

Dokumentet er delvist forældet.

Grundet ændringer i projektets design er dele af dette dokument forældet.

Juni 2022

Det er ikke længere ambitionen at gennemføre afprøvning af Beslutningsstøtten på faktiske sager i dette projekt. Dermed er enhver beskrivelse af Pilottest 2 og Lodtrækningsforsøg forældet.

Baggrunden for ændringerne er beskrevet i "Notat om ændringer i projekt Underretninger i Fokus", som er tilgængelig på Trygfondens Børneforskningscenters hjemmeside <https://childresearch.au.dk/udsatte-boern-unge-og-familier/projekter/underretninger-i-fokus> og UC viden <https://www.ucviden.dk/da/projects/underretninger-i-fokus>.

Projektleder
Line Berg

Konsekvensanalyse vedrørende databeskyttelse

Dataansvarlige:

VIA University College (herefter VIA) og TrygFondens Børneforskningscenter, Aarhus Universitet (herefter AU). VIA og AU er selvstændige dataansvarlige for forskningsprojektets behandling af personoplysninger.

Udfyldt af:

Anne Marie Villumsen (VIA) og Michael Rosholm (AU)

Ansvarlig kontaktperson:

Anne Marie Villumsen (VIA) og Michael Rosholm (AU)

Repræsentant for ledelsen:

Andreas Rasch-Christensen (VIA) og Anne Marie Villumsen (VIA) og Niels Haldrup (AU)

Databeskyttelsesrådgivere:

Kristian Hohwü Nielsen (VIA) og Søren Broberg Nielsen (AU)

System, der anvendes til projektet:

Beslutningsstøtten

Versionsstyring

Version, dato	Bemærkninger
V.1, [26.02.2021]	Udarbejdelse af første version af konsekvensanalysen
V.2, [opdateringsdato]	[Indsæt bemærkning om årsag til opdatering]
V.3, [opdateringsdato]	[Indsæt bemærkning om årsag til opdatering]



TRYGFONDENS
BØRNEFORSKNINGSCENTER
INSTITUT FOR ØKONOMI
AARHUS UNIVERSITET



VIA University
College

Indholdsfortegnelse

SAMMENFATNING	3
1. INDLEDNING OG BAGGRUND	6
1.1 Konsekvensanalysens baggrund	6
1.2 Fordeling af dataansvar	7
1.3 Konsekvensanalysens formål	9
2. TRIN 1: SYSTEMATISK BESKRIVELSE AF BEHANDLINGEN AF PERSONOPLYSNINGER	10
2.1 Behandlingens formål	10
2.2 Behandlingens karakter	11
2.3 Behandlingens omfang	23
2.4 Sammenhæng og kontekst for behandlingen af personoplysninger i projektet	29
2.5 Modtagere af personoplysninger	30
2.6 Opbevaringsperiode for personoplysningerne i projektet	31
3. TRIN 2: INDDRAGELSE AF RELEVANTE INTERESSEENTER	32
3.1 Inddragelse af VIAs og AU's databeskyttelsesrådgiver (DPO)	32
3.2 Indhentning af de registreredes eller deres repræsentanters synspunkter	33
4. TRIN 3: PROJEKTETS LOVLIGHED, NØDVENDIGHED OG PROPORTIONALITET	35
4.1 Principperne om lovlighed, rimelighed og gennemsigtighed	35
4.2 Princippet om formålsbegrænsning	39
4.3 Princippet om dataminimering	40
4.4 Princippet om rigtighed	40
4.5 Princippet om opbevaringsbegrænsning	41
4.6 Princippet om integritet og fortrolighed (sikkerhed og robusthed)	41
4.7 De registreredes rettigheder	42
4.8 Databehandlere	43
4.9 Overførsel til tredjelande og/eller internationale organisationer	45
5. TRIN 4: IDENTIFIKATION, EVALUERING OG HÅNDTERING AF RISICI	45
5.1 Valg af evalueringskriterier for sandsynlighed og konsekvens	45
5.2 Risici	48
5.3 Residualrisiko	63
6. TRIN 5: KONKLUSION OG LEDELSESGODKENDELSE	65
6.1 Samlede residualrisiko og høring af Datatilsynet	65
6.2 VIAs ledelses godkendelse af konsekvensanalysen	66
6.3 AU's ledelses godkendelse af konsekvensanalysen	67
7. FASTLÆGGELSE AF EN IMPLEMENTERINGSPLAN	68
8. BILAG TIL KONSEKVENSANALYSEN	70
9. TRIN 6: AJOURFØRING AF KONSEKVENSANALYSEN	70

SAMMENFATNING

Formålet med denne konsekvensanalyse er at beskrive den behandling af personoplysninger, som VIA University College og TrygFondens Børneforsikringscenter, Aarhus Universitet, foretager i forskningsprojekterne "Beslutningsstøtte ved underretninger" og "Underretninger i fokus".

Den overordnede konklusion på konsekvensanalysen er, at residualrisikoen ved behandling af personoplysningerne i projektet er **lav til mellem med overvægt til lav risiko**. VIA og AU vurderer med andre ord, at de identificerede risici gennem mitigerende foranstaltninger, der allerede er eller vil blive implementeret, kan nedbringes til et acceptabelt niveau, så behandlingen ikke vil føre til en høj risiko for de registreredes rettigheder og frihedsrettigheder. Det er derfor også VIAs og AU's vurdering, at høring af Datatilsynet ikke er påkrævet.

De væsentligste risici ved behandlingen af personoplysninger i projektet er risikoen for ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (risiko nr. 4), risikoen for uautoriseret adgang til oplysningerne (risiko nr. 7) samt risikoen for manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset (risiko nr. 8).

Risiko nr. 4 er og vil fortsat blive imødekommet ved gennemgående træning af modellen med store mængder repræsentative data samt ved test og fravalg af variable med henblik på at sikre, at modellerne ikke giver anledning til prædiktioner baseret på bestemte parametre. Risikoen er og vil desuden fortsat blive nedbragt ved anvendelse af kvalificeret personale med et indgående statistisk kendskab til udvikling af modellerne.

Risiko nr. 7 er og vil fortsat blive nedbragt ved både bruger- og adgangsstyring, logning og kryptering.

Risiko nr. 8 er imødekommet ved, at VIA og AU har fastlagt dataansvar for begge piloter, og at hver part vil indgå de fornødne aftaler med relevante samarbejdspartner.

VIAs og AU's databeskyttelsesrådgivere bemærker, at det på baggrund af de i implementeringsplanen anførte mitigerende foranstaltninger, og forudsat, at disse implementeres som anført, er deres opfattelse,

at behandlingen sker i overensstemmelse med databeskyttelsesforordningen og supplerende bestemmelser i databeskyttelsesloven. De dataansvarlige skal løbende rapportere om gennemførelse af de mitigerende handlinger og er desuden forpligtet til at opdatere konsekvensanalysen løbende og forelægge enhver ændring af konsekvensanalysen for databeskyttelsesrådgiverne, hvor denne ændring medfører en øget risiko for den registreredes rettigheder og frihedsrettigheder.

VIA og AU's ledelse har på ovenstående baggrund godkendt konsekvensanalysen.

VIA og AU har for at imødekomme de identificerede risici fastlagt følgende implementeringsplan:

Afhjælpende foranstaltninger, der skal gennemføres	Frist for gennemførelse	Status på gennemførelse	Ansvarlig person
Uddannelse af socialrådgivere, øget transparens omkring inputdata og vægtning heraf samt træning og test af Beslutningsstøtten	Før afprøvning i Pilot #2 og indledningsvist under afprøvning i Pilot #2	Endnu ikke gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)
Dokumentation af Beslutningsstøttens formål, datagrundlag, modelvalg, performance mv.	01-03-21	Gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)
Retningslinjer for pseudonymisering	15-02-21	Gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)
Retningslinjer for periodisk evaluering af proportionaliteten	15-02-21	Gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)

Indgåelse af videregivelsesaftale mellem AU og VIA for Pilot #1	Videregivelsesaftale mellem AU og VIA for Pilot #1 er er indgået den 27. januar 2021.	Endnu ikke igangsat	Andreas Rasch-Christensen (VIA)
Indgåelse af videregivelsesaftale mellem AU og VIA for Pilot #2	Resten af de nævnte foranstaltninger skal være implementeret senest før videregivelse eller overførsel af data.		Michael Rosholm (AU)
Indgåelse af videregivelsesaftale mellem AU og samarbejdskommune for Pilot #2			
Indgåelse af databehandlersaftale(r) med databehandler(e) for Pilot #2			

Det er VIAs og AU's databeskyttelsesrådgivere, der er ansvarlige for at kontrollere, om implementerings-skridtene er foretaget. Det er de dataansvarlige, der er ansvarlige for at ajourføre og vedligeholde denne konsekvensanalyse.

1. INDLEDNING OG BAGGRUND

1.1 Konsekvensanalysens baggrund

VIA University College (herefter VIA) og TrygFondens Børneforsikringscenter, Aarhus Universitet (herefter AU), igangsatte i 2017 forskningsprojekterne ”Beslutningsstøtte ved underretninger” og i 2019 ”Underretninger i fokus” (begge herefter benævnt projektet). Et af de overordnede formål med projektet er at udvikle et beslutningsstøtteredskab (herefter Beslutningsstøtten), der kan bistå kommunernes socialfaglige praksis, herunder socialrådgivere, med at vurdere underretninger om børn og unges mistrivsel. Tanken er, at Beslutningsstøtten skal understøtte socialrådgiverens sagsbehandling ved – på baggrund af de tilgængelige oplysninger i sagen – at udregne en konkret risikoscore, der indikerer risikoen for et barns eller en ung persons mistrivsel. Risikoscoren har i denne forbindelse til formål at skærpe socialrådgiverens beslutningsgrundlag, og den må alene anvendes som beslutningsstøtte. I løbet af forskningsprojektet skal Beslutningsstøtten afprøves med fokus på redskabets præcision, anvendelighed, påvirkningskraft samt betydning for socialrådgivernes arbejde.

Delprojektet ”*Beslutningsstøtte ved underretninger*” består af pilotforsøget Pilot #1, der udover selve udviklingen af Beslutningsstøtten inkluderede en egentlig afprøvning og evaluering af Beslutningsstøtten. Afprøvningen blev foretaget med to samarbejdskommuner, Silkeborg Kommune og Hjørring Kommune, og fandt sted i perioden november 2018 til februar 2019.

VIA og AU besluttede på baggrund af erfaringerne fra Pilot #1, at Beslutningsstøtten skulle revideres. Revisionen har imidlertid været så omfattende, at VIA og AU vurderer, at Beslutningsstøtten skal afprøves ved et nyt pilotforsøg (Pilot #2), inden redskabet afprøves i større skala i form af et RCT/lodtrækningsforsøg. I Pilot #2 vil afprøvningen alene blive foretaget i én kommune. Pilot #2 og RCT/lodtrækningsforsøget hører samlet under delprojektet ”*Underretninger i Fokus*”.

VIA og AU har vurderet, at VIA og AU har pligt til at udarbejde en konsekvensanalyse for projektet. Baggrunden herfor er, at Beslutningsstøtten bygger på maskinlæring, hvor navnlig risiko for manglende transparens i forhold til systemets processer, risiko for de facto fuldautomatiske afgørelser samt risiko for fejl i Beslutningsstøtten øger risikoen for krænkelse af fysiske personers rettigheder og frihedsrettigheder. Derudover forudsætter afprøvningen af Beslutningsstøtten behandling af personoplysninger om

sårbare fysiske personer, herunder børn og unge, samt behandling af følsomme personoplysninger, hvilket ligeledes øger risikoen for fysiske personers rettigheder og frihedsrettigheder.

Konsekvensanalysen vil alene inddrage den behandling, der har og fortsat finder sted i forbindelse med Pilot #1, og som vil finde sted i forbindelse med Pilot #2. Konsekvensanalysen vil derimod ikke adressere den behandling, der påtænkes at finde sted i forbindelse med RTC/lodtrækningsforsøget, idet rammerne for denne fase af forskningsforsøget endnu ikke er tilstrækkeligt fastlagt. Der vil ske en ajourføring af konsekvensanalysen, inden projektet overgår til denne og eventuelle efterfølgende faser.

1.2 Fordeling af dataansvar

VIA og AU er enerådige parter i forskningsprojektet, og hver part behandler personoplysningerne i projektet til egne formål og med egne hjælpemidler. VIA og AU må således anses som selvstændige dataansvarlige i relation til deres behandling af personoplysninger i projektet.

VIA og AU igangsatte forskningsprojektet med det overordnede formål at udvikle et beslutningsstøtteredskab til socialrådgiveres vurdering af underretninger om børns og unges mistrivsel. Hvor VIAs fokus er kvalitativt på Beslutningsstøttens anvendelighed for socialrådgiverne og familierne, herunder om Beslutningsstøtten forbedrer sagsbehandlingen, er AU's fokus dog rettet mod den kvantitative forskning, herunder hvorvidt det er statistisk muligt at forudsige mistrivsel. Disse respektive delformål har afspejlet sig i parternes arbejdsdeling, herunder i forhold til behandlingen af personoplysninger. VIA har således fokuseret på de kvalitative aspekter af forsøget, herunder blandt andet interview- og observationsdata, mens AU har fokuseret på de kvantitative aspekter, herunder bl.a. udarbejdelse af den statistiske model på baggrund af registerdata og udarbejdelse af deskriptive analyser på baggrund af kvantitative data fra pilotafprøvningen mv.

VIA og AU har i lyset af de adskilte fokusområder desuden ikke i fuldt omfang behandlet de samme data. Det er f.eks. alene AU, der behandler data fra Danmarks Statistiks registre mv., og kun VIA, der har adgang til rådata fra interviews og observationer. VIA og AU har desuden anvendt forskellige hjælpemidler til behandlingen af personoplysninger. AU har f.eks. anvendt statistikprogrammer og statistiske analysemetoder, mens VIA f.eks. har anvendt programmer til kodning af kvalitative interviews.

VIA og AU har løbende orienteret hinanden om valg af bl.a. variabler, emner til interviewguides og hjælpemidler. Denne orientering har dog alene haft karakter af sparring og vidensdeling, og hver part har således selvstændigt truffet de endelige beslutninger, herunder selvstændigt fastlagt hvilke hjælpemidler, der bedst anvendes til behandlingen af oplysningerne.

Både Silkeborg Kommune og Hjørring Kommune agerede under Pilot #1 desuden som selvstændige dataansvarlige.

VIA og AU vil i Pilot #2 fortsat agere som selvstændige dataansvarlige, hvilket tillige vil gælde for den deltagerkommune, som vil indgå i Pilot #2.

1.2.1 Nærmere om VIAs og AU's arbejdsdeling

AU har ansvaret for at udvikle og validere den statistiske algoritme, som Beslutningsstøtten er baseret på. AU har og vil i den forbindelse fortsat udvikle modellerne i Danmarks Statistiks forskningsserver på grundlag af pseudonymiserede registerdata, hvortil alene AU har adgang. AU har desuden stået for den løbende validering af algoritmen.

AU anvendte i forbindelse med udviklingen og afprøvningen i Pilot #1 IT-virksomheden Indiesoft, der havde til opgave at udarbejde den software, der skulle anvendes i Hjørring Kommune og Silkeborg Kommune. Indiesoft havde desuden til opgave at udtage dataudtræk fra Beslutningsstøttens afprøvning i kommunerne og overføre dette dataudtræk til AU, så AU kunne udarbejde deskriptive statistiske analyser på baggrund heraf.

VIA gennemførte forud for udviklingen af modellen i Pilot #1 kvalitative forstudier, herunder interviews med socialrådgivere fra 4 kommuner samt et kvalitativt studie af et antal underretninger modtaget i Silkeborg Kommune og Hjørring Kommune. Dette for at konkretisere den beslutningsproces, Beslutningsstøtten skulle bistå med at skærpe. VIA gennemførte desuden løbende interviews og observationer under Pilot #1. Indiesoft videregav på vegne af AU desuden dataudtrækket fra afprøvningen til VIA, for at VIA kunne undersøge, hvordan Beslutningsstøtten påvirkede socialrådgivernes beslutningsproces.

1.3 Konsekvensanalysens formål

Formålet med denne konsekvensanalyse vedrørende databeskyttelse er at beskrive den behandling af personoplysninger, som VIA og AU foretager i forbindelse med projektet.

Konsekvensanalysen har endvidere til formål at afdække de risici, der er forbundet med behandlingen af personoplysninger i projektet samt bidrage til at håndtere disse risici for fysiske personers rettigheder og frihedsrettigheder ved at vurdere dem og fastlægge effektive foranstaltninger til at afhjælpe dem.

Denne konsekvensanalyse er udarbejdet i overensstemmelse med minimumskravene til en konsekvensanalyses indhold i databeskyttelsesforordningens artikel 35, stk. 7, samt i bilag 2 til Artikel 29-Gruppens (nu: Det Europæiske Databeskyttelsesråd) vejledning om konsekvensanalyser.¹

Konsekvensanalysen skal således bidrage til at sikre overholdelse af databeskyttelsesforordningens² og databeskyttelseslovens³ regler, ligesom den er en væsentlig forudsætning for overholdelse af forordningens grundlæggende princip om ansvarlighed – dvs. dokumentation for overholdelse af forordningens regler, jf. forordningens artikel 5, stk. 2, og artikel 24.

¹ Artikel 29-Gruppen (nu Det Europæiske Databeskyttelsesråd, herefter forkortet "EDPB"): Retningslinjer for konsekvensanalyse vedrørende databeskyttelse (DPIA) og bestemmelse af, om behandlingen "sandsynligvis indebærer en høj risiko" i henhold til forordning (EU) 2016/679, WP 248, rev. 01, revideret og endeligt vedtaget den 4. oktober 2017.

² EUROPA-PARLAMENTETS OG RÅDETS FORORDNING (EU) 2016/679 af 27. april 2016 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger og om ophævelse af direktiv 95/46/EF (generel forordning om databeskyttelse) (EØS-relevant tekst).

³ Lov nr. 502 af 23. maj 2018 om supplerende bestemmelser til forordning om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger og om fri udveksling af sådanne oplysninger (databeskyttelsesloven).

2. TRIN 1: SYSTEMATISK BESKRIVELSE AF BEHANDLINGEN AF PERSONOPLYSNINGER

2.1 Behandlingens formål

Antallet af underretninger om mistrivsel blandt børn og unge er tredoblet i løbet af det seneste årti. Disse underretninger indeholder ofte få konkrete oplysninger, hvilket gør det vanskeligt for socialrådgiveren at vurdere barnets risiko for mistrivsel. Dette er problematisk, idet børn og unges udvikling og trivsel har afgørende betydning for deres muligheder senere i livet. Kommunernes håndtering af underretninger, hvor der udtrykkes bekymring for et barn eller en ung, har derfor væsentlig betydning for barnets eller den unges fremtid.

VIA og AU ønsker at imødegå ovenstående udfordringer og undersøge, hvorvidt det er muligt at udvikle et beslutningsstøtteredskab, der kan skærpe beslutningsgrundlaget og understøtte socialrådgiverens vurdering af underretninger om børns og unges mistrivsel.

Formålet med behandlingen af personoplysninger i projektet er således at undersøge:

1. Hvorvidt den prædiktive risikomodel, som Beslutningsstøtten bygger på, er egnet til at forudsige mistrivsel,
2. hvordan socialrådgiverne oplever at arbejde med Beslutningsstøtten, samt
3. hvordan borgerne oplever anvendelsen af Beslutningsstøtten.

VIA og AU har, som led i projektet, gennemført Pilot #1 i perioden fra november 2018 til februar 2019 med henblik på at afprøve Beslutningsstøttens egnethed som beslutningsstøtteredskab. VIA og AU har siden foretaget en række grundlæggende ændringer af Beslutningsstøtten, og VIA og AU vil nu afprøve, hvad disse ændringer betyder for den socialfaglige praksis, herunder hvorvidt Beslutningsstøtten i sin nuværende form er egnet til at indgå i RCT/lodtrækningsforsøget, eller hvorvidt Beslutningsstøtten skal tilpasses yderligere, forinden den afprøves i en større skala.

VIA's og AU's behandling af personoplysningerne i projektet sker udelukkende i statistisk og forskningsmæssigt øjemed, og hverken VIA eller AU vil senere anvende oplysningerne til andre formål.

2.2 Behandlingens karakter

2.2.1 Udvikling

Beslutningsstøtten baserer sig på statistiske modeller, der er udviklet ved brug af maskinlæring. Udviklingen af modellerne har fundet sted i Danmarks Statistiks forskningsserver på grundlag af pseudonymiseret registerdata, herunder data fra danske underretningssager. For en nærmere opregning af de registre og variabeltyper, som Beslutningsstøtten udvikles på baggrund af, se Bilag A, hvis indhold ligeledes er gengivet i punkt 2.3.2. Det er AU's forventning, at de ved at oplære og træne modellerne med denne type og mængde af data kan skabe en statistisk algoritme, der er egnet til at forudsige barnets eller den unges risiko for fremtidig "mistrivsel". Ved udviklingen af modellerne har AU alene anvendt de former for registerdata, der også vil være tilgængelige for socialrådgiverne i kommunerne, når de behandler disse underretningssager efter de gældende regler herom.

Udviklingen af modellerne er sket gradvist og har således ændret sig i takt med, at VIA og AU har opbygget erfaring, herunder ved hjælp af Pilot #1. Fælles for modellerne er dog, at de er statiske. Som følge heraf er de ikke i stand til at udvikle sig på baggrund af nye inputdata efter ibrugtagning i praksis, herunder under pilotforsøgene. Derudover er det væsentligt at understrege, at modellerne alene består af en matematisk formel (algoritme) uden personoplysninger. Så snart modellerne er færdigudviklede i Danmarks Statistiks forskningsserver, udtager AU algoritmen, men ikke de personoplysninger, som algoritmen er blevet udarbejdet på baggrund af. Derudover vil de data, som Beslutningsstøtten afprøves på, ikke blive lagret i modellen. Således vil der ikke forekomme samkøring af Danmarks Statistiks oplysninger og kommunernes oplysninger, ligesom samkøring af personoplysninger på tværs af forvaltninger ikke vil finde sted.

AU har som et led i valideringen af modellens egnethed til at forudsige mistrivsel forud for Pilot #2 gennemført en række analyser med henblik på at sikre, at modellerne er følsomme over for andre saglige mål for mistrivsel, herunder f.eks. fremtidige sigtelser for kriminalitet og fremtidige somatiske diagnoser. Disse analyser er alene gennemført med henblik på at sikre modellernes validitet. Data anvendt til validering, herunder oplysninger om helbred eller kriminalitet, indgår således ikke i selve modellerne, og dermed ikke i den konkrete udregning af risikoscoren. AU har ved hjælp af tests i udviklingsfaserne desuden verificeret, at modellerne ikke giver anledning til prædiktioner baseret på usaglige parametre, som f.eks. etnicitet. Det bemærkes i den forbindelse, at der ligeledes er gennemført en forvaltningsretlig

legalitetskontrol af Beslutningsstøtten, hvor det blandt andet er sikret, at modellerne ikke inddrager ulovlige variable (hensyn) efter forvaltningsretten.

2.2.1.1 Pilot #1

Til brug for Pilot #1 udviklede AU i perioden september 2017 - oktober 2018 en lineær statistisk model i Danmarks Statistiks forskningsserver. AU valgte den lineære statistiske model, idet denne form for model er relativ enkel at implementere og forklare. Derudover valgte AU at udvikle modellen ved brug af POST-LASSO-metoden, idet denne anerkendte statistiske metode indebærer målrettet udvælgelse af de mest nødvendige variable.

POST-LASSO-metoden består overordnet set af to trin: variabeludvælgelse og estimering af modellen. I forbindelse med variabeludvælgelsen bliver systemet "straffet" for at have for mange variable. Som følge heraf udvælger det alene de variable, der er mest nødvendige for at nå et outcome. Variable med begrænset forklaringskraft bliver således fravalgt for at sikre, at datagrundlaget ikke er større end nødvendigt. Variabeludvælgelsen skete på baggrund af data fra alle danske underretningssager i perioden april 2014-2015. Efterfølgende blev modellen estimeret, dvs. at vægtningen af de forskellige variable blev fastlagt. Vægtningen blev fastlagt på baggrund af samme grundlag som variabeludvælgelsen.

Forud for udviklingen af modellen gennemførte VIA et kvalitativt studie af alle underretninger modtaget i Silkeborg Kommune og Hjørring Kommune i perioden januar til februar 2017. Formålet hermed var at konkretisere den beslutningsproces, Beslutningsstøtten skulle bistå med at skærpe. Underretningerne blev videregivet til VIA af Silkeborg Kommune og Hjørring Kommune. Kommunerne havde inden videregivelsen effektivt anonymiseret underretningerne.

VIA foretog i maj og juni 2017 forud for udviklingen af modellen kvalitative interviews i 4 kommuner: Hjørring, Silkeborg, Århus og Holstebro. Formålet med interviewene var at afdække den beslutningsproces, det er at vurdere underretninger. Der blev fx spurgt til, hvordan vurderingen af underretninger foregår i den pågældende kommune (har de fx et kategoriseringssystem), og hvordan forskellige informationer vægtes, når socialrådgiveren skal vurdere, hvor bekymrende indholdet af underretningen er, og hvor hurtigt der bør ageres på det.

Børn og unges udsathed eller mistrivsel er et komplekst mål, der ikke direkte kan observeres i de kommunale systemer eller i nationale registre. VIA og AU blev derfor nødt til at finde en indikator, der var egnet til at identificere børn og unge, der er i risiko for at mistrives. Konkret valgte VIA og AU at operationalisere ”mistrivsel” til mindst et af følgende tre outcomemål:

- Anbringelse uden for hjemmet (inden for seks måneder efter modtagelse af en underretning),
- De sociale myndigheder modtager en grov underretning (inden for seks måneder efter modtagelse af en underretning), eller
- Iværksættelse af en alvorlig § 52-foranstaltning⁴ (inden for seks måneder efter modtagelse af en underretning).

