NY DOKUMENTATIONS-PRAKSIS VED BRUG AF MACHINE LEARNING

PROJEKT OMSORG OG NÆRVÆR I ÆLDREPLEJEN

THISTED KOMMUNE

Evaluering af Machine Learning

Projekt Styrket omsorg og nærvær i ældreplejen 2021-2022.

Rapporten er udarbejdet af:

Mette Rægaard Lindemann Larsen, projektleder og digitaliseringskonsulent, Thisted Kommune.

Projektteamet i Systematic og Thisted Kommune

Miriam Lerkenfeld, Innovationskonsulent, Hard Work, Mind & Magic

Illustrationer: Aleksander Weis Klinke

Udgivet august 2022.

INDHOLD

TEKNISK DOKUMENTATION

Baggrund

I januar 2021 igangsatte Thisted Kommune projektet "Styrket omsorg og nærvær i ældreplejen via øget rehabilitering og intelligent teknologi." Projektet er blevet finansieret af puljemidler fra Sundhedsstyrelsen til at styrke omsorg og nærvær i ældreplejen. Projektets formål har været at undersøge om kommunen kan frigive tid til omsorg og nærvær i ældreplejen ved at tilvejebringe øget kvalitet i dokumentationen og en mere effektiv tilrettelæggelse af den samlede rehabiliteringsindsats i Thisted Kommune.

I denne rapport evalueres om machine learning er en teknologi, der har et anvendelsespotentiale i omsorgsjournalsystemet Cura. Systematic er leverandør af Cura, og har deltaget i projektet via et OPI-samarbejde med Thisted Kommune. Der er lavet flere evalueringer, som er tilgængelige på www.fritældreliv.dk

Resume

Thisted kommune har i et innovationsprojekt undersøgt og afprøvet, hvordan en arbejdsgang for dokumentation kan forbedres med machine learning. Målet er at få en mere brugervenlig dokumentationspraksis og samtidig højne kvaliteten i dokumentationen. I denne rapport præsenteres det udviklede machine learning-koncept og samt resultaterne fra innovationsprojektet.

Udfordringer

I foråret 2021 blev der i projektet lavet observationer og interviews med medarbejdere, for at undersøge, hvordan medarbejderne oplevede den eksisterende dokumentationspraksis. Fundene i feltarbejdet blev samlet i en indsigtsrapport, der kan tilgås på www.fritældreliv.dk.

Medarbejderne oplevede arbejdsgangen for dokumentation som tidskrævende og besværlig, fordi der skal skrives flere steder. De havde særligt svært ved at få et overblik i journalen.

For det meste ligger der relevant viden under observationskort - men det sker også, at en observation er skrevet i et forkert kort. Jeg bliver derfor nødt til at læse alle observationskort igennem for at være sikker på, at jeg har fået det hele med. Det betyder, at jeg læser en masse irrelevant
- Terapeut

Udfordringen blev bekræftet i en journalaudit, der blev lavet i efteråret 2021. Her blev der gennemgået 10 tilfældige journaler og i 7 ud af 10 tilfælde var der dokumenteret i de forkerte observationskort.

Nationalt bliver der stillet krav til dokumentationen, da alle kommuner skal dokumentere efter metoden Fællessprog III (herefter FSIII). Målet med FSIII er at få mere struktur i data, så dokumentation bliver nemmere at finde og få et overblik over. Det betyder også, at dele af dokumentationen skal skrives flere steder, hvilket betyder dokumentationen opleves mere fragmenteret. Den tidligere nævnte journalaudit konkluderede, at FSIII-metoden ikke altid følges i kommunen.

Der er 91 tilgængelige observationskort for ældreområdet, som Thisted Kommunes hjemmeplejere kan vælge at dokumentere i, i Cura. Observationstyperne knytter sig til såkaldte ”tilstande” jf. FSIII. 21 af observationskortene er defineret af FSIII og kan ikke fjernes. Årsagen til der er 70 mere, er at der skal være observationskort til fx blodsukkermålinger, vægt, kontakt til læge osv. Ved at det er mere specifikt, kan dokumentationen også bedre målrettes. På den ene side, er det derfor godt at have mange observationskort. Udfordringen ved at have mange er, at det er svært for medarbejderne at finde og vælge de korrekte observationskort i løsningen.

En anden central udfordring er at en stor del af medarbejderne i hjemmeplejen, i forskellig grad, har svært ved at læse og skrive. Det betyder, at de bruger langt tid på at finde og skrive dokumentation.

På baggrund af de indledende analyser blev det tydeligt, at der er et behov for at gentænke dokumentationspraksissen i ældreplejen. I projektet har vi derfor undersøgt, hvordan Thisted kommune kan understøtte medarbejderne langt bedre i dokumentationen og følge FSIII, og samtidig højne kvaliteten, så det også efterfølgende bliver nemmere at genfinde relevant information i omsorgsjournalen.

Citater fra feltarbejdet

Jeg kan ikke se, hvilke dokumentationsmuligheder, der er under de enkelte observationskort - f.eks. ville dokumentere brandfare hos en borger med demens, men jeg var i tvivl om, hvor det skulle ligge. Er det en observation til hukommelse eller ejendom?
- Hjemmeplejer

Jeg er nervøs for ikke at skrive det rigtige, de rigtige steder
- Hjemmeplejer

Jeg bruger funktionsniveau/egenomsorg som det primære observationskort - det er ikke rigtigt, men det er nemt
- Hjemmeplejer

FSIII er skyld i, at vi har 12 observationsskabeloner
- Visitator

Hvad er afprøvet?

Formålet har været at prøve og forbedre dokumentationspraksissen, i omsorgsjournalsystemet Cura, ved hjælp af machine learning. Machine Learning er en teknologi, hvor computeren trænes til at finde regler og sammenhænge i et større datasæt. Når computeren kan genkende mønstre, kan den komme med anbefalinger til, hvad der er det korrekte at gøre i forskellige scenarier. Baseret på udfordringerne beskrevet i feltarbejdet, har fokus for dette projekt været at undersøge om machine learning kan bidrage til følgende:

- Gøre det lettere at dokumentere et besøg i de korrekte observationstyper, så der arbejdes efter FSIII-metoden.
- Hjælpe hjemmeplejerne med at dokumentere kort og præcist.
- Optimere dokumentationsprocessen så hjemmeplejeren kan spare tid (og frustration) og hermed sikre at der dokumenteres oftere og i bedre kvalitet.

Arbejdsgangsbeskrivelse - nuværende og fremtidig

Når hjemmeplejeren observerer ændringer i tilstanden hos borgeren, og det afviger fra det normale besøg, så skal dette dokumenteres i et observationskort. Observationskortet er et notat, der lægges i journalen, og som kan genfindes af den næste medarbejder, og/eller sendes som opgaver til andre medarbejdere. Praksis er i dag, at medarbejderne dokumenterer deres observationer efter besøget hos borgeren ved at vælge en specifik observation, som de finder ved at søge på navnet på observationen. Herefter udfyldes observationen med de nødvendige informationer. Hvis man ønsker at have lignende information i en anden observationskort gentages processen, så der eksempelvis registreres information i observationer som 'Fald' og 'Smerter'.

Ideen med at benytte machine learning er, at medarbejderen i stedet skriver i et fritekstfelt. På baggrund af friteksten, foreslår systemet, hvor observationerne skal placeres. Formålet er at arbejdsgangen bliver nemmere, og at observationskortene bliver placeret i mere korrekte observationskort, så kvaliteten øges og dokumentation efterfølgende nemmere kan genfindes. Ideen er afbilledet illustrativt i nedenstående.

Dokumentation med og uden machine learning

Figur 1 viser, hvordan dokumentationen foregår i dag og med kunstig intelligens/machine learning. Prikkerne er dokumentationen, der skal placeres det rigtige sted i journalen. Venstre side viser, at prikkerne ikke kommer i de rigtige mapper, og at det derfor bliver svært at genfinde. Højre side af figuren viser, at dokumentationen vha. machine learning kommer i de rigtige mapper, og at det bliver nemmere at genfinde relevant dokumentation.

Dokumentation i dag med manuelt valg

Fremtidig dokumentation med kunstig intelligens

Nuværende arbejdsgang

uden machine learning for hjemmepleje

1. Medarbejder åbner Cura.
2. I Cura trykkes på et + for at lave en ny observation.
3. Medarbejderne skal søge efter den relevante observation. Der 91 observationer og navnet på observationen skal skrives korrekt, for at findes frem.
4. Medarbejder får observationskort frem
5. Medarbejderne indtaster dokumentation
6. Medarbejder gemmer/sender observation

Hvis medarbejder skal lave endnu en observation gentages arbejdsgangen

Outcome: observationer ligger hovedsageligt i en lille gruppe af observationer og er ikke inddelt korrekt. Det er svært at genfinde relevant information

Fremtidig arbejdsgang

ved brug af machine learning for hjemmepleje

1. Medarbejder åbner Cura
2. Medarbejder trykker på + for at lave nye observationer
3. Medarbejder skriver observation i feltet
4. Baseret på input forslår Cura 5 observationer, som kan vælges med mulighed for at vælge flere forslag.
5. Medarbejder godkender placeringer af tekst på 1 eller flere relevante observationer
6. Observationer gemmes/sendes

Outcome: Observationerne bliver spredt ud på mere korrekt steder, hvilket øger den faglige refleksion og skaber et bedre overblik, når dokumentationen skal anvendes igen.

Brugertest

Metode

I løbet af 6 måneder blev der arbejdet intensivt med det koncept, der vil blive beskrevet nærmere i nedenstående. Konceptet er blevet udviklet i samarbejde med medarbejdere hjemmeplejen.

Der blev afholdt 6 workshops. Til workshops deltog medarbejdere fra hjemmeplejen, udviklerteamet, en Cura ekspert samt Thisted Kommunes projektleder.

Deltagerne på workshopsene har været forskellige over de første workshops for senere at have de samme deltagere med, således at de bedre kunne følge udviklingen af konceptet og dermed kunne reflektere bedre over det. Deltagerne har været et repræsentativt udsnit af de ansatte i hjemmeplejen med forskel i alder, etnicitet, erfaring osv.

De første to workshops blev udført på en måde, hvor der var forberedt en række forskellige mockups (skærmbilleder lavet i et designprogram) samt interviewguides. Deltagerne blev guidet igennem forskellige scenarier og stillet spørgsmål i relation til deres forståelse af konceptet, arbejdsgangen og brugergrænsefladerne for at designeren kunne spore sig ind på det mest intuitive design, der bedst understøttede arbejdsflowet i deltagernes hverdag.

Den efterfølgende workshop bestod af to dele, hvoraf den første del bygger på principper fra Participartory Design, hvor deltagerne skulle forestille sig og visualisere et fremtidsscenarie ved hjælp af forskellige værktøjer. Anden del bestod i en test at den første version af prototype-appen, hvor en tidlig version af machine learning-motoren til at foreslå observationer blev testet. På baggrund af tilbagemeldingerne fra medarbejderne blev designet videreudviklet, testet og tilpasset i de efterfølgende workshops.

Til sidst blev appen driftstestet hos medarbejdere i hjemmeplejen.

Konceptbeskrivelse

Begreber brugt i den følgende tekst:

Tabel 1.12

Term	Betydning
Observation	Dokumentationstype i Cura som foretages efter besøg. Denne kan være knyttet til en ydelse.
Korrekt observation	Teksten i observationen skal være relevant for den valgte observation. Dokumentationen skal være kort og præcis, og skal ikke forholde sig til andet indhold end den valgte observation; medmindre dette er gjort eksplicit.
God kvalitet i observationer	Teksten skal være dokumenteret i den/ de korrekte observation(er), og teksten skal være kort, præcis og forståelig.
Prototype-appen	Der blev udviklet en projekt-prototypen i projektperioden, der vil blive henvist til som prototype-appen. Det er en wep-app, der er brugt til at teste konceptet og afprøve brugergrænsefladen.

Fra feltarbejde, hjemmeplejer anvender tablet i bilen, marts 2021

Dokumentation i én proces efter besøg

Konceptet bygger på, at brugeren skal have én fælles indgang til dokumentation efter et besøg. Brugeren noterer alt, som de vil dokumentere i fritekst; enten via talegenkendelse eller ved indtastning af tekst. Prototypen, som blev udviklet i projektperioden, fokuserede udelukkende på tekst. Havde der været mere tid i projektet, ville der også være blevet arbejdet videre med indtaling af tekst. Dette blev testet i begyndelsen af projektet, men blev dog fraprioriteret pga. manglende ressourcer. I den udviklede prototype kan der derfor kun indtastes tekst.

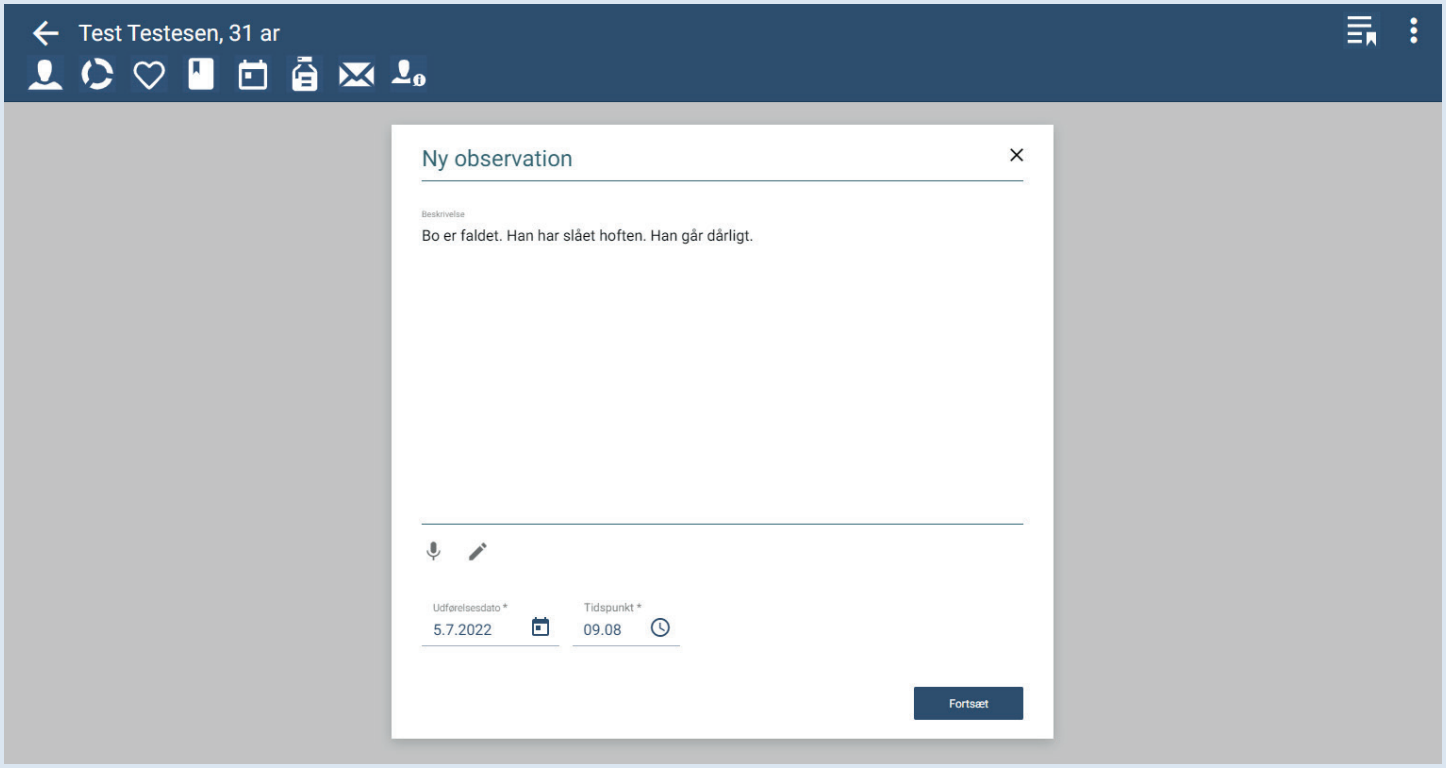
Efter indtastning af teksten, får brugeren forskellige forslag til, hvor teksten kan placeres i relevante observationstyper herefter får brugeren forskellige forslag til, hvor teksten kan placeres i relevante observationstyper. Denne proces bygger på en maskinlæringsteknologi kaldet natural language processing (NLP).

Understøttelse af valg af korrekt observationstype(r)

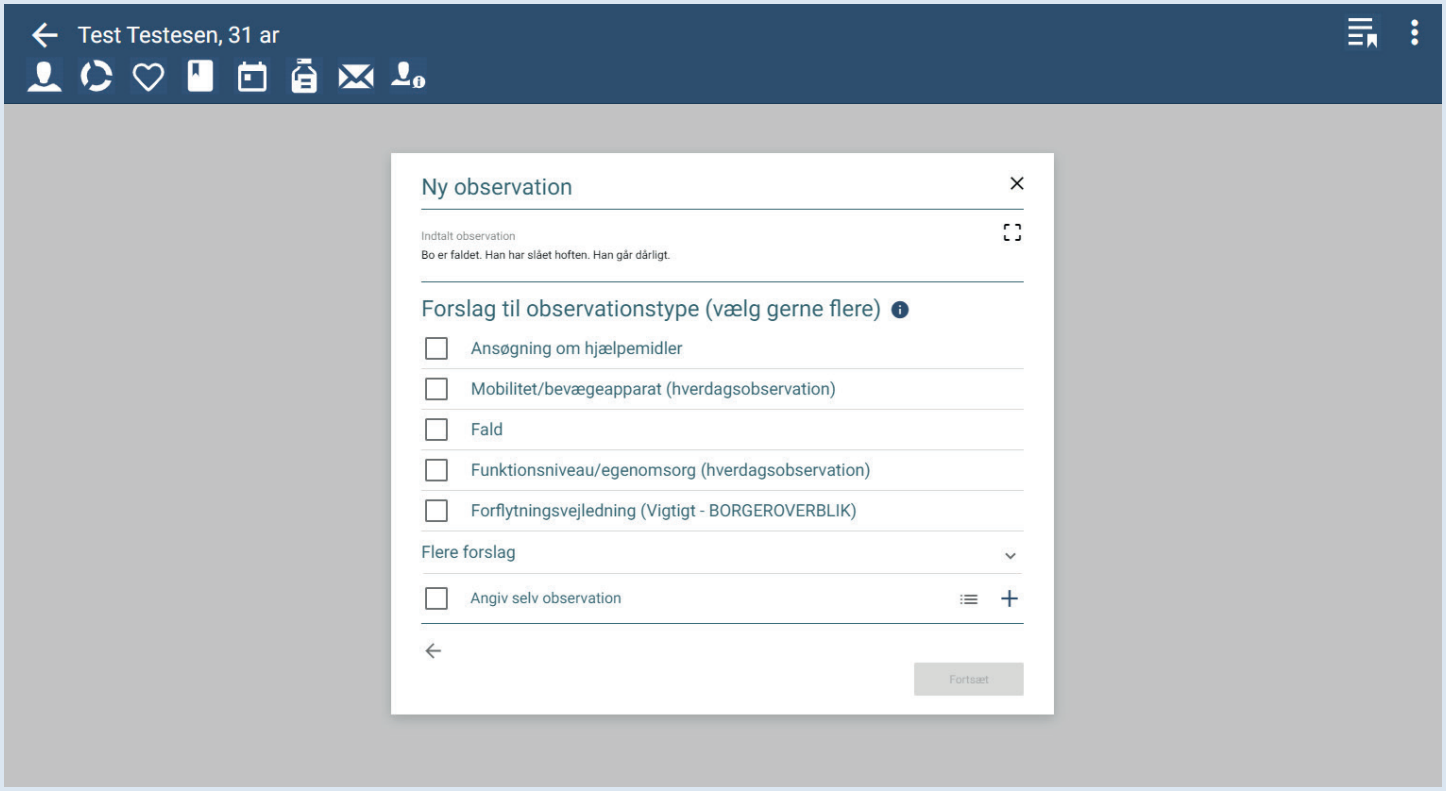
Efter at have beskrevet alt, som brugeren ønsker at dokumentere i fritekstfeltet bliver de præsenteret for en række forslag til observationstyper, som teksten kan placeres i. Forslagene er baseret på en NLP-maskine, som sammenligner brugerens tekst i forhold til historiske data, som allerede eksisterer i systemet i forskellige observationstyper. Dette betyder, at de historiske data danner grundlag for, hvordan systemet foreslår relevante observationstyper.

Hvis systemet på sigt integreres i Cura vil det tilpasse sig brugeres historiske og fremtidige data. Således vil systemet udvikle sig, hvis brugerne for eksempel begynder at bruge nye observationstyper; systemet vil lære dette og tilpasse sig, så denne givne observationstype vil blive forslået oftere.

Brugerstudierne viste, at dokumentation som fritekst understøtter de eksisterende arbejdsprocesser og behov hos hjemmeplejerne. Ligeledes var muligheden for at kunne vælge flere observationstyper til den samme tekst direkte efterspurgt fra flere brugere, og blev vel modtaget i brugertestene. Forslagene understøttede hjemmeplejeren til at reflektere over valg af observationstype i brugertestene. En lang række deltagere i brugertestene fremhævede, at systemet var hjælpsomt og brugervenligt – alene på baggrund af den refleksion, forslagene satte i gang, og at det blev nemmere at vælge flere observationer på én gang. Projektet betragter dette, som en kvalitets-forbedring af dokumentationen, samt en forbedring af brugeroplevelsen med Cura.



Skærmbillede 1.I dette eksempel indtaster brugeren alt, som vedkommende ønsker at dokumentere i fritekst. Skærmbilledet er fra prototype-appen.



Skærmbillede 2. Systemet præsenterer forslag til observationstyper, hvor teksten skal dokumenteres. ud fra friteksten. Skærmbilledet er fra prototype-appen.

Understøttelse af korrekt dokumentation i observationer

Efter brugeren har valgt en eller flere observationstyper, præsenteres brugeren for skærmbillede 3.

Systemet benytter en anden NLP-motor til at foreslå hvilke sætninger, som hører til hvilke observationer. Brugerstudiet viser, at den anvendte farvekodning virker umiddelbar intuitiv for brugeren. Brugeren kan vælge at tilpasse fremhævningen i teksten for at definere præcist, hvad der skal med i den givne observation. Brugeren trykker på hele sætninger, og hele sætningen farvemarkeres (se eksempel på skærmbillede 3 på side 17).

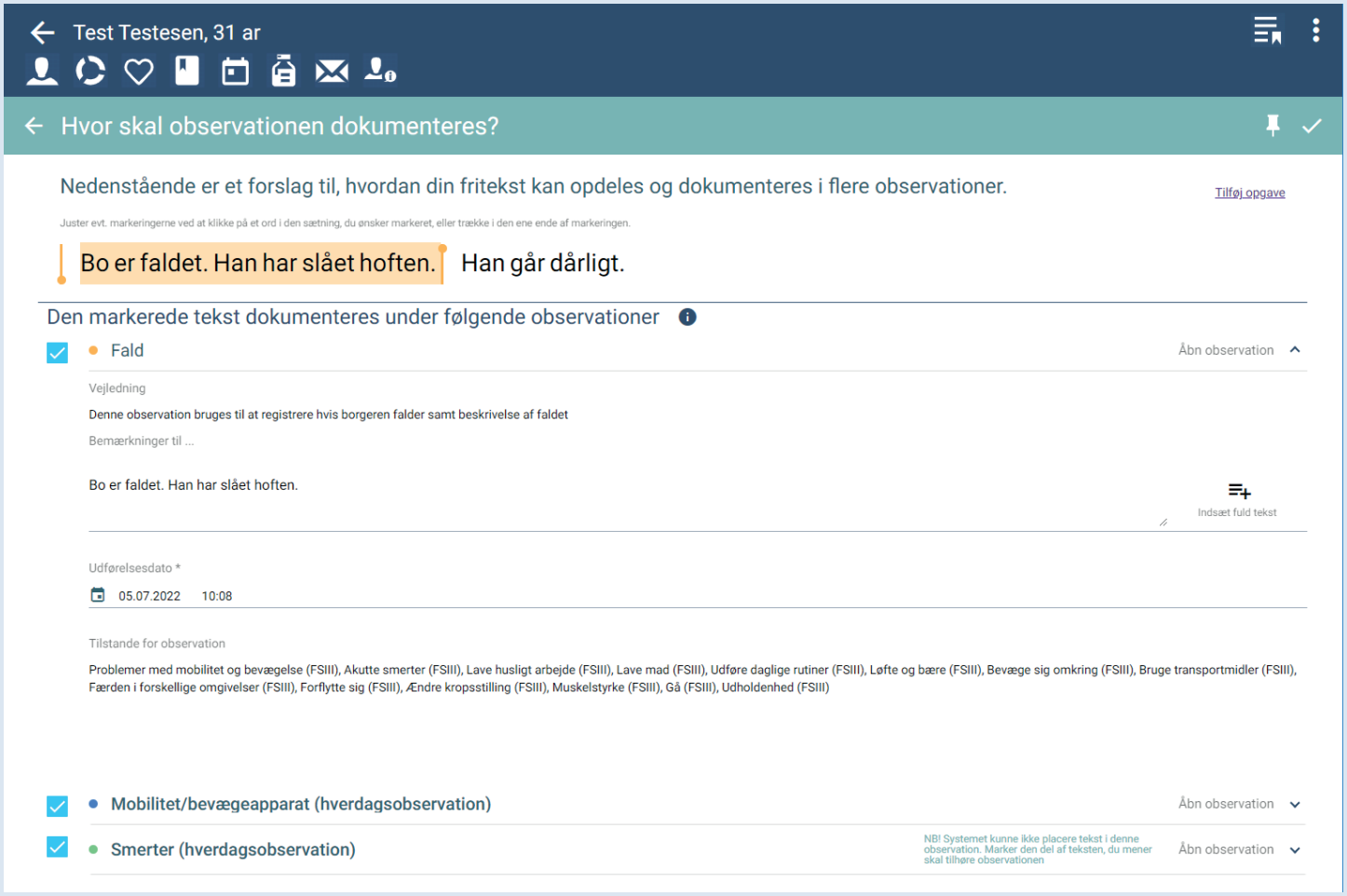
Inden da var det forsøgt at brugeren kunne "trække" i teksten. Dette fungerede dog ikke, og var meget besværligt at gøre på mobile enheder. På trods af at der er en markant forbedring, med den nye markeringsfunktionalitet, mente mange af brugerne, at farvemarkeringsfunktionaliteten kunne være bedre. Det er også en udfordring, at der markeres i hele sætninger, da ikke alle i hjemmeplejen, er stærk i korrekt tegnsætning. Udfordringen er, at hvis medarbejderne ikke bruger punktummer, vil det ikke være muligt for medarbejderen at opdele teksten efterfølgende. Der bør derfor arbejdes videre med farvemarkeringen, så det bliver mere brugervenligt og intuitivt. I den forbindelse vil det desuden være relevant at arbejde med at hjælpe brugerne til korrekt tegnsætning, eksempelvis via machine learning.

Overblik over tekst i de forskellige observationer

Som der fremgår af skærmbillede 4, kan man til sidst få et overblik over teksterne, og hvilke observationer de er blevet placeret i.

Det bliver mere præcist, når man skriver under én, og man taber ikke noget, og der er den samme historie. Vi er mange gange usikre på, under hvad vi skal skrive det. [Det nye koncept] giver bedre kvalitet for borgeren.
- Hjemmeplejer

Prototypen bruger kun fritekst-feltet for alle observationer, og den placerer alt tekst i dette felt. Konceptet inkluderer dialogbokse for de rette observationer, og for fremtidigt arbejde kunne man kigge på at få placeret de specifikke tekster i de relaterede felter i observationen. Herved guides brugeren gennem udfyldning af felterne i observationen. Nedenstående skitserer hvorledes dette kunne udfoldes. Alt, som systemet placerer i observationen, er fremhævet med gult. Ligeledes er de felter, som endnu ikke er udfyldt, men påkrævet for at kunne afslutte observationen.



Skærmbillede 3. Systemet kommer med forslag til, hvor teksten specifikt skal dokumenteres i den eller de valgte observationer. Skærmbilledet er fra prototype-appen.

Skærmbillede 4. Brugeren kan se alle relevante observationer og hver observation har en farvekode, som relaterer sig til farvekoden i teksten. Skærmbilledet er fra prototype-appen.

Brugerstudier viste, at det gav værdi for brugerne at skrive fritekst, som systemet efterfølgende assisterede med at placere i de relevante observationer. Det viste sig at det gav et godt overblik over alt, der skulle dokumenteres, og at flowet gav en god sammenhæng i arbejdsgangen, der var værdifuld for brugerne. Flere udtalte, at konceptet var smart, og at det ville spare tid i dokumentationen.

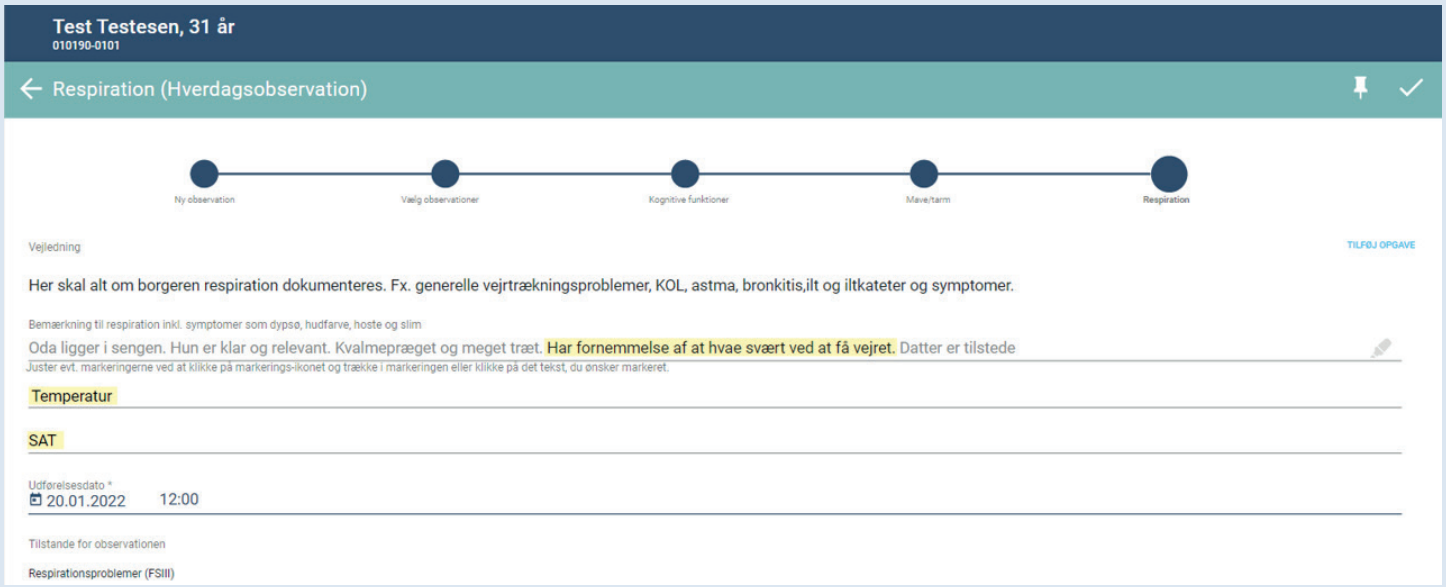
Der var lidt usikkerhed ift. den udviklede funktionalitet med farvemarkering, og hvor den øverste observation var foldet ud, gav mening for alle brugere.

Det viste sig i nogle tilfælde at være svært at forstå, at man fordelte teksten ud med farvemarkeringer, når kun den øverste farvemarkering og observationsdialog var foldet ud. I en tidligere test blev systemet testet, hvor skærmbillede 4 var synligt før skærmbillede 3. I dette tilfælde var det tydeligere for brugerne, hvordan teksten var fordelt ud, fordi man kunne se flere farvemarkeringer og flere observationsdialoger (foldet sammen) på én gang.

Dog var der udfordringer med dette koncept i form af især overlappende farver, og ændring af de enkelte farvemarkeringer, hvorfor den endelige version blev som beskrevet ovenfor. Det er dog en mulighed projekt-teamet mener bør afsøges nærmere med flere brugertest.

Der blev også testet et flow, hvor der kun blev åbnet en observation op af gangen. Dette kan sikre, at der er fokus på alle felter bliver udfyldt, og medarbejderne forholder sig til én observation ad gangen, og dermed sikrer en bedre kvalitet. Udfordringen ved denne arbejdsgang viste sig, at flowet kunne virke længere for medarbejderen. Det er dog en mulighed, der bør undersøges nærmere.

Dette er eksempler på, at der skal arbejdes videre med konceptet og brugergrænsefladen. Farvemarkeringsdelen kan optimeres og flowet er endnu ikke helt perfekt. Der vil derfor skulle arbejdes videre med brugergrænsefladen med flere brugertest, før det er klar til at sætte det i udvikling og dernæst i produktion.



Skærmbillede 5 skitserer hvordan en guidet udfyldning af alle relevante felter i en observation kan integreres.

RESULTATER

Driftstest i hjemmeplejen

Driftstest i hjemmeplejen

I projektet er der udført en driftstest i hjemmeplejen. Her afprøvede 7 medarbejdere i løbet af to dage at dokumentere ved hjælp af prototype-appen. Feltstudierne er foretaget på en måde, hvor den nye prototype-app er testet efter besøg hos borgere. Samtidig blev testen foretaget så nær realtid som muligt i stil med deres dokumentation i dag. Årsagen hertil var at få en så reel test som muligt nær brugernes daglige arbejdsgang. Medarbejderen dokumenterede først i det rigtige Cura, og dernæst vha. prototype-appen. De fik ingen undervisning i den nye arbejdsgang, og dette var for at undersøge om løsningen var intuitiv for medarbejderne. De kvalitative resultater fra testen er gennemgået med beskrivelsen af konceptet og dets fordele ovenfor – mere kvantitative målinger beskrives nedenfor.

Dokumentationen er efterfølgende blevet evalueret af Thisted Kommunes systemadministrator/ dokumentationsekspert. Der er blevet undersøgt følgende:

- Hvilke dokumentationskort, der er dokumenteret i og om disse var korrekte.
- Om dokumentationen er forståelig, kort og præcis.
- Om inddelingen af dokumentationen/ sætningerne er korrekt

Læsevejledning

Der var i alt blevet gemt dokumentation 12 gange i driftstesten. Det var muligt at finde baseline målinger i det rigtige Cura i 8 af eksemplerne, og der er på baggrund af dette lavet nedenstående diagram.

X-aksen viser cases/dokumentationer, mens y-aksen viser antallet af observationer, der er valgt. Inden for hver case, er der 4 søjler. Søjle 1 viser med farver, hvilke observationer, der er korrekt ifølge Cura-eksperten. Anden søjle viser baseline-målinger, altså hvad medarbejderen valgte at dokumentere i det rigtige Cura. Den tredje søjle viser, hvad der bliver valgt vha. prototype-appen. Den sidste søjle viser hvilke observationstyper, der var i top 5.

Driftstesten viser, at der i 7 ud af 8 tilfælde skal vælges mere end én observation, hvilket understøtter udfordringen med, at der ofte skal dokumenteres i mere end én observationsdialog.

Præcision: I 6 ud af de 8 cases, er de korrekte observationer i top 5. I de resterende to cases, er det kun én observation der mangler. Det tyder derfor på, at modellen er meget præcis på disse observationstyper, der har været behov for i driftstesten.

Kvalitet: der ses et kvalitetsløft i dokumentationen i case 1-4. Case 2 bør nævnes, fordi der vælges 4 observationer mod 2 i baseline målingen. Vha. prototypeappen sker der en forbedring, fordi der vælges 2 korrekte til, mens medarbejderen fastholder at vælge de to forkerte, som også skete i baseline målingen. Der ses en forværring i case 5, fordi der vælges en observation for meget, men appen sikrer dog stadig at de korrekte vælges. Case 7-8 er uændret. Samlet set tyder de 8 cases på et kvalitetsløft, og at machine learning kan påvirke medarbejdernes valg af observationer.

Der er nogle standardobservationer, der vælges til som ekstra observationer. Det er fx psykosocialt og funktionsniveau/egenomsorg. Det er de to observationstyper, der i dag i Cura oftest bliver benyttet af medarbejderne som en slags almene observationskort. For at undgå at dette sker i fremtiden, kan man vha. modellen nedtone disse observationer, så de ikke er i top 5. En anden mulighed er at kommunikere til medarbejderne, at de ikke som udgangspunkt skal vælge disse typer til, men at der skal være et fagligt grundlag, for at vælge dem.

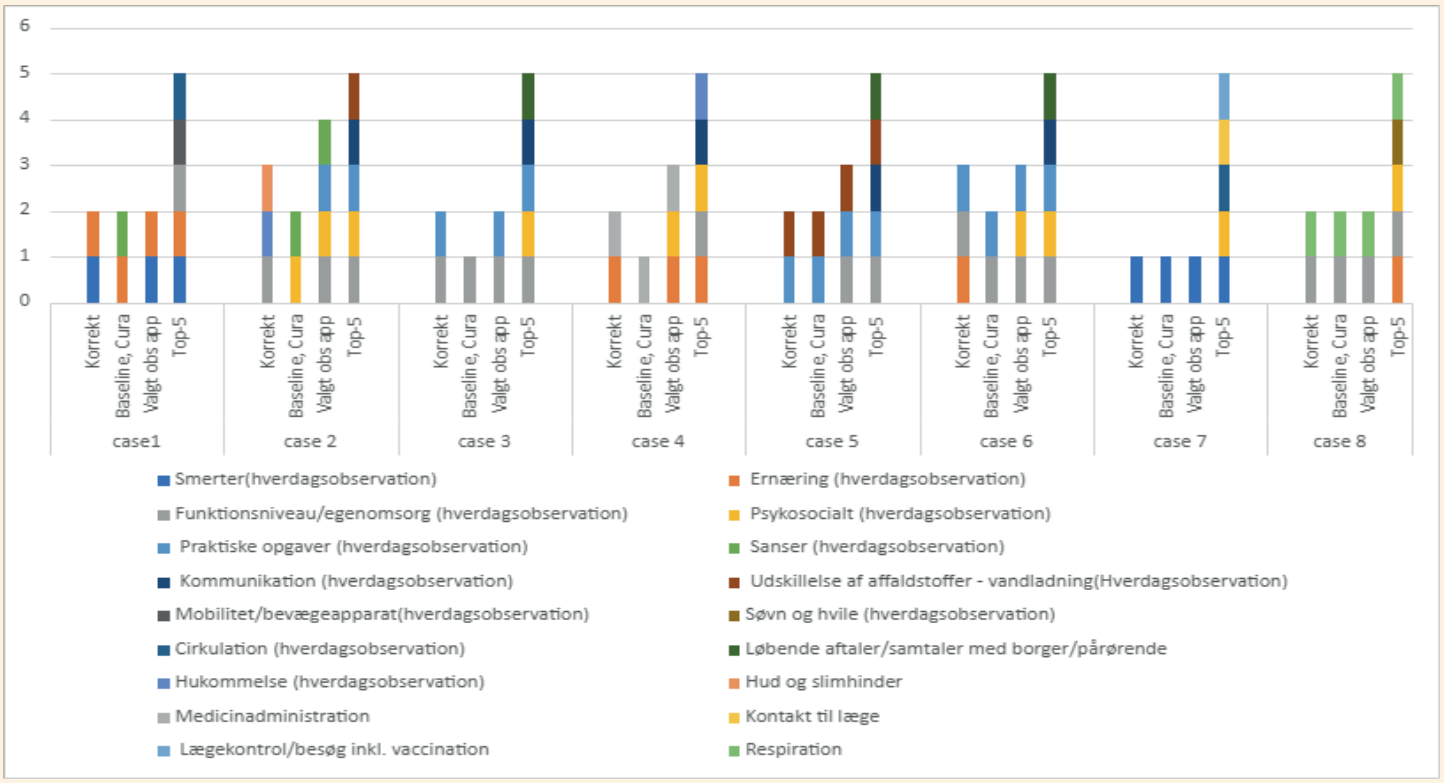


Diagram 1 viser, hvordan der blev dokumenteret i driftstesten, både i det almindelige Cura og via prototype-appen.

Baseline og effektmål – opnås målene?

Udover driftstesten blev der også foretaget baseline og effektmålinger af dokumentationen. Dette var blandt andet for at undersøge medarbejdernes påvirkning, og om kvaliteten vil blive forbedret.

10 medarbejdere fra hjemmeplejen blev udvalgt til målingerne. Målingen blev foretaget med 5 medarbejdere ad gangen. Desværre blev 2 af de 10 medarbejdere skiftet ud mellem baseline og effektmål, så det betyder, at det er sværere at sammenligne, da der er forskel i hvordan medarbejderne dokumenterer.

Der var afsat 2 timer til begge målinger, og der var udarbejdet 10 fiktive cases. Det var de samme cases, der anvendtes til både baseline og effektmåling, så data kunne sammenlignes 1:1. Der var ca. 8-10 uger imellem målingerne, hvorfor det vurderedes, at de samme cases kunne anvendes for at sikre et godt sammenligningsgrundlag. Hver medarbejder er blevet sammenlignet i baseline og effektmål, så vi kan se om deres valg har forandret sig afhængig af dokumentationsflowet med og uden machine learning.

Baseline målingen:

- Medarbejderne skulle dokumentere i Cura, som de gør i dag, og dernæst gemme dokumentationen. Medarbejderne sad sammen, men blev bedt om ikke at spørge hinanden til råds.

Effektmålingen

- Medarbejderne skulle dokumentere via prototype-appen, og dernæst gemme dokumentationen. Medarbejderne sad sammen og blev bedt om ikke at spørge hinanden til råds. Derudover valgte de test-ansvarlige ikke at give en introduktion og undervisning i projekt-protypen for at vurdere, hvor intuitivt systemet var. Da medarbejderne aldrig havde arbejdet i appen før, tog det længere tid at dokumentere på denne måde, særligt i starten, da de først skulle lære, hvordan man gjorde.

Derefter blev baseline – og effektmålene sammenlignet, og der blev evalueret på følgende:

- Hvilke observationer blev valgt, og var de korrekte?
- Kvaliteten af dokumentationen – var det forståeligt, kort og præcist?

