



Automatisk klassifikation af utilsigtede hændelser fra Dansk Patientsikkerhedsdatabase - et system baseret på Naive Bayes



Titel:

Automatisk klassifikation af
utilsigtede hændelser fra Dansk
Patientsikkerhedsdatabase
- et system baseret på Naive Bayes

Tema:

Specialeprojekt

Projektperiode:

P10, forårssemester 2012

Projektgruppe: 1073

Deltagere:

Anders Schierup
Christian Bruun Jensen
Duy Thien Van

Vejleder:

Ole Hejlesen

Oplagstal: 5

Sidetal: 93

Bilag: CD-ROM

Afsluttet: 1. juni 2012

Synopsis:

Siden loven om patientsikkerhed trådte i kraft i 2004, har det været lovpligtigt at indberette utilsigtede hændelser til Dansk Patientsikkerhedsdatabase (DPSD). I 2010 blev der rapporteret ca. 35.000 hændelser, hvilket er en tidobling siden indførelsen af rapporteringspligten. I dag klassificeres rapporter ved en manuel retrospektiv gennemgang fortaget af risikomanagere og ansatte ved Patientombuddet.

I dette projekt undersøges potentialet for anvendelse af automatisk tekstklassifikation til kategorisering af IT-relaterede utilsigtede hændelser indrapporteret til DPSD. Et system kaldet UTH-Finder blev udviklet, som ved anvendelse af en Naive Bayes algoritme kunne processere og klassificere DPSD rapporter.

Et balanceret træningssæt på 482 rapporter (241 IT-relaterede og 241 andre hændelser) og tre testsæt på hver 1.000 rapporter blev anvendt til at teste systemet. Resultatet af testen viste en nøjagtighed og specificitet på 95 % ved klassifikation af IT-relaterede utilsigtede hændelser.

Det konkluderes med udviklingen af UTH-Finder at automatisk tekstklassifikation har potentiale for anvendelse til kategorisering af utilsigtede hændelser fra DPSD.

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



Title:

Automatic classification of adverse events from the Danish Patient Safety Database
- a system based on Naive Bayes

Subject:

Master's thesis

Project period:

P10, spring semester 2012

Project group: 1073

Participants:

Anders Schierup
Christian Bruun Jensen
Duy Thien Van

Supervisor:

Ole Hejlesen

Copies: 5

Pages: 93

Annexe: CD-ROM

Finished: 1 June 2012

Abstract:

Since the law of patient safety came into force in Denmark in 2004 it has been required by law to report adverse events to the Danish Patient Safety Database (DPSD). 35.000 incidents were reported in 2010, which is a tenfold increase since the introduction of the law. Adverse events are currently classified by a manual retrospective review by risk managers and employees at Patientombudet.

This project investigated the potential of using automatic text classification to categorize IT-related adverse events reported to DPSD. A system called UTH-Finder was developed, which is able to process and classify DPSD reports by using a Naive Bayes algorithm.

A balanced training set containing 482 reports (241 IT-related and 241 other events) and three test set each containing 1.000 reports was used to test the system. The system classified IT-related adverse events with an accuracy and specificity of 95 %.

With the development of UTH-Finder it is concluded that automatic text classification has a potential to be used in categorization of adverse events from DPSD.

*The reports content is freely accessible,
but publication (with references) may be made
only after agreement with the authors.*

Forord

Denne rapport er udarbejdet af projektgruppe 1073 på 4. semester af kandidatuddannelsen i Sundhedsteknologi i perioden 1. februar 2012 til 1. juni 2012 ved Aalborg Universitet.

Formålet med dette projekt er at bidrage med viden om potentialet for anvendelse af statistisk tekstklassifikation til klassificering af utilsigtede hændelser. Derudover bidrages der med viden om hvordan et system til klassificering af utilsigtede hændelser fra Dansk Patientikkerhedsdatabase (DPSD) kan implementeres, hvilket vil understøtte regionale risikomanagere, samt analytikere ved DPSDs arbejde imod en mere konsistent klassifikation af utilsigtede hændelser.

Rapporten henvender sig til regionale risikomanagere, ansatte ved DPSD, vejledere, medstuderende og andre med interesse for hvorledes systemer til automatisk klassificering af utilsigtede hændelser kan udvikles og sikre en konsistent klassificering på tværs af regionerne. Desuden henvender rapporten sig til Australian Institute of Health Innovation på University of New South Wales, Australien, idet projektet er en videreudvikling fra et 3. semesters kandidatprojekt udarbejdet i samarbejde med Dr. Farah Magrabi og Mei-Sing Ong.

Vi vil rette en tak til Patientombuddet, som villigt har stillet lokaler og data fra DPSD til rådighed. Specielt vil vi rette en tak til MD, ph.d. Kim Lyngby Mikkelsen for faglig sparring under vores ophold ved Patientombuddet. Ligeledes vil vi takke vores forsøgspersoner for deres medvirken i usability testene, som bidrog med indsigtfulde kommentarer, der sikrede et udbytterigt resultat.

Aalborg Universitet, juni 2012.

Anders Schierup

Christian Bruun Jensen

Duy Thien Van

Læsevejledning

Rapporten er opbygget i fire hoveddele:

- Del I: Problemanalyse
- Del II: Problemløsning
- Del III: Syntese
- Del IV: Appendiks

Projektgruppen anbefaler at rapporten læses samlet og fortløbende. Appendiks indeholder fem kapitler indeholdende supplerende oplysninger, som understøtter rapportens indhold.

Referencer er angivet efter Harvardmetoden på følgende måde [Efternavn, årstal], hvor forfatterens efternavn og publikationsår er angivet i den firkantede parentes. Referencer er angivet både før og efter punktum. Såfremt en reference står før et punktum, henvises der til den pågældende sætning, mens hvis referencen står efter et punktum, henvises der til hele afsnittet. Figurer og tabellers referencenummer refererer til hvilket kapitel de tilhører og hvilken placering de har heri. Eksempelvis refererer "Figur 2.1" til figur 1 i kapitel 2.

I rapporten refereres der til et 9. semesterprojekt af Anders Schierup, Jakob Skrivers Routhe og Christian Bruun Jensen omhandlende udvikling af en metode, samt software til automatisk tekstklassifikation af utilsigtede hændelser fra MAUDE- og AIMS-databaserne. Denne rapport refereres således [Schierup et al., 2011].

Bagerst i rapporten er vedlagt en cd-rom indeholdende en eksekverbar version af UTH-Finder, samt kildekode i form af Java-filer. En elektronisk PDF-udgave af projekt-rapporten findes ligeledes på den vedlagte cd-rom. Data anvendt til udvikling, træning og test af systemet er udleveret af DPSD. Rapporterne er af fortrolig karakter og er derfor ikke medtaget.

Indhold

1	Indledning	1
2	Metode	5
2.1	Metodeinddeling og litteraturreview	5
I Problemanalyse		
3	Rapporteringssystemer til utilsigtede hændelser	13
3.1	Anvendelse af rapporteringssystemer i sundhedssektoren	13
3.2	Sanktionerende og ikke-sanktionerende rapporteringssystemer	14
3.3	Opbygning af et rapporteringssystem til utilsigtede hændelser	15
4	Rapportering af utilsigtede hændelser i Danmark	19
4.1	Dansk Patientsikkerhedsdatabase	19
4.2	Indrapporteringsforløbet af en utilsigtet hændelse	20
4.3	Struktur af utilsigtet hændelsesrapport	20
5	Statistisk tekstklassifikation af utilsigtede hændelser	23
5.1	Anvendelse af machine learning algoritmer til tekstklassifikation	23
6	Problemformulering	25
II Problemløsning		
7	Systemudvikling	29
7.1	Metode til systemudvikling	29
7.2	Kravworkflow	29
7.3	Analyseworkflow	33
7.4	Implementering	37
7.5	Opsamling på systemudvikling	43
8	Systemvalidering	45
8.1	Resultater fra læringstest	45
8.2	Resultater fra effektivitetstest	46
III Syntese		
9	Diskussion	51
10	Konklusion	55
11	Perspektivering	57

Litteratur	59
IV Appendiks	
A Indrapportering til Dansk Patientsikkerhedsdatabase	63
A.1 Rapporteringsskemaets opbygning	63
B Test af systemet	69
B.1 Typer af test	69
B.2 Integrationstest	69
B.3 Systemtest	72
C Naive Bayes	73
D Usability test	75
D.1 Evaluering af brugervenlighed og -oplevelse	75
D.2 Målsætning med usability test	75
D.3 Forsøgsprotokol for Incident Tracker	76
D.4 Resultater fra Incident Tracker	78
D.5 Design af det nye system	83
E Forsøgsprotokol for systemvalidering	85
E.1 Formål	85
E.2 Præprocessering	85
E.3 Fremstilling af trænings- og testsæt	86
E.4 Klassificering	87
E.5 Resultater	88
E.6 Databehandling	91

Indledning

I de seneste 20 år har den sundhedspolitiske debat været fokuseret på kvalitetsudvikling af sundhedsvæsenets ydelser. Sundhedsstyrelsen udviklede i 1993 en fælles national strategi for kvalitetsudvikling. Den blev i 1999 fulgt op med etablering af Det Nationale Råd for kvalitetsudvikling, hvis formål er at koordinere kvalitetsudviklingsaktiviteter, igangsætte nye initiativer og udarbejde en revideret national strategi for kvalitetsudvikling i sundhedsvæsenet. [DSKS, 2003]

Internationale studier havde i perioden 1991-2001 udført strukturerede gennemgange af journaler fra somatiske sygehuse. Studierne estimerede forekomsten af utilsigtede hændelser og evnen til at forebygge disse. På baggrund af publikationerne iværksatte DSI Institut for Sundhedsvæsen i samarbejde med en række amter den første danske journalaudit med det formål at undersøge forekomsten af utilsigtede hændelser på somatiske sygehuse i Danmark. Undersøgelsen blev publiceret i 2001 og viste at der skete utilsigtede hændelser ved 9 % af alle indlæggelser. Af disse utilsigtede hændelser var 40 % fejl, der kunne være forhindret ved bedre brug af eksisterende viden og teknologi, mens de resterende 60 % dækkede over andre komplikationer. [DSI, 2001]

På baggrund af undersøgelsen udgav Det Nationale Råd for Kvalitetsudvikling og Sundhedsstyrelsen i 2002 en strategi for kvalitetsudvikling med fælles mål og handleplan for perioden 2002 til 2006. Strategiens overordnede mål var at sikre en høj professionel standard, effektiv ressourceudnyttelse, minimal patientrisiko, høj patienttilfredshed og helhed i patientforløb, hvori forebyggelse af fejl og utilsigtede hændelser var af essentiel karakter. [Sundhedsstyrelsen, 2002]

Patientombuddets definition og håndtering af utilsigtede hændelser kom senere i Sundhedslovens § 198, stk. 4, hvor en utilsigtet hændelse defineres som:

”Ved en utilsigtet hændelse forstås en begivenhed, der forekommer i forbindelse med sundhedsfaglig virksomhed, herunder præhospital indsats, eller i forbindelse med forsyning af og information om lægemidler. Utilsigtede hændelser omfatter på forhånd kendte og ukendte hændelser og fejl, som ikke skyldes patientens sygdom, og som enten er skadevoldende eller kunne have været skadevoldende, men forinden blev afværget eller i øvrigt ikke indtraf på grund af andre omstændigheder.”

[Sundhedsministeriet, 2010]

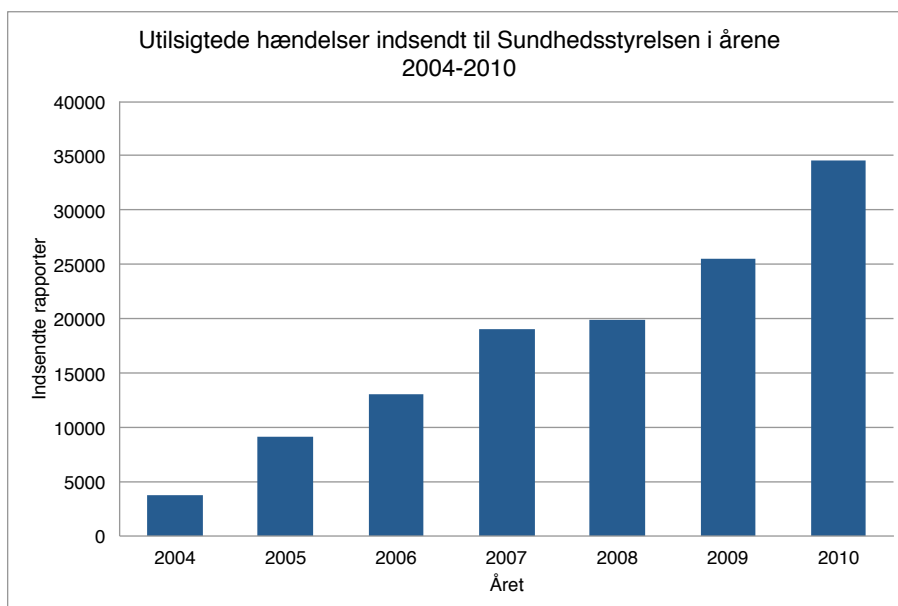
I 2004 trådte loven om patientsikkerhed i kraft. Det betød at sundhedspersoner fremover skulle indrapportere utilsigtede hændelser, som enten de selv var involveret i eller som de observerede hos kollegaer i sygehusvæsenet. Med denne lov blev der systematisk skabt opmærksomhed om forebyggelse af utilsigtede hændelser. [Sundhedsstyrelsen, 2002]

I sundhedsloven er sundhedspersoner defineret som:

”Ved sundhedspersoner forstås i denne lov personer, der er autoriseret i henhold til særlig lovgivning til at varetage sundhedsfaglige opgaver, og personer, der handler på disses ansvar”.

[Sundhedsministeriet, 2010]

I Danmark indrapporteres utilsigtede hændelser til Dansk Patientsikkerhedsdatabase (DPSD), som varetages af Patientombuddet. Siden 2004, hvor det blev lovpligtigt for sundhedspersoner at indrapportere utilsigtede hændelser i sygehusvæsenet, har antallet af rapporterede hændelser været stigende. Figur 1.1 viser at der i 2010 blev indrapporteret 34.418 hændelser, hvor der til sammenligning i 2004 blev indrapporteret 3.626 hændelser. [Sundhedsstyrelsen, 2010]



Figur 1.1: Opgørelse over den sekundære rapportering af utilsigtede hændelser indsendt til Sundhedsstyrelsen i årene 2004-2010. Den sekundære rapportering foretages af regionen, kommunen eller det private sygehus. [Sundhedsstyrelsen, 2010]

Analysen af de indrapporterede hændelser foregår på nuværende tidspunkt som en manuel, retrospektiv gennemgang af de indsendte rapporter. Denne metode er en ressourcekrævende proces og med den udvidede patientsikkerhedsordning, som trådte i kraft pr. 1. september 2010, forventes det at antallet af indrapporterede hændelser vil

stige yderligere i fremtiden. [Johnson, 2003; Sundhedsstyrelsen, 2010]
Udvidelsen af patientsikkerhedsordningen betyder at indrapporteringspligten udvides til at omfatte praksissektoren, den kommunale sundhedssektor, apotekerne og det præhospitale område, samt muligheden for at patienter og pårørende ligeledes kan indrapportere utilsigtede hændelser. [Sundhedsstyrelsen, 2010]

Klassificering af utilsigtede hændelser i forskellige kategorier er en af udfordringerne ved analysen [Johnson, 2003]. Klassificeringen foretages af forskellige personer med forskellige medicinske baggrunde, som ikke er eksperter i klassificering. Dette kan gøre analysen besværlig, da nogle rapporter ikke bliver korrekt klassificeret. Dette indikerer, at der er brug for mere effektive metoder til klassificering, hvormed udtræk af information til forebyggelse af utilsigtede hændelser kan forbedres.

Efterhånden som brugen af IT i sundhedssektoren øges, sker der en stigning i antallet af fejl relateret til IT [Sittig og Singh, 2011]. Brugen af IT i sundhedsvæsenet kan fremme sikkerheden og nedbringe antallet af fejl, men studier har vist at IT betragtes som en faktor, der øger risikoen for utilsigtede hændelser [Weiner et al., 2007].

Implementeringen af den Elektroniske Patientjournal (EPJ) betragtes som værende essentiel i forbedringen af sundhedsvæsenet og bliver derfor brugt i et større omfang. En rapport fra US Institute of Medicine (IOM) beskriver indflydelsen af IT i sundhedsvæsenet. IOM konkluderer at der er brug for tiltag, som kan gøre brugen af IT mere sikker, herunder et system som kan identificere og analysere utilsigtede hændelser relateret til IT. [Warden, 2012]

Tidligere studier af Ong et al. [2010] og Schierup et al. [2011] har vist, at automatisk kategorisering ved brug af statistisk tekstklassificering har potentiale som erstatning for den manuelle kategorisering.

Schierup et al. [2011] udviklede et system til evaluering af algoritmer til automatisk klassificering af utilsigtede hændelser. Resultatet af projektet viste at statistisk tekstklassificering er en valid metode til identificering af utilsigtede hændelser relateret til IT.

For at kunne udvikle et system som faciliterer automatisk klassificering af utilsigtede hændelser, er det en forudsætning at vide, hvordan indrapporteringen er struktureret, herunder hvordan en utilsigtet hændelse indberettes i et system, hvordan en rapport er struktureret og hvilke informationer der gemmes.

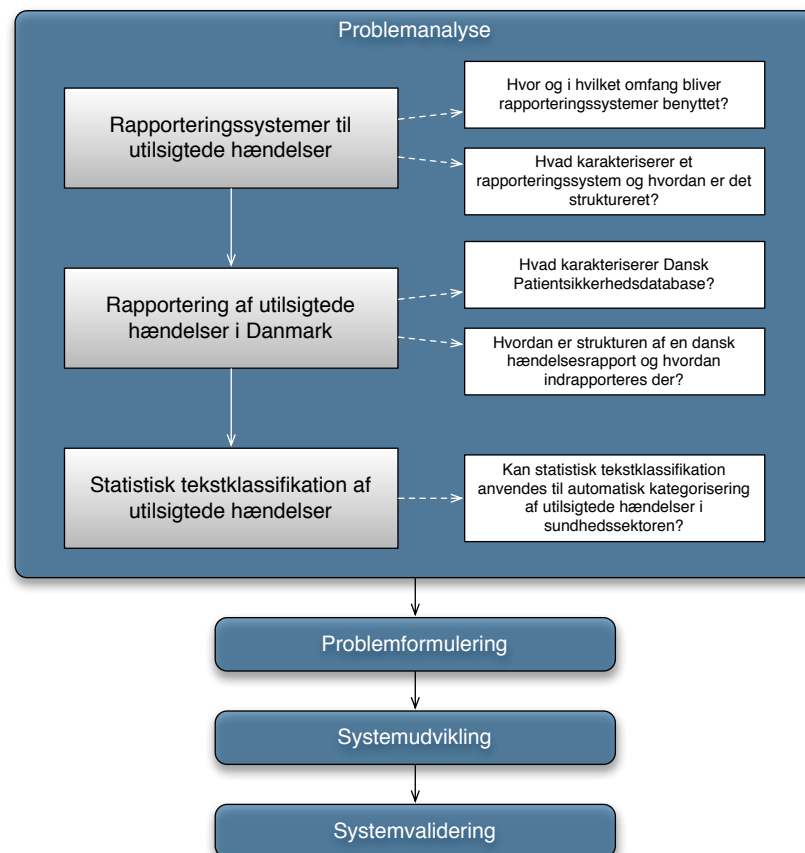
Således er den initierende problemstilling givet ved:

"Hvad karakteriserer det danske indrapporteringssystem for utilsigtede hændelser og hvordan er en rapport struktureret?"

I dette kapitel beskrives metoden, der anvendes til at besvare den initierende problemstilling: "Hvad karakteriserer det danske indrapporteringssystem for utilsigtede hændelser og hvordan er en rapport struktureret?"

2.1 Metodeinddeling og litteraturreview

Projektets analyse har til formål at besvare den initierende problemstilling. Analysen består af tre dele, hvor hver del indeholder en række operationelle spørgsmål, der tilsammen besvarer den initierende problemstilling, se Figur 2.1.



Figur 2.1: Oversigt over metoden, der anvendes til besvarelse af den initierende problemstilling.

Først foretages en generel analyse af rapporteringssystemer, herunder hvilke typer der findes og hvordan systemerne er opbygget. Dernæst undersøges det hvordan det danske indrapporteringssystem er organiseret. Problemanalysen afsluttes med et review af tidligere studier, der anvender statistisk tekstklassifikation til kategorisering af utilsigtede hændelser. Slutteligt vil den endelige problemformulering opstilles. Figur 2.1 illustrerer en oversigt over metoden, som anvendes til besvarelse af den initierende problemstilling.