Tidsintervallet på seks måneder skyldtes begrænsninger i de tilgængelige registerdata.⁵

Baseret på den tilgængelige registerdata udvalgte modellens algoritme herefter følgende ni variable, som den fandt, forklarede det meste af variationen i de tre ovenstående outcomemål, og som således havde den største forklaringskraft:

- Barnets alder
- Antal tidligere underretninger inden for de seneste 90 dage
- Antal tidligere underretninger inden for de seneste 180 dage
- Om nærværende underretning vedrørende kriminalitet (hos barnet)

⁴ Det følger af servicelovens § 52, at kommunalbestyrelsen kan iværksætte hjælp inden for forskellige typer af tilbud, når barnet eller den unge har behov for særlig støtte. På nuværende tidspunkt kan følgende foranstaltninger anses for alvorlige § 52-foranstaltninger: Etablering af aflastningsordning i en almen plejefamilie, i en forstærket plejefamilie, i en specialiseret plejefamilie eller i en netværksplejefamilie, på et opholdssted eller på en døgninstitution (servicelovens § 52, stk. 3, nr. 5), Udpegning af en fast kontaktperson for barnet eller den unge eller for hele familien (servicelovens § 52, stk. 3, nr. 6), Formidling af praktiktilbud hos en offentlig eller privat arbejdsgiver for den unge og i den forbindelse udbetaling af godtgørelse til den unge (servicelovens § 52, stk. 3, nr. 7) og Meddelelse af børne- og ungepålæg (servicelovens § 57 b).

⁵ Begrænsningerne skyldes, at underretningsregisteret (BUU) alene indeholder data tilbage fra april 2014, hvortil kommer, at der er en vis forsinkelse mellem, at underretningerne modtages og registres i kommunen, til de er tilgængelige i Danmarks Statistiks registre. AU havde på tidspunktet, hvor modellen til Pilot #1 blev udviklet, derfor alene adgang til data fra april 2014 og 2015. AU havde brug for mindst et halvt års underretningsdata før modtagelse af underretningen, og ovenstående begrænsninger i datagrundlaget tillod derfor alene AU at kigge 6 måneder frem efter modtagelse af underretningen.

- Om nærværende underretning vedrørende overgreb eller vold mod barnet
- Om nærværende underretning vedrørende forældres misbrug
- Antal af § 52-foranstaltninger inden for de seneste 180 dage
- Antal tidligere anbringelser inden for det seneste år
- Antal tidligere anbringelser inden for de seneste 5 år

Forud for afprøvningen af Beslutningsstøtten indledte VIA og AU et samarbejde med kommunernes datavarehuse, der er administreret af henholdsvis EG og KMD, med henblik på at sikre, at de udvalgte variable var defineret på samme måde som de data, AU har haft tilgængelige via Danmarks Statistiks registre. I de tilfælde, hvor der var uoverensstemmelser, foretog Indiesoft en omkodning af data på vegne af AU (f.eks. ved at tælle antal tidligere underretninger inden for de seneste 90 dage). Dette for at sikre, at data i praksis operationaliseres på samme måde, som i de registre, modellen er udviklet på, og dermed også modellens egnethed til at indgå i Pilot #1.

2.2.1.2 Pilot #2

På baggrund af de erfaringer, som blev gjort i Pilot #1, besluttede VIA og AU at udarbejde nye statistiske modeller med henblik på at forbedre Beslutningsstøttens præcision. De nye modeller skal udvikles på baggrund af registerdata fra april 2014-2018, herunder i alt 506.718 underretninger fra den pågældende periode. Således ønsker VIA og AU at træne modellerne med en væsentlig større mængde data, end hvad der var tilfældet i Pilot #1.

I de nye statistiske modeller operationaliseres ”*mistrivsel*” alene til ét enkelt outcomemål:

- Anbringelse uden for hjemmet (inden for 12 måneder efter modtagelse af en underretning).

De nye modeller skal alene trænes til at prædiktere barnets eller den unges risiko for anbringelse inden for 12 måneder. Baseret på de tilgængelige registerdata skal modellerne beslutte, hvilke variable der har betydning for målet om anbringelse. Der skal således ske en variabeludvælgelse ved brug af supervised maskinlæring. I modsætning til Pilot #1 skal de nye modeller inddrage flere variable end modellen i Pilot #1. VIA og AU har valgt at øge antallet af variable med henblik på at gøre modellen mere præcis samt

for at give socialrådgiverne et mere fyldestgørende indblik i datagrundlaget og hermed styrke redskabets funktion som beslutningsstøtte.

De præcise modeller, der skal anvendes til Pilot #2, er endnu ikke fastlagt. Dog overvejer VIA og AU at anvende modeller baseret på beslutningstræer, herunder anerkendte statistiske modeller/metoder som f.eks. XGBoost og Random Forest.

Til brug for validering af algoritmens performance har AU i henholdsvis april 2019⁶ og februar 2020 til brug for Pilot #2 desuden fået videregivet registerdata fra Styrelsen for IT og Sundhedsdatastyrelsen.⁷ Data er sendt direkte til Danmarks Statistik, som har pseudonymiseret data.

2.2.2 Afprøvning

Det er indledningsvis værd at bemærke, at Beslutningsstøtten alene er et beslutningsstøtteredskab, og at Beslutningsstøtten under ingen omstændigheder må anvendes til de facto at træffe fuldautomatiske afgørelser. VIA og AU har og vil fortsat sikre, at Beslutningsstøtten ikke benyttes til at træffe fuldautomatiske afgørelser ved at undervise kommunens socialrådgivere i brug af Beslutningsstøtten samt i, hvordan Beslutningsstøtten kommer frem til sine risikoscorer. Desuden skal det bemærkes, at Beslutningsstøtten i Pilot #1 alene blev anvendt, og i Pilot #2 fortsat alene vil blive anvendt på de data, kommunen er i besiddelse af, og som dermed allerede er i forvaltningens sagsbehandlingssystem. Der indsamles ikke nye data til brug for sagen, og der forekommer ikke samkøring af data på tværs af kommunale forvaltninger.

2.2.2.1 Pilot #1

Den første afprøvning af Beslutningsstøtten blev som nævnt gennemført i perioden november 2018 til februar 2019 i henholdsvis Silkeborg Kommune og Hjørring Kommune.

⁶ Efterfølgende videregivelser vil ske hvert efterår, typisk i oktober/november måned, indtil projektet afsluttes.

⁷ AU har fra Styrelsen for IT og Læring fået videregivet data fra "Den Nationale Trivselsmåling", de nationale test i dansk og matematik i folkeskolen og data om skolefravær. AU har desuden fået videregivet data fra "Sundhedsstyrelsens Centrale Odontologiske Register (SCOR)" af Sundhedsdatastyrelsen.

AU testede forud for implementeringen af Beslutningsstøtten og selve afprøvningen Beslutningsstøtten på en række konkrete underretningssager videregivet fra kommunerne for at sikre, at algoritmen inddrog de rigtige data, og at der blev beregnet en korrekt risikoscore for fremtidig mistrivsel. AU's databehandler, Indiesoft, foretog desuden et dataudtræk efter igangsættelsen af pilotafprøvningen, hvorefter AU tjekkede, at risikoscoren blev beregnet rigtigt.

I forbindelse med afprøvningen i Pilot #1 blev begge kommuners socialrådgivere instrueret i at vurdere og behandle underretningssagerne på almindelig vis. Rådgiverne skulle således registrere den modtagne underretning i kommunens sagsbehandlingssystem som sædvanligt og foretage den socialfaglige vurdering inden ibrugtagningen af Beslutningsstøtten.

Socialrådgiveren blev efterfølgende bedt om at åbne Beslutningsstøtten, der automatisk (Silkeborg Kommune) eller efter indtastning af barnets eller den unges CPR-nummer (Hjørring Kommune) indhentede de relevante oplysninger⁸ fra underretningssagen i sagsbehandlingssystemet – alt afhængig af kommunens tekniske setup. I Beslutningsstøtten blev socialrådgiveren bedt om at vurdere, hvor bekymret socialrådgiveren var for barnets eller den unges trivsel og udvikling på en skala fra 1-10. Dernæst skulle socialrådgiveren begrunde sin vurdering. Socialrådgiverens vurdering byggede i denne forbindelse på den konkrete underretning samt den viden, som socialrådgiveren ellers havde om barnet eller den unge.

Efterfølgende beregnede Beslutningsstøtten risikoen for, at barnet eller den unge inden for de kommende seks måneder 1) ville blive anbragt uden for hjemmet, 2) at kommunen ville modtage en grov underretning, og/eller 3) at der ville blive iværksat en alvorlig § 52-foranstaltning. Beslutningsstøtten præsenterede herefter risikoscoren for socialrådgiveren. Konkret konverterede Beslutningsstøtten den beregnede sandsynlighed til deciler, hvor en risikoscore på 10 angiver den højeste risiko, mens 1 er den laveste. En risikoscore på 10 svarer således til, at barnet eller den unge er blandt de 10 % børn, der underrettes om, og som har den højeste risiko for at opleve en af ovenstående tre udfald. Omvendt vil en risikoscore på 1 betyde, at barnet eller den unge er blandt de 10 % børn, der underrettes om, som har den laveste risiko for at opleve en af ovenstående tre udfald.

⁸ En nøjagtig angivelse af, hvilke oplysninger Beslutningsstøtten indhentede fra kommunernes sagsbehandlingssystem, fremgår af ”20200204_Data og modelbeskrivelse vedrørende projekt Beslutningsstøtte ved underretninger”, der er bilagt denne konsekvensanalyse som Bilag B.

Socialrådgiveren blev slutteligt bedt om at revurdere, hvor bekymret vedkommende var for barnet eller den unges trivsel og udvikling på en skala fra 1-10 samt begrunde vurderingen med udgangspunkt i alt tilgængelig information, herunder den nu beregnede risikoscore. Formålet med dette trin var at følge, hvorvidt og eventuelt hvordan socialrådgiveren ændrede sin indledningsvise vurdering ved brug af Beslutningsstøtten. Hverken Beslutningsstøttens risikoscore eller socialrådgiverens begrundelse blev journaliseret på sagen, og efter endt brug af Beslutningsstøtten havde hverken den relevante socialrådgiver eller andre sagsbehandlere i kommunerne adgang til socialrådgiverens vurdering af barnet eller den unge i Beslutningsstøtten. Alene AU's databehandler Indiesoft havde adgang til systemets data og kunne trække rapporter på disse. Under afprøvningen videregav begge kommuner således løbende data til AU's databehandler Indiesoft.

Endelig fulgte VIA nogle af socialrådgiverne gennem interviews og observationer i løbet af Pilot #1. Der blev i forbindelse med de enkelte interviews ikke drøftet konkrete sager, men alene socialrådgiverens vurdering af Beslutningsstøttens anvendelighed.

2.2.2.2 Pilot #2

VIA og AU besluttede på baggrund af erfaringerne fra Pilot #1 at foretage en omfattende revision af Beslutningsstøtten. VIA og AU besluttede først og fremmest at ændre Beslutningsstøttens statistiske model. Der er redegjort for de konkrete modelændringer i denne konsekvensanalyses punkt 2.2.1. VIA og AU vil dog også foretage fire andre ændringer, der ligeledes har betydning for behandlingen af personoplysningerne. De fire ændringer, som VIA og AU påtænker at afprøve i Pilot #2, er følgende:

- Ændring af workflow

VIA og AU vil fremadrettet lade socialrådgiveren tilgå Beslutningsstøtten som et af de første skridt efter at have modtaget og læst underretningen. Derved vil Beslutningsstøtten indgå i opstarten af sagsbehandlingen, og således *før* socialrådgiveren har foretaget sin socialfaglige vurdering. Denne ændring sker i pilot #2, fordi det nu er tid at afprøve beslutningsstøtten sådan, som det er intentionen at gøre i et lodtrækningsforsøg. Dette skal afspejle den måde, den påtænkes brugt i socialfaglig praksis. I pilot #1 handlede det om at finde ud af, hvorvidt en beslutningsstøtte kunne udvikles, og hvordan socialrådgiverne oplevede at bruge den. I pilot #2 handler det også om at afprøve beslutningsstøtten i en konkret socialfaglig praksis.

- Øget informationsniveau

Pilot #1 viste, at mange af socialrådgiverne fandt, at Beslutningsstøtten ikke gav tilstrækkeligt med information til, at de kunne bruge risikoscoren i deres beslutningsproces. Da Beslutningsstøtten alene udregnede en risikoscore uden yderligere begrundelse herfor, fandt socialrådgiverne det usikkert, hvilke informationer Beslutningsstøtten havde inddraget og vægtet, og hvilke supplerende informationer, socialrådgiverne eventuelt måtte besidde. VIA og AU har med henblik på at sikre størst mulig transparens derfor besluttet at 1) implementere en dataoversigt, der opsummerer al data, som Beslutningsstøtten har anvendt til at beregne sin risikoscore, og 2) at præsentere socialrådgiverne for et variabelindeks, der giver dem indblik i, hvilke typer af informationer, der har trukket scoren op eller ned. VIA og AU vil i Pilot #2 således undersøge, om socialrådgiverne oplever, at denne ændring i informationsniveauet tilbyder tilstrækkelig med information, samt om ændringen er et brugbart bidrag i beslutningsprocessen. Endvidere vil VIA og AU undersøge, hvordan de ekstra informationer bliver anvendt i socialrådgivernes vurdering af underretningerne for at sikre, at Beslutningsstøtten rent faktisk (og på korrekt vis) understøtter de beslutninger, socialrådgiverne skal træffe. Rådgiverens indsigt i, hvilke data risikoscoren er udregnet på baggrund af, vil desuden fremme socialrådgiverens vurdering af, om der er fejl i data, eller om der er nye væsentlige oplysninger i sagen, der ikke er registreret. Socialrådgiveren vil således altid have flere oplysninger tilgængelige end det informationsgrundlag, som Beslutningsstøtten har udregnet sin score på baggrund af.

- Socialrådgiverens vurderinger af risikoscore fjernes

I Pilot #1 blev socialrådgiveren bedt om at vurdere og give underretningen en risikoscore, inden vedkommende modtog Beslutningsstøttens risikoscore samt efter at have set den beregnede score. Afprøvningen viste, at socialrådgiverne i meget begrænset omfang ændrede deres initiale risikovurdering. VIA og AU har ikke været i stand til entydigt at konkludere, hvad dette skyldtes. Som følge af afprøvningens resultater har VIA og AU dog valgt ikke at bede socialrådgiverne om deres egne vurderinger af risikoscorer i Pilot #2.

- Mulighed for at lade Beslutningsstøtten følge sagen

Pilot #1 pegede på, at det var vanskeligt for socialrådgiverne at anvende Beslutningsstøtten i den indledningsvise vurdering af underretninger, da der i lovgivningen gælder en tidsfrist på 24 timer. I løbet af de første 24 timer skal socialrådgiveren have vurderet, hvorvidt barnets sundhed eller udvikling er i fare, samt om der derfor er behov for iværksættelse af akutte foranstaltninger

over for barnet eller den unge.⁹ For at fremme anvendeligheden af Beslutningsstøtten ønsker VIA og AU at lade Beslutningsstøtten følge sagen og gøre systemet tilgængeligt for socialrådgiverne fra modtagelse af underretning til beslutning om § 50-undersøgelse¹⁰ eller forebyggende § 11, stk. 3-forløb¹¹. Det vil således også være muligt for socialrådgiverne at anvende Beslutningsstøtten sammen med familierne. Det vil dog altid være op til socialrådgiveren at vurdere, i hvilket omfang Beslutningsstøtten skal inddrages i den socialfaglige vurdering og i dialogen med familien.

Herudover skal Beslutningsstøtten i Pilot #2 anvendes på *alle* underretningssager med undtagelse af følgende typer underretningssager:

1. Underretningssager vedrørende børn uden CPR-numre, herunder f.eks. ufødte
2. Underretningssager vedrørende børns handicap, herunder f.eks. ansøgning om støtte, hjælpemidler mv.

Afprøvningen og forudgående tests vil herudover i øvrigt foregå som i Pilot #1 – med undtagelse af, at Pilot #2 alene skal afprøves i én kommune.

⁹ § 155, stk. 2, i bekendtgørelse nr. 789 af 1. august 2019 af lov om social service.

¹⁰ § 50-undersøgelsen er en børnefaglig undersøgelse, der finder sted med hjemmel i servicelovens § 50, når det må antages, at et barn eller en ung trænger til særlig støtte, herunder på grund af nedsat fysisk eller psykisk funktionsevne.

¹¹ Det følger af servicelovens § 11, stk. 3, at kommunen kan tilbyde følgende forebyggende indsatser: 1) Konsulentbistand, herunder familierettede indsatser, 2) Netværks- eller samtalegrupper, 3) Rådgivning om familieplanlægning og 4) Andre indsatser, der har til formål at forebygge et barns eller en ungs eller familiens vanskeligheder.

2.2.3 *Evaluering*

2.2.3.1 *Pilot #1*

Efter afprøvningen i Pilot #1 blev al data fra Beslutningsstøtten, herunder personhenførbare data, trukket ud af systemet af Indiesoft og overført til AU og VIA. Dataudtrækkene indeholdt ikke de konkrete underretninger, men alene de data, systemet havde anvendt til at udregne risikoscoren – dvs. de valgte variable – samt socialrådgivernes vurderinger og deres begrundelser herfor. Formålet med udtrækkene og deres videregivelse var at evaluere pilotafprøvningen, herunder Beslutningsstøttens egnet- og anvendelighed. VIA og AU modtog begge det fulde dataudtræk i forbindelse med evalueringen, og Indiesoft videregav således dataene til VIA på vegne af AU. Det var dog alene AU, der modtog backuppen fra afprøvningen.

Evalueringen bestod derudover af de interviews og observationer, som VIA havde foretaget under afprøvningen. Det var alene VIA, der indsamlede og opbevarede interview- og observationsdata.

I forbindelse med evalueringen delte VIA og AU slutteligt aggregerede og pseudonymiserede data, f.eks. opsamlinger fra interviews eller grafer.

AU har i forbindelse med evalueringen valgt, at pilotafprøvningens data ikke skal lægges på Danmarks Statistiks forskningsserver, idet dataene er af meget begrænset omfang.

2.2.3.2 *Pilot #2*

Efter afprøvningen i Pilot #2 vil Indiesoft på samme måde som under Pilot #1 lave et komplet dataudtræk fra systemet. Dette dataudtræk vil dog alene blive overført til AU. Dataudtrækket vil indeholde de behandlede data, herunder personhenførbare data, samt systemets udregnede risikoscorer. Dataudtrækket fra Pilot #2 vil rumme flere typer af data, end hvad der var tilfældet i Pilot #1, idet Beslutningsstøtten i Pilot #2 vil inddrage væsentligt flere variable til beregning af risikoscoren.

Der vil under evalueringen af Pilot #2 også blive inddraget data fra surveys, interviews og observationer, der er foretaget under afprøvningen. Det vil her fortsat alene være VIA, der indsamler og opbevarer disse survey-, interview- og observationsdata.

I forbindelse med evalueringen af Pilot # 2 vil VIA og AU slutteligt dele aggregerede og pseudonymiserede data, f.eks. opsamlinger fra interviews eller grafer.

Det er endnu ikke afklaret, om dataene fra Pilot #2 skal lægges på Danmarks Statistiks forskningsserver efter endt afprøvning.

2.2.4 Særligt om aktiver af betydning for behandlingen

2.2.4.1 Det forudgående kvalitative studie

I forbindelse med det kvalitative forstudie, der fandt sted inden udviklingen af Beslutningsstøtten, blev de anonymiserede underretninger manuelt videregivet fra Hjørring Kommune og Silkeborg Kommune til VIA. Disse underretninger findes alene fysisk, og de er således aldrig blevet indscannet eller lignende. De fysiske underretninger opbevares i et aflåst skab, som alene den nuværende projektleder har adgang til.

Data fra VIAs initiale interviews blev optaget på en selvstændig lydoptager. Efter interviewene var afsluttet, blev optagelserne flyttet over på VIAs individuelle forskerdrev og slettet fra den pågældende lydoptager. Data kan således kun tilgås af den nuværende projektleder, da de ligger på vedkommendes individuelle forskerdrev.

2.2.4.2 Udvikling og afprøvning af Pilot #1

I forbindelse med udviklingen af Beslutningsstøtten har AU behandlet personoplysninger centralt på Danmarks Statistiks forskningsserver. Personoplysningerne har på intet tidspunkt forladt forskningsserveren, og efter endt udvikling er alene den matematiske formel (algoritmen) blevet taget ud.

Under afprøvningen i Pilot #1 har Beslutningsstøtten indhentet data fra henholdsvis Hjørring Kommunes datavarehus, KMD, samt Silkeborg Kommunes sagsbehandlingssystem, EG Sensum. I forbindelse

med afprøvningen blev dataene opbevaret på henholdsvis Hjørring Kommunes hardware i Hjørring og Indiesofts hardware hos Amazon i Irland.

I Silkeborg Kommune blev data hentet via en webservice fra EG's servere via en SSL-krypteret forbindelse. Beslutningsstøtten var installeret på Indiesofts servere. I Hjørring Kommune blev data hentet fra KMD's datavarehus via en databaseforbindelse over en VPN-tunnel. Beslutningsstøtten var installeret på Hjørring Kommunes egne servere.

Det stykke software, der indsamlede og behandlede data fra kommunernes sagsbehandlingssystem samt præsenterede socialrådgiverne for en risikoscore på baggrund af en algoritme udviklet i projektet, er udviklet af Indiesoft, der ligeledes har stået for at implementere Beslutningsstøtten i kommunerne. Til udvikling af softwaren er benyttet traditionelle Microsoft teknologier, herunder Windows operativsystem og IIS webserver. Herudover er der udviklet skræddersyet software på en Microsoft ASP.NET Core (C#) og Microsoft SQL Server platform.

Det er alene udvalgte socialrådgivere, der har haft adgang til Beslutningsstøtten i Pilot #1. De udvalgte socialrådgivere har ikke haft mulighed for at se hinandens brug af Beslutningsstøtten, og efter endt workflow har den enkelte sagsbehandler ikke haft mulighed for at genbesøge vurderingerne i systemet.

Data, der efter Pilot #1 blev overført til AU til brug for evaluering af afprøvningen, bliver opbevaret på et sikkert drev, hvor kun projektgruppens AU-ansatte medlemmer samt Indiesoft¹² har adgang. Drevet ligger på AU's servere, der hostes af Global Connect A/S i Skanderborg.¹³ AU's medarbejdere kan kopiere data til eget eller eksternt udstyr fra det sikre drev. Alle hændelser logges.

I VIA opbevares videregivet data på sikrede forskningsservere. Data opbevares enten på et fælles drev med adgang for flere eller på et drev udelukkende tilknyttet den enkelte medarbejder. Hvis flere skal have adgang, og data lægges på et fælles forskningsdrev, er det kun forskergruppens medlemmer fra VIA,

¹² Indiesofts adgang til det sikrede drev er alene åbent i forbindelse med afprøvningerne, hvor Indiesoft skal kunne levere udtræk af data. I mellemliggende perioder mellem afprøvninger lukkes Indiesofts adgang til drevet.

¹³ GlobalConnect A/S behandler ikke data på vegne af AU og varetager alene en housing rolle, hvor de leverer opbevaringslokale, køling og strøm til AU's egne servere.

der har adgang. VIAs medarbejdere vil i relation til projektet alene anvende forsknings-pc'ere. En forsknings-pc i VIA kan kun anvende et USB-drev, hvis USB-drevet er krypteret.

Data fra VIAs løbende observationer og interviews blev indsamlet in persona og optaget som lydfiler enten på separate lydoptagere eller som optagelse i et online møde. De er efterfølgende opbevaret på individuelle forskerdrev tilhørende de deltagende i forskergruppen fra VIA.

2.2.4.3 Udvikling og afprøvning af Pilot #2

De samme forhold, som er beskrevet ovenfor, forventes at gælde for Pilot #2.

2.3 Behandlingens omfang

2.3.1 Kategorier af registrerede

2.3.1.1 Pilot #1

VIA og AU har og vil i forbindelse med henholdsvis udviklingen, afprøvningen og evalueringen i relation til Pilot #1 behandle oplysninger om følgende kategorier af registrerede:

1. Børn og unge med modtaget underretning
2. Barnets eller den unges søskende (hvor disse er omtalt i underretningerne, som Silkeborg Kommune og Hjørring Kommune videregav til VIA i forbindelse med det kvalitative studie)
3. Barnets eller den unges forældre

VIA og AU har i forbindelse med afprøvningen og evalueringen desuden behandlet oplysninger om følgende kategori:

4. Medarbejdere, der anvender Beslutningsstøtten

Under den indledningsvise udvikling blev Beslutningsstøtten som beskrevet i punkt 2.2.1 trænet med registerdata fra Danmarks Statistik, herunder i alt 120.000 underretningssager fra perioden april 2014-2015, hvorefter Beslutningsstøtten under afprøvningen i Pilot #1 blev anvendt på cirka 200 børn og unge.

2.3.1.2 Pilot # 2

VIA og AU vil i forbindelse med henholdsvis udviklingen, afprøvningen og evalueringen i relation til Pilot #2 behandle oplysninger om følgende kategorier af registrerede:

1. Børn og unge med modtaget underretning
2. Barnets eller den unges søskende
3. Barnets eller den unges forældre

VIA og AU vil i forbindelse med afprøvningen og evalueringen desuden behandle oplysninger om følgende kategori:

4. Medarbejdere, der anvender Beslutningsstøtten

Beslutningsstøtten vil blive trænet med registerdata fra april 2014-2018, herunder i alt 506.718 underretningssager. VIA og AU forventer at anvende Beslutningsstøtten på cirka 100-200 husstande i Pilot #2.

2.3.2 *Typer af personoplysninger*

2.3.2.1 Pilot #1

I forbindelse med udviklingen, afprøvningen og evalueringen af Beslutningsstøtten er og vil der blive behandlet både almindelige personoplysninger, følsomme personoplysninger, oplysninger om strafbare forhold samt fortrolige oplysninger. Hvilke oplysninger, der behandles, varierer afhængig af kategorien af registrerede.¹⁴

Der behandles følgende oplysninger om "Børn og unge med modtaget underretning":

- Barnets CPR-nummer, køn*, alder og navn
- Tidspunkt for modtagelse af underretning

¹⁴ Oplysninger markeret med * er alene anvendt i udviklingen af algoritmen til Pilot #1 på Danmarks Statistiks forskerserver. Oplysningerne har ikke været anvendt til beregning af de konkrete risikoscorer i forbindelse med afprøvningen.

- Antal tidligere underretninger de seneste 90 og 180 dage
- Antal tidligere forebyggende foranstaltninger de seneste 180 dage
- Paragrafkode samt dato for forebyggende foranstaltninger
- Antal tidligere anbringelser det seneste år og de seneste fem år
- Dato for anbringelse samt handlekommune*
- Bopælskommune* og flytninger*
- Underretningsårsag(er):
 - o Misbrug hos barnet/den unge
 - o Kriminalitet hos barnet/den unge
 - o Seksuelt krænkende adfærd hos barnet/den unge
 - o Selvskadende adfærd hos barnet/den unge
 - o Selvmordstanker eller –forsøg
 - o Social isolation
 - o Skoleproblemer hos barnet/den unge, f.eks. fravær
 - o Anden bekymrende adfærd hos barnet/den unge, f.eks. udadreagerende adfærd
 - o Betydelig eller varigt nedsat fysisk eller psykisk funktionsevne hos barnet/den unge
 - o Sundhedsforhold hos barnet/den unge
 - o Overgreb mod barnet/den unge, f.eks. seksuelle eller voldelige overgreb
 - o Anden form for omsorgssvigt over for barnet/den unge
 - o Misbrug hos forældre
 - o Kriminalitet hos forældre
 - o Anden bekymrende adfærd hos forældre
 - o Betydelig eller varigt nedsat funktionsevne hos forældre
 - o Højt konfliktniveau eller vold i hjemmet mellem voksne
 - o Utilstrækkelig omsorg fra forældre
 - o Fogedsag, hjemløshed eller udsættelse fra bolig
 - o Andet
- Underretningstype*
- Den beregnede risikoscore

Der behandles følgende oplysninger om "Barnets eller den unges søskende":

- Oplysninger om barnets eller den unges søskende, der fremgår af underretningerne, som Silkeborg Kommune og Hjørring Kommune videregav til VIA i forbindelse med det kvalitative studie

Der behandles følgende oplysninger om "Barnets eller den unges forældre":

- Forældres PNR-nummer*, køn* og alder*
- Forældres bopælskommune* og flytninger*
- Forældres baggrund (indvander, efterkommer eller dansk oprindelse)*
- Oplysninger om barnets eller den unges forældre, der fremgår af underretningsårsagen mv.

Der behandles følgende oplysninger om "Medarbejdere, der anvender Beslutningsstøtten":

- Medarbejderens e-mailadresse til brug for logning og auditering
- Medarbejderens initial og afsluttende risikoscore samt medarbejderens begrundelser herfor
- Interview- og observationsdata

2.3.2.2 Pilot #2

I forbindelse med udviklingen, afprøvningen og evalueringen af Beslutningsstøtten i Pilot #2 vil der blive behandlet både almindelige personoplysninger, følsomme personoplysninger, oplysninger om strafbare forhold samt fortrolige oplysninger.

Hvilke oplysninger, der behandles, varierer afhængig af kategorien af registrerede.

Der behandles følgende personoplysninger om "Børn og unge med modtaget underretning":

- Barnets CPR-nummer og alder
- Tidspunkt for modtagelse af underretning
- Antal tidligere underretninger de seneste to år
- Antal tidligere alvorlige underretninger de seneste to år
- Antal tidligere forebyggende foranstaltninger de seneste to år
- Antal tidligere alvorlige forebyggende foranstaltninger de seneste to år
- Paragrafkode samt dato for forebyggende foranstaltninger

- Antal tidligere anbringelser de seneste to år
- Antal tidligere tvangsanbringelser de seneste to år
- Dato for anbringelse samt handlekommune
- Underretningsårsag(er):
 - o Misbrug hos barnet/den unge
 - o Kriminalitet hos barnet/den unge
 - o Skoleproblemer hos barnet/den unge, f.eks. fravær
 - o Anden bekymrende adfærd hos barnet/den unge, f.eks. udadreagerende adfærd
 - o Betydelig eller varigt nedsat fysisk eller psykisk funktionsevne hos barnet/den unge
 - o Sundhedsforhold hos barnet/den unge
 - o Overgreb mod barnet/den unge, f.eks. seksuelle eller voldelige overgreb
 - o Anden form for omsorgssvigt over for barnet/den unge
 - o Misbrug hos forældre
 - o Kriminalitet hos forældre
 - o Anden bekymrende adfærd hos forældre
 - o Betydelig eller varigt nedsat funktionsevne hos forældre
 - o Højt konfliktniveau eller vold i hjemmet mellem voksne
 - o Utilstrækkelig omsorg fra forældre
 - o Fogedsag, hjemløshed eller udsættelse fra bolig
 - o Andet
- Underretningstype
- Den beregnede risikoscore
- Oplysninger om underretningens alvorlighed
- Antal flytninger
- Hvorvidt barnet er nytilflytter
- Typen af husstand, som barnet bor i (bor med begge forældre, bor med mor, der er i et nyt parforhold, bor med enlig mor, bor med far, der er i et nyt parforhold, bor med enlig far eller bor ikke hos forældrene)
- Hvorvidt ingen mor/far er registreret

Der behandles følgende oplysninger om "Barnets eller den unges søskende":

- Oplysninger om barnets eller den unges søskende, der fremgår af underretningsårsagen mv.
- CPR-nummer for barnets søskende (juridiske søskende)
- Antal tidligere underretninger for barnets søskende (juridiske søskende)
- Antal tidligere forebyggende foranstaltninger for barnets søskende (juridiske søskende)
- Antal tidligere anbringelser for barnets søskende (juridiske søskende)

Der behandles følgende oplysninger om "Barnets eller den unges forældre":

- CPR-nummer og alder
- Hvorvidt forælderen er barnets mor eller far (juridiske forældre)
- Civilstatus
- Antal børn
- Antal partnere, med hvem forælderen har fået børn
- Alder ved barnets fødsel
- Hvorvidt forælderen er gået bort
- Oplysninger om barnets eller den unges forældre, der fremgår af underretningsårsagen

Der behandles følgende oplysninger om "Medarbejdere, der anvender Beslutningsstøtten":

- Medarbejderens e-mailadresse til brug for logning og auditering
- Interview-, observations- og surveydata

AU arbejder i forbindelse med Pilot #2 med at validere algoritmens performance på baggrund af pseudo-nomiseret registerdata fra Danmarks Statistik, Styrelsen for IT og Læring og Sundhedsdatastyrelsen. Disse registerdata er ikke anvendt til at udvikle modellen, men alene at evaluere dens performance. Der behandles alene personoplysninger om "Børn og unge med modtaget underretning" i forbindelse med arbejdet med at validere modellerne. Følgende oplysninger er anvendt:

- Hvorvidt barnet/den unge har været offer for en straffelovsforbrydelse samt dato for hændelsen og type
- Hvorvidt barnet/den unge har været sigtet for en straffelovsforbrydelse samt dato for hændelsen og sigtelsen
- Somatiske eller psykiske diagnoser

- Dato hvor barnet/den unge har været henvist til psykologisk behandling eller indlæggelse
- Svar fra den nationale trivselsmåling i folkeskolen
- Ulovligt skolefravær
- Tandsundhed

2.4 Sammenhæng og kontekst for behandlingen af personoplysninger i projektet

I løbet af forskningsprojektet bliver der primært behandlet oplysninger om børn og unge, der kan være i risiko for at mistrives samt disses forældre og søskende. VIA og AU er bevidste om, at de behandler oplysninger om en række sårbare registrerede, hvorfor VIA og AU udviser særlig agtpågivenhed ved behandlingen af personoplysningerne. I tråd hermed er Beslutningsstøtten udformet efter principperne bag *"Databeskyttelse gennem design"* og *"Databeskyttelse gennem standardindstillinger"*. Både AU og VIA har desuden mange års erfaring med behandling af personoplysninger i forbindelse med forskningsprojekter, herunder med behandling af følsomme personoplysninger om sårbare registrerede.

VIA og AU har ikke en direkte relation til de registrerede, men stiller alene Beslutningsstøtten til rådighed for kommunerne. I forbindelse med træningen af modellen har AU alene anvendt registerdata og har således heller ikke her haft en relation til de registrerede.

Kommunerne er selvstændige dataansvarlige for så vidt angår de oplysninger, som VIA og AU får videregivet fra kommunerne, og den registreredes sædvanlige rettigheder gælder således primært i forhold til kommunen. De registrerede har på baggrund heraf som udgangspunkt ikke mulighed for direkte at påvirke VIAs og AU's behandling af personoplysninger i projektet, dog vil f.eks. en berettiget anmodning om berigtigelse af personoplysninger få indvirkning på VIAs og AU's behandling.

Det forhold, at det er kommunernes socialrådgivere, der skal anvende Beslutningsstøtten, kan indebære en risiko for, at Beslutningsstøtten anvendes i strid med sit formål. Beslutningsstøtten er som redegjort for i punkt 2.2.2 alene et beslutningsstøtteredskab og må ikke anvendes til de facto at træffe fuldautomatiske afgørelser. VIA og AU har og vil fortsat imødegå dette gennem uddannelse af kommunerne og deres socialrådgivere.

2.5 Modtagere af personoplysninger

2.5.1 Pilot #1

AU overlod i forbindelse med Pilot #1 personoplysninger i projektet til følgende databehandlere:

- o Indiesoft.

Personoplysningerne blev overladt med henblik på at kunne 1) udvikle softwaren og implementere denne hos kommunerne, 2) drifte softwaren, herunder yde support for medarbejderne, der anvender systemet, samt bidrage til kvalitetssikring af redskabet, 3) logge informationer om, hvordan systemet virker, hvordan det anvendes, herunder af hvem og hvornår og 4) overføre data fra afprøvningen til AU.

AU videregav desuden et dataudtræk fra Beslutningsstøttens afprøvning i Pilot #1 til VIA.

VIA og AU videregav ikke projektets personoplysninger til andre dataansvarlige i løbet af Pilot #1.

2.5.2 Pilot #2

VIA og AU forventer fortsat at anvende Indiesoft som databehandler i Pilot #2. AU vil dog ikke videregive et dataudtræk fra Beslutningsstøtten til VIA. Derudover har AU overladt oplysninger fra Styrelsen for IT og Læring samt Sundhedsdatastyrelsen til Danmarks Statistik. Det er endnu ikke afklaret, hvorvidt AU vil overlade data fra afprøvningen i Pilot #2 til Danmarks Statistik.

VIA og AU forventer i øvrigt at publicere de endelige forskningsresultater i et anerkendt videnskabeligt tidsskrift el.lign. Data anvendt i publikationer vil være aggregerede, så enkelte individer ikke kan identificeres. VIA og AU vil, hvis anonymisering mod forventning ikke er mulig, indhente en tilladelse fra Datatilsynet før publiceringen, jf. databeskyttelseslovens § 10, stk. 3, nr. 3.

2.6 Opbevaringsperiode for personoplysningerne i projektet

2.6.1 Pilot #1

VIA opbevarer interviewdata fra det kvalitative forstudie i 5 år efter indgåelse af aftale om publicering, idet dette er foreskrevet hos flere videnskabelige tidsskrifter med henblik på eventuel verificering. Det samme gør sig gældende for de anonymiserede underretninger fra forstudiet.

AU har foreløbig adgang til Danmarks Statistiks forskerserver til og med den 31. december 2025. AU vil dog søge om at få denne periode forlænget, så AU har adgang til data indtil 5 år efter indgåelse af aftale om publicering. Træningsdata, der ikke længere er relevante for projektets formål, slettes fra AU's projektmappe hos Danmarks Statistik.

Indiesoft opbevarede personoplysninger på sin Microsoft SQL Serverdatabase fra det tidspunkt, de blev registreret i systemet, og indtil det tidspunkt, hvor dataene blev overført til VIA og AU, og pilotprojektet blev lukket ned. Indiesoft foretog løbende backup af serverdatabasen, der blev slettet efter 70 dage. Backuppen blev inden da overført til AU. Lokal opbevaring ("caching") under anvendelse af systemet var slået fra, og det var således ikke muligt at eksportere data til lokale medier, harddiske eller lign. Data blev alene opbevaret i databasen.

Hjørring Kommune opbevarede dataene på deres eget hardware i løbet af afprøvningen.

Efter afprøvningen blev data overført til VIA og AU til brug for evaluering. Data slettes senest 31. marts 2021.

Der foretages løbende backup af data på henholdsvis AU's og VIA's servere. Hos AU bliver backuppen automatisk slettet efter 3-6 måneder – afhængig af logningstypen. Hos VIA slettes backups af servere senest efter 33 dage.

Data, som undtagelsesvist opbevares på eksterne enheder af VIA eller AU (f.eks. krypterede USB-stik eller SD-kort) slettes, så snart formålet med at flytte data til den eksterne enhed er opfyldt, f.eks. videregivelse af aggregerede data mellem AU og VIA.

2.6.2 Pilot #2

VIA og AU forventer, at ovenstående fortsat vil gøre sig gældende for Pilot #2. VIA og AU forventer at gemme dataene fra Pilot #2 i 5 år efter indgåelse af aftale om publicering, idet dette som nævnt er foreskrevet hos flere videnskabelige tidsskrifter. Efter 5 år anonymiseres eller slettes dataene uigenkaldeligt. Hos AU har den interne projektleder ansvaret for rettidig sletning af data. Hos VIA sikres rettidig sletning ved, at VIAs forskere modtager en påmindelse om at slette eller anonymisere data, så snart forskningsprojektet er afsluttet.

3. TRIN 2: INDDRAGELSE AF RELEVANTE INTERESSETER

3.1 Inddragelse af VIAs og AU's databeskyttelsesrådgiver (DPO)

Det følger af databeskyttelsesforordningen, at den dataansvarlige skal rådføre sig med sin databeskyttelsesrådgiver i forbindelse med udarbejdelsen af konsekvensanalyser.

3.1.1 VIAs databeskyttelsesrådgiver

Er VIAs databeskyttelsesrådgiver blevet hørt og inddraget i forbindelse med udarbejdelsen af denne konsekvensanalyse?

☒ Ja

☐ Nej

Databeskyttelsesrådgiveren har følgende bemærkninger:

VIAs DPO oplyser, at han er enig i de anførte opmærksomhedspunkter og konklusioner i konsekvensanalysen, og at han i øvrigt vil sikre udarbejdelse og implementering af de nødvendige foranstaltninger.

3.1.2 AU's databeskyttelsesrådgiver

Er AU's databeskyttelsesrådgiver blevet hørt og inddraget i forbindelse med udarbejdelsen af denne konsekvensanalyse?

☒ Ja

☐ Nej

Databeskyttelsesrådgiveren har følgende bemærkninger:

"På baggrund af de mitigerende foranstaltninger, der er anført i implementeringsplanen, og forudsat, at disse implementeres som anført, er det min opfattelse, at behandlingen sker i overensstemmelse med databeskyttelsesforordningen og supplerende bestemmelser i databeskyttelsesloven.

Det bemærkes, at det er den dataansvarlige, der skal forestå implementering af de mitigerende foranstaltninger og løbende foretage ajourføring af konsekvensanalysen. Den dataansvarlige skal løbende rapportere om gennemførelse af de mitigerende handlinger. Den dataansvarlige er desuden forpligtet til at forelægge enhver ændring af konsekvensanalysen for databeskyttelsesrådgiveren, hvor denne ændring medfører en øget risiko for den registreredes rettigheder og frihedsrettigheder.

Databeskyttelsesenheden ved Aarhus Universitet vil føre kontrol med implementeringen."

3.2 Indhentning af de registreredes eller deres repræsentanternes synspunkter

Det følger af databeskyttelsesforordningens artikel 35, stk. 9, at det i nogle tilfælde er relevant at indhente de registreredes eller deres repræsentanternes synspunkter. Dette afhænger af en konkret vurdering, som skal foretages og dokumenteres under dette afsnit.

Er de registreredes eller deres repræsentanternes synspunkter indhentet og inddraget i forbindelse med udarbejdelsen af denne konsekvensanalyse?

☒ Ja

☐ Nej

VIA og AU oprettede i 2017 en socialfaglig følgegruppe med henblik på at inddrage relevante synspunkter fra relevante interessenter, herunder repræsentanter for de registrerede.¹⁵ VIA og AU inviterer løbende interessenter i takt med, at VIA og AU får kendskab til disse. Følgegruppen mødes som udgangspunkt én gang årligt. Desuden vil de registrerede blive hørt i Pilot #2 i forbindelse med det tilknyttede ph.d.-projekt. Det tilknyttede ph.d.-projekt har fokus på inddragelse af børn/unge og forældre i forbindelse med underretninger. Her inddrages de registrerede og får mulighed for at udtrykke deres oplevelse af brugen af Beslutningsstøtten. Det er for nuværende planen, at de inddrages via interviews. Viden fra børn/unge og deres forældre er, kombineret med viden fra socialrådgiverne, vægtige elementer i, hvorvidt Beslutningsstøtten anbefales efter endt projekt.

VIA og AU har løbende overvejet, hvordan VIA og AU herudover kan sikre, at projektets behandling af personoplysninger finder sted under hensyntagen til de registrerede og disses evt. repræsentanternes synspunkter. Oplysninger om projektet – herunder delprojekterne ”*Beslutningsstøtte ved underretninger*” og ”*Underretninger i fokus*” – har løbende været offentliggjort på VIAs og AU’s hjemmesider samt gennem annonceringer på sociale medier, herunder bl.a. LinkedIn. VIA og AU har i forbindelse med den løbende offentliggørelse genereret en offentlig debat, hvorved der er fremkommet en række relevante synspunkter, herunder betragtninger omkring datakvalitet og risikoen for de facto fuldautomatiske afgørelser mv. VIA og AU har taget disse synspunkter til efterretning og inddraget synspunkterne ved planlægningen af den videre behandling.

¹⁵ Pr. dags dato deltager følgende interessenter: Kommunernes Landsforening, Socialstyrelsen, Dansk Socialrådgiverforening, Børns Vilkår, ForældreLANDsforeningen, Foreningen af Socialchefer i Danmark. Derudover deltager en repræsentant fra Hjørring Kommune og Silkeborg Kommune.

4. TRIN 3: PROJEKTETS LOVLIGHED, NØDVENDIGHED OG PROPOR- TIONALITET

4.1 Principperne om lovlighed, rimelighed og gennemsigtighed

VIA og AU behandler personoplysningerne i projektet lovligt, rimeligt og på en gennemsigtig måde i forhold til den registrerede, jf. databeskyttelsesforordningens artikel 5, stk. 1, litra a.

4.1.1 *Lovlighed*

4.1.1.1 *VIA's og AU's behandlingshjemmel*

Behandlingen af projektets almindelige ikke-følsomme personoplysninger er lovlig, idet behandlingen er nødvendig for udførelsen af en opgave i samfundets interesse, jf. databeskyttelsesforordningens artikel 6, stk. 1, litra e, smh. databeskyttelseslovens § 6, stk. 1. Idet projektets behandling af personoplysninger sker i statistisk og videnskabeligt øjemed med henblik på at imødegå mistrivsel hos børn og unge, vurderes dette retsgrundlag at dække VIA's og AU's behandling af almindelige personoplysninger.

Det følger af databeskyttelseslovens § 10, stk. 1, at oplysninger, som nævnt i databeskyttelsesforordningens artikel 9, stk. 1 (følsomme personoplysninger), og artikel 10 (oplysninger om strafbare forhold), må blive behandlet, hvis dette alene sker med henblik på at udføre statistiske eller videnskabelige undersøgelser af væsentlig samfundsmæssig betydning, og hvis behandlingen er nødvendig af hensyn til udførelsen af undersøgelserne.

Formålet med behandlingen af personoplysningerne er, som der er redegjort for i denne konsekvensanalyse punkt 2.1, at undersøge, hvorvidt det er muligt at udvikle et beslutningsstøtteredskab, der kan skærpe beslutningsgrundlaget i kommunernes vurdering af underretninger om børn og unges mistrivsel. Projektet vurderes således at være af væsentlig samfundsmæssig betydning.

VIA og AU har desuden vurderet, at de involverede oplysninger er nødvendige af hensyn til udførelsen af undersøgelserne. Hvis Beslutningsstøtten skal være effektiv og saglig, skal den udvikles, herunder trækkes, med de samme data, som er tilgængelige for socialrådgiverne i kommunerne, herunder med en mængde, der er tilstrækkelig til at danne en repræsentativ og saglig algoritme. VIA og AU har til brug

for træningen af modellen samt det kvalitative forstudie alene anvendt pseudonymiseret eller effektivt anonymiseret data. Validering af Beslutningsstøtten kræver desuden, at den afprøves på socialrådgivernes faktiske sagsbehandling, herunder aktuelle underretningssager og de heri inkluderede personoplysninger, hvorfor denne del ligeledes vurderes nødvendig.

VIA og AU behandler de registreredes CPR-numre i løbet af projektet i overensstemmelse med databeskyttelseslovens § 11, stk. 1, hvorefter offentlige myndigheder må behandle oplysninger om personnummer med henblik på en entydig identifikation eller som journalnummer.

På baggrund af ovenstående har VIA og AU vurderet, at den nævnte behandling af personoplysninger i projektet kan finde sted med hjemmel i databeskyttelsesforordningens artikel 6, stk. 1, litra e, databeskyttelseslovens § 10, stk. 1, da behandlingen alene sker med henblik på at udføre statistiske eller videnskabelige undersøgelser af væsentlig samfundsmæssig betydning samt med hjemmel i databeskyttelseslovens § 11, stk. 1.

For så vidt angår den behandling, der fandt sted før databeskyttelsesforordningens og databeskyttelseslovens ikrafttræden, behandlede VIA og AU oplysningerne i projektet med hjemmel i de tilsvarende bestemmelser i persondataloven, dvs. med hjemmel i den dagældende persondatalovs §§ 6, stk. 1, nr. 5, 10, stk. 1 samt 11, stk. 1.

4.1.1.2 Særligt om videregivelse af personoplysninger fra Silkeborg Kommune og Hjørring Kommune

AU og VIA har forud for videregivelsen fra Silkeborg Kommune og Hjørring Kommune indgået videregivelsesaftaler med begge kommuner. Disse videregivelsesaftaler blev indgået i henholdsvis december 2018 og oktober 2018. AU vil med samarbejdskommunen i Pilot #2 indgå en videregivelsesaftale analogt¹⁶ med videregivelsesbekendtgørelsen.¹⁷

¹⁶ Idet kommunerne ikke behandler de aktuelle personoplysninger med hjemmel i databeskyttelseslovens § 10, stk. 1 og 2, finder videregivelsesbekendtgørelsen ikke direkte anvendelse. AU vil dog indgå en aftale med den relevante kommune efter principperne i bekendtgørelsen.

¹⁷ Bekendtgørelse 2019-12-18 nr. 1509 om videregivelse af personoplysninger omfattet af databeskyttelseslovens § 10, stk. 1 og 2.

4.1.1.3 Særligt om videregivelse af personoplysninger mellem AU og VIA

VIA og AU indgik forud for deling af data ikke en videregivelsesaftale eller anden dokumentation for, at videregivelserne levede op til de daværende krav¹⁸, hvilket VIA og AU rettelig burde have gjort. VIA og AU har rettet op på dette ved at indgå en erklæring, der dokumenterer, at videregivelserne er sket i overensstemmelse med den nugældende videregivelsesbekendtgørelse. VIA og AU vil desuden sikre, at de i pilot #2 indgår en videregivelsesaftale i overensstemmelse med videregivelsesbekendtgørelsen, inden de deler data mellem dem.

4.1.1.4 Særligt om videregivelse af personoplysninger fra Styrelsen for IT og Sundhedsdatastyrelsen

AU har forud for videregivelsen fra Styrelsen for IT og Sundhedsdatastyrelsen i Pilot #2 indgået videregivelsesaftaler med begge styrelser. Aftalerne blev indgået i marts og december 2019 samt i januar 2020 i overensstemmelse med videregivelsesbekendtgørelsen.

4.1.1.5 Særligt om viderebehandling til andre formål

Det fremgår af databeskyttelseslovens § 10, stk. 2, at de oplysninger, der er omfattet af stk. 1, ikke senere må behandles i andet end videnskabeligt eller statistisk øjemed. Det samme gælder behandling af andre

¹⁸ Da databeskyttelsesforordningen trådte i kraft den 25. maj 2018, blev den dagældende persondatalovs § 10, stk. 3, erstattet af databeskyttelseslovens § 10. Databeskyttelseslovens § 10 rummer ikke et krav om generel forudgående tilladelse. Dog fastsatte Datatilsynet et krav om indgåelse af en forudgående videregivelsesaftale. Herefter skulle den afgivende dataansvarlige modtage en skriftlig begrundet erklæring fra den modtagende dataansvarlige, der som minimum beskrev følgende forhold:

1. At oplysningerne konkret var nødvendige for modtagerens undersøgelse.
2. At oplysningerne alene ville blive brugt i statistisk eller videnskabeligt øjemed.
3. At oplysningerne ved undersøgelsens afslutning ville blive slettet, anonymiseret eller tilintetgjort, så det efterfølgende ikke er muligt at identificere enkeltpersoner, eller alternativt vil blive overført til opbevaring i arkiv efter reglerne i arkivlovgivningen.

Disse indholdsmæssige minimumskrav fremgik alene af Datatilsynets hjemmeside, men de kunne både anses som en udmøntning af bemyndigelsesbestemmelsen i databeskyttelseslovens § 10, stk. 4, og databeskyttelsesforordningens princip om ansvarlighed.

Den 1. januar 2020 trådte den nugældende videregivelsesbekendtgørelse i kraft. Bekendtgørelsen tilkendegiver i §§ 8-9 de indholdsmæssige krav til videregivelseserklæringer. Det væsentligste indhold af disse krav stemmer overens med retstilstanden før den 1. januar 2020. Dog følger det af videregivelsesbekendtgørelsen, at erklæringer fremadrettet også skal dokumentere, at den modtagende dataansvarlige overholder bekendtgørelsens §§ 10-14 (den modtagende dataansvarliges forpligtelser).

oplysninger, som alene foretages i statistisk eller videnskabeligt øjemed efter databeskyttelsesforordningens artikel 6. VIA og AU vil ikke senere behandle projektets personoplysninger til andre formål.

VIA og AU vil desuden ikke videregive oplysningerne i personhenførbare form til andre dataansvarlige. Databeskyttelseslovens § 10, stk. 3, om tilladelse til videregivelse til tredjemænd i visse situationer samt reglerne om videregivelse i stk. 4, som nærmere udmøntet i videregivelsesbekendtgørelsen, er således ikke relevante.

4.1.2 Rimelighed og gennemsigtighed

VIA og AU har i løbet af projektet gennemført en række foranstaltninger for at sikre, at projektets datagrundlag og Beslutningsstøtten ikke er præget af urimelige skævheder, der kan føre til ulovlig forskelsbehandling.

Forud for Pilot #1 gennemførte VIA et kvalitativt studie af en række underretninger modtaget i Silkeborg Kommune og Hjørring Kommune med henblik på at konkretisere, hvilken beslutningsproces, Beslutningsstøtten skulle bistå med at skærpe. Dette kvalitative studie skulle sikre, at modellens udformning og sigte stemte overens med socialrådgivernes praktiske behov.

Forud for afprøvningen og implementeringen af Beslutningsstøtten i Pilot #1 indledte VIA og AU desuden et samarbejde med kommunernes datavarehuse med henblik på at sikre, at de udvalgte variable var defineret på samme måde som de data, AU har haft tilgængelige via Danmarks Statistiks registre. AU gjorde dette for at sikre modellens og datas egnethed.

AU har forud for Pilot #2 som andet led i valideringen af modellernes egnethed gennemført en række analyser med henblik på at sikre, at modellerne er følsomme over for andre saglige mål for mistrivsel. AU har ved hjælp af tests i udviklingsfaserne desuden verificeret, at modellerne ikke giver anledning til prædiktioner baseret på usaglige parametre, som f.eks. etnicitet.

VIA og AU har derudover forud for Pilot #2 gennemført en række foranstaltninger for at sikre, at de anvendte modeller understøtter et tilstrækkeligt niveau af transparens. I Pilot #2 planlægger AU at anvende beslutningstræer, der er mere præcise og komplekse matematiske modeller end den lineære model.

Til trods for deres øgede kompleksitet understøtter disse modeller fortsat et tilstrækkeligt niveau af systemtransparens, idet det bl.a. er muligt at få indsigt i modellernes vægtning af de anvendte variabler.

Som en yderligere foranstaltning med henblik på transparens har VIA og AU, som redegjort for i punkt 2.2.2, besluttet at øge informationsniveauet til socialrådgiverne under Pilot #2 for at skabe transparens omkring, hvilke oplysninger, Beslutningsstøtten har inddraget ved udregningen af risikoscoren, og hvordan disse oplysninger er blevet vægtet. Denne transparens vil gøre socialrådgiveren i stand til at vurdere, om der er fejl i data, eller om der er nye væsentlige oplysninger i sagen, der ikke er registreret. VIA og AU har desuden et væsentligt fokus på at uddanne kommunerne i korrekt brug af Beslutningsstøtten, så denne ikke anvendes til de facto at træffe automatiske afgørelser.

4.2 Princippet om formålsbegrænsning

De personoplysninger, der indsamles via Beslutningsstøtten eller til brug for udvikling, herunder træning af Beslutningsstøtten, indsamles alene med det overordnede formål at skabe et redskab, der kan skærpe beslutningsgrundlaget og understøtte socialrådgiverens vurdering af underretninger om børns og unges mistrivsel, jf. punkt 2.1. VIA og AU indsamler således oplysningerne i projektet til et på forhånd fastsat, afgrænset samt legitimt formål og vil ikke senere viderebehandle oplysninger på en måde, der er uforenelig med dette formål, jf. databeskyttelsesforordningens artikel 5, stk. 1, litra b.

Videregivelserne fra samarbejdskommunerne, Styrelsen for IT og Læring og Sundhedsdatastyrelsen anses desuden for at være forenelig med disse parterers oprindelige indsamlingsformål, idet både VIAs og AU's behandling alene sker til et videnskabeligt formål i overensstemmelse med artikel 89, stk. 1, hvilket ifølge databeskyttelsesforordningens artikel 5, stk. 1, litra b, ikke anses for at være uforeneligt med de oprindelige formål.

VIA og AU har løbende offentliggjort oplysninger om projektet på blandt andet AU's hjemmeside og på UC viden, og de registrerede er herigennem informeret om behandlingens formål.

4.3 Princippet om dataminimering

Personoplysningerne, som VIA og AU behandler i projektet, er tilstrækkelige, relevante og begrænset til, hvad der er nødvendigt i forhold til det formål, hvortil de behandles, jf. databeskyttelsesforordningens artikel 5, stk. 1, litra c.

VIA og AU har vurderet, at de involverede oplysninger er nødvendige af hensyn til udførelsen af undersøgelserne. Som redegjort for i punkt 4.1.1.1, skal Beslutningsstøtten for at være effektiv og saglig udvikles, herunder trænes, med de samme data, som er tilgængelige for socialrådgiverne i kommunerne og med en mængde, der er tilstrækkelig til at danne en repræsentativ og saglig algoritme. VIA og AU har til brug for træningen af modellen samt det kvalitative forstudie alene anvendt pseudonymiseret eller effektivt anonymiseret data. Validering af Beslutningsstøtten kræver desuden, at den afprøves på socialrådgivernes faktiske sagsbehandling, herunder den type af data, som Beslutningsstøtten er trænet med.

4.4 Princippet om rigtighed

Det følger af databeskyttelsesforordningens artikel 5, stk. 1, litra d, at personoplysninger skal være korrekte og om nødvendigt ajourførte; der tages ethvert rimeligt skridt for at sikre, at personoplysninger, der er urigtige i forhold til de formål, hvortil de behandles, straks slettes eller berigtiges.

For at sikre overholdelsen af dette princip om rigtighed har VIA og AU alene anvendt personoplysninger indsamlet fra offentligt anerkendte kilderegistre til udviklingen af Beslutningsstøtten. Desuden er der ved afprøvningen alene indsamlet oplysninger fra kommunernes sagsbehandlingssystemer, og således alene oplysninger, som almindeligvis indgår i socialrådgiveres sagsbehandling, herunder den socialfaglige vurdering. Personoplysningerne må således som udgangspunkt anses for at være af en sådan kvalitet og rigtighed, at dataene vil kunne anvendes af VIA og AU i projektet umiddelbart uden yderligere kvalitetskontrol til brug for udvikling og afprøvning af Beslutningsstøtten.

I relation til Pilot #2 vil den øgede transparens i forhold til, hvilke oplysninger, der er inddraget til udregning af risikoscoren, og hvordan disse oplysninger er blevet vægtet, desuden bistå ved socialrådgivernes vurdering af, om der er fejl i oplysningerne.

4.5 Princippet om opbevaringsbegrænsning

Det følger af databeskyttelsesforordningens artikel 5, stk. 1, litra e, at personoplysninger skal opbevares på en sådan måde, at det ikke er muligt at identificere de registrerede i et længere tidsrum end det, der er nødvendigt til de formål, hvortil de pågældende personoplysninger behandles. Derudover fremgår det, at personoplysninger kan opbevares i længere tidsrum, hvis personoplysningerne alene behandles til arkivformål i samfundets interesse, til videnskabelige eller historiske forskningsformål eller til statistiske formål i overensstemmelse med artikel 89, stk. 1, under forudsætning af, at der implementeres passende tekniske og organisatoriske foranstaltninger, som denne forordning kræver for at sikre den registreredes rettigheder og frihedsrettigheder.

VIA og AU vurderer, at projektets personoplysninger bliver opbevaret på en sådan måde, at det ikke er muligt at identificere de registrerede i et længere tidsrum end det, der er nødvendigt til at tilgodese behandlingsformålet. Der er tidligere nærmere redegjort for projektets opbevaringsperioder samt begrundelsen herfor under konsekvensanalysens punkt 2.6, hvortil der henvises.

4.6 Princippet om integritet og fortrolighed (sikkerhed og robusthed)

VIA og AU vurderer, at projektets personoplysninger bliver behandlet på en måde, der sikrer tilstrækkelig sikkerhed for de pågældende personoplysninger, herunder beskyttelse mod uautoriseret eller ulovlig behandling og mod hændeligt tab, tilintetgørelse eller beskadigelse, under anvendelse af passende tekniske eller organisatoriske foranstaltninger – i overensstemmelse med databeskyttelsesforordningens artikel 5, stk. 1, litra f.

VIA og AU har i forbindelse med tilrettelæggelsen af projektet foretaget en risikovurdering vedrørende behandlingssikkerhed i overensstemmelse med kravene hertil i databeskyttelsesforordningens artikel 32. Med udgangspunkt heri har VIA og AU fastsat passende sikkerhedsforanstaltninger for VIA og AU, der løbende er blevet overholdt. VIA og AU har og vil desuden fortsat stille de nødvendige krav til fremtidige databehandlere om behandlingssikkerhed samt sikre sig ret til at føre det nødvendige tilsyn med disse og deres underdatabehandlere. Der henvises herved til risikovurderingen, der er vedlagt denne konsekvensanalyse som Bilag B.

4.7 De registreredes rettigheder

4.7.1 Registreredes ret til underretning (oplysningspligten)

VIA og AU behandler som led i projektet alene personoplysninger, som VIA og AU har indhentet fra andre kilder end de registrerede selv.

Det følger af databeskyttelsesforordningens artikel 14, at de registrerede som udgangspunkt har ret til en række oplysninger om behandlingen af deres personoplysninger, uanset at oplysningerne er indsamlet fra andre kilder end den registrerede selv. Oplysningspligten finder dog ikke anvendelse, hvis det vil kræve en uforholdsmæssig stor indsats af den dataansvarlige at give oplysningerne, jf. databeskyttelsesforordningens artikel 14, stk. 5, litra b. Der henses her blandt andet til antallet af registrerede, oplysningernes alder samt eventuelle nødvendige garantier, der er stillet. Undtagelsen vil navnlig være relevant i tilfælde, hvor behandlingen sker til videnskabelige eller statistiske formål.

AU har til brug for udviklingen af Beslutningsstøtten indsamlet pseudonymiseret registerdata fra april 2014-2018, herunder i alt 506.718 underretningssager. Idet dataene er effektivt pseudonymiserede, kender AU ikke identiteten på de registrerede, hvorfor en underretning af disse ikke umiddelbart kan finde sted. Det følger af databeskyttelsesforordningens artikel 11, stk. 1, at den dataansvarlige ikke er forpligtet til at beholde, indhente eller behandle yderligere oplysninger for at kunne identificere den registrerede. Af denne grund er det AU's vurdering, at oplysningspligten i dette tilfælde kan undtages. AU gør i denne forbindelse opmærksom på, at projektet løbende er blevet offentliggjort på blandt andet VIAs og AU's respektive hjemmesider, hvorved de registrerede indirekte har haft mulighed for at blive informeret om behandlingen.

Det vurderes dog, at VIA og AU har oplysningspligt i forhold til de oplysninger, de har fået videregivet i forbindelse med afprøvningen og evalueringen i Pilot #1. Dette skyldes blandt andet, at de registrerede i dette tilfælde kan identificeres af VIA og AU samt det forhold, at der tale om et relativt lavt antal registrerede (cirka 200 børn og unge). VIA og AU har endnu ikke iagttaget denne oplysningspligt i forhold til Pilot #1, og oplysning af de registrerede er således alene givet via den løbende offentliggørelse på hjemmesider mv. VIA og AU vil i samarbejde med samarbejdskommunerne imidlertid sikre, at denne oplysningspligt bliver iagttaget ift. Pilot #1.

For så vidt angår den forestående afprøvning og evaluering i Pilot #2, vil der ligeledes alene indgå oplysninger om cirka 100-200 husstande, som VIA og AU kan identificere uden at indhente yderligere oplysninger. VIA og AU er således også her forpligtet til at oplyse de registrerede om behandlingen.

Oplysningspligten vil i Pilot #2 i praksis mest hensigtsmæssigt løses ved, at kommunerne underretter de registrerede på vegne af VIA og AU i forbindelse med, at kommunerne informerer om deres egen videregivelse til VIA og AU.

4.7.2 Øvrige rettigheder

Databeskyttelsesforordningens artikel 15 (indsigtsretten), artikel 16 (berigtigelsesretten), artikel 18 (ret til begrænsning) og artikel 21 (indsigelsesretten) finder ikke anvendelse ved behandlingen af personoplysningerne i projektet, da oplysningerne udelukkende behandles i videnskabeligt og statistisk øjemed, jf. databeskyttelseslovens § 22, stk. 5.

Tilsvarende finder retten til sletning efter forordningens artikel 17 ikke anvendelse, jf. databeskyttelsesforordningens artikel 17, stk. 3, litra d.

Artikel 19 (underretningspligt i forbindelse med berigtigelse eller sletning af personoplysninger eller begrænsning af behandling) samt artikel 20 (ret til dataportabilitet) finder ikke anvendelse, og artikel 22 (rettigheder i forbindelse med automatiske individuelle afgørelser, herunder profilering) vurderes ikke at være relevant i forbindelse med forskningsprojektet, idet der ikke forekommer fuldautomatiske afgørelser, men alene beslutningsstøtte.

4.8 Databehandlere

Det er alene AU, som har, og som fortsat vil anvende databehandlere i projektet.

4.8.1 Pilot #1

AU anvendte i Pilot #1 følgende databehandlere:

- Indiesoft.

- Danmarks Statistik

4.8.1.1 *Indiesoft*

AU og Indiesoft indgik en databehandleraftale forud for, at Indiesoft påbegyndte databehandling på vegne af AU. I databehandleraftalen var der blandt andet sikret kontrolforanstaltninger i form af årlig revisionserklæring, alternativt en egenerklæring. AU var derudover berettiget til – for egen regning – at udpege en uafhængig ekspert til at undersøge, om Indiesoft overholdt sine forpligtelser i forhold til databehandleraftalen. AU har i Pilot #1 dog ikke ført tilsyn med Indiesoft grundet pilotafprøvningens varighed. Indiesoft slettede alle personoplysninger fra Pilot #1 fra Indiesofts systemer og opbevaringsenheder efter overførslen til VIA og AU.

I overensstemmelse med den indgåede databehandleraftale benyttede Indiesoft sig under Pilot #1 af følgende underleverandør (underdatabehandler), som der var indgået en underdatabehandleraftale med:

- Amazon Web Services Ireland Limited.

4.8.1.2 *Danmarks Statistik*

AU og Danmarks Statistik havde ligeledes indgået en databehandleraftale forud for, at Danmarks Statistik påbegyndte databehandlingen på vegne af AU. Danmarks Statistik havde i aftalen forpligtet sig til én gang årligt at indhente en revisionserklæring fra en uafhængig tredjepart om Danmarks Statistiks overholdelse af kravene i Databeskyttelsesforordningen. Erklæringen var tilgængelig på Danmarks Statistiks hjemmeside.

Danmarks Statistik benyttede ingen underdatabehandlere i forbindelse med den pågældende databehandling.

4.8.2 *Pilot #2*

AU vil i Pilot #2 fortsat anvende Danmarks Statistik som databehandler. AU har herudover endnu ikke endeligt afklaret, hvilken/hvilke databehandler(e) AU ellers vil benytte sig af i Pilot #2. Dog forventer AU, at Indiesoft fortsat vil indgå i denne pilotafprøvning som databehandler.

Til brug for Pilot #2 vil AU (og hvis relevant VIA) sikre, at der bliver indgået alle nødvendige databehandleraftaler i overensstemmelse med databeskyttelsesforordningens artikel 28, hvori der blandt andet er sikret den nødvendige behandlingssikkerhed, adgang til kontrol mv.

4.9 Overførsel til tredjelande og/eller internationale organisationer

VIA og AU overfører ikke personoplysninger til tredjelande eller internationale organisationer i forbindelse med projektets behandling af personoplysninger.

5. TRIN 4: IDENTIFIKATION, EVALUERING OG HÅNDTERING AF RISICI

I dette trin i konsekvensanalysen skal VIA og AU identificere behandlingsaktivitetens risici for de registreredes rettigheder og frihedsrettigheder (risikoidentifikation), evaluere og beskrive disse risici ud fra deres sandsynlighed og alvor (risikoevaluering) samt håndtere risiciene ved hjælp af identifikation af afhjælpende foranstaltninger, jf. databeskyttelsesforordningens artikel 35, stk. 7, litra c-d. Formålet med de afhjælpende foranstaltninger er at nedbringe de identificerede risici til et acceptabelt niveau. De typiske risikostyringsstrategier vil være enten at eliminere, reducere eller acceptere den identificerede risiko.

5.1 Valg af evalueringskriterier for sandsynlighed og konsekvens

En risiko kan defineres som et scenarie, der beskriver en hændelse og konsekvenserne heraf, som vurderes i forhold til alvor og sandsynlighed. I forbindelse med risikoevalueringen skal der derfor foretages en vurdering af risikoens sandsynlighed og konsekvens. Vurderingen skal foretages for hver enkelt identificeret risiko set ud fra den registreredes perspektiv, men på et objektivt grundlag.

I denne konsekvensanalyse anvendes følgende evalueringskriterier for sandsynlighed:

4	Forventet: Det forventes, at hændelsen vil forekomme, herunder f.eks.: - Man har erfaring med hændelsen inden for de sidste 12 måneder - Hænder jævnligt hos andre offentlige myndigheder og private virksomheder (omtales ofte i pressen)
3	Moderat sandsynligt: Det er moderat sandsynligt, at hændelsen vil forekomme, herunder f.eks.: - Man har erfaring med hændelsen, men ikke inden for de sidste 12 måneder - Kendes fra andre offentlige myndigheder og private virksomheder i Danmark (omtales i pressen)
2	Mindre sandsynligt: Hændelsen forventes ikke at forekomme, herunder f.eks.: - Ingen erfaring med hændelsen - Kendes fra få andre offentlige myndigheder og private virksomheder, men ikke i Danmark
1	Usandsynligt: Det anses for næsten udelukket, at hændelsen nogensinde kan forekomme, herunder f.eks.: - Ingen erfaring med hændelsen - Kendes fra få andre offentlige myndigheder og private virksomheder, men ikke i Danmark

Tabel 1. Evalueringskriterier for sandsynlighed

I denne konsekvensanalyse anvendes følgende evalueringskriterier for konsekvens¹⁹:

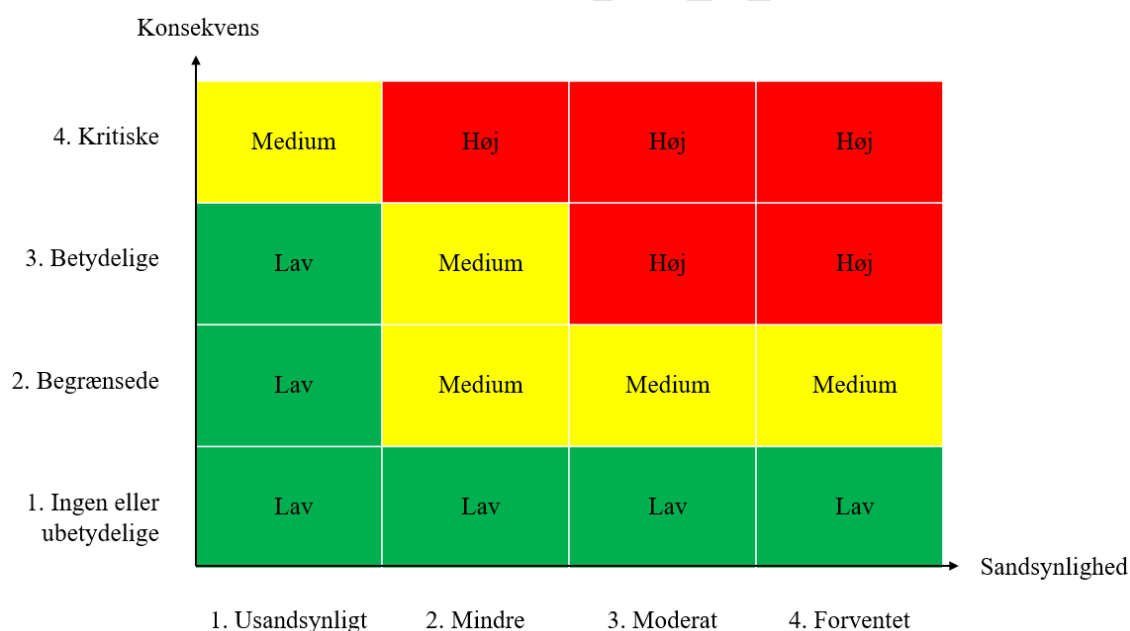
4	Kritiske konsekvenser: De registrerede kan opleve kritiske konsekvenser, som de ikke nødvendigvis kan overvinde, f.eks. økonomisk nød som betydelig gæld eller manglende evne til at arbejde, langsigtede psykiske eller fysiske lidelser, død mv.
3	Betydelige konsekvenser: De registrerede oplever betydelige konsekvenser, som de kan overvinde om end med alvorlige vanskeligheder, f.eks. identitetstyveri eller -svig, finansielle tab,

¹⁹ Se f.eks. punkt A.2 i Annex A til ISO/IEC 29134:2017, Information technology - Security techniques - Guidelines for privacy impact assessment; Datatilsynet m.fl., Vejledning om behandlingssikkerhed og databeskyttelse gennem design og standardindstillinger, juni 2018, s. 9, samt Digitaliseringsstyrelsen, Vejledning i vurdering af offentlige it-projekters potentielle konsekvenser for privatlivet, maj 2013, s. 12.

	blacklisting af banker, ejendomsskade, tab af beskæftigelse, stævning, forværring af sundhedstilstanden, tab af fortrolighed af personoplysninger, der er omfattet af tavshedspligt mv.
2	Begrænsede konsekvenser: De registrerede oplever begrænsede konsekvenser, som de vil være i stand til at overvinde med få vanskeligheder, f.eks. ekstra omkostninger, nægtelse af adgang til forretningstjenester, manglende forståelse, frygt, stress, mindre fysiske påvirkninger mv.
1	Ingen eller ubetydelige konsekvenser: De registrerede bliver enten ikke påvirket eller udsættes alene for få generende konsekvenser, som de uden problemer kan håndtere, f.eks. tidsforbrug brugt på at genindtaste oplysninger, irritationer, dårlig brugeroplevelse mv.

Tabel 2. Evalueringskriterier for konsekvens

Når evalueringskriterierne for sandsynlighed og konsekvens er fastlagt, kan hver enkelt identificeret risiko vurderes og kortlægges på et såkaldt risikokort. I denne konsekvensanalyse anvendes følgende risikokort:



Figur 1. Risikokort

5.2 Risici

VIA og AU har som angivet i punkt 4.6 foretaget en risikovurdering vedrørende behandlingssikkerhed i overensstemmelse med kravene hertil i databeskyttelsesforordningens artikel 32. Med udgangspunkt heri har VIA og AU fastsat passende sikkerhedsforanstaltninger for VIA og AU, der løbende er blevet overholdt. VIA og AU har og vil desuden fortsat stille de nødvendige krav til fremtidige databehandlere om behandlingssikkerhed samt sikre sig ret til at føre det nødvendige tilsyn med disse og deres underdatabehandlere. Der henvises herved til risikovurderingen, der er vedlagt denne konsekvensanalyse som Bilag B.

5.2.1 *Risiko nr. 1: Beslutningsstøtten kan føre til de facto fuldautomatiske afgørelser*

5.2.1.1 *Beskrivelse*

Beslutningsstøtten er alene et beslutningsstøtteredskab, der skal understøtte socialrådgiverens beslutning om, hvordan en underretning om et barns eller den unges mistvivl skal håndteres. Beslutningsstøtten må under ingen omstændigheder anvendes til de facto at træffe fuldautomatiske afgørelser. Der kan dog være risiko for, at socialrådgiverne ikke foretager en reel vurdering og efterprøvelse af Beslutningsstøttens risikoscore, dvs. at socialrådgiverne ukritisk lægger Beslutningsstøttens risikoscore til grund ("rubber stamping"), og herved ikke selvstændigt afgør, hvordan underretningen skal håndteres. Dette kan have flere årsager, f.eks. at socialrådgiveren ikke forstår, hvordan Beslutningsstøtten er nået frem til risikoscoren (manglende transparens). Den manglende efterprøvelse kan også skyldes, at socialrådgiveren stoler på Beslutningsstøtten og dermed stopper med at bruge sin egen kritiske sans og undlader at overveje, om risikoscoren kan være forkert ("automation bias"). Risikoen for de registrerede vil her være, at der ikke foretages saglig sagsbehandling, herunder at der ikke udøves de fornødne tiltag til beskyttelse af et barn eller ung i mistvivl, eller at der benyttes for indgribende foranstaltninger i forhold til disse.

5.2.1.2 *Foranstaltning nr. 1: Uddannelse af socialrådgiverne*

VIA og AU har og vil fortsat reducere denne risiko ved at uddanne socialrådgiverne med henblik på at sikre, at de forstår, hvordan Beslutningsstøtten fungerer og kommer frem til risikoscoren, herunder Beslutningsstøttens begrænsninger samt give socialrådgiverne klare kriterier for, hvornår de forventes at se bort fra Beslutningsstøttens risikoscore.

VIA og AU uddannede forud for afprøvningen i Pilot #1 socialrådgiverne i, at Beslutningsstøtten aldrig måtte stå alene, men alene skulle indgå i den socialfaglige vurdering sammen med de andre elementer, som socialrådgiveren almindeligvis baserer sin vurdering på, herunder lovgivning, teoretisk og forskningsbaseret viden, erfaring samt den lokale kontekst for arbejdet. VIA og AU uddannede desuden socialrådgiverne i Beslutningsstøttens begrænsninger, herunder at Beslutningsstøtten ikke har adgang til den samme mængde informationer som socialrådgiverne, der har adgang til flere informationer via pro-sateksten i journalen mv. Socialrådgiverne og deres ledere blev undervist over to gange, hvor anden undervisningsgang var en konkret træning i brug af modellen på pseudonymiserede underretninger. Dette med henblik på konkret oplæring i brugen af Beslutningsstøtten, herunder hvordan Beslutningsstøtten måtte benyttes i socialrådgiverens beslutningsproces. Det bemærkes, at Beslutningsstøtten i Pilot #1 først blev anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen.

VIA og AU vil forud for afprøvningen i Pilot #2 uddanne socialrådgiverne og deres ledere på samme vis som i Pilot #1. Derudover vil VIA og AU bistå kommunerne med at etablere et organisatorisk set-up, hvor socialrådgiverne løbende sikres faglig ledelsesmæssig understøttelse. Dette f.eks. ved at teamlederen i projektperioden løbende indgår i dialog med socialrådgiverne vedrørende anvendelse af redskabet i konkrete sager eller løbende tager stikprøver ud til fælles faglig sparring.

5.2.1.3 Foranstaltning nr. 2: Transparens omkring inputdata og vægtning heraf

VIA og AU vil desuden reducere risikoen for automatiske afgørelser ved transparens, herunder ved at sikre at datagrundlaget og Beslutningsstøttens vægtning af data er transparent for socialrådgiveren.

VIA og AU har besluttet at øge informationsniveauet til socialrådgiverne under Pilot #2 for at skabe transparens omkring, hvilke oplysninger Beslutningsstøtten har inddraget ved udregningen af risikoscoren, og hvordan disse oplysninger er blevet vægtet. Dette gøres ved at 1) implementere en datooversigt, der opsummerer alt data, som Beslutningsstøtten har anvendt til at beregne sin risikoscore, og 2) at præsentere socialrådgiverne for variabel indeks, der giver dem indblik i, hvilke typer af informationer, der har trukket scoren op/ned. Denne transparens vil blandt andet gøre socialrådgiveren i stand til at vurdere, om der er fejl i data, eller om der er nye væsentlige oplysninger i sagen, der ikke er registreret.

Det bemærkes, at socialrådgiverne i forbindelse med Pilot #1 ikke fik en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke fik forklaret risikoscoren.

Hvis evalueringen af Pilot #2 peger på, at Beslutningsstøttens risikoscorer ikke er tilstrækkeligt transparente og forklarlige, vil VIA og AU om nødvendigt redesigne Beslutningsstøtten forud for RCT/Iodtrækningsforsøget.

5.2.2 Risiko nr. 2: Manglende forklarlighed og transparens i forhold til Beslutningsstøttens risikoscore.

5.2.2.1 Beskrivelse

Det er nødvendigt, at VIA og AU kan forklare, hvordan Beslutningsstøtten når frem til sine risikoscorer (outputs), når Beslutningsstøtten anvendes som sagsunderstøttende værktøj. Dette indebærer, at VIA og AU skal kunne forklare Beslutningsstøttens bagvedliggende logik, samt at der skabes transparens omkring datagrundlaget, som en konkret score er baseret på.

Hvis VIA og AU ikke er i stand til at forklare, hvordan Beslutningsstøtten virker og kvaliteten af Beslutningsstøttens output samt rationale bag dette, skaber det risiko for, at fejl ved den foreslåede risikoscore ikke opdages. Risikoen for de registrerede vil her også være, at der ikke foretages saglig sagsbehandling, herunder at der ikke udøves de nødvendige tiltag til beskyttelse af et barn eller ung i mistrivsel, eller at der benyttes indgreb, hvor dette ikke var berettiget, herunder at der benyttes for indgribende foranstaltninger i forhold til et barn eller en ung.

5.2.2.2 Foranstaltning nr. 1: Transparens omkring inputdata og vægtning heraf

VIA og AU har, som netop redegjort for, besluttet at reducere risikoen ved at øge informationsniveauet til socialrådgiverne under Pilot #2 og herved skabe transparens omkring, hvilke oplysninger, Beslutningsstøtten har inddraget ved udregningen af risikoscoren, samt hvordan disse oplysninger er blevet vægtet.

Som netop anført fik socialrådgiverne i forbindelse med Pilot #1 ikke en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke fik forklaret risikoscoren.

VIA og AU vil desuden vejlede kommunerne i at journalisere risikoscoren samt grundlaget herfor.

5.2.2.3 Foranstaltning nr. 2: Dokumentation af Beslutningsstøttens formål, datagrundlag, modelvalg, performance mv.

VIA og AU vil desuden øge Beslutningsstøttens transparens ved at udarbejde et notat over følgende emner i relation til Pilot #1:

1. Beslutningsstøttens formål.
2. Hvilke datasæt, der er anvendt til udviklingen af modellen, og hvorfor disse er udvalgt.
3. Oplæringen af modellen, herunder hvordan trænings- og valideringsdatasættene er udvalgt og deres fordeling.
4. Hvilke algoritmer, der er testet og deres performance.
5. En redegørelse for endeligt valg af model og variable samt den bagvedliggende logik for netop denne model.
6. Variablenes vægtning.
7. Omregning til deciler.
8. Nødvendigheden af at bruge individdata.

Notatet vil blive ajourført i forhold til de ændringer, som følger af Pilot #2.

5.2.3 *Risiko nr. 3: Fejl i modellen (algoritmen)*

5.2.3.1 *Beskrivelse*

Hvis den algoritme, som Beslutningsstøtten baserer sig på, er fejlbehæftet, vil resultaterne fra Beslutningsstøtten (output) ligeledes være fejlbehæftede. Hvis Beslutningsstøttens algoritme er fejlbehæftet, er der risiko for usaglig påvirkning af socialrådgiverens sagsbehandling.

5.2.3.2 *Foranstaltning nr. 1: Test og validering forud for afprøvning*

Risikoen for fejl i modellen er reduceret ved følgende test og valideringer:

Forud for udviklingen af modellerne gennemførte VIA et kvalitativt studie af alle underretninger modtaget i Silkeborg Kommune og Hjørring Kommune i perioden januar til februar 2018. Formålet hermed var at konkretisere, hvilken beslutningsproces, Beslutningsstøtten skulle bistå med at skærpe. AU har forud for Pilot #2, som et led i valideringen af modellernes egnethed til at forudsige mistrivsel, desuden gennemført en række analyser med henblik på at sikre, at modellerne er følsomme over for andre saglige mål for mistrivsel, herunder f.eks. fremtidige sigtelser for kriminalitet og fremtidige somatiske diagnoser.

Beslutningsstøtten blev forud for Pilot #1 desuden testet på en række konkrete underretningssager i kommunerne med henblik på at sikre, at algoritmen inddrog de rigtige data og i øvrigt fungerede som tiltænkt. AU's databehandler, Indiesoft, foretog desuden et dataudtræk efter igangsættelsen af pilotafprøvningen, hvorefter AU tjekkede, at risikoen blev beregnet rigtigt. AU vil foretage samme tests forud for og under afprøvningen i Pilot #2. AU vil desuden forud for afprøvningen validere algoritmens performance på baggrund af registerdata fra Danmarks Statistik, Styrelsen for IT og Læring og Sundhedsdatastyrelsen. AU vil desuden løbende tage stikprøver, hvor det undersøges, om Beslutningsstøtten beregner scoren rigtigt.

Det forhold, at modellerne er statiske, afhjælper desuden risikoen for, at der opstår utilsigtede fejl under afprøvningen af Beslutningsstøtten.

5.2.3.3 *Foranstaltning nr. 2: Gentræning eller rekalkibrering af algoritmen*

Risikoen for fejl er desuden reduceret ved gentræning og rekalkibrering af algoritmen.

VIA og AU valgte på baggrund af erfaringerne fra Pilot #1 således at reducere ”*mistrivsel*” til ét enkelt outcomemål samt at udvide antallet af variable. Dette med henblik på at gøre algoritmen i stand til at skabe mere præcise prædiktioner, der giver et mere dækkende helhedsbillede for de kommunale socialrådgivere. Algoritmen skal i overensstemmelse hermed trænes på baggrund af en væsentlig større datamængde, end det var tilfældet forud for Pilot #1.

VIA og AU vurderede, at det ikke var nødvendigt at foretage gentræning eller rekalkibrering af algoritmen under Pilot #1, idet denne afprøvning alene forløb over en periode på fire måneder. Det er ligeledes VIAs og AU's vurdering, at det ikke er nødvendigt at gennemføre gentræning eller rekalkibrering under Pilot #2, idet denne afprøvning alene vil forløbe over 1-3 måneder. Modellen vil dog blive gentrænnet eller rekalkibreret forud for RTC/lodtrækningsforsøget, hvis afprøvningen i Pilot #2 viser et behov herfor. VIA og AU vil desuden løbende analysere og overveje, om modellen kan tilpasses/og eller forbedres i lyset af nye metoder og ny viden på området.

5.2.3.4 Foranstaltning nr. 3: Transparens omkring inputdata og vægtning heraf

VIA og AU vil, som beskrevet i de to forudgående afsnit, øge informationsniveauet til socialrådgiverne under Pilot #2, og herved give disse indsigt i hvilke oplysninger, Beslutningsstøtten har inddraget ved udregningen af risikoscoren samt hvordan disse oplysninger er blevet vægtet. Det er VIAs og AU's vurdering, at denne transparens vil øge sandsynligheden for, at fejl i algoritmen opdages, og derfor vil reducere risikoen for en usaglig påvirkning af socialrådgiverne.

Som tidligere beskrevet fik Socialrådgiverne i Pilot #1 ikke en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke blev givet samme indsigt i Beslutningsstøttens datagrundlag mv.

5.2.3.5 Foranstaltning nr. 4: Kvalificeret personale

VIA og AU har og vil desuden fortsat reducere risikoen for fejl i algoritmen ved at anvende fagligt kvalificeret personale til udvikling af de statistiske modeller. VIA og AU har i denne forbindelse blandt andet

sikret, at den primære udvikler har et indgående statistisk kendskab, hvilket både er opnået igennem uddannelse og erfaring. Derudover indgår der i projektet også andre medarbejdere med dybt statistisk kendskab.

5.2.4 Risiko nr. 4: Ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias).

5.2.4.1 Beskrivelse

Hvis Beslutningsstøttens datasæt – både under udvikling og drift – er præget af historisk (utilsigtet) skævhed, vil systemet med høj sandsynlighed også medføre utilsigtet direkte eller indirekte ulovlig forskelsbehandling. Selvom rådata (inputdata) registreres korrekt, kan datasættet muligvis ikke være fuldt repræsentativt, eller analysen af dataene kan indeholde skjulte skævheder. Ulovlig forskelsbehandling kan også opstå som følge af en forkert udvikling og programmering af algoritmer.

5.2.4.2 Foranstaltning nr. 1: Gennemgående træning af Beslutningsstøtten

Risikoen er søgt reduceret ved at træne Beslutningsstøtten på store mængder repræsentative registerdata fra Danmarks Statistik, inden Beslutningsstøtten er taget i anvendelse i praksis, herunder på et stort antal underretningssager.

Forud for afprøvningen og implementeringen af Beslutningsstøtten i Pilot #1 indledte VIA og AU desuden et samarbejde med kommunernes datavarehuse for at sikre, at de udvalgte variable var defineret på samme måde som de data, AU har haft tilgængelige via Danmarks Statistiks registre. AU vil ligeledes validere de udvalgte variable forud for afprøvningen i Pilot #2.

5.2.4.3 Foranstaltning nr. 2: Test og fravalg af variable

Risikoen er desuden søgt reduceret ved at sikre, at modellerne ikke giver anledning til prædiktioner baseret på bestemte parametre. Det er for Pilot #2 f.eks. besluttet, at modellen ikke skal medtage oplysninger om køn, etnicitet og socioøkonomisk status for at forhindre, at modellen viderefører en uhensigtsmæssig forskelsbehandling på baggrund af disse parametre.

VIA og AU vil ligeledes sikre, at ovenstående parametre ikke inkluderes i Pilot #2. VIA og AU vil desuden under udviklingen af modellen samt i evalueringen af afprøvningsfasen teste, om modellen uagtet bevæger sig i en retning, som kan tyde på uhensigtsmæssig forskelsbehandling.

Idet Beslutningsstøttens bagvedliggende model er statistisk, er det ikke nødvendigt at teste modellen hyppigere, idet den ikke ændrer sig i lyset af evt. nye mønstre i de data, den behandler i praksis.

5.2.4.4 *Foranstaltning nr. 3: Kvalificeret personale*

VIA og AU har og vil som anført i punkt 5.2.3.5 fortsat anvende personale med et indgående statistisk kendskab til udvikling af modellerne, hvilket bistår til at reducere risikoen for, at forkert udvikling og programmering fører til ulovlig forskelsbehandling.

5.2.5 ***Risiko nr. 5: Behandling af for mange eller ikke nødvendige oplysninger***

5.2.5.1 *Beskrivelse*

Projektet indebærer behandling af blandt andet følsomme personoplysninger samt CPR-numre. Hvis Beslutningsstøtten udvikles, uden at der sker en tilstrækkelig vurdering af, om behandlingen af personoplysningerne til udvikling eller anvendelse af systemet er begrænset til det nødvendige, relevante og tilstrækkelige, medfører dette en risiko for overtrædelse af princippet om dataminimering. I punkt 4 er der foretaget en vurdering af behandlingens nødvendighed og proportionalitet, herunder at der er behov for at behandle de enkelte typer af personoplysninger.

5.2.5.2 *Foranstaltning nr. 1: Dokumentation for vurdering af proportionaliteten i databehandlingen*

Risikoen vil blive reduceret derved, at AU vil udarbejde dokumentation for, at AU alene har og fortsat kun vil inddrage personoplysninger af betydning for projektets formål. AU vil i den forbindelse dokumentere deres overvejelser omkring, hvorfor brug af fiktive, anonymiserede og aggregerede oplysninger ikke er tilstrækkelige til oplæring, modeludvikling og test, jf. punkt 5.2.2.3.

5.2.5.3 *Foranstaltning nr. 2: Dokumentation for modelvalg*

AU vil i relation til Pilot #2 desuden undersøge og dokumentere, at der anvendes den matematiske model, som giver den bedst mulige prædiktion baseret på færrest data, jf. 5.2.2.3.

5.2.5.4 *Foranstaltning nr. 3: Retningslinjer for pseudonymisering*

VIA og AU vil desuden reducere risikoen ved at udarbejde og implementere retningslinjer for, hvornår data kan og skal pseudonymiseres.

5.2.5.5 *Foranstaltning nr. 4: Retningslinjer for periodisk evaluering af proportionaliteten*

AU vil desuden udarbejde og implementere retningslinjer for, hvor ofte model og data skal gennemgås med henblik på at vurdere, om data stadig er nødvendige i lyset af formålet. Formålet med denne foranstaltning er at sikre, at modellen til enhver tid alene bygger på det nødvendige datagrundlag, og herved alene behandler relevant data i drift.

5.2.6 **Risiko nr. 6: Dataophobning**

5.2.6.1 *Beskrivelse*

VIA og AU behandler personoplysningerne i projektet med henblik på at udvikle et beslutningsstøttende redskab, der er egnet til at skærpe socialrådgivernes beslutningsgrundlag ved underretninger om børn og unges mistrivsel. Hvis VIA og AU fortsætter med at opbevare oplysningerne efter formålets ophør, udgør dette en risiko for krænkelse af de registreredes rettigheder.

5.2.6.2 *Foranstaltning nr. 1: Overordnede retningslinjer for sletning*

Risikoen er reduceret ved, at både VIA og AU sikrer effektiv sletning i overensstemmelse med de retningslinjer og procedurer for sletning, der er angivet i punkt 2.6. VIA og AU vil således senest slette data 5 år efter indgåelse af aftale om publicering, således som det er foreskrevet hos flere videnskabelige tidsskrifter i forbindelse med publicering og eventuel verificering.

De implicerede systemer, hvor personoplysningerne opbevares og behandles, understøtter en effektiv og

uigenkaldelig sletning.

AU har i databehandleraftalerne med henholdsvis Indiesoft og Danmarks Statistik desuden betinget sig sletning efter AU's nærmere instruks. AU vil betinge sig samme vilkår under Pilot #2.

5.2.7 Risiko nr. 7: Uautoriseret adgang til oplysningerne

5.2.7.1 Beskrivelse

I forbindelse med både Pilot #1 og Pilot #2 kan der være en risiko for, at uvedkommende får adgang til de data, der indsamles via Beslutningsstøtten. Dette kan medføre et brud på persondatasikkerheden med potentielle krænkelse af de registreredes rettigheder og frihedsrettigheder til følge. Uvedkommende udgøres i denne henseende både af udefrakommende, der tvinger sig adgang til oplysningerne samt VIAs og AU's egne medarbejdere, Indiesofts medarbejdere og kommunernes medarbejdere, når disse ikke har et nødvendigt arbejdsbetinget behov for at have adgang til oplysningerne.

Det vil desuden udgøre en risiko, hvis uvedkommende under Pilot #2 får adgang til de statistiske modeller, da en påvirkning af disse kan medføre fejl i algoritmen og derved skabe risiko for, at Beslutningsstøtten påvirker sagsbehandlingen på usaglig vis.

Uautoriseret adgang til oplysningerne kan være et resultat af såvel et utilstrækkeligt eksternt som internt sikkerhedsniveau. Uautoriseret adgang dækker tillige over tilfælde, hvor Indiesoft uretmæssigt anvender oplysningerne til egne formål eller uretmæssigt sender oplysningerne til underleverandører, herunder til tredjelande.

5.2.7.2 Foranstaltning nr. 1: Bruger- og adgangsstyring

Beslutningsstøtten anvender den seneste udgave af Microsofts løsning til autorisation og adgangskontrol. Brugere oprettes ved, at der sendes et engangslink til en side, hvor de skal oprette et komplekst kodeord. Linket i mailen udløber efter kort tid. Gennem brug af roller er systemet opbygget således, at det er muligt at styre, hvilke brugere, der må oprettes og ændre personoplysninger, og hvilke brugere, der kan oprette socialrådgivere og trække rapporter.

AU havde i databehandleraftalen med Indiesoft vedrørende Pilot #1 desuden instrueret Indiesoft i at begrænse adgangen til personoplysningerne til de medarbejdere, for hvem det var nødvendigt at have adgang for at opfylde Indiesofts forpligtelser over for AU. Indiesoft var desuden instrueret om at behandle oplysningerne i projektet fortroligt og ikke uberettiget at udnytte, videregive eller på anden måde videreformidle de fortrolige oplysninger.

De data, der blev videregivet til AU efter Pilot #1, opbevares på et sikkert drev, hvor kun projektgruppens AU-ansatte medlemmer samt René Karberg fra Indiesoft²⁰ har adgang. Michael Rosholm (data owner) og Sanne Dalgaard Toft (data manager) kan tildele og fjerne personers adgang til mappe samt styre læse- og skriverettigheder. Adgangen til det sikre drev opdateres løbende.

AU's medarbejdere skal desuden benytte VPN med to-faktor autentifikation ved fjernadgang (fx hjemmearbejdsplads). Den ene faktor er brugernavn og password, mens den anden faktor er afhængig af, hvem medarbejderen er, og hvorvidt medarbejderen benytter udstyr, der er privat eller ejet af AU.

AU's medarbejdere kan kopiere data til eget eller eksternt udstyr fra det sikre drev, men denne overførsel logges, hvilket gør AU i stand til følge op, hvis der er mistanke om, at en medarbejder eller konsulent misbruger dennes adgang, jf. punkt 5.2.7.3.

I VIA opbevares data enten på et fælles drev med adgang for flere eller på et drev udelukkende tilknyttet den enkelte medarbejder. Hvis flere skal have adgang, og data lægges på et fælles forskningsdrev, er det kun forskergruppens medlemmer fra VIA, der har adgang. VIAs medarbejdere vil i relation til projektet alene anvende forsknings-pc'er. At arbejde på en forsker-pc indebærer bl.a., at der ikke kan opnås lokal-administratorrettigheder. Adgangen gives af VIA service, hvilket vil sige, at adgangskontrollen er centralt administreret.

Det er alene udvalgte socialrådgivere, der har haft adgang til Beslutningsstøtten i Pilot #1. De udvalgte socialrådgivere har ikke haft mulighed for at se hinandens brug af Beslutningsstøtten, og efter endt workflow har den enkelte sagsbehandler heller ikke haft mulighed for at genbesøge vurderingerne i systemet.

²⁰ René Karbergs/Indiesofts adgang til det sikrede drev er alene åbent i forbindelse med afprøvningerne, hvor René Karberg/Indiesoft skal kunne levere udtræk af data. I mellemliggende perioder mellem afprøvninger lukkes René Karbergs/Indiesofts adgang til drevet.

De samme forhold som beskrevet ovenfor vil fortsat gøre sig gældende i forhold til Pilot #2. Den enkelte sagsbehandler vil dog kunne genbesøge vurderingerne i systemet, idet der som nævnt i punkt 2.2.2.2 vil være mulighed for at lade Beslutningsstøtten følge sagen.

5.2.7.3 Foranstaltning nr. 2: Logning

I AU logges al databehandling på det sikre drev. Data, som undtagelsesvist opbevares på en ekstern enhed (fx krypteret USB eller SD-kort), logges ikke. Dog logges det, at dataene er blevet overført til den eksterne enhed. Loggen krypteres efter ét år og opbevares herefter af AU i 20 år, jf. AU's generelle retningslinjer om logning.

Indiesoft blev under Pilot #1 instrueret i at foretage fuldstændig logning på enhver behandling af personoplysningerne i systemet, hvorved kommunernes færden i systemet ligeledes blev logget. Loggen fra Pilot #1 blev slettet af Indiesoft, da pilotprojektet blev lukket ned, og backup blev slettet efter 70 dage. Indiesoft overførte dog loggen til AU. AU har efterfølgende slettet denne log. Indiesoft overførte samtidigt dataudtrækket fra afprøvningen til AU og VIA, der opbevarer denne i overensstemmelse med retningslinjerne anført i punkt 2.6.

I VIA logges alle transaktioner til og fra forskningsserverne. Logningen opbevares i et år.

VIA og AU kan med logningen føre kontrol med hvem, der har tilgået data, og hvornår de har gjort dette. Logningen forhindrer ikke direkte uautoriseret adgang til personoplysninger, men den gør VIA og AU i stand til at følge op, hvis der er mistanke om, at en medarbejder eller konsulent misbruger vedkommenes adgang udover det arbejdsbetingede behov. Derudover kan logningen have en præventiv effekt, hvis medarbejdere gøres bevidste om, at deres handlinger på de sikrede drev og servere bliver loggede og potentielt gennemgået med henblik på kontrol.

VIA og AU vil følge samme retningslinjer som angivet ovenfor i Pilot #2.

5.2.7.4 Foranstaltning nr. 3: Kryptering

Der anvendes anerkendt kryptering mellem AU og AU's eksterne leverandører. AU's servere overholder i øvrigt alle sikkerhedskrav.

VIA's forskningsservere er BitLocker-krypterede, og adgang til disse kræver både en forsker-pc og VPN-adgang. VIA's forsker-pc'er kan alene anvende et USB-drev, hvis det er krypteret.

I forbindelse med afprøvningen i Pilot #1 blev dataene opbevaret på henholdsvis Hjørring Kommunes hardware i Hjørring og Indiesofts hardware hos Amazon i Irland. Alle dataene blev krypterede under transporten via Internettet, og navn samt CPR-nummer blev opbevaret krypteret i databaserne. AU vil i Pilot #2 ligeledes stille krav om kryptering til sin(e) databehandler(e).

I det omfang, det bliver nødvendigt at udveksle personhenførbare data mellem AU og VIA i Pilot #2, vil data sendes krypteret med sikker post med digital signatur eller via krypteret USB/SD-kort (f.eks. krypteret med 7-zip eller BitLocker).

Disse foranstaltninger bidrager samlet set til at reducere risikoen for, at uvedkommende eksterne aktører får uberettiget adgang til personoplysningerne, herunder f.eks. med henblik på misbrug eller forvanskning af data.

5.2.8 Risiko nr. 8: Manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset

5.2.8.1 Beskrivelse

Det er af afgørende betydning, at VIA og AU har et klart overblik over rækkevidden af deres dataansvar, når de behandler personoplysningerne i projektet. Dette gælder både i forhold til databehandlere, men også i forhold til andre tredjeparter såsom Danmarks Statistik, samarbejdskommunerne mv. Manglende overblik over parternes dataansvar kan medføre risici for de registrerede, da det risikeres, at hverken VIA og AU påtager sig ansvar for en del af behandlingen, herunder at VIA og AU ikke indgår de aftaler, som er nødvendige i henhold til databeskyttelsesretten.

5.2.8.2 Foranstaltning nr. 1: Fastlæggelse af dataansvar

Som redegjort for i punkt 1.2.1 er VIA og AU selvstændige dataansvarlige og vil fortsat agere med denne ansvarsfordeling. VIA's og AU's dataansvar er herved fastlagt for begge piloter.

5.2.8.3 *Foranstaltning nr. 2: Indgåelse af de nødvendige aftaler*

AU behandler, herunder opbevarer, fortsat de oplysninger, som blev indsamlet i forbindelse med Pilot #1. VIA har slettet de data, der blev delt i forbindelse med Pilot #1. VIA og AU vil sikre, at de i Pilot #2 indgår en videregivelsesaftale i overensstemmelse med videregivelsesbekendtgørelsen, inden de evt. deler data mellem dem.

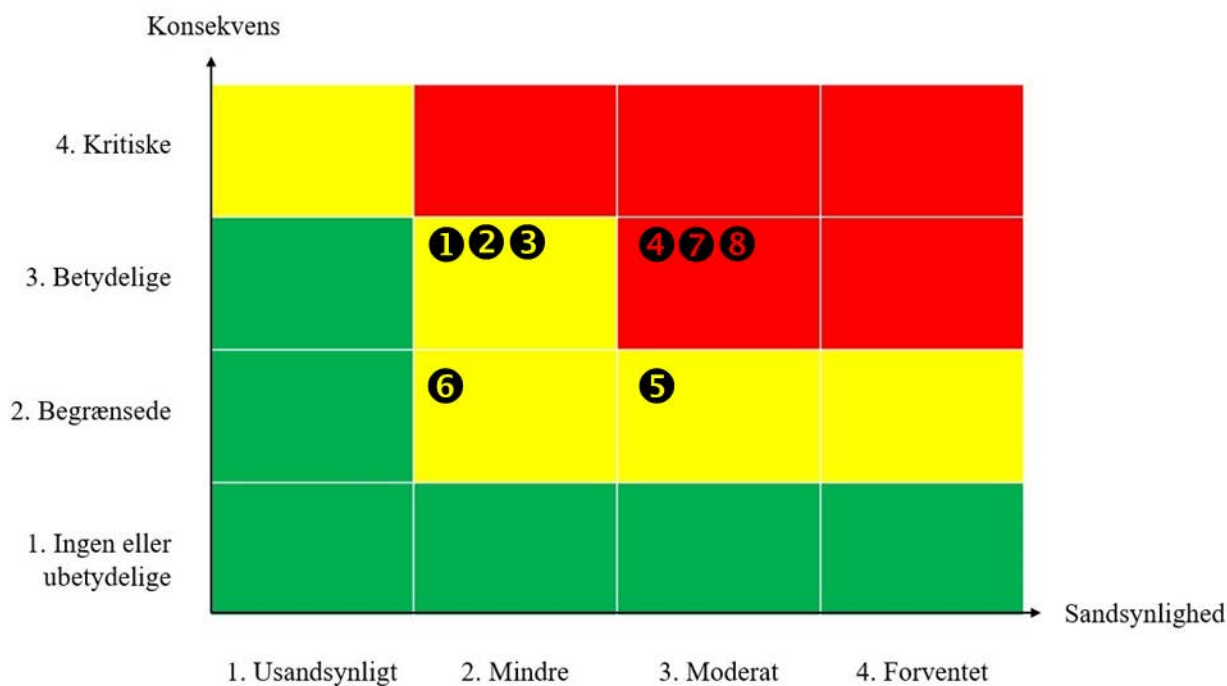
AU vil i henhold til Pilot #2 desuden indgå en videregivelsesaftale med den valgte samarbejdskommune.

VIA og AU vil ligeledes sikre, at hver part indgår de(n) nødvendige databehandleraftale(r), jf. databeskyttelsesforordningens artikel 28, når det endeligt er afgjort, hvilken/hvilke databehandler(e), de vil gøre brug af under pilotforsøget.

5.2.9 *Placering af risici i risikokort før og efter afhjælpende foranstaltninger*

5.2.9.1 *Risikokort inden afhjælpende foranstaltninger*

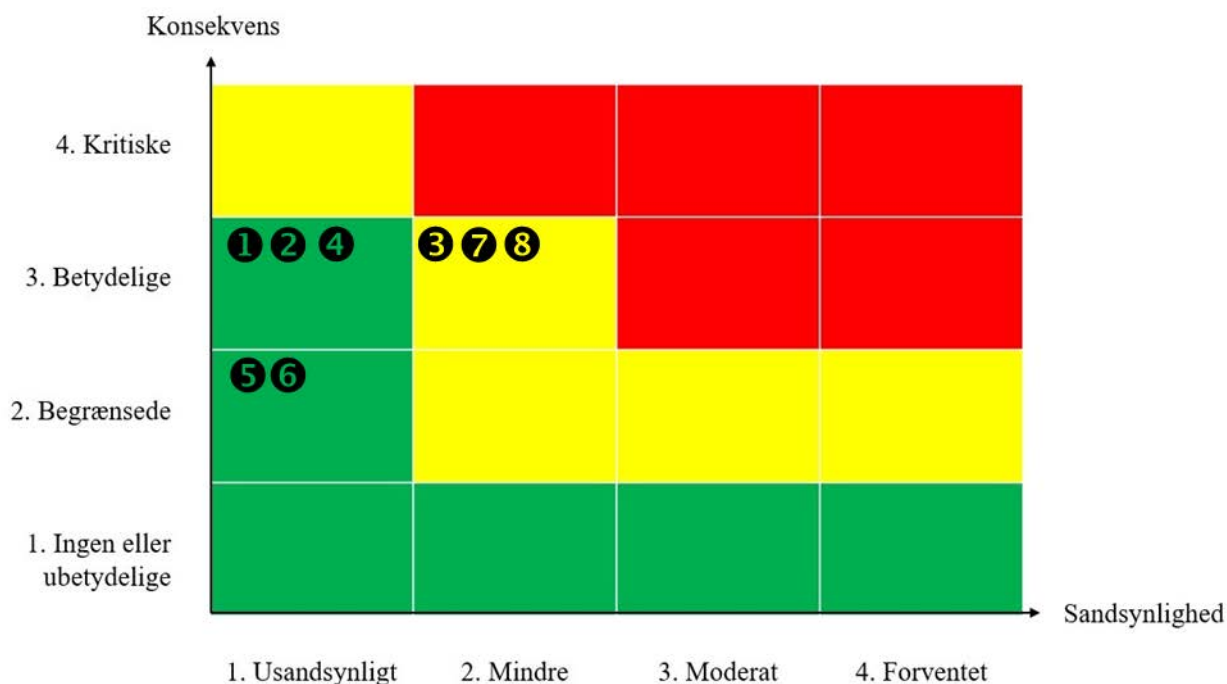
Nedenfor er de identificerede 8 risici blevet vurderet og placeret i et risikokort, der angiver risicienes potentielle konsekvenser og sandsynlighed *inden* VIAs og AU's gennemførelse af de identificerede afhjælpende foranstaltninger:



Figur 2. Udfyldt risikokort inden afhjælpende foranstaltninger (initialrisiko)

5.2.9.2 Risikokort efter afhjælpende foranstaltninger

Nedenfor er de identificerede 8 risici blevet vurderet og placeret i et risikokort, der angiver risicienes potentielle konsekvenser og sandsynlighed *efter* VIAs og AU's gennemførelse af de identificerede afhjælpende foranstaltninger:



Figur 3. Udfyldt risikokort efter afhjælpende foranstaltninger (restrisiko)

5.3 Residualrisiko

Risiko nr.	Valg af foranstaltning for at håndtere risiko	Effekt på risiko	Restrisiko	Er foranstaltningerne implementeret?
1	Uddannelse af socialrådgiverne	Risikoen er blevet reduceret	Lav	De er undervejs og vil være implementeret, inden Pilot #2 påbegyndes.
	Transparens omkring inputdata og vægtning heraf			
2	Transparens omkring inputdata og vægtning heraf	Risikoen er blevet reduceret	Lav	De er undervejs og vil være implementeret, inden Pilot #2 påbegyndes.

	Dokumentation af Beslutningsstøttens formål, datagrundlag, modelvalg, performance mv.			
3	Test og validering forud for afprøvnings	Risikoen er blevet reduceret	Medium	Ja, de blev implementeret inden Pilot #1 og vil fastholdes i Pilot #2.
	Gentræning/-kalibrering af algoritmen			
	Transparens omkring inputdata og vægtning heraf			
	Kvalificeret personale			
4	Gennemgående træning af Beslutningsstøtten	Risikoen er blevet reduceret	Lav	Ja, de blev implementeret inden Pilot #1 og vil fastholdes i Pilot #2.
	Tests og fravalg af variable			
	Kvalificeret personale			
5	Dokumentation for vurdering af proportionaliteten i databehandlingen	Risikoen er blevet reduceret	Lav	De er undervejs og vil være implementeret inden Pilot #2 påbegyndes.
	Dokumentation for modelvalg			

	Retningslinjer for pseudonymisering			
	Retningslinjer for periodisk evaluering af proportionaliteten			
6	Overordnede retningslinjer for sletning	Risikoen er blevet reduceret	Lav	Ja, de blev implementeret inden Pilot #1 og vil fastholdes i Pilot #2.
7	Bruger- og adgangsstyring	Risikoen er blevet reduceret	Medium	Ja, de blev implementeret inden Pilot #1 og vil fastholdes i Pilot #2.
	Logning			
	Kryptering			
8	Fastlæggelse af dataansvar	Risikoen er blevet reduceret	Medium	De er undervejs og vil være implementeret, inden Pilot #2 påbegyndes.
	Indgåelse af nødvendige aftaler			

Tabel 4. Residualrisikoscore

6. TRIN 5: KONKLUSION OG LEDELSESGODKENDELSE

6.1 Samlede residualrisiko og høring af Datatilsynet

På baggrund af de samlede residualrisikoscoringer *efter* afhjælpende foranstaltninger, som VIA og AU har vurderet og angivet under håndteringen af de enkelte identificerede risici, vurderes den samlede residualrisiko for behandlingen af personoplysninger i projektet at være **lav til mellem** med overvægt til lav risiko.

☐ Høj

☐ Medium

☒ Lav

Som følge heraf har det ikke været vurderet nødvendigt for VIA og AU at høre Datatilsynet forud for projektets behandling af personoplysninger.

6.2 VIAs ledelses godkendelse af konsekvensanalysen

☒	Godkendt	Behandlingsaktiviteten kan påbegyndes, hvis de afhjælpende foranstaltninger i implementeringsplanen bliver gennemført.
☐	Betinget godkendt	Behandlingsaktiviteten kan kun påbegyndes, hvis nærmere beskrevne ændringer foretages. Efter den betingede godkendelse skal ledelsen præsenteres for en ny, revideret konsekvensanalyse med henblik på endelig godkendelse.
☐	Ikke godkendt	Behandlingsaktiviteten kan ikke gennemføres.

Ovenstående valg af godkendelse har følgende begrundelse:

Indeværende konsekvensanalyse indeholder efter VIAs vurdering en hensigtsmæssig og kompetent bearbejdning af projektets forløb. Analysen er desuden udarbejdet i samarbejde med eksterne eksperter inden for feltet, og behandler de forventede punkter, hvorfor VIAs ledelse godkender konsekvensanalysen.

Vurderer VIAs ledelse, at konsekvensanalysen skal offentliggøres enten helt eller delvist?

- ☒ Ja, konsekvensanalysen skal offentliggøres i sin helhed
- ☐ Ja, konsekvensanalysen skal delvist offentliggøres
- ☐ Nej, konsekvensanalysen skal ikke offentliggøres

Ovenstående valg i forhold til offentliggørelse har følgende begrundelse:

VIA's ledelse mener, at konsekvensanalysen bør offentliggøres for at skabe transparens i forbindelse med denne type af projekter.

6.3 AU's ledelses godkendelse af konsekvensanalysen

☒	Godkendt	Behandlingsaktiviteten kan påbegyndes, hvis de afhjælpende foranstaltninger i implementeringsplanen bliver gennemført.
☐	Betinget godkendt	Behandlingsaktiviteten kan kun påbegyndes, hvis nærmere beskrevne ændringer foretages. Efter den betingede godkendelse skal ledelsen præsenteres for en ny, revideret konsekvensanalyse med henblik på endelig godkendelse.
☐	Ikke godkendt	Behandlingsaktiviteten kan ikke gennemføres.

Ovenstående valg af godkendelse har følgende begrundelse:

Indeværende konsekvensanalyse indeholder efter AU's vurdering en hensigtsmæssig og kompetent bearbejdning af projektets forløb. Analysen er desuden udarbejdet i samarbejde med eksterne eksperter inden for feltet, og behandler de forventede punkter, hvorfor AU's ledelse godkender konsekvensanalysen.

Vurderer AU's ledelse, at konsekvensanalysen skal offentliggøres enten helt eller delvist?

☒ Ja, konsekvensanalysen skal offentliggøres i sin helhed

☐ Ja, konsekvensanalysen skal delvist offentliggøres

☐ Nej, konsekvensanalysen skal ikke offentliggøres

Ovenstående valg vedrørende offentliggørelse har følgende begrundelse:

AU's ledelse mener, at konsekvensanalysen bør offentliggøres for at skabe transparens i forbindelse med denne type af projekter.

7. FASTLÆGGELSE AF EN IMPLEMENTERINGSPLAN

Konsekvensanalysen har vist, at der skal foretages en række mitigerende foranstaltninger for at nedbringe de identificerede risici til et acceptabelt niveau. Nedenfor følger en plan for gennemførelsen af disse foranstaltninger, herunder med angivelse af de ansvarlige personer herfor.

Risiko-nummer	Afhjælpende foranstaltninger, der skal gennemføres	Frist for gennemførelse	Status på gennemførelse	Ansvarlig person
Angiv nummeret på den risiko, som den afhjælpende for-	Beskriv den relevante afhjælpende foranstaltning.	Angiv den valgte frist for gennemførelse.	Angiv status på gennemførelsen af den afhjælpende foranstaltning, f.eks.	Angiv det fulde navn på den person, der er ansvarlig for at gennemføre den nævnte afhjæl-

anstaltning knytter sig til.			"ikke påbegyndt", "i proces" eller "færdig.	pende foranstaltning inden den valgte frist.
1, 3 og 4	Uddannelse af socialrådgivere, øget transparens omkring inputdata og vægtning heraf samt træning og test af Beslutningsstøtten.	Før afprøvelse i Pilot #2 og indledningsvist under afprøvelse i Pilot #2	Endnu ikke gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)
2	Dokumentation af Beslutningsstøttens formål, datagrundlag, modelvalg, performance mv.	01-03-21	Gennemført.	Andreas Rasch-Christensen (VIA) Michael Rosholm (AU)
5	Retningslinjer for pseudonymisering.	15-02-2021	Gennemført.	Andreas Rasch-Christensen (VIA)
	Retningslinjer for periodisk evaluering af proportionaliteten.			Michael Rosholm (AU)
8	Indgåelse af videregivelsesaftale mellem AU og VIA for Pilot #1.	Videregivelsesaftale mellem AU og VIA for Pilot #1 er indgået den 27. januar 2021. Resten	Endnu ikke gennemført	Andreas Rasch-Christensen (VIA)
	Indgåelse af videregivelsesaftale mellem AU og VIA for Pilot #2.			Michael Rosholm (AU)
	Indgåelse af videregivelsesaftale mellem AU og samarbejdskommune for Pilot #2.			

	Indgåelse af databehandleraftale(r) med databehandler(e) for Pilot #2.	af de nævnte foranstaltninger skal være implementeret senest før videregivelse eller overførsel af data.		
--	--	--	--	--

8. BILAG TIL KONSEKVENSANALYSEN

Følgende bilag vedlægges eventuelt til konsekvensanalysen:

- ☐ Dataflow over behandlingsaktivitetens forløb
- ☐ Forudgående tærskelanalyse af behandlingsaktiviteten, dvs. vurdering af, hvorfor der skal laves en konsekvensanalyse
- ☐ Databeskyttelsesrådgiverens bemærkninger
- ☐ Indhentede synspunkter fra registrerede eller deres repræsentanter
- ☐ Datatilsynets rådgivning i forbindelse med høringsprocessen
- ☒ Andre (indsættes i kommentarboks nedenfor):

Bilag A: Registre fra Danmarks Statistik, der er anvendt under udviklingen af Beslutningsstøtten

Bilag B: Risikovurdering af behandlingen af persondata i forbindelse med forskningsprojekterne "Beslutningsstøtte ved underretninger" og "Underretninger i fokus"

Bilag C: Mitigerende foranstaltninger for forskningsprojekterne Beslutningsstøtte ved underretninger og Underretninger i fokus

9. TRIN 6: AJOURFØRING AF KONSEKVENSANALYSEN

VIA og AU vil regelmæssigt gennemgå konsekvensanalysen vedrørende databeskyttelse og de behandlingsaktiviteter, som vurderes i denne, jf. databeskyttelsesforordningens artikel 35, stk. 11. Det vil som

minimum ske, når der sker en ændring af risikoen ved behandlingen af aktiviteten og forud for RTC/lod-trækningsforsøget igangsættes.

Denne konsekvensanalyse vedrørende databeskyttelse skal fremadrettet ajourføres efter følgende procedure:

Andreas Rasch-Christensen (forskningschef, VIA) og Michael Rosholm (professor, AU) er ansvarlige for opdateringen af denne konsekvensanalyse.

De dataansvarlige skal løbende rapportere om gennemførelse af de mitigerende handlinger samt forelægge enhver ændring af konsekvensanalysen for databeskyttelsesrådgiverne.

Konsekvensanalysen skal revideres, hvor der er en ændring i projektet, der medfører en øget risiko for den registreredes rettigheder og frihedsrettigheder, som minimum helårligt.

Bilag A: Registre fra Danmarks Statistik, der er anvendt under udviklingen af Beslutningsstøtten

I det følgende oplystes de registre, der er anvendt til AU's udvikling af Beslutningsstøtten for henholdsvis Pilot #1 og Pilot #2. Idet kun udvalgte variable fra registrene er blevet anvendt til udviklingen, er disse variables navn og indhold kort angivet.

1. UDVIKLINGEN AF PILOT #1

1.1 Register 1: Underret_tryk1415 (svarer til det nuværende "Børn og unge – Underretninger")

Følgende variable er anvendt inden for dette register:

- o PNR (Personnummer)
- o BAGGRUND1-BAGGRUND20 (Baggrund 1-20 for underretningen)
 - o BAGGRUND1 Kriminalitet (den unge)
 - o BAGGRUND2 Misbrug (den unge)
 - o BAGGRUND3 Seksuelt krænkende adfærd (den unge)
 - o BAGGRUND4 Selvskadende adfærd
 - o BAGGRUND5 Udadreagerende adfærd
 - o BAGGRUND6 Indadreagerende adfærd
 - o BAGGRUND7 Selvmordstanker eller -forsøg
 - o BAGGRUND8 Social isolation
 - o BAGGRUND9 Ulovligt skolefravær
 - o BAGGRUND10 Kriminalitet i hjemmet
 - o BAGGRUND11 Misbrug i hjemmet
 - o BAGGRUND12 Omsorgssvigt overfor barnet
 - o BAGGRUND13 Seksuelt overgreb mod barnet
 - o BAGGRUND14 Voldeligt overgreb mod barnet
 - o BAGGRUND15 Udsættelsessag/boligforhold

- o BAGGRUND16 Psykisk funktionsnedsættelse hos barn
- o BAGGRUND17 Psykisk funktionsnedsættelse hos forældre
- o BAGGRUND18 Fysisk funktionsnedsættelse hos barn
- o BAGGRUND19 Fysisk funktionsnedsættelse hos forældre
- o BAGGRUND20 Andet UNDERRETNINGSDATO (Underretningsdato)
- o UNDERRETNINGSTYPE (Underretningstype)

1.2 Register 2: BEF (Befolkningen)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- PNRM (Personnummer på mor)
- PNRF (Personnummer på far)
- FOED_DAG (Fødselsdato)
- KOEN (Køn)
- KOM (Kommunekode)
- IE_TYPE (Indvandrere, efterkommere, personer med dansk oprindelse)

1.3 Register 3: BUFO (Børn og unge forebyggende foranstaltninger)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- PGF (Paragrafkode)
- SAG_VFRA (Startdato for hændelsen)
- SAG_VTIL (Slutdato for hændelsen)

1.4 Register 4: BUAH (Børn og unge anbragte hændelsesregister)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
 - EFFEKTDATO (Effektueringsdato)
 - HAENDELSE (Hændelse i anbringelsessag)
 - NYKOM (Handlekommune)
-

1.5 Register 5: HUST (Husstandsoplysninger)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- BOPIKOM (Adressen i kommunen)
- YEAR (Året hvor den pågældende information er registeret)

2. PILOT #2

2.1 Register 1: BUU (Børn og unge – Underretninger)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- BAGGRUND1-BAGGRUND16 (Baggrund 1-16 for underretningen)
 - o BAGGRUND1 Misbrug hos barn/ung
 - o BAGGRUND2 Kriminalitet hos barn/ung
 - o BAGGRUND3 Skoleproblemer hos barn/ung fx fravær
 - o BAGGRUND4 Anden bekymrende adfærd hos barn/ung fx udadreagerende adfærd
 - o BAGGRUND5 Betydelig eller varig nedsat fysisk eller psykisk funktionsevne hos barn/ung
 - o BAGGRUND6 Sundhedsforhold hos barn/ung
 - o BAGGRUND7 Overgreb mod barn/ung fx seksuelt eller voldeligt
 - o BAGGRUND8 Anden form for omsorgssvigt over for barn/ung
 - o BAGGRUND9 Misbrug hos forældre
 - o BAGGRUND10 Kriminalitet hos forældre
 - o BAGGRUND11 Anden bekymrende adfærd hos forældre
 - o BAGGRUND12 Betydelig eller varigt nedsat fysisk eller psykisk funktionsevne hos forældre
 - o BAGGRUND13 Højt konfliktniveau eller vold i hjemmet mellem voksne
 - o BAGGRUND14 Utilstrækkelig omsorg fra forældre
 - o BAGGRUND15 Fogedsag, hjemløshed eller udsættelse fra bolig
 - o BAGGRUND16 Andet
- UNDERRETNINGSDATO (Underretningsdato)
- UNDERRETNINGSTYPE (Underretningstype)

2.2 Register 2: Underret_tryg1415 (svarer til det nuværende "Børn og unge – Underretninger")

Følgende variable er anvendt inden for dette register:

- o PNR (Personnummer)
- o BAGGRUND1-BAGGRUND20 (Baggrund 1-20 for underretningen)
 - o BAGGRUND1 Kriminalitet (den unge)
 - o BAGGRUND2 Misbrug (den unge)
 - o BAGGRUND3 Seksuelt krænkende adfærd (den unge)
 - o BAGGRUND4 Selvskadende adfærd
 - o BAGGRUND5 Udadreagerende adfærd
 - o BAGGRUND6 Indadreagerende adfærd
 - o BAGGRUND7 Selvmordstanker eller -forsøg
 - o BAGGRUND8 Social isolation
 - o BAGGRUND9 Ulovligt skolefravær
 - o BAGGRUND10 Kriminalitet i hjemmet
 - o BAGGRUND11 Misbrug i hjemmet
 - o BAGGRUND12 Omsorgssvigt overfor barnet
 - o BAGGRUND13 Seksuelt overgreb mod barnet
 - o BAGGRUND14 Voldeligt overgreb mod barnet
 - o BAGGRUND15 Udsættelsessag/boligforhold
 - o BAGGRUND16 Psykisk funktionsnedsættelse hos barn
 - o BAGGRUND17 Psykisk funktionsnedsættelse hos forældre
 - o BAGGRUND18 Fysisk funktionsnedsættelse hos barn
 - o BAGGRUND19 Fysisk funktionsnedsættelse hos forældre
- o BAGGRUND20 Andet UNDERRETNINGSDATO (Underretningsdato)
- o UNDERRETNINGSTYPE (Underretningstype)

2.3 Register 3: BEF (Befolkningen)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- PNRM (Personnummer på mor)
- PNRF (Personnummer på far)
- FOED_DAG (Fødselsdato)
- FM_MARK (Forældremarkering)
- CIVST (Civilstand)

2.4 Register 4: BUFO (Børn og unge forebyggende foranstaltninger)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- PGF (Paragrafkode)
- SAG_VFRA (Startdato for hændelsen)
- SAG_VTIL (Slutdato for hændelsen)

2.5 Register 5: BUAH (Børn og unge anbragte hændelsesregister)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- EFFEKTDATO (Effektueringsdato)
- HAENDELSE (Hændelse i anbringelsessag)

2.6 Register 6: DOD (Døde i Danmark)

Følgende variable er anvendt inden for dette register:

- o PNR (Personnummer)
- o DODDATO (Dato for dødsfald)

2.7 Register 7: FLYT (Flytninger i Danmark)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- HAEND_DATO (Dato for flytningen)

3. PILOT #2 – VALIDERING AF MODELLEN

AU arbejder i forbindelse med Pilot#2 med at validere algoritmens performance på baggrund af registerdata fra Danmarks Statistik, Styrelsen for IT og Læring og Sundhedsdatastyrelsen. Disse variable er ikke anvendt til at udvikle modellen, men alene at evaluere dens performance. Variablene oplistes nedenfor.

3.1 Register 1: KROF (Ofre for straffelovsforbrydelser)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- YEAR (Året hvor den pågældende information er registeret)
- OFR_GERFRADT (Gerningens startdato)
- OFR_GER7 (Offeret gerningsindholdskode DS)

3.2 Register 2: KRSI (Kriminalstatistik sigtelser)

Følgende variabler er anvendt inden for dette register:

- PNR (Personnummer)
- YEAR (Året hvor den pågældende information er registeret)
- SIG_SIGTDTO (Sigtelsesdato)
- SIG_GER1DTO (Gerningsdag, startdato)

3.3 Register 3: LPRADMDIAG (Landspatientregisteret)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- YEAR (Året hvor den pågældende information er registeret)
- C_DIAG (Diagnosekode)

3.4 Register 4: PSYK_ADM (Landspatientregistret psykiatri - administrative oplysninger)

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- RECNUM (LPR-identnummer)
- D_HENDTO (Henvisningsdato)
- D_INDDTO (Indlæggelsesdato)

3.5 Register 5: PSYK_DIAG (Landspatientregistret psykiatri - diagnoser)

Følgende variable er anvendt inden for dette register:

- RECNUM (LPR-identnummer)
- C_DIAG (Diagnosekode)

3.6 **Register 6: Trivsel_2015_2018 (De nationale trivselsmålinger, data fra Styrelsen for IT og Læring)**

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- AAR (Året hvor den pågældende information er registeret)
- Q1 - Q40 (Spørgsmålene er tilgængelige på Børne- og Undervisningsministeriets hjemmeside, her: <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/trivselsmaaling>)

3.7 **Register 7: Fravaer_2011_2018 (Skolefravær, data fra Styrelsen for IT og Læring)**

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- DageUlovFra (Antal dage med ulovligt fravær)
- Skoleaar (Skoleår)
- Maaned (Måned for registrering af fravær)
- DageAktiv (Antal skoledage i måneden)

3.8 **Register 8: T_SCOR_1995_2018 (Tandlægedata, fra Sundhedsdatastyrelsen)**

Følgende variable er anvendt inden for dette register:

- PNR (Personnummer)
- DATO (Dato for undersøgelse)
- GINGIV (Tandkødsbetændelse)
- I ALT_TD (Summen af antallet af tilstedeværende temporære tænder og antallet af frembrudte permanente tænder)
- DEF_T_3 (Antal temporære og permanente tænder med kode 1,2,4,5,6 indberettet DeMFT-index)

Bilag B: Risikovurdering af behandlingen af persondata i forbindelse med forskningsprojekterne ”Beslutningsstøtte ved underretninger” og ”Underretninger i fokus”

1. INDLEDNING

Denne risikovurdering omhandler VIA's og AU's behandling af personoplysninger i forbindelse med forskningsprojekterne ”Beslutningsstøtte ved underretninger” og ”Underretninger i fokus” (begge herefter benævnt projektet).

Nærværende risikovurdering er foretaget med henblik på at fastsætte niveauet for passende tekniske og organisatoriske sikkerhedsforanstaltninger i overensstemmelse med Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 (Databeskyttelsesforordningen), artikel 32.

2. RISIKOVURDERINGEN

Det er på baggrund af nedenstående metode vurderet, at behandlingen af personoplysninger i forbindelse med projektet indebærer en **mellem** risiko for de registrerede, hvis personoplysninger behandles.

		Konsekvens		
		Lav	Mellem	Høj
Sandsynlighed	Lav			X
	Mellem			
	Høj			

Farvenøgle



Lav risiko



Mellem risiko



Høj risiko

3. METODE

Risikovurderingen er foretaget i overensstemmelse med ENISA's Handbook on Security of Personal Data Processing¹ og Datatilsynets Vejledende tekst om risikovurdering af juni 2019². Modellen baserer sig på den *konsekvens*, som et brud på persondatasikkerheden vurderes at have for den registrerede, sammenholdt med den *sandsynlighed*, der vurderes at foreligge for et sådant brud.

VIA's og AU's ledelse har forholdt sig til denne risiko og accepteret den, herunder på baggrund af de fastsatte sikkerhedsforanstaltninger for behandlingen i forbindelse med projektet.

Risikovurderingen er foretaget i overensstemmelse med den matrix, der fremgår af Datatilsynets Vejledende tekst om risikovurdering af juni 2019, hvorefter konsekvens og sandsynlighed tillægges lige stor vægt i vurderingen af den samlede risiko forbundet med behandlingen.

3.1 Konsekvens

Konsekvensvurderingen er foretaget ud fra parametrene *fortrolighed*, *integritet* og *tilgængelighed*. Den samlede konsekvensvurdering baserer sig på det parameter, der indebærer den største konsekvens for den registrerede.

3.1.1 Fortrolighed

Vurderingen af konsekvensen af manglende fortrolighed omfatter, hvilken konsekvens, det har for den registrerede, at persondata vedrørende denne kommer uberettigede til kendskab.

Det er i relation til projektet vurderet, at manglende fortrolighed kan have en **høj** konsekvens for den registrerede.

Vurderingen er baseret på følgende primære forhold:

¹ ENISA's Handbook on Security of Personal Data Processing kan tilgås her <https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing>

² Datatilsynets vejledning kan tilgås her: <https://www.datatilsynet.dk/media/7900/vejledende-tekst-om-risikovurdering.pdf>

I tilfælde af brud på fortroligheden vil uvedkommende kunne få kendskab til de registreredes meget fortrolige oplysninger, herunder de behandlede børn og forældres CPR-numre, sundhedsforhold, evt. kriminelle forhold og evt. overgreb mv. De forventede konsekvenser af et sådan brud er betydelige, idet bruddet kan få vidtrækkende konsekvenser for både børnenes og forældrenes velfærd samt mentale helbred. De registrerede består hovedsageligt af udsatte og potentielt udsatte børn og deres forældre, hvorfor der er et særligt beskyttelseshensyn at varetage i forhold til fortroligheden af deres oplysninger.

3.1.2 *Integritet*

Vurderingen af konsekvensen af manglende integritet omfatter, hvilken konsekvens, det har for den registrerede, at den behandlede persondata om vedkommende ikke er korrekt. Det er på baggrund af nedenstående vurderet, at der er en **mellem** konsekvens for den registrerede forbundet med manglende integritet.

Vurderingen er baseret på følgende primære forhold:

Hvis dataenes integritet bliver kompromitteret, kan det være problematisk, idet Beslutningsstøttens risikoscore direkte afhænger af dataene, herunder at dataene er retvisende. Kompromitterede data vil potentielt kunne påvirke sagsbehandlingen på usaglig vis. Dog er rækkevidden af denne påvirkning begrænset af, at Beslutningsstøtten i Pilot #1 alene blev anvendt efter den almindelige sagsbehandling, dvs. efter, at socialrådgiveren havde taget stilling til sagen på baggrund af de foreliggende oplysninger. I Pilot #2 er rækkevidden ligeledes begrænset af, at socialrådgiveren får et ark over de betydningsfulde variable og Beslutningsstøttens vægtning af disse. I denne forbindelse øges socialrådgiverens mulighed for at gennemskue de anvendte data, herunder hvorvidt de er retvisende eller forkerte. Hvis det viser sig, at der har været brud på dataenes integritet, skal den registrerede ikke foretage sig noget aktivt, da dataene blot kan hentes på ny fra kommunens database. Under hensyntagen til disse faktorer vurderes det, at brud på integriteten vil have en mellem konsekvens for den registrerede.

3.1.3 *Tilgængelighed*

Vurderingen af konsekvensen af manglende tilgængelighed omfatter, hvilken konsekvens, det har for den registrerede, at denne ikke har adgang til de personoplysninger, der behandles i forbindelse med Beslutningsstøtten.

Det er på baggrund af nedenstående vurderet, at der er en **lav** konsekvens for den registrerede forbundet med manglende tilgængelighed af de personoplysninger, der behandles i forbindelse med Beslutningsstøtten.

Vurderingen er baseret på følgende primære forhold:

Hvis Beslutningsstøtten skulle være utilgængelig, har det ikke særlig betydningsfulde konsekvenser, idet Beslutningsstøtten alene bistår sagsbehandlingen. Systemet indgår ikke som en uundværlig del af sagsbehandlingen, og sagsbehandlingen kan således finde sted upåvirket, selvom Beslutningsstøtten skulle være midlertidigt utilgængelig.

3.2 Sandsynlighed

Sandsynlighedsvurderingen er foretaget ud fra parametrene *Netværk og tekniske behandlingsmidler*, *Processer og procedurer relateret til behandlingsaktiviteten*, *Personer involveret i behandlingsaktiviteten*, *Sektorspecifikke trusler og behandlingens omfang*. Den samlede sandsynlighedsvurdering er baseret på helheden af de fire parametre.

3.2.1 *Netværk og tekniske behandlingsmidler*

Det er på baggrund af nedenstående vurderet, at der foreligger en **lav** sandsynlighed for brud på persondatasikkerheden som følge af behandlingens netværk og behandlingsmidler.

Vurderingen er baseret på følgende primære forhold:

Idet Beslutningsstøtten er en cloud-baseret løsning, bliver data transporteret over internettet. Transporten sikres ved at kryptere indholdet. Så snart dataene er modtaget af AU og VIA, bliver internettet ikke længere anvendt. For at tilgå AU's sikre drev skal man være på AU's lokale netværk – enten ved fysisk tilstedeværelse hos AU eller via en VPN-forbindelse med to-faktor autentifikation. På samme måde kan man alene få adgang til VIA's drev ved fysisk tilstedeværelse på VIA's lokation eller gennem VPN. Herudover har VIA og AU opsat adgangskontrol til deres drev, der sikrer, at alene relevante medarbejdere har adgang.

Beslutningsstøtten er eksternt forbundet med samarbejdskommunernes databaser, hvor dataene trækkes over i systemet. Internt er systemet alene forbundet med backupsystemet.

Beslutningsstøtten er designet, implementeret og bliver vedligeholdt i overensstemmelse med best practice på området, og i tråd hermed har udviklerne af systemet haft fokus på overholdelse af databeskyttelsesforordningens principper om "databeskyttelse gennem design" og "databeskyttelse gennem standardindstillinger". Best practice anvendes for at drage nytte af andres erfaringer og for at undgå at begå de samme fejl, som andre har begået. Best practice ændrer sig i takt med lovgivningen på området og den teknologiske udvikling, og det er derfor også en indirekte vej til at sikre en mere sikker og tidssvarende løsning. Som et eksempel på, hvordan best practice er anvendt ved design af Beslutningsstøtten, kan nævnes "Caching". Beslutningsstøtten er således konfigureret, så personoplysninger ikke efterlades i webbrowserens cache eller på harddisken hos brugere af systemet. Dette gælder, uanset om data tilgås via computer, tablet eller telefon. Denne konfiguration er ikke standard for webteknologier, men den er implementeret og testet i Beslutningsstøtten.

Endelig har VIA og AU iværksat tiltag for at hindre, at uautoriserede kan tilgå de lokationer, hvorpå der behandles personoplysninger. For det første behandles personoplysningerne som altovervejende udgangspunkt alene digitalt på henholdsvis VIA's og AU's lukkede servere, som beskrevet ovenfor. For det andet har alene den nuværende projektleder adgang til de fysiske personoplysninger i form af underretninger, der opbevares i et aflåst skab hos VIA.

3.2.2 *Processer og procedurer relateret til behandlingsaktiviteten*

Det er på baggrund af nedenstående vurderet, at der foreligger en **lav** sandsynlighed for brud på persondatasikkerheden som følge af processer og procedurer relateret til behandlingsaktiviteten.

Vurderingen er baseret på følgende primære forhold:

Rolle- og ansvarsfordelingen for databehandlingen er fast defineret i de indgåede samarbejdsaftaler mellem AU og VIA. Derudover finder databehandlingen hos AU og VIA sted under hensyntagen til interne retningslinjer herom, f.eks. AU's "Tjekliste til informationssikkerhed" og VIA's "Informationssikkerhedspolitik". I overensstemmelse med AU's og VIA's interne retningslinjer har alene en meget begræn-

set persongruppe adgang til forskningsprojektets personoplysninger hos VIA, AU og Indiesoft. Adgangsstyringen foretages i denne forbindelse af enkelte kompetente medarbejdere med indgående kendskab til projektet.

Hos AU er det muligt at kopiere data fra det sikre drev til eget eller eksternt udstyr. Dog logges alle hændelser af denne karakter. Derudover slettes data, som undtagelsesvis opbevares på en ekstern enhed (f.eks. et krypteret USB-stik eller SD-kort), så snart formålet med flytningen af dataene er opfyldt. Hos VIA er det alene muligt at kopiere data fra forsknings-pc'erne til et USB-drev, hvis drevet er krypteret. Det skal desuden bemærkes, at der dannes logfiler vedrørende enhver behandling i selve Beslutningsstøtten og vedrørende behandling hos AU og VIA.

3.2.3 *Personer involveret i behandlingsaktiviteten*

Det er på baggrund af nedenstående vurderet, at der foreligger en *lav* sandsynlighed for brud på persondatasikkerheden som følge af de personer, der er involveret i behandlingsaktiviteten.

Vurderingen er baseret på følgende primære forhold:

Personkredsen, der anvender Beslutningsstøtten, er afgrænset til de mest nødvendige medarbejdere hos VIA, AU og Indiesoft, og i samarbejdskommunerne anvender alene udvalgte socialrådgivere værktøjet. De medarbejdere, der behandler Beslutningsstøttens personoplysninger hos VIA, AU og Indiesoft, har omfattende erfaring med at arbejde med informationssikkerhed. VIA IT har i flere år understøttet beskyttelsen af såvel almindelige som følsomme personoplysninger i forbindelse med forskningsprojekter på VIA's servere. AU og AU's medarbejdere har tilsammen flere årtiers erfaring med behandling af personoplysninger i forbindelse med forskningsprojekter. Endelig har René Karberg, der repræsenterer Indiesoft i forbindelse med forskningsprojektet, over 10 års erfaring med at arbejde med informationssikkerhed.

Både AU og VIA har som dataansvarlige betydelig fokus på at lagre og slette projektets personoplysninger på forsvarlig vis. AU opbevarer personoplysningerne fra pilotafprøvningerne m.m. på et sikkert og lukket drev, der befinder sig på AU's servere, der hostes af Global Connect A/S i Skanderborg. Disse servere overholder alle sikkerhedskrav, og der anvendes anerkendt kryptering mellem AU og eksterne leverandører. Der foretages løbende backup af data på serverne. Backups slettes efter 3-6 måneder af-

hængig af logningstypen. Projektets personoplysninger slettes uigenkaldeligt fra serveren 5 år efter indgåelse af aftale om publicering – i overensstemmelse med, hvad der er foreskrevet hos flere videnskabelige tidsskrifter i forbindelse med publicering og evt. verificering.

I forbindelse med afprøvningen i Pilot #1 opbevarede Hjørring Kommune oplysningerne på eget hardware i Hjørring, mens AU's databehandler Indiesoft opbevarede personoplysningerne fra Silkeborg Kommune i deres Microsoft SQL Server database hos Amazon i Irland. Alle dataene blev krypterede under transporten, og navn samt CPR-nummer opbevares ligeledes krypteret i databasen. Så snart pilotafprøvningen blev lukket ned, slettede Indiesoft databasens personoplysninger uigenkaldeligt. Endvidere blev backups af databasen slettet efter 70 dage. Det er endnu ikke fastlagt, hvordan personoplysninger vil blive opbevaret under afprøvningen i Pilot #2.

VIA opbevarer personoplysningerne fra pilotafprøvningerne m.m. på sikrede Bitlocker-krypterede forskningsservere, der alene kan tilgås via en forsker-PC og VPN-adgang. Det er kun forskningsgruppens medlemmer, der har adgang til fælles forskningsdrev – og VIA foretrækker generelt, at personoplysninger opbevares på personlige sikrede drev i videst muligt omfang. Der foretages løbende backups af data på forskningsserverne. Backups slettes senest efter 33 dage. VIA sletter ligeledes projektets personoplysninger uigenkaldeligt efter 5 år af samme årsager som angivet af AU.

3.2.4 Sektorspecifikke trusler og behandlingens omfang

Det er på baggrund af nedenstående vurderet, at der foreligger en *lav* sandsynlighed for brud på persondatasikkerheden.

Vurderingen er baseret på følgende primære forhold:

Undervisnings- og forskningssektoren, som både VIA og AU fungerer indenfor, vurderes overordnet set ikke at være særligt udsat for målrettede cyberangreb. Derudover har hverken forskningscentre ved VIA eller AU været udsat for udefrakommende sikkerhedsbrud indenfor de seneste år. I løbet af afprøvningerne skal Beslutningsstøtten behandle personoplysninger om i alt 400 børn og unge samt deres forældre og søskende, og der er dermed ikke tale om en meget omfattende databehandling. Data anvendt til udvikling af Beslutningsstøtten og evaluering af dens performance var og vil blive pseudonymiseret.

Mitigerende foranstaltninger

Vedr. forskningsprojekterne

Beslutningsstøtte ved underretninger &

Underretninger i Fokus

Marts 2021



INDHOLDSFORTEGNELSE

1.	Indledning.....	1
1.1	Baggrund	1
1.2	Formål.....	2
1.3	De identificerede risici.....	2
2.	Risiko nr. 1: Beslutningsstøtten kan føre til de facto automatiske afgørelser.....	5
2.1	Foranstaltning nr. 1: Uddannelse af socialrådgiverne.....	5
2.2	Foranstaltning nr. 2: Transparens omkring inputdata og vægtning heraf	6
3.	Risiko nr. 2: Manglende forklarlighed og transparens i forhold til Beslutningsstøttens risikoscore	7
3.1	Foranstaltning nr. 1: Transparens omkring inputdata og vægtning heraf	7
3.2	Foranstaltning nr. 2: Dokumentation af processer og datasæt, herunder træningsdata, testdata og algoritmer mv	7
3.2.1	Processer og datasæt vedr. Pilot#1.....	7
3.2.2	Processer og datasæt vedr. Pilot#2	11
4.	Risiko nr. 3: Fejl i modellen (algoritmen)	16
4.1	Foranstaltning nr. 1: Test og validering forud for afprøvning	16
4.2	Foranstaltning nr. 2: Gentræning eller recalibrering af algoritmen	16
4.3	Foranstaltning nr. 3: Transparens omkring inputdata og vægtning heraf	17
4.4	Foranstaltning nr. 4: Kvalificeret personale	17
5.	Risiko nr. 4: Ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias)	18
5.1	Foranstaltning nr. 1: Gennemgående træning af Beslutningsstøtten	18
5.2	Foranstaltning nr. 2: Test og fravalg af variable.....	18
5.3	Foranstaltning nr. 3: Kvalificeret personale	18
6.	Risiko nr. 5: Behandling af for mange eller ikke nødvendige oplysninger.....	19
6.1	Foranstaltning nr. 1: Dokumentation for vurdering af proportionaliteten i databehandlingen	19
6.2	Foranstaltning nr. 2: Dokumentation for modelvalg	26

6.3	Foranstaltning nr. 3: Retningslinjer for pseudonymisering	29
6.4	Foranstaltning nr. 4: Retningslinjer for evaluering af proportionaliteten i databehandlingen	31
7.	Risiko nr. 6: Dataophobning.....	32
7.1	Foranstaltning nr. 1: Overordnede retningslinjer for sletning	32
8.	Risiko nr. 7: Uautoriseret adgang til oplysningerne	33
8.1	Foranstaltning nr. 1: Bruger og adgangsstyring.....	33
8.2	Foranstaltning nr. 2: Logning	34
8.3	Foranstaltning nr. 3: Kryptering	35
9.	Risiko nr. 8: Manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset	36
9.1	Foranstaltning nr. 1: Fastlæggelse af dataansvar	36
9.2	Foranstaltning nr. 2: Indgåelse af de nødvendige aftaler	36
10.	Litteratur	37

1. Indledning

1.1 Baggrund

VIA University College (herefter VIA) og TrygFondens Børneforskningscenter, Aarhus Universitet (herefter AU), igangsatte i 2017 forskningsprojekterne "Beslutningsstøtte ved underretninger" og i 2019 "Underretninger i fokus". Et af de overordnede formål med projekterne er at udvikle et beslutningsstøtteredskab (herefter Beslutningsstøtten), der kan bistå kommunernes socialfaglige praksis, herunder socialrådgivere, med at vurdere underretninger om børn og unges mistrivsel. Tanken er, at Beslutningsstøtten skal understøtte socialrådgiverens sagsbehandling ved – på baggrund af de tilgængelige oplysninger i sagen – at udregne en konkret risikoscore, der indikerer risikoen for et barns eller en ung persons mistrivsel. Børn og unges mistrivsel er imidlertid et komplekst mål, som ikke direkte kan observeres i de kommunale systemer eller i nationale registre. Vi har derfor været nødt til at støtte os til en indikator, der kan hjælpe os med at identificere børn og unge, der er i risiko for at mistrives. I projektet "Beslutningsstøtte ved underretninger" anvendte vi et samlet mål for mistrivsel, som vi redegør for i afsnit 3.2.1 nedenfor. I projektet "Underretninger i fokus" har vi valgt anbringelse som indikator. Det redegør vi for i afsnit 3.2.2.

Risikoscoren har til formål at skærpe socialrådgiverens beslutningsgrundlag, og den må alene anvendes som beslutningsstøtte. I løbet af forskningsprojektet skal Beslutningsstøtten afprøves med fokus på redskabets præcision, anvendelighed, påvirkningskraft samt betydning for socialrådgivernes arbejde.¹

Som en del af projektet har AU og VIA i samarbejde med Poul Schmith udarbejdet en risikovurdering og konsekvensanalyse efter GDPR artikel 32 og 35 for forskningsprojekterne *Beslutningsstøtte ved underretninger* og *Underretninger i Fokus*. Risikovurderingen og konsekvensanalysen kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#). Vi har i arbejdet med konsekvensanalysen identificeret en række risici, som vi, sammen med Poul Schmith, vurderer kan nedbringes gennem mitigerende foranstaltninger, så behandlingen i forskningsprojekterne ikke vil medføre en høj risiko for de registrerede.

Den overordnede konklusion på konsekvensanalysen er, at hvis de identificerede mitigerende foranstaltninger gennemføres, vil residualrisikoen ved behandling af personoplysningerne i projektet være lav til mellem med overvægt til lav risiko for de registrerede.

¹ Projektet er beskrevet yderligere i en kort projektbeskrivelse, som kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#) (TrygFondens Børneforskningscenter, 2021a).

Behandlingen af personoplysninger i projekterne, og forskellene dem imellem, er beskrevet systematisk i den udarbejdede konsekvensanalyse (Poul Schmith, VIA University College og Aarhus Universitet, 2021), og vil derfor ikke blive behandlet i dette notat.

1.2 Formål

Formålet med notatet er at adressere de risici, vi identificerede i konsekvensanalysen vedrørende databehandling og modelvalg, med mitigerende foranstaltninger med henblik på at reducere disse risici.

1.3 De identificerede risici

I konsekvensanalysen blev der identificeret i alt otte risici i behandlingen. Det drejer sig om:

- 1) Beslutningsstøtten kan føre til de facto fuldautomatiske afgørelser
- 2) Manglende forklarlighed og transparens i forhold til Beslutningsstøttens risikoscore.
- 3) Fejl i modellen (algoritmen)
- 4) Ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias).
- 5) Behandling af for mange eller ikke nødvendige oplysninger
- 6) Dataophobning
- 7) Uautoriseret adgang til oplysningerne
- 8) Manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset

De identificerede risici vil blive nedbragt ved følgende mitigerende foranstaltninger angivet i Tabel 1 nedenfor. De enkelte risici og foranstaltninger udfoldes enkeltvis i notatet, og gennemgangen vil følge rækkefølgen i tabellen. Foranstaltninger markeret med stjerne * adresseres også i konsekvensanalysen, og store dele af disse afsnit eller hele afsnittet er således direkte kopieret fra konsekvensanalysen.

Tabel 1: Risici og mitigerende foranstaltninger

Risiko	Mitigerende foranstaltninger	Hvor de adresseres
Risiko nr. 1 <i>Beslutningsstøtten kan føre til de facto fuldautomatiske afgørelser</i>	Foranstaltning nr. 1 Uddannelse af socialrådgiverne Foranstaltning nr. 2 Transparens omkring inputdata og vægtning heraf	Adresseres i afsnit 2.1* Adresseres i afsnit 2.2*
Risiko nr. 2 <i>Risiko for manglende forklarlighed og transparens i forhold til Beslutningsstøttens risikoscore og datagrundlag</i>	Foranstaltning nr. 1 Transparens omkring Beslutningsgrundlag og inputdata og vægtning heraf Foranstaltning nr. 2 Dokumentation af processer og datasæt, herunder træningsdata, testdata og algoritmer mv.	Adresseres i afsnit 3.1* Adresseres i afsnit 3.2
Risiko nr. 3 <i>Fejl i modellen (algoritmen)</i>	Foranstaltning nr. 1 Test og validering forud for afprøvning Foranstaltning nr. 2 Gentræning eller rekalkibrering af algoritmen Foranstaltning nr. 3 Transparens omkring inputdata og vægtning heraf Foranstaltning nr. 4 Kvalificeret personale	Adresseres i afsnit 4.1* Adresseres i afsnit 4.2* Adresseres i afsnit 4.3* Adresseres i afsnit 4.4*
Risiko nr. 4 <i>Ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias)</i>	Foranstaltning nr. 1 Gennemgående træning af Beslutningsstøtten Foranstaltning nr. 2 Test og fravalg af variable Foranstaltning nr. 3 Kvalificeret personale	Adresseres i afsnit 5.1* Adresseres i afsnit 5.2* Adresseres i afsnit 5.3*
Risiko nr. 5 <i>Risiko for behandling af for mange eller ikke nødvendige oplysninger</i>	Foranstaltning nr. 1. Dokumentation for vurdering af proportionaliteten i databehandlingen. Foranstaltning nr. 2. Dokumentation for modelvalg Foranstaltning nr. 3. Retningslinjer for pseudonymisering Foranstaltning nr. 4. Retningslinjer for periodisk evaluering af proportionaliteten	Adresseres i afsnit 6.1 Adresseres i afsnit 6.2 Adresseres i afsnit 6.3 Adresseres i afsnit 6.4

Risiko	Mitigerende foranstaltninger	Hvor de adresseres
Risiko nr. 6 Dataophobning	Foranstaltning nr. 1 Overordnede retningslinjer for sletning	Adresseres i afsnit 7.1*
Risiko nr. 7 Uautoriseret adgang til oplysningerne	Foranstaltning nr. 1 Bruger- og adgangsstyring	Adresseres i afsnit 8.1*
	Foranstaltning nr. 2 Logning	Adresseres i afsnit 8.2*
	Foranstaltning nr. 3 Kryptering	Adresseres i afsnit 8.3*
Risiko nr. 8 Manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset	Foranstaltning nr. 1 Fastlæggelse af dataansvar	Adresseres i afsnit 9.1*
	Foranstaltning nr. 2 Indgåelse af de nødvendige aftaler	Adresseres i afsnit 9.2*

2. Risiko nr. 1: Beslutningsstøtten kan føre til de facto automatiske afgørelser

Beslutningsstøtten er alene et beslutningsstøtteredskab, der skal understøtte socialrådgiverens beslutning om, hvordan en underretning om et barns eller den unges mistvivsel skal håndteres. Beslutningsstøtten må under ingen omstændigheder anvendes til de facto at træffe fuldautomatiske afgørelser. Der kan dog være risiko for, at socialrådgiverne ikke foretager en reel vurdering og efterprøvelse af Beslutningsstøttens risikoscore, dvs. at socialrådgiverne ukritisk lægger Beslutningsstøttens risikoscore til grund ("rubber stamping"), og herved ikke selvstændigt afgør, hvordan underretningen skal håndteres. Dette kan have flere årsager, f.eks. at socialrådgiveren ikke forstår, hvordan Beslutningsstøtten er nået frem til risikoscoren (manglende transparens). Den manglende efterprøvelse kan også skyldes, at socialrådgiveren stoler på Beslutningsstøtten og dermed stopper med at bruge sin egen kritiske sans og undlader at overveje, om risikoscoren kan være forkert ("automation bias"). Risikoen for de registrerede vil her være, at der ikke foretages saglig sagsbehandling, herunder at der ikke udøves de fornødne tiltag til beskyttelse af et barn eller ung i mistvivsel, eller at der benyttes for indgribende foranstaltninger i forhold til disse.

2.1 Foranstaltning nr. 1: Uddannelse af socialrådgiverne

VIA og AU har og vil fortsat reducere denne risiko ved at uddanne socialrådgiverne med henblik på at sikre, at de forstår, hvordan Beslutningsstøtten fungerer og kommer frem til risikoscoren, herunder Beslutningsstøttens begrænsninger samt give socialrådgiverne klare kriterier for, hvornår de forventes at se bort fra Beslutningsstøttens risikoscore.

VIA og AU uddannede forud for afprøvningen i Pilot # 1 socialrådgiverne i, at Beslutningsstøtten aldrig måtte stå alene, men alene skulle indgå i den socialfaglige vurdering sammen med de andre elementer, som socialrådgiveren almindeligvis baserer sin vurdering på, herunder lovgivning, teoretisk og forskningsbaseret viden, erfaring samt den lokale kontekst for arbejdet. VIA og AU uddannede desuden socialrådgiverne i Beslutningsstøttens begrænsninger, herunder at Beslutningsstøtten ikke har adgang til den samme mængde informationer som socialrådgiverne, der har adgang til flere informationer via prosateksten i journalen mv. Socialrådgiverne og deres ledere blev undervist over to gange, hvor anden undervisningsgang var en konkret træning i brug af modellen på pseudonymiserede underretninger. Dette med henblik på konkret oplæring i brugen af Beslutningsstøtten, herunder hvordan Beslutningsstøtten måtte benyttes i socialrådgiverens beslutningsproces. Det bemærkes, at Beslutningsstøtten i Pilot #1 først blev anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen.

VIA og AU vil forud for afprøvningen i Pilot #2 uddanne socialrådgiverne og deres ledere på samme vis som i Pilot #1. Derudover vil VIA og AU bistå kommunerne med at etablere et organisatorisk set-up, hvor socialrådgiverne løbende sikres faglig ledelsesmæssig understøttelse. Dette f.eks. ved at teamlederen i projektperioden løbende indgår i dialog med socialrådgiverne vedrørende anvendelse af redskabet i konkrete sager eller løbende tager stikprøver ud til fælles faglig sparring.

2.2 Foranstaltning nr. 2: Transparens omkring inputdata og vægtning heraf

VIA og AU vil desuden reducere risikoen for automatiske afgørelser ved hjælp af transparens, herunder ved at sikre at datagrundlaget og Beslutningsstøttens vægtning af data er tilgængeligt for socialrådgiveren.

VIA og AU har besluttet at øge informationsniveauet til socialrådgiverne under Pilot #2 for at skabe transparens omkring, hvilke oplysninger Beslutningsstøtten har inddraget ved udregningen af risikoscoren, og hvordan disse oplysninger er blevet vægtet. Dette gøres ved at 1) implementere en dataoversigt, der opsummerer alle data, som Beslutningsstøtten har anvendt i beregningen af risikoscoren, og 2) præsentere socialrådgiverne for variabelindeks, der giver dem indblik i, hvilke typer af informationer, der har trukket scoren op/ned. Denne transparens vil blandt andet gøre socialrådgiveren i stand til at vurdere, om der er fejl i data, eller om socialrådgiveren har adgang til væsentlige oplysninger i sagen, som ikke indgår i datagrundlaget for risikoscoren.

Det bemærkes, at socialrådgiverne i forbindelse med Pilot #1 ikke fik en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes at ske i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke fik forklaret risikoscoren.

Hvis evalueringen af Pilot #2 peger på, at Beslutningsstøttens risikoscores ikke er tilstrækkeligt transparente og forklarlige, vil VIA og AU om nødvendigt re-designe Beslutningsstøttens brugerflade forud for RCT/lodtrækningsforsøget.

3. Risiko nr. 2: Manglende forklarlighed og transparens i forhold til Beslutningsstøttens risikoscore

Det er nødvendigt, at VIA og AU kan forklare, hvordan Beslutningsstøtten når frem til sine risikoscorer (outputs), når Beslutningsstøtten anvendes som sagsunderstøttende værktøj. Dette indebærer, at VIA og AU skal kunne forklare Beslutningsstøttens bagvedliggende logik samt at der skabes transparens omkring datagrundlaget, som en konkret score er baseret på.

Hvis VIA og AU ikke er i stand til at forklare, hvordan Beslutningsstøtten virker og kvaliteten af Beslutningsstøttens output samt rationalet bag dette, skaber det risiko for, at fejl ved den foreslåede risikoscore ikke opdages. Risikoen for de registrerede vil her også være, at der ikke foretages saglig sagsbehandling, herunder at der ikke udøves de nødvendige tiltag til beskyttelse af et barn eller ung i mistrivsel, eller at der benyttes indgreb, hvor dette ikke var berettiget, herunder at der benyttes for indgribende foranstaltninger i forhold til et barn eller en ung.

3.1 Foranstaltning nr. 1: Transparens omkring inputdata og vægtning heraf

VIA og AU har, som netop redegjort for, besluttet at reducere risikoen ved at øge informationsniveauet til socialrådgiverne under Pilot #2 og herved skabe transparens omkring, hvilke oplysninger, Beslutningsstøtten har inddraget ved udregningen af risikoscoren samt hvordan disse oplysninger er blevet vægtet.

Som netop anført fik socialrådgiverne i forbindelse med Pilot #1 ikke en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik, som det forventes i Pilot #2, derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke fik forklaret risikoscoren.

VIA og AU vil desuden vejlede kommunerne i at journalisere risikoscoren samt grundlaget herfor.

3.2 Foranstaltning nr. 2: Dokumentation af processer og datasæt

3.2.1 Processer og datasæt vedr. Pilot#1

Hvis Beslutningsstøtten skal være en hjælp for socialrådgiverne til gavn for børn, unge og deres familier, er det afgørende, at vi kan forklare, hvordan Beslutningsstøtten når frem til sine risikoscores; herunder Beslutningsstøttens bagvedliggende logik (hvordan virker den?), hvilke data der inddrages, samt kvaliteten af det output, Beslutningsstøtten producerer. Denne transparens omkring Beslutningsstøttens beregnede risikoscore

er vigtig, da der ellers vil være risiko for, at socialrådgiverne ikke opdager fejl i data, fejl i systemet, eller det kan føre til mistolkning af Beslutningsstøttens output. I værste fald vil dette kunne betyde, at der ikke foretages en saglig sagsbehandling, herunder at der ikke udøves de nødvendige tiltag til beskyttelse af et barn eller ung i mistrivsel, eller at der benyttes indgreb, hvor dette ikke var berettiget, herunder at der benyttes for indgribende foranstaltninger.

I dette afsnit beskriver og dokumenterer vi processerne og datasættene anvendt i Pilot#1 med henblik på at øge Beslutningsstøttens transparens. Processerne og datasættene anvendt i Pilot#2 behandles i afsnittet herefter, i afsnit 3.2.2.

Beslutningsstøttens formål

Formålet med pilotprojektet var at udvikle et statistisk redskab til at understøtte socialrådgivernes arbejde med vurdering af underretninger. Ideen var at samle og udnytte de data, kommunerne allerede har adgang til i deres forskellige administrative systemer, og som de har lov til at anvende. Håbet er, at det statistiske redskabs vurdering af barnets risiko for mistrivsel kan understøtte socialrådgivernes arbejde med at identificere hvilke børn og unge, der har behov for særlig støtte, og hvilke der ikke har. Det statistiske redskab vil fungere som et tillæg til den viden, socialrådgiverne allerede besidder.

Formålet med Beslutningsstøtten i Pilot#1 var således at undersøge, om en statistisk model baseret på machine learning kunne udvikles, implementeres og formidles til socialrådgiverne på en måde, så forudsætningerne for at forbedre fagligheden og sagligheden i sagsbehandlingen kunne tilvejebringes.

Data anvendt til udvikling

De overordnede registre og data, som blev anvendt til udviklingen af algoritmen i Pilot#1, fremgår af Tabel 2 i afsnit 6.1. Her er også redegjort for baggrunden for, at netop disse variable er udvalgt. De eksakte variable og registre som indgår i den færdigudviklede algoritme er angivet i tabel 1 i TrygFondens Børneforskningscenter (2020). Algoritmen fremlægges nedenfor.

Data anvendt til træning og validering

Der blev anvendt omkring 120.000 underretninger fra perioden 2014-15 til at træne og validere modellen. 70 procent tilfældigt udvalgte underretninger blev anvendt til træning af modellen, og de resterende 30 procent til validering.

Testede algoritmer

I Pilot#1 afprøvedes kun en Post-LASSO model, idet hensigten i Pilot#1 primært var at teste om det overhovedet i praksis kunne lade sig gøre at udvikle, implementere og formidle modellen. Modellen var trænet til at prædiktere risikoen for, at en eller flere af følgende tre udfald ville indtræffe inden for 180 dage efter modtagelse af nærværende underretning:

- (1) At kommunen modtager en grov underretning² med bekymring om barnet/den unge.
- (2) Om kommunen iværksætter en forebyggende foranstaltning for barnet/den unge.
- (3) Om barnet/den unge bliver anbragt.

Modellernes performance

Der blev ikke publiceret analyser af modellens prædiktive egenskaber i Pilot#1, men de lå typisk med en såkaldt AUC score på omkring 75 procent. Det indebærer, at modellen i 75 procent af tilfældene vil kunne skelne et barn hvor et af de tre ovenstående udfald indtræffer fra et barn, hvor ingen af de tre udfald indtræffer, på baggrund af modellens prædiktioner.

Valg af model og variabelnes vægtning

Den endelige model er beskrevet i TrykFondens Børneforskningscenter (2020). Den gengives herunder. Risikoscoren blev beregnet som:

$$\begin{aligned} \text{RS} = & 0,1175 + 0,0541 * ((\text{Barnets alder}-9,6308)/4,8251) \\ & + 0,0122 * ((\text{Antal tidligere underretninger, seneste 90 dage}-0,4919)/1,0475) \\ & + 0,0472 * ((\text{Antal tidligere underretninger, seneste 180 dage}-0,7724)/1,4455) \\ & + 0,0068 * ((\text{Type 2 underretning: Kriminalitet hos barn/ung}-0,0130)/0,1131) \\ & + 0,0212 * ((\text{Type 7 underretning: Overgreb mod barn/ung f.eks. seksuelt eller voldeligt}-0,0208)/0,1428) \\ & + 0,0245 * ((\text{Type 9 underretning: Misbrug hos forældre}-0,0616)/0,2403) \end{aligned}$$

² Vi har anvendt Danmarks Statistiks operationalisering af en grov underretning, som er en af følgende tre underretningsårsager: overgreb mod barn/ung f.eks. seksuelt eller voldeligt, misbrug hos forældre og kriminalitet hos forældre (Danmarks Statistik, 2018).

$$\begin{aligned}
& + 0,0116 \quad * \quad ((\text{Antal forebyggende foranstaltninger, seneste 180 dage}-0,0306)/0,1790) \\
& - 0,0002 \quad * \quad ((\text{Antal anbringelser uden for hjemmet, seneste år}-0,0251)/0,1777) \\
& + 0,0094 \quad * \quad ((\text{Antal anbringelser uden for hjemmet, seneste fem år}-0,0680)/0,3364)
\end{aligned}$$

Modellen er valgt ud fra de kriterier, at den havde en rimelig forudsigelseskraft, og at den var relativt enkel at implementere i den kommunale forvaltning og formidle til socialrådgiverne.

Omregning til deciler

Den beregnede risikoscore er efterfølgende blevet re-skaleret til en skala fra 1 til 10. Følgende cut-offs er anvendt, idet disse udgør decil-afgrænsningerne i de prædikterede risikoscores i datasættet:

Score = 1,	hvis RS ligger i intervallet	] - ∞; 0.0094122]
Score = 2,	hvis RS ligger i intervallet	] 0.0094122; 0.0420653]
Score = 3,	hvis RS ligger i intervallet	] 0.0420653; 0.0649243]
Score = 4,	hvis RS ligger i intervallet	] 0.0649243; 0.0878977]
Score = 5,	hvis RS ligger i intervallet	] 0.0878977; 0.1103221]
Score = 6,	hvis RS ligger i intervallet	] 0.1103221; 0.1312139]
Score = 7,	hvis RS ligger i intervallet	] 0.1312139; 0.1450047]
Score = 8,	hvis RS ligger i intervallet	] 0.1450047; 0.1863387]
Score = 9,	hvis RS ligger i intervallet	] 0.1863387; 0.2367061]
Score = 10,	hvis RS ligger i intervallet	] 0.2367061; + ∞ [

Nødvendigheden af at bruge individdata

Som det også begrundes i afsnit 6.1 nedenfor, er det ikke muligt at udvikle modellen uden anvendelse af individdata. Afprøvningen i kommunerne måtte også nødvendigvis ske på rigtige sager, da vi havde behov for at teste for mulige fejl, mangler og uhensigtsmæssigheder i modellens implementering i kommunerne³. Fx kunne der opstå fejl ved indlæsning af data fra en ny underretninger, eller ved indhentning af information fra socialforvaltningens databaser, lige som kodningen af variablene i forhold til de tilgængelige data bedst kunne afprøves på 'rigtige' underretninger.

³ Afprøvning foregik efter endt sagsbehandling for ikke at influere herpå.

3.2.2 Processer og datasæt vedr. Pilot#2

Som vi redegjorte for i indledningen til afsnit 3.2.1 er det afgørende, at vi kan forklare, hvordan Beslutningsstøtten når frem til sine risikoscores, hvis Beslutningsstøtten skal være en hjælp for socialrådgiverne og dermed være til gavn for børn, unge og deres familier. I dette afsnit beskriver og dokumenterer vi processerne og datasættene anvendt i Pilot#2 med henblik på at øge Beslutningsstøttens transparens.

Beslutningsstøttens formål

I lighed med det første udførte pilotprojekt er formålet med Pilot#2 at udvikle et statistisk redskab, som kan understøtte socialrådgivere med en systematisk risikovurdering i underretningssager. Beslutningsstøtten anvender administrative data, der knytter sig til de underretningssager, som værktøjet skal anvendes på. Dette er data som socialrådgiverne allerede er i besiddelse af. Værktøjet kan derfor ikke erstatte socialfaglig viden, men blot være et supplement til arbejdet med at skelne mellem de børn og unge, som har brug for særlig støtte, og dem der ikke har.

Håbet med Beslutningsstøtten, der er udviklet til Pilot#2, er at den kan give mere præcise risikovurderinger i underretningssager vedrørende børn og unge. Dette håb er blandt andet velbegrundet i det faktum, at den nye beslutningsstøtte er udviklet på baggrund af et større datamateriale. Antallet af inputvariable er forøget ganske betragteligt, og samtidig kan vi udnytte en længere tidsdimension, da underretningsdataene i de nationale registre hos Danmarks Statistik denne gang er tilgængelig for hele perioden fra april 2014 til december 2018. Desuden er den statistiske metodik blevet forbedret i forhold til den første version af Beslutningsstøtten. Dette kommer til udtryk i anvendelsen af en ny outcomevariabel (anbringelse inden for et år), samt anvendelsen af en ny statistisk model, XGBoost, som i udviklingsfasen har vist sig som værende den mest nøjagtige blandt de testede modeltyper.

Data anvendt til udvikling

Til udvikling af modellen til Pilot#2 er anvendt en række forskellige registre, som befinder sig på en forskningsserver hos Danmarks Statistik. De anvendte registre og tilhørende variable er angivet i Tabel 2 i afsnit 6.1. Som det også var tilfældet for modeludviklingen til Pilot#1 anvendes kun information, som socialrådgiverne allerede har adgang til og som lovligt kan anvendes til vurdering af underretninger. Med udgangspunkt i de oplistede variabler i Tabel 2 har vi konstrueret en lang række af forklarende variable, som udgør input til den prædiktive risikomodel. Helt konkret har vi konstrueret 126 forklarende variable, der udgør den informationsmængde, som modellen har til rådighed. Overordnet set kan disse variable inddeles i fem kategorier:

1. Basal information vedrørende det barn, underretningen vedrører (alder og antal flytninger)
2. Information vedrørende den modtagne underretning

3. Underretnings- og anbringelseshistorik samt information om tidligere og nuværende forebyggende foranstaltninger for det underrettede barn.
4. Underretnings- og anbringelseshistorik samt information om tidligere og nuværende forebyggende foranstaltninger for det underrettede barns søskende.
5. Basal information vedrørende forældrene til det barn, underretningen vedrører (eksempelvis: alder ved barnets fødsel, civilstand og antal børn)

Data anvendt til træning og validering

Til at udvikle den nye version af Beslutningsstøtten har vi anvendt 173.044 underretningssager om 90.644 forskellige børn og unge modtaget i perioden fra og med april 2016 til og med december 2017. Ud af disse underretninger er 120.395 (ca. 70%) blevet brugt til at oplære modellen, mens de resterende 52.649 underretninger (ca. 30%) er blevet brugt til at evaluere modellens styrke i forhold til at forudsige fremtidige anbringelser. Det faktum, at vi betragter perioden fra april 2016 til december 2017, gør os i stand til at have en komplet underretningshistorik på minimum to år i alle sagerne, da det historiske datagrundlag for underretninger går tilbage til april 2014. Dette er den primære årsag til at modellen er oplært på baggrund af underretninger for netop denne periode.

Som nævnt er det overordnede formål med Beslutningsstøtten at identificere mistrivsel blandt underrettede børn og unge. Den statistiske model, der danner grundlag for Beslutningsstøtten, er designet til at forudsige fremtidige anbringelser. Som det diskuteres i afsnit 6.1 drejer mistrivsel sig om meget mere end hvorvidt et barn eller ung anbringes uden for hjemmet. For at undersøge, hvorvidt der er en sammenhæng mellem Beslutningsstøttens beregnede risikoscores og andre alternative mål for mistrivsel, har vi foretaget en ekstern validering af modellen ved brug af de registre og data, som er opstillet i Tabel 3 i afsnit 6.1. Som det fremgår, er der til den eksterne validering anvendt data, der knyttet sig til kriminalitet, fysisk og mental sundhed, fraværs- og trivselsdata fra den danske folkeskole samt tandlægedata fra Sundhedsdatastyrelsen. Disse data er *udelukkende* anvendt til evaluering af modellen, og altså *ikke* anvendt i oplæringsfasen af modellen. Det vil altså sige, at modellens output ikke afhænger af disse data. De førnævnte data er udelukkende anvendt til at etablere, at der er en sammenhæng mellem Beslutningsstøttens risikoscores og mistrivsel i en mere bred forstand.

Testede algoritmer og modelegenskaber

Som det fremgår af afsnit 6.2, har vi i forbindelse med forberedelsen af Pilot#2 testet fire forskellige statistiske modeller. De fire modeller er (listet efter kompleksitet) den lineære sandsynlighedsmodel, en logistisk regressions model med variabel udvælgelse (LASSO), en random forest model og en XGBoost model. Evalueringen af disse fire modeltyper viser, at XGBoost algoritmen opnår den højeste AUC-værdi, hvilket er den metrik, som vi har valgt at anvende til at vurdere modellernes prædiktive egenskaber. AUC-værdien for XGBoost modellen er på 84,08%.

Valg af model

Modellen der danner grundlag for Beslutningsstøtten i Pilot#2 er machine learning algoritmen XGBoost, som beskrives yderligere i afsnit 6.2 nedenfor. Den primære årsag til valget af netop denne model, er, at denne model opnår den højeste AUC-værdi blandt de testede algoritmer (se Tabel 4, afsnit 6.2). Da de testede modeller som udgangspunkt anvender de nøjagtigt samme inputvariable, er det naturligt at anvende den modeltype, som er bedst til at forudsige mistrivsel. Intuitivt kan XGBoost metodens gode egenskaber i denne sammenhæng forstås ved, at denne algoritme er specielt god til at modellere komplekse ikke-lineære relationer i data. Da en beslutning om en anbringelse uden for hjemmet typisk er et udfald, som skyldes et samspil mellem mange forskellige faktorer, er det ikke overraskende, at XGBoost modellens forudsigelser viser sig som værende de mest præcise.

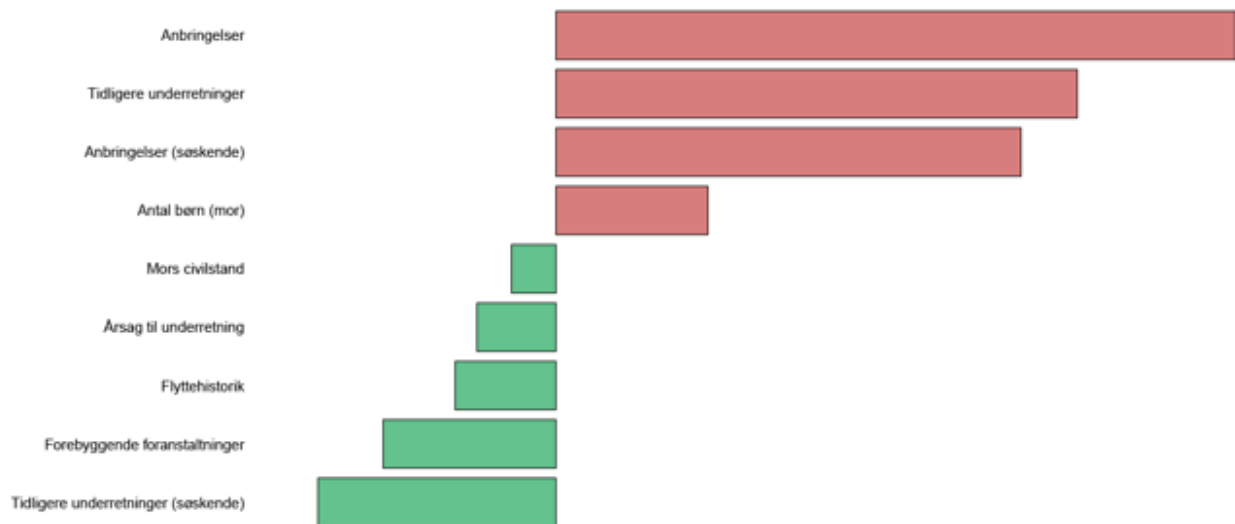
For at sikre at Beslutningsstøtten er så "fair" som muligt, er det valgt, at modellen ikke skal have kendskab til den etniske oprindelse af de børn og de familier, som værktøjet bruges på. Ved at udelade denne type information fra modellen, sikres det, at en given risikovurdering fra Beslutningsstøttens side ikke kan tilskrives barnets etniske oprindelse. På tilsvarende vis indgår der ikke direkte informationer i modellen, der indikerer barnets/den unges køn eller familiens socioøkonomiske status.

Derudover har vi også undersøgt sammenhængen mellem Beslutningsstøttens risikovurderinger og de faktiske anbringelsesrater i vores historiske datamateriale. Disse analyser viser, at der kun er lille systematisk forskel mellem anbringelsesraterne for børn af vestlig og ikke-vestlig oprindelse, når man har taget højde for hvilken risikoscore værktøjet har tildelt børnene. Det vil sige, at et barn af vestlig oprindelse med en given risikoscore har lige så stor sandsynlighed for at blive anbragt som et barn af ikke-vestlig oprindelse, der også har fået tildelt den samme risikoscore. Derimod viser vores post-analyse, at børn med lav socioøkonomisk status har højere anbringelsesrater end børn med høj socioøkonomisk status, selv når man har taget højde for Beslutningsstøttens risikovurdering. Algoritmen indeholder ingen informationer om socioøkonomisk status, og skelner således ikke mellem børn fra familier med lav eller høj socioøkonomisk status. Det betyder, at hvis alt andet er lige, vil børnene/de unge få beregnet samme risikoscore.

Variablenes vægtning

En af ulemperne ved at anvende XGBoost metodikken er, at det er vanskeligt at kvantificere, hvorledes de enkelte inputvariable har bidraget til den endelige prædiction. Dette er en vigtig faktor, for at sådanne værktøjer kan blive en succes i praksis. Til at "forklare" hvordan de enkelte variable vægtes af modellen, påtænker vi i Pilot#2 at anvende såkaldte SHAP-værdier (Lundberg og Lee, 2017). I Figur 1 nedenfor vises et eksempel på hvordan SHAP-værdier kan forklare hvilke typer af information, der har trukket modellens prædiction henholdsvis op og ned. Dette skaber en øget transparens omkring Beslutningsstøtten, da den enkelte socialrådgiver

direkte får kendskab til, hvad værktøjet har identificeret som de væsentligste risikofaktorer for hvert enkelt underretningssag.



Figur 1: SHAP værdier for en enkelt modelprædiktion.

Omregning til deciler

Baseret på XGBoost modellens prædikterede anbringelsessandsynligheder beregnes en heltallig risikoscore der varierer mellem 1 og 10. Risikoscoren udregnes ved at betragte decil-afgrænsningerne af anbringelsessandsynlighederne inden for hvert alderstrin. Eksempelvis vil alle underretninger blandt tiårige, hvis prædikterede sandsynlighed ligger under den første decil-grænse for tiårige blive tildelt en risikoscore på 1, og på tilsvarende vis vil en underretning der tilhører den tiende decil-gruppe blandt tiårige få tildelt den højeste risikovurdering på 10. Ved at bruge aldersspecifikke decil-afgrænsninger sikres, at risