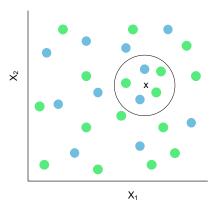
Beslutningsreglen kan beskrives som betinget sandsynlighed:

Beslut
$$\omega_1$$
 Hvis $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$; Ellers Beslut ω_2

Resultaterne for en 10-fold krydsvalidering af Bayes klassifikationsalgoritme med de seks udvalgte features giver et positivt F-mål på 0,824 og et negativt F-mål på 0,957, se figur 6.5. En af fordelene ved Bayes klassifikationsalgoritme er at hele processen, kun skal forløbe en gang, og at den er robust i for hold til invariante features. Derimod er en af ulemperne, at den kræver det statistisk uafhængighed imellem features. [Vilalta 2008]

K-Nærmeste Nabo

KNN klassificerer en rapport x ved at tilskrive det en label, som oftest fremkommer i blandt de k nærmeste naboer i featurerummet, dette ses på figur 6.6. Det betyder, at beslutningen er baseret på undersøgelse af de k nærmeste naboer, hvor det er flertallet, der bestemmer hvilken klasse rapporten x tilhører. Værdien for k bør være ulige for at sikre, at de to sandsynligheder ikke er ens. [Duda et al. 2001] 10-fold krydsvalidering af KNN klassifikationsalgoritme med



Figur 6.6: KNN starter med et træningsrapport x og vokser i en sfærisk region, indtil den har indelukket k træningsrapporter og klassificere rapporterne ud fra flertallet. I dette tilfælde hvor k=5, vil træningsrapporten x komme til at tilhøre klassen for de grønne punkter. [Duda $et\ al.\ 2001$] - modificeret

de seks udvalgte features giver et positivt F-mål på 0,733 og et negativt F-mål på 0,944, se figur 6.7. Disse mål viser, at KNN er dårligere til at klassificere de positive resultater sammenlignet med Bayes klassifikationsalgoritme. Hvis valget stod imellem Bayes og KNN ville Bayes klassifikationsalgoritme være at fortrække, da den præsterer bedre. Fordelen ved

KNN					
SP Rate	FP Rate	Præsision	Recall	F-Mål	Klasse
0.588	0.028	0.833	0.588	0.69	Positiv
0.972	0.412	0.908	0.972	0.939	Negativ
Forudsagt					
Positiv	Negativ		TI		
√ 10	X 7	Positiv	Faktis		
A	A CO		<u></u>		

Figur 6.7: Resultaterne for de seks udvalgte features klassificeret ved hjælp af KNN klassifikationsalgoritme. Øverst vises valideringsmålene og nederst blandingsmatricen.

KNN er at det er en robust klassifikationsalgoritme, hvis klasserne ikke er lineær adskillige. Hvorimod en ulempe er at den kræver høj beregningskraft idet et similaritetsmål skal beregnes for hver UTH-rapport. Yderligere håndterer den ikke ubalancerede klasser så godt.

Support Vector Machines

SVM deler featurerummet med en beslutningsgrænse i flere dimensioner;

$$g(x) = w^t x + w_0$$

i det binære klassifikations tilfælde, hvor w er vægtningsvektoren og w_0 er biasen eller tærskelværdien. En tilsvarende beslutningsregel er:

Beslut
$$\omega_1$$
 Hvis $g(x) > 0$; Ellers ω_2

Hvis g(x) = 0, vil x tilskrives begge kategorier. Diskriminantfunktionen g(x) beskriver distancen fra x til hyperplanet. Hyperplanets orientering bestemmes af w, og placeringen af w_0 . I dette tilfælde kan x bestemmes ved:

$$x = x_p + r \frac{w}{\|w\|}$$

hvor x_p er en normal projektion af x på H, og r er den ønskede algebraiske distance. [Duda $et\ al.\ 2001$]

SVM klassifikationsalgoritmen med en 10-fold krydsvalidering med de seks udvalgte features giver et positivt F-mål på 0,833 og et negativt F-mål på 0,956, se figur 6.8. På baggrund af de positive F-mål vil SVM klassifikationsalgoritmen ved en 10-foldskrydsvalidering være den af de afprøvede algoritmer, som opnår den bedste klassifikation af de positive UTH-rapporter. En af SVM-algoritmens fortrin er, at den er god til at håndtere ubalancerede klasser. Hvorimod hvis klasserne ikke er lineært adskillige er den ikke anvendelig. [Vapnik 1995]

6.3.1 Valg af Klassifikationsalgoritme

Efter gennemgang af de tre hyppigst anvendte klassifikationsalgoritmer inden for tekstklassifikation, er fordele og ulemper listet i forhold til den aktuelle problemstilling. En oversigt af

SVM					
SP Rate	FP Rate	Præsision	Recall	F-Mål	Klasse
0.882	0.042	0.833	0.882	0.847	Positiv
0.958	0.118	0.971	0.958	0.965	Negativ
Forudeagt					

Foruasagi				
Positiv	Ne	gativ		п
√ 15	×	2	Positiv	akt
X 3	4	68	Negativ	isk

Figur 6.8: Resultaterne for de seks udvalgte features klassificeret ved hjælp af SVM klassifikationsalgoritme. Øverst vises valideringsmålene og nederst blandingsmatricen.

F-mål							
Klassse	Bayes	KNN	SVM				
Positiv	0.848	0.69	0.857				
Negativ	0.965	0.939	0.965				

Figur 6.9: F-målene for klassifikationsalgoritmerne Bayes, KNN og SVM.

dette kan ses i tabel 6.4. Ud fra 10-fold krydsvalideringsresultaterne, er det muligt at vælge endelig klassifikationsalgoritme, se figur 6.9. Når en sådan udvælgelse fortages, er det vigtig at have fokus på hvad formålet med klassifikationen er. I dette tilfælde er det at få klassificeret så mange som muligt af de positive rapporter korrekt, hvilket betyder at F-målet for den positive klasse skal være så højt så muligt.

Algoritme	Fordele	Ulemper
Bayes	Udførelsen er hurtig, idet den kun kræver en enkelt gennemkørsel af data. Robust ift irrelevante features. [Vilalta 2008]	Kræver statistisk uafhængighed i mellem features. [Vilalta 2008]
K-NN	Robust hvis klasserne ikke er lineær adskillige. [Feldman & Sanger 2006]	Kræver høj beregningskraft similariteten skal beregnes for hvert rapport. [Feldman & Sanger 2006]
SVM	God til ubalancerede klasser, håndtering af højdimentionelt featurerum. [Vapnik 1995]	Hvis klasserne ikke er lineært adskillige. [Vapnik 1995]

Tabel 6.4: De tre klassifikationsalgoritmer, med hver deres fordele og ulemper.

Det ses i figur 6.9, at Bayes og SVM giver et positivt F-mål, der ligger tæt op af hinanden. Dette understøtter, at KNN ikke håndterer ubalancerede klasser så godt, idet F-målet for den negative klasse ligger højt og den positive klasse ligger lavere. For SVM klassifikationen ses, at det positive F-mål er højere end ved Bayes, hvorimod der ved det negative F-mål ses

et højere mål for Bayes end for SVM. Med udgangspunkt i at formålet med klassifikationen er at klassificere så mange så muligt af de positive UTH-rapporter korrekt, og SVM er den klassifikationsalgoritme som opnår det bedste resultat, udvælges den til at blive benyttet til den endelige EPM klassifikationsmodel. Det er ligeledes tidligere vist, at SVM klassifikationsalgoritmen giver lovende resultater ved tekstklassifikation [Savova et al. 2007, Joachims 1998].

SVM algoritmen er intolerant overfor ubalancerede klasser i træningssæt [Drucker et al. 1999]. Dette er en egenskab, som er vigtig, for den klassifikationsalgoritme, som skal anvendes til at adskille UTH-rapporter med søgeordet 'EPM'. De fleste klassifikationsalgoritmer forsøger at reducere fejlraten ved at klassificere til den klasse som på forhånd har det højeste antal instanser på bekostning af den mindste klasse, hvilket ses ved både Bayes og KNN. SVM adskiller mønstrene i et høj dimensionalt featurerum, og på den måde gør den brugbar for ubalancerede klasser. [Drucker et al. 1999]

Klassifikationsfejl for Træningsrapporter

I tilfælde, som her, hvor der er en begrænset mængde UTH-rapporter tilstede er det vigtig at kigge på de fejlklassificerede UTH-rapporter. Dette gøres ud fra 10-fold krydsvalideringen for dermed at afgøre hvorvidt det er muligt at forbedre modellen og dermed klassificeringen. De fire fejlklassificerede falsk negative UTH-rapporter, havde sagsnummer xxxxxx5, xxxxxx6, xxxxxx7 og xxxxx8. og de to fejlklassificerede falsk positive rapporter havde sagsnummer xxxxxx9 og xxxxx10. Da der i problemstillingen primært er fokus på korrekt klassifikation af de positive UTH-rapporter, er det valgt at se nærmere situationerne hvor disse fejlklassificeres.

Feature	xxxxxx9	xxxxx10
1) Udvalgte ord	0	0
2) Udvalgte Fraser	0	0
5) EPM i Hændelsesbeskrivelsen	1	0
8) EPM i Tiltagsbeskrivelsen	0	1
12) Potentiel SAC-score	2	3
15a) Statistiske Positive Verber	1	1

Tabel 6.5: De to UTH-Rapporter, der bliver fejlklassificeret til negative, selv om de er positive.

I tabel 6.5 vises featureværdierne for to sagsnumre i henhold til de udvalgte seks features. I rapporten med sagsnummer xxxxxx9 har eksperterne markeret sætningen: Det oplyses, at EPM var "lukket" på bestillingstidspunkt, hvorfor der er udfyldt rekvisition, som karakteristisk for den positive klasse. Dette har ført til, at frasen 'EPM var lukket' er blevet tilføjet i featuren 'Udvalgte Fraser', men årsagen til at denne frase ikke bliver registret som et ettal i featuren, hvilket skyldes, at "lukket"er sat i anførelsestegn. Denne problematik opstår på grund af, at rapportøren er i tvivl om hvorvidt 'lukket' er det rigtige ord at bruge. En mulighed for at sikre at denne feature fremover bliver registret korrekt er at fjerne

alle anførelsestegn i rapporterne i forbindelse med præprocesseringen. Featurekombinationen 0,0,1,0,2,1 fremkommer i seks andre sager, som alle er negative sager, hvorimod featurekombinationen 1,0,1,0,2,1 ikke findes i andre sager, og der er derfor potentiale for, at den ville blive klassificeret som positiv, hvis den havde fået et ettal i featuren Udvalgte Fraser.

I den anden fejlklassificerede rapport med sagsnummer xxxxx10, er der i fritekstfeltet 'Formodede Årsager' af eksperterne markeret Hændelsen skyldes EPM's struktur, hvilket betyder, at 'epm struktur' er tilføjet til Udvalgte Fraser, men på grund af en lemmetiseringsfejl bliver 'EPM's struktur' lemmatiseret til 'epm's' og 'struktur', hvilket betyder, at den heller ikke bliver registreret som værende tilstede i de udvalgte fraser. Hvis denne sag havde været registreret som værende tilstede ville featurevektoren være blevet '1,0,0,3,1,1'. En anden UTH-rapport har også denne featurevektor, og den er klassificeret som positiv.

Efter processen med at undersøge årsagen til de fejlklassificerede rapporter er fuldført, er de to punkter blevet korrigeret og resultatet af dette kan ses på figur 6.10, hvor det ses at det positive F-mål er blevet forbedret markant.

SVM					
SP Rate	FP Rate	Præsision	Recall	F-Mål	Klasse
1	0.056	0.8	1	0.889	Positiv
0.944	0	1	0.944	0.971	Negativ
Forudsagt					
Positiv	Negativ		THE STATE OF		
√ 17	X 0	Positiv	Faktis		
Y	A 65	Mogativ	is.		

Figur 6.10: Nye UTH-rapporter bliver klassificeret ved, at de seks features bliver udtrukket for hver UTH-rapport og herefter klassificeret ved brug af SVM klassifikationsalgoritmen til enten at tilhøre den positive klasse eller den negative klasse.

Den færdige årsagsklassifikationsmodel er nu konstrueret, den klassificerer UTH-rapporterne til én af to klasser. Modellen består af; NLP-præprocessering, seks features og en SVM klassifikationsalgoritme. Hermed er use-casen 'Find EPM i UTH-rapporter' og 'Klassificer Rapporter ved brug af NLP' blevet implementeret.

Kapitel 7

Udvikling af Brugergrænseflade

For at udvikle et samlet system, der opfylder alle de krav, som Enhed for Patientsikkerhed har, er det nødvendigt at udvikle en brugergrænseflade, således at brugeren kan interagere med klassifikationsmodellen.

Brugergrænsefladen skal gøre det muligt at udføre følgende use-cases: 'Åbn Tidligere Årsagsbestemmelse, 'Vælg Gemt Fil', 'Foretag EPM Årsagsbestemmelse', 'Vælg Kildefil', 'Indtast Årsag', 'Se Positivt Klassificerede Rapporter', 'Se Negativt Klassificerede Rapporter', 'Sorter Rapporter' og 'Udskriv Rapporter' i henhold til design af systemet, som tidligere er vist på figur 4.2 i kapitel 4. Disse use-cases er således alle blevet opstillet ud fra brugerens krav til systemet.

Et godt design af en grafisk brugergrænseflade er karakteriseret ved at det er nemt at lære, effektivt at bruge samtidigt med at det giver brugeren en god oplevelse [Molich 1999]. Ved design af brugergrænseflader tales der om brugervenlighed, som dækker over fem kvalitetskomponenter [Nielsen 1998]:

Let at lære: Systemet bør være let at lære, således at brugeren let kan komme i gang med at anvende systemet. Dette kommer sig for eksempel til udtryk i, hvor nemt det er for brugeren at fuldføre basale opgaver første gang de anvender designet.

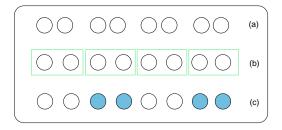
Effektivt at bruge: Systemet bør være effektivt at bruge således, at når brugeren først har lært designet at kende, kan vedkommende hurtigt udføre basale opgaver.

Let at huske: Systemet bør være nemt at huske, således at brugeren kan vende tilbage til systemet, efter ikke at have anvendt det over en periode, og dermed have let ved at genskabe deres færdigheder.

Antallet af fejl: Systemet bør have en lav fejlhyppighed, således at brugeren kun laver få fejl ved brug af systemet, og hvis de laver fejl bør disse være lette at rette. Der må således ikke opstå katastrofale fejl, der gør at brugeren ikke kan fortsætte med at bruge systemet.

Tilfredshed: Systemet bør være tilfredsstillende at anvende. Det vil sige, at brugeren finder det behageligt at anvende designet.

Ved systemer der anvendes flere gange dagligt er effektivitet vigtigt, hvorimod hvis systemet anvendes sjældent er indlæring og genindlæring af større betydning [Molich 1999].



Figur 7.1: De grundlæggende principper i brugervenligt design; (a) Nærhed , (b) Lukkethed og (c) Lighed. [Molich 1999]

7.1 Brugervenligt Design

Den teoretiske tilgang for at opnå et brugervenligt design ved at tage udgangspunkt i 3 overordnede principper, der alle er med til at skabe et enkelt og logisk design af brugergrænsefladen. Disse er nærhed, lukkethed og lighed, der er illustreret på figur 7.1 [Molich 1999]. Nærhed: Ved at placere elementer der logisk set hører sammen tæt ved hinanden, opnås et bedre og mere intuitivt design, idet disse elementer vil opfattes som sammenhørende. Det kan for eksempel være at placere en knap, der hører til et bestemt tekstfelt så tæt på tekstfeltet som muligt. Generelt bør elementer placeres til højre for det element det tilhører. Hvis det ikke kan lade sig gøre kan elementet alternativt placeres nedenfor. Afstanden imellem grupperne bør skabes ved blanke linier eller søjler, således at der skabes store mellemrum imellem elementer, der logisk set ikke hører sammen, se figur 7.1 (a).

Lukkethed: Elementer der står indenfor den samme ramme, opfattes som hørende samme. Dette illustreres ved at placere en ramme omkring de sammenhørende elementer, se figur 7.1 (b).

Lighed: Der er en tendens til, at brugeren vil koble elementer sammen, der ligner hinanden meget. Elementerne på en brugergrænseflade bør placeres så elementernes kanter flugter med så mange andre sammenhørende elementer som muligt. Dette kan enten være i forhold til den venstre eller højre kant eller toppen eller bunden af elementet, se figur 7.1 (c).

Ved at anvende disse tre overordnede principper i design af brugergrænseflader, vil det blive lettere for brugeren at anvende systemet, idet den grafiske udformning er logisk, hvilket er med til at øge tilfredsheden, når systemet er i brug.

7.1.1 Virkemidler

Udover de tre grundlæggende principper i logisk design findes der en række virkemidler, der kan bidrage til et mere brugervenligt design. Disse er brug af titel, dialogbokse, hjælpefunktioner, ikoner og værktøjstips [Molich 2003], og vil alle blive gennemgået nærmere i det følgende.

Titel: Titlen på systemets brugergrænseflade skal være sigende i forhold til indhold eller

funktion. På hver enkelt brugergrænseflade placeres titlen i den øverste ramme på vinduet. Titlen kan suppleres med kodeværdier som for eksempel navnet på den fil brugeren har åbnet. [Molich 1999]

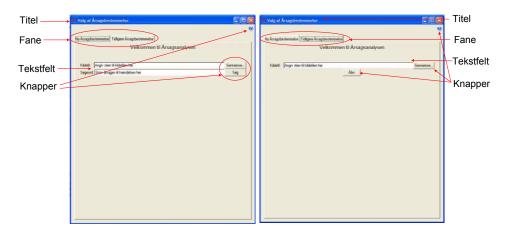
Dialogbokse: For at guide brugeren af systemet igennem de forskellige funktionaliteter kan det være anvendeligt med dialogbokse. Meddelelsen i dialogboksene bør være konstruktiv, præcis og undskyldende. Overordnet kan meddelelser inddeles i 3 grupper: Fejlmeddelelser, advarsler og informative beskeder. Ved en fejlmeddelelse kan systemet ikke fortsætte behandlingen af den aktuelle opgave før brugeren har foretaget en korrektion. En fejlhandling kan for eksempel være, hvis brugeren ønsker at påbegynde en ny årsagsbestemmelse, men har gemt at specificere et søgeord. I sådant et tilfælde bør der være en dialogboks, der gør brugeren opmærksom på dette. En advarsel bruges, når det er sandsynligt, at der kan opstå en uønsket situation. Først når brugeren har bekræftet, at meddelelsen er læst kan vedkommende forsætte. Dette er for eksempel anvendeligt hvis brugeren klikker på afslut knappen, for dermed at sikre at brugeren ikke lukker systemet i utide. Endeligt kan der bruges informative meddelelser, der fortæller brugeren, at alt går som det skal. [Molich 1999]

Hjælpefunktion: En indbygget hjælpefunktion kan ligeledes oplyse brugeren om systemets funktionalitet. Hjælpefunktioner fremkommer, hvis brugeren selv ønsker det. Dette kan for eksempel være ved at klikke på en hjælp-knap, som skal være tilgængelig på hver eneste brugergrænseflade. [Molich 1999]

Ikoner: Brugen af ikoner kan øge brugervenligheden, såfremt betydning af ikonet er entydigt. Ikoner kan være nemmere at overskue end tekst, og kan dermed være med til at skabe en følelse af overblik for brugeren, og dermed øge brugerens tilfredshed med systemet. Brugen af ikoner kunne være hensigtsmæssig i forhold til funktionaliteter såsom gem, udskriv og åbn fil, da dette er kendte ikoner, som brugeren kender fra andre systemer. [Molich 1999]

Værktøjstip: Værktøjstip skal give en kortfattet hjælp til aktive elementer, det vil sige de elementer som brugeren kan klikke på. Tippet skal fremkomme, når brugeren lader markøren stå over elementet uden at klikke i mindst 1 sekund, og forsvinder når brugeren flytter markøren væk fra elementet eller senest efter 15 sekunder. [Molich 1999] Dette er for eksempel anvendeligt, når brugeren skal indtaste søgeordet, der angiver årsagen til den utilsigtede hændelse, idet det kan være gavnligt for brugeren at få oplyst, hvorledes dette søgeord skal specificeres.

Sprogbrug: Det er vigtigt at være opmærksom på sprogbruget i systemet. Teksten skal formuleres i et sprog som de brugere, der skal anvende systemet, forstår. Dette kan opnås ved at anvende termer, der på forhånd er kendte for brugeren. Generelt bør der bruges aktiv form ved tiltale, idet der herved ikke er tvivl om hvorvidt, det er systemet eller brugeren, der skal handle. [Molich 1999]



Figur 7.2: Brugergrænseflade til Ny Årsagsbestemmelse.

Figur 7.3: Brugergrænsefladen til Tidligere Årsagsbestemmelse.

7.2 Systemets Brugergrænseflader

Design af brugergrænsefladerne til systemet til årsagsklassifikation er foretaget i samråd med Enhed for Patientsikkerhed. At involvere brugeren tidligt i forløbet er afgørende for at sikre en hensigtsmæssig udformning af designet [dialogdesign.dk 2007]. Der er foreløbigt blevet afholdt tre møder med repræsentanter for Enhed for Patientsikkerhed. På første møde er systemets anvendelse blevet diskuteret, hvorefter brugeren på andet møde blev præsenteret for et mock-up, som er et papirbaseret udkast til design af brugergrænsefladerne. Derved havde brugeren mulighed for at komme med kommentarer inden den egentlige udvikling skulle påbegyndes. Endeligt er brugeren blevet præsenteret for en prototype, hvortil de har mulighed for at afprøve systemets funktionaliteter, og komme med kommentarer hertil, hvilket er beskrevet yderligere i kapitel 10 på side 59.