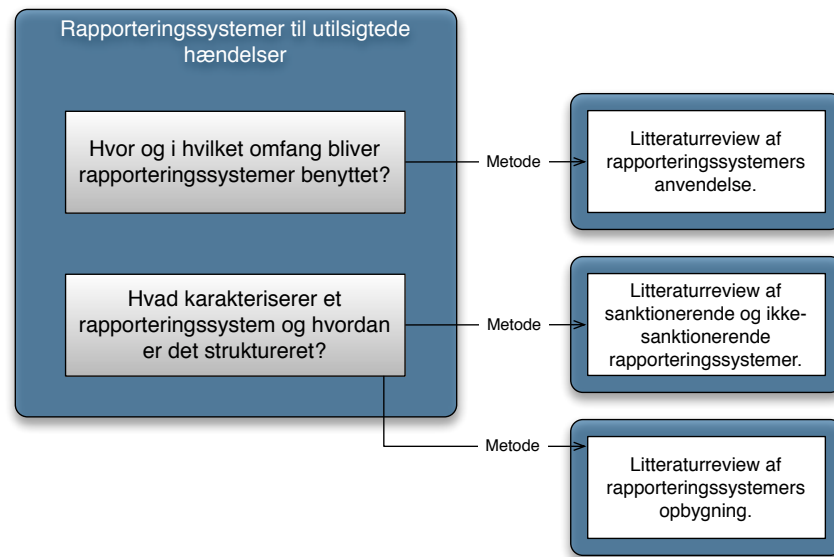
Litteraturreviews blev foretaget i forbindelse med indsamling af viden til de tre analyseafsnit. Den biomedicinske database *PubMed* blev hovedsageligt benyttet, da den udelukkende indeholder peer-reviewed artikler. Litteraturreviewet blev udført ud fra krydssøgning på søgeord i *PubMed*. Søgeordene var primært nøgleord fundet i hovedartikler indenfor emnet. *Google Scholar* blev også benyttet til litteratursøgning, og de inddragede artikler blev endvidere suppleret med *PubMeds* autogenerede liste over relaterede artikler. Dermed blev artikler inddraget, som ikke blev fundet i litteratursøgningen ud fra de valgte nøgleord. Referencetracking blev anvendt til at sikre validiteten af de inddragede artikler. Denne metode sikrede desuden at originale kilder blev anvendt. Derved kunne litteratur, der underbyggede eller udfordrede den gængse opfattelse af emnet opnås. Den egentlige litteratursøgning blev afsluttet primo marts, men fortsatte som en iterativ proces, sideløbende med udarbejdelsen af analysen og besvarelse af problemformuleringen.

2.1.1 Rapporteringssystemer til utilsigtede hændelser

Analysen har til formål at besvare to operationelle spørgsmål:

- Hvor og i hvilket omfang bliver rapporteringssystemer benyttet?
- Hvad karakteriserer et rapporteringssystem og hvordan er det struktureret?

Ved at besvare disse spørgsmål opnås en viden som er nødvendig for at forstå hvor og hvorfor et system til klassificering af utilsigtede hændelser skal implementeres og hvordan det kan understøtte processen. Derudover fokuseres der på hvordan forskellige typer af systemer er struktureret, og hvordan denne strukturering kan have indflydelse på mængden af indberetninger samt udbyttet af det indberettede data.



Figur 2.2: Oversigt over de metoder som er benyttet til besvarelse af de to operationelle spørgsmål, der tilknytter sig til kapitlet "Anvendelse af rapporteringssystemer". Analysen foretages igennem litteraturreviews af rapporteringssystemers benyttelse, sanktionerende og ikke-sanktionerende rapporteringssystemer, samt et litteraturreview af rapporteringssystemers opbygning.

De operationelle spørgsmål blev besvaret ved at foretage et litteraturreview. Ved spørgsmålet "Hvor og i hvilket omfang bliver rapporteringssystemer benyttet?", undersøges det hvor rapporteringssystemer anvendes internationalt, både indenfor sundhedssektoren og i andre brancher. Derfor er danske søgeord udeladt i litteraturreviewet. Litteraturreviewet blev gennemført som belyst i 2.1 med anvendelse af følgende søgeord:

- Incident reporting systems
- Adverse events
- Advers events
- Near miss

Disse søgeord resulterede i en bred søgning, hvilket resulterede i at mange ikke-relevante artikler blev fundet. For at indsnævre søgefeltet, blev søgeordene kombineret med "experiences" og "reporting". Med spørgsmålet "Hvad karakteriserer et rapporteringssystem og hvordan er det struktureret?" fokuseres der på den internationale anvendelse og opbygning af rapporteringssystemer i sundhedsvæsenet med enkelte sammenligninger til det danske system. Her undersøges de forskellige typer af rapporteringssystemer inden for sundhedssektoren ved et litteraturreview af sanktionerende og ikke-sanktionerende rapporteringssystem. Her er førnævnte nøgleord anvendt, samt følgende nøgleord i både den engelske og danske udgave:

- Patient safety
- Voluntary reporting
- Accountability
- Confidentiality
- Anonymity

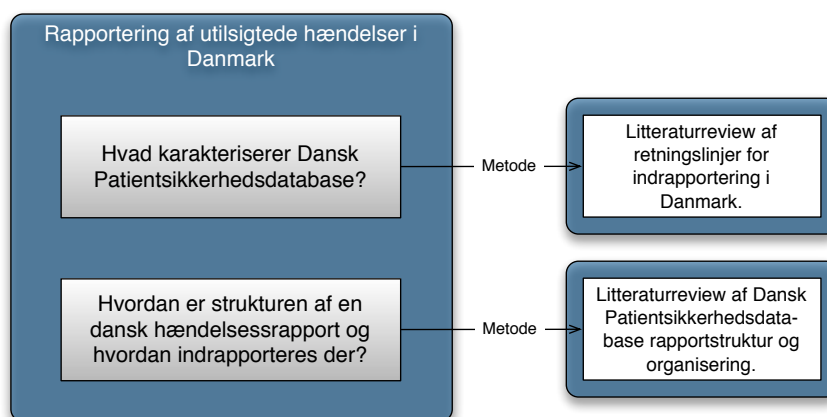
Slutteligt blev der foretaget et litteraturreview angående rapporteringssystemers opbygning. De udvalgte artikler blev gennemgået systematisk med henblik på at besvare de to operationelle spørgsmål. Analysemetoden til "Anvendelse af rapporteringssystemer" er illustreret på figur 2.2.

2.1.2 Rapportering af utilsigtede hændelser i Danmark

Analysen har til formål at besvare to operationelle spørgsmål:

- Hvad karakteriserer Dansk Patientsikkerhedsdatabase?
- Hvordan er strukturen af en dansk hændelsesrapport og hvordan indrapporteres der?

Formålet er at opnå viden om det danske indrapporteringssystem, en viden som gør det muligt at udvikle et system til klassificering af utilsigtede hændelser, der understøtter den danske patient sikkerhedsdatabase. Således tages der udgangspunkt i følgende operationelle spørgsmål:



Figur 2.3: Metode der benyttes til besvarelse af de operationelle spørgsmål i kapitlet "Rapportering af utilsigtede hændelser i Danmark". Analysen foretages med litteraturreviews af dansk oprindelse for at afdække retningslinjer, struktur og organisering i det danske rapporteringssystem.

Besvarelsen af de operationelle spørgsmål vil bidrage med viden, om hvordan databasen er organiseret, og hvordan denne indgår som en integreret del af Det Nationale Råd for kvalitetsudvikling og Sundhedsstyrelsens strategi for kvalitetsudvikling og

minimal patientrisiko. Et litteraturreview blev foretaget for at besvare de to operationelle spørgsmål. Ved spørgsmålet "Hvad karakteriserer Dansk Patientsikkerhedsdatabase?" fokuseres der på hvordan indrapportering er lovmæssigt forankret i det danske sundhedsvæsen, hvem der skal rapportere og det danske rapporteringssystems formål.

Den ønskede viden er specifik for DPSD, hvorfor informationen hovedsageligt blev fundet via relevante organisationers hjemmesider og publikationer. Derudover var kun danske udgivelser relevante i forhold til analysen af det danske indrapporteringssystem, hvilket betød at der udelukkende blev anvendt danske nøgleord til krydsøgninger. De anvendte nøgleord er listet nedenfor:

- Dansk Patientsikkerhedsdatabase
- Det danske indrapporteringssystem
- Elektroniske rapporteringssystem
- Utilsigtede hændelser
- Primærsektor
- Sekundærsektor
- Anonym

Det blev vurderet at litteratur publiceret før 2002 ikke var relevant, da det var første år, der kom fokus på udviklingen af systemer til indrapportering af utilsigtede hændelser i Danmark.

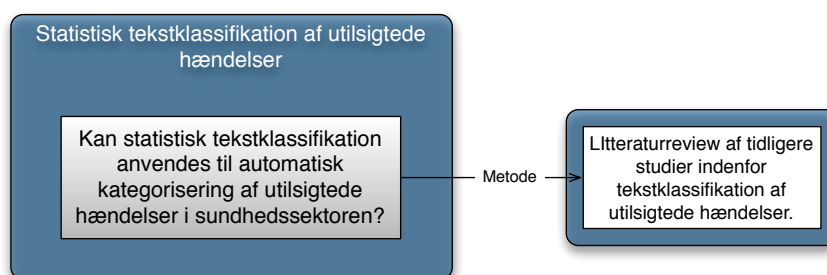
Med det operationelle spørgsmål "Hvordan er strukturen af en dansk rapporteringsrapport og hvordan indrapporteres denne?" fokuseres der på hvordan indrapporteringen er organiseret i Danmark, herunder hvordan en utilsigtet hændelse rapporteres til databasen, samt hvilke oplysninger der gemmes. Litteraturreviewet er baseret på Sundhedsstyrelsens årsrapport 2010 og DPSDs hjemmeside.

Viden opnået ved besvarelsen af de operationelle spørgsmål er essentiel og giver en forståelse af hvilke dataformater, der skal understøttes og hvilke oplysninger, der er tilgængelige. Ved at analysere hvilke personer der kan indberette i systemet opnås en viden om terminologiske sammensætning i databasen. Terminologien afhænger af rapportøren. Er det udelukkende fagpersonale der kan indberette, vil databasen hovedsageligt indeholde fagtermer. Hvis det i et indberetningssystem også er muligt for patienter og pårørende at rapportere, vil databasen også indeholde ikke fagspecifikke termer. Viden om anvendte terminologi skal anvendes ved udvikling af en tekstklassificeringsalgoritmen. Figur 2.3 illustrerer analysemetoden anvendt til kapitlet "Rapportering af utilsigtede hændelser i Danmark".

2.1.3 Statistisk tekstklassifikation af utilsigtede hændelser

Schierup et al. [2011] undersøgte potentialet for anvendelse af automatisk tekstklassifikation af utilsigtede hændelser. Afsnittet tager udgangspunkt i følgende operationelle spørgsmål:

- Kan statistisk tekstklassifikation anvendes til automatisk kategorisering af utilsigtede hændelser i sundhedssektoren?



Figur 2.4: Metode der benyttes til besvarelse af det operationelle spørgsmål i kapitlet "Statistisk tekstklassifikation af utilsigtede hændelser".

Besvarelsen af det operationelle spørgsmål blev foretaget gennem et litteraturreview og bidrager med viden om tidligere erfaringer opnået med statistisk tekstklassifikation, heriblandt om statistisk tekstklassifikation er en valid metode til at klassificere utilsigtede hændelser.

Tidligere studier af Ong et al. [2010], Ong et al. [2012] og Schierup et al. [2011] påviste potentiale for anvendelse af automatisk tekstklassifikation til kategorisering af utilsigtede hændelser i sundhedssektoren. Litteraturreviewet vil således tage udgangspunkt i de nævnte artikler. Figur 2.4 viser analysemetoden anvendt til kapitlet "Statistisk tekstklassifikation af utilsigtede hændelser".

D E L



Problemanalyse

Rapporteringssystemer til utilsigtede hændelser

Dette kapitel har til formål at besvare projektets første problemstilling: "Hvor og i hvilket omfang bliver rapporteringssystemer benyttet?" og "Hvad karakteriserer et rapporteringssystem og hvordan er det struktureret?". Derfor analyseres rapporteringssystemer til indsamling af utilsigtede hændelser. Endvidere redegøres der for det sanktionerende og ikke-sanktionerende rapporteringssystem samt en model til opbygning af et indrapporteringssystem.

3.1 Anvendelse af rapporteringssystemer i sundhedssektoren

Rapporteringssystemer til utilsigtede hændelser har været benyttet af højrisikoindustrier som kernekraft-, fly- og petroleumindustrien i mere end 30 år. Flere undersøgelser har vist, at udviklingen af disse systemer har været med til at nedbringe antallet af fejl og dermed højne sikkerheden. Endvidere er det vist at implementering af rapporteringssystemer til utilsigtede hændelser er billigere end prisen på de ulykker de forhindrer, hvilket giver systemerne en god cost benefit. [Barach og Small, 2000; Johnson, 2003]

I løbet af det sidste årti har adskillige organisationer i lande verden over etableret lokale og nationale rapporteringssystemer til sundhedssektoren. Formålet med et rapporteringssystem er at indsamle viden om systemfejl, medicineringsfejl og andre risici, som opstår under et behandlingsforløb. Sundhedsorganisationer eller individer kan med fordel benytte rapporteringssystemer i større organisationer, da det gør det muligt at modtage brugbar viden opnået ved generalisering og analysering af lignende hændelser på tværs i organisationen. [WHO, 2005]

Således opnås en øget bevidsthed om patientsikkerhed, en bedre forståelse af omfanget samt en profil af de sikkerhedsmæssige mangler. [Pham et al., 2010]

Endvidere fremhæver IOM rapporteringssystemer til utilsigtede hændelser som en nøglestrategi til en forbedret patientsikkerhed i sundhedsvæsenet. [Kohn et al., 2000]

Rapporteringssystemer kan ikke benyttes til kvantificering af omfanget af utilsigtede hændelser, men kan ses som et kvalitativt redskab. De indsamlede rapporter afspejler karakteren af de utilsigtede hændelser og kan benyttes til at identificere risikoområder og der fra udlede nye sikkerhedsinitiativer. [Bjørn et al., 2009]

3.2 Sanktionerende og ikke-sanktionerende rapporteringssystemer

Der skelnes mellem to typer af rapporteringssystemer. Rapporteringssystemer, som har til formål at holde de involverede personer ansvarlige, er typisk af obligatorisk art og sanktionerende. Er formålet derimod kvalitetssikring, vil rapporteringssystemet typisk være ikke-sanktionerende. [Kohn et al., 2000; Wu et al., 2002]

3.2.1 Det sanktionerende rapporteringssystem

I sanktionerende rapporteringssystemer er sundhedspersonalet forpligtet til at rapportere. Systemet fokuserer hovedsageligt på alvorlige fejl i sundhedssektoren og er typisk begrænset til en forhåndsdefineret liste, såsom uventet dødsfald, reaktioner som følge af transfusion eller fejloperationer. Hændelser bliver typisk efterforsket af et ministerium, som forsøger at blotlægge persondata for de indblandede og information om hændelsen. På baggrund af de indsamlede oplysninger kan den årsagsgivende faktor for fejlen findes. Dermed kan regeringen stadfæste sit ansvar som tilsynsførende for hospitalerne og sikre at befolkningen får den bedst mulige behandling. [Nagamatsu et al., 2009; WHO, 2005]

Hovedparten af de sanktionerende rapporteringssystemer straffer hospitalerne med bødestraf, sanktioner og andet, og kan forlange en øjeblikkelig forbedring. Sanktionerende rapporteringssystemer er effektive så længe at sanktionsmidlerne giver hospitalerne incitament til kvalitetsforbedringer af patientsikkerheden. Dog kan læringen i et sanktionerende system blive undermineret, hvis rapportøren frygter repressalier, som følge af indrapporteringen. [Nagamatsu et al., 2009]

3.2.2 Det ikke-sanktionerende rapporteringssystem

I ikke-sanktionerende rapporteringssystemer foretages rapporteringen typisk på frivillig basis og de indrapporterede hændelser dækker et bredere område af hændelser, sammenlignet med det sanktionerende rapporteringssystem. Ved indberetning til ikke-sanktionerende rapporteringssystemer indsendes rapporter anonymt eller fortroligt til en ekstern gruppe, der sørger for dataopsamling, analyse og feedback. Det fremmer rapporteringsvilligheden, da der ingen risiko er for, at rapportøren kan blive retsforfulgt. [Barach og Small, 2000; WHO, 2005]

Ved anonym rapportering er det ikke muligt for analytikerne at kontakte rapportøren for yderligere information, og upålidelige rapportering kan finde sted. Dog vil anonymitet kunne betyde, at tilliden til systemet hurtigere kan blive opbygget og dermed hurtigere kan føre til forbedringer på baggrund af de indsamlede rapporter. [Barach og Small, 2000]

Ikke-sanktionerende rapporteringssystemer kaldes også læringssystemer og er designet til at fremme fortsatte forbedringer i sygehusvæsenet. Det opnås ved at identificere temaer, reducere variation i forskellige behandlinger ved at fremhæve best practice og stimulere omfattende forbedringer. Ved hjælp af ekspertanalyser af de omkringliggende forhold er anbefalinger til systemforbedringer lavet for at forøge systemets ydeevne og reducere fejl og uheld. [WHO, 2005]

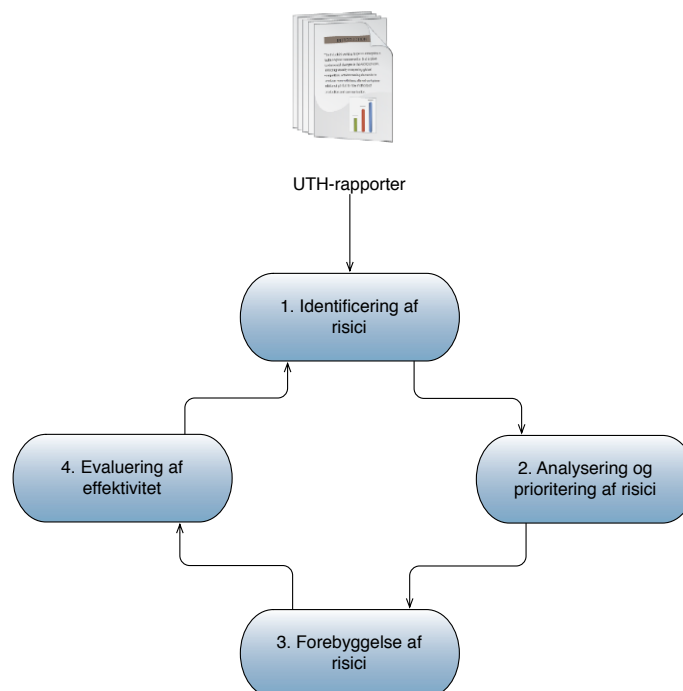
Læringsystemer har fokus på hændelser, der har forårsaget minimal eller ingen skade (near misses) med det formål at identificere og afhjælpe svagheder i systemer, der kan forårsage skade i fremtiden. Baggrunden for at opsamle utilsigtede hændelser og near misses skyldes deres mange lighedspunkter. Grundlæggende adskilles de ved at near misses bliver forhindret i at udvikle sig til en utilsigtet hændelse. [Nagamatsu et al., 2009; Wu et al., 2002]

Rapportering af near misses giver tilmed flere fordele end rapportering af utilsigtede hændelser. Utilsigtede hændelser sker relativt sjældent, mens near misses bliver rapporteret hyppigere og kan derfor analyseres kvantitativt. [Barach og Small, 2000]

Studier har vist at systemer med frivillig anonymiseret rapportering, sandsynligvis er mere effektiv end obligatorisk rapportering. Det afspejler sig også i de implementerede systemer, hvor hovedparten i dag er af frivillig karakter. [Kohn et al., 2000; Wu et al., 2002]

3.3 Opbygning af et rapporteringssystem til utilsigtede hændelser

Pham et al. [2010] har opstillet en model, der beskriver forløbet i et indrapporteringssystem. Modellen består af fire elementer der til sammen bidrager til at identificere og forstå nye sikkerhedsproblemer, se figur 3.1. I det følgende vil hvert element i modellen blive beskrevet.



Figur 3.1: Forløbet i et rapporteringssystem. Element 1: Data fra utilsigtede hændelsesrapporter indsamles i rapporteringssystemet. Element 2: Data analyseres and prioriteres. Element 3: På baggrund af analysen bliver fremtidige risici forebygget. Element 4: Effektiviteten af forebyggelsesmetoderne bliver evalueret. [Pham et al., 2010]

3.3.1 Identificering af risici

Første element af forløbet i et rapporteringssystem omhandler identificering af risici. Elementet afhænger af flere faktorer såsom hvilke slags fejl der rapporteres. Som tidligere nævnt håndterer de enkelte systemer vidt forskellige hændelser. Systemer kan være designet til udelukkende at håndtere specifikke hændelser, eksempelvis fejl, der fører til dødsfald, mens andre håndterer hændelser med alle typer udfald. [Pham et al., 2010]

En anden faktor er rapportøren, der typisk er klinikere eller administrationspersonale. Klinikere har ofte en bedre viden om den specifikke hændelse, mens administrationspersonale har et større overblik over alle hændelser. [Pham et al., 2010]

Flere systemer, herunder det danske system, har implementeret muligheden for at patienter og pårørende kan indberette utilsigtede hændelser, som de har været involveret i. Det giver et nyt perspektiv, som hidtil har været overset og kan bidrage med viden om kvaliteten på den udførte behandling. [Pham et al., 2010]

3.3.2 Analysering og prioritering af risici

Analysen foregår som en individuel gennemgang af den enkelte rapport. For at give et bedre overblik, inddeles hændelsesrapporterne i kategorier efter eksempelvis hændelsestype, medicinsk speciale og hændelsens udfald. Ved at lave et sammendrag af flere utilsigtede hændelser i en database, er det muligt at finde gennemgående tendenser. [Pham et al., 2010]

Prioriteringen af hændelserne foregår efter forskellige principper. I nogle systemer bruges en SAC-matrix score (Safety Assessment Code), hvor en hændelse bliver tildelt en score i forhold til skadens omfang og risiko for gentagelse. SAC-matrix scoren defineres afhængigt af det implementerende system. [Pham et al., 2010]

I Danmark har Dansk Selskab for Patientsikkerhed indført SAC-matrix scoren, der benyttes til at analysere en utilsigtet hændelse, se Tabel 3.1 [DSP, 2012].