Der var i alt 10 cases, og meget stor forskel på, hvor mange cases der blev nået af de forskellige medarbejdere. Der er desværre kun 3 cases fra både baseline og effektmål, der blev nået af næsten alle medarbejderne, og nedenstående konklusioner er derfor baseret på et lille datagrundlag, og kan blot vise tendenser.

Læsevejledning

På x-aksen kan man se brugerne, der er angivet med "Korrekt", A, B, C osv., og med farveangivelsen kan man se, hvad de har valgt i baseline- og effektmålingen. "Korrekt" søjlen er Cura ekspertens svar og anses her som facit. Bemærk bruger D og F ikke er med i målingerne, da vi ikke har resultaterne fra prototype-appen. Bruger J og K har ikke baseline målinger, men er taget med, for at have flest mulige resultater ift. effekten af prototype-appen. I sidste søjle pr. bruger, kan man se hvilke observationer, der var i top 5 i prototype-appen.

I første case har cura-eksperten vurderet, at der bør dokumenteres i følgende observationer: Fald, smerter, Hud/slimhinder/udslæt. I nedenstående diagram, er de angivet med farverne blå, orange og grå under "korrekt". Derudover kan man også vælge "Mobilitet/bevægeapparat", der er gul som, afhængig af hvordan der dokumenteres, også kan anses som korrekt.

I case 2 vurderer Cura-eksperten, at der skal dokumenteres i praktiske opgaver (blå) og Hukommelse (orange).

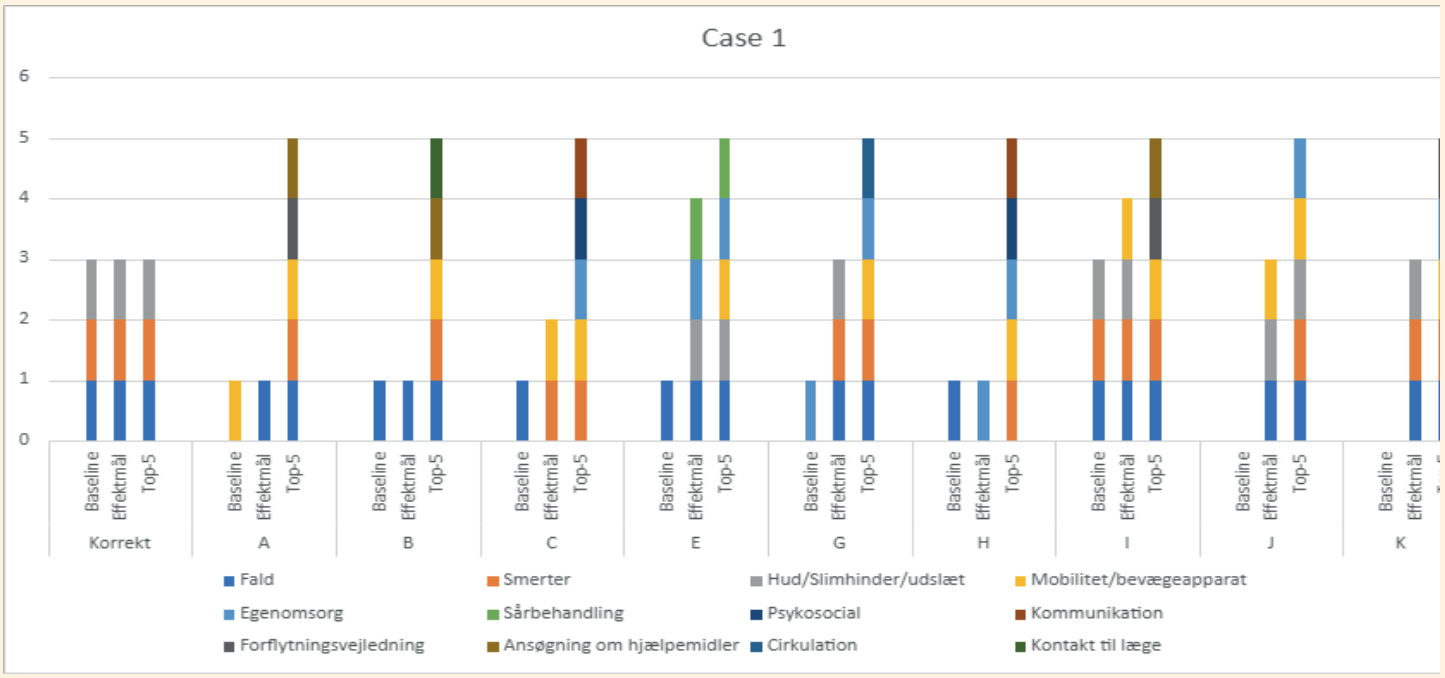


Diagram 2 viser, baseline og effektmål i case 1 for 8 medarbejdere angivet med A, B, C osv. I de første 3 søjler, er der angivet, hvilke observationstyper, der er korrekt.

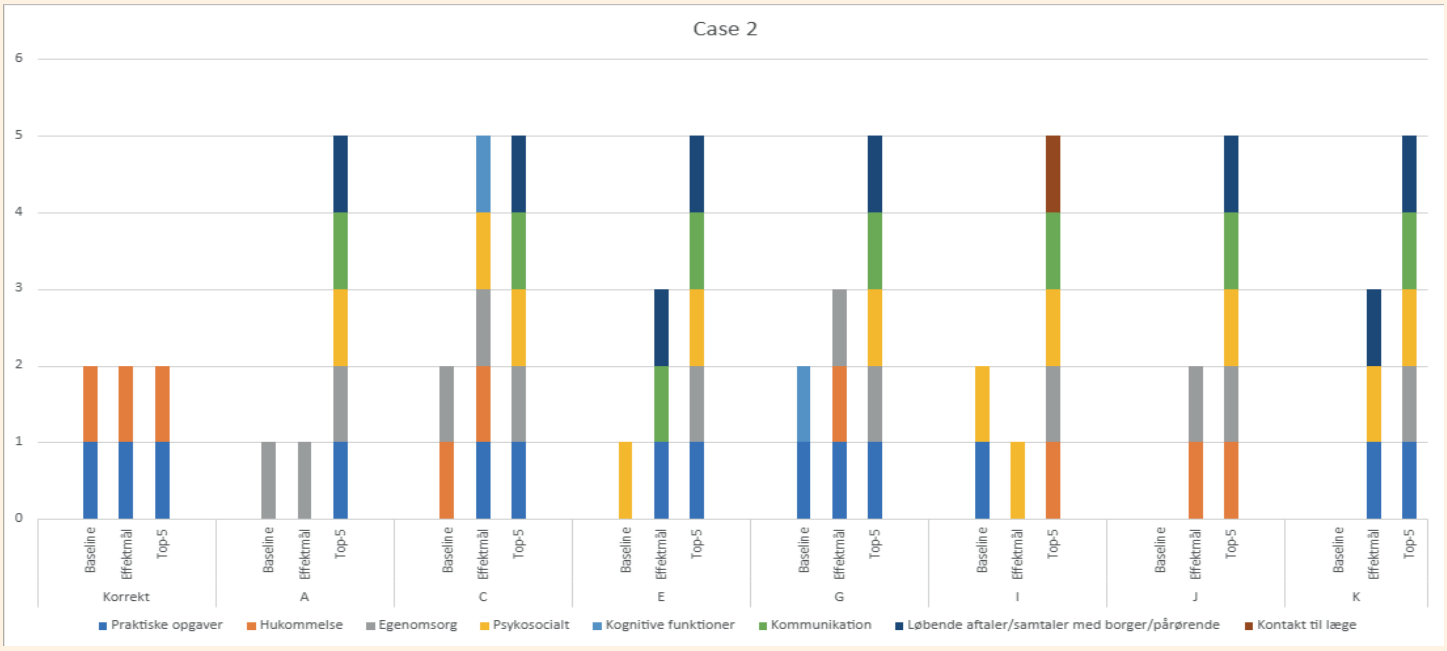


Diagram 3 viser, baseline og effektmål i case 2 for 8 medarbejdere angivet med A, B, C osv. I de første 3 søjler, er der angivet, hvilke observationstyper, der er korrekt.

I case 3 er det besøgsændring (blå) og medicinadministration (orange) der skal vælges. Nogle medarbejdere har i stedet for besøgsændring valgt "Ud af huset aftaler" (grå). Valget kan godt forsvares, ifølge Cura-eksperten, og kan derfor også anses som korrekt.

Baseret på målingerne kan der overordnet set konkluderes følgende tendenser:

- Vi påvirker medarbejderne til at vælge flere observationer, men også nogle gange for mange. Det kan man fx se på bruger C og E.
- Vi påvirker medarbejderne til at vælge de rigtige observationer i flere tilfælde, og ser en øgning i kvaliteten hos nogle brugere, fx bruger G har en tydelig kvalitetsforbedring i case 1 og 2. Hos få brugere ser vi ikke nogen ændring i valg af observationer i baseline og effektmål. I case 3 ses en forværring hos nogle medarbejdere. Disse tyder på, at prototype-appen ikke har en høj nøjagtighed på netop disse to observationstyper.
- Der ses en mindre kvalitetsforbedring i data, særligt i case 1, hvor flere vælger korrekt.
- Der er en tendens til at medarbejderne vælger observationerne "Funktionsniveau/egenomsorg" og "Psykosocialt" og "Kommunikation" som ekstra observationer. Disse observationer bruges i dag meget ofte af medarbejderne, som en slags standardobservationer.
- Det tyder også på, at der udforskes flere observationer, sandsynligvis fordi medarbejderne præsenteres for flere muligheder, og at det er muligt at vælge flere på én gang.

Der er en tendens, der viser, at man med machine learning kan påvirke medarbejderne i deres valg. Præcisionen på nogle observationstyper kan være bedre, så medarbejderne uanset observationstyper bliver ansporet til at vælge de korrekte observationer. Et forslag til videreudvikling er et lille "i" (informationstegn) i top-5, hvor medarbejderne kan læse vejledningsteksten, så de kan læse hvad observationen indeholder, inden de vælger observationer.

Konklusion baseret på data

Der ses en tendens til at kvaliteten øges vha. prototype-appen. Det er interessant at der ses et tydeligere kvalitetsløft i de rigtige data fra driftstesten, fremfor i de fiktive cases og målinger i baseline- og effektmål. Årsagen er højst sandsynligt, at appen er bygget efter rigtige data i Thisted Kommune, og at driftstesten viser et mere realistisk billede af, hvordan der dokumenteres i praksis end de fiktive cases. Hvis der skal laves flere målinger i fremtiden, anbefales det er gøre det som via driftstesten, så det gøres med rigtige data. Det tyder dog også på, at der er nogle observationstyper, som prototype-appen, skal være bedre til at foreslå.

Der har ikke været undervisning forbundet med brug af appen. Det vurderes, at der vha. undervisning i prototype-appen, FSIII og observationstyper, og hvornår disse skal bruges, kan opnås et markant kvalitetsløft i data, fordi appen netop understøtter at flere observationstyper kan vælges på en gang og foreslår de relevante typer.

Evaluering af prototype-appen ift. kvalitet

I evalueringen blev kvaliteten af observationerne også vurderet. Thisted Kommunes Cura-ekspert, har vurderet kvaliteten af dokumentationen i både baseline- og effektmåling. På baggrund af dette kan der konkluderes følgende tendenser:

- I prototype-appen starter medarbejderen med at beskrive konteksten, og herefter beskrive handlinger. Konteksten bliver kopieret inden i alle observationer, og handlingerne ift. de enkelte observationer adskilles. De enkelte observationer er derfor forståelige, og kontekst er næsten altid med, selvom der arbejdes i flere observationer på en gang.
- Medarbejderne er gode til at lave kort og præcis dokumentation - både i baseline og effekt. De skriver ikke for meget i forbindelse med effektmålingen, selvom det er fritekstfelter.
- Generelt god kvalitet i observationerne beskrevet via appen. Projektteamet ved, at der sker kopiering fra observationer i Cura i dag, men har ikke set det i baseline målinger.
- Få gange dobbeltdokumentation - nogle

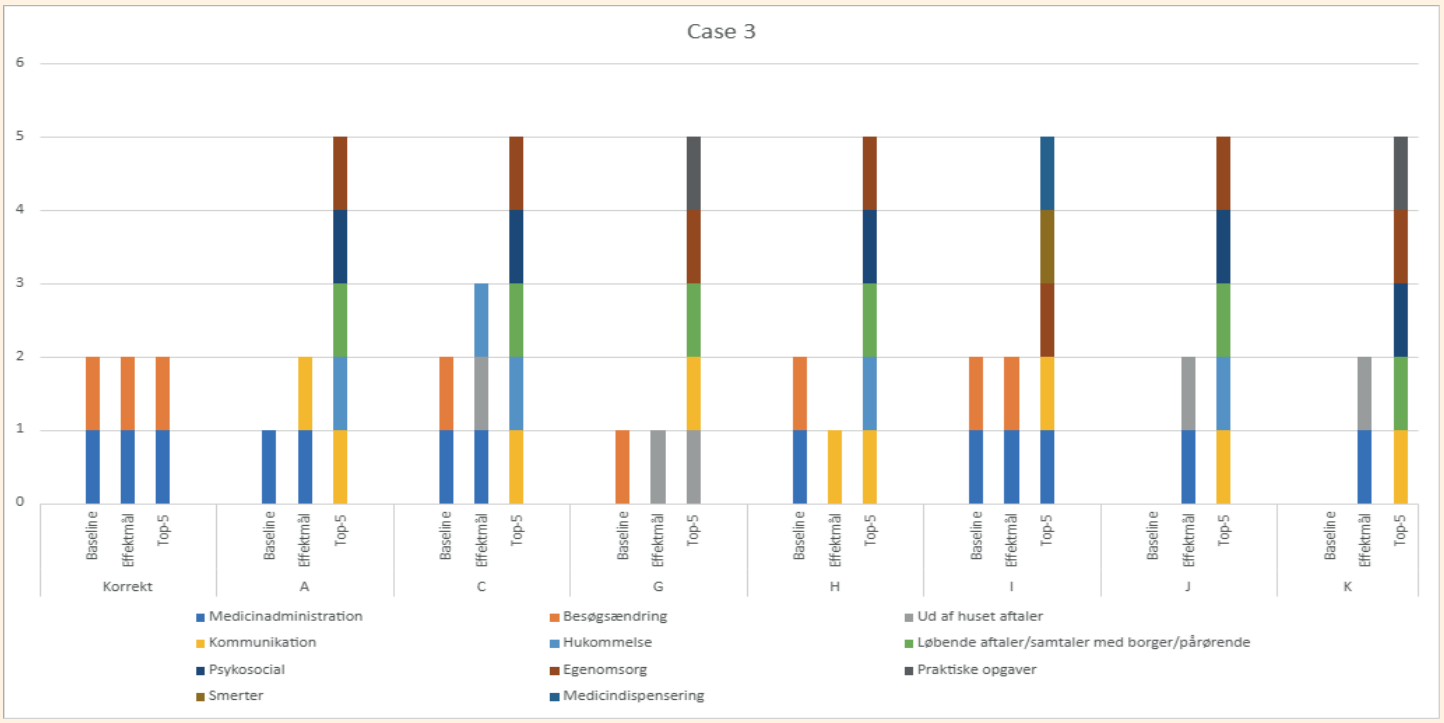


Diagram 4 viser, baseline og effektmål i case 3 for 8 medarbejdere angivet med A, B, C osv. I de første 3 søjler, er der angivet, hvilke observationstyper, der er korrekt.

vælger ekstra observationer for en sikkerheds skyld. Overflødige observationer der vælges, er eksempelvis funktionsniveau/egenomsorg, psykosocial og kommunikation.