Brugergrænsefladen som fremkommer, når systemet åbnes, kaldes 'Valg af Årsagsbestemmelse' og den brugergrænseflade, hvorpå klassifikationerne kan ses, kaldes 'Resultat af Årsagsbestemmelse'.

7.2.1 Valg af Årsagsbestemmelse

Brugergrænsefladen 'Valg af Årsagsbestemmelse' består af to faner, idet der både er mulighed for at foretage en ny årsagsbestemmelse, indikeret ved fanen 'Ny Årsagsbestemmelse' samt åbne en allerede gemt årsagsbestemmelse vist ved fanen 'Tidligere Årsagsbestemmelse'. Brugen af faner er ofte at foretrække, idet de er overskuelige og nemme at forstå, såfremt systemet har flere parallelle funktionaliteter. Det er vigtigt at alle sammenhængende funktionaliteter placeres på samme faneblad og at fanebladene er uafhængige, da inddelingen i faner ellers blot vil være til gene for brugeren. [dialogdesign.dk 2007] De parallelle funktionaliteter, der er implementeret kan ses på figur 7.2 og 7.3. I følge kravene til systemet skal brugeren have mulighed for at specificere et søgeord samt en kildefil, når der skal foretages en årsagsbestemmelse. Dette er muligt på brugergrænsefladen 'Valg af Årsagsbestemmelse' under fanen 'Ny Årsagsbestemmelse'. Når vinduet åbnes er fanen 'Ny Årsagsbestemmelse' valgt som udgangspunkt, idet brugeren oftest har brug for at foretage en ny søgning, frem

for at åbne en allerede gemt søgning.

Brugeren har mulighed for at specificere kildefilen, der indeholder UTH-rapporterne ved direkte at indskrive stien til filen i et tekstfelt. Alternativt kan brugeren browse sig frem til den korrekte placering ved at klikke på 'Gennemse...'-knappen. De tre punktummer efter 'Gennemse' indikerer, at knappen ikke udfører sin funktionalitet umiddelbart, idet der åbnes et dialogvindue, hvori det er muligt at vælge den ønskede fil. Det er valgt at implementere begge disse muligheder, idet det tilgodeser såvel den erfarne bruger som nybegynderen.

Derudover skal brugeren vælge hvilket søgeord, der ønskes som årsag til den utilsigtede hændelse. Da systemet er et 'proof-of-concept', skal der i dette tilfælde indtastes 'EPM', 'EMP' eller begge dele, hvis der indtaste andre ord vil en advarselsboks advare om at søgningen ikke kan fortages. For at starte selve årsagsbestemmelsen skal brugeren som det sidste klikke på 'Søg'-knappen.

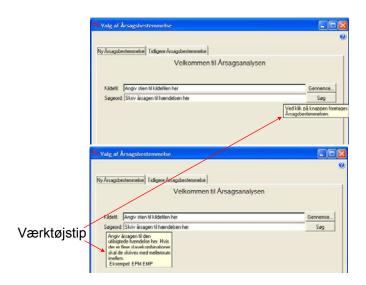
Hvis brugeren i stedet ønsker at åbne en allerede gemt årsagsbestemmelse skal der klikkes på fanen 'Tidligere Årsagsbestemmelse'. I dette tilfælde skal brugeren kun specificere den fil, der ønskes åbnet, idet der i en gemt årsagsbestemmelse allerede er valgt et søgeord. Dette gøres på samme måde, som når der foretages en ny årsagsbestemmelse med mulighed for direkte at indskrive stien til filen eller finde den ved at browse. For at åbne filen skal der klikkes på knappen 'Åbn'.

Generelt er det vigtigt, at brugeren ikke er i tvivl om konsekvensen af sine handlinger [Molich 2003]. Et af virkemidlerne der er brugt i designet hertil er at tage højde for at anvende tre punktummer på knapper, der ikke umiddelbart udfører en funktionalitet. Ligeledes er der som virkemiddel implementeret værktøjstip til de knapper der vil udføre en funktionalitet med det samme, knapperne 'Søg' og 'Åbn' og feltet, hvor årsagen til hændelsen skal angives. Eksempler på dette ses på figur 7.4. Derudover er der implementeret dialogbokse, der vejleder brugeren i tre situationer: 1) Hvis der søges uden at der forinden er specificeret et søgeord, 2) Hvis der søges med en ugyldig kildefil og 3) Når brugeren ønsker at lukket vinduet. Et eksempel på dette er vist på figur 7.5. For yderligere at hjælpe brugeren er der implementeret en hjælpefunktion, der tilgås ved klik på ikonet med spørgsmålstegn øverst i højre hjørne. Herved åbnes en dialogboks, der beskriver brugergrænsefladens funktionaliteter.

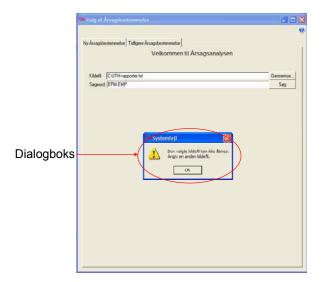
Ved udformning af årsagsbestemmelses brugergrænseflade opfyldes de følgende use-cases 'Åbn Tidligere Årsagsbestemmelse', 'Vælg Gemt Fil', 'Foretag EPM Årsagsbestemmelse', 'Vælg Kildefil' og 'Indtast Årsag'.

7.2.2 Resultat af Årsagsbestemmelsen

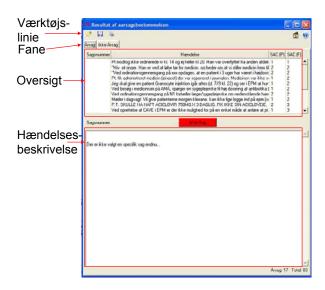
Resultatet af årsagsbestemmelsen er fremstillet på en brugergrænseflade kaldet 'Resultat af Årsagsbestemmelsen', der ligesom i 'Valg af årsagsbestemmelse' er inddelt i to faner; én med rapporter, hvor søgeordet er årsag til den utilsigtede hændelse, og én med de resterende rapporter, hvilket fremgår af figur 7.6. Denne implementering er valgt, da principperne i de to resultater er ens, idet brugeren har de samme muligheder ved henholdsvis rapporterne med EPM som årsag og de resterende rapporter. Det der adskiller fanerne er indholdet, det vil sige de rapporter brugeren har adgang til. Fanebladet med de positivt klassificerede rapporter er valgt som den fane, der vises, når brugergrænsefladen åbnes, da det er disse rapporter brugeren oftest vil være interesseret i. Der er fra brugeren side et ønske om at blive præsenteret for en oversigt såvel som hver enkelt rapport. Overvejelserne i forbindelse med



 $\begin{tabular}{ll} \textbf{Figur 7.4:} På figuren ses eksempler på anvendelse af værktøjstips for brugergrænsefladen 'Valg af Årsagsbestemmelse'. \\ \end{tabular}$



 $\textbf{Figur 7.5:} \ P \& \ figuren \ ses \ eksempler \ p \& \ anvendelse \ af \ dialogboks \ for \ brugergrænsefladen 'Valg \ af \ \& ragsbestemmelse'.$



Figur 7.6: Brugergrænseflade til de positivt klassificerede rapporter. Brugergrænsefladen for de negativt klassificerede rapporter er tilsvarende, blot at der er valgt en anden fane.

denne funktionalitet har blandt andet inkluderet teorien om nærhed, lukkethed og lighed, samt at der undersøgt hvilke programmer, der anvender lignende funktionaliteter. Princippet i at få vist en oversigt, samtidig med at få vist et markeret element fra oversigten, kendes fra mail-klient programmer, som MS Outlook Express og Thunderbird, hvor oversigten ses i øverste halvdel af skærmbilledet og det valgte element ses i nederste halvdel af skærmbilledet. Ligeledes kendes princippet, hvor en træstruktur i vestre side viser oversigten og det enkelte element er vist i højreside, hvor der er mulighed for at korrigere hvad størrelsesforholdet skal være mellem disse to dele. Det er blandt andet produkter, som anvendes i forbindelse med programudvikling, såsom MatLab og Eclipse, som anvender denne struktur. Ud fra disse betragtninger er valgt at implementere 'Resultat af Årsagsbestemmelsen' med inspiration fra mail-klient programmer, idet det er den opbygning som selv meget uerfarne brugere ville kunne håndtere.

Oversigten af 'Resultat af Årsagsbestemmelse' indeholder alle informationer om den enkelte rapport, det vil sige sagsnummer, hændelsesbeskrivelse, og den potentielle og faktuelle SAC-score. Ved at klikke på et overskriften 'Sagsnummer', 'SAC (P)' eller 'SAC (F)' er det muligt at sortere rapporterne efter den givne overskrift. Den valgte sortering vises ved, at den pågældende overskrift markeres. For at opfylde kravet om visning af hele hændelsesbeskrivelsen kan brugeren klikke på en rapport, og derved få vist hele teksten i et tekstfelt placeret nederst i vinduet. Kombinationen på brugergrænsefladen af en oversigt samtidigt med mulighed for at se alle oplysninger om en rapport, er hensigtsmæssigt, idet den opfylder alle brugerens behov. Selve teksten i tekstfeltet er inddelt i afsnit for at skabe et større overblik for brugeren.

Som tidligere nævnt er den endelige klassifikation først foretaget, når brugeren af systemet har gennemgået systemets bud på klassifikationer. Såfremt en hændelse er klassificeret forkert har brugeren mulighed for at flytte hændelsen til den anden fane ved at klikke på knappen 'Afvis Sag'. Herved blive hændelsen flyttet til den modsatte fane, og sagsnummeret vises med grønt, for at indikere at brugeren har godkendt denne ene klassifikation. Knappen 'Afvis

Sag' har en rød baggrundsfarve, da rød er en farve der oftest anvendes til at indikere noget forkert, hvorimod baggrunden for sagsnummeret skifter til grønt efter korrektionen, da dette er en farve, der forbindes med noget korrekt.

Endeligt er der en værktøjslinie allerøverst med ikoner. Værktøjslinier er karakteriseret ved ikoner, der illustrerer funktionaliteter. Afhængigt af programmets størrelse, bør der være mulighed for og tilføje eller fjerne værktøjslinier [useit.com 2008]. Antallet af funktionaliteter på grænsefladen 'Resultat af Årsagsbestemmelse' er fem og det er derfor valgt at lade dem være tilgængelig hele tiden. De fem funktionaliteter er: mulighed for at gemme klassifikationen, åbne en tidligere årsagsbestemmelse, udskrive hændelserne, vende tilbage til startvinduet og få hjælp. Når brugeren klikker på printikonet på brugergrænsefladen for 'Resultat af Årsagsbestemmelsen', vil der fremkomme en brugergrænseflade, der viser rapporterne som de vil blive udskrevet. Herved har brugeren mulighed for at udskrive en til standardprinteren.

Ligeledes er det muligt at vende tilbage til vinduet for valg af årsagsbestemmelse ved at klikke på ikonet 'Hjem'. Programmet afsluttes ved klik på krydset øverst i højre hjørne. Værktøjslinien er valgt frem for en menulinie, idet menulinien ville bestå af de samme funktionaliteter, men ikke lige så hurtig at tilgå.

For brugergrænsefladen 'Resultat af Årsagsbestemmelsen' er der ligeledes implementeret værktøjstips og dialogbokse for at vejlede brugeren, når systemet anvendes. Ved udformning af denne brugergrænseflade opfyldes de følgende use-cases 'Foretag EPM Årsagsbestemmelse', 'Se Positivt Klassificerede Rapporter', 'Se Negativt Klassificerede Rapporter', 'Filtrér Rapporter' og 'Udskriv Rapporter'.

Herved er der blevet udviklet en brugergrænseflade til systemet, således at brugeren har mulighed for at se en grafisk fremstilling af klassifikationsresultatet, hvilket var en nødvendighed for at de havde mulighed for at vurdere klassifikationen.

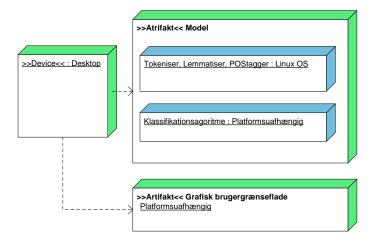
Kapitel 8

Integration of NLP-System

Efter de forskellige dele til systemet er blevet udviklet, skal disse integreres i et samlet NLP-system. Hvorledes dette er gjort, er beskrevet i dette kapitel, samtidigt med at det vurderes, hvor i rapporteringsforløbet ved Enhed for Patientsikkerhed systemet bør se sin anvendelse.

Allerede fra start, under konstruktion af use-casediagrammet beskrevet i kapitel 2, var det klart, at for at opfylde Enhed for Patientsikkerheds krav til systemet krævede det anvendelse af andre systemer i form aktørerne; NLP-værktøjer og Klassifikationsalgoritme. Formålet med dette kapitel er, at illustrere hvordan disse systemer er integreret til et samlet årsagsbestemmelsessystem.

I UML-notation, er et deploymentdiagram, et diagram, som viser systemets fysiske dele, hardware, og eksekverbart software. På deploymentdiagrammet på figur 8.1, anvendes noder, hvor hardware er illustreret ved teksten »Device« i form af en desktop og eksekverbar software med notationen »Artifakt« i form af en model og brugergrænseflade. Ligeledes skal der i deploymentdiagrammet tages højde for, hvilken platform den eksekverbare software skal køres på. [Hamilton & Miles 2006]



 $\textbf{Figur 8.1:} \ \, \textbf{Deployment diagrammet } \ \, \textbf{viser systemets hardware som } \, \textbf{\texttt{N}Device} \, \textbf{\texttt{@}}, \ \, \textbf{\texttt{og eksekverbar software som}} \, \textbf{\texttt{w}Artifakt} \, \textbf{\texttt{@}}.$

8.1 Integration af Model og Brugergrænseflade

Hardwaredelen består af en computer, hvor det for Enhed for Patientsikkerheds vedkommende vil være deres desktop, hvorpå de foretager årsagsanalyser til daglig. De eksekverbare miljøer er, delt op i en modeldel og grafisk brugergrænsefladedel.

8.1.1 Model

Modellen består af fire eksekverbare filer; hvoraf de tre af filerne er NLP-værktøjerne fra CST; tokeniser, lemmatiser og POS-tagger. Den sidste eksekverbare fil er klassifikationsalgoritmen. Af funktionaliteter under denne del ligger yderligere præprocessering, og featureudtræk, med idet de ikke er selvstændige eksekverbare filer, skal de ikke tages med på deploymentdiagrammet.

NLP-værktøjer: NLP-værktøjerne kan kun eksekveres på en computer med et installeret Linux operativsystem. Dette skyldes, at de er forkompilet på en Linux platform med de lingvistiske resurser fra CST. Dette stiller krav til de resterende eksekverbare miljøer, idet de også skal være i stand til at blive eksekveret på en Linux platform.

Inputtet til NLP-værktøjerne er UTH-rapporterne som er præprocesseret for at holde samme struktur som rapporterne har i den database, hvorfra de kommer.

Klassifikationsalgoritme: Softwaret Weka, der er anvendt til klassifikationsalgoritmen kan eksekveres med både Linux og Windows som operativsystem, og det er derfor installeret på Linux, for at kunne blive integreret med NLP-værktøjerne.

Inputtet til klassifikationsalgoritmen er featurevektorerne for hver UTH-rapport.

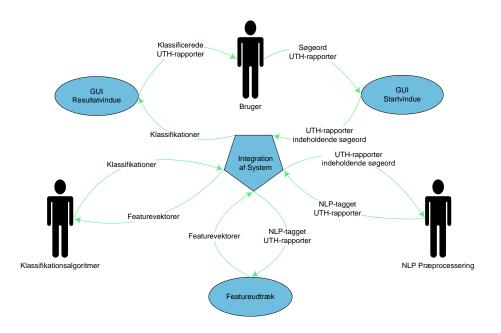
8.1.2 Brugergrænseflade

Grænsefladen til systemet er lavet grafisk for at opfylde kravene fra Enhed for Patientsikkerhed. Hvis systemet skulle eksekveres fra en kommandolinie med et output til for eksempel en fil, ville det ikke være muligt at opfylde dette krav. Brugergrænsefladen er ikke på nuværende tidspunkt lavet som en selvstændig eksekverbar fil, men det er muligt, og er medtaget på deploymentdiagrammet for at vise vigtigheden af adskillelsen imellem arkitetkturlagene. [Hamilton & Miles 2006]

8.2 Eksekvering af Systemet

For at det samlede system kan eksekveres, kæver det, at en Python-fortolker er installeret, da hovedparten af systemet er skrevet i programmeringssproget Python. Derudover kræver det interaktion fra brugerens side, hvorefter de enkelte processer bliver kaldt fra en overordnet fil. Dette princip er illustreret på figur 8.2. Hovedprocessen er vist ved en femkant, eftersom fem processer kaldes fra denne overordnede proces. Mændene illustrerer aktørerne identificeret i use-casediagrammet i kapitel 4. Det er valgt at inddrage dem på figuren, for at vise hvilken sammenhæng de har med det udviklede system.

Input fra brugeren består af et søgeord og de UTH-rapporter der skal søges i. Herefter findes de UTH-rapporter, der indeholder søgeordet og disse bliver brugt som input til NLP-præprocesseringen. Når NLP-præspocesseringen er færdig, er det muligt at udtrække features og efterfølgende klassificere UTH-rapporterne ved brug af klassifikationsalgoritmen. Når disse resultater er præsenteret for brugeren er årsagsbestemmelsen afsluttet.



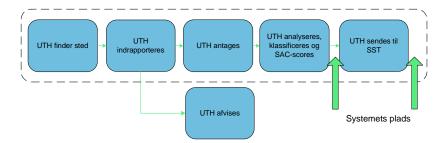
Figur 8.2: Integration af de enkelte dele, således at der opnås et samlet system er foretaget i en main-fil, hvorfra eksekveringen starter med input fra brugeren. Her illustreret ved et input-outputdiagram. Mændene repræsenterer de aktører, der blev identificeret i use-casediagrammet, imens ovalerne repræsenterer de enkelte processer.

8.3 Systemets Plads i Rapporteringsforløbet

Når årsagsbestemmelsen er integreret til et samlet system kan der ses på hvor i rapporteringsforløbet, det ses anvendt af Enhed for Patientsikkerhed. Det nuværende rapporteringsforløb blev beskrevet i kapitel 2 i form af hvad der skete med rapporten fra den blev indrapporteret til den endte ved Sundhedsstyrelsen. Det samlede system vil først komme i anvendelse, når Enhed fra Patientsikkerhed arbejder med flere UTH-rapporter samtidigt. Det er altså ikke ved behandlingen af den enkelte rapport, at systemet vil se sig anvendt, men derimod når analyse af den enkelte rapport er afsluttet. Det vil sige, at fra en UTH bliver indrapporteret til, at den er færdigbehandlet med henblik på analyser og klassifikation af SAC-scorer, vil rapporten ikke blive årsagsklassificeres ved brug af systemet. Dette er illustreret på figur 8.3, der viser de forskellige processer i rapporteringsforløbet. Der hvor systemet vil blive brugt er angivet ved de grønne pile, imens den stiplede linie indikerer, at dette er en proces, der vil forløbe mange gange. Efter UTH-rapporterne er blevet analyseret, klassificeret og SAC-scoret er der intet til hindre for hvornår systemet anvendes, idet det kan ønskes at foretage

årsagsklassifikationer på rapporter, der er indsendt til Sundhedsstyrelsen, såvel som rapporter som endnu ikke er blevet indsendt til Sundhedsstyrelsen.

Input til systemet bliver derved de rapporter, der ligger gemt i Region Hovedstadens database.



Figur 8.3: Figuren viser de processer, der er indeholdt i rapporteringsforløbet. Den stiplede boks angiver, at dette er processer, der forløber for mange processer, imens de grønne pile indikerer de steder, hvor det samlede system vil blive anvendt.

Outputtet fra systemet, det vil sige klassifikationerne af UTH-rapporterne i henhold til en specifik årsag, vil typisk se sig anvendt i forhold til udarbejdelse af regionale retningslinier indenfor det område søgeordet dækkede over. Disse retningslinier vil typisk blive udformet som temarapporter eller såkaldte 'pas-på-meddelelser', der distribueres ud til sundhedspersonalet i regionen, for dermed at forebygge lignende hændelser.

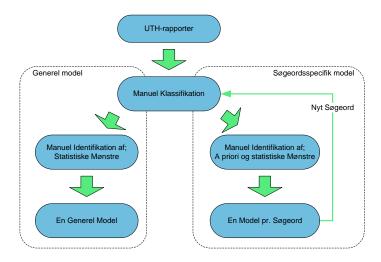
Herved er alle delene i systemet blevet integreret til et samlet system, der kan tilgås via en grafisk brugergrænseflade.

Kapitel 9

Generel Klassifikationsmodel

Indtil nu er der udviklet et system, der klassificerer hvorvidt et specifikt søgeord, 'EMP', er årsag til en utilsigtet hændelse eller ej. I dette kapitel vil det blive gennemgået, om det er muligt at udvide dette til en generel model for hvilket som helst søgeord.

Der kan tages udgangspunkt i to metoder, når en model skal opbygges. En metode, hvor en model er baseret på statistisk viden eller en metode, hvor modellen baseres på a priori viden. Figur 9.1 illustrerer forskellen på en generel statistisk model, hvor den generelle model kun kræver identifikation af mønstre én gang, og en model bygget på a priori viden, der kræver identifikation af mønstre, hver gang den skal benyttes til et nyt søgeord. I det følgende vil princippet bag mønstre for disse to typer af modeller blive gennemgået.



Figur 9.1: Forskellen mellem fremgangsmåden i en generel model, som anvender statistisk information og en model bygget i forhold til ét specifikt søgeord, der anvender a priori information.

9.1 A Priori Mønstre

Modellen med EPM som søgeord indeholder seks features, hvoraf de to af featurerne bygger på a priori viden fra eksperter. Disse features bestod af henholdsvis udvalgte ord og fraser. For bare ét enkelt søgeord, kræver det en stor arbejdsindsats, at få identificeret a priori viden

bestående af ord og fraser, der danner grundlag for deres beslutning om hvilken klasse UTH-rapporten tilhører. Der kan derved stilles spørgsmålstegn ved, om klassifikationen bidrager med nok og fremtidig vil blive benyttet af tilstrækkeligt mange personer sammenlignet med den arbejdsbyrde det er at udtrække viden for alle tænkelige søgeord.

Under antagelse af, at det ville kunne betale sig, at udvikle en model indeholdende flere søgeord, vil det kræve at eksperterne fra en brugergrænseflade ville kunne oprette et nyt søgeord, og under gennemlæsning af UTH-rapporterne markere hvilken klasse den pågældende rapport tilhørte og hvilke ord og fraser, der karakteriserede denne klasse. Ud fra dette ville det være muligt at bygge en model til hvert søgeord, og slutteligt præsentere nøjagtigheden for prædiktionerne af modellen. Modellen vil dog ikke blive tilsvarende den model, der er udviklet for EPM som søgeord, da eksperternes markering af karakteristiske ord og fraser har gennemgået en yderligere evaluering, idet de ikke er blevet anvendt som features direkte.

9.2 Statistiske Mønstre

Der er af projektgruppen i forbindelse med undersøgelse af data gjort flere tiltag for at finde generelle statistiske mønstre for henholdsvis den positive og negative klasse. Måden hvorpå årsagen til den utilsigtede hændelse fremkommer, kan inddeles i tre mønstre:

- 1. Årsagen er eksplicit udskrevet i teksten
- 2. Årsagen er eksplicit udskrevet i teksten men på varierende form
- 3. Årsagen er implicit i teksten

Hvis årsagen er eksplicit udskrevet i UTH-rapporten som '..EPM er årsag..' eller på varierende form som '..EMP gav anledning til..' vil det være muligt at klassificere dokumentet efter simple regler så som:

Hvis én af følgende fraser er til stede, så klassificer UTH-rapporten som positiv:

- [SØGEORD] [VERBUM]* [årsag|skyld|anledning|grund|som følge af] [PRÆP]*
- [hændelse] [VERBUM]* [årsag|skyld|anledning|grund|som følge af] [SØGEORD]

Hvor * er den gængse notation i et regulært udtryk for at ingen eller flere af det forrige element er til stede og | er notationen for det logiske udtryk ELLER. [VERBUM] kan være et hvilket som helst verbum, og [PRÆP] kan være et vilkårligt forholdsord. Disse grammatiske typer identificeres ved brug af POS-taggeren. Et eksempel på den første frase i reglen kunne være: 'EPM er årsag til', hvorimod et eksempel på den anden frase kunne være 'hændelsen var som følge af EPM'. Men eftersom årsagen ofte ikke er udtrykt eksplicit i de UTH-rapporter, som der er gennemarbejdet i dette projekt, har det ikke være muligt at klassificere efter denne form for regler.

Årsagen fremkommer derimod ofte implicit. Eksempler hvor søgeordet EPM er årsag, men at det kun fremgår implicit i teksten er;

"'Efterfølgende har det ikke været muligt at oprette pt i EPM under personens rigtige identitet."

I denne hændelsesbeskrivelse ses det for eksempel, at frasen 'ikke været muligt at oprette' henviser til, at der har været problemer med EPM, men ikke at det har været årsag til en utilsigtet hændelse. Et andet eksempel på implicit udtryk af årsag er givet ved denne sætning:

"Ændring af EPM kodning for granocyte (og neupogen for den sags skyld), så der under volumina står 2 stk OG 1 ml."

I denne UTH-rapport er der eksplicit udtrykt et ønske om en ændring af EPM, hvilket implicit indikerer, at der er problemer med EMP, men stadig ikke henviser til, at det er årsagen. Klassifikation med generelle statistiske mønstre, hvor årsagen er implicit udtrykt i teksten, vil kun være mulig hvis, der er en markant forskel på de rapporter, hvor søgeordet er årsag frem for de rapporter, hvor de ikke er årsag. Hvorvidt denne form for klassifikation er muligt, er undersøgt ved hjælp af algoritmen 'Bag of Words', se appendiks C, som ofte anvendes til dokumentklassifiaktion [Witten & Frank 2005, Weiss et al. 2005]. Denne algoritme gav dog ikke et tilfredsstillende resultat.

En anden statistisk tilgang er forsøgt, hvor alle tre på hinanden følgende ord er udskrevet, og de der fremkommer med en frekvens på over 4 er anvendt som statistiske features, hvilket heller ikke bidrager til adskillelse af de positive og negative klasser.

Ud fra ovenstående overvejelser må det konstateres, at det ikke er muligt at konstruere en generel model til klassifikation af UTH-rapporter uden inddragelse af a priori viden om hvert enkelt søgeord. Dette skyldes grundlæggende at samhørigheden indenfor hver klasse er for lav, mens koblingen klasserne imellem er for høj.

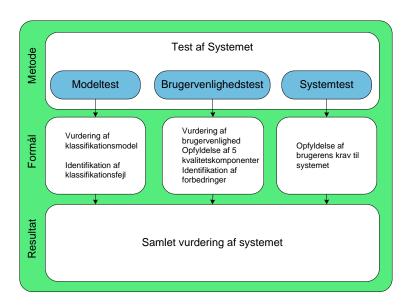
Ud fra ovenstående analyse er gjort overvejelser i henhold til, hvorvidt det er muligt at udvikle en generel model til årsagsklassifikation af utilsigtede hændelser. Efterfølgende vil klassifikationsmodellen med EPM som søgeord blive testes, således at modellens anvendelighed kan vurderes.

Kapitel 10

Test og Resultater

Efter integration af det samlede system er det muligt at teste systemet i henhold til både klassifikationsmodel, brugergrænseflade og krav til systemet. I kapitlet vil metoderne til test blive gennemgået såvel som resultaterne af test.

Test udføres med henblik på, at validere systemet og derved skabe et grundlag for hvilke dele af systemet, der fungerer godt og hvilke dele af systemet der skal forbedres. Det er valgt at udføre tre tests 1) Test af klassifikationsmodellen, 2) Systemtest, og 3) Brugervenlighedstest, se figur 10.1. Det er vigtigt at teste klassifikationsalgoritmen for sig selv, da det her er muligt at vurdere anvendeligheden af NLP i forhold til UTH-rapporterne. Test af det samlede system følger SPU-metoden og vil fokusere på hvorvidt brugerens krav til systemet er opfyldt. Disse krav er primært brugergrænsefladeorienterede, og derfor testes klassifikationsalgoritmen for sig. Til sidst udføres en brugervenlighedstest med henblik på at få vurderet systemets brugervenlighed, og ud fra disse resultater øge brugervenligheden af systemet.



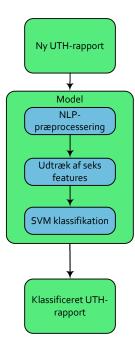
Figur 10.1: Oversigt over hvorledes systemet vil blive testet, og hvad formålet er med de enkelte tests.

10.1 Test af EPM Klassifikationsmodel

Formålet med at teste klassifikationsmodellen er, at få et mål for hvordan modellen klassificerer UTH-rapporter, der ikke er anvendt til træning. Til dette formål anvendes de tidligere randomiserede testrapporter, som består af en tredje del af de tilgængelige rapporter fra start. Dette svarer til henholdsvis 35 negative og 8 positive UTH-rapporter, der ikke har indgået i konstruktionen af modellen.

10.1.1 Metode

Figur 10.2 illustrerer hvordan UTH-rapporterne skal klassificeres, ved brug af klassifikationsmodellen, uden at der på forhånd er viden tilstede om hvilken én af klasserne den enkelte
UTH-rapport tilhører. For at kunne validere hvordan modellen udfører klassifikationerne, er
det midlertidigt nødvendigt at have viden om, hvilken klasse UTH-rapporten tilhører for
at kunne afgøre om modellen klassificerer korrekt. Denne klassifikation er manuelt udført
af eksperter. Under udførsel af testen, må der således ikke tages højde for hvilken klasse
UTH-rapporten tilhører.



Figur 10.2: Nye UTH-rapporter bliver klassificeret ved, at de seks features bliver udtrukket for UTH-rapporten og ved hjælp af SVM klassifikationsalgoritmen bliver tildelt til enten den positive klasse eller den negative klasse

10.1.2 Resultater

At måle en klassifikationsalgoritmes opnåede resultat er naturligt at gøre ved fejlrate, eftersom klassifikationsalgoritmen enten klassificerer en utilsigtet hændelse korrekt eller forkert.

Fejlraten er derfor proportionen af fejlklassificerede UTH-rapporter i forhold til det samlede antal UTH-rapporter. Fejlraten for de 43 testrapporterne er 9.3 % for de fire fejlklassificerede rapporter og sucessraten er 90.6% for 39 korrekt klassificerede UTH-rapporter. Det indikerer, at fremtidige UTH-rapporter vil blive klassificeret tæt på denne successrate, f. For en bruger af systemet ville det være fordelagtigt at vide hvor tæt, på denne successrate systemet vil klassificere.

Sandsynlighed for at den uklassificerede rapport ligger inden for et interval z, når der er givet en vis sikkerhed c, fx 90%, nøjagtighed, hvor z er en tabelværdi for normalfordelinger. Der et derfor under antagelse af, at antallet af positive og negative rapporter er normalfordelt. Ligning 10.1 angiver sandsynligheden Pr for at en stokastisk variabel X, med en middelværdi på 0 og som ligger indenfor konfidensområde på 2z.

$$Pr[-z \le X \le z] = c \tag{10.1}$$

Ved at indsætte værdier ind i ligning 10.1 fås ligning 10.2. Med 90% nøjagtighed kan det siges inden for hvilket udfald den stokastiske variabel, UTH-rapporten, kommer til at ligge, hvor 1.65 er fundet ved tabelopslag.

$$Pr[-1.65 \le \text{UTH-rapport} \le 1.65] = 90\%$$
 (10.2)

Yderligere afhænger nøjagtigheden af resultatet også af antallet at testrapporter N. Ud fra disse parametre, f, z, c, og N kan det beregnes indenfor hvilket konfidensområde udfaldet af klassifikationen forekommer, ved brug af ligning 10.3

$$p = \left(f + \frac{z^2}{2N} \pm z \cdot \sqrt{\frac{f}{N} - \frac{f^2}{N} + \frac{z^2}{4 \cdot N^2}}\right) / \left(1 + \frac{z^2}{N}\right) =$$
(10.3)

$$\left(0.9285 + \frac{1.65^2}{2 \cdot 42} \pm 1.65 \cdot \sqrt{\frac{0.9285}{42} - \frac{0.9285^2}{42} + \frac{1.65^2}{4 \cdot 42^2}}\right) / \left(1 + \frac{1.65^2}{42}\right) = [0.8336 : 0.9711]$$
(10.4)

Det betyder, at det med 90% nøjagtighed kan siges, at antallet af korrekt klassificerede UTH-rapporter kommer til at ligge imellem 83% og 97%. Denne information er relevant for brugeren af systemet, for at have en ide om, hvor mange klassifikationsfejl, der kan være tilstede

Figur 10.3 viser blandingsmatricen for testrapporterne, hvor antallet af rapporter i hver klasse ses i forhold til, hvordan de blev klassificeret. Det ses her, at alle de negative UTH-rapporter klassificeres korrekt, hvilket giver en falsk negativ (FN) rate på 0. Til sammenligning bliver 4 ud 8 af de positive UTH-rapporter bliver klassificeret korrekt, hvilket betyder, at de positive rapporter får en sand positiv (SP) rate på 0,5, se figur 10.3. Et plot af den sande positive rate mod den falske positive rate kaldes en *Receiver Operating Characteristic* (ROC) kurve. Formålet med ROC kurven er at få en grafisk repræsentation af den binære klassifikationsalgoritme. Den sande positive rate, som der plottes, er givet ved:

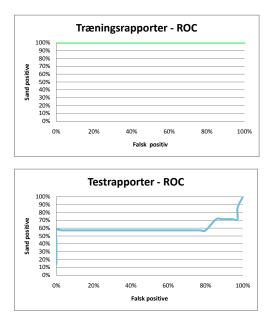
Sand Positive Rate =
$$\frac{\text{Sand Positive}}{\text{Positive}} = \frac{\text{Sand Positive}}{\text{Sand Positive} + \text{Falsk Negative}}$$
 (10.5)

I forhold til resultaterne vist på figur 10.3, vil det kun betyde et punkt i (1;0,57). Derfor udregnes værdierne til plottet under klassifikationen således, at det bliver udregnet en brøkdel af gangene svarende til antallet af testrapporter. [Witten & Frank 2005]

TEST					
SP Rate	FP Rate	Præsision	Recall	F-Mål	Klasse
0.5	0	1	0.5	0.667	Positiv
1	0.5	0.897	1	0.946	Negativ
TEST					
Positiv	Negativ		TI C		
4	× 4	Positiv	Faktisk		
X 0	√ 35	Negativ	<u>Š</u>		

Figur 10.3: Øverst ses blandingsmatricen, der viser at 4 ud af 8 positive rapporter klassificeres korrekt samt alle 35 negative rapporter klassificeres korrekt. Nederst ses resultaterne for klassifikationen, udtryk ved Sand positiv (SP), Falsk negativ (FP), Præcision, Recall og F-mål

Den bedst mulige klassifikationsalgoritme vil give en ROC-kurve med en vandret linie, som vil starte i (0;1) og stoppe (1;1). Det er et resultat, hvor 100% af de sande positive er klassificeret korrekt. Figur 10.4 viser et eksempel på ROC kurve for træningsrapporterne. Et helt uafhængigt og tilfældig gæt, uden brug af klassifikationsalgoritme, vil give et punkt på en diagonal linie, fra punktet (0;0) til (1;1), kaldet diskriminationslinien. For at have en brugbar klassifikation, kræver det således, at punkterne på ROC-kurven ligger over diskriminationslinien [Witten & Frank 2005]. Figur 10.4 viser et plot af både den ideelle, ROC-kurve i form af træningrapporter og den faktiske ROC-kurve i form af testrapporter. Ud fra dette ses, at der opnås en bedre ROC-kurve for træningssættet end testsættet.



Figur 10.4: Den øverste kurve er hvordan den ideelle ROC kurve ser ud, her taget ud fra træningsrapporter, hvor 100% er klassificeret korrekt. Den nederste kurve er ROC-kurven for testrapporterne.

Indenfor NLP er det ofte anvendt at evaluere en algoritme, ved at sammenligne den med et baselinesystem. I denne rapport er det valgt at sammenligne med den mest anvendt dokumentklassifikationsalgoritme, BOW, for UTH-testrapporterne. Se figur 10.5 for en sammenligning her imellem. Med et positive F-mål på 0 for BOW og et positiv F-mål på 0,667 for systemets klassifikationsalgoritme ses det, at systemets klassifikationsalgoritme præsterer langt bedre end en almindelig dokumentklassifikation, både i den negative og positive klasse.

Formålet med at evaluere i forhold til et baselinesystem er også at se om den mest basale form for algoritme er i stand til at klassificere nogle UTH-rapporter korrekt, som den mere avancerede algoritme ikke i stand til. Dette er ikke tilfælde her, idet systemets klassifikationsalgoritme klassificerer alle negative rapporter korrekt og de fire fejlklassificerede positive rapporter heller ikke klassificeres korrekt af baselinesystemet.



Figur 10.5: Resultaterne for baselinesystemet ses øverst og under ses testresultaterne

For at finde årsagen til de fejlklassificerede rapporter, undersøges featurevektorerne for de fire UTH-rapporter, der fejlklassificeres. I testrapporterne var det som tidligere omtalt 4, der blev klassificeret til negativ, selvom de var positive. Det handler om sagsnummer xxxxx11, xxxxx12, xxxxx13 og xxxxx14. Eksperterne har identificeret følgende ord og fraser i testrapporterne:

xxxxx11: Efterfølgende har det ikke været muligt at oprette pt i EPM under hans rigtige identitet.

xxxxx12: EPM systemet er nede og der er rod i "nødsystemet"

xxxxx13: Fordi med skal godkendes fra overgang fra amb til afd

xxxxx14: Disse to rettelser bevirker, at det på administrationssiden viser, at konakion skal gives 3 gange: 1 ml IM, 1 ml IV og 2 ml IV. På min ordinationsside ses kun en ordination, og de øvrige kan ikke slettes på administrationssiden.

Den første fejlklassificerede rapport med sagsnummer 'xxxxx11', er svær at klassificere ud fra den sætning, som eksperterne har markeret. Rapportøren har dog i feltet 'Formodede

årsager til hændelsen' skrevet '? Program problemer?', en frase som havde været oplagt at inkludere som a priori viden. Dette kan indikere, at der ikke har været tilstrækkeligt med positive træningsrapporter, for at få integreret den nødvendige viden.

Den næste fejlklassificerede rapport med sgsnummer 'xxxxx12', er af samme karakter som den ovenstående, hvor der i forvejen er en frase som lyder 'epm er nede', men på grund af at der i denne rapport er skrevet 'epm systemet er nede', er der ikke tilstrækkelige information til klassifikationsalgoritmen.

Den tredje fejlklassificerede rapport med sagsnummer 'xxxxxx13' er af en anden karakter. Ud fra den frase som eksperterne har markeret er det ikke muligt at konkludere, at det var EPM der er, har været årsag til hændelsen. Ved gennemlæsning af hele sagen er der stadig tvivl om hvilken klasse denne rapport tilhører. Det understreger, at det kan være vanskeligt for en algoritme at klassificere, når årsag ikke er eksplicit udtrykt i teksten.

Ligeledes er den fjerde og sidste fejlklassificerede rapport med sagsnummer 'xxxxx14' svær at identificere som en positiv rapport på baggrund af den markerede sætning fra eksperterne. Ordene 'administrationsside' og 'ordinationsside' tyder på, at hændelsen omhandler EPM.

Hermed er modellen blevet testet. Næste test er systemtesten, hvori klassifikationsalgoritmen er inkluderet.

10.2 Systemtest

Systemtesten er en test af det samlede system, hvilket vil sige både modellen samt brugergrænsefladen. Testen har til formål at verificere om brugerens krav til systemet er blevet opfyldt. Som input til systemet anvendes testsættet med UTH-rapporter.

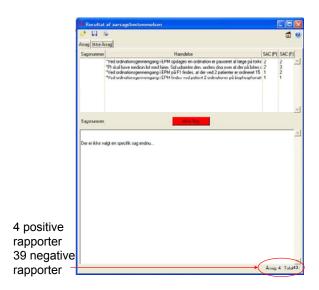
Der findes to typer af testteknikker: white box og black box. Til systemtesten bruges black box, der tager udgangspunkt i systemet funktionalitet frem for white box, der fokuserer på systemets programmel. Der drages således ikke brug af viden om hvorledes systemet er opbygget ved brug af black box teknikken. [Biering-Sørensen et al. 2002]

Testen foretages ved at teste hvert enkelt krav til systemet for sig, for til sidst at vise en samlet oversigt over hvorvidt kravene er opfyldt eller ej.

Krav 1: Systemet skal klassificere UTH-rapporter til én af to mulige klasser; enten hvor søgeord er årsag til hændelsen, eller hvor søgeordet blot indgår.

For at teste hvorvidt dette krav er opfyldt, anvendes filen med UTH-testrapporter. Denne fil indeholder 8 positive og 35 negative UTH-rapporter. Ved at specificere denne kildefil på brugergrænsefladen for 'Valg af Årsagsbestemmelse' med søgeordene 'EPM EMP', fås resultaterne vist på figur 10.6, der viser de positivt klassificerede UTH-rapporter. Heraf ses, at systemet klassificerer alle rapporter til én af de to mulige klasse, men det ses også at ikke alle rapporter bliver klassificeret korrekt, idet der samlet nu er 4 rapporter under fanebladet 'Årsag' ud af i alt 43 rapporter. Dermed må det konkluderes, at systemet er i stand til at klassificere alle rapporter, men at det ikke er alle der bliver klassificeret korrekt, hvorfor kravet kun er Delvist Opfyldt.

Krav 2: Systemet skal give brugeren mulighed for at angive en kildefil indeholde UTH-rapporter.



Figur 10.6: Brugergrænsefladen viser fanebladet for de positivt klassificerede rapporter. Nederst ses at de positive rapporter udgør 4 ud af 43 rapporter.

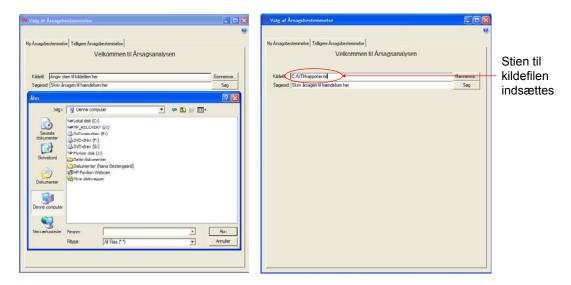
Dette krav er testes ved at angive kildefilen på de to måder, hvorpå det er muligt. 1) Ved at skrive stien 2) Ved at browse efter den korrekte placering. På figur 10.7 ses hvorledes det er muligt for brugeren at browse efter den korrekte placering, hvorefter stien placeres i det felt, hvor brugeren alternativt kunne indskrive stien direkte, hvilket ses i figur 10.8. Idet det er muligt at skrive direkte i tekstfeltet samt browse efter den korrekte placering, hvorved stien bliver angivet i tekstfeltet, konkluderes det, at dette krav er *Opfyldt*.