Score-matrix	Katastrofal	Betydende	Moderat	Minimal/ingen
Hyppig	3	3	2	1
Mindre hyppig	3	2	1	1
Sjælden	3	2	1	1
Meget sjælden	3	2	1	1

Tabel 3.1: Denne Safety Assessment Code matrix score-tabel anvendes til vurdering af en utilsigtet hændelse i Danmark. Afhængig af skadens omfang og hyppighed tildeles en passende talværdi. Hvis en hændelse opnår en score på 1 eller 2 vil hændelsen være genstand for mindre analyser, mens hændelser med en score på 3 betegnes som alvorlig. [DSP, 2012]

Skadens omfang og hyppighed er betydende for om der bliver foretaget en kerneårsagsanalyse. En kerneårsagsanalyse er en metode, der bruges til at finde udløsende faktorer for en utilsigtet hændelse med henblik på at identificere hvilke tiltag, der skal gøres for at minimere risikoen for en gentagelse af samme skadelige hændelse. Kerneårsagsanalysen kan give værdifuld information, men er samtidig en tidskrævende

proces og egner sig bedst til systemer med en begrænset datamængde. [Pham et al., 2010]

3.3.3 Forebyggelse af risici

Der findes forskellige tiltag til forebyggelse af en utilsigtet hændelse, som er identificeret fra indrapporteringssystemet. Ofte benyttede forebyggelsesmetoder er publicering af forebyggende materialer, udvikling af tjeklister til brug ved standardopgaver, oprettelse af diskussionsfora samt udførelse af grundige analyser af utilsigtede hændelser. [Pham et al., 2010]

3.3.4 Evaluering af effektivitet

Ved evaluering af effektiviteten ved indførelse af forebyggende tiltag foretages der en kvalitativ analyse. Der evalueres over en længere periode om antallet af samme fejltype er faldet efter indførelsen. [Pham et al., 2010]

Rapportering af utilsigtede hændelser i Danmark

De initierende problemstillinger "Hvad karakteriserer Dansk Patientsikkerhedsdatabase?" og "Hvordan er strukturen af en dansk hændelsesrapport og hvordan indrapporteres der?" besvares i dette kapitel. Det danske rapporteringssystem for utilsigtede hændelser analyseres, herunder systemets opbygning, loven om patientsikkerhed og indrapporteringsforløbet.

4.1 Dansk Patientsikkerhedsdatabase

DPSD indsamler og overvåger utilsigtede hændelser i det danske sundhedsvæsen. Ved at indsamle data fra utilsigtede hændelser er det muligt at analysere og skabe en systematisk læring af utilsigtede hændelser. Rapporteringssystemer til utilsigtede hændelser er udråbt som en af nøglestrategierne til en øget patientsikkerhed. [Kohn et al., 2000]

Den 1. januar 2004 trådte lov om patientsikkerhed i kraft. Med denne lov, der senere er indarbejdet i sundhedsloven, blev Danmark det første land i verden, som indførte en national rapporteringspligt for utilsigtede hændelser i et fortroligt, ikke sanktionerende og lokalt forankret system med henblik på at opnå læring. [Bjørn et al., 2009]

I 2010 blev patientsikkerhedsloven udvidet, således ansatte i praksissektoren, apoteksektoren, det præhospitale område, samt den kommunale sundhedssektor også blev omfattet af rapporteringspligten. Dermed har alt autoriseret sundhedspersonale, ambulancebehandlere, apotekspersonale og andre der handler på disses ansvar, pligt til at rapportere utilsigtede hændelser, som nævnt i indledningen. [Sundhedsstyrelsen, 2010] Alle ovennævnte rapporteringspligtige personer skal rapportere utilsigtede hændelser som observeres, herunder både hændelser de selv er impliceret i og hændelser de observerer eller bliver gjort opmærksom på fra andre i forbindelse med udøvelsen af deres faglige gerning. [Sundhedsstyrelsen, 2010]

4.1.1 Stillingskategori af de indberettede hændelser

I 2010 lavede DPSD en opgørelse over hvilken stillingskategori indberetteren tilhørte. 56 % af de indberettede rapporter stammede fra sygeplejersker og andet plejepersonale, mens lægeligt personale indberettede 12 % af det samlede antal rapporter. Stillingskategorierne for henholdsvis andet klinisk personale og andet personale tilknyttet primær- og sekundærsektoren stod hver for 8 % af de indberettede rapporter. [Sundhedsstyrelsen, 2010]

4.2 Indrapporteringsforløbet af en utilsigtet hændelse

I den danske sundhedssektor foretages en rapportering af utilsigtede hændelser i to trin. I første trin rapporteres der til regionen, kommunen eller privathospitalet, hvor rapportøren kan vælge at rapportere anonymt eller ikke-anonymt. Hvis rapportøren vælger at indberette ikke-anonymt, vil kontaktinformationerne kun videregives til de personer i samme region, kommune eller privathospital, som behandler hændelsen. [Sundhedsstyrelsen, 2010]

I næste trin behandles den indgivne rapport af regionen, kommunen eller privathospitalet inden for en begrænset tidsperiode på 90 dage. Derefter videresendes rapporten til DPSD, hvor kontaktoplysninger og personlig data om rapportøren slettes. Endvidere er det muligt for patienter og pårørende at rapportere utilsigtede hændelser. [Sundhedsstyrelsen, 2010]

I 2010 blev 85 % af de rapporterede hændelser indberettet ikke-anonymt. Dermed var det muligt for regioner, kommuner og privathospitalerne at få yderligere information og lave opfølgninger på størstedelen af hændelserne. Baseret på hyppigheden af rapporterede hændelser, der ikke foretages anonymt, kunne det tolkes at hovedparten af de indberetningspligtige er bevidste om, at rapporteringssystemet er ikke-sanktionerende. [Sundhedsstyrelsen, 2010]

4.3 Struktur af utilsigtet hændelsesrapport

I sundhedssektoren betegnes en utilsigtet hændelsesrapport som et dokument, der dokumenterer detaljer om en hændelse, som har forårsaget skade eller kunne have haft forvoldt skade på en patient. Formålet med at indrapportere en utilsigtet hændelse er at få de nøjagtige detaljer om hændelsen, mens den fortsat er i frisk erindring hos de involverede parter. Derfor skal en hændelse indrapporteres umiddelbart efter hændelsens indtræden. [Runciman et al., 2007]

I Danmark har sundhedsprofessionelle, patienter og pårørende mulighed for at indrapportere en utilsigtet hændelse til DPSD. Indrapporteringen kan ske elektronisk via Patientombuddets hjemmeside, eller ved at der udskrives en papirblanket, som udfyldes og efterfølgende sendes til regionen hvor hændelsen fandt sted. Selve rapporteringen sker i et semistruktureret skema, idet der både anvendes strukturerede felter og felter til fritekst. [Sundhedsstyrelsen, 2010]

4.3.1 Semistrukturerede rapporteringsskemaer

DPSD anvender semistrukturerede indberetningsskemaer til indsamling af data, se Appendix A. Det semistrukturerede skema er en kombination af strukturerede datafelter og ustrukturerede fritekstfelter. Fra strukturerede felter kan der foretages et direkte dataudtræk, da data kan tolkes og indgå i en kontekst, hvor det bidrager med information idet feltet udfyldes. Således er udtræk af informationer fra strukturerede felter enkel, da der ikke kræves yderligere analyse for at det kan indgå en meningsfuld kontekst. Da det strukturerede felt angives i et prædefineret format bliver der pålagt restriktioner for rapportøren, hvilket kan minimere detaljegraden og vigtig information kan gå tabt. [Johnson et al., 2008]

Det ustrukturerede fritekstfelt giver rapportøren formulerings- og detaljefrihed og tillader derved en hurtig måde at dokumentere en hændelse. Fritekstfeltet kan derfor indeholde vigtig information, som går tabt i den overfladiske detaljegrad fra det strukturerede skema. Dette sker dog på bekostning af muligheden for udtræk af data, som besværliggøres ved at hvert enkelt ord ikke er informationsgivende, men skal indgå i en sammenhæng for at danne en forståelse og gøre dataudtræk mulig [Johnson et al., 2008]. For at udtrække informationer fra fritestfelter er en analyse nødvendigt. Studier har vist at en manuel gennemgang af fritest er en tidskrævende proces og kræver anonymisering for at beskytte patienten og indberetterens fortrolighed. Derfor er der kommet fokus på systemer til automatisk analysering af fritest med henblik på kategorisering af utilsigtede hændelser. [Wang et al., 2012]

Statistisk tekstklassifikation af utilsigtede hændelser

Dette kapitel har til formål at foreligge evidens for potentialet af automatisk kategorisering ved anvendelse af machine learning algoritmer, samt brugen af disse til at klassificere utilsigtede hændelser i sundhedssektoren. Det gøres ved at svare på spørgsmålet "Kan statistisk tekstklassifikation anvendes til automatisk kategorisering af utilsigtede hændelser i sundhedssektoren?"

5.1 Anvendelse af machine learning algoritmer til tekstklassifikation

Erfaringerne med anvendelse af statistisk tekstklassifikation til automatisk kategorisering af utilsigtede hændelser i sundhedssektoren er fortsat begrænsede. Flere studier har imidlertid vist potentiale for anvendelse af metoden. Et studie af Ong et al. [2010] undersøgte potentialet for anvendelse af to machine learning algoritmer, Naive Bayes og Support Vector Machine (SVM), til kategorisering af utilsigtede hændelser. Studiet anvendte data fra det australske Advanced Incident Management System (AIMS) relateret til utilstrækkelig klinisk overdragelse og misidentifikation af patienter.

Naive Bayes havde en gennemsnitlig nøjagtighed på 88,91 % ved klassificering af utilsigtede hændelser relateret til kliniske overdragelser, mens SVM havde en gennemsnitlig nøjagtighed på 86,69 % ved klassificering af samme type hændelse. For utilsigtede hændelser, som skyldtes misidentifikation af patienter, blev der opnået en gennemsnitlig nøjagtighed på 96,37 % for Naive Bayes, mens SVM havde en gennemsnitlig nøjagtighed på 94,56 %, se Tabel 5.1. [Ong et al., 2010]

Algoritme	Utilstrækkelig klinisk overdragelse (%)	Misidentifikation af patienter (%)
Naive Bayes	88,9	96,4
Support Vector Machine	86,7	94,6

Tabel 5.1: Den opnåede nøjagtighed fra et studie af Ong et al. [2010], hvor machine learning algoritmerne Naive Bayes og Support Vector Machine blev testet som tekstklassifikationsalgoritmer på AIMS-data med fokus på utilstrækkelig klinisk overdragelse og misidentifikation af patienter.

Et andet studie af Ong et al. [2012] undersøgte potentialet for anvendelse af statistisk tekstklassifikation til automatisk klassificering af ekstrem-risiko hændelser, som er en hændelse der er tildelt SAC-scoren 1 i det australske AIMS-system. I dette studie blev der ligeledes anvendt Naive Bayes og SVM til klassificeringen.

Resultaterne viste at Naive Bayes havde en nøjagtighed på 80 %, mens SVM opnåede en nøjagtighed på 86,7 %, se Tabel 5.2. [Ong et al., 2012]

Algoritme	Ekstrem-risiko hændelser (%)
Naive Bayes	80,0
Support Vector Machine	86,7

Tabel 5.2: Resultaterne viser den opnåede nøjagtighed ved anvendelse af Naive Bayes og Support Vector Machine til automatisk klassificering af ekstrem-risiko hændelser i AIMS-systemet. [Ong et al., 2012]

Forud for dette projekt blev der udviklet et system til automatisk kategorisering af utilsigtede hændelser relateret til IT i sundhedssektoren. Kategoriseringen blev foretaget på data fra AIMS-databasen og FDAs Manufacturer and User Facility Device Experience (MAUDE) database, hvor der blev anvendt Naive Bayes og SVM til kategoriseringen. For AIMS-data blev der opnået en nøjagtighed på 92 % og 94 % for henholdsvis Naive Bayes og SVM. Ved kategorisering af MAUDE-data var nøjagtigheden 95 % for Naive Bayes og 93 % for SVM. [Schierup et al., 2011]

Algoritme	IT-hændelser fra AIMS-database (%)	IT-hændelser fra MAUDE-database (%)
Naive Bayes	92,0	95,0
Support Vector Machine	94,0	93,0

Tabel 5.3: Nøjagtigheden opnået ved kategorisering af utilsigtede hændelser relateret til IT i AIMS- og MAUDE-databaserne. [Schierup et al., 2011]

I alle tre studier konkluderes det at anvendelse af statistisk tekstklassifikation har potentiale som værende en valid metode til klassificering af utilsigtede hændelser. Statistisk tekstklassifikation vil fungere som en forbedring af den nuværende manuelle tilgang. Derudover kan det konkluderes, at metoden kan anvendes til klassificering af andre typer hændelser og indrapporteringssystemer. [Ong et al., 2010, 2012; Schierup et al., 2011]

Problemformulering

Siden 2004, hvor loven om patientsikkerhed trådte i kraft, har antallet af indberettede utilsigtede hændelser været stigende. Således er antallet af indberettede utilsigtede hændelser i perioden 2004-2010 blevet tidoblet. Med udvidelsen af patientsikkerhedsordningen i 2010 forventes det at antallet af indrapporterede utilsigtede hændelser vil stige yderligere, da indberetningspligten udvides til også at omfatte praksissektoren, den kommunale sundhedssektor, apotekerne og det præhospitale område. Dertil har patienter og pårørende ligeledes mulighed for at indrapportere.

På baggrund af erfaringer i andre højrisikoindustrier, hvor indrapportering har vist sig brugbart, er indrapportering af utilsigtede hændelser i den danske sundhedssektor blevet implementeret og sat i system. En utilsigtet hændelse indrapporteres på papirform eller elektronisk via DPSDs portal. Den efterfølgende analyse foregår ved en manuel, retrospektiv gennemgang af rapporterne. Denne metode er en ressourcekrævende proces og da rapportering ikke foretages af eksperter inden for klassifikation, bliver hændelser ofte klassificeret forkert. Dette indikerer, at der er behov for andre og mere effektive analysemetoder for at imødekomme den stigende datamængde.

Flere studier har konkluderet at anvendelse af machine learning til automatisk tekstklassifikation af utilsigtede hændelser i sundhedssektoren har potentiale som et alternativ til den manuelle retrospektive analyse. Studierne har kun anvendt data fra AIMS- og MAUDE-databaserne med henblik på klassificering af forskellige typer utilsigtede hændelser. Dermed er der endnu ikke udført studier baseret på klassifikation af utilsigtede hændelser fra DPSD.

Denne rapport fokuserer på klassifikation af utilsigtede hændelser, der involverer brugen af IT i den danske sundhedssektor. Dermed kan der opnås en bedre forståelse af hvordan utilsigtede hændelser opstår. Formålet er udvikling af et brugervenligt system, som ved brug af statistisk tekstklassifikation automatisk kan klassificere IT-relaterede hændelser.

På baggrund af analysen opstilles følgende problemformulering:

”Hvordan kan et system udvikles med henblik på at klassificere IT-relaterede utilsigtede hændelser fra Dansk Patientsikkerhedsdatabase?”

D E L



Problemløsning

Systemudvikling

I dette kapitel beskrives udvikling af systemet til klassificering af utilsigtede hændelser. Til dokumentation af udviklingsforløbet anvendes Unified Process og Unified Modeling Language som værktøj til at beskrive systemets funktionalitet, arkitektur og design. Tests relateret til Unified Process findes i Appendiks B.

7.1 Metode til systemudvikling

Formålet med projektet er at udvikle et machine learning system, der understøtter automatisk klassifikation af utilsigtede hændelser fra DPSD. Udviklingen af systemet dokumenteres med udgangspunkt i metoderne Unified Process (UP) og Unified Modelling Language (UML). UP er en iterativ metode, der har til formål at beskrive systemets funktioner, mens anvendelse af UML gør det muligt at skabe overblik over systemet ved at specificere delene og de indbyrdes relationer. Målet med dokumentationsprocessen er at støtte programmøren i udviklingen af en software applikation ved at beskrive systemets overordnede struktur og funktionaliteter. De workflows, som giver programmøren et generelt overblik over systemudviklingen, er krav og analyse. Fra kravworkflowet er det valgt at beskrive systemkrav, use case-diagram og use case-specifikationer, mens aktivitetsdiagrammer og analyseklasser beskrives i analyseworkflowet. Dokumentation af de udførte tests relateret til UP findes i Appendiks B.

7.2 Kravworkflow

7.2.1 Systembeskrivelse

Formålet med systemet er at kunne udføre automatisk klassifikation af utilsigtede hændelser relateret til IT. Systemet skal derfor kunne klassificere hændelser i to kategorier, IT og andet. Programmet skal designes til forskere og analytikere, som monitorerer indrapporterede hændelser med henblik på at identificere hvilke type hændelser, der er relateret til IT, samt at finde tendenser heri, med henblik på at forbedre patientsikkerheden.

Systemet skal kunne håndtere filer i .csv formatet, da DPSD bruger denne standard til at lagre de indberettede rapporter. Tekstklassifikationsalgoritmen Naive Bayes anvendes til klassifikation. Denne algoritme anvender supervised learning og systemet skal derfor kunne trænes ved at indlæse en model for de to klassifikationskategorier, også kaldet en træningsfil. Endvidere skal det være muligt at klassificere ukendte utilsigtede hændelser ved at indlæse en testfil.

For at gøre systemet anvendeligt og skabe værdi for slutbrugeren skal systemet kunne vise et preview af de hændelser, der er blevet klassificeret som værende relateret til IT. Derudover skal det være muligt at gemme resultatet i en tekstfil, således yderligere analysering kan foretages i eksterne programmer.

7.2.2 Kravspecifikation

På baggrund af systembeskrivelsen er der opstillet en kravspecifikation, som danner systemets rammer. Kravene er opdelt i funktionelle og ikke-funktionelle krav. De funktionelle krav er en direkte beskrivelse af systemets funktionaliteter. De ikke-funktionelle krav beskriver de rammer og begrænsninger som systemet skal udvikles indenfor.

Funktionelle krav

- Systemet skal kunne håndtere utilsigtede hændelser fra DPSD
- Systemet skal kunne håndtere .csv-filer
- Systemet skal kunne klassificere utilsigtede hændelser i to kategorier, IT og andet
- Systemet skal kunne foretage en automatisk klassifikation ved anvendelse af Naive Bayes tekstklassifikation algoritme
- Systemet skal kunne give et output i form af en tekstfil, indeholdende hændelser klassificeret som IT-relaterede fejl
- Systemet skal indeholde en intuitiv brugergrænseflade

Ikke-funktionelle krav

- Systemet skal designes til forskere og analytikere regionerne
- Systemets sprog skal være dansk
- Systemet skal være brugervenligt
- Systemet skal være platformuafhængigt

7.2.3 Use case-diagram

Et use case diagram illustrerer kommunikationsforholdene i et system. Kommunikation kan opstå mellem en aktør og en use case eller use cases imellem. Use case-diagrammet viser systemets primære elementer og processer, samt hvordan aktøren vil have systemet til at agere. En use case er defineret som:

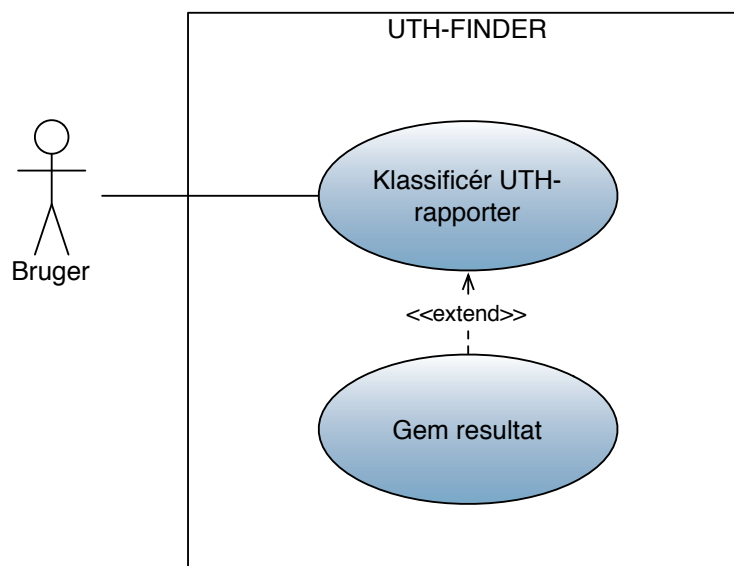
- En afsluttet proces
- Værdiskabende
- Igangsat af en aktør

Da en use case er igangsat af en aktør, er den skrevet ud fra dennes synspunkt. Derfor er funktioner, der ikke inddrager aktøren, ikke beskrevet i et use case diagram.

Use case diagrammet for det samlede system er illustreret på figur 7.1. Diagrammet viser at det er muligt for aktøren, der interagerer med systemet, at klassificere en utilsigtet hændelse, hvorefter det er muligt at gemme resultatet.

For at kunne imødekomme kravspecifikationen er følgende use cases defineret:

- ID 1: Klassificér UTH-rapporter
- ID 2: Gem resultat



Figur 7.1: Systemet illustreret ved et use case-diagram.