- Medarbejderne er gode til at dele sætninger op i prototype-appen, så de passer til observationen. Selvom valget af observation i nogle tilfælde var forkert, var tanken bag inddeling korrekt.
- I nogle tilfælde bliver teksten delt mere ud end i baseline-målingerne. At teksten deles ud er første skridt ift. at dokumentere i FSIII. Generelt vurderes det, at appen understøtter, at der arbejdes bedre efter FSIII-metoden.

Generelt ses der et kvalitetsløft, særligt fordi arbejdsgangen sikrer at medarbejderne deler deres tekst mere ud, så der arbejdes mere efter FSIII metoden. Når data placeres korrekt, vil det på sigt også betyde, at der kan fås et bedre overblik over data i journalen.

Hvad synes brugerne om løsningen?

Som tidligere beskrevet har medarbejderne været med til at udvikle og påvirke løsningen fra ide til udvikling. Undervejs er der kommet mange vigtige input, der har præget løsningen. Da demoen var færdigudviklet var der i alt 16 medarbejdere, der prøvede løsningen. 9 af dem prøvede løsningen i forbindelse med baseline og effektmål, mens de resterende 7 prøvede appen til driftstest. Medarbejderen havde ikke fået undervisning i appen.

Disse medarbejdere fik et spørgeskema, og her blev de bedt om at vurdere løsningen ift. brugervenlighed, kvalitet, faglig refleksion og mulighed for at spare tid ved at dokumentere på denne måde.

I spørgeskemaet, blev der stillet spørgsmål omkring tid, brugervenlighed, overblik, kvalitet og faglig refleksion. På x-aksen kan man se spørgsmålet, der er stillet. På y-aksen kan man se antal der har svaret. Hvis søjlen er blå er medarbejderen enig i udsagnet, orange er delvist enig, og grå er uenig.

Generelt er der rigtig god og positiv medarbejderfeedback. Mange af medarbejderne har været meget positive om løsningen. Det bør også nævnes, at få medarbejdere har været mere negative om løsningen. Årsagen har været, at det har været svært med en anderledes dokumentationspraksis. Andre har været bekymrede for om automatikken ved brug af machine learning kunne føre til fejl.

Brugervenlighed

Diagram 5 viser, at medarbejderne vurderer, at løsningen er brugervenlig, og mere brugervenlig end Cura er i dag. 12 af de adspurgte medarbejdere tror også, at flere vil dokumentere i Cura, fordi løsningen er mere brugervenlig. Vedr. farvemarkerings-funktionaliteten, så er der 7 der er uenige i, at der er simpelt at farvemarkere teksten, så det tyder på, at der bør ses nærmere på denne funktion.

Overblik

Diagram 6 viser, at medarbejderne er enige i, at løsningen giver et godt overblik over mulige observationer, og at det vil være nemmere at finde tilbage til informationer i Cura, fordi løsningen sikrer, at der vælges de korrekte observationskort.

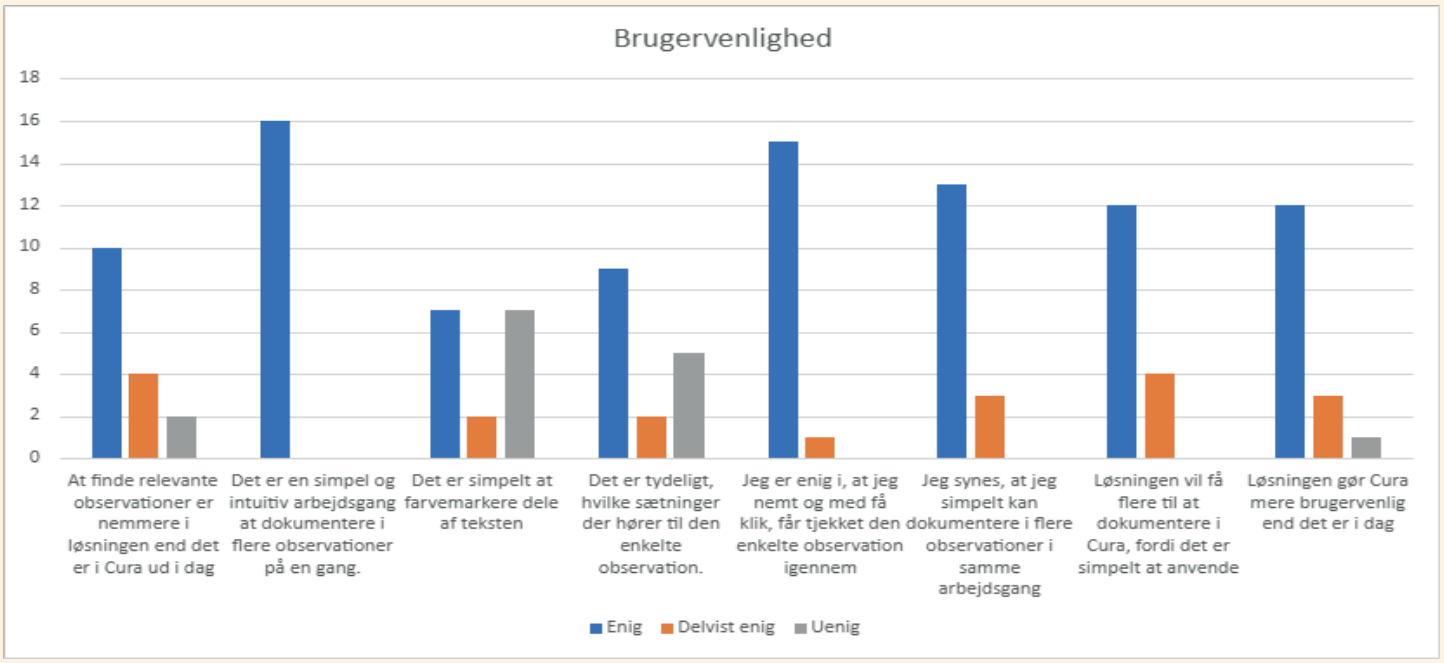


Diagram 5 viser 16 medarbejderes besvarelse ift. brugervenlighed. Udsagnene står beskrevet på x-aksen, og y-aksen viser antallet af medarbejdere. De forskellige søjler, angiver om medarbejderne er enige, delvist enige eller uenige i udsagnet.

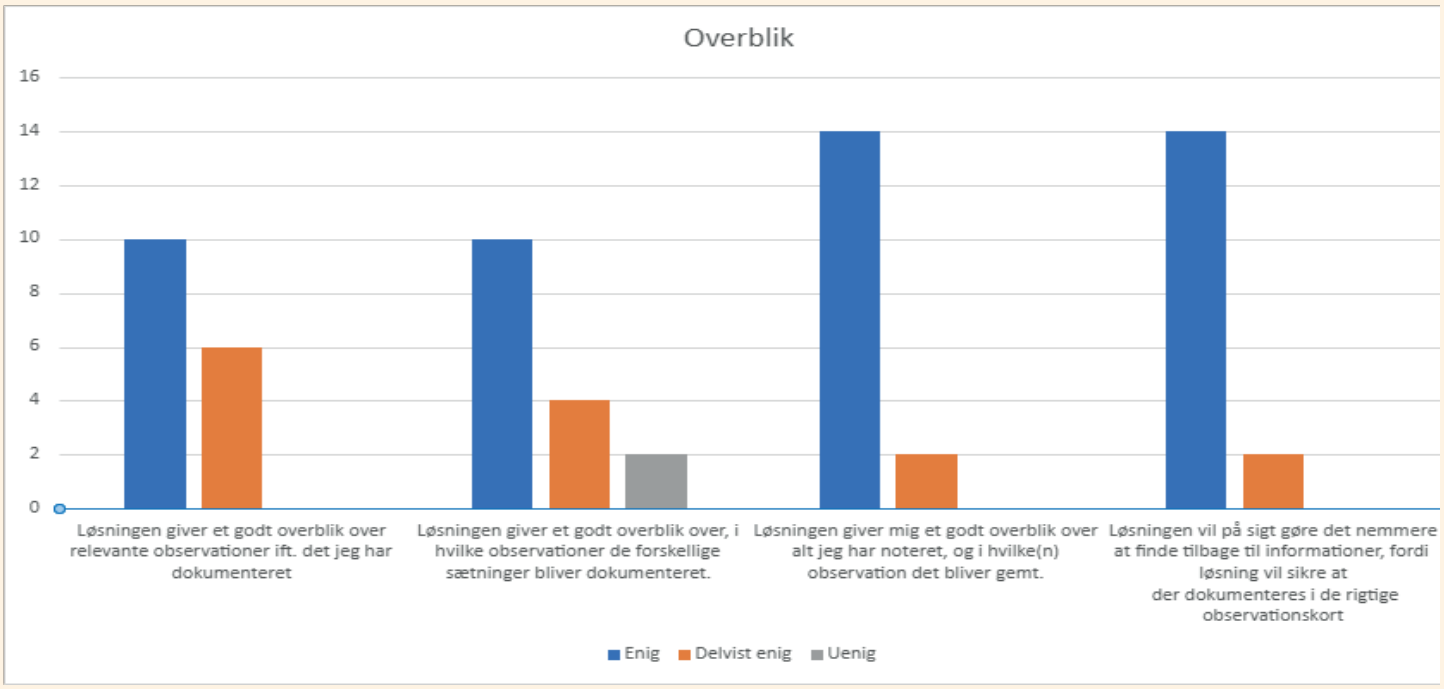


Diagram 6 viser 16 medarbejderes besvarelser ift. overblik. Udsagnene står beskrevet på x-aksen, og y-aksen viser antallet af medarbejdere. De forskellige søjler, angiver om medarbejderne er enige, delvist enige eller uenige i udsagnet.

Faglig refleksion

Ifølge medarbejderne bliver de inspireret til at vælge flere observationer end de gør i Cura i dag, hvilket også er målet, fordi tendensen i dag er, at der vælges få observationer, og ofte de samme typer. Medarbejderne synes ikke, at de bliver inspireret til at anvende andre typer observationer via prototype-appen. Derudover er der flere medarbejdere, der ikke forholdt sig til vejledningsteksten i observationerne, hvilket også afspejles i det sidste spørgsmål i nedenstående diagram. Her kan der eksempelvis arbejdes med at tydeliggøre vejledningsteksten i løsningen, så medarbejderen vha. Cura tjekker, at der dokumenteres det korrekte sted.

Når medarbejderne brugte løsningen, kunne man dog se, at der skete en faglig refleksion, når de valgte observationer. De valgte ikke blot den første observation på listen, men reflekterede over, hvilke de valgte. En medarbejder spurgte fx også til i en fiktiv case, om en borger havde sukkersyge, fordi observationen "blodsukkermåling" fremkom i top 5. Vurderingen er derfor, at løsningen understøtter en vis grad af faglig refleksion, men ifølge medarbejderne, kan de understøttes endnu mere. Der kan bl.a. arbejdes med, at vejledningsteksterne i observationerne kan tydeliggøres, hvilket kan hjælpe medarbejderne til at få udfyldt observationerne korrekt.

Kvalitet

I diagram 8 ses det at 12 ud af 16 medarbejderne vurderer, at de i højere grad vil vælge de korrekte observationstyper vha. Prototype-appen og at kvaliteten i dokumentationen vil blive øget. Medarbejderne er enige i, at de skal tjekke det enkelte observationskort igennem, inden dokumentationen gemmes.

Tid

At spare tid, var ikke det primære mål ved at udvikle denne løsningen. Men medarbejderne er enige, eller delvist enige i, at de kan spare tid ved at benytte løsningen.

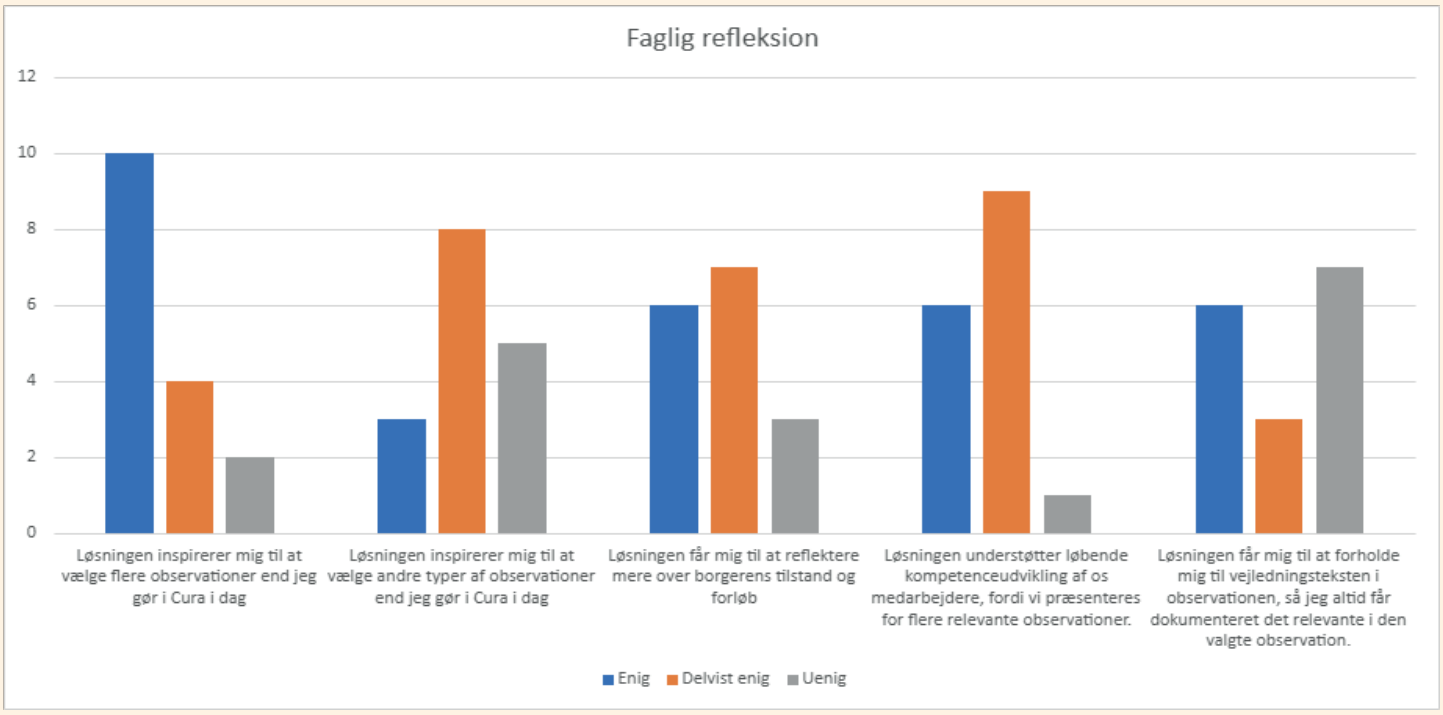


Diagram 7 viser 16 medarbejderes besvarelser ift. Faglig refleksion. Udsagene står beskrevet på x-aksen, og y-aksen viser antallet af medarbejdere. De forskellige søjler, angiver om medarbejderne er enige, delvist enige eller uenige i udsagnet.

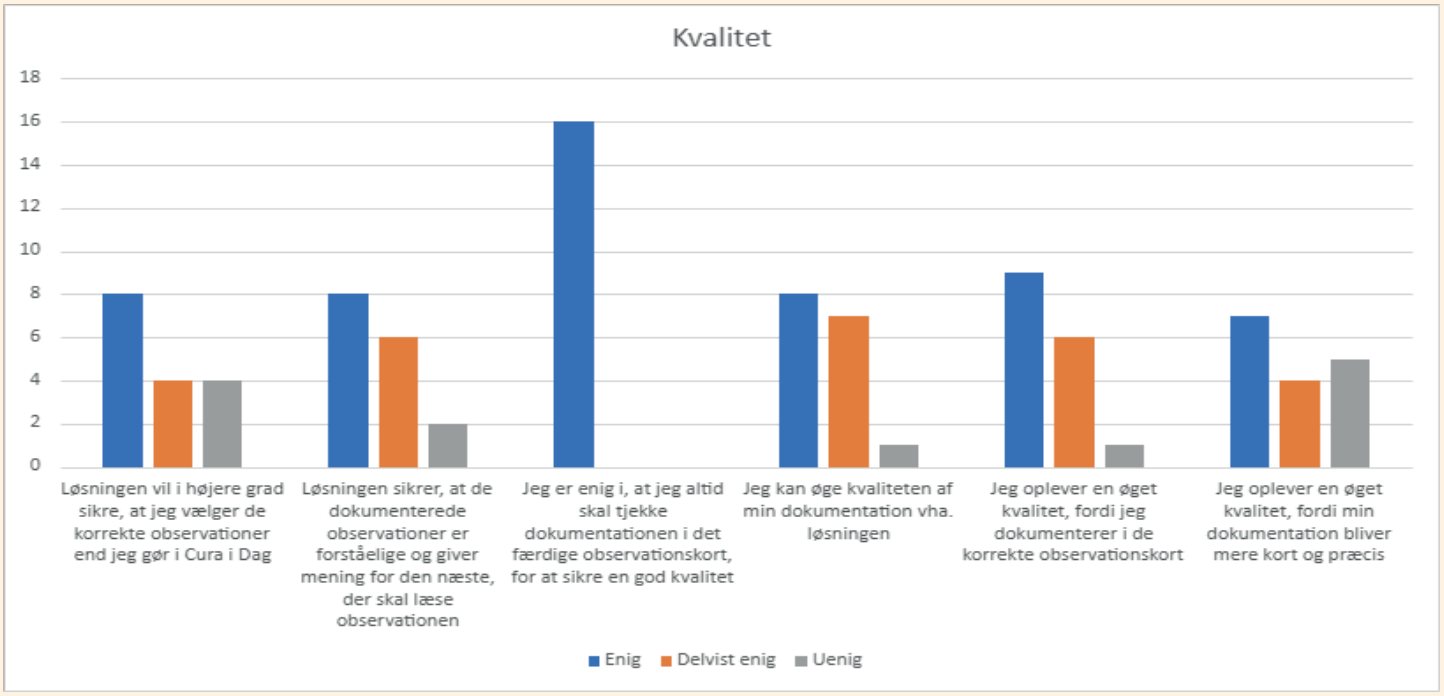


Diagram 8 viser 16 medarbejderes besvarelser ift. kvalitet. Udsagnene står beskrevet på x-aksen, og y-aksen viser antallet af medarbejdere. De forskellige søjler, angiver om medarbejderne er enige, delvist enige eller uenige i udsagnet.

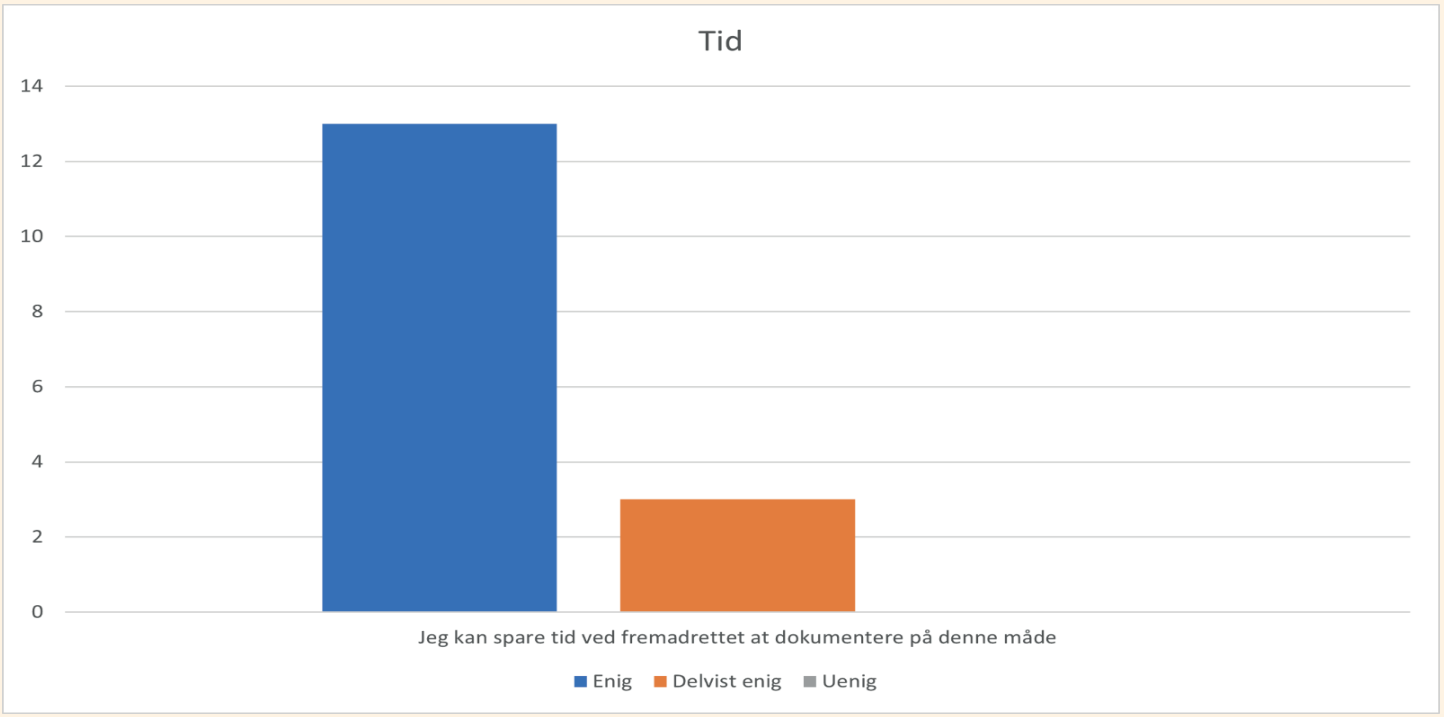


Diagram 9 viser 16 medarbejderes besvarelser ift. tid. Udsagnet står beskrevet på x-aksen, og y-aksen viser antallet af medarbejdere. De forskellige søjler, angiver om medarbejderne er enige, delvist enige eller uenige i udsagnet.

Konklusion på brugertest

Overordnet viser resultaterne, at machine learning virker. Der er et stort potentiale for at forbedre brugeroplevelsen og øge kvaliteten af dokumentationen. De fleste medarbejdere vurderede også, at det ville spare dem tid at dokumentere på denne måde, hvilket dog ikke har været det primære fokus for dette projekt. På baggrund af projektet kan der konkluderes følgende:

- Vi ser en tendens til, at der vælges flere og mere korrekte observationer, og ser derfor et stort potentiale i at øge kvaliteten af dokumentationen.
- Medarbejderne er meget positive for løsningen, og de oplever en mere simpel brugergrænseflade.
- Det er vigtigt at løsningen har en ret god præcision, så medarbejderne ikke blot vælger de forkerte observationstyper. Derudover skal der ske undervisning i dokumentationsflowet, observationstyper og FSIII.
- Medarbejderne vurderer, at de kan spare tid vha. løsningen
- Medarbejderen begynder på baggrund af forslagene at reflektere mere over deres valg af observation, samt at der skal vælges flere.
- Vi ser at medarbejderne begynder at arbejde mere efter FSIII-metoden, fordi teksterne bliver delt mere op.

Ved at gentænke dokumentationspraksissen, kan der konkluderes, at der er et stort potentiale ved at bruge machine learning i omsorgsjournalsystemerne. Det anbefales dog, at der arbejdes videre med dette koncept med mere test i samarbejde med brugerne, da "farverne" og flowet kan gøres mere intuitivt jf. tilbagemeldingerne fra brugertests. De nuværende brugertest har dog allerede fået en rigtig god feedback fra medarbejderne, og det ser lovende ud ift. at arbejde mere efter FSIII metoden, da det ansporer brugerne til at vælge flere observationer og få inddelt teksterne. Det bør undersøges nærmere, hvordan man efterfølgende tilgår dataene. Skal man fx kunne se hele den oprindelige tekst, der er "tagget" med observationstyperne, eller blot de enkelte observationstyper – eller begge dele. Det har der desværre ikke været tid til at undersøge nærmere i dette projekt.

Dette er kun første skridt i arbejdet med at gøre dokumentationen mere simpel for medarbejderne vha. machine learning. Under arbejdet er der kommet mange flere ideer fra deltagere i projektet, til hvordan man kan ellers kan indtænke og bruge machine learning i systemerne.

Hvis medarbejderne skal understøttes bedre i deres dokumentationspraksis end de bliver i dag, så kan machine learning være med til at sikre, en bedre arbejdsgang i systemerne, der også vil bidrage til en højere kvalitet. I dag oplever medarbejderne, at data skal deles meget op pga. FSIII. Struktur og opdeling er dog vigtig, fordi det gør det nemmere at genfinde og genbruge data.

At indtænke machine learning i vores dokumentationssystemer kan blive revolutionerende i fremtiden, fordi det kan guide medarbejderne i dokumentationen. Det kan hjælpe dem til at strukturere data i et mere brugervenligt dokumentationsflow. Det kræver dog, at der afsættes flere ressourcer, mere tid og udvikling, og at der arbejdes mere med brugergrænsefladerne i samarbejde med medarbejderne.

Undervejs i projektet, er der kommet flere ideer til, hvor man ellers kan benytte machine learning. Ideer til videreudvikling kunne være følgende:

- Automatisk indsætte tekst i besøgsplanen observationen, når det er relevant, så medarbejderen ikke skal skrive det samme to steder. Besøgsplanen er der, hvor medarbejderne kan se, hvad borger skal have hjælp til.
- Machine learning kan forbedre brugeroplevelsen yderligere ved at understøtte tegnsætning og retstavning, da flere medarbejdere har læse- og skrivevanskeligheder.
- Maskinlæring kan hjælpe med at slette unødvendige informationer (tomme ord) og/ eller fremhæve de vigtigste informationer.
- Tilføj spørgsmål som tilpasser sig den tekst, der skrives som fritekst i dokumentationen. I stil med en chatbot for at understøtte at brugerne kommer omkring alle relevante detaljer.

- At der laves et ”smart summary” til den enkelte medarbejder. Når medarbejderen fx skal besøge borger, er nødvendige oplysninger samlet et sted, hvor relevant info fra besøgsplan og seneste observationer vises – eller evt. kan læses op, så køretiden kan udnyttes.

TEKNISK DOKUMENTATION


CURA

Valg af observationer

Denne del af evalueringen beskriver den tekniske dokumentation af projektet, og beskriver det arbejde udviklerteamet bestående af data-scientists, har lavet, for at kunne teste og afprøve konceptet.

Den tekniske dokumentation er inddelt i to. Første del beskriver den tekniske dokumentation ift. valget af observationer, så der fremkommer en top 5, baseret på det skrevne fritekst. Dette er afbilledet i brugergrænsefladen på skærmbillede 2, som tidligere vist i rapporten. Anden del af den tekniske dokumentation er vedr. sætningsplacering. Når der er skrevet fritekst og valgt relevante observation, skal modellen dernæst foreslå hvilke sætninger der hører til de valgte observationstyper. Dette er tidligere vist på skærmbillede 3 og 4.

Udfordringer

Der er i alt 342 forskellige observationskort i Cura i Thisted Kommune. Der er 91 tilgængelige for ældreområdet, som Thisteds hjemmeplejere kan vælge at dokumentere i, i Cura. 70 af observationstyperne er oprettet af Thisted Kommune, mens de 21 observationstyper skal benyttes pga. dokumentations-metoden Fællessprog III. Fordelen ved at have mange observationer er, at de kan gøres konkrete, så der kun dokumenteres det mest nødvendige. Ulempen er, at det er svært for medarbejderne at huske navnene på de 91 observationstyper, samt hvad de indebærer. Det betyder, at mange i stedet vælger den samme observation igen og igen, og kvaliteten derfor ikke er som ønsket.

Diagram 10 viser hvor mange observationsskemaer, medarbejderne bruger. Y-aksen viser antallet af medarbejdere, og x-aksen viser antallet af observationskort, der har brugt det specifikke kort. Det viser, at næsten 200 medarbejdere kun har anvendt ét observationskort i Cura.

I diagram 10 kan vi også se tendensen ift. at medarbejderne ofte benytter de samme tre observationstyper, som er blevet nævnt tidligere i rapporten. Her ses tydeligt, at medarbejderne har tendens til at benytte hvad de anser som "standardobservationer" frem for at vælge de rigtige.

Ved forkert placering bliver det sværere at genfinde informationen og få et overblik i Cura. Til et besøg kan det godt være, at der skal dokumenteres i tre forskellige observationer, hvilket også er tidskrævende. Som et forslag for at optimere denne proces vil vi foreslå at bruge Machine Learning (ML) til at foreslå passende observationer. ML handler om at give en computer evnen til at lære datas sammenhæng med definerede labels. Disse trænede modeller kan bruges på data, som ikke før er set af modellen, og sætte det label på som passer bedst.

Ved brug af Natural Language Processing (NLP), en gren inden for machine learning som kan forstå og bearbejde menneskeligt sprog og tekst, kan man foreslå relevante observationstyper til brugeren baseret på den fritekst, de skriver ind i Cura. Dette medfører, at dokumentationsprocessen bliver ændret, så man dokumenterer hele sit besøg i ét samlet fritekstfelt, frem for at vælge en observationstype på forhånd og dokumentere heri.

Målet for denne del af projektet var således at kunne identificere den rigtige observation ud fra en given tekst, som er dokumenteret i Cura.

Hvilken type hardware er brugt?

Denne type hardware er brugt for at kunne gennemføre arbejdet.

- Thisted Kommune: NVIDIA GeForce RTX3090 24GB6 STRIX+ AMD Ryzen™ Processors 9 5900X + DDR 128GB + HDD 2TB
- Systematic: NVIDIA Quadro P5000 + Intel(R) Xeon(R) Gold 5118+ SSD 1TB + NMVE SSD 1TB + DDR 64GB

Data

Data stammer fra Cura (med fokus på hjemmeplejen) i Thisted Kommune i perioden 03.02.2018 til 19.07.2021. Data indeholder i alt 14.356.154 antal rækker, hvor én række repræsenterer dokumentation fra et felt i en observationsformular. Der blev kun brugt data, som

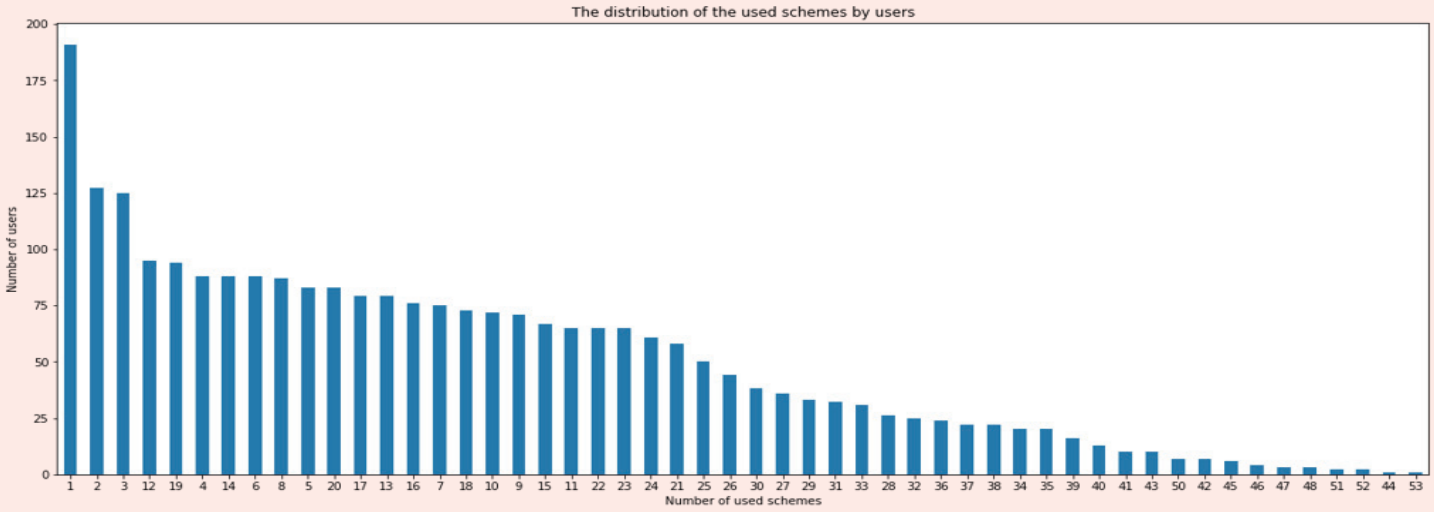


Diagram 10 - Antallet af gange, som de forskellige observationer er anvendt i det eksisterende datasæt

havde et fritekstfelt i Curas observationsformularer. Det resulterede i 84 observationskort i Thisted, som indeholder et fritekstfelt. Data bestod således af et dokumentations-id, fritekstfeltet, observationstype, dato for dokumentation og hjemmeplejerens navn (antal rækker: 668.348).

Preprocessing

Data skulle gøres klar til modeltræning. Klargøring af data bestod af at rengøre teksten i fritekstfelterne: fjerne rækker, hvor der ikke er tekst i fritekstfeltet, fjerne tal fra teksten og beregne længden af teksten. Hvis en tekst bestod af mindre end fem ord blev det fjernet fra datasættet. Der fandtes også dubletter, dvs. ens rækker. Når en bruger åbner en specifik observationskort, vil der oprettes en ny version af dokumentationen, selv hvis der ikke er lavet ændringer i formularen. Disse versioner blev derfor fjernet fra datasættet, så der ikke var dubletter.

For at forbedre kvaliteten af data valgte projektteamet yderligere at bruge et udpluk af det fulde dataset, baseret på udvalgte brugere. Da der er stor forskel på, hvor mange forskellige observationer de forskellige brugere benytter, har Thisted Kommunes superbruger udvalgt nogle hjemmehjælpere til at være superbrugere i Cura dokumentation. Dette valg af superbrugere blev valgt tidligt i projektet, og indebærer 28 forskellige brugere. I dette datasæt var der i alt dokumenteret i 66 forskellige observationer. Dette superbrugerdatasæt bestod af 19.261 rækker efter klargøring af data.

På et senere tidspunkt har projektteamets data scientists selv udvalgt bestemte hjemmehjælpere som superbrugere – denne gang baseret på statistik. Kriterierne for at være superbruger indbefattede således, at en bruger som minimum havde anvendt mere end 21 forskellige observationer, når de dokumenterer i Cura, og igennem tiden har registreret mere end 100 observationer. Derudover var der krav til at antallet af registreringer for de to hyppigst anvendte observationer ikke måtte udgøre mere end 50% af det totale antal registreringer. Med disse kriterier endte vi med at have 731 superbrugere. Data bestod af 416.355 rækker, hvor 80 forskellige observationer er repræsenteret i datasættet. De 20 mest anvendte observationer kan ses på diagram 11. Dette datasæt har også hjulpet på modellernes performance, som beskrevet i afsnit 5.3.

Næste skridt bestod af at lave embeddings: en mapping fra tekst til vektorer af tal. Vi brugte forskellige embeddings: LaBSE, Laser, TF-IDF og Roberta. Når disse embeddings var lavet, og de respektive observationer var tilkoblet de enkelte vektorer, blev data opdelt i et trænings-, valideringss- og testsæt. Træningssættet blev brugt til at træne modeller på, og valideringssættet blev brugt til at vurdere og finpudse modellerne med under træning. Testsættet er et uafhængigt datasæt, som

modellen ikke ser i træningsperioden. Data blev sorteret efter oprettelsesdato, fra ældste til nyeste dokumentation. De sidste 2000 rækker, det nyeste dokumentation i Cura, blev brugt som testsæt.

Udover de traditionelle ML-modeller, hvor ovenstående embeddings blev brugt, kiggede vi på sprogmodeller, som går ind under deep learning-kategorien. Disse sprogmodeller er dybere modeltyper, som har evnen til at lære mere avancerede koncepter - heriblandt fra sprogrelaterede projekter. Disse modeller kræver også, at man anvender embeddings, som her er trænet sammen med den model, man anvender.

Der blev også eksperimenteret med dataaugmentering, hvor man kunstigt tilføjer mere data baseret på allerede eksisterende datapunkter. F.eks. kan man fjerne nogle af ordene i en tekst, erstatte navneord med andet eller blande ordene rundt i teksten. Dette øgede dog ikke model performance. Derfor blev det af DS-teamet vurderet, at det ikke var en fordel at anvende augmentering yderligere.

Modeller

Målet var at udvikle en model, som kan tage en tekstbeskrivelse af et borgerbesøg og identificere hvilken observationstype, teksten omhandler. DS-teamet afprøvede forskellige machine learning modeller til denne opgave. Baseret på den første version af superbruger-datasættet, var logistisk regression og Linear Discriminant Analysis (LDA) med LaBSE embeddings de to højest performende modeller med en nøjagtighed på hhv. 55,2% og 54,8%. Nøjagtighed, eller accuracy, er et mål for hvor god en model er til at identificere de rigtige observationer. Det betyder således, at den logistiske regression var i stand til at foreslå den korrekte observation på 55% af datapunkterne i testsættet.

DS-teamet ønskede at gøre modellerne bedre, og derfor blev der afprøvet andre metoder, bl.a. hierarkisk modellering og ensemble modeller. Det boostede dog ikke model performance betydeligt. Deep learning (DL), er en gren af Machine Learning som forsøger at efterligne, hvordan signaler bevæger sig i hjernen, og dermed være i stand til at lære mere avancerede principper og koncepter. På baggrund af disse styrker var DL næste skridt. Der blev brugt en allerede trænet DL sprogmodel til at lære sproget brugt i Cura-dokumentation. Dette gav os en accuracy på 63%, hvilket er betydeligt højere end hvad der hidtil har været opnået.
Med data fra den anden version af superbrugere (baseret på statistiske valg) kunne modellen opnå accuracy på 83%. Dette er den højeste accuracy, der er blevet opnået til denne opgave, og det er derfor også denne model, som kører i demo-applikationen i dag.

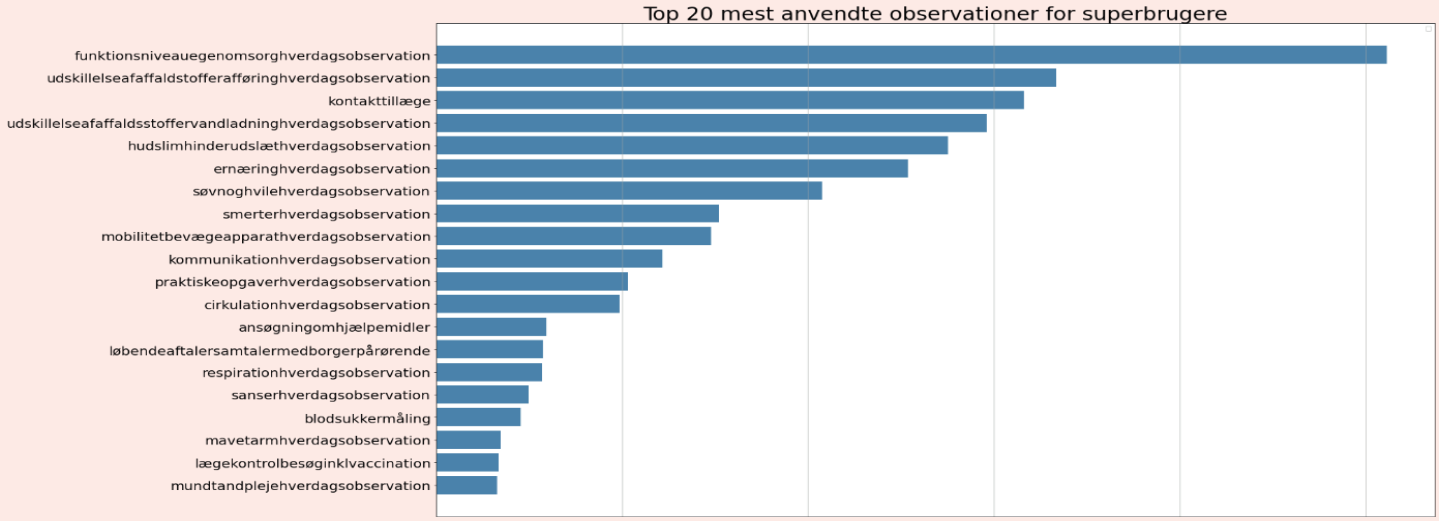


Diagram 11 - De 20 mest anvendte observationer for superbrugere (baseret på statistiske valg) i Thisteds hjemmepleje.

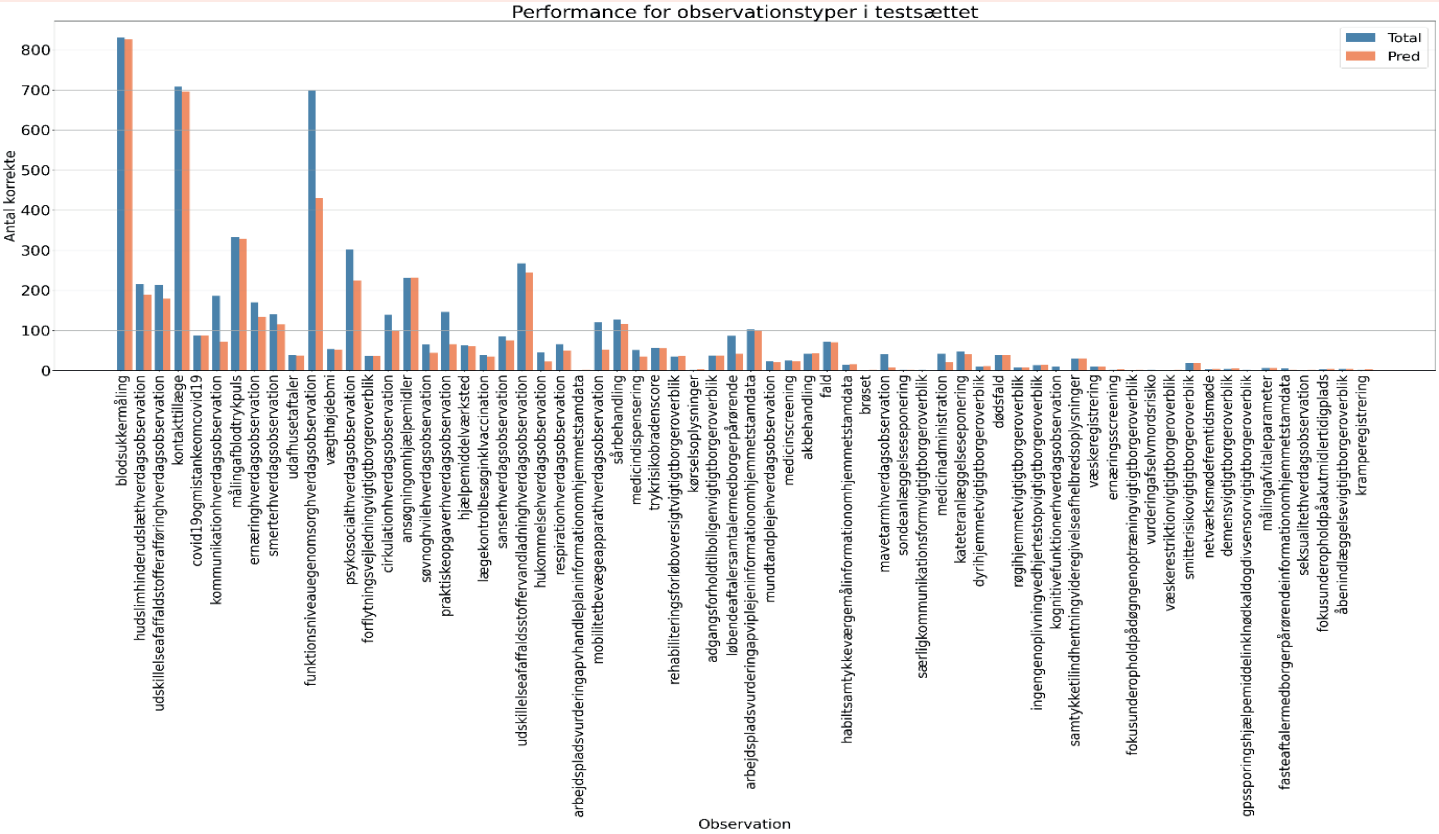


Diagram 12 - Blå: Antallet af datapunkter der tilhører en observation. Orange: Antallet af gange modellen har valgt korrekt.

Fejlanalyse

For at forstå modellens prædiktioner, kiggede vi ind i de fejl, den laver. F.eks. kan man kigge ind i, om der er nogle observationer, som modellen er bedre til at identificere end andre, og om der er nogle observationstyper, som modellen har svært ved at differentiere mellem. For at analysere på de fejl modellen laver, blev resultaterne for de forskellige observationstyper undersøgt nærmere. Resultatet heraf kan ses på diagram 12.

I diagrammet repræsenterer de blå bjælker det antal gange den pågældende observation findes i testsættet. Den orange bjælke repræsenterer, hvor mange gange observationen blev identificeret. Ved at dykke ned i denne figur kan det ses, at der er mange observationer, som modellen er god til at klassificere. Men der findes også observationer, som den har svært ved at identificere. Blandt de observationer, som modellen har svært ved at identificere, er bl.a. 'Funktionsniveau- og egenomsorg' og 'Kommunikation'. Modellen opnår en accuracy på henholdsvis 62% og 38% på disse observationer. Dette betyder, at modellen kun er i stand til at klassificere 'Kommunikation' rigtigt 38% af gangene, hvor den faktisk burde klassificere teksten som 'Kommunikation'. De lave accuracies kan være et tegn på støj i data: disse to observationer kan være vage i definition, og kan derfor dække over mange forskellige observationer fra hjemmehjælperne. Dette understøttes af, at Cura-eksperter i Thisted Kommune godt kender til denne fejltype i dokumentationen. Her er det tidligere påvist i journalaudits, at medarbejdere fx ofte bruger observationen funktionsniveau/ egenomsorg som en standard observation. Af de observationer, som modellen er god til at identificere, kan f.eks. 'Ansøgning om hjælpemidler' og 'Kontakt til læge' fremhæves, med accuracy på henholdsvis 100% og 98%.

Delkonklusion - valg af observationer

Der var støj i data, som bl.a. kommer fra dokumentation i de forkerte observationsformularer. Ved at benytte superbruger-datasæt kunne vi reducere støjen i data, da superbrugerne repræsenterer en dokumentation som var af høj kvalitet. Dette resulterede også i højere model performance.

På baggrund af opgavens kompleksitet var DL-modellerne overlegne, sammenlignet med traditionelle machine learning-modeller. DL-sprogmodellen, trænet på det store superbrugerdatasæt, opnåede en accuracy på 83%. Det vil sige, at modellen i 83% af tilfældene kunne identificere den rigtige observation til en given tekst. Baseret på fejlanalysen viser det, at modellen generelt har en god performance. Dog er der enkelte observationstyper, hvor modellen ikke performer på samme niveau. Det kan være grundet de vage definitioner og den brede anvendelse af brugerne.

Sætningsplacering

Som tidligere beskrevet, foreslås der en anden måde at dokumentere i Cura på, end man gør i dag: i stedet for at vælge en observation på forhånd og dokumentere i denne formular, foreslås der i dette projekt, at man beskriver hele sit borgerbesøg i ét samlet fritekstfelt. Herefter kan flere forskellige observationer blive valgt som relevante for den givne dokumentation.

En beskrivelse af et helt besøg vil dog ikke altid være relevant for alle valgte observationer, og derfor forsøger sætningsplacering automatisk at placere hver individuel sætning i den rigtige observationstype. Det er tiltænkt, at denne metode kan spare brugeren tid, da man ikke behøver at oprette en ny observationsformular til hver individuel observation.

Data

Sætningsplacerings-opgaven går ud på at klassificere, hvilken observation en sætning omhandler. Som klargøring til denne opgave, brugte vi superbruger-datasættet defineret af Cura-eksperter/systemadministrator. Dette indebar 28 superbrugere, som tilsammen havde dokumenteret i 66 forskellige observationer. Teksten blev opdelt per sætning. Modsat tidligere opgave (afsnit 5) indeholder data nu ikke den samlede beskrivelse af et borgerbesøg, men sætningerne herfra. Dette gjorde, at vi kunne betragte opgaven som en binær klassifikationsopgave: tilhører en given sætning observationstype X (ja eller nej)?

Preprocessing

For at klargøre et datasæt til denne opgave, var der brug for 'sande' og 'falske' tekster: sande tekster betyder, at den observationstype, der står ved en sætning, er korrekt. Falske tekster betyder, at en sætning og observationstype ikke passer sammen. For at skabe disse tekster, blev teksten fra et borgerbesøg opdelt i individuelle sætninger. Herefter blev de enkelte sætninger sat sammen med et True mærkat. For at lave de falske sætninger, blev sætningerne matchet op med tilfældige forkerte observationstyper og sat sammen med et False mærkat. Med disse mærkater kan modellen lære, hvilke sætninger der er klassificeret korrekt, og hvilke som er forkerte.

Ligesom i tidligere opgave, blev der fjernet duplikater fra data, og sætninger kortere end fem ord blev fjernet. Det resulterede i 154.671 datapunkter, heraf blev de sorteret efter tid, og de sidste 16.000 blev brugt som testsæt. Det resterende data blev splittet i et trænings- og valideringssæt (80/20 split), resulterende i hhv. 111.772 og 27.944 datapunkter.

Modeller

Forskellige DL-sprogmodeller er blevet afprøvet for at håndtere sætningsplacerings-opgaven. Sprogmodellerne har været prætrænede med enten dansk eller flere sprog. Data science teamet har brugt transfer learning til at gøre modellerne specifikke til den givne opgave.

DL-modellerne er tunge, og de to største udfordringer med disse modeller var 1) det tager tid at træne modellerne, og 2) dette er på bekostning af accuracy. De modeller, som gav højest accuracy, var også dem, som tog længst tid om at træne og give svar tilbage til brugeren. Omvendt, de modeller som gav lavere accuracy var også dem,

der var hurtige til at træne og give et svar tilbage til brugeren. Gennem forskellige evalueringer med Cura-eksperter og brugertest med folk fra hjemmeplejen, blev det tydeligt, at tid er en vigtig faktor. Der skulle helst ikke gå mange sekunder før modellen skulle komme med en prædiktion, også selvom modellen ikke var lige så nøjagtig, som hvis en langsommere model var blevet valgt. Den model, som i dag bliver brugt i demo-applikationen, er derfor en dansk sprogmodel ved navn Ælæctra. Ælæctra har en accuracy på 64.97%, men den langsommere model (multilingvistisk) havde en accuracy på 67.99%.

Fejlanalyse

Som tidligere nævnt, forekommer der støj i data, og en af årsagerne kan være, at brugerne ofte vælger observationer, som de er vant til at bruge, i stedet for at lede efter en mere passende observationstype i Cura. For at evaluere, om brugerne har valgt den mest hensigtsmæssige observationstype, har en Cura-ekspert fra Thisted Kommune vurderet, hvilke observationstyper, som passer til 100 tilfældige sætninger. Udover at sammenligne ekspertens valg med, hvad brugerne havde valgt, blev eksperten også holdt op imod, hvad sætningsscorings-modellen ville vælge. Modellen og Cura-eksperten valgte den samme observation 74% af gangene. Det var dog ikke muligt at sammenligne modellen med Cura-brugerne, da brugerne ikke har haft mulighed for at sætte den samme tekst i flere observationer, hvilket det foreslåede workflow i dette projekt bygger på (ét samlet fritekst-felt).

Delkonklusion - sætningsplacering

Den model, som gav en høj performance samtidig med at være hurtig til at sende en prædiktion til brugerne, blev vurderet som bedst under de givne omstændigheder. Dette var den danske sprogmodel Ælæctra med en accuracy på 64.97%. Det betyder, at i 65% af tilfældene kan modellen identificere, om en sætning tilhører en observation eller ej.

Når en bruger har dokumenteret et borgerbesøg i demoapplikationen, og valgt de relevante observationer, kan hver sætning nu matches op med de valgte observationer. I 65% af tilfældene er sætningsscorings-modellen i stand til at mappe sætningerne til de valgte observationer.

Hvordan håndteres nye observationer i Cura?

På sigt vil Thisted Kommune gerne være i stand til at håndtere, hvis observationer fjernes eller tilføjes til dokumentationsprocessen i Cura. Der ændres løbende i felterne i observationstyperne, ligesom der også kan være behov for at slette og oprette nye observationstyper. For at denne prototype skal være i stand til at håndtere dette, har DS-teamet leveret forskellige løsningsforslag hertil.

Når der skal fjernes en observationstype, kan observationen blot blive fjernet fra de forslag, som demo-applikationen giver i første omgang. Senere kan der trænes en ny model, hvor de fjernede observationstyper ikke indgår i træningen.

For at tilføje observationstyper foreslås det, at modellen trænes på ny. Dette kræver, at en ekspert laver et antal eksempler på sætninger til den nye observationstype, da disse skal bruges under modeltræning. Problemet med denne metode er, at det kan være besværligt at beslutte det korrekte antal sætninger, der skal bruges for at opnå en god model performance. Der er lavet en indledende analyse af problemet, og det viser sig, at der kan være stor forskel på, hvor mange eksempler som er nødvendige. Det tyder på, at der skal bruges mellem 10-40 eksempler, afhængig af hvilken model der vælges. Denne mængde kan udvides over tid, som den nye observationstype bliver brugt af brugerne. Når flere eksempler er kommet ind i systemet, kan modellen trænes igen og lære fra disse eksempler. Det forventes, at dette ville forbedre performance for den enkelte observationstype.

Forslag til videreudvikling af NLP-sporet

Der er forskellige muligheder for at forbedre datakvalitet og modelarbejdet. Udviklingsforslag fra DS-teamet lyder bl.a. på:

1. Der er forskellig slags støj i data. Dels fordi brugerne har mange besøg pr. dag, og ikke har tid nok til grundig dokumentation. Et andet problem kan være at nogle af brugerne ikke er danske, og kan have problemer med at skrive dokumentation i korrekt dansk. For at mindste støjen kan der laves yderligere forbehandling af data, såsom at rette tegnsætning, stavefejl og grammatiske fejl.
2. I dette projekt har der været fokus på fritekstfelterne, men der findes også andre slags data, som datoer og numerisk data. Ved også at træne modeller til denne slags data og give forslag hertil, kunne det være en yderligere hjælp for brugerne.
3. Som nævnt er datasættet ubalanceret, dvs. observationerne er ikke ligeligt fordelt. Hovedproblemet med ubalancerede datasæt er, at det er svært at observere den sande accuracy både over de observationer, der findes mange af, og dem der findes få af. Ved at bruge ML teknikker som resampling (upsampling og downsampling) kan performance af modellerne muligvis forbedres.

Kirstine fra hjemmeplejen tester teknologier på Teknologi-dagen i juni 2021. De øvrige på billedet er fra Systematic og Hard Work, Mind & Magic


KIRSTINE
JAN