Krav 3: Systemets input skal være kompatibelt med databasen hvorfra UTH-rapportene skal hentes.

For at teste dette er det undersøgt hvad der sker, hvis der angives en fil, hvis format ikke er kompatibelt med databasen, hvorfra rapporterne hentes. I dette tilfælde bør systemet melde en fejl, for at gøre brugeren opmærksom på, at det ikke er et anerkendt filformat. Resultatet af denne test ses i figur 10.9. Herved ses at systemet melder en fejl til brugeren, hvorfor det konkluderes, at dette krav er *Opfyldt*.

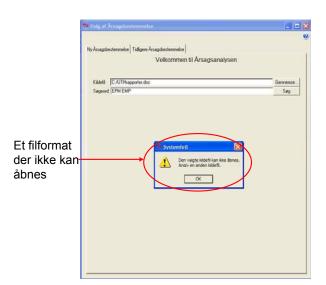
Krav 4: Systemet skal give brugeren mulighed for at angive et søgeord, der specificerer den årsag, der skal klassificeres efter med mulighed for flere stavekombinationer af årsagen.

Dette testes ved at foretage tre årsagsbestemmelser i filen med testrapporter. Én hvor 'EPM' er søgeord, én hvor 'EMP' er søgeord, og én hvor både 'EPM' og 'EMP' er søgeord. Herved ses, at årsagsbestemmelsen resulterer i 45 rapporter, når 'EPM' eller 'EPM EMP' er søgeord, men kun fire rapporter hvis 'EMP' er søgeord. Dette skyldes at alle rapporter i testfilen har medicinmodulet betegnet som EPM, dog har fire af disse rapporter også medicinmodulet benævnt som EMP. Dermed er kravet om mulighed for at angive et søgeord *Opfyldt*, idet der på brugergrænsefladen er implementeret et felt, hvori brugeren kan indskrive årsagen. Yderligere er kravet om mulighed for flere søgeord opfyldt.



Figur 10.7: Figuren viser dialogboksen, hvor brugeren har mulighed for at browse sig frem til kildefilens placering.

Figur 10.8: Figuren viser brugergrænsefladen, hvor stien til kildefilen er blevet angivet i feltet hertil.



Figur 10.9: Figuren viser at der fremkommer en advarselsboks, hvis den specificerede kildefils format ikke er kompatibelt med Region Hovedstadens database.

Krav 5: Systemet skal klassificere alle positive UTH-rapporter korrekt.

Dette krav er testet i det forrige afsnit, se kapitel 10 på side 59, i form af test af klassifikationsmodellen. Resultatet af denne test var, at ikke alle positive rapporter blev klassificeret korrekt, hvorfor det må konkluderes at dette krav er *Ikke Opfyldt*.

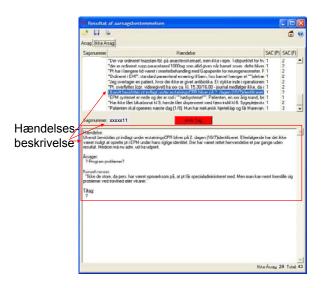
Krav 6: Systemet skal give mulighed for at se både de positive og de nega-

Krav	Opfyldt	Delvist Opfyldt	Ikke Opfyldt
Systemet skal klassificere UTH-rapporter til én af to mulige klasser; enten hvor søgeord er årsag til hændelsen, eller hvor søgeordet blot indgår.		√	
Systemet skal give brugeren mulighed for at angive en kildefil indeholde UTH-rapporter.	√		
Systemets input skal være kompatibelt med databasen hvorfra UTH-rapporterne skal hentes.	√		
Systemet skal give brugeren mulighed for at angive et søgeord der specificerer den årsag, der skal klassificeres efter med mulighed for flere stavekombinationer af årsagen.	√ 		
Systemet skal klassificere alle positive rapporteringer korrekt.			√
Systemet skal give mulighed for at se både de positive og negative klassificerede UTH-rapporter i en oversigt, samt den individuelle rappport.	√		
Systemet skal give brugeren mulighed for at korrigere klassifikationerne ved at flytte en rapport fra positiv til negativ klasse og omvendt såfremt en klassificering er forkert.	√ 		
Systemet skal kunne sortere klassifikationerne i henhold til SAC-score og sagsnummer.	√		
Systemet skal kunne gemme en klassifikation.			
Systemet skal kunne vise og udskrive de klassificerede UTH-rapporter enten på papir eller til en pdf fil.	√		

Tabel 10.1: Status over hvorvidt brugerens krav til systemet er opfyldt eller ej.

tive klassificerede UTH-rapporter i en oversigt, samt den individuelle rapport.

I kraft af at brugergrænsefladen er implementeret som den er, er dette krav *Opfyldt*, idet der øverst er en oversigt over alle rapporter i den pågældende klasse. Ved at klikke på en enkelt rapport vises hændelsesbeskrivelsen i tekstfelten nedenfor, hvilket fremgår af figur 10.10. Her



Figur 10.10: Figuren viser at det er muligt både at se rapporterne i en oversigt øverst, men ved klik på for eksempel sagsnummer 'xxxxx11', vises den tilhørende hændelsesbeskrivelse.

ses, at ved markering af sagsnummer 'xxxxx11', vises den tilhørende hændelsesbeskrivelse i tekstfeltet.

Krav 7: Systemet skal give brugeren mulighed for at korrigere klassifikationerne ved at flytte en rapport fra en positiv til en negativ klasse og omvendt såfremt en klassificering er forkert.

Dette krav testes ved at klikke på 'Afvis Sag'-knappen for rapporten med sagsnummeret 'xxxxx11', der er en fejlklassifikation. Dette er vist på figur 10.11. Efter klik på 'Afvis Sag'-knappen skal rapporten flyttes til fanebladet for positive klasse. Af figur 10.12 fremgår det, at rapporten er blevet flyttet, hvorfor det må konkluderes, at kravet er *Opfyldt*.

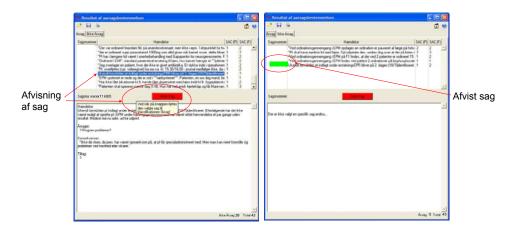
Krav 8: Systemet skal kunne sortere klassifikationerne i henhold til SAC-score og sagsnummer.

Dette krav er ligeledes testet på resultatet af testfilen. Testen er udført på både de negative og positive rapporter, hvoraf resultatet af testen på de negative klassifikationer. Rapporterne sorteres alt efter om 'Sagsnummer, 'SAC (P)' eller 'SAC (F)' er markeret, hvorfor det kan konkluderes, at dette krav er *Opfyldt*.

Krav 9: Systemet skal kunne gemme en klassifikation.

Dette krav er testes ved at flytte den fejlklassificerede rapport med sagsnummer 'xxxxx11' i testfilen og derefter gemme den nye klassifikation under et nyt navn. Når den fil åbnes skulle der gerne være fem klassifikationer i det positive faneblad, og 36 klassifikationer i det negative faneblad. Test af dette viser at den redigerede testfil gemmes som tiltænkt, hvorfor det konkluderes, at kravet er *Opfyldt*.

Krav 10: Systemet skal kunne vise og udskrive de klassificerede UTH-rapporter, enten på papir eller til en pdf fil.



Figur 10.11: På figuren er sagsnummer 'xxxxx11' markeret og er klar til afvisning, idet dette er en fejlklassifikation. Dette gøres ved at klikke på den røde knap.

Figur 10.12: Figuren viser at rapport med sagsnummer 'xxxxx11' er blevet afvist og derved flyttet til det modsatte faneblad, hvor den nu er markeret med grønt.

Det testes hvorvidt systemet giver mulighed for at vise udskrift i form af en pdf fil. Dette gøres ved at klikke på 'print'-knappen. Idet computeren hvorpå testen blev foretaget indeholder et pdf-genereringsprogram, at det muligt at udskrive til en pdf fil, hvorved et udskrift af rapporterne vises. Alternativt kan brugeren vælge at udskrive til en fysisk printer i stedet for pdf-genereringsprogrammet. Herved skulle dokumentet blive printet på den tilkoblede printer, hvilket er tilfældet, hvorfor det må konkluderes, at dette krav er *Opfyldt*. Det skal dog nævnes, at ved klik på print-knappen, bliver klassifikationerne på nuværende tidspunkt vist i et editorvindue, hvorefter brugeren endnu engang skal klikke på print, førend årsagsbestemmelsen printes.

En oversigt over status på de enkelte krav til systemet er samlet i tabel 10.1. Heraf ses at 9 ud af 12 krav er opfyldt, imens 2 er delvist opfyldt og ét enkelt krav ikke er opfyldt.

10.3 Brugervenlighedstest

Formålet med en test af brugervenligheden, er at få brugerens opfattelse af hvorledes et system fungerer. I denne sammenhæng er det især vigtigt, at se på hvor brugervenligt det pågældende system er, da dette er en vigtig forudsætning for en senere anvendelse af systemet.

Brugervenlighed måles ud fra antallet af katastrofer, der opstår, når en bruger anvender systemet. Jo færre katastrofer jo højere brugervenlighed. En katastrofe opstår, idet to brugere støder på det samme kritiske problem i en brugergrænseflade, hvor et kritisk problem er defineret ved [dialogdesign.dk 2007]:

- 1. Brugeren er ude af stand til at fortsætte sit arbejde med systemet uden at få hjælp fra en anden person, f.eks. en kollega, en superbruger eller en konsulent.
- 2. Brugeren føler, at systemets opførsel er stærkt irriterende eller irrationel.
- 3. Der er en kritisk forskel mellem det, som brugeren tror systemet gør, og det som systemet rent faktisk gør.

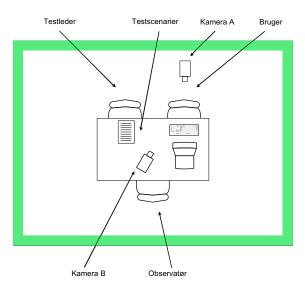
Kritiske problemer bør rettes omgående. Et udtryk for brugervenligheden er yderligere givet ved antallet af alvorlige problemer og kosmetiske problemer. Et alvorligt problem giver lejlighedsvis anledning til katastrofer, og bør ligeledes rettes, imens et kosmetisk problem blot bør rettes ved lejlighed. [Molich 1999]

10.3.1 Metode

Til afprøvning af brugervenligheden af nye systemer er metoderne Tænke højt afprøvning, Heuristisk inspektion og Konstruktiv interaktion særligt velegnet. Ved tænke højt testen får brugeren stillet en række realistiske opgaver, som skal løses ved brug af systemet. Imens brugeren løser opgaverne skal vedkommende tænke højt, det vil sige, sige hvad denne tænker på, hvad der tvivl om, hvad der forventes systemet vil gøre næst, hvordan denne fortolker en fejlmeddelelse og så videre. Ved heuristisk inspektion vurderes brugergrænsefladen af en række 'vurderingsmænd', der noterer hvilke fejl de måtte støde på undervejs. Konstruktiv interaktion er en variant af tænke højt afprøvningen. Her arbejder to brugere sammen om at løse opgaverne. [Molich 1999]

Det er i dette projekt valgt at foretage en tænke højt afprøvning, idet det prioriteres at afprøve systemet på reelle brugere frem for udenforstående vurderingsmænd, da anvendelse af systemet kræver viden om håndtering af UTH-rapporter. At det vælges at foretage en tænke højt test frem for heuristisk inspektion grunder i, at brugeren under normale omstændigheder vil anvende systemet alene, hvorfor det vurderes at dette også er den mest hensigtsmæssige måde at teste systemet på.

Der er en række fordele ved tænke højt testen, såfremt den bliver udført korrekt. Der opnås et billede af brugerens forventninger, adfærd og præferencer i konkrete situationer, der afspejler hvorledes systemet vil bliver anvendt i virkeligheden. Det, at brugeren forsøger mundtligt, at udtrykke sine tanker giver mulighed for at få problemer i designet afdækket, og samtidigt kan der opnås forståelse for, hvorfor de opstår. Ved brug af metoden er det ligeledes muligt at måle mere objektive kriterier såsom fejl og misforståelser i forbindelse med bestemte funktionaliteter i systemet. Metoden mister dog en del af sin brugbarhed, hvis brugeren ikke tænker højt undervejs. [kommunikationsforum.dk 2002] Det er i tænke højt testen valgt at afprøve systemet på to brugere, der begge er ansat ved Enhed for Patientsikkerhed, og derfor er vant til at arbejde med UTH-rapporter. Testen fandt sted i et mødelokale ved Enhed for Patientsikkerhed, og hver test tog ca. 15 minutter. Testen foregik på en bærbar computer. Med henblik på at kunne efterbehandle testen, blev det besluttet at filme situationen. Alle bevægelser på skærmen blev optaget ved hjælp af et softwareprogram samtidigt med, at brugeren blev filmet, hvilket er illustreret på figur 10.13. Herved er det efterfølgende muligt at genskabe, hvad der skete under testen ud fra lyd og billeder, hvilket er hensigtsmæssigt, idet det ellers kan være svært at huske alt der fandt sted under en test [Molich 1999]. Under den efterfølgende analyse af testen, er optagelsen af brugeren og skærmbilledet afspillet samtidigt, for derved at opfange brugerens ansigtsudtryk ved de forskellige episoder, ud fra hvilke bevægelser der var med musen på skærmen.



Figur 10.13: Figuren viser brugeropstillingen for testen, hvor kamera A optager musens bevægelser på brugerens skærm, og kamera B optager brugeren.

Testscenario

Det er i systemet vigtigt at have fokus på at teste de interaktionsfunktionaliteter, der er specielle for netop dette systems brugergrænseflade. De funktionaliteter, der er særlige for brugergrænsefladen er brugen af faner, afvisning af forkert klassificerede rapporter, sortering af rapporterne i forhold til enten sagsnummer eller SAC-score, samt visning af hændelsesforløb for hver enkelt rapport. For at teste disse funktionaliteter, er der udarbejdet et testscenario, som brugerne skal igennem. Det er vigtigt, at scenariet er så virkelighedstro som muligt, for at sikre at systemet bliver testet i henhold til den måde det er tiltænkt anvendt [Molich 1999].

Brugeren blev stillet overfor 12 opgaver:

1. Foretag en årsagsbestemmelse ud fra filen 'brugertest.txt' beliggende på 'skrivebordet'. Det ønskes at afgøre hvorvidt EPM/EMP er årsag til de utilsigtede hændelser eller ej.

Det testes om brugeren kan finde ud af at angive en gyldig kildesti samt anvende flere stavekombinationer af søgeordet.

2. Læs første hændelse, hvor EPM er årsag til den utilsigtede hændelse.

Det testes om brugeren forstår at klikke på en rapport for at få vist den tilhørende tekst. Ligeledes er det interessant at se, hvor brugeren klikker for at få vist hændelsesbeskrivelsen.

3. Hvor mange rapporter findes der, hvor EPM ikke er årsag til den utilsigtede hændelse?

Det undersøges om informationerne om antallet af rapporter er placeret hensigtsmæssigt. Ligeledes bør brugeren under denne opgave gøre brug af fanerne for at komme over til fanebladet med rapporter, hvor EPM ikke er årsag til hændelsen.

4. Hvor mange rapporter med EPM som årsag har en potentiel risikoscore på 3?

For at finde ud af dette kan brugeren enten vælge direkte at tælle antallet, eller at sortere rapporterne efter SAC P, for at få vist dem sammen.

5. Gem årsagsbestemmelsen på 'skrivebordet'.

Der testes om brugeren har opdaget funktionaliteten til at gemme klassificeringer.

6. Vend tilbage til startvinduet.

Der testes hvorvidt brugeren har opdaget 'hjem'-funktionaliteten.

7. Åbn årsagsbestemmelsen du gemte under den tidligere øvelse.

Tester at brugeren har lagt mærke til, at der er to faner i startvinduet, og om denne kan angive den korrekte sti.

8. I hvor mange rapporter er EPM årsag til hændelsen?

Det ønskes at gøre brugeren opmærksom på dette for, at denne senere kan se, at dette tal ændrer sig.

9. Er sagsnummer xxxxx15 klassificeret korrekt? Hvis ikke skal dette rettes.

For at løse denne opgave bør brugeren anvende faner, sortering og læsning af tekstfelt. Rapporten er i den negative klasse og er en fejlklassifikation, hvorfor rapporten skal afvises.

10. Har antallet af rapporter, hvor EPM er årsag til de utilsigtede hændelser, ændret sig?

Brugeren bør nu se, at antallet har ændret sig siden sidst denne kiggede på det.

11. Print årsagsbestemmelsen.

Tester om brugeren genkender print-funktionaliteten.

12. Luk printerprogrammet og programmet til årsagsbestemmelsen ned.

Tester om brugeren kan finde ud af at lukke programmet og ikke bare vende tilbage til startvinduet.

10.3.2 Resultater

For at give en oversigt over hvorledes testscenariet forløb, er resultatet af disse fremstillet i tabel 10.2. Begge brugere formåede at løse alle opgaver på nær én. Alle opgaverne er i tabellen forsynet med et symbol, der angiver om opgaven var let eller svær at løse. Løst uden problemer angives ved $\sqrt{\ }$, men løst med problemer angives ved ∇ . Opgaver, der kun kunne løses ved hjælp fra testlederen angives ved \div . En mere detaljeret beskrivelse af testscenariets forløb er at finde i appendiks E.

Som det ses af tabel 10.2 var der problemer med opgave 1, 4, 6, 9 og 11. Da brugerne blev bedt om at foretage en ny årsagsbestemmelse, blev begge brugere forvirret over at vindueskiftet imellem 'Valg af Årsagsbestemmelse' og 'Resultat af Årsagsbestemmelse' forårsagede et par sekunder, hvor intet vindue blev vist. Denne situation er ikke opstået tidligere under test af systemet, og skyldes det softwareprogram, der blev brugt til at optage brugerens bevægelser

Opgave	Bruger 1	Bruger 2
1: Foretag ny årsagsbestemmelse	∇	÷
2: Læs hændelsesbeskrivelse	\checkmark	
3: Antal ikke årsag rapporter	\checkmark	
4: Antal årsag rapporter med SAC P på 3	∇	∇
5: Gem rapporter	√	√
6: Vend tilbage til startvindue	÷	
7: Åbn tidligere årsagsbestemmelse	\checkmark	
8: Antal årsag rapporter	\checkmark	
9: Specifik sagsvurdering	∇	
10: Antal rapporter efter afvisning		
11: Print årsagsbestemmelse	∇	∇
12: Luk programmet		

Tabel 10.2: Oversigt over hvorledes brugerne løste opgaverne, hvor $\sqrt{}$ angiver, at den blev løst uden problemer, og ∇ angiver at der var problemer med løsning af opgaven, og \div at det krævede hjælp for at få gennemført opgaven.

på skærmbilledet. Dette testscenarie havde ikke været afprøvet inden brugertesten blev fortaget, hvilket ikke er hensigtsmæssigt. Yderligere ved Bruger 2, poppede terminalvinduet hvorfra systemet eksekveres op foran brugergrænsefladen, og testlederen måtte fjerne terminalvinduet igen. Denne fejl ses om kritisk, idet testlederen måtte bryde ind i testen. Da brugerne blev bedt om at finde antallet af UTH-rapporter, hvor den potentielle SAC-score var tre, havde begge brugere problemer. Disse bundede hovedsageligt i, at systemet kun kan sortere hændelserne efter stigende orden. Begge brugere havde ønsket en funktionalitet, der gjorde det muligt at sortere hændelserne i faldende orden også.

Ligeledes opstod der tvivl, da brugerne blev bedt om at vende tilbage til startvinduet, hvor Bruger 1 i stedet valgte at lukke programmet, mens Bruger 2 ikke vidste hvad betegnelsen startvindue betød. Bruger 1 fik hjælp af testlederen til at starte programmet igen. Bruger 2 gættede sig dog frem til det korrekte ikon, hvorefter denne vendte tilbage til startvinduet.

Da brugerne blev bedt om at korrigere en forkert klassificeret hændelse blev Bruger 1 i tvivl om denne skulle klikke på 'Afvis Sag' knappen, da denne forbandt dette med at sagen forsvinder helt, hvilket denne ikke ønskede. Denne misforståelse bunder i brug af forkert terminologi, idet brugeren forbinder det at afvise en sag med den proces, der finder sted når Enhed for Patientsikkerhed modtager en ny UTH-rapport, hvor de skal vurdere hvorvidt rapporten er en utilsigtet hændelse eller ej. Dette indikerer dermed, at der bør angives en anden tekst på knappen, idet brugeren ikke må være i tvivl om hvad de forskellige funktionaliteter indeholder.

Begge brugere fandt det ligeledes kompliceret at printe årsagsbestemmelsen. For det første opstod der forvirring om hvad årsagsbestemmelsen var, og for det andet var der tvivl om hvorvidt rapporterne var blevet printet, idet de blot blev vist i et særskilt vindue, da der

ikke var tilknyttet en printer til computeren under testen. Dette burde der have været taget højde for under udarbejdelsen af testscenarier.