Aktørbeskrivelse

Brugeraktøren repræsenterer en forsker eller analytiker, der arbejder med analysering af utilsigtede hændelser. Typisk har brugeren en sundhedsfaglig baggrund og er ansat som sagsbehandler ved enten regionerne eller Patientombuddet.

7.2.4 Use case specifikationer

ID 1: Klassificér UTH-rapporter

Denne use case tillader brugeren at klassificere utilsigtede hændelser fra DPSD. Klassifikationen sker ved at brugeren indlæser en model for de to kategorier i en træningsfil. Hændelser som skal klassificeres indlæses fra en tekstfil. Efter klassifikationen bliver resultatet visualiseret for brugeren.

Use Case: Klassificér UTH-rapporter
ID: 1
Aktører: Bruger
Beskrivelse: Brugeren ønsker at klassificere utilsigtede hændelser relateret til IT og få visualiseret resultatet af klassifikationen.
Pre-betingelser: Brugeren har indlæst en modelfil og en testfil.
Hovedforløb: 1. Brugeren indlæser en modelfil. 2. Brugeren indlæser en testfil. 3. Brugeren vælger en a priori sandsynlighed svarende til den virkelige fordeling. 4. Brugeren trykker "Test". 5. Resultatet visualiseres på brugergrænsefladen. <i>extend:</i> Gem resultat.
Alternative forløb: 1. Brugeren har ikke indlæst korrekt model- eller testfil. Testen kan ikke udføres før korrekte filer er indlæst.
Post-betingelser: UTH er klassificeret og resultatet visualiseres.

Table 7.1: Use case-specifikation for "Klassificér UTH-rapporter".

ID 2: Gem resultat

Denne use case gør det muligt for brugeren, at gemme en tekstfil indeholdende de klassificerede hændelser med relation til fejl på IT udstyr. Således kan der foretages yderligere analyse i eksterne programmer.

Use Case: Gem resultat
ID: 2
Aktører: Bruger
Beskrivelse: Brugeren ønsker at gemme resultatet af klassifikationen.
Pre-betingelser: Klassifikationen er udført.
Hovedforløb: <i>extend:</i> Klassificér UTH-rapporter 1. Brugeren ønsker at gemme resultatet. 2. Brugeren vælger sti og filnavn.
Alternative forløb: 1. Klassifikationen er ikke udført og resultatet kan ikke gemmes.
Post-betingelser: Resultatet er gemt i en tekstfil.

Tabel 7.2: Use case-specifikation for "Gem resultat".

7.3 Analyseworkflow

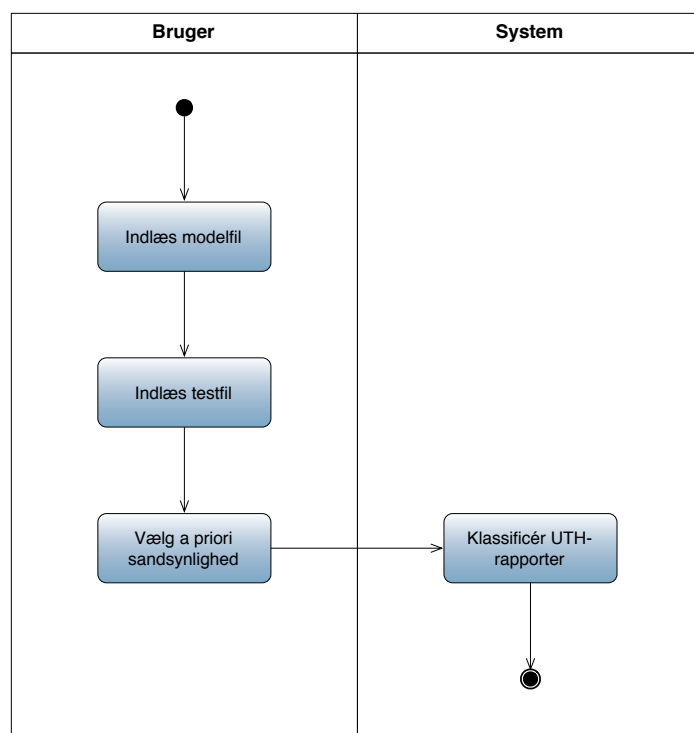
7.3.1 Aktivitetsdiagrammer

Aktivitetsdiagrammer modellerer en proces som en aktivitet og beskriver systemets flow, herunder hvor og hvornår systemet skal udføre en opgave. Diagrammet afbilder flowet af aktiviteter i en proces ved hjælp af handlinger i systemet. Dette kan enten ske mellem bruger og systemet eller internt i systemet.

Det følgende visualiserer aktivitetsdiagrammer for de implementerede use cases og giver overblik over systemets flow. Aktivitetsdiagrammerne er opstillet i *swim lanes*, som afgrænser en aktivitet til hvor i systemet den udføres. Således er brugerrelaterede interaktioner angivet under *Bruger* og systemrelaterede aktiviteter placeret under *System*. Formålet er give et overblik over hvilke opgaver der foregår på systemniveau og hvilke opgaver der inkluderer brugerinteraktion.

ID 1: Klassificér UTH-rapporter

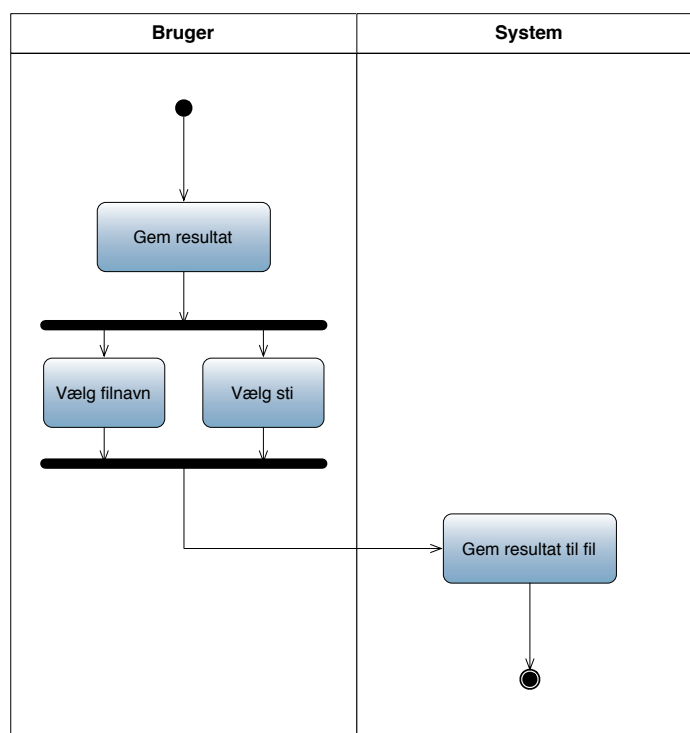
Aktiviteten "Klassificér UTH-rapporter" starter når brugeren åbner systemet. Brugeren træner systemet ved at indlæse en model. Efterfølgende indlæses en testfil, der angives en a priori sandsynlighed og klassifikationen startes. Aktiviteten afsluttes idet klassifikationen er fuldført og resultatet er visualiseret.



Figur 7.2: Aktivitetsdiagram for use casen "Klassificér UTH-rapporter".

ID 2: Gem resultat

Aktiviteten "Gem resultat" kan initieres når klassifikationen er foretaget. Det sker hvis brugeren ønsker at gemme resultatet. Aktiviteten kræver at brugeren indtaster et filnavn, samt vælger hvor filen skal gemmes. Når resultatet er gemt i en fil afsluttes aktiviteten.



Figur 7.3: Aktivitetsdiagram for use casen "Gem resultat".

7.3.2 Analyseklasser

Formålet med analyseklasser er at give systemudvikleren et overblik over de vigtigste klasser, deres metoder og attributter. I modsætning til klasser på designniveau indeholder analyseklasserne kun metoder og attributter, som kan give det nødvendige overblik. Af hensyn til udviklingsværktøjets syntaks er analyseklasserne beskrevet og udviklet på engelsk. Systemet består af følgende tre analyseklasser:

- GUI
- trainClassifier
- classifyUTH

Analyseklasse: GUI

Analyseklassen "GUI" dækker over den grafiske brugergrænseflade, se Tabel 7.3. Klassen indeholder informationer om de grafiske komponenter, herunder knapper, rulpaneler og tekstfelter. Samtidig sørger GUI for at starte klassifikationen ved at kal-

de metoderne *runTraining()* og *runTest()*, og visualisere resultatet ved at kalde metoden *showResult()*. Derudover indeholder klassen hjælpemetoder som *saveResult()*, som kan kaldes fra andre klasser.

«Boundary» GUI
trainingFilePath testFilePath aprioriValue result outputPath fileName
runTrainig() runTest() showResult() saveResult() readResult()

Tabel 7.3: Analyseklasse "GUI".

Analyseklasse: trainClassifier

Analyseklassen "trainClassifier" står for træningen af systemet, se Tabel 7.4. Systemet trænes ved at kalde metoden, *trainClassifier()*, der endvidere kalder hjælpemetoden *readTrainingFile()*, som læser et træningsfilen.

«Control» trainClassifier
trainingFilePath trainingAttributtes attributteProbability
readTrainingFile() calculateAttributProbability() trainClassifier()

Tabel 7.4: Analyseklasse "trainClassifier".

Analyseklasse: classifyUTH

Analyseklassen "classifyUTH" er den klasse, der indeholder metoderne som bruges til at klassificere en UTH. Klassifikationen foretages ved at kalde metoderne *calculateITProbability()*, *calculateOtherprobability()* og *compareProbability()*. Derudover kaldes hjælpemetoderne *readTestFile()*, som læser testsættet og *readAttributeProbability()*, der henter modellens attributter som bruges ved klassifikationen. Analyseklassen ses i Tabel 7.5.

«Control» classifyUTH
testFilePath aprioriValue testAttributes attributeProbability ITProbability otherProbability result
readTestFile() readAttributProbability() calculateITProbability() calculateOtherProbability() compareProbabilities()

Tabel 7.5: Analyseklasse "classifyUTH".

7.4 Implementering

7.4.1 Implementering af Naive Bayes algoritme

Den implementerede algoritme bygger på Naive Bayes tekstklassifikationsalgoritmen, som er yderligere forklaret i appendiks C.

Træning af systemet

Til træning af systemet blev "bag-of-word" princippet benyttet, hvilket betyder at der for hver klasse oprettes en liste af ord, som er indeholdt i klassens rapporter også kaldet en "bag". I det implementerede program oprettes to ordlister svarende til de to klasser, IT og andet. For at et ord kan optages i ordlisten og dermed indgå i klassifikationen, skal ordet bestå af minimum tre bogstaver. Dette kriterium er fastsat for at undgå fyldord, som ikke direkte kan relateres til den enkelt klasse og derfor blot vil optræde som "støj" i klassifikationen. Hvert af de inkluderede parametre i de to ordlister tillægges en værdi, som beskriver sandsynligheden for at det pågældende ord optræder i en rapport tilhørende en bestemt klasse. Sandsynligheden er ikke afhængig af ordfrekvensen, men af antallet af rapporter hvori ordet optræder. Eksempelvis beregnes sandsynligheden for at ordet "EPM" i en rapport tilhørende klasse C_1 på følgende formel:

$$P(\text{"epm"}) \text{ optræder i } C_1 = \frac{\# \text{ dokumenter i } C_1 \text{ indeholdende "epm"}}{\text{Total \# dokumenter i } C_1} \quad (7.1)$$

Hvis et ord kun optræder i den ene klasse udføres laplace smoothing, hvilket betyder, at ordet bliver tilføjet den anden klasse med en sandsynlighed svarende til, at ordet kun optræder i en enkel rapport. Trænings sættet er fuldent, når begge ordlister har fået beregnet sandsynligheder for samtlige parametre.

Klassifikation af testsæt med systemet

Ved dannelse af et fuldentt træningssæt er systemet klar til at klassificere et testsæt. Hver rapport i testsættet klassificeres som tilhørende den klasse, hvor følgende argument maksimeres:

$$v = \max_{v_j \in V} P(v_j) \prod P(a_i | v_j) \quad (7.2)$$

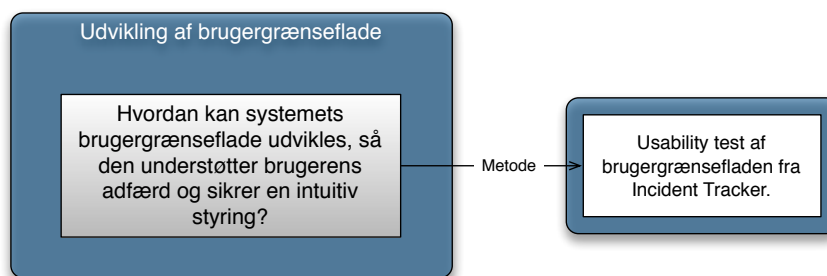
I praksis udføres denne klassifikation af rapporten ved at oprette en ordliste med samtlige ord, der opfylder de tidligere nævnte udelukkelseskriterier. Hver parameter i ordlisten sammenlignes med de to klassers ordlister dannet i træningssættet og sandsynligheden for parameteren noteres for begge klasser. Hvis en parameter fra testsættet ikke findes i træningssættet udelukkes denne parameter. Til slut beregnes den samlede sandsynlighed for at rapporten tilhører de respektive klasser efter følgende formel:

$$P(a_1, a_2, \dots, a_n | v_j) = \prod P(a_i | v_j) \quad (7.3)$$

Den klasse der maksimerer den samlede sandsynlighed vælges som klassen, hvor den testede rapport tilhører.

7.4.2 Design af brugergrænseflade

Til udvikling af en brugergrænseflade til systemet er der taget udgangspunkt i projektet af Schierup et al. [2011], hvor der blev udviklet et system kaldet Incident Tracker til præprocessering af data fra AIMS- og MAUDE-databaserne. Klassifikationen blev efterfølgende foretaget i klassifikationsprogrammet Weka. Til udvikling af brugergrænsefladen er erfaringer og forbedringsforslag fra Incident Tracker opsamlet ved at udføre en usability test, se Figur 7.4.



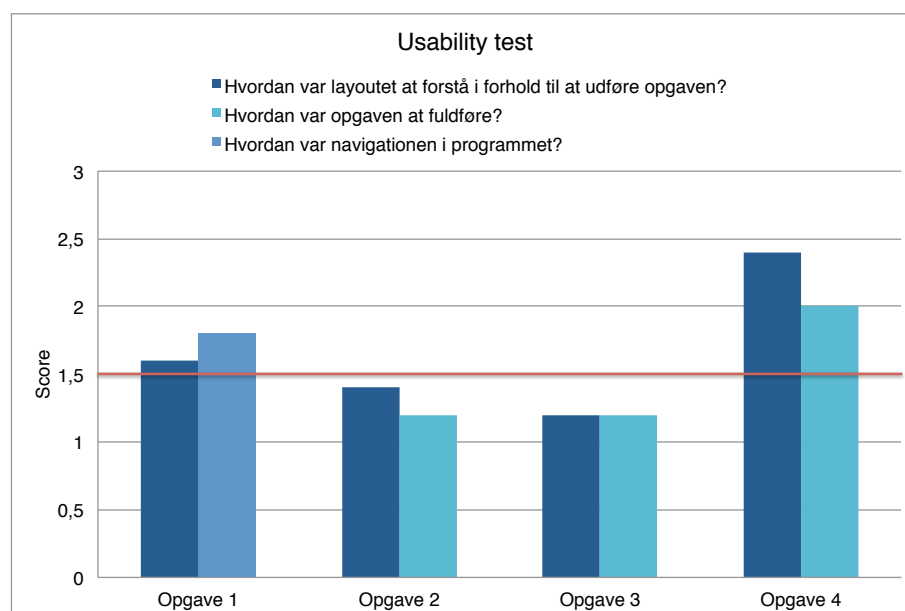
Figur 7.4: For at forbedre brugervenligheden foretages en usability test af brugergrænsefladen af "Incident Tracker" udviklet af Schierup et al. [2011].

Usability test er en metode, der bliver brugt, til at evaluere interaktionsdesign ved at teste brugergrænsefladen på en testsperson. Metoden giver således direkte input til hvordan testspersonen anvender systemet og er derfor et mål for brugbarheden og brugervenligheden af det specifikke system. [Sharp et al., 2002]

På baggrund af usability testen vil der blive udviklet en brugervenlig grænseflade til systemet. Forsøgsprotokol og resultater fra usability testen er beskrevet i Appendiks D.

Sammendrag af usability test

I usability testen blev testpersonerne bedt om at fuldføre en række opgaver i Incident Tracker relateret til klassifikationen af utilsigtede hændelser. Efterfølgende blev testpersonerne bedt om at tildele sværhedsgraden af den stillede opgave en score mellem 1 og 3, hvor 1=let, 2=middel, 3=svær, samt * hvis testpersonen havde brug for assistance til løsning af opgaven. Hvis den gennemsnitlige score for en opgave oversteg 1,5 skulle der foretages ændringer i systemet, således det imødekommer brugernes behov. For sikre detaljerede input fra brugeren blev testpersonerne bedt om give en forklarende tekst i forbindelse med den tildelte score. På figur 7.5 ses den gennemsnitlige score for de stillede opgaver.



Figur 7.5: Gennemsnitlige score for de stillede opgaver. Opgave 1: Åbn UTH-Finder. Opgave 2: Indlæs træningssæt og testsæt. Opgave 3: Vælg a priori sandsynlighed. Opgave 4: Test og gem resultatet. Den røde linje ved scoren 1,5 illustrerer threshold for, hvornår ændringer i systemet skal foretages.

Figur 7.5 giver overblik over hvilke opgaver, som kræver særlig opmærksomhed på forbedringer, der imødekommer brugerens behov.

Specielt opgaven med klassifikation af de utilsigtede hændelser i Weka var vanskelig for testpersonerne at fuldføre. Der var en tendens blandt deltagerne til at miste overblikket, da løsningen af en enkelt opgave krævede anvendelse to forskellige programmer. Derudover kommenterede flere testpersoner, at det var vanskeligt at anvende Weka, da brugergrænsefladen var kompleks og det indeholdte for mange funktioner, som ikke blev brugt. Flere deltagere kommenterede, at det var vanskeligt at navigere i programmet. Det var svært for brugeren at vide hvor i programmets flow brugeren befandt sig.

Flere testpersoner kommenterede, at det var svært at navigere i programmet, da det var svært at følge hvor i et forløb brugeren befandt sig på et givent tidspunkt.

Enkelte testpersoner kommenterede, at mange steps skulle gennemføres i Incident Tracker før klassifikationen var fuldført, hvilket med fordel kunne nedbringes til færre trin. Ligeledes fandt flere testpersoner indlæsningen af de nødvendige filer til klassifikationen besværlig. Det var vanskeligt for testpersonerne at identificere de korrekte filer, da flere filer skulle sammensættes til en enkelt fil før klassifikationen kunne igangsættes. Testpersonerne efterspurgte en funktion til automatisk indlæsning af de relevante filer, således at kun første fil blev indlæst af brugeren og resten blev indlæst automatisk af systemet.

Generelt var der blandt brugerne enighed om at layoutet på brugergrænsefladen var enkelt og overskueligt. Desuden var testpersonerne ikke i tvivl om hvordan en opgave skulle udføres, da anvendelsen af "tooltips" og tekst på knapperne understøttede opgavens udførelse. Desuden pointerede testpersonerne at brugen af feedback fra systemet var en god feature. Eksempelvis blev en succesfuld afsluttet proces markeret med et grønt "tick". Således kunne brugeren altid se om en opgave var fuldført korrekt.

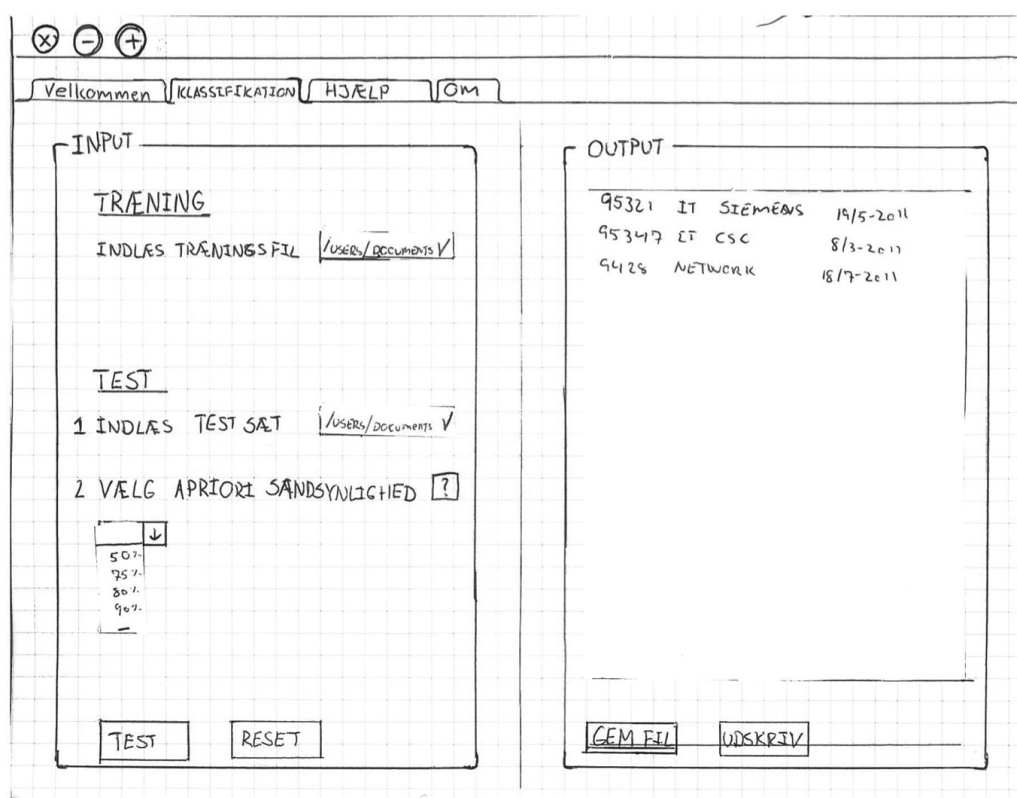
Problem	Prioritering	Problem	Løsningforslag
1	Høj	Klassificering med Weka var svært, grundet utallige muligheder i Wekas brugergrænseflade	Inkorporer tekstklassificeringsfunktioner i programmet istedet.
2	Høj	Uoverskueligt at benytte to programmer.	Inkorporer tekstklassificeringsfunktioner i programmet istedet.
3	Høj	Brugeren af programmet føler sig ikke tilstrækkelig guidet i programmet.	Ændre layoutet, gør hjælpetekster mere tydelige, samt lav en udførlig skriftlig og videovejledning til programmet.
4	Høj	Der er for mange trin brugeren skal igennem før der er dannet en fil, der kan bruges til den videre klassificering.	Nedbring antallet af trin ved at benytte bla. gem-dialogbokse.
5	Mellem	Brugeren ved ikke altid hvilken fil der skal findes i vælg-dialogboksene	Indskriv i dialogboksen, hvilken fil brugeren skal søge efter.