Problemerne i opgave 1, 4, 6, 9 og 11 kan inddeles i henholdsvis kritiske problemer, alvorlige problemer og kosmetiske problemer, hvilket fremgår af tabel 10.3. Problematikken med at startvinduet forsvandt, og der gik et par sekunder inden resultatvinduet kom frem, vurderes som et kritisk problem, idet begge brugere havde problemer, men da denne situation ikke ville opstå under almindeligt brug vurderes problemet til kosmetisk. Problemet med sortering i forhold til den potentielle SAC-score vurderes som et kosmetisk problem, da det er et irritationsmoment for brugeren, hvor brugeren stadig kunne løse opgaven. Sortering begge veje ville være hensigtsmæssigt, men er ikke afgørende for løsning af opgaven. At begge brugere har problemer med at vende tilbage til startvinduet, vurderes som et kritisk problem, idet det i den ene brugers tilfælde betød, at denne blev afbrud i sin opgave, da programmet lukkede ned. Da begge brugere har problemer må det betegnes som en katastrofe, og bør straks udbedres. Scenariet med at afvise en rapport såfremt den er fejlklassificeret vurderes ligeledes til at være et kritisk problem, da systemet gjorde noget andet end hvad brugeren forventede. Endeligt er printfunktionen vurderet som et alvorligt problem, da brugeren stadig kunne fuldføre opgaven på trods af de givne omstændigheder.

Opgave	Kritisk Problem	Alvorligt Problem	Kosmetisk Problem
1:Foretage ny årsagsbestemmelse			X
4: Antal årsag rapporter med SAC P på 3			X
6: Vend tilbage til startvindue	X		
9: Specifik sagsvurdering	X		
11: Print årsagsbestemmelse		X	

Tabel 10.3: Oversigt over hvor alvorlige de fire problemer som brugerne stødte på under løsning af opgaverne.

Forbedringer til Systemet

Efter udførslen af testen blev systemet diskuteret i fællesskab. Ud af dette blev der listet en række punkter, der fungerede godt samt punkter, der kunne forbedres. Følgende punkter blev af brugerne nævnt til at fungere godt:

God hændelseslæser: Måden med opdeling af Resultat af Årsagsbestemmelsen i en oversigtsdel og en uddybende hændelsesbeskrivelsesdel fungerede ifølge brugeren rigtig godt, og de var ikke i tvivl om hvorledes disse skulle anvendes.

Værktøjstips: Begge brugere følte sig vejledt af værktøjstipsene, når de havde brug for hjælp.

Markering af afviste sager: At markere de hændelser som brugeren har korrigeret klassifikationen af med grønt, fandt brugerne godt, idet det gav dem et godt overblik over

hændelserne.

Opdeling af rapporter i printudskriftet: Brugerne syntes at udskriftsprintets layout var overskueligt, idet det var nemt at adskille de enkelte sager og deres SAC-scorer.

Af områder der kan forbedres er placeringen af 'Søg'-knappen på fanen for 'Ny Årsagsbestemmelse'. Bruger 1 fandt det uhensigtsmæssigt, at søg-knappen er placeret ved siden af feltet, hvor årsagen til den utilsigtede hændelse skal indtastes, da disse to funktionalitetsmæssigt ikke er forbundet med hinanden. Det er derfor valgt at placere søg-knappen nedenfor årsagsfeltet, efter samme princip som på fanebladet 'Tidligere Årsagsbestemmelse'.

Bruger 1 blev forvirret over, at der skulle angives et mellemrum imellem søgeordet, hvis dette kunne have flere forskellige stavemåder, da denne forbandt et mellemrum med et logisk AND, og der i søgning med flere søgeord egentlig ønskes et logisk OR. Det er valgt at bibeholde den oprindelige måde at søge på, da det vurderes at det blot vil blive mindre intuitivt at anvende systemet, hvis brugeren skulle skrive 'ELLER' imellem hvert eneste søgeord. Yderligere understøttes denne beslutning af, at Bruger 2 angav flere søgeord ved brug af mellemrum efter at have fået hjælp fra værktøjstippet. Der er dog efterfølgende blevet implementeret en hjælpetekst på begge faneblade, der vejleder brugeren, så denne ikke er afhængig af værktøjstips.

Bruger 2 undrede sig over, da denne ønskede at åbne en tidligere årsagsbestemmelse, skulle klikke på en Åbn-knap to gange. Første gang i Åbn-dialogboksen, og anden gang på selve brugergrænsefladen for fanebladet 'Tidligere Årsagsbestemmelse'. Det er valgt at placere en 'Åbn'-knap på brugergrænsefladen, da brugeren også skal have mulighed for direkte at indtaste stien til filen med den tidligere årsagsbestemmelse og derefter åbne filen, hvorfor denne knap ikke kan undgås.

Begge brugere mente, at titlen på brugergrænsefladerne burde ændres til 'Årsagsklassifikation' frem for 'Årsagsbestemmelse', da systemet ikke bestemmer årsagen, men blot klassificerer rapporter på baggrund af specificerede årsager, hvilket er blevet rettet.

Når brugeren skal åbne en tidligere årsagsbestemmelse, bliver vedkommende bedt om at specificere kildefilen. Her mente Bruger 1, at 'kildefil' var en misvisende betegnelse, idet der er forskel på den fil, der åbnes, når der foretages en ny årsagsbestemmelse, og når brugeren skal åbne en allerede gemt årsagsbestemmelse. Det er derfor valgt at ændre begrebet til 'Hændelser', idet brugeren mente, at det var mere passende.

Når brugergrænsefladen for resultatet af årsagsbestemmelsen vises, vil der i det nederste tekstfelt stå, at der endnu ikke er valgt en specifik sag. På dette tidspunkt er brugeren end ikke klar over, at dette er et tekstfelt, hvorfor Bruger 1 mener, at det er bedre at lade dette tekstfelt være tomt, idet det andet blot er en unødvendigt oplysning, der forvirrer mere, end den gavner. Det er derfor valgt at slette denne tekst.

Som tidligere nævnt var der problemer forbundet med 'Afvis Sag'-knappen. Denne var implementeret ved brug af rød baggrundsfarve for at indikere, at hvis en specifik rapport var klassificeret forkert kunne den afvises ved denne knap. Begge brugere opfatter dog dette som en meget voldsom farve at anvende, idet det signalerer fare og skaber et unødigt blikfang. Det er derfor valgt at fjerne baggrundsfarven. Ligeledes er teksten på knappen blevet ændret til henholdsvis 'Flyt Sag til Ikke Årsag' og 'Fly Sag til Årsag'. Bruger 1 forbandt det, at afvise en sag med den proces, der finder sted, når Enhed for Patientsikkerhed modtager en ny UTH-rapport, og de skal vurdere hvorvidt sagen er en utilsigtet hændelse eller ej. Da det

ikke er dette, der sker ved klik på knappen årsagsbestemmelsen, er teksten blevet ændret for at undgå tvetydige betydninger.

Begge brugere fandt det uhensigtsmæssigt, at det kun var muligt at sortere sagerne i stigende orden. Det er derfor valgt at implementere sortering efter faldende orden også.

Begge brugere var gentagne gange i tvivl om hvilken fane de befandt sig i, både på brugergrænsefladen for 'Valg af Årsagsbestemmelse' og 'Resultat af Årsagsbestemmelse'. For at undgå dette kan en mulig løsning være en tydeligere markering af det faneblad brugeren befandt sig i, for eksempel ved at give dette en anden farve.

Som tidligere nævnt havde begge brugere besvær med at vende tilbage til startvinduet. Bruger 2 fordi denne ikke viste hvad begrebet 'startvindue' dækkede over, og Bruger 1 fordi denne ikke var opmærksom på det tilhørende ikon. En løsning på dette kunne i følge Bruger 2 være større ikoner, samt en anden betegnelse for 'startvindue', hvilket kunne være 'startside'. Ligeledes bør 'hjem'-knappen placeres sammen med de andre funktionaliteter såsom åbn, gem og print, frem for at være placeret sammen med hjælp-knappen.

Bruger 1 fandt det uhensigtsmæssigt, at det samlede antal af rapporter og antallet af rapporter i det pågældende faneblad var placeret i bunden af brugergrænsefladen. I stedet ville det være at foretrække, hvis oplysningen var placeret sammen med det, det hang sammen med, nemlig oversigten, hvorfor det er valgt at flytte antallet derop.

I tabel 10.4 vises en oversigt over de ændringer brugerne havde til systemet, samt hvorvidt disse ændringer er blevet indført eller ej. Disse ændringer fører til et ændret udseende af de to brugergrænseflader. Brugergrænsefladerne for 'Valg af Årsagsklassifikation' ses på figur 10.14 og 10.15 og brugergrænsefladerne for 'Resultat af Årsagsklassifikation' ses på figur 10.16 og 10.17.

Udført
$\sqrt{}$
-
-
\checkmark
√
$\sqrt{}$
$\sqrt{}$
\checkmark
-

Tabel 10.4: Oversigt over hvilke forbedringer der er blevet lavet til systemet på baggrund af brugernes ønske, hvor $\sqrt{}$ angiver at forbedringen er udført, og - angiver at det ikke er udført.



Valg af Arsagsklassifikation

Velkommen til Arsagsklassifikationen

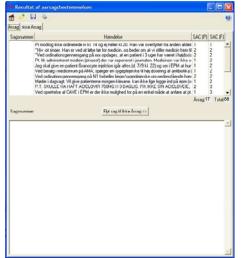
Hendelser (Argy sten til hendelsere her

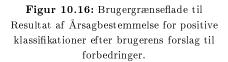
Argir tilen til den knagsklassifikation,

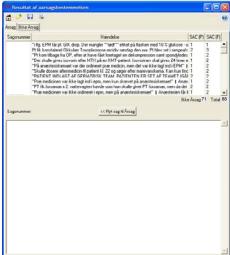
dv. eriskt er ålere

Figur 10.14: Brugergrænseflade til Ny Årsagsklassifikation efter brugerens forslag til forbedringer.

Figur 10.15: Brugergrænsefladen til en Tidligere Årsagsbestemmelse efter brugerens forslag til forbedringer.







Figur 10.17: Brugergrænsefladen til Resultat af Årsagsbestemmelse for negative klassifikationer efter brugerens forslag til forbedringer.

Vurdering af Systemets Brugergrænseflader

Ud fra ovenstående test af systemets brugergrænseflader er det muligt at vurdere hvorvidt designet af disse ud fra de fem kvalitetskomponenter, der blev opstillet i kapitel 7, Let at lære, Effektivt at bruge, Let at huske, Antallet af fejl og Tilfredshed.

Let at lære: Systemet opfylder kravet om, at det skal være let at lære. Under tænke højt testen formåede begge brugere at opnå en forståelse for hvad systemet kunne og hvilke funktionaliteter det indeholdt. Det er derved rimeligt at antage, at enhver bruger med de samme forhåndskundskaber som de to brugere, vil kunne lære at anvende systemet indenfor en kortere indlæringsperiode. Det vurderes, at én af årsagerne til at systemet er let at lære, er at strukturen for systemets brugergrænseflade er genkendelig fra for eksempel brugerens mailklientprogram.

Effektivt at bruge: Systemet effektivitet er svært at vurdere ud fra én enkelt tænke højt test. Effektiviteten af systemet bør sammenlignes med hvorledes brugeren vurdere UTH-rapporter uden anvendelse af systemet. Ved brug af systemet bliver brugeren præsenteret for langt færre UTH-rapporter end ellers, idet størstedelen af rapporterne frasorteres som negative hændelsesrapporter. Dog bliver nogle af de positive rapporter klassificeret forkert hvorfor effektiviteten bør vurderes ud fra, hvorvidt dette er acceptabelt.

Let at huske: Ifølge brugerne er funktionaliteterne i systemet lette at huske. Det er dog vanskeligt at vurdere, hvorvidt dette er tilfældet, idet brugere kun har haft mulighed for at anvende systemet en enkelt gang. Men ud fra, at det det gik forholdsvis let med at lære systemets funktionaliteter at kende første gang de brugte det, vurderes det at systemet er let at huske, hvilket yderligere understøttes af at det er opbygget efter et princip, de kender.

Antallet af fejl: Under test af systemet opstod der kun få deciderede fejl, hvor brugeren ikke var i stand til at fortsætte med anvendelsen af systemet. Dette var da brugeren skulle vende tilbage til startvinduet. Derudover var der en række områder, der kunne forbedres for at højne den generelle tilfredshed med systemet, men disse var ikke områder, der resulterede i fejl. Det må derved konkluderes at systemet er designet således, at risikoen for at lave fejl er minimeret.

Tilfredshed: Som nævnt ovenfor er der en række områder, der kan forbedres, hvilket vil have en betydelig effekt på den generelle tilfredshed med systemet. Brugerne fandt dog systemet anvendeligt, og opfattede det som en god hændelseslæser, hvorfor det må konkluderes at systemet var tilfredsstillende at anvende i forhold til den konkrete problemstilling.

De tre tests, der er udført på systemet, har alle bidraget til en forbedring af systemet. Såfremt der var flere UTH-rapporter til rådighed ville det ligeledes være muligt at forbedre klassifikationsalgoritmen yderligere.

Kapitel 11

Diskussion

Dette kapitel har til formål at diskutere de opnåede resultater i henhold til metoden, hvorpå et NLP-baseret klassifikationssystem er blevet konstrueret.

Rapportering af utilsigtede hændelser foregår i Region Hovedstaden ved at udfylde semistrukturerede indrapporteringsskemaer på intranettet. Fire af felterne i indrapporteringsskemaerne er fritekstfelter, mens de resterende seks felter er strukturerede, hvorfra det er muligt at udtrække information direkte. Hvis det skal være muligt, at genanvende den information, som bliver udfyldt i fritekstfelterne i rapporteringskemaerne er det nødvendigt at finde en metode, der øger tilgængeligheden af information. På nuværende tidspunkt genanvendes informationen, ved manuel gennemlæsning og klassifikation af rapporterne, hvilket fortages af en risikomanager tilknyttet den regionale enhed. Formålet med at udtrække informationen er at få lavet regionale retningslinier og temarapporter, der kan være med til at øge patientsikkerheden ved at hindre kommende lignende utilsigtede hændelser. Hvis det er muligt, at finde en automatisk metode, som ved sammenhængende forståelse af ord kan udtrække information, vil det lette arbejdsgangen for risikomanageren samtidigt med, at klassifikationen af rapporterne blive mere konsistent.

En metode til udviklingen af et automatisk system, som er i stand til at arbejde med fritekst, er NLP. I dette projekt er NLP afprøvet på UTH-rapporter med det formål at opnå information om hvorvidt et specifikt søgeord, EPM, er årsag til den utilsigtede hændelse eller ej. At afprøve metoden i forhold til et enkelt søgeord, skal kaste lys over om, det er muligt at udvide metoden til at omhandle flere søgeord.

UTH-Rapporter

Analysen af UTH-rapporterne viser, at UTH-rapporter er inkonsistente, idet de både indeholder stavefejl, indtastningsfejl og ufuldendte sætninger. Denne inkonsistens kan være årsag til, at de 88 træningsrapporter ikke er repræsentative for UTH-rapporter indeholdende søgeordet EPM. Det er forsøgt at ensrette rapporterne ved at rette stavefejl og forkortningsfejl, hvilket dog stadig ikke giver en stagnering forhold til antallet af unikke ord. Så på trods af korrektionen af UTH-rapporterne er 88 rapporter til træning stadig ikke nok til, at de er repræsentative for rapporter indeholdende søgeordet 'EPM'. Dette kan skyldes, at utilsigtede hændelser indrapporteres af alle faggrupper, og der er dermed meget stor forskel på hvilke ord, der anvendes. Ligeledes er det sjældent den samme person, der indrapporterer flere hændelser, hvorfor ordforrådet yderligere varierer. Derudover kan sprogbruget bære præg af manglende

indsigt i terminologien for software og computerbrug. For eksempel ses, at der anvendes anførselstegn eller blokbogstaver, hvis rapportøren er tvivl om, hvorvidt et pågældende ord er korrekt at anvende. Hvis alle sundhedspersoner, der indrapporterede hændelserne havde indsigt i terminologien om brug af software og computer, ville variationen i sproget blive reduceret og dermed mindre kompliceret arbejde med. Dette er dog ikke sandsynligt at antage, hvorfor problematikken må håndteres anderledes. Yderligere, er der ingen retningslinier for UTH-rapporter i forhold til, om der anvendes latinske eller danske medicinske betegnelser, hvilket dermed også øger mængden af unikke ord og dermed yderligere gør det vanskeligt at lave en statistisk repræsentation af UTH-rapporterne.

Måden hvorpå risikomanageren foretager manuelle klassifikationer af UTH-rapporter kan variere, hvorved der kan være forskel på, hvor mange rapporter én risikomanager vil klassificere som positive sammenlignet med en anden risikomanager. Dette skal dog ikke ses som et udtryk for manglende indsigt fra risikomanagerens side, men snarere et udtryk for UTH-rapporternes, til tider, meget tvetydige karakter. Ved en automatisering af processen, ville UTH-rapporterne blive klassificeret efter de samme parametre hver gang.

Anvendelse af NLP-værktøjer

De tre NLP-værktøjer, tokeniser, lemmatiser og POS-tagger, som blev anvendt til, at bidrage med grammatisk information til UTH-rapporterne, ligger med en nøjagtighed på 94-98% [cst.dk 2008]. På trods af en høj nøjagtighed vil der stadig være fejl i NLP-værktøjerne, som bliver propageret igennem hele klassifikationsmodellen, hvilket kan resultere i klassifikationsfejl. Dette ses for eksempel ved udtræk af de statistiske substantiver og verber, hvor adverbiet 'så', bliver POS-tagget til verbet 'så'. Dette bevirker, at ordet kommer til at fremkomme som et unikt verbum for den negative klasse af UTH-rapporter, hvor det i virkeligheden er et adverbium. En anden fejl, som skyldes NLP-værktøjerne, er hvor en ureglementeret forkortelse for et præparat 'AB' bliver lemmatiseret forkert. Da lemmatiseren ikke kender grundformen af 'AB', angiver den grundformen ved at tilføje et 'e' hvorved grundformen bliver til abe, hvilket heller ikke bør være en del af de unikke ord for de negative substantiver. Hvis de danske værktøjer derimod var trænet på medicinske tekster, ville der være en mulighed for at undgå denne form for fejl. At træne på medicinske tekster er en omfattende opgave. Hvis en *POS-tagger* skulle trænes på medicinske tekster ville det kræve et ordkorpus på minimum 250.000 taggede ord [cst.dk 2008]. En medicinsk træning af NLP-værktøjerne kan betyde øget præcision af klassifikationen.

Tilgængeligheden af danske NLP-værktøjer er yderligere begrænset af, hvor få mennesker, der anvender det danske sprog. Dette ses ved, at der findes medicinske engelsksprogede værktøjer, som er udviklet til at håndtere medicinske ord. Yderligere eksisterer der også engelsksprogede værktøjer, som er i stand til at genkende verbumsfraser. Det kan være en fordel at anvende en verbumsfrasegenkender i NLP-processeringen, idet verbet er et vigtigt element i sætningen, og ofte kan bidrag med yderligere forståelse af sætningen.

Resultat af Klassifikationen

Konstruktionen af klassifikationsmodellen er lavet ud fra udvælgelse af både statistiske og a priori features, på UTH-træningsrapporter. De statistiske features bygger på automatisk udtrækning af forskellene imellem den positive og negative klasse. A priori viden er fundet ud fra de ord og fraser, som eksperterne manuelt har identificeret i træningsrapporterne.

Problematikken, især blandt de statistiske features er, at UTH-rapporterne ikke er repræsentative for klasserne. Dette kan bevirke en overtilpasning af modellen, hvilket kan betyde at featureudvælgelse udvælger for få eller de forkerte features. For at undersøge om en overtilpasning har fundet sted, ville det have været brugbart med både et trænings-, test- og valideringssæt, da dette ville give mulighed for endnu en featureudvælgelse på testsættet. Herved kunne det kontrolleres, om det var de samme features, som blev udvalgt på testsættet.

Efter featureudvælgelsen er der blevet valgt en klassifikationsalgoritme. Valget stod imellem tre ofte anvendte algoritmer til tekstklassificering; Bayes klassifier, K-nærmeste Nabo og Support Vector Machine. Den klassifikationsalgoritme, som gav det bedste resultat på træningsrapporterne, ved en 10-foldskrydsvalidering, var SVM-algoritmen. En af fordelene ved SVM-algoritmen frem for de to andre klassifikationsalgoritmer er, at det er en lineær klassifikationsalgoritme, som er i stand til at håndtere ubalancerede klasser. At den er lineær er en fordel, idet tekst ofte fremkommer som lineært adskilleligt [Joachims 1998].

I forhold til søgeordet EPM er der langt flere UTH-rapport, hvor søgeordet blot indgår, men ikke er årsag, sammenlignet med antallet af rapporter, hvor EPM er årsag. Denne ubalance mellem klasserne betyder, at den klasse med færrest i ofte vil blive negligeret i en klassifikation frem for klassen med flest rapporter.

Ligeledes kunne der også med fordel have været anvendt et valideringssæt i denne situation for at undersøge om valget af klassifikationsalgoritme ville ændres i forhold til de rapporter, der ikke havde været anvendt til træning. Det er vurderet, at hvis de 131 UTH-rapporter skulle deles i både et trænings-, test- og valideringssæt ville der være for lidt data at træne på. Derfor vil det kræve flere UTH-rapporter med søgeordet EPM i, for at kunne anvende et valideringssæt.

Resultaterne fra den endelige klassifikationsmodel viser, at alle negative UTH-rapporter klassificeres korrekt, hvorimod halvdelen af de positive UTH-rapporter klassificeres korrekt. Halvdelen af de positive er ikke et tilfredsstillende resultat for Enhed for Patientsikkerhed, idet de havde et krav om, at alle positive UTH-rapporter skal klassificeres korrekt. Det er dog vigtigt at bemærke, at F-målet for den positive klasse ligger på 0,667, hvilket skyldes, at de alle de negative UTH-rapporter er korrekt klassificeret. F-målet for den negative klasse ligger på 0,946, som er et meget tilfredsstillende mål for en klassifikation. Havde det været formålet kun at klassificere de negative UTH-rapporter, ville dette være opfyldt.

Ved analyse af fejlklassifikationerne for den positive klasse ses en tendens til, at der kunne have været opnået et højere klassifikationsmål, såfremt klassifikationsmodellen havde været trænet på flere positive UTH-rapporter, således disse repræsenterede hele domænet. Her falder spørgsmålet igen tilbage på hvor mange rapporter, det vil kræve for at få et repræsentativt mål for en klasse, og om det er muligt grundet den store variation, der er i sprogbruget i rapporterne. Hvis det ikke er muligt kan det siges, at det ville være fordelagtigt, at arbejde med retningslinier for hvordan UTH-rapporterne skal udformes. Årsagen til, at der ikke

er nogle retningslinier for indrapporteringerne i dag, grunder i ønsket om at få så mange som muligt af hændelserne indrapporteret. Ved at stille krav til indrapporteringen af UTH-rapporterne kan der opnås mere struktur i fritekstfelterne. Dette skal dog gøres således, at rapporteringsprocessen ikke bliver mere kompleks. Måden hvorpå teksten er formuleret i dag lægger op til en beskrivende forklaring, hvorimod hvis spørgsmålene blev omformuleret til at være mere konkrete, er det tænkeligt at svarerne på spørgsmålene bliver mere konkrete og eksplicitte, hvorved et NLP-baseret system ville have lettere ved at identificere information. Det kunne for eksempel være ved at ændre teksten 'Hvorfor skete det?' i rapporteringsskemaet til 'Hvad var årsagen?' og tilføje en vejledningstekst i fritekstfeltet, der kunne guide brugeren i forhold til at få struktur på fritekstfeltet.

Det kan diskuteres, hvorvidt det er muligt at udvide den nuværende klassifikationsmodel til at kunne klassificere i henhold til ethvert søgeord. Test af modellen viser, at det ikke er muligt at konstruere en generel model blot ud fra statistiske features. Dette skyldes, at der er for få forskelle imellem henholdsvis de positive og negative rapporter. Så udarbejdelsen af en model vil inkluderer a priori viden om hvert enkelt søgeord. Dette resulterer i et omfattende arbejde pålagt eksperterne, og det bør vurderes om tiden til udvikling at en sådan model tilsvarer udbyttet af det endelige klassifikationssystem.

Resultat af Systemtest

Overordnet viste systemtesten, at 8 ud af i alt 10 krav var opfyldt. De krav der ikke blev opfyldt omhandler, at ikke alle positive rapporter bliver klassificeret korrekt. Det er derfor nødvendigt, at diskutere hvorvidt systemet er brugbart eller ej. Da systemet skal anvendes til sekundær brug af data, er klassifikationsfejl ikke direkte forbundet med reduceret patientsikkerhed. Sandsynligheden for, at risikomanagere overser positive rapporter når, de gennemlæser store mængder af UTH-rapporter, eksisterer, og det er derfor vigtigt, at vurdere om systemets reduktion i antallet af negative rapporter opvejer de positive rapporter, der ikke bliver klassificeret korrekt. Da brugeren altid vil kigge systemets klassifikationer igennem, kunne systemet designes således, at alle alvorlige hændelser, det vil sige hændelser med potentiel og faktuel SAC-score på 3 altid blev klassificeret positivt. Derved ville alvorlige hændelser altid blive endeligt verificeret af eksperter, der derved kunne vurdere om de var en positiv hændelse eller ej, samtidigt med at antallet af negative rapporter stadig blev reduceret kraftigt, hvorved brugeren ville spare tid.

Test af systemet viser, at brugergrænsefladen fungerer godt som en hændelseslæser, hvorfor denne prototype fremtidig kunne udvikles til at håndtere risikomanagerens gennemlæsning af rapporterne. På nuværende tidspunkt printes rapporterne ud på papir og gennemlæses. Fordelen ved at anvende brugergrænsefladen frem for papirudskrift, er muligheden for at have de positive og de negative UTH-rapporter adskilt. Yderligere kan klassifikationen gemmes og arbejdet kan forsættes på et andet tidspunkt.

Resultat af Brugervenlighedstest

Udførelsen af brugervenlighedstesten gav anledning til områder, som skulle forbedres. Det var blandt andet ændring af en knap, der gav anledning til forvirring om hvilken funktion-

alitet, der lå bag. Yderligere blev der tilføjet en funktion, der gjorde det muligt at sortere rapporterne i begge retninger. Når det er første gang systemet afprøves af rigtige brugere, ses der hurtigt en tendens til hvilke funktionaliteter, der har behov for at blive forbedret. Brugervenlighedstesten blev udført på to brugere, og der ville derfor ikke være behov for at teste på flere brugere, førend disse fejl blev rettet. Derefter kunne det være en fordel, at inddrage flere brugere til afprøvning af systemet efter udbedring af de groveste fejl. At udvikle et system er en iterativ proces, som kræver en høj involvering af brugeren, hvilket også har været tilfældet under udvikling af denne prototype.

Test af brugervenligheden blev udført af to brugere, der har været involveret i udformningen af systemets funktionalitet. Dette kan betyde, at resultatet af brugervenlighedstesten er farvet af, at brugerne på forhånd har haft et indblik i, hvad systemet kan. Dette understøtter blot yderligere at kommende test af systemet bør udføres på brugere, der ikke på forhånd har kendskab til systemet.

Herved er det blevet diskuteret hvilke udfordringer der i forbindelse med at designe et NLP-baseret system til klassifikation af utilsigtede hændelser og hvilke resultater, der er opnået med dette formål.

Kapitel 12

Konklusion

 $I\ dette\ afsnit\ konkluderes\ på\ rapportens\ problemstilling,\ der\ omhandlede\ konstruktion\ af\ et\ NLP-baseret\ system.$

Formålet med dette projekt er, at undersøge hvorvidt det er muligt at konstruere et NLP-baseret system til klassifikation af årsagen til utilsigtede hændelser. Rapporteringsskemaerne til utilsigtede hændelser er semistrukturerede, hvilke gør det svært at genanvende den information, der er tilgængelig i fritekstfelterne. Det er undersøgt, hvorvidt der er muligt som et 'proof-of-concept' at identificere, om EPM er årsag til den utilsigtede hændelse eller ej ved at opnå information om teksten i fritekstfelter ved brug af NLP-teknikker. En automatisk identifikation kan øge tilgængeligheden af information i UTH-rapporterne, hvorved risikomanagere ville spare tid i forbindelse med udarbejdelse af fremtidige retningslinier.

Modellen, udviklet til årsagsklassifikation, er baseret på såvel statistiske som a priori features. Til konstruktion af disse features er der anvendt NLP, med henblik på at opnå mere information om teksten i kraft af grammatik.

Resultatet af klassifikationsmodellen viser, at alle UTH-rapporter, hvor EPM blot indgår, men ikke er årsag til den utilsigtede hændelse, klassificeres korrekt. Derimod bliver kun halvdelen af de UTH-rapporter, hvor EPM er årsag til hændelsen klassificeret korrekt. Resultatet af klassifikationen skyldes primært kvaliteten af UTH-rapporterne. Rapporterne er generelt præget af inkonsistens i form af stavefejl, forkortningsfejl og ufuldendte sætninger, hvilket gør det vanskeligt at anvende NLP. Ved gennemgang af klassifikationsfejlene, ses det, at det er muligt at korrigere for halvdelen af fejlene. Fejl der dermed sandsynligvis kunne have været undgået, hvis der havde været flere UTH-rapporter til rådighed ved træning af modellen. Det kan derfor konkluderes, at der er potentiale for forbedring af modellen ved inddragelse af flere UTH-rapporter.

For at give risikomanageren mulighed for at vurdere klassifikationsmodellen er det valgt at implementere modellen i et samlet NLP-system med en dertilhørende brugergrænseflade. Ud fra test af det samlede system, ses det, at risikomanageren får mulighed for at arbejde med UTH-rapporterne på en ensrettet måde, idet der er mulighed for at flytte rapporterne fra den ene klasse til den anden samt gemme årsagsklassifikationerne, hvilket ellers ikke er muligt i dag. Ud fra dette kan det konkluderes, at hensigten med brugergrænsefladen er opfyldt, men for en fremtidig anvendelse, er det nødvendigt at udvide klassifikationsmodellen til at kunne klassificere andre søgeord end blot EPM. Konstruktion af en generel model, er vanskelig, idet det er nødvendigt at inddrage a priori viden i modellen. Såfremt denne a priori viden er tilgængelig, vil det være muligt at klassificere UTH-rapporterne i henhold til andre årsager.

Kapitel 13

Perspektivering

I dette kapitel vurderes anvendelsen af et NLP-baseret system til klassificering af årsagen til UTH i et fremtidigt perspektiv. Herunder hvordan systemet kan udvides og forbedres, således at anvendeligheden af det øges.

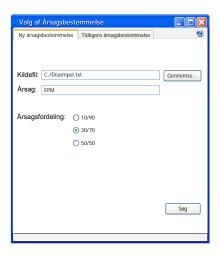
Det udviklede system til årsagsanalyse er på nuværende tidspunkt et proof-of-concept. det betyder, at der er en række områder, der bør forbedres eller udvides førend, der kan være tale om egentlig anvendelse af systemet i praksis.

En udvidelse til systemet kan være at udvide til flere søgeord end EPM. Det er tænkeligt, at fordelingen af positive og negative rapporter er anderledes end ved EPM, der er et eksempel på en ubalanceret fordeling. Det er i den forbindelse sandsynligt, at andre klassifikationsalgoritmer vil være at foretrække frem for SVM algoritmen, der primært udmærker sig ved ubalancerede klasser. Hvis brugeren på forhånd har en fornemmelse af fordelingen af positive og negative rapporter kunne systemet designes således, at brugeren inden en årsagsbestemmelse angav fordelingen i en procentsats, hvorefter systemet valgte den klassifikationsalgoritme, der formodentlig var bedst egnet. Dette kunne implementeres på brugergrænsefladen for 'Valg af Årsagsklassifikation' ved at angive forskellige fordelinger med tilhørende afkrydsningsbokse, hvor brugeren så kunne afkrydse den boks, der passede bedst, se figur 13.1. I den forbindelse kunne man også forestille sig, at brugeren vil foretage flere årsagsbestemmelser på samme kildefil og ud fra det sammenligne resultaterne.

Da der sjældent er en klassifikationsmodel, der klassificerer 100 % korrekt, kunne det være anvendeligt, hvis systemet blev udvidet til at lære af brugerens afvisninger af fejlklassificerede rapporter. Herved ville klassifikationsmodellen blive adaptiv, hvilket ville forbedre klassifikationen i forhold til hvis UTH-rapporterne ændrer karakter med tiden.

Set i lyset af, at det er problematisk at lave en klassifikation, kunne det måske være en fordel at udvide systemet fra at være en binær klassifikation til at indeholde tre klasser hvoraf de to af klasserne, stadig var den positive og den negative klasse, men hvor der var tilføjet en klasse, med usikre klassifikationer, som krævede at brugeren skulle tage stilling til hvilken klasse den pågældende UTH-rapport tilhørte.

Med henblik på udvidelser til systemets funktionalitet, kunne det være hensigtsmæssigt at give brugeren mulighed for at kommentere en klassifikation af de enkelte rapporter i brugergrænsefladen for 'Resultat af Årsagsbestemmelse'. Dermed har brugeren mulighed for at forklare, hvorfor en specifik klassifikation er blevet korrigeret, hvilket kan være anvendeligt i forhold til hvis en kollega har brug for at se nærmere på en årsagsklassifikation fortaget af en kollega.



Figur 13.1: På figuren ses et mockup af hvorledes brugergrænsefladen kunne se ud, hvis brugeren skulle have mulighed for at angive fordeling af positive og negative rapporter inden årsagsbestemmelsen.

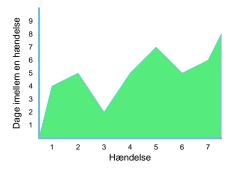
Der kunne ligeledes arbejdes med markering af specifikke ord eller fraser i tekstfeltet, der viser hændelsesbeskrivelsen. Disse ord og fraser kunne være søgeordet, såvel som de fraser eksperterne på forhånd har markeret som karakteristiske. Det er tænkeligt, at en sådan markering kunne give en hvis værdi til brugeren i forhold til, om denne skulle korrigere klassifikationen, da denne derved kunne få i indblik i systemets begrundelse for klassifikationen.

I databasen hvori Region Hovedstadens UTH-rapporter er ligger findes der flere oplysninger om den enkelte utilsigtede hændelse end blot sagsnummer, hændelsesbeskrivelse og SAC-scorer. Heriblandt er de vigtigste tidspunkt og dato for hændelsen samt sted for hændelsen. Dette er ligeledes informationer, der bør vises i de to oversigter over henholdsvis positive og negative rapporter på brugergrænsefladen for 'Resultat af årsagsbestemmelse'. Idet disse oplysninger fremkommer i strukturerede felter i UTH-rapporten vil det være muligt at udtrække dem direkte og dermed vise dem på brugegrænsefladen ligesom sagsnummer og SAC-score.

For at lette anvendelsen af systemet, bør der laves en direkte forbindelse til databasen således, at systemet automatisk udvælger de rapporteringer brugeren ønsker, uden at brugeren forinden er nødt til at eksportere dem til en separat fil. På brugergrænsefladen skal det således være muligt for brugeren at specificere, hvilke rapporter han ønsker klassificeret ud fra angivelse af en bestemt tidsperiode for indrapporteringerne.

Under forudsætning af at de ovenstående forbedringer og udvidelser tilføjes systemet, er det interessant at undersøge i hvilke situationer systemet ellers kunne anvendes. Som tidligere nævnt er det relevant at klassificere UTH-rapporter, med henblik på generering af temarapporter omhandlende specifikke årsager. Der er dog en anden mulighed, hvor det ligeledes ville være interessant at anvende systemet. Region Hovedstaden har længe haft et ønske om at behandle deres hændelsesrapporter kvantitativt, frem for blot de kvalitative gennemlæsninger. Til dette formål ville systemet kunne anvendes, idet det ligeledes kunne fungere som en form for overvågningssystem, hvor antallet af nye forekomne hændelser blev plottet i forhold til antallet af dage imellem hændelserne. Systemet skulle derved være udvidet til at håndtere andre søgeord end blot EPM. Hvis det var tilfældet, ville der kunne udpeges for eksempel 10

årsagstyper, der ønskes overvåget. Ud fra denne overvågning ville det dermed være muligt at konstruere en kurve over udviklingen efter princippet vist på figur 13.2.



Figur 13.2: På figuren ses princippet i en kurve for udviklingen i antallet af nyforekomne UTH-rapporter.

Herved havde Enhed for Patientsikkerhed mulighed for at opnå en fornemmelse af om der sker et fald i antallet af en specifik hændelse efter et konkret tiltag i forhold til at forhindre det. Tendensen for dette skulle være, at kurven efterhånden blev placeret højre oppe samtidigt med at den generelt ville stige.

Eftersom samarbejdet har været med Enhed for Patientsikkerhed, der udgør den regionale enhed for Region Hovedstaden. Dette betyder at systemet ikke umiddelbart vil kunne udvides til de resterende fire regioner, idet de anvender dansk patientsikkerhedsdatabase (DPSD) til indrapportering af utilsigtede hændelser. DPSD er et tiltag under Sundhedsstyrelsen, som i øjeblikket er under opgradering til en nyere version, hvor det er muligt at forestille sig at en fremtidig udgave af systemet ville kunne forbindes med det.

Litteratur

- [Biering-Sørensen et al. 2002] Biering-Sørensen, S., Hansen, F. O., Klim, S. & Madsen, P. T. (2002) Håndbog i Struktureret Program-Udvikling. Ingeniøren|Bøger. ISBN 87-571-1046-8.
- [Brennan et al. 1991] Brennan, T., Leape, L., Laird, N., Hebert, L., Localio, A. & Lawthers, A. (1991) Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study i. N Engl J Med, 6 (324).
- [Buckland & Gey 1994] Buckland, M. & Gey, F. (1994) The relationship between Recall and Precision. Journal of the American Society for Information Science, 45 (1).
- [Colas & Brazdil 2006] Colas, F. & Brazdil, P. (2006) On the Behavior of SVM and Some Older Algorithms in Binary Text Classification Tasks. Book Series: Lecture Notes in Computer Science, 4188, 45–52.
- [cst.dk 2008] (2008) CST's online demoer. http://cst.dk/online/. Københavns Universitet.
- [Dale et al. 2000] Dale, R., Moisl, H. & Somers, H. (2000) Handbook of Natural Language Processing. Marcel Dekker. ISBN 0-8247-9000-6.
- [dialogdesign.dk 2007] (2007) Dialog Design ved Rolf Molich. http://www.dialogdesign.dk.
- [dpsd.dk 2008] (2008) Dansk Patient Sikkerhedsdatabase. http://www.dpsd.dk/Publikationer %20mv/Nyhedsbreve/januar%202008.aspx.
- [Drucker et al. 1999] Drucker, H., Wu, D. & Vapnik, V. N. (1999) Support Vector Machines for Spam Categorization. *Transactions on Neural Networks*, 10 (5).
- [Duda et al. 2001] Duda, R. O., Hart, P. E. & Stork, D. G. (2001) Pattern Classification. Weiley Interscience. ISBN 0-471-05669-3.
- [Egebart & Bjørn 2008] Egebart, J. & Bjørn, B. Interview med Enhed for Patientsikkerhed, Region Hovedstaden.
- [Eriksson et al. 2004] Eriksson, H.-E., Penker, M., Lyons, B. & Fado, D. (2004) UML Toolkit 2. Whiley Publishing Inc. ISBN:0-471-46361-2.

LITTERATUR LITTERATUR

[Feldman & Sanger 2006] Feldman, R. & Sanger, J. (2006) The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press. ISBN 0-521-83657-3.

- [Hall 1999] (1999) Correlation-based Feature Subset Selection for Machine Learning. PhD dissertation, Department of Computer Science, University of Waikato. Hall, M. A.
- [Hamilton & Miles 2006] Hamilton, K. & Miles, R. (2006) Learning UML 2.0. O'Reilly. ISBN 0-596-00982-8.
- [Hansen 2005a] Finn Overgaard Hansen (2005a). Spu-uml note, systematisk programud-vikling. http://www.nyttf.dk/download/SPU-UML.pdf. 2005.
- [Hansen 2005b] Jørgen Hansen (2005b). Vejledning om rapportering af utilsigtede hændelser i sygehusvæsenet. 2005.
- [Hayward & Hofer 2001] Hayward, R. & Hofer, T. (2001) Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *JAMA*, 4 (286).
- [Joachims 1998] Joachims, T. (1998) Text Categorization with Support Vector Machines: Learning with Many Relevant Features. ECML-98 10th European Conference on Machine Learning, 10.
- [Johnson et al. 2008] Johnson, S. B., Bakken, S., Dine, D., Hyun, S., Mendonça, E., Morrison, F., Bright, T., Vleck, T. V., Wrenn, J. & Stetson, P. (2008) An Electronic Health Record Based on Structured Narrative. J Am Med Inform Assoc, 15.
- [kommunikationsforum.dk 2002] (2002) Tænke-højt-testen forklaret for dummies. http://www.kommunikationsforum.dk/artikler.asp?articleid=5381.
- [Liddy 2001] Liddy, E. (2001) Natural Language Processing, In Encyclopedia of Library and Information Science, 2nd Ed. Marcel Decker, Inc.
- [Lipczak & Schiøler 2001] Lipczak, H. & Schiøler, T. (2001) Rapportering af utilsigtede hændelser. *Ugeskrift for Læger*, 39.
- [Marcus et al. 1993] Marcus, M. P., Marcinkiewicz, M. A. & Santorini, B. (1993) Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics, 19 (2).
- [Matthews 2006] Matthews, M. (2006) Improving Biomedical Text Categorisation with NLP. Proceedings of the SIGs, The Joint BioLINK-Bio-Ontologies Meeting, Aug (1).
- [Molich 1999] Molich, R. (1999) Brugervenlige edb-systemer. Teknisk Forlag. ISBN 87-571-1647-4.
- [Molich 2003] Rolf Molich (2003). Gui standard retningslinjer for hvordan brugergrænse-flader skal se ud. Hovedstadens Sygehusfællesskab. 2003.
- [Mortensen & Dunker 2004] Mortensen, H. & Dunker, L. Amternes organisering af patientsikkerhedsarbejdet. Amtsrådsforeningen.

LITTERATUR LITTERATUR

[Nielsen 1998] Nielsen, J. (1998) Useablity Engineering. Academic Press. ISBN 0-12-518406-9.

- [patientsikkerhed.dk 2008a] (2008a) Patientsikkerhed Fakta om Utilsigtede Hændelser. http://www.patientsikkerhed.dk/Fakta om utilsigtede haendelser.
- [patientsikkerhed.dk 2008b] (2008b) Patientsikkerhed Begreber. http://www.patientsikkerhed.dk/dk/patientsikkerhed/begreber/.
- [ranks.nl 2008] (2008) Danish Stopwords. http://www.ranks.nl/stopwords/danish.html.
- [Region Hovedstaden 2008] Region Hovedstaden, E. f. P. (2008) Patientsikkerhed i Region Hovedstaden Handlingsplan 2008-2010. Enhed for Patientsikkerhed. ISBN 87-92195-04-0.
- [retsinformation.dk 2007] (2007) Bekendtgørelse om rapportering af utilsigtede hændelser i sygehusvæsenet. https://www.retsinformation.dk/Forms/R0710.aspx?id=11376.
- [Safran et al. 2007] Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C. & Detmer, D. E. (2007) Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association, 14 (1).
- [Savova et al. 2007] Savova, G. K., Ogren, P. V., Duffy, P. H., Buntrock, J. D. & Chute, C. G. (2007) Mayo Clinic NLP System for Patient Smoking Status Identification. JAMIA, 10.
- [Schiøler 2001] Schiøler, T. (2001) Patientsikkerhed Begivenhedsrelaterede begreber. Ugeskrift for Læger, 39.
- [Schiøler et al. 2001] Schiøler, T., Lipczak, H. & Lilj, B. (2001) Utilsigtede hændelser på danske sygehuse. Sygeplejersken, 38.
- [sikkerpatient.dk 2008] (2008) Viden om Patientsikkerhed. http://www.sikkerpatient.dk/Viden-om.aspx.
- [Somers 2004] Somers, H. (2004) Emperical Approaches to Natural Language Processing. Marcel Dekker inc. ISBN 0-8742-9000-6.
- [Sundhedsstyrelsen 2007] Sundhedsstyrelsen (2007) Årsrapport 2006 DPSD. Sundhedsstyrelsen. ISBN 87-7676-460-5.
- [Sygehusfællesskab 2005] Sygehusfællesskab, H. (2005) Patientsikkerhed i H:S Handlingsplan II 2005-2008. Hovedstadens Sygehusfællesskab. ISBN 87-7858-248-2.
- [useit.com 2008] (2008) Jakob Nielsen on Useablity and Web design. http://www.useit.com/papers/sun/icons.html.
- [Vapnik 1995] Vapnik, V. (1995) The Nature of Statistical Learning Theory. Springer- Verlag. ISBN 0387987800.
- [Vilalta 2008] Vilalta, R. Scaling up the accuracy of naïve-bayes classifiers: a decision-tree hybrid. Advanced Artifical Intelligence.

LITTERATUR LITTERATUR

[Vingtoft & Pedersen 2007] Vingtoft, S. & Pedersen, S. S. (2007) Analyse og krav til itunderstøttelse af rapportering af utilsigtede hændelser. Mediq.

- [Weiss et al. 2005] Weiss, S. M., Indurkhya, N., Zhang, T. & Damerau, F. J. (2005) Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer. ISBN 0-387-95433-3.
- [Winston 1970] Winston, R. (1970) Managing the Development of Large Software Systems. Proceedings of IEEE WESCON 26, August (9).
- [Witten & Frank 2005] Witten, I. H. & Frank, E. (2005) Data Mining: Practical machine learning tools and techniques. Elsevier. ISBN 0-12-088407-0.

Appendix A

Risikovurdering af Utilsigtede Hændelser

Dette kapitel gennemgår hvorledes utilsigtede hændelser risikovurderes i henhold til hændelsens alvorlighed og hyppighed.

Utilsigtede hændelser (UTH) kan inddeles i ni forskellige kategorier. Disse er vist i tabel A.1 sammen med de dertilhørende hyppighedsprocenter for UTH indrapporteret i 2007. [dpsd.dk 2008]

Utilsigtet Hændelse	Hyppighed (%)
Hændelser i forbindelse med medicinering	34
Hændelser i forbindelse med operative eller invasive indgreb	5
Fald	14
Selvmord og selvmordsforsøg	1
Hændelser i forbindelse med anæstesiprocedurer	2
Forveksling eller fejlkommunikation	14
Hændelser i forbindelse med kontinuitetsbrud	10
Hjertestop eller uventet dødsfald	1
Øvrige hændelser	19

Tabel A.1: Inddelingen af UTH og deres hyppighed for rapporteringer fra 2007. [dpsd.dk 2008]

Risikoen forbundet med UTH vurderes ud fra det internationale Safety Assessment Code (SAC) system. Risikoscoren udregnes ud fra to aspekter: Alvorlighedsgraden (skadens omfang) og hændelsens hyppighed (sandsynligheden for gentagelse). [Sundhedsstyrelsen 2007]

Alvorlighedsgraden

Alvorlighedsgraden er inddelt i fire kategorier: Katastrofal, betydende, moderat og minimal. Betydningen af disse kategorier vil blive beskrevet i det følgende.

Katastrofal: En af følgende:

- Død
- $\bullet\,$ Varigt betydende funktionsstab/handicap, hvor mén-graden er $\geq 15\%$ for én patient

Betydende: En eller flere af følgende:

- Varige funktionsstab/handicap, hvor mén-graden ≤ 15% for minimum én patient
- Betydelig øget udrednings/behandlingsintensitet, herunder overflyttelse til intensiv observation/dialyse/koronarafsnit for en patient
- Ved flere patienter skadet af samme begivenhed: Lettere øget udredning/behandlingsintensitet eller forlænget indlæggelsestid for minimum 3 patienter.

Moderat: En eller flere af følgende:

- Lettere øget udredningsarbejde/behandlingsintensitet som blev håndteret på en stamafdeling for én patient
- Ved en til to patienter skadet af samme begivenhed: Forlænget indlæggelsesvarighed.

Minimal: Ingen eller minimal betydning for patienten

Hændelsens Hyppighed

Hyppigheden inddeles ligeledes i 4 kategorier: Hyppig, mindre hyppig, sjælden og meget sjælden.

Hyppig: Vurderes at forekomme op til flere gange indenfor et år på det pågældende hospital.

Mindre hyppig: Vurderes at forekomme igen inden for et til to år på det pågældende hospital.

Sjælden: Vurderes at forekomme igen inden for to til fem år på det pågældende hospital.

Meget sjælden: Vurderes at forekomme igen inden for fem til 30 år på det pågældende hospital.

SAC Matrix

Ud fra SAC matricen, der er vist i tabel A.2 udregnes en risikoscore. Risikoscoren kan antage værdierne 1, 2 og 3. [Sundhedsstyrelsen 2007]

	Alvorlighedsgrad			
Hyppighed	Katastrofal	Betydende	Moderat	Minimal
Hyppig	3	3	2	1
Mindre hyppig	3	2	1	1
Sjælden	3	2	1	1
Meget sjælden	3	2	1	1

Tabel A.2: SAC matrice til risikovurdering af UTH, hvor 3 indikerer høj risiko og 1 lav risiko. [Sundhedsstyrelsen 2007]

En faktuel score på 3 vil udløse en kerneårsagsanalyse. Dog kan sådan analyse fravælges, hvis en lignende hændelse for nylig er blevet kerneårsagsanalyseret, og hvor det vil have større værdi at understøtte implementeringen af handlingsplaner. En kerneårsagsanalyse udmønter sig i systematisk at skrælle lag på lag af en hændelse for herved at nå ind til den inderste kerne/de egentlige årsager til, at patienten blev påført en skade. En potentiel score på 3, samt faktuel og potentiel score på 2 vil bevirke en aggregeret kerneårsagsanalyse, der er en analyse af flere hændelser af samme karakter, rapporteret over en given periode. Endeligt vil en faktuel score på 1 blot blive håndteret lokalt, hvor hændelsen er sket. Hændelserne kan også indgå i aggregerede kerneårsagsanalyser. [Region Hovedstaden 2008]

3% af de indsendte hændelser i 2006 blev rapporteret som 'meget alvorlige', hvilket vil sige at de havde en faktuel risikoscore på 3. 22% af hændelserne havde en faktuel risikoscore på 2, mens de resterende 75% af hændelserne havde en faktuel score på 1. [Sundhedsstyrelsen 2007].

Appendix B

Natural Language Processing

Dette appendiks støtter op om teorien der anvendes indenfor NLP. Hvilke sprogniveauer der anvendes og hvilke grammatiske forkortelser der er brugt under udviklingen af systemet.

Natural Language Processing (NLP) er et forsøg på at få computere til at analysere tekst ved hjælp af teorier og teknologier. Der findes forskellige definitioner af hvad NLP er. E.D. Liddy [Liddy 2001] definere det således:

NLP er en række af teoretiske 'beregningsteknikker' til at analysere og repræsentere fritekst på et eller flere niveauer af sproglig analyse. Dette med det formål at opnå menneskelignende sprogprocessering til en række opgaver og applikationer.

Ud fra denne definition ses, at NLP er tiltænkt til at indgå i sammenhæng med andre opgaver og applikationer. Det giver ikke mening at tale om at NLP som enkeltstående, det bliver nødt til at være i forhold til noget som for eksempel en tekst eller tale. På nuværende tidspunkt er det generelle mål for NLP, i for hold til tekst, at opfylde følgende tre opgaver;

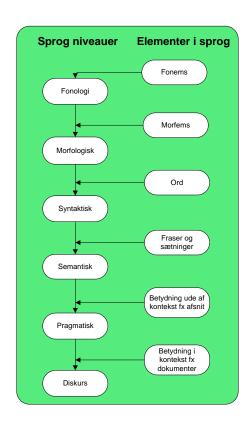
- 1. Omformulere en tekst
- 2. Oversætte en tekst til et andet sprog
- 3. Besvare spørgsmål om konteksten af en tekst

NLP kan kun til dels opfylde disse tre punkter, fordi der ikke kan opnås 100% nøjagtighed indenfor opgaverne, alt afhængigt af teksten sværhedsgrad. I forhold til medicinske tekster, bliver denne opgave endnu mere kompleks på grund af at medicinske tekster ofte indeholder både latinsk ord, danske ord og navne på præparater. For at lave værktøjer, der er optimeret til medicinske tekster ville det kæve et medicinsk ordkorpus, til træning af værktøjerne. [Liddy 2001]

Sprogniveauer

For at forklare hvordan NLP anvendes på tekster, er det hensigtsmæssigt at tage udgangspunkt i hvilke sprogniveauer, der eksisterer. Figur B.1 illustrere de seks niveauer

Fonotik og Fonologi: Fonetik er studier og klassifikation af sproglyde, det vil sige hvilke lyde, der anvendes i sprog, hvilke egenskaber disse lyde har og hvordan de frembringes.



Figur B.1: De firkantede bokse til venstre illustrerer sprogniveauet og de til højre et konkret eksempel i forhold til tekst. [Dale $et\ al.\ 2000$]-modificeret

Fonologi er kontrastforholdene imellem de fundamentale sproglyde både indenfor og på tværs af ord. [Liddy 2001] For eksempel udtales 'd' i drengen og sand ikke ens.

Morfologi: Omhandler analyse af ordenes interne struktur. Morfologien for et ord er ordets struktur eller form. Den mindste meningsfulde sprogenhed i et ord kaldes et morfem. Et eksempel på et ords morfologi kan være ordet; 'præprocessering', der kan inddeles i tre morfemer; præfikset 'præ', roden 'proces', og suffikset 'ring'. Betydningen af morfemerne er den sammen på tværs af ord, hvilket betyder, at mennesker er i stand til at opdele ord i morfemer og ved sammensætning af disse forstå nye ord og deres betydning. Denne evne til at forstå morfemer gøres brug af i simple grammatiske regler. For eksempel indikerer verber, som ender på 'ed', at det omhandler en datidshandling. [Liddy 2001]

Syntaks: Viden om strukturerne imellem sætninger og ord. For at opnå syntaktisk viden kræves både grammar og en parser. En grammar er en væsentlig måde at specificere et uendeligt regelret sprog med et endeligt input. En grammar består af et regelsæt, som kaldes rewrite rules. Reglerne bruges til at udlede en streng fra en anden ved substring replacement. En parser er et program for beslutningsprocessen og en inputstreng af symboler er en sætning på et sprog.

Syntaks er en vigtig del af sproget, fordi rækkefølgen af ordene bidrager til forståelse af sætningen. For eksempel de to sætninger: 'Manden spiser koen.' og 'Koen spiser manden.' adskiller sig kun i rækkefølgen af ordene, men har vidt forskellig betydning. [Liddy 2001] Syntaks inkluderer, at hvert enkelt ord tilskrives en 'part-of-speach' (POS) tag for hvert ord. *POS-taggen* for en af de to store ordklasser, substantiver, angiver således, at ordet er et substantiv, bestemt eller ubestemt og hvilket grammatisk tal ordet er til stede i. I den anden store ordklasse, verber, angives, at det er et verbum, og hvilken tid dette det er til stede i. Den samlede oversigt over de 50 *POS-tags*, der anvendes på dansk, kan ses i tabel B.1 og tabel B.2. [cst.dk 2008]

Semantik: Viden om ord og sætningers betydning. Semantikken bestemmes ud fra den mulige betydning af en sætning. Dette gøres ved at fokusere på interaktionen imellem ordniveauers betydning i en sætning. Når et ord har flere en betydning kræves det, at det bliver gjort entydigt. Semantisk entydighed tillader kun en betydning af ord. For eksempel navneordet 'bakke' kan betyde enten en bakke til at bærere ting på, eller det kan betyde en bakke i et landskab. Det vil altså kræve information fra resten af sætningen for at finde ud af, hvad betydningen af ordet bakke er. [Liddy 2001]

Pragmatik: Pragmatik omhandler sprogbrug og i hvilken kontekst det bruges. Målet er at gøre det klar, hvordan ekstra betydning læses ind i en tekst uden, at det står der eksplicit. [Liddy 2001] Dette kan for eksempel være om en sætning er ment som en ordre eller en konstatering.

Diskurs: Viden om lingvistiske enheder større end en enkelt sætning. Hvor syntaks og semantik arbejder med sætningsenheder, arbejder det sidste niveau i sprogbrug med tekstenheder længere end en sætning. [Liddy 2001]

Tag	Betydning	Eksempel
N_DEF_SING	substantiv bestemt, singularis,	hyldeblomsten
N_DEF_SING_GEN	substantiv, bestemt, singularis, genitiv	hyldeblomstens
N_DEF_PLU	substantiv, bestemt, pluralis	hyldeblomsterne
N_DEF_PLU_GEN	substantiv, bestemt, pluralis, genitiv	hyldeblomsternes
N_INDEF_SING	substantiv, ubestemt	hyldeblomst
N_INDEF_SING_GEN	substantiv, ubestemt, genitiv	hyldeblomsts
N_INDEF_PLU	substantiv, ubestemt, pluralis	hyldeblomster
N_INDEF_PLU_GEN	substantiv, ubestemt, pluralis, genitiv	hyldeblomsters
N_SING	substantiv, singularis	messias
N_SING_GEN	substantiv, singularis, genitiv	messias's
N_PLU	substantiv, pluralis	boligsøgende
N	substantiv	rub, særdeleshed, zig-zag
$V_{\perp}INF$	verbum, infinitiv	samle
V_PRES	verbum, præsens (nutid)	samler
V_PAST	verbum, præteritum (datid)	samlede
V_INF_PAS	verbum, infinitiv, passiv	samles
V_PRES_PAS	verbum, præsens, passiv	samles
V_PAST_PAS	verbum, præteritum, passiv	samledes
V_PARTC_PAST	verbum, participium præteritum	samlet
V_PARTC_PRES	verbum, participium præsens	samlende
V_GERUND	verbum, gerundium	samlen
V_IMP	verbum, imperativ	saml
EGEN	proprium	Eddie
EGEN_GEN	proprium, genitiv	Østrigs
ADJ	almindeligt adjektiv	økonomiske
ADV	adverbium	ud, ind, overfor
NAMEX_ADV[1]	dato	24.=maj=2005*DATE
NUM	kardinal adjektiv	to
NUM_ORD	ordinal adjektiv	3.

Tabel B.1: POS tags for substantiver, verber, proprium (egennavn), adjektiv og adverbier

Tag	Betydning	Eksempel
PRON_REC	reciprokt pronomen	hinanden
PRON_REC_GEN	reciprokt pronomen, genitiv	hinandens
PRON_DEMO	demonstrativ pronomen	den, det, de, denne, dette, disse
PRON_DEMO_GEN	demonstrativ pronomen genitiv	dennes, disses
PRON_UBST	ubestemt pronomen	en, et, noget, man, noget, nogen anden, andet, andre
PRON_UBST_GEN	ubestemt pronomen genitiv	ens, andres
PRON_POSS	possessivt pronomen	min, din
PRON_PERS	personligt pronomen	jeg, det, de, sig
PRON_INTER_REL	interrogativt/relativt pronomen	hvem, hvad, hvilke
PRON_INTER_REL_GEN	interrogativt/relativt pronomen genitiv	hvis
PRÆP	præposition	i, ad, på
UNIK	unik	som, der, at
SKONJ	sideordnende konjunktion	og, men
UKONJ	underordnende konjunktion	at, fordi, om, mens
INTERJ	lydord/udråbsord	nå, ak, næh
FORK	forkortelse	el., dr.
TEGN	interpunktionstegn	;;:!?()
FORM	formler	1986:5, 16V, OX99-11
SYMBOL	symboler	\$, *, ğ
XX	andet	fejl mm.
NO_CAT	Ord med tvivlsom kategori	Sparto, rabundus, adhoc

Tabel B.2: POS tags for alle pronomier, præpositioner, unikke, konjunktioner, lydord, forkortelser, tegn, formler, symboler og andet.

Appendix C

Bag of Words Algoritme

I dette kapitel beskrives en standardmetode til tekstklassificering, en algoritme kaldet Bag of Words. Algoritmen testes i forhold til både trænings- og testsættet med UTH-rapporter.

Dokumentklassifikation er en metode, der ønsker at tilskrive et givent dokument en af flere på forhånd definerede kategorier. Indenfor dokumentklassifikation er en de hyppigst anvendte metoder 'Bag of Words' eller en variant heraf. Bag of Words er en model, hvor en tekst er repræsenteret ved en uordnet samling af ord uden hensynstagen til hverken grammatik eller ordenernes rækkefølge. [Witten & Frank 2005]

Metode

Når modellen skal udvikles kan metoden inddeles i flere trin; præprocessering, featureudvælgelse og klassifikation. Disse trin vil blive gennemgået i det følgende.

Præprocessering

I præprocesseringen er antallet er mulige ord blevet reduceret, for dermed at reducere dimensionaliteten af featurerummet. Særlige ord er ekskluderet fra algoritmen, såkaldte stopord. Dette er ord hvis mening er kontekstuafhængig, som for eksempel og, men og fordi. En samlet liste af de 101 anvendte stopord fremgår af figur C.1 [ranks.nl 2008]. Ligeledes er teksten blevet omsat til kun at bestå af stammen af de forekomne ord. Stammen til hvert enkelt ord er fundet ved brug af en lemmatiseralgoritme udviklet af CST [cst.dk 2008]. På baggrund af en række regler og en ordbog, der begge afspejler forholdet mellem ordformer og grundformer bestemmes den enkelte stamme. Ved at bruge sådan en algoritme undgås multiple variationer af det samme ord.

Featureudvælgelse

Hvert element i featurevektoren indikerer et ords tilstedeværelse eller ej, hvor 1 angiver at det pågældende ord er til stede og 0 indikerer, at det ikke er tilstede. Featurevektoren genereres ud fra træningsdata, hvorved der opnås en vektor med 1421 forskellige ord, der enten kan være tilstede eller ej. Ordene er arrangeret alfabetisk hvor de første 10 ord som eksempel er: abe, absolut, abstinensbeh, abstinensbehandling, abstinenser, abstinensmedicin, abstinensscoringsskema, abstinsmedicin, acetrapid og acicloveie. Ud fra denne liste af ord genereres der

)
(af	alle	andet	andre	at
	begge	da	de	den	denne
	der	deres	det	dette	dig
	din	dog	du	ej	eller
	en	end	ene	eneste	enhver
	et	fem	fire	flere	fleste
	for	fordi	forrige	fra	få
	før	god	han	hans	har
	hendes	her	hun	hvad	hvem
	hver	hvilken	hvis	hvor	hvordan
	hvorfor	hvornår	i	ikke	ind
	ingen	intet	jeg	jeres	kan
	kom	kommer	lav	lidt	lille
	man	mand	mange	med	meget
	men	mens	mere	mig	ned
	ni	nogen	noget	ny	nyt
	nær	næste	næsten	og	ор
	otte	over	på	se	seks
	ses	som	stor	store	syv
	ti	til	to	tre	ud
	var)
_	_				

Figur C.1: Oversigt over de anvendte stopord. [ranks.nl 2008]

Figur C.2: Eksempel på featurevektorer før og efter de reduceres til sparse data format. [Witten & Frank 2005]

en unik featurevektor for hver eneste UTH-rapport. Udover de 1421 features tilføjes endnu en feature, en såkaldt identifikationsfesture, der angiver sagsnummeret for den pågældende rapportering.

Da mange af featurene vil antage værdien 0, idet et pågældende ord ikke er tilstede i rapporten transformeres featurevektorerne til sparse data. Et eksempel på det sparse format er vist på figur C.2. De første 2 linier indikere den originale featurevektor, hvorimod de nederste to linier viser de tilhørende sparse featurevektorer. Alle nullerne er blevet fjernet og i stedet indikerer det første element nu pladsen for den første feature, der har en værdi. Det betyder at klasse A's featurevektor har en værdi på 26 på plads 1.

Klassifikation

Som klassifikationsalgoritme er Support Vector Machine valgt, da dette er en anerkendt algoritme indenfor tekstklassificering. Til dette formål er softwaren Weka anvendt, der giver adgang til forskellige klassifikationsalgoritmer. [Witten & Frank 2005]

Resultater

Algoritmen testes på to måder. Ud fra en 10-fold krydsvalidering på træningsættet og test på testsættet. Ved en 10-fold krydsvalidering bliver datasættet inddelt i 10 undersæt, hvor algoritmen testes af 10 omgange. For hver omgang holdes ét sæt udenfor og bruges derved som testsæt, hvor de resterende ni sæt anvendes til træning af modellen. Dette forløber nu 10 gange og det samlede resultat af algoritmen skal ses som en kombination af resultaterne for de 10 omgange.

Ud fra 10-fold krydsvalidering opnås resultaterne, som fremgår af tabel C.1 og C.2, hvoraf det ses at de positive rapporter kun bliver klassificeres korrekt i 11,1% af tilfældene, hvorimod de negative rapporteringer klassificeres korrekt i 89,6 % af tilfælde. Da det i dette projekts problemstilling er vigtigt at undgå falsk positiv klassificerede rapporteringer, kan Bag of Words algoritmen ikke antages for tilstrækkelig, da procentdelen af de positivt klassificerede rapporteringer er for lav.

Præcision	Recall	F-Mål	Klasse
0,5	0,059	0,105	Positiv
0,814	0,986	0,892	Negativ

Tabel C.1: Resultaterne fra 10-fold krydsvalidering af Bag of Words algoritmen på træningssættet

a	b	<- Klassificeret som
1	16	$\mathbf{a} = \mathbf{Positiv}$
1	70	$\mathbf{b} = \mathbf{Negativ}$

Tabel C.2: Blandingsmatrice for 10-fold krydsvalidering af Bag of Words algoritmen på træningssættet

Som nævnt tidligere er algoritmen ligeledes testet på testsættet. Resultaterne af denne test fremgår af tabel C.3 og C.4, hvilket yderligere understøtter at Bag of Words, ikke er en brugbar algoritme til klassifikation af UTH-rapporter, da ingen af de positive rapporter i dette tilfælde bliver klassificeret korrekt.

Præcision	Recall	F-Mål	Klasse
0	0	0	Positiv
0,805	0,943	0,868	Negativ

 $\textbf{Tabel C.3:} \ \operatorname{Resultaterne} \ \operatorname{fra} \ 10\text{-fold} \ \operatorname{krydsvalidering} \ \operatorname{af} \ \operatorname{Bag} \ \operatorname{of} \ \operatorname{Words} \ \operatorname{algorithmen} \ \operatorname{på} \ \operatorname{tests} \\ \text{\approxtest} \ \operatorname{test} \ \operatorname{algorithmen} \ \operatorname{på} \ \operatorname{tests} \\ \text{\approxtest} \ \operatorname{algorithmen} \ \operatorname{på} \ \operatorname{algorithmen} \ \operatorname{på} \ \operatorname{tests} \\ \text{\approxtest} \ \operatorname{algorithmen} \ \operatorname{på} \ \operatorname{algor$

a	b	<- Klassificeret som
0	8	$\mathbf{a} = \mathbf{Positiv}$
2	33	$\mathbf{b} = \mathbf{Negativ}$

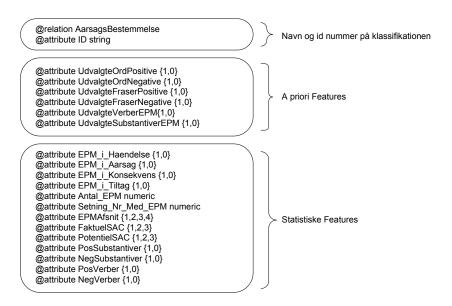
Tabel C.4: Blandingsmatrice for 10-fold krydsvalidering af Bag of Words algoritmen på testsættet

Appendix D

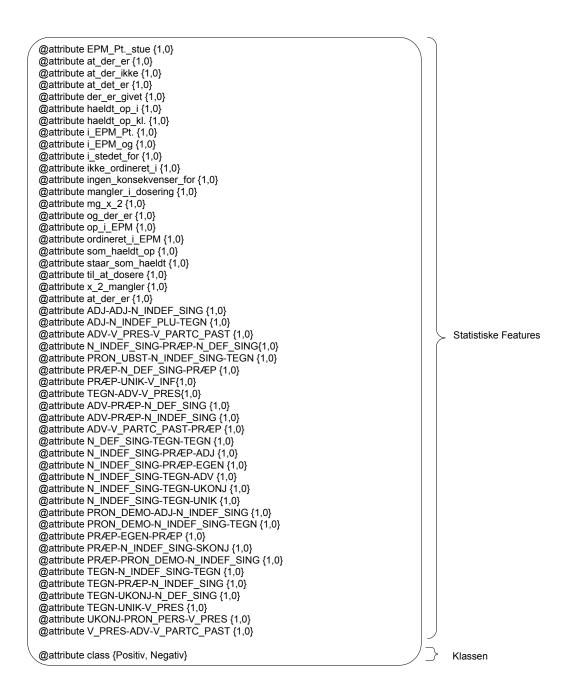
Udtrukket Features

Dette appendiks viser, hvilke features der er uddraget, inden den endelige featureudvælgelse finder sted. Ligeledes vises der en samling over de ord og fraser, der er identificeret ud fra eksperternes markeringer i UTH-træningsrapporter.

Et eksempel på hvorledes, både a priori og de statistiske features bliver repræsenteret inden klassifikation kan ses på figur D.1, figur D.2 og figur D.3. Den første linie '@relation' er navnet for klassifikation. Ud for hver '@attribute' er featurens navn, og tallene inde i tuborgklammerne viser de mulige featureværdier. Efter '@attribute' er angivet '@class', som er input til om rapporten tilhører en positiv eller negativ klasse. I testsituationer, kan det ikke angives, da den ikke vides på forhånd. Efter '@class' kommer '@data' hvor hver featurevektor tilsvarer en rapport, hvor det første tal angiver sagsnummeret og efterfølgende værdien for hver af featuerne, '@attribute', som forekommer. De ord og fraser som eksperterne har markeret i UTH-rapporterne er vist i figur D.4.



Figur D.1: @relation er navnet på klassifikationen, @attribute er features. Til højre er markeret, hvilke der er a priori og statistiske features. Figuren fortsættes på figur D.2.



Figur D.2: Statistiske features fortsat fra figur D.1 og fortsættes på figur D.3. De første 22 features angiver treords fraser, mens de resterende features angiver treords POS. Nederst angives klassen.

@data $\mathsf{ID12}, \mathsf{1}, \mathsf{1}, \mathsf{1}, \mathsf{0}, \mathsf{1}, \mathsf{0}, \mathsf{1}, \mathsf{0}, \mathsf{1}, \mathsf{0}, \mathsf{4}, \mathsf{2}, \mathsf{2}, \mathsf{3}, \mathsf{1}, \mathsf{1}, \mathsf{1}, \mathsf{1}, \mathsf{0}, \mathsf$

Featurevektorerne

Figur D.3: @data angiver at efterfølgende tilsvarer hver UTH-rapport en featurevektor. Her er eksemplet vist for 13 UTH-rapporter, 6 negative og 7 positive, og er fortsat fra figur D.2.

faneblad, logge, serverfejl, mus, dispenseringsoversigt, lukke, advarselsboks, poppe, advare, systemfejl, tekniskfejl, fane, software Udvalgte Negative Ord ordination, ordinere, ord, ord., ordinering, etiket, adm, adm., administrere, administrering, administration, dok., dok, dokumentere, dokumentering, dokumentation, disp, disp., dispensering, dispensation, dispensere, sep., sep, seponere, seponering, seponation, blandecentral Udvalgte Positive Fraser (eknisk fejl, advare ikke, automatisk seponere, automatisk dispensere, automatisk administrer, advare ikke, epm struktur, trykke forkert, taste forkert, ændring PRÆP epm Udvalgte Negative Fraser (ikke være overensstemmelse, tjekke ikke PRÆP epm, så forkert, fejl PRÆP journal, ikke lægge PRÆP epm, ikke lægge PRÆP epm

Figur D.4: De ord og fraser fra henholdsvis den positive og den negative klasse, som eksperterne har identificeret i UTH-træningsrapporterne.

Appendix E

Tænke Højt Test

I dette kapitel gives et referat af hvad de to brugere gjorde og sagde under tænke højt testen. Denne information er opnået ud fra efterbehandling af den optagede video og skærmoptagelse. Citater fra brugerne er angivet med kursiv

Bruger 1

Foretag en årsagsbestemmelse ud fra filen 'brugertest.txt' beliggende på 'skrivebordet'. Det ønskes at afgøre hvorvidt EPM/EMP er årsag til de utilsigtede hændelser eller ej.

Brugeren starter med at se hvad der er på begge faneblade på brugergrænsefladen for 'Valg af Årsagsbestemmelse'. Herefter finder brugeren frem til kildefilen ved at klikke på gennemseknappen. Brugeren læser med det samme værktøjstippet, da han skal angive årsagen til hændelsen. Brugeren siger 'så forsvandt billedet - hmm', da startvinduet forsvinder. Da det næste vindue kommer frem er brugeren lidt i tvivl om systemet er færdig med at finde rapporter, men bliver overbevist da han læser titlen på vinduet: 'Nå, her står resultat af årsagsbestemmelsen - okay'.

Læs første hændelse, hvor EPM er årsag til den utilsigtede hændelse.

Brugeren ser med det samme at det ikke er hele hændelsen, der står beskrevet i oversigten: 'Men hvis jeg klikker på den kan det være, at den kommer frem', hvorefter han læser hændelsen igennem. Brugeren klikker på nogle af de andre sager og kommenterer: 'Det er en meget fin oversigt'.

Hvor mange rapporter findes der, hvor EPM ikke er årsag til den utilsigtede hændelse?

Brugeren ser, at der er to faner og finder ud af, at han befinder sig på fanebladet 'Årsag', hvorefter han lokaliserer antallet af rapporterne nederst på brugergrænsefladen. 'Jeg havde måske forventet at lede efter det her' siger brugeren og peger under oversigten med alle positive rapporter.

Hvor mange rapporter med EPM som årsag har en potentiel risikoscore på 3? Brugeren sikrer sig, at han er på det rigtige faneblad, hvorefter han klikker på SAC P overskriften: 'Mon ikke jeg kan sortere dem, ved at klikke på dem'. Brugeren klikker flere gange

for at få den til at sortere omvendt: 'Der havde jeg så håbet på at den ville sortere omvendt ved at klikke på den igen' men går i stedet i gang med at tælle hvor mange rapporter der er. Brugeren prøver herefter at højreklikke, for derved at se om der kommer andre muligheder for sortering frem, dog uden held.

Gem årsagsbestemmelsen på 'skrivebordet'.

Brugeren er i tvivl om hvad årsagsbestemmelsen er: 'Årsagsbestemmelsen, det må næsten være det jeg har her', hvorefter han finder ikonet for 'Gem' og gemmer herefter klassifikationen på skrivebordet.

Vend tilbage til startvinduet.

'Det kommer vi jo nok ved at lukke heroppe' siger brugeren, hvorefter han klikker på krydset øverst til højre på brugergrænsefladen. Herefter lukker hele programmet ned og testlederen er nødt til at starte det op igen.

Åbn årsagsbestemmelsen du gemte under den tidligere øvelse.

Ja, så skal jeg jo nok ikke have en kildefil, siger brugeren idet han står på fanebladet på 'Ny Årsagsbestemmelse', hvorefter han klikker på fanebladet for 'Tidligere Årsagsbestemmelse'. Brugeren undrer sig over, at kildefilen hedder det samme begge steder, men vælger at angive stien til den fil han gemte før, hvorefter der klikkes på 'Åbn'-knappen.

I hvor mange rapporter er EPM årsag til hændelsen?

Brugeren kan fra en tidligere opgave huske hvor denne information er at finde, og løser derved opgaven med det samme.

Er sagsnummer xxxxx15 klassificeret korrekt? Hvis ikke skal dette rettes.

Brugeren forsøger først at finde sagsnummeret i fanebladet for 'Årsag'. Dette gør han ved at sortere i forhold til sagsnumre. Da han ikke finder rapporten her, klikker han på det modsatte faneblad og sorterer ligeledes disse rapporter efter sagsnummer. Kommer frem til den korrekte sag ved at scrolle på musen. Brugeren bliver i tvivl om hvilket faneblad han befinder sig i: 'Spørgsmålet er om jeg er i Årsag eller Ikke Årsag', hvorefter han klikker på begge faner, for at finde ud af at han er i 'Ikke Årsag'. Da brugeren finder ud af at rapporten er klassificeret forkert siger han: 'Så ved jeg ikke om vi skal sige afvis sag. Det virker så voldsomt, som om det slet ikke er en gyldig sag', men vælger at klikke på den alligevel. Da han klikker over på fanebladet for 'Årsag' ser han at den afviste rapport er blevet grøn: 'Nej, og nu er den blevet grøn - hvor smart. Det gør det tydeligt og nemt for mig at finde den igen.'.

Har antallet af rapporter hvor EPM er årsag til de utilsigtede hændelser ændret sig?

Brugeren ser med det samme at der er kommet en sag mere.

Print årsagsbestemmelsen.

Brugeren klikker med det samme på ikonet for print. Da printet kommer frem i et andet

vindue bliver brugeren forvirret: 'Kan det være her den er printet'. Det vælger brugeren at antage, at det er, hvorefter han ser nærmere på printet: 'Det er meget fint at der er de der inddelinger. Det gør det nemt for mig at se dem.'

Luk printerprogrammet og programmet til årsagsbestemmelsen ned.

Brugeren lukker begge programmer ned ved at klikke på krydset.

Bruger 2

Foretag en årsagsbestemmelse ud fra filen 'brugertest.txt' beliggende på 'skrivebordet'. Det ønskes at afgøre hvorvidt EPM/EMP er årsag til de utilsigtede hændelser eller ej.

Brugeren starter med at finde kildefilen: 'Det tror jeg godt at jeg kan finde ud af', men bliver lidt forvirret, da han skal finde frem til skrivebordet i 'Åbn ny fil'-dialogen: 'Uha, det var godt nok ikke ret overskueligt', men finder filen. Da årsagen skal indtastes, ser han værktøjstippet, hvorefter han finder ud af, at der kan skrives flere stavekombinationer. Da vinduet forsvinder, bliver han overrasket idet terminalvinduet hvorfra systemet kaldes popper op foran brugergrænsefladen for systemet. Testlederen er nødt til at gribe ind og finde resultatvinduet frem, hvorefter brugeren udbryder: Hov - nå så får jeg det simpelthen smidt i hovedet her.

Læs første hændelse, hvor EPM er årsag til den utilsigtede hændelse.

Brugeren klikker på fanerne for at finde ud af hvilket faneblad han befinder sig i. Da han finder ud af at han er i 'Årsag' fanebladet klikker han på sagsnummeret for den første rapport og læser den tilhørende tekst.

Hvor mange rapporter findes der, hvor EPM ikke er årsag til den utilsigtede hændelse?

Brugeren finder frem til informationen nederst på brugergrænsefladen og vælger at regne differencen på antallet af rapporter og antallet af positive rapporter ud, for derved at finde ud af hvor mange rapporter, der var klassificeret negativt, i stedet for at skifte fane.

Hvor mange rapporter med EPM som årsag har en potentiel risikoscore på 3?

Brugeren klikker på overskriften for SAC P flere gange: 'Ser der ville jeg jo have forventet at den sorterede den anden vej'. 'Nu ved jeg så ikke rigtig', men brugeren beslutter sig for at scrolle ned i bunden af listen, og tælle antallet af rapporter med SAC P på 3.

Gem årsagsbestemmelsen på 'skrivebordet'.

Brugeren går i stå: $\emptyset h...$ Testlederen spørger hvad han forventer at finde, hvorefter brugeren selv får øje på gem-ikonet: 'Jeg ville nok lede efter noget hvor der også stod noget tekst'. Herefter finder brugeren frem til skrivebordet og får gemt årsagsbestemmelsen.

Vend tilbage til startvinduet.

Brugeren skal nu vende tilbage til startvinduet, men spørger til det vindue han befinder sig

i: 'Er det det her?'. Testlederen forklarer, at startvinduet er der hvor han startede, hvorefter brugeren får øje på et ikon, han tror det kan være. Først da dialogboksen spørger ham om han ønsker at vende tilbage til startvinduet, ved han at han er på rette vej.

Åbn årsagsbestemmelsen du gemte under den tidligere øvelse.

Brugeren får øje på fanebladet tidligere årsagsbestemmelse, men bliver forvirret over, at der er en Åbn-knap: 'Tidligere Årsagsbestemmelser ligger jo lige der, men hvorfor er der en åbn-knap der'. Han beslutter sig for at nok skal bruge den senere, hvorefter han klikker på gennemse. Brugeren finder det underligt at han skal klikke på Åbn to gange.

I hvor mange rapporter er EPM årsag til hændelsen?

Brugeren ser med det samme informationen nederst på brugergrænsefladen.

Er sagsnummer xxxxx15 klassificeret korrekt? Hvis ikke skal dette rettes.

Brugeren starter med at sortere efter sagsnummer. Igen forsøger han at sortere begge veje. Han ser, at den ikke er i den fane, hvor han står, hvorefter han klikker på den anden. Han finder frem til den korrekte sag. Brugeren forventer nu, at han skal klikke på afvis sag, hvorefter den vil blive flyttet til det andet faneblad: 'Ja så blev den væk. Jeg tjekker lige om den er ovre i den anden så'.

Har antallet af rapporter hvor EPM er årsag til de utilsigtede hændelser ændret sig?

Brugeren ser nu at der er kommet en hændelse mere. Han finder frem til hændelsen og ser at den er blevet markeret med grøn: 'Er det så fordi at jeg har været inde og lave om på den - det er smart'.

Print årsagsbestemmelsen.

Brugeren klikker på ikonet for print: 'Det må være den der'. Brugeren kigger på printet og går ud fra at 'positiv' betyder at søgeordet var årsag, hvorefter han printervinduet ned.

Luk printerprogrammet og programmet til årsagsbestemmelsen ned.

Brugeren afslutter programmet ved at klikke på krydset øverst i højre hjørne: Ønsker du at afslutte programmet - ja joeh..Jamen det var jo slet ikke så svært som jeg troede..

Bilag A

Region Hovedstadens Rapporteringsskema

I dette bilag vises, hvorledes Region Hovedstadens rapporteringsskema er struktureret. Rapporteringsskemaet består af to sider. Disse fremgår af figur A.1 og A.2.

Rapportering af: UTH

Utilsigtet Hændelse

Hvad er en utilsigtet hændelse?

En utilsigtet hændelse er en begivenhed, der medførte skade eller kunne have medført skade og som er en følge af undersøgelse, behandling eller pleje eller mangel herpå, og som ikke skyldes patientens underliggende sygdom.

Alle klinisk betydende hændelser hvor der skete patientskade eller kunne være sket skade, og hvor der er risiko for gentagelse og potentiale for læring skal rapporteres.

Hvordan rapporterer jeg en utilsigtet hændelse?

Udfyld skemaet nedenfor så udførligt og præcist som muligt. Alle felter er obligatoriske (patientoplysninger dog kun, hvis der er mindst 1 involveret patient).

Hvad sker der med min rapport?

Når du har udfyldt skemaet og klikket på send-knappen, bliver din rapport gemt i en database. Herefter kontakter systemet de relevante personer, dvs. ledelsen og den patientsikkerhedsansvarlige på afdelingen/klinikken samt hospitalets risikomanager. Hændelsen bliver herefter analyseret, kategoriseret og anonymiseret inden overførsel til H:S database for utilsigtede hændelser.

Alvorlige hændelser bliver gennemgået ved en kerneårsagsanalyse, hvor et til lejligheden nedsat udvalg analyserer hændelsen for at finde ud af, hvordan man kan forebygge lignende hændelser i fremtiden.

Mindre alvorlige hændelser vil ofte indgå i aggregerede kerneårsagsanalyser, hvor man analyserer flere hændelser med fællestræk.

Du skal ikke forvente at få "svar" på din rapport. Men du kan altid henvende sig til den patientsikkerhedsansvarlige, hvis du har spørgsmål eller forslag til patientsikkerhedsarbejdet på din egen afdeling.

Hvornår skete det?

Angiv dato og tidspunkt for hændelsen. Hvis hændelsen strækker sig over længere tid, angiv da starttidspunktet. Hvis du ikke er sikker på tidspunktet, angiv da dit bedste skøn eller tidspunktet, hvor hændelsen blev opdaget.

Dato:	Sklokken ca.:	
Hvor skete det? Angiv hospital og afdeling, hvor h hvor hændelsen begyndte.	ændelsen fandt sted. Hvis flere steder	er involverede, angiv da stedet
Her kan du skrive eventuelle sup	plerende oplysninger om afdeling/klinik	·.
S .	sse eller telefonnummer herunder. Hvi oruge din rapport fordi det så ikke vil va	3 1 3
Navn:	Email:	Telefonnr.:
112111		
Rolle:		
	•	

Figur A.1: Figuren viser side 1 af Region Hovedstadens rapporteringsskema.

Hvornår skete det? Beskriv hændelsens forløb så udførligt og nøjagtigt som muligt. I dette felt skal du kun skrive, hvad der
skete. Husk at medtage navne på fx medicin og udstyr, der var involveret i hændelsen.
Hvorfor skete det?
Her kan du forklare, hvilke årsager du tror var medvirkende til, at hændelsen kunne ske. Det kan fx dreje sig omkommunikationsproblemer, mangelfuld oplæring, uhensigtsmæssigt arbejdsmiljø, fejl ved apparatur, uklare retningslinier eller svigtende kontrolforanstaltninger. En detaljeret medicinliste kan være relevant fx i forbindelse med fald og medicineringsfejl.
Hvilke konsekvenser havde hændelsen for patienten? Beskriv hændelsens konsekvenser for den eller de berørte patienter, fx fysiske eller psykiske skader, forlænget indlæggelse eller øget udredning. Ved nærhændelser eller hændelser uden patientskade skrives blot: "Ingen"
Skirves blot. Ingen
Hvordan, tror du, man kunne undgå lignende hændelser i fremtiden?
Giv dine forslag til, hvordan man kan undgå lignende hændelser i fremtiden. Det kan fx dreje sig om
ændrede procedurer, bedre oplæring, andet udstyr eller indførelse af kontrolforanstaltninger.
Patientoplysninger
Du skal her opgive nogle data om den patient, der var (mest, hvis flere) udsat for den utilsigtede
hændelse: Antal: Patientens køn • Angiv patienttype • Alder:

 ${\bf Figur~A.2:}~{\rm Figuren~viser~side~2~af~Region~Hoved stadens~rapportering sskema}.$