Tabel 7.6: Prioriteret oversigt over indsatsområdet til design af brugergrænsefladen for det udviklede system.

På baggrund af usability testens spørgeskema, hvor testpersoner blev bedt om at tildele en opgaves sværhedsgrad med et antal point, blev der udvalgt fem indsatsområder i prioriteret rækkefølge til forbedring af brugergrænsefladen, således designet kan imødekomme brugerens oplevelse af systemet. Tabel 7.6 viser de fem indsatsområder samt deres respektive løsningsforslag:

Design af prototype

På baggrund af erfaringerne fra usability testen blev der fremstillet en Lo-Fi prototype, se Figur 7.6. Prototypen blev designet efter følgende principper:



Figur 7.6: Den udviklede prototype af brugergrænsefladen udviklet som en papirmodel.

Klarhed:

Brugergrænsefladen skal være overskuelig og intuitiv. Dette gøres ved at lade brugergrænsefladens flow følge den naturlige læseretning. Således er det element som aktiveres først placeret øverst på siden, hvorefter næste element i flowet er placeret umiddelbart under. Derudover er det kun de væsentlige elementer, som er placeret på grænsefladen, mens yderligere funktioner er skjult som dialogbokse, som aktiveres når det er nødvendigt.

Genkendelighed:

For øge genkendeligheden, har nøgleordet været konsistens. Dette gælder for komponenter på brugergrænsefladen. Feks. er komponenter med samme funktion altid ens. Således ligner alle knapper hinanden, hvor dens funktion er beskrevet i tekst på selve

knappen. Positiv feedback er markeret med grønt, mens fejl er markeret med rødt.

Feedback:

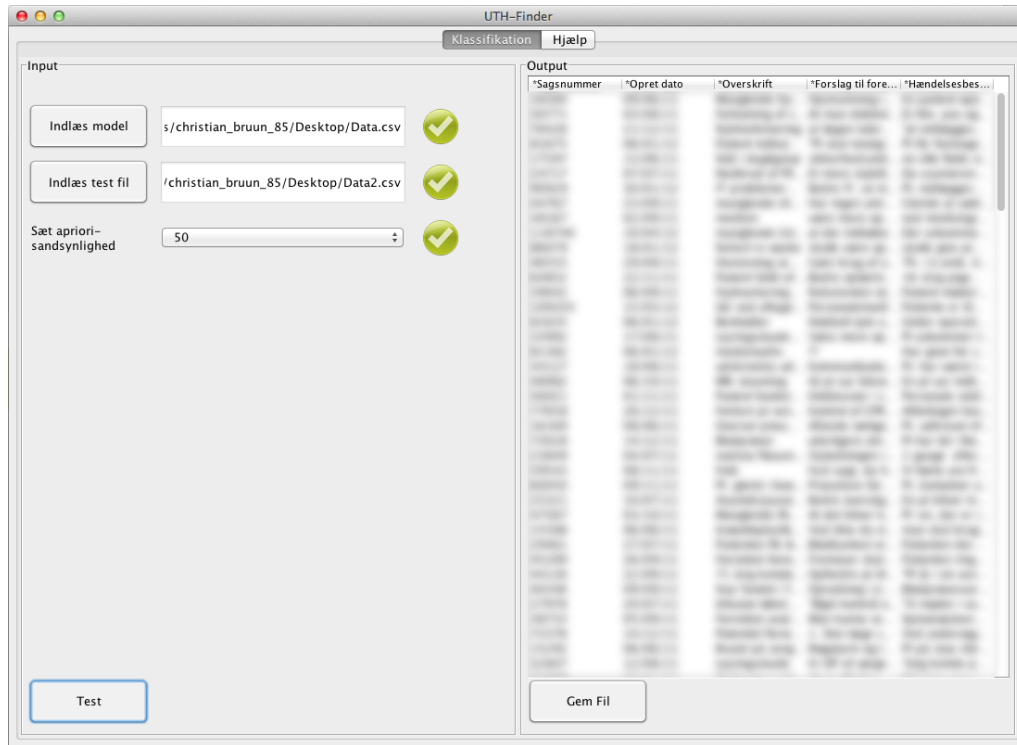
Systemet skal give feedback til brugeren. Ethvert succesfuldt afsluttet trin markeres med et grønt tick, som giver brugeren en indikation på at et trin er gennemført. Derudover skal kun elementer som knapper, der er involveret i det flow være aktive, således brugeren ikke kan springe obligatoriske processer over. Når en proces er succesfuldt afsluttet gøres den næste aktiv.

Brugervenlighed:

Systemet skal være brugervenligt. For at sikre brugervenligheden må der ikke være tvivl om, hvordan programmet skal bruges. Dette forhindres ved hjælp af intuitive komponenter som har navne, der klart beskriver dens funktionalitet. For at evaluere anvendeligheden af designprincipperne og modtage de sidste forbedringsforslag blev forsøgspersonerne bedt om at udføre samme opgave på prototypen. Denne test gav ikke anledning til yderligere ændringer. Med udgangspunkt i prototypen blev den endelige brugergrænseflade implementeret.

7.4.3 Implementering af brugergrænseflade

Figur 7.7 viser den implementerede brugergrænseflade. Dette vindue indeholder alle funktionerne til klassifikationen. Modelfilen indlæses ved at trykke på knappen *Indlæs model*, hvorefter der åbnes en stifinder, hvor modelfilen kan vælges. Testfilen indlæses ved at trykke på knappen *Indlæs testfil* og igen åbnes en stifinder, hvor test filen kan vælges. A priori sandsynlighed kan vælges i en dropdown menu, som standard er sat til 50 %. Klassifikationen startes ved at trykke på *Test*. Resultater af klassifikationen vises i tekstfeltet til højre. Ved at trykke på *Gem fil* åbnes en stifinder, hvor filnavn og sti kan specificeres hvorefter filen gemmes. Fanebladene *Hjælp* og *Om* indeholder henholdsvis hjælp til programmet og informationer om programmet.



Figur 7.7: Den implementerede brugergrænseflade til systemet.

7.5 Opsamling på systemudvikling

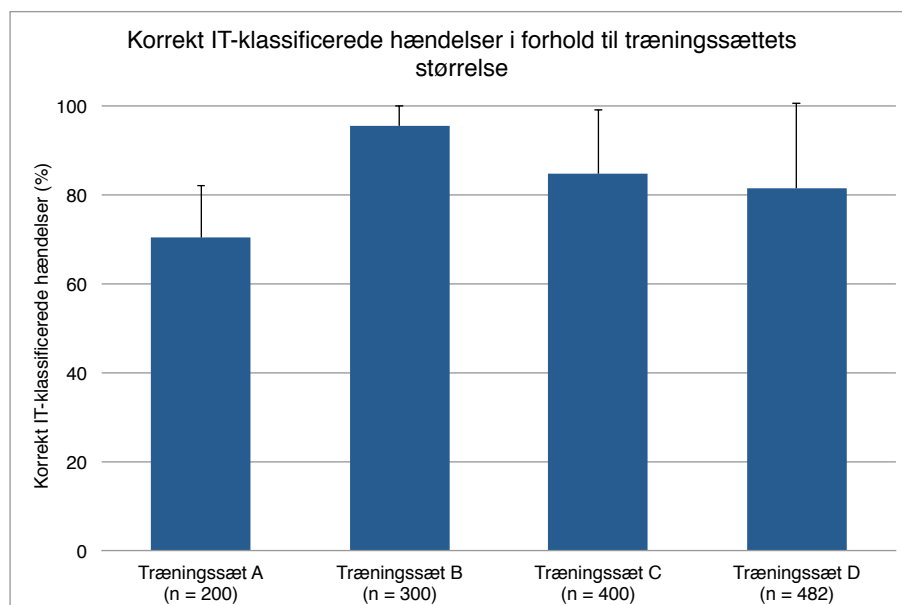
Integrationstesten blev udført på det færdige system for at evaluere om systemet fungerede efter hensigten. Testen viste at systemet gav det forventede output i forhold til det givne input. Det blev derfor konkluderet, at systemet fungerede efter hensigten. En samlet systemtest evaluerede om de funktionelle krav var imødekommet af systemet. Resultatet af denne test viste at samtlige krav var imødekommet. Samlet konkluderes det derfor at systemdesign og implementering imødekommer samtlige krav, og det er derfor lykkedes at implementere et system med de ønskede funktionaliteter.

Systemvalidering

I dette kapitel præsenteres testresultaterne for systemets evne til at klassificere IT-relaterede utilsigtede hændelser ved anvendelse af en Naive Bayes algoritme. En forsøgsprotokol med tilhørende udregninger kan findes i Appendiks E.

8.1 Resultater fra læringstest

Formålet med læringstesten var at se, om hvorvidt en ændring i træningssættets størrelse ville have en indflydelse på klassifikationen. Det blev dannet fire balancerede træningssæt med henholdsvis 100, 150, 200 og 241 IT-relaterede utilsigtede hændelser, som efterfølgende er blevet testet på tre testsæt med hver indeholdende 1000 utilsigtede hændelser, se Figur 8.1.



Figur 8.1: Sammenligning af de fire træningssæts evne til at klassificere IT-relaterede utilsigtede hændelser. Antallet af korrekt klassificerede hændelser er illustreret som et gennemsnit af de tre testsæt, der hver indeholdte 1000 utilsigtede hændelser. Standardafvigelsen er angivet øverst i hver søjle.

Læringstesten viste at Naive Bayes algoritmen præsterede bedst ved brug af træningssæt B med en korrekt klassifikation på 95,51 % og en standardafvigelse på 4,35 %, jævnfør Figur 8.1.

8.2 Resultater fra effektivitetstest

Testen havde til formål at evaluere systemets effektivitet. Tre testsæt med hver indeholdende 1000 utilsigtede hændelser blev udført med et balanceret træningssæt indeholdende 241 IT-relaterede utilsigtede hændelser (træningssæt D).

8.2.1 Testsæt A

I dette testsæt blev der ved manuel gennemgang fundet 21 IT-relaterede utilsigtede hændelser. Systemet klassificerede 61 hændelser, hvoraf 13 var korrekt klassificerede. I Figur 8.2 opstilles en confusionsmatrice med resultatet af testen.

		Ekspertklassificeret		
		IT	Andre	Total
Algoritme klassificeret	IT	13	48	61
	Andre	8	931	939
	Total	21	979	1000

Figur 8.2: Confusionsmatrice for testsæt A. Ved ekspertklassificeret forstås manuel gennemgang af rapporterne. Den anvendte classifier blev trænet med et balanceret træningssæt indeholdende 241 IT-relaterede utilsigtede hændelser (træningssæt D).

8.2.2 Testsæt B

I dette testsæt blev der ved manuel gennemgang fundet 23 IT-relaterede utilsigtede hændelser. Systemet klassificerede 63 hændelser, hvoraf 19 var korrekt klassificerede. I Figur 8.3 opstilles en confusionsmatrice med resultatet af testen.

		Ekspertklassificeret		
		IT	Andre	Total
Algoritme klassificeret	IT	19	44	63
	Andre	4	933	937
	Total	23	977	1000

Figur 8.3: Confusionsmatrice for testsæt B. Ved ekspertklassificeret forstås manuel gennemgang af rapporterne. Den anvendte classifier blev trænet med et balanceret træningssæt indeholdende 241 IT-relaterede utilsigtede hændelser (træningssæt D).

8.2.3 Testsæt C

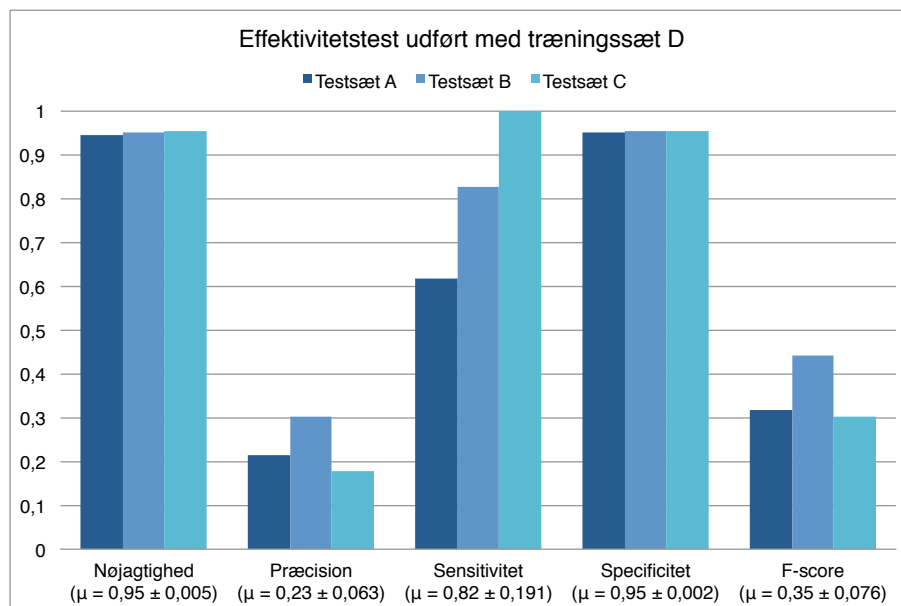
I dette testsæt blev der ved manuel gennemgang fundet 10 IT-relaterede utilsigtede hændelser. Systemet klassificerede 56 hændelser, hvoraf 10 var korrekt klassificerede. I Figur 8.4 opstilles en confusionsmatrice med resultatet af testen. Systemet klassificerede alle 10 IT-relaterede utilsigtede hændelser samt 46 falsk positive.

		Ekspertklassificeret		
		IT	Andre	Total
Algoritme klassificeret	IT	10	46	56
	Andre	0	933	933
	Total	10	979	1000

Figur 8.4: Confusionsmatrice for testsæt C. Ved ekspertklassificeret forstås manuel gennemgang af rapporterne. Den anvendte classifier blev trænet med et balanceret træningssæt indeholdende 241 IT-relaterede utilsigtede hændelser (træningssæt D).

8.2.4 Sammenligning af statistiske parametre

I forbindelse med effektivitetstesten er der udregnet fem statistiske parametre. Figur 8.5 sammenligner parametrene nøjagtighed, præcision, sensitivitet, specificitet og F-score mellem hvert testsæt. Resultatet viser en gennemsnitlig nøjagtighed på 95 % med en standardafvigelse på 0,5 %. For udregningsmetoder og ubehandlet data henvises der til Appendiks E.



Figur 8.5: Parametrene for nøjagtighed, præcision, sensitivitet, specificitet og F-score for testsæt A, B og C med tilhørende middelværdi og standardafvigelse. Den anvendte classifier blev trænet med et balanceret træningssæt indeholdende 241 IT-relaterede utilsigtede hændelser (træningssæt D).

Tabel 8.1 viser resultaterne for hvert testsæt, hvor parametrene nøjagtighed, præcision, sensitivitet, specificitet og F-score er udregnet.

Testsæt	Nøjagtighed	Præcision	Sensitivitet	Specificitet	F-score
Testsæt A	0,94	0,21	0,62	0,95	0,31
Testsæt B	0,95	0,30	0,83	0,95	0,44
Testsæt C	0,95	0,18	1,00	0,95	0,30

Tabel 8.1: *Individuelle resultater for nøjagtighed, præcision, sensitivitet, specificitet og F-score fra effektivitetstesten for testsæt A, B og C. Den anvendte classifier blev trænet med et balanceret trænings-sæt indeholdende 241 IT-relaterede utilsigtede hændelser (trænings-sæt D).*

D E L



Syntese

Diskussion

Siden loven om patientsikkerhed trådte i kraft i 2004, har det været lovpligtigt at indberette utilsigtede hændelser i den danske sundhedssektor. Hændelserne rapporteres til DPSD, hvis formål er at understøtte patientsikkerheden ved at indsamle, analysere og formidle viden og dermed skabe systematisk læring. I 2010 blev der rapporteret ca. 35.000 hændelser, hvilket er en tidobling siden indførelsen af rapporteringspligten. For at sikre den systematiske læring er det nødvendigt at øge tilgængeligheden af information i rapporterne. Det gøres ved at klassificere i forhold til hændelsestype. På nuværende tidspunkt foretages denne klassifikation ved en manuel retrospektiv gennemgang af hændelserne. [Sundhedsstyrelsen, 2010]

Grundet det stigende antal indrapporteringer er der opstået efterspørgsel på systemer, der automatisk kan klassificere utilsigtede hændelser. Udvikling af automatiske systemer kan effektivisere analyseprocessen samt sikre en konsistent klassifikation. Rapporteringen af utilsigtede hændelser foretages i semistrukturerede indrapporteringsskemaer, hvoraf fire felter indeholder fritekst. For at udtrække information fra felterne er det nødvendigt at foretage en tekstanalyse. Dette projekt har til formål at udvikle et system, der automatisk kan klassificere IT-relaterede utilsigtede hændelser fra DPSD ved brug af en Naive Bayes algoritme. Algoritmen er valgt, da flere studier har anvendt den til tekstklassifikation og opnået gode resultater. [Ong et al., 2010, 2012; Schierup et al., 2011]

9.1 Anvendelse af Naive Bayes algoritme

Klassifikationen af de utilsigtede hændelser blev foretaget ved anvendelse af Naive Bayes algoritmen. Algoritmen udmærker sig ved at kræve et lille træningssæt, men kan samtidig håndtere et højdimensionelt input [Nigam et al., 2000; Witten et al., 2011]. Eftersom forekomsten af IT-relaterede utilsigtede hændelser blev fundet til at være 1,5 %, har det været vigtigt at anvende en algoritme, der kræver få træningsdata. Naive Bayes på den antagelse, at data ikke er indbyrdes korreleret. En mulighed kunne være at implementere en algoritme, som ikke bygger på denne antagelse for at se, om det kan have indflydelse på klassifikationen.

9.2 Resultater

Læringstesten blev udført for at se, om en forøgelse i træningssættets størrelse havde indflydelse på klassifikationen. Testen viste at træningssæt B, C og D klassificerede over 80 % korrekt, hvor træningssæt B, med et balanceret træningssæt på 150 IT-relaterede hændelser, præsterede bedst med 95,51 %. Træningssæt A klassificerede 70,35 % kor-

rekt, hvilket kunne skyldes at modellen var baseret på et træningssæt med få rapporter. Fælles for alle træningssæt var den store standardafvigelse som resultat af, at klassifikationen kun blev udført med tre testsæt for hvert træningssæt. For at finde tendenser, der kan beskrive, hvilken indflydelse mængden af træningsdata har i forhold til antallet af korrekte klassificerede, er det nødvendigt at inkludere flere testsæt. Da der ikke blev fundet tydelige tendenser blandt testsættets størrelse, blev det valgt at udføre effektivitetstesten på det størst mulige træningssæt. Det blev vurderet at det største træningssæt ville danne den mest repræsentative model for de to klasser.

Resultatet af effektivitetstesten viste en gennemsnitlig nøjagtighed på 95 % korrekt IT-klassificerede hændelser og en præcision på 23 %. I projektet af Schierup et al. [2011] blev der opnået en nøjagtighed og præcision på 95 % på data klassificeret fra AIMS-databasen. Den store forskel i præcisionen mellem de to systemer kan skyldes at der i Danmark indrapporteres utilsigtet hændelser fra alle faggrupper samt fra patienter og pårørende, der selv har været udsat for eller vidne til en utilsigtet hændelse. Dermed kan der være stor variation i de anvendte terminologier i rapporterne. Eksempelvis er patienter og pårørende ofte ikke bekendte med terminologien, der anvendes i forbindelse med sundheds-IT. Således varierer ordforrådet mellem lægmandssprog og fagtermer, hvilket kan vanskeliggøre den manuelle klassifikation og skabe ”støj” ved den automatisk klassifikation. Ved at udelukke patienter og pårørende fra at indrapportere til databasen vil der til en vis grad kunne skabes en mere ensartet terminologianvendelse. Dette er dog ikke sandsynligt at antage, da patienter og pårørende i nogle tilfælde kan opdage utilsigtede hændelser, som sundhedspersonalet ikke var opmærksomme på og derfor skal problematikken håndteres anderledes.

Den høje nøjagtighed og lave præcision, som er opnået med den anvendte algoritme, betyder at systemet identificerer størstedelen af de IT-relaterede hændelser (sand positive), hvilket er på bekostning af en forholdsvis stor mængde falsk positive resultater. Det kan således diskuteres om systemet har potentiale for praktisk anvendelse i klassifikationsprocessen af utilsigtede hændelser.

9.3 Anvendelsespotentiale

Da det udviklede system ikke anvender data i direkte forbindelse med behandling af patienter, vil fejlklassifikationer ikke være ensbetydende med en reduceret patientsikkerhed. Sandsynligheden for at risikomanagerne fejlklassificerer eller overser IT-relaterede rapporter er eksisterende, hvilket er en konsekvens af den store mængde rapporter de gennemlæser. Det er derfor oplagt at diskutere hvorvidt konsekvenserne af det antal IT-relaterede rapporter, som klassificeres forkert (falsk negative), kan opvejes af reduktionen i rapporter, som ikke er relateret til IT. Analysen af de utilsigtede hændelser bruges af Patientombuddet til at understøtte patientsikkerheden. På baggrund af analysen udgives forskellige publikationer, herunder OBS-meddelelser, temarapporter og årsrapporter, som kan forebygge utilsigtede hændelser i fremtiden. Patientombuddets arbejde afhænger ikke af at samtlige IT-relaterede utilsigtede hændelser identificeres, da der hovedsageligt fokuseres mere på analyse af tendenser og fejltyper end af specifikke fejl. Risikomanageren analyserer de klassificerede hændel-

ser ved manuel gennemgang og har derved mulighed for at afvise en fejlagtig klassifikation.

Ved systemtesten blev det bevist at antallet af ikke IT-relaterede hændelser i gennemsnit var reduceret med en faktor ti. Dermed vil risikomanageren kunne spare tid i analyseprocessen. Da det på nuværende tidspunkt er forskellige risikomanagere, som manuelt klassificerer de indrapporterede hændelser, vil der være en naturlig variation blandt rapporter der er klassificeret som IT-relaterede. Denne variation kan gøre det vanskeligt at lave en statistisk repræsentation af rapporterne. Ved anvendelse af automatisk klassifikation kan der opnås en mere konsistent klassifikation, da rapporterne bliver bedømt efter samme parametre hver gang.

9.4 Udvikling af brugergrænseflade

At udvikle et brugervenligt system er en iterativ proces, som kræver høj brugerinvolvement. Derfor blev der udført to usabilitytest. Første test havde til formål at samle erfaringer fra systemet udviklet af Schierup et al. [2011]. Efter fem testpersoner havde fuldført testen, blev der vist en tendens til at testpersonerne fandt det vanskeligt at foretage en klassifikation med klassifikationsprogrammet Weka, samt at mange trin skulle fuldføres før en klassifikation var udført. Derfor blev det besluttet, at det ikke var nødvendigt at teste på flere personer, førend disse udfordringer var imødekommet.

På baggrund af erfaringerne blev der udviklet en prototype af brugergrænsefladen til det udviklede system, som imødekom testpersonernes input. Efterfølgende blev der udført en usability test på prototypen. For at udgå at testpersonerne havde et forudgående kendskab til systemet, og dermed kunne skabe bias i testen, blev de udvalgt fra et kriterium, om at de ikke tidligere havde haft kendskab til systemet. Testpersonerne var udvalgt blandt medstuderende med ingeniørmæssige erfaringer herunder udvikling af IT. Da slutbrugerne er risikomanagere, som ofte har en klinisk baggrund, har disse personer andre forudsætninger for anvendelsen af programmet, hvorfor det vil være relevant at teste systemet på slutbrugeren før en endelig brugergrænseflade implementeres.

Konklusion

I dette projekt blev potentialet for anvendelse af automatisk tekstklassifikation til kategorisering af IT-relaterede utilsigtede hændelser undersøgt. Et system kaldet UTH-Finder blev udviklet med det formål at klassificere rapporter fra DPSD ved brug af en klassifikationsalgoritme baseret på Naive Bayes.

Under udviklingen stillede Patientombuddet 31.184 rapporter til disposition, hvoraf halvdelen af data blev brugt til træning og de resterende til test af systemet. Ved træning af systemet blev der sammensat et balanceret træningssæt indeholdende 241 IT-relaterede hændelser. Hændelserne blev udvalgt ved en manuel gennemgang af 15.592 rapporter, hvor forekomsten af IT-relaterede hændelser blev bestemt til 1,5 %.

Systemtesten blev foretaget for at undersøge systemets evne til at klassificere IT-relaterede utilsigtede hændelser. Testen viste lovende resultater ved klassifikation af de anvendte testsæt, hvor der blev opnået en gennemsnitlig nøjagtighed og specificitet på 95 %. Systemet udmærker sig ved at kunne klassificere hændelser korrekt og dermed reducere mængden af hændelser, som risikomanageren manuelt skal gennemlæse. Endvidere vil systemet skabe en mere konsistent klassificering.

Det konkluderes med udviklingen af UTH-Finder at automatisk tekstklassifikation har potentiale for anvendelse til kategorisering af utilsigtede hændelser fra DPSD. De opnåede resultater berettiger til yderligere undersøgelser med fokus på forbedring af algoritmen samt at undersøge, hvorvidt modellen kan udvides til at omfatte flere hændelsestyper.

Perspektivering

UTH-Finder er i dets nuværende fase et proof-of-concept system. Dermed bør en række forbedringer foretages før systemet kan anvendes i praksis. Systemet er i stand til at klassificere med en høj nøjagtighed. Til gengæld er præcisionen forholdsvis lav, hvorfor et forbedringstiltag kunne være at optimere klassifikationsalgoritmen. Forbedring af algoritmen kan ske ved en overvejelse af, hvorledes parametre til dannelse af en model kan udvælges, herunder hvordan diversiteten mellem de anvendte terminologier behandles, samt hvordan der kan tages højde for stave- og slåfejl.

Systemets algoritme er på nuværende tidspunkt designet til at klassificere mellem to klasser. En betydelig forbedring vil være at udvide systemet fra den nuværende binære klassifikation til en flerstrengt klassifikation. Ved at håndtere flere klasser simultant vil der samtidig kunne opnås en højere grad af automatisering.

Da diversiteten i de overordnede fejkategorier, som IT, medicin og fald med flere, spænder over mange specifikke fejltypen, bør den praktiske anvendelse overvejes. Eksempelvis kan en IT-relateret hændelse skyldes tyveri af hardware, men også være forårsaget af nedbrud på netværket. Den store diversitet af indholdet i en overordnet kategori gør det vanskeligt at foretage statistiske analyser af enkelte fejltypen. Derfor vil yderligere underklassificering af specifikke fejltypen for hver kategori kunne sikre en mere praktisk tilgang til analysen.

Ligeledes kan en dynamisk klassifikationsmodel kan med fordel implementeres således, at brugeren kan oplære systemet ved manuelt, at verificere de klassificerede rapporter. Dermed vil systemet kunne tilpasse sig de sproglige forandringer, der afspejler sig i rapporterne, hvilket i sidste ende vil forbedre klassifikationsnøjagtigheden.

Potentialet for at anvende systemet, som et analyseredskab med henblik på identifikation og håndtering af risikoområder, er stort. Der vil i første omgang kunne skabes en mere konsistent klassifikation, da alle hændelser klassificeres efter samme parametre. Den konsistente klassifikation vil betyde, at det efterfølgende er nemmere at aggregere på data. Derudover vil den øgede automatisering kunne effektivisere arbejdet for risikomanagerne, som derved ikke skal læse store mængder af data igennem.

Litteratur

- Barach, P. og Small, S. D. (2000). Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems. *BMJ*, 320:759-763.
- Bjørn, B., Anhøj, J. og Lilja, B. (2009). Rapportering af utilsigtede hændelser. *Ugeskr Læger*, 171:1677-1680.
- DPSD (2012). *Rapporter en utilsigtet hændelse*. Dansk Patientsikkerhedsdatabase. URL https://dpsd.csc-scandihealth.com/Form/PublicSubmission.aspx?form=DPSD_Public.
- DSI (2001). *Utilsigtede hændelser på danske sygehuse*. DSI Institut for Sundhedsvæsen.
- DSKS (2003). *Sundhedsvæsenets kvalitetsbegreber og -definitioner*. Dansk Selskab for Kvalitet i Sundhedssektoren (DSKS).
- DSP (2012). *SAC Matrix score systemet*. Dansk Selskab for Patientsikkerhed. URL <http://patientsikkerhed.dk/index.php?id=470>.
- Johnson, C. W. (2003). How will we get the data and what will we do with it then? issues in the reporting of adverse healthcare events. *Qual Saf Health Care*, 12:ii64-ii67.
- Johnson, S. B., Bakken, S., Dine, D., Hyun, S., Mendonca, E., Morrison, F., Bright, T., Vleck, T. V., Wrenn, J. og Stetson, P. (2008). An electronic health record based on structured narrative. *J Am Med Inform Assoc*, 15:54-64.
- Kohn, L. T., Corrigan, J. M. og Donaldson, M. S. (2000). *To err is human: Building a safer health system*. National Academies Press.
- Nagamatsu, S., Kami, M. og Nakata, Y. (2009). Healthcare safety committee in japan: Mandatory accountability reporting system and punishment. *Current Opinion in Anaesthesiology*, 22:199-206.
- Nigam, K., McCallum, A. K. og Mitchell, T. (2000). Text classification from labeled and unlabeled documents using em. *Machine Learning*, 39:103-134.
- Ong, M.-S., Magrabi, F. og Coiera, E. (2010). Automated categorisation of clinical incident reports using statistical text classification. *Qual Saf Health Care*, 19:1-7.
- Ong, M.-S., Magrabi, F. og Coiera, E. (2012). Automated identification of extreme-risk events in clinical incident reports. *J Am Med Inform Assoc*, 1:1-9.
- Pham, J. C., Gianci, S., Battles, J., Beard, P., Clarke, J. R., Coates, H., Donaldson, L., Eldridge, N., Fletcher, M., Goeschel, C. A., Heitmiller, E., Hensen, J., Kelley, E., Loeb, J., Runciman, W., Sheridan, S., Wu, A. W. og Pronovost, P. J. (2010). Establishing a global learning community for incident-reporting systems. *Qual Saf Health Care*, 19:446-451.
- Runciman, B., Merry, A. og Walton, M. (2007). *Safety and Ethics in Healthcare - A Guide to Getting it Right*. Ashgate Publishing Company.

- Schierup, A., Routhe, J. og Jensen, C. B. (2011). Automated categorisation of patient safety incident reports related to health information technology using text classification.
- Sharp, H., Rogers, Y. og Preece, J. (2002). *Interaction Design: Beyond Human-Computer Interaction*. Wiley.
- Sittig, D. F. og Singh, H. (2011). Defining health information technology-related errors. *Arch Intern Med*, 171:1281-1284.
- Sundhedsministeriet (2010). *Sundhedsloven*. Indenrigs- og Sundhedsministeriet.
- Sundhedsstyrelsen (2002). *National strategi for kvalitetsudvikling i sundhedsvæsenet*. Sundhedsstyrelsen.
- Sundhedsstyrelsen (2010). *Årsrapport 2010 - DPSD, Dansk Patientsikkerhedsdatabase*. Sundhedsstyrelsen.
- Wang, Z., Shah, A. D., Tate, A. R., Denaxas, S., Shawe-Taylor, J. og Hemingway, H. (2012). Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. *PLoS ONE*, 7:1-7.
- Warden, G. L. (2012). *Health IT and patient safety: Building safer systems for better care*. The National Academies Press.
- Weiner, J. P., Kfuri, T., Chan, K. og Fowles, J. B. (2007). "e-iatrogenesis": The most critical unintended consequence of cpoe and other hit. *J Am Med Inform Assoc*, 14:387-388.
- WHO (2005). *WHO draft guidelines for adverse event reporting and learning systems*. World Health Organization.
- Witten, I. H., Frank, E. og Hall, M. A. (2011). *Data mining - practical machine learning tools and techniques*(third edition). Elsevier.
- Wu, A. W., Pronovost, P. og Morlock, L. (2002). Icu incident reporting systems. *Journal of Critical Care*, 17:86-94.

D E L



Appendiks



Indrapportering til Dansk Patientsikkerhedsdatabase

I dette appendiks gennemgås indrapporteringsskemaet for sundhedsprofessionelle til Dansk Patientsikkerhedsdatabase.

A.1 Rapporteringsskemaets opbygning

Rapporteringen sker i et semistruktureret skema idet der både anvendes strukturerede felter og felter til fritekst. Skemaet består af følgende otte sektioner [DPSD, 2012]:

- Hvad skete der?
- Hvor og hvornår skete det?
- Hvad var konsekvensen?
- Forslag til forebyggelse
- Oplysninger om patienten
- Var andre faktorer involveret i hændelsen?
- Kontaktoplysninger
- Vedhæft en fil

A.1.1 Hvad skete der?

Denne sektion består af to obligatoriske fritekstfelter, se Figur A.1. I det første felt skal der skrives en overskrift, som er dækkende for hændelsen. Hvis f.eks. hændelsen omhandler en patient, der faldt, kan ordet "fald" og en kort beskrivelse af hvor det skete indgå i overskriften. Det andet felt skal indeholde en beskrivelse af hændelsen. Det undlades at skrive stillingsbetegnelser, navne, CPR-numre eller andre personhenførbare data i dette felt. Bortset fra denne begrænsning er der ingen retningslinjer for feltets indhold. Afhængig af indberetteren kan beskrivelsen være alt fra en enkelt sætning til en flere siders lang beretning. [DPSD, 2012]

▼ 1. Hvad skete der?

Husk: Skriv aldrig navn, CPR-nummer eller andre personhenførbare oplysninger i nogle af de beskrivende felter.

Overskrift ★ [?]

Beskriv hvad der skete ★ [?]

Figur A.1: Skærbillede af sektionen "Hvad skete der?" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.2 Hvor og hvornår skete det?

Beskrivelsen af tid og sted består af fire strukturerede felter, hvor der er implementeret en obligatorisk søgefunktion til at finde de relevante oplysninger om lokalitet, postnummer, sygehus og afdeling, se Figur A.2. Feltet, der beskriver lokaliteten, er det eneste obligatoriske. Der kan angives et opdagelsessted, hvis ikke hændelsen blev opdaget på det sted hvor den skete. Dette er ligeledes et struktureret felt med en obligatorisk søgefunktion. [DPSD, 2012]

▼ 2. Hvor og hvornår skete det?

Søg hændelsessted Søg...

Lokalitet ★ [?] Postnummer [?]

Opdagelsessted

Hvis hændelsen blev opdaget et andet sted end hvor den skete - så angiv venligst denne lokalitet her nedenfor, ved at klikke på "Tilføj"

Navn

Ingen data

Tilføj Aben Slet

Dato og tidspunkt

Dato: [?] [?] Tidspunkt: [?] [?]

23-05-2012

Kendt

Kendt

Figur A.2: Skærbillede af sektionen "Hvor og hvornår skete det?" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.3 Hvad var konsekvensen?

Her skal to obligatoriske felter udfyldes, se Figur A.3. Først et fritekstfelt, der beskriver, hvilke følger hændelsen havde for patienten. Det undlades at skrive navne, CPR-numre eller andre personhenførbare data i dette felt. Der er ingen begrænsninger for indholdet i dette felt. Det andet felt er et struktureret felt, hvor alvorligheden af hændelsen angives ud fra seks valgmuligheder. [DPSD, 2012]



▼ 3. Hvad var konsekvensen?

Beskriv hvad den umiddelbare konsekvens af hændelsen var: * [?]

Hvor alvorlig var hændelsen? * [?]

Figur A.3: Skærbillede af sektionen "Hvad var konsekvensen?" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.4 Forslag til forebyggelse

I det obligatoriske fritekstfelt skal der beskrives hvilke tiltag, der vil kunne forebygge lignende hændelser, se Figur A.4. Her udlades det at skrive navne, CPR-numre eller andre personhenførbare data. [DPSD, 2012]



▼ 4. Forslag til forebyggelse

Beskriv dine forslag * [?]

Figur A.4: Skærbillede af sektionen "Forslag til forebyggelse" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.5 Oplysninger om patienten

Denne sektion indeholder fem valgfrie strukturerede felter, der beskriver den involverede patient, se Figur A.5. Her ønskes oplysninger om navn, CPR-nummer, køn, alder, patientstatus og hvis flere patienter er involveret. [DPSD, 2012]

5. Oplysninger om patienten

Patientens navn og CPR-nummer anvendes kun til, at din lokale sagsbehandler kan få hændelsens omstændigheder yderligere belyst. Oplysningerne slettes automatisk inden sagen indsendes til Patientombuddet.

Patientens navn [?] Patientens CPR-nummer (Husk bindestreg) [?]

Patientens køn: [?] Patientens alder: [?] Patientstatus: [?] Flere patienter involveret [?]

Figur A.5: Skærbillede af sektionen "Oplysninger om patienten" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.6 Var andre faktorer involveret i hændelsen?

Sektionen indeholder tre felter til angivelse af faktorerne "Medicin", "Medicinsk udstyr eller hjælpemidler" og "Røntgen eller anden stråling", se Figur A.6. For alle tre felter er der implementeret en søgefunktion til angivelse af eventuelt involverede faktorer. [DPSD, 2012]

6. Var andre faktorer involveret i hændelsen?

Oplysningerne i dette afsnit bruges dels til at identificere de hændelser, som skal behandles af særlig sagkyndige på området, dels til at sikre, at vi får tilstrækkeligt detaljeret oplysninger om hændelsen.

Medicin

Handelsnavn	Indholdsstof	Adm.vej	Styrke	ATC
Ingen data				

Tilføj | Åben | Slet

Medicinsk udstyr eller hjælpemidler

Udstyrs type	Handelsnavn	Model
Ingen data		

Tilføj | Åben | Slet

Røntgen eller andet stråling

Strålingstype [?]

Figur A.6: Skærbillede af sektionen "Var andre faktorer involveret i hændelsen?" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.7 Kontaktoplysninger

Fire felter til angivelse af navn, e-mail, telefonnummer og faggruppe skal udfyldes i denne sektion, se Figur A.7. Hvis rapportøren ønsker at være anonym kan felterne undlades. [DPSD, 2012]

7. Kontaktoplysninger

Dine kontaktoplysninger skal bruges til at din lokale sagsbehandler kan indhente eventuelle supplerende oplysninger om hændelsen og omstændighederne omkring den. Kontaktoplysningerne vil ikke blive videregivet og de vil blive slettet automatisk før rapporten indsendes til Patientombuddet.

Hvis du skriver din e-mailadresse, så vil du modtage en bekræftigelse på modtagelsen af din rapport.

Navn [?] Email [?] Tlf. [?] Faggruppe [?]

Figur A.7: Skærbillede af sektionen "Kontaktoplysninger" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]

A.1.8 Vedhæft en fil

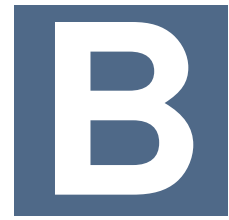
Rapportøren har i denne sektion mulighed for at vedhæfte en fil med yderligere oplysninger, se Figur A.8. [DPSD, 2012]

8. Vedhæft en fil

Vedhæftet fil: Der er ikke valgt nogen fil

Maks. filstørrelse: 3 MB

Figur A.8: Skærbillede af sektionen "Vedhæft en fil" fra DPSDs indrapporteringsskema for sundhedsprofessionelle. [DPSD, 2012]



Test af systemet

Følgende appendiks dokumenterer test relateret til Unified Process, som blev udført i forbindelse med systemudviklingen.

B.1 Typer af test

Test af et system har til formål at undersøge, hvorvidt et system fungerer efter hensigten, herunder lokalisere og udbedre fejl som ikke blev opdaget under udviklingen. De mest omfattende test, der bliver udført, er systemtest og integrationstest. Disse udføres efter udviklingen af en prototype for at evaluere om systemet imødekommer kravspecifikationen. Unit- og regressionstest udføres løbende under udviklingen og vil derfor ikke blive dokumenteret. Følgende test er således blevet udført:

- Unittest
- Regressionstest
- Integrationstest
- Systemtest

B.2 Integrationstest

Integrationstesten er en samlet afprøvning af det færdigudviklede system. Det testes således om komponenter og klasser arbejder sammen efter hensigten. Integrationstesten har kun fokus på input og output af systemet, men tager ikke hensyn til systemets opbygning.

B.2.1 Metode

Der foretages en systematisk test af alle scenarier fra de implementerede use cases. Således testes systemets mainflow og alternative flows. Fejl vil udbedres løbende og samtidig undersøges det, om udbedringer af fejl påvirker programeksekveringen. Da fejl løbende udbedres vil resultatet af testen være fra systemets sidste iteration, hvor samtlige fejl er udbedret. De testede use cases er som følger:

- ID1: Klassificér UTH-rapporter
- ID2: Gem resultat

ID1: Klassificér UTH-rapporter

ID1: Klassificér UTH-rapporter
Formål: Det testes om systemet kan klassificere UTH indlæst fra en tekstfil.
Metode: Der indlæses en testfil indeholdende UTH-rapporter som skal klassificeres. Der angives en a priori sandsynlighed svarende til den virkelige fordeling af kategorierne.
Forudsætninger: 1. Systemet er trænet ved indlæsning af modelfil.
Input: 1. Indlæsning af modelfil 2. Indlæsning af testfil 3. A priori sandsynlighed
Forventet output: Udskrift af de IT-klassificerede hændelser på brugergrænsefladens tekstfelt.
Resultat: Efter klassifikationen udskrives de IT-klassificerede hændelser på brugergrænsefladen.

Tabel B.1: Resultat fra integrationstesten af use casen "Klassificér UTH-rapporter".

ID2: Gem resultat

ID2: Gem resultat
Formål: At teste om det er muligt at gemme en tekstfil med de IT-kategoriserede hændelser, så yderligere analysering kan foretages i eksterne programmer.
Metode: Efter klassifikationen vælges der, at der skal gemmes til en fil.
Forudsætninger: 1. Klassifikationen er udført.
Input: 1. Filnavn 2. Sti
Forventet output: En tekstfil indeholdende IT-klassificerede hændelser.
Resultat: Tekstfil indeholdende IT-klassificerede hændelser gemt under specificeret filnavn og sti.

Tabel B.2: Resultat fra integrationstesten af use casen "Gem resultat".

B.2.2 Konklusion på integrationstest

Integrationstesten viste for use casen "Klassificér UTH-rapporter" at korrekt indlæsning af modelfil, testfil og a priori sandsynlighed gav det forventede output. For use casen "Gem resultat" blev det forventede output ligeledes opnået. Dermed konkluderes det at systemet virkede efter hensigten.

B.3 Systemtest

Formålet med systemtesten er at undersøge om kravspecifikationen imødekommes af systemet. Testen udføres på det færdige system og er en black box test hvor kun input og output kendes på forhånd.

B.3.1 Metode

Systemtesten tager udgangspunkt i systemets funktionelle krav. Hver enkelt krav gennemgås, for at evaluere om systemet imødekommer disse krav. Lever systemet ikke op til kravspecifikationen, vil dette blive udbedret løbende, hvorfor dokumentationen vil være på baggrund af systemets sidste iteration.

Krav	Resultat
Systemet skal kunne håndtere utilsigtede hændelser fra DPSD.	Kravet er imødekommet, systemet indlæser rapporter fra DPSD.
Systemet skal kunne håndtere .csv-filer.	Kravet er imødekommet, systemet indlæser model og test-filer af formatet .csv.
Systemet skal kunne klassificere utilsigtede hændelser i to kategorier, IT og andet.	Kravet er imødekommet testfilens indhold kategoriseres i de to kategorier.
Systemet skal kunne foretage en automatisk klassifikation ved anvendelse af Naive Bayes tekstklassifikation algoritmen.	Kravet er imødekommet, Naive Bayes algoritmen er implementeret i systemet.
Systemet skal kunne give et output i form af en tekstfil, indeholdende hændelser klassificeret som IT-relaterede fejl.	Kravet er imødekommet, resultatet af klassifikationen kan gemmes i en tekstfil.
Systemet skal indeholde en intuitiv brugergrænseflade.	Kravet er imødekommet, systemet indeholder en brugergrænseflade, som er udviklet på baggrund af input fra brugeren.

Tabel B.3: Resultat af systemtest.

B.3.2 Konklusion på systemtest

Det konkluderes at systemtesten imødekom kravspecifikationen.



Naive Bayes

Naive Bayes classifier tilhører Bayes familien og bygger på antagelsen af, at der ikke er korrelation mellem data. Naive Bayes bliver ofte brugt i tekstklassifikation, da den er hurtig og nøjagtig. Derudover præsterer den godt på små trænings sæt og med test sæt af høj dimensionalitet. Ved anvendelse af bayesianske algoritmer kombineres træningsdata med a priori viden om en hændelse for at bestemme en a posteriori sandsynlighed for en hypotese. Således er det muligt at finde den mest sandsynlige hypotese på baggrund af træningsdata. Naive Bayes bygger på Bayes rule, som er givet ved:

$$p(h|D) = \frac{p(D|h)p(h)}{p(D)} \quad (\text{C.1})$$

hvor

- $P(h)$ = a priori sandsynlighed for hypotese h
- $P(D)$ = a priori sandsynlighed for træningsdata D
- $P(h|D)$ = sandsynlighed for h givet D
- $P(D|h)$ = sandsynlighed for D givet h

På baggrund af træningsdata kan Naive Bayes algoritmen beregne sandsynligheden for to hypoteser givet antagelsen, at data ikke er internt korreleret. Den naive antagelse kan matematisk beskrives som:

$$v = \max_{v_j \in V} P(v_j | a_1, a_2, \dots, a_n) \quad (\text{C.2})$$

$$= \max_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n) P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (\text{C.3})$$

$$= \max_{v_j \in V} P(a_1, a_2, \dots, a_n | v_j) P(v_j) \quad (\text{C.4})$$

$$P(a_1, a_2, \dots, a_n | v_j) = \prod P(a_i | v_j) \quad (\text{C.5})$$

Hypotesen som opfylder følgende argument vælges:

$$v = \max_{v_j \in V} P(v_j) \prod P(a_i | v_j) \quad (\text{C.6})$$

Dette betyder at hypotesen som maksimerer den samlede sandsynlighed vælges mens den anden hypotese afvises.



Usability test

I det følgende appendix beskrives de to usability test, der blev fortaget. Første usability test blev udført på Incident Tracker. På baggrund af resultaterne blev der udarbejdet en Lo-Fi prototype af brugergrænsefladen til det udviklede system. Ligeledes blev der fortaget en usability test på systemet.

D.1 Evaluering af brugervenlighed og -oplevelse

For at udvikle et brugervenligt produkt er det nødvendigt at identificere fejl og mangler som forringer brugervenligheden og brugeroplevelsen. Grundlæggende vil brugeren have et system, der er intuitivt, effektivt, sikkert og tilfredsstillende. Derfor er det betydningsfuldt at vide, hvad der skal evalueres og hvornår i processen evalueringen skal finde sted. Dette kræver en viden om brugerens behov, hvilket skal inddrages i alle faser af systemudviklingen. Er der forståelse af brugernes behov reflekteres dette klart i det gode design. Brugerens behov kan opnås gennem et bredt spektrum af forskellige test-, evaluerings- og interviewmetoder, der hver henvender sig til forskellige stadier i udviklingen. Ligeledes vil brugeren give bedre og mere detaljeret feedback, når tidligere designidéer er blevet implementeret, hvilket udløser flere forbedringer. Dermed opstår der et positivt feedbacksystem, hvor brugerens behov og inddrivelsen af denne har en hovedrolle i systemudviklingen. [Sharp et al., 2002]

Tognazzini er førende indenfor udvikling af Interaction Design og har udpeget følgende grunde til at investere i brugertest [Sharp et al., 2002]:

- Problemer bliver rettet før produktet er udsendt og ikke efter.
- Udviklingsholdet kan forholde sig til virkelige problemer og ikke kunstige.
- Ingeniører kan kode istedet for at debattere.
- Udviklingsperioden bliver stærkt reduceret.

D.2 Målsætning med usability test

Usability test er en af hovedmetoderne inden for evaluering af softwareprototyper. Ved usability test vurderes den typiske brugers udførelse af nogle specifikke forberedte opgaver der typisk vil blive udført i softwaren. Under testen bliver brugerens interaktion med softwaren observeret og videooptaget. Disse optagelser bruges senere til at identificere fejl, tidsforbrug og brugeradfærd. Endvidere anvendes brugertilfredshedsspørgeskemaer og interviews for at få resultater af mere kvalitativ art. Evalueringen foregår typisk derefter ved optælling af omfanget af fejl, den anvendte tid for en opgave samt ved at inddrage og diskutere de indsamlede erfaringer fra det kvalitative data. Usability

test er en stærkt kontrolleret form for test og foregår hovedsageligt i laboratorielignende omgivelser. Alle udenforliggende forstyrrelser er ikke tilladt og dermed simuleres brugerens omgang med programmet ikke som en ægte hverdagsituation. Usability test er hovedsagelig af kvantitativ karakter. Dog kan data opsamlet i en usability test ikke bruges til statistiske analyser, da antallet af involverede testpersoner er begrænset. Et minimumsniveau for godkendelse kan dermed specificeres og holdes op mod tidligere usability test. [Sharp et al., 2002]

I første usability test undersøges programmet Incident Tracker udviklet af Schierup et al. [2011]. Formålet med Incident Tracker var at udvikle et program, der automatisk kunne processere og analysere data fra forskellige rapporteringssystemer og dermed ikke havde brugervenlighed som førsteprioritet. Efterfølgende præsentationer af programmet har vist, at det ikke er intuitivt og simpelt som først antaget, hvorfor en usability test er foretaget for at belyse dette aspekt. Hensigten med denne test er at se, hvordan interfacet i Incident Tracker kan forbedres.

Erfaringer fra usability testen af Incident Tracker implementeres i en Lo-Fi prototype af det nye system. Dernæst foretages der en usability test af Lo-Fi prototypen. Formålet med denne test var ikke at finde yderligere forbedringer, men at evaluere de forbedringer, der allerede var foretaget. På baggrund af de opnåede resultater udvikles den endelige version af systemet.

D.3 Forsøgsprotokol for Incident Tracker

Forsøgene blev udført af Anders Schierup, Christian Bruun Jensen og Duy Thien Van ved Aalborg Universitet. Ved første usability test blev data opsamlet med videooptagelse og notater under testseancen samt ved spørgeskema og interview efter testen. Notater blev benyttet til at indfange detaljer, som ikke kom med på videoen. Det efterfølgende spørgeskema og interview gjorde det muligt at få kvantificeret besvarelserne og få kvalitative data. I spørgeskemaet blev en række spørgsmål stillet i relation til de opgaver, der skulle løses. Testbrugeren blev bedt om at besvare spørgsmålene med en kort forklarende tekst samt en score omhandlende sværhedsgraden af opgaven. Scoren rangerende fra 1-3, hvor 1 = let, 2 = middel, 3 = svær og * (stjerne) hvis testbrugeren havde brug for assistance. Såfremt en opgave fik en gennemsnitsscore på 1,5 skulle en detaljeret analyse af problemet finde sted. Resultatet kan findes i Tabel D.1. De nævnte problemer blev efterfølgende prioriteret i forhold til alvorlighed, hvilket også angav løsningsrækkefølgen. Resultater kan findes i Tabel D.2. Til usability testen blev fem forsøgspersoner rekrutteret.

Den anden usability test blev foretaget på Lo-Fi prototypen og præsenteret for fem forsøgspersoner. Prototype blev evalueret med besvarelse af spørgeskemaet i Tabel D.3.

D.3.1 Forsøgsudførelse

Følgende forsøgsprotokol henvender sig til den første usability test hvor Incident Tracker testes. Udførelse af den anden usability test vil kort blive gennemgået i afsnit D.5.1.

Forsøgspersonen blev anvist en stol foran en dertil opstillet computer og blev briefet i testens formål og forløb forud for testens begyndelse. Inden testen fik forsøgspersonen fem minutter til at gøre sig bekendt med programmet. Under testen blev forsøgspersonen bedt om løbende at give kommentarer på hvad forsøgspersonen så på, tænkte og fandt forvirrende. Forsøgspersonen blev stillet en række struktureret opgaver, som alle er beskrevet i afsnit D.3.2. Den forsøgsansvarlige noterede detaljer undervejs angående forsøgsudførelsen. Hvis forsøgspersonen løb ind i problemer, trådte den forsøgsansvarlige til med en hjælpende hånd. Ved endt test blev forsøgspersonen bedt om at udfylde et spørgeskema, hvor hver opgave skulle gives en score baseret på sværhedsgraden. Herefter blev forsøgspersonen interviewet.

D.3.2 Instruktioner

Følgende tekst blev læst op for testbrugeren under usability testen. Teksten indeholder en forklaring af programmet og opgaverne, der skal løses:

I denne test er vi interesseret i at opnå en forståelse af de problemer, som en bruger oplever når Incident Tracker benyttes. Incident Tracker er et statistisk værktøj, der kan benyttes til at præprocessere data fra to forskellige rapporteringssystemer til utilsigtede hændelser til machine learning programmet Weka. Dermed kan Incident Tracker benyttes til at danne et trænings sæt, som kan træne en tekstklassifier og præprocessere ukendt data fra forskellige databaser. Data fra denne test skal bruges til at udvikle et nyt system, der skal behandle data fra DPSD.

Vi vil bede dig om at løse følgende opgaver, så vi kan opnå en bedre brugergrænseoplevelse ved at lokalisere og belyse uforudsete problemer. Når du udfører følgende opgaver, er det vigtigt for os, at høre hvad du tænker. Derfor vil vi bede dig om at fortælle os, hvad du kigger på, hvad du tænker og hvad dig forvirrer dig, når du løser opgaverne.

Opgave 1: Åbn Incident Tracker

Din første opgave er at åbne programmet og bruge fem minutter på at gøre dig bekendt med det. Husk at fortælle os hvad du kigger og tænker på, imens du går på opdagelse.

Opgave 2: Opret en labelled MAUDE-fil

I denne opgave skal en MAUDE-træningsfil konstrueres, så den er klar til at blive benyttet i Weka. MAUDE-filer er gemt i csv-filformatet. En utilsigtet hændelse i MAUDE er opdelt i to filer. En tekstfil og en devicefil, som hver er underopdelt i to filtyper, positive og negative. Begge filer og deres underkategorier skal benyttes til at danne en træningsfil, som senere kan benyttes i Weka.

Opgave 3: Anvend den oprettede MAUDE-fil til at teste klassiferingen i Weka

I denne opgave skal klassiferingen fra MAUDE-data klargøres til Weka. Det gøres ved anvendelse af den oprettede træningsfil.

Opgave 4: Åbn en fil i Weka og klassificér den med Naive Bayes

Åbn en fil i Weka og klassificér den med Naive Bayes.

D.3.3 Spørgeskema og interview

Efter endt test fik forsøgspersonen udleveret et spørgeskema. Forsøgspersonerne, som er medtaget i denne undersøgelse, har alle en civilingeniørmæssig baggrund og har derfor erfaringer med brug af IT-systemer. Forsøgspersonerne blev bedt om at vurdere sværhedsgraden for den stillede opgave hvor 1 = let, 2 = middel, 3 = svær, og *(stjerne) = forsøgspersonen havde brug for assistance.

Ved interviewet er der benyttet et semistruktureret interview. Det indeholder kendetegn fra både struktureret og ustruktureret interviews ved både at benytte åbne og lukkede spørgsmål. Et overordnet sæt spørgsmål er udformet på forhånd med det formål at berøre de samme emner i hvert interview og ud fra dette forme samtalen i den retning interviewerens og den interviewede ønsker. Følgende spørgsmål indgik i interviewet efter en kort introduktion:

1. Hvordan er førstehåndsindtrykket?
2. Er det nemt at navigere i programmet?
3. Er velkomstskræmen overskuelig?
4. Modtages der den information du har brug for?
5. Er det tydeligt på velkomstskræmen hvad programmets formål er?
6. Er sproget på knapperne let at forstå?
7. Er layoutet passende programmet igennem?
8. Er der meget unødvendigt information i programmet?
9. Hvordan fungerede oprettelsen af labelled data i MAUDE?
10. Virkede det logisk at benytte to programmer?
11. Er der en tydelig rød tråd igennem programmet?
12. Har du nogle design ideer der kan bruges i dette program?

D.4 Resultater fra Incident Tracker

Ved analyse af data opnået ved denne usability test er der først dannet et overblik over data, som skiller sig ud. Dernæst er alt andet data analyseret. Denne sektion indeholder resultater opsamlet fra spørgeskema og interviews og afsluttes med en opsamling over de problemer, som blev synlige ved usability testen af Incident Tracker.

D.4.1 Spørgeskema

Forsøgsperson:	1	2	3	4	5	Gns.
Køn	M	K	K	K	K	
Alder	25	24	25	23	25	24,4
Opgave 1: Åbn Incident Tracker						
- Var opgaven klart defineret?	<i>ja</i>	<i>ja</i>	<i>nej</i>	<i>ja</i>	<i>ja</i>	
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	2	1	1	1	3	1,6
- Hvordan var navigationen i programmet?	2	2	1	1	3	1,8
Opgave 2: Opret en labelled MAUDE-fil						
- Var opgaven klart defineret?	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	1	1	2	1	2	1,4
- Hvordan var opgaven at fuldføre?	1	1	1	1	2	1,2
Opgave 3: Anvend den oprettede MAUDE-fil til at teste klassifieren i Weka						
- Var opgaven klart defineret?	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	1	1	1	1	2	1,2
- Hvordan var opgaven at fuldføre?	1	1	1	1	2	1,2
Opgave 4: Åbn en fil i Weka og klassificér den med Naive Bayes						
- Var opgaven klart defineret?	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>ja</i>	
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	2	3	3*	2	2	2,4
- Hvordan var opgaven at fuldføre?	2	3	3*	1	1	2,0

Tabel D.1: Resultatet fra spørgeskemaet der blev benyttet som et led af usability testen af Incident Tracker. Forsøgspersonerne blev bedt om at løse fire opgaver i testen og efterfølgende vurdere sværhedsgraden for hver opgave. Skalaen der blev benyttet var: 1 = let, 2 = middel, 3 = svær, og *(stjerne) = forsøgspersonen havde brug for assistance.

D.4.2 Interviews

En transkribering af de semistrukturerede interviews, foretaget i usability testen af Incident Tracker, er inkluderet nedenfor. De korresponderende spørgsmål er anført i D.3.3.

Forsøgsperson 1

1. Nogenlunde overskueligt. Man kunne gøre mere for at guide brugeren. Gør teksten læselig. Gør at man kan se, om man er inde i det ene eller det andet i AIMS- og MAUDE-systemerne. Arranger create labelled MAUDE-data og MAUDE-system knapperne anderledes.
2. Man kan hurtigt komme frem og tilbage, men det er ikke altid lige nemt, at finde ud af hvor man er. Man kan fare vild.
3. Velkomstkærmen er fin bortset fra arrangeringen af knapperne og den grå tekst.
4. Måske noget guide omkring hvilke filer, der skal findes. Hvis der findes en navngivningsstandard så brug denne til at finde filerne. Der er fin mængde information bortset fra de steder det mangler.
5. Sproget på knapperne er okay, simpelt og overskueligt
6. Fint layout gennem programmet. MAUDE er ret enkelt i og med, at der er en liste, som man skal gennemføre. Listen kan gøres kortere, hvis der er en filnavnsstandard. Så vælger man første fil og så finder/foreslår den selv resten eller hvad der nu giver mening.
7. Lidt unødvendigt information i programmet. Hvis det grå i velkomstmenuen er vigtigt burde det være læseligt. Universitetslogoerne burde være i Om-menuen.
8. Man kunne gøre MAUDE labelled data nemmere. Man kunne gøre sådan, at enter filename selv fjerner det skrevne. Step 5 og 6 kan sættes sammen i en dialogboks.
9. Logisk nok at bruge to programmer, hvis det ene program ikke understøtter oversættelse af data til det rette filformat. Færre programmer, færre trin og mere automatik. Færre klik er altid godt.
10. Navigation er kastet ind på samme måde alle steder, så det er godt, at de er på samme sted. Typisk er tilbageknappen og øvrig navigation i toppen. At man skal starte med knap nummer 4 og derefter nummer 2 virker ikke logisk. Trinene er fine. Dog lidt mærkeligt med trin 1 efterfølgende trin 1.
11. Navigation oppe i toppen. Drop logoer. Farven på teksten er ikke hensigtsmæssig. I loading-boksene må der godt stå, hvad man skal findes. Kan eventuel gøres i arkivformat. Det virker ret godt med tooltips.

Forsøgsperson 2

1. Simpel og overskueligt, få valgmuligheder, sigende navne, hvis man er inde for feltet.
2. Incident Tracker-delen var nemt og overskueligt, sværere i Weka og brug for assistance.
3. Ja, overskuelig.
4. Ja, knapperne er fine.
5. Manglende information i Weka. Incident Tracker er fin. Svært at gøre noget forkert.
6. Weka var uoverskueligt og kræver noget træning.
7. Der var en fin mængde information i Incident Tracker modsat Weka.
8. Labelled data oprettelsen var fin, men krævede lidt forklaring.
9. Virker uoverskueligt med brug af to programmer.

10. Rød tråd igennem Incident Tracker.
11. Måden at gemme filer på i Incident Tracker. Findes der ikke en standard?

Forsøgsperson 3

1. Rigtig flot. Måske ændre beskrivelsesteksten, da den er utydelig.
2. Ja.
3. Ja, overskuelig.
4. Knapperne er fine.
5. Ja.
6. Sproget på knapperne er lette at forstå.
7. Pænt layout.
8. Ingen unødvendig information. Måske gemme help-knappen væk.
9. Skrifttype skal ikke være større end selve navnet.
10. Weka var fin nok. Godt med dialogboks.
11. Logisk nok at benytte to programmer, selvom Weka er sværere at navigere rundt i.
12. Ikke helt rød tråd. Mangler info/hjælp til programmet (skal være i selve programmet).
13. Ikke rigtig. Fjern de tre prikker.

Forsøgsperson 4

1. Overskueligt. Fin med en indledning. Brugervenlig opdeling af AIMS og MAUDE.
2. Ja, nemt at navigere. Lidt problemer skyldes Mac. Lade åbn-dialogboksen starte i en foruddefineret mappe.
3. Overskuelig i Incident Tracker, men kræver lidt mere tilvænning i Weka.
4. Ja, formålet er klart.
5. Sproget på knapperne er gode og med forklaringer. Trin 1 til 7. Behøver ikke at stå forklaring på knapperne. Weka mindre godt, men ok.
6. Meget enkelt og overskueligt.
7. Nej, ikke noget, måske teksten i Load-knappen (jf. spørgsmål 5).
8. Fint.
9. Ja, logisk. Knappen til Weka skal dog virke.
10. Ja, okay.
11. Hvis man kender til opgaven. Lidt manglende info i startskærmen.
12. Kun det med knappen "Open file in Weka".

Forsøgsperson 5

1. Kedelige infomationstekst. Udelad det gerne.
2. Svært at navigere rundt. Der mangler information.
3. Velkomstskærmen er overskuelig, men kan arrangeres bedre. Læg infomations-teksten i bunden og henvist til et link. Mærkeligt med Weka-knappen i midten. Det bør stå for sig selv.
4. Mangler generelt information. Fine knappenavne, men der mangler tekstbeskrivelse.

5. Okay.
6. Create label-knapper er fine, men mangler info omkring, hvad man kan ved at klikke på MAUDE- og AIMS-knapperne.
7. Programmet skal kunne køre i fuldskærm. Bortset fra infomationsteksten er det overskueligt. Lidt for tomt/kedeligt på de andre sider.
8. Ja.
9. Åbn-boksene skal bare hedde "Åbn" og ikke selve beskrivelsen.
10. Nej, ulogisk.
11. Ikke tydelig, men okay. Savner mere information.
12. Tekst generelt for lille. Bør gøres større, så brugeren bedre kan læse, hvad der står.

D.4.3 Opsamling

Svarene opsamlet fra spørgeskemaet gav udtryk for at det største problem var i relation til at åbne en fil i Weka og klassificere denne. Alle forsøgspersoner kritiserede Wekas layout og kaldte det for uorganiseret og svært at finde rundt i. Endvidere var selve opgaven svær at udføre. Weka har et utal af funktioner inden for machine learning og kan benyttes til et væld af forskellige opgaver. Dette afspejler sig også i interfacet, hvilket er forvirrede blandt forsøgspersonerne. Svarene fra spørgeskemaet gav ligeledes udtryk for, at Incident Tracker havde et dårligt layout, hvor navigationen var svær at forstå.

Problem	Prioritering	Problem	Løsningforslag
1	Høj	Klassificering med Weka var svært, grundet utallige muligheder i Wekas brugergrænseflade	Inkorporer tekstklassificeringsfunktioner i programmet istedet.
2	Høj	Uoverskueligt at benytte to programmer.	Inkorporer tekstklassificeringsfunktioner i programmet istedet.
3	Høj	Brugeren af programmet føler sig ikke tilstrækkelig guidet i programmet.	Ændre layoutet, gør hjælpetekster mere tydelige, samt lav en udførlig skriftlig og videovejledning til programmet.
4	Høj	Der er for mange trin brugeren skal igennem før der er dannet en fil, der kan bruges til den videre klassificering.	Nedbring antallet af trin ved at benytte bla. gem-dialogbokse.
5	Mellem	Brugeren ved ikke altid hvilken fil der skal findes i vælg-dialogboksene	Indskriv i dialogboksen, hvilken fil brugeren skal søge efter.

Table D.2: Prioriteret oversigt over problemer der skal løses i Incident Tracker.

I interviewet kom det frem at knappernes rækkefølge ikke gav logisk mening og ikke lignede, hvad man kendte fra andre programmer. Desuden blev det klart at teksten, hvis mening var at hjælpe brugeren, ikke blev læst på grund af tekstens teksttype og farve. Tabel D.2 viser resultaterne af usability testen. Resultaterne bruges til at udvikle det nye system, hvor prioritetsvægtningen angiver alvorligheden af en løsning.

D.5 Design af det nye system

Baseret på usability testen af Incident Tracker har projektgruppen designet softwaren til at støtte brugeren endnu mere ved kun at inkorporere de funktioner, som brugeren kommer til at benytte ved automatisk tekstklassifikation og dermed fjerne afhængigheden af Weka. Endvidere bliver knapperne inddelt i klare grupperinger til hver funktion. Der ydes en indsats for at nedbringe antallet af trin, som brugeren skal igennem for at klassificere data, Desuden inkorporeres gem-dialogbokse. Der oprettes en brugermanual i programmet, som både guider brugeren ved hjælp af video og tekst. Endvidere kommer hjælpeteksten i programmet til at fremstå tydeligere, klarere og med forståelig simplicitet.

D.5.1 Forsøgsprotokol af det nye system

Usabilitytesten af Lo-Fi prototypen forløber med samme fremgangsmåde som den foregående, men med enkelte undtagelser. Da målsætningen med testen er at evaluere forbedringernes effektivitet og ikke at forbedre yderligere på systemet, er der kun foretaget spørgeskemaundersøgelse og et udokumenteret interview.

I spørgeskemaet vil fem forsøgspersoner blive bedt om at give en score omhandlende sværhedsgraden af opgaven imellem 1-3, hvor 1= let, 2= middel, 3= svær, samt * (stjerne) hvis forsøgspersonen havde brug for assistance. Udfra gennemsnitsresultatet af denne score er et kvantitativ resultat produceret af det nyudviklede system, hvilket benyttes som et succeskriterie for systemet.

Forsøgsudførelse og de givne instruktioner er sammenlignelige med sektion D.3.1 og D.3.2.

D.5.1.1 Resultater fra usability test af Lo-Fi prototypen for det nye system

Resultater opsamlet ved usabilitytesten af Lo-Fi prototypen er angivet i Tabel D.3. Samtlige forsøgspersoner nævnte i interviewet, at systemet var enkelt at bruge. Layout og navigation var simpelt og logisk og opfyldte systemets krav til fulde. Endvidere blev fanebladsopbygningen fremhævet.

Forsøgsperson:	1	2	3	4	5	Gns.
Køn	M	M	K	M	K	
Alder	26	24	22	23	21	23,3
Opgave 1: Åbn UTH-Finder						
- Hvordan var layoutet at forstå?	1	1	1	1	1	1,0
- Hvordan var navigationen i programmet?	1	1	1	1	1	1,0
Opgave 2: Indlæs træningsæt og testsæt						
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	1	1	1	1	2	1,2
- Hvordan var opgaven at fuldføre?	1	1	1	1	1	1,0
Opgave 3: Vælg a priori sandsynlighed						
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	1	1	1	1	1	1,0
- Hvordan var opgaven at fuldføre?	1	1	2	1	1	1,2
Opgave 4: Test og gem resultatet						
- Hvordan var layoutet at forstå i forhold til at udføre opgaven?	1	1	1	1	1	1,0
- Hvordan var opgaven at fuldføre?	1	1	1	1	1	1,0

Tabel D.3: Resultatet fra spørgeskemaet, der blev benyttet som et led af usability testen af det nye system. Forsøgspersonerne blev bedt om at løse fire opgaver i testen og efterfølgende vurdere sværhedsgraden for hver opgave. Skalaen der blev benyttet var: 1 = let, 2 = middel, 3 = svær, og *(stjerne) = forsøgspersonen havde brug for assistance

D.5.2 Opsamling

Spørgeskemaundersøgelsen viste at systemet layoutmæssigt var fyldestgørende for forsøgspersonerne. Endvidere var systemets funktioner til klassifikation af utilsigtede hændelser logiske og simple at bruge.



Forsøgsprotokol for systemvalidering

Det følgende beskriver udførelsen af forsøgene til systemvalideringen herunder formål, præprocessering, metode og resultater. I dette afsnit præsenteres kun rå data, mens det behandlede data dokumenteres i kapitel 8.

E.1 Formål

Formålet med forsøget er at evaluere systemets anvendelighed og effektivitet ved klassificering af IT-relaterede til IT utilsigtede hændelser. Endvidere vil det blive testet, hvilken indflydelse træningssettets størrelse har på systemets ydelse.

E.2 Præprocessering

Til evaluering af systemet blev der anvendt rapporter fra DPSD. Der blev foretaget et udtræk fra databasen af samtlige rapporter indberettet fra regionerne i perioden 01-01-2011 til 30-04-2012, hvilket svarer til 31.184 rapporter. En rapport indeholder følgende 22 parametre, hvor hver parameter svarer til et felt i en tabel:

- Overskrift
- Hændelsesbeskrivelse
- Lokalitet
- Sygehus
- Afdeling
- Postnummer
- Følger for patienten
- Alvorligheden af hændelsen
- Forslag til forebyggelse
- Patientens navn
- Patientens CPR-nummer
- Patientens køn
- Patientens alder
- Patientstatus
- Flere patienter involveret
- Medicin
- Medicinsk udstyr
- Hjælpemidler
- Rapportørens navn
- Rapportørens email
- Rapportørens telefonnummer

- Rapportørens faggruppe
- Vedhæftet fil

Parametrene er en blanding af strukturerede felter og fritekstfelter. Det vurderes at fritekstfelterne er de eneste parametre, som kan indgå i klassifikationen, hvorfor følgende parametre er blevet udvalgt:

- Overskrift
- Hændelsesbeskrivelse
- Forslag til forebyggelse

For at kunne identificere rapporterne blev sagsnummer og oprettelsesdato medtaget, men indgik ikke i klassifikationen. De udleverede rapporter var gemt i .csv filformatet. For at undgå fejl ved deling af parametrene blev samtlige kommaer fjernet således, at en rapport blev indlæst som en samlet tekststreng. Ligeledes blev alle skilletegn, matematiske tegn, numeriske tegn og specielle karakterer fjernet. Ord med to eller færre karakterer blev vurderet som værende fyldord og blev ligeledes fjernet. I Tabel E.1 ses et fiktivt eksempel på en rapport med de valgte parametre, som er klar til at blive klassificeret.

Sagsnummer	Oprettelsesdato	Overskrift	Forslag til forbedring	Hændelsesbeskrivelse
30389	10/08/11	epj nede	system der virker	epj gik ned det utænkelige som ikke kan ske nemlig den skrivebeskyttede del også gik ned var faktum

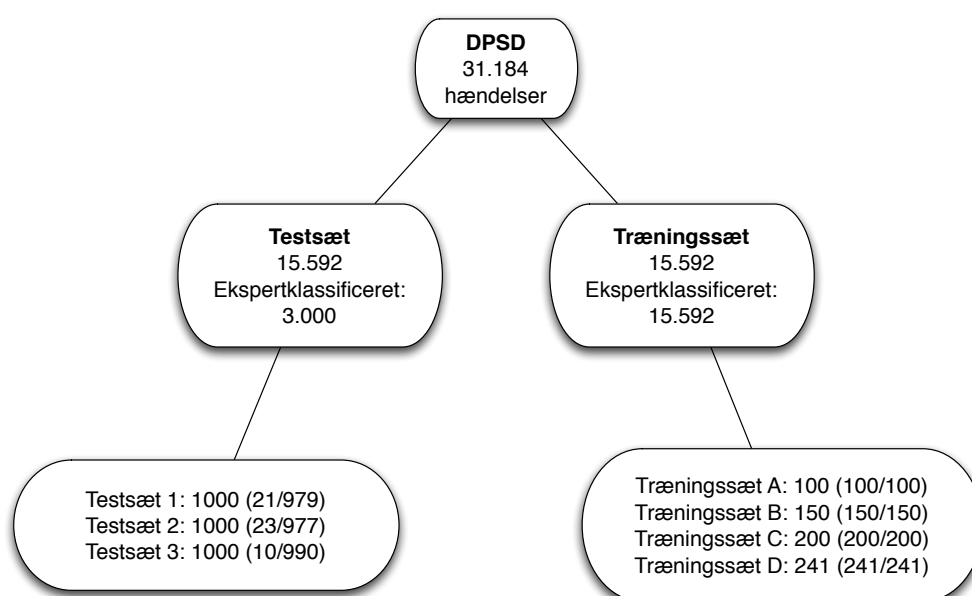
Tabel E.1: *Opdigtet sample af en IT-relateret hændelse.*

E.3 Fremstilling af trænings- og testsæt

Til klassifikationen blev Naive Bayes algoritmen anvendt, hvilket er en supervised learning algoritme, som kræver forudgående træning før klassifikation kan foretages. Træningssettet blev fremstillet fra halvdelen af de tilgængelige hændelser svarende til 15.592 rapporter. Samtlige rapporter i træningssættet blev manuelt klassificeret i de to kategorier, IT og andet. 241 rapporter blev klassificeret som værende relateret til IT svarende til en forekomst på 1,5 %. Systemet blev trænet med balancerede træningsset indeholdende henholdsvis 100, 150, 200 og 241 IT-relaterede hændelser. Det skal således forstås at eksempelvis et træningsset på 100 IT-relaterede hændelser også indeholder 100 andre hændelser, som ikke er relateret til IT. IT-delen af træningssettene blev dannet ved et randomiseret udtræk fra de 241 IT-klassificerede hændelser. Således blev der ved træningssættet indeholdende 100 IT-klassificerede hændelser udtrukket 100 tilfældige hændelser fra det samlede antal IT-klassificerede hændelser. Den anden del af træningssættene indeholdende andre hændelser, som ikke var relateret til IT,

blev dannet ved et randomiseret udtræk af 15.351 hændelser. Det samme gør sig gældende for træningssættene med 150, 200 og 241 IT-relaterede hændelser, se Figur E.1. Det er derfor sandsynligt at samme hændelserrapport kan indgå i flere træningssæt.

Testsættet blev sammensat af de resterende 15.592 rapporter. Der blev sammensat tre testsæt af 1000 rapporter benævnt som A, B og C. Disse blev sammensat ved et randomiseret udtræk, hvilket betyder at samme rapport kunne indgå i mere end et testsæt, se figur E.1. For at kunne evaluere systemets ydelse, blev samtlige 3000 rapporter fra de tre testsæt klassificeret og tildelt en label svarende til den aktuelle kategori. Den tildelte label indgik ikke i klassificeringen, men blev udelukkende brugt til den efterfølgende evaluering.



Figur E.1: *Det udleverede data opdeles i to dele, hvor den ene del bruges til at udarbejde træningsdata og den anden til testdata. Alt data i første del ekspertklassificeres og bruges derefter til et balanceret træningssæt med 50 % IT-relaterede hændelser og 50 % andre hændelser. Ud fra træningssættet udtrækkes fire træningssæt med stigende samplestørrelse. Den anden halvdel benyttes til testsæt. Heraf er tre randomiserede testsæt udtrukket, som hver er blevet ekspertklassificeret. De udarbejdede trænings- og testsæt benyttes til at finde læringskurve og ydeevne.*

E.4 Klassificering

Klassificeringen blev foretaget ved at træne systemet med de fremstillede træningssæt og efterfølgende indlæse testsættet som ønskes klassificeret. Da træningssættene er balancerede sættes a priori sandsynlighed til 50 %, hvilket også er systemets prædefinerede værdi. For hvert indlæst træningssæt blev alle testsæt klassificeret. Efter endt klassificering blev resultatet gemt i en tekstfil og efterfølgende indlæst i Microsoft Excel for videre behandling.

E.5 Resultater

Ved hvert testsæt blev systemet trænet med henholdsvis sæt træningssæt A, B, C og D, hvorfor der for hvert testsæt er opstillet fire confusionmatricer.

E.5.1 Testsæt A

Tabel E.2 viser fordelingen af hændelser i testsæt A:

IT	Andet	Total
21	979	1000

Tabel E.2: Kategorifordeling for testsæt A.

Nedenstående tabeller E.3, E.4, E.5 og E.6 viser resultater for klassificering af testsæt A ved anvendelse af balancerede træningssæt med henholdsvis 100, 150, 200 og 241 IT-relaterede hændelser:

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	12	66
	Andet	9	913

Tabel E.3: Resultat for testsæt A klassificeret med et balanceret træningssæt på 100 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	20	177
	Andet	1	802

Tabel E.4: Resultat for testsæt A klassificeret med et balanceret træningssæt på 150 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	15	108
	Andet	6	871

Tabel E.5: Resultat for testsæt A klassificeret med et balanceret træningssæt på 200 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	13	48
	Andet	8	931

Tabel E.6: Resultat for testsæt A klassificeret med et balanceret trænings sæt på 241 IT-relaterede hændelser opstillet i confusionmatrice.

E.5.2 Testsæt B

Tabel E.7 viser fordelingen af hændelser i testsæt B:

IT	Andet	Total
23	977	1000

Tabel E.7: Kategorifordeling for testsæt B.

Nedenstående tabeller E.8, E.9, E.10 og E.11 viser resultater for klassificering af testsæt B ved anvendelse af balancerede trænings sæt med henholdsvis 100, 150, 200 og 241 IT-relaterede hændelser:

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	17	52
	Andet	6	925

Tabel E.8: Resultat for testsæt B klassificeret med et balanceret trænings sæt på 100 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	21	195
	Andet	2	782

Tabel E.9: Resultat for testsæt B klassificeret med et balanceret trænings sæt på 150 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	19	108
	Andet	4	869

Tabel E.10: Resultat for testsæt B klassificeret med et balanceret træningssæt på 200 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	19	44
	Andet	4	933

Tabel E.11: Resultat for testsæt B klassificeret med et balanceret træningssæt på 241 IT-relaterede hændelser opstillet i confusionmatrice.

E.5.3 Testsæt C

Tabel E.12 viser fordelingen af hændelser i testsæt C:

IT	Andet	Total
10	990	1000

Tabel E.12: Kategorifordeling for testsæt C.

Nedenstående tabeller E.13, E.14, E.15 og E.16 viser resultater for klassificering af testsæt C ved anvendelse af balancerede træningssæt med henholdsvis 100, 150, 200 og 241 IT-relaterede hændelser:

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	8	61
	Andet	2	929

Tabel E.13: Resultat for testsæt C klassificeret med et balanceret træningssæt på 100 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	10	203
	Andet	0	787

Tabel E.14: Resultat for testsæt C klassificeret med et balanceret træningssæt på 150 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	10	109
	Andet	0	881

Tabel E.15: Resultat for testsæt C klassificeret med et balanceret træningssæt på 200 IT-relaterede hændelser opstillet i confusionmatrice.

		Ekspertklassificeret	
		IT	Andet
Algoritmeklassificeret	IT	10	46
	Andet	0	944

Tabel E.16: Resultat for testsæt C klassificeret med et balanceret træningssæt på 241 IT-relaterede hændelser opstillet i confusionmatrice.

E.6 Databehandling

Evalueringen af systemet tager udgangspunkt i følgende tests:

- Læringstest
- Effektivitetstest

Forud for databehandlingen er det opnåede resultater opstillet i en confusionsmatrice. I confusionsmatricen kan der ske en direkte aflæsning af sand positiv (TP) og sand negativ (TN) værdier, samt type 1 (FP) og type 2 (FN) fejl. I Tabel E.17, ses et eksempel på en confusion matrice.

		Ekspertklassificeret		Total
		IT	Andet	
Algoritmeklassificeret	IT	TP	FP	$TP + FP$
	Andet	FN	TN	$FN + TN$
Total		$TP + FP$	$FN + TN$	N

Tabel E.17: Eksempel på confusionmatrice, som bruges til at bestemme træningssættets indflydelse på antallet af korrekt klassificerede rapporter. Den procentvise andel af korrekt klassificerede rapporter bestemmes som $TP/(TP+FN)$.

E.6.1 Læringstest

Formålet med læringstesten var at undersøge størrelsen på træningssættets indflydelse på antallet af korrekt klassificerede hændelser. Således undersøges det hvor stor en procentdel af de ekspertklassificerede IT-hændelser, der blev klassificeret ved anvendelse af den automatiske klassifikation, svarende til antallet af sand positive (TP), se Tabel E.17. Den procentvise andel findes ved følgende formel:

$$\text{Sensitivitet} = \frac{TP}{(TP + FN)} \cdot 100 \quad (\text{E.1})$$

E.6.2 Effektivitetstest

Effektivitetstesten af systemet har til formål at undersøge systemets ydeevne. Der udføres følgende tests:

Nøjagtighed

Nøjagtigheden er et statistisk mål for, hvor god systemet er til at klassificere hændelserne korrekt. Således testes det om systemet korrekt klassificerer en IT-hændelse (TP) eller en hændelse tilhørende kategorien andet (TN). Nøjagtigheden findes med følgende formel:

$$\text{Nøjagtighed} = \frac{TP + TN}{(TP + FP + FN + TN)} \quad (\text{E.2})$$

Præcision

Præcisionen er et mål for, hvor godt et system er til finde de sandt positive (TP) hændelser. Bliver der fundet mange falsk positive hændelser falder systemets præcision. Systemets præcision findes med følgende formel:

$$\text{Præcision} = \frac{TP}{(TP + FP)} \quad (\text{E.3})$$

Sensitivitet

Systemets sensitivitet viser, hvor godt systemet er til at finde de sandt positive (TP) rapporter. Således vil en lav sensitivitet være ensbetydende med mange oversete IT-relaterede hændelser. Sensitiviteten findes på følgende måde:

$$\text{Sensitivitet} = \frac{TP}{(TP + FN)} \quad (\text{E.4})$$

Specificitet

Specificiteten angiver systemets evne til at frasortere hændelser, som ikke er relateret til IT og findes med følgende formel:

$$\text{Specificitet} = \frac{TN}{(TN + FP)} \quad (\text{E.5})$$

F-score

F-score er et kombineret mål for præcision og sensitivitet, der ofte benyttes til søgealgoritmer, klassifikation og machine learning. Det bruges til teste nøjagtigheden.

F-scoren findes som:

$$\text{F-score} = \frac{(\textit{precision} \cdot \textit{sensitivitet})}{(\textit{precision} + \textit{sensitivitet})} \quad (\text{E.6})$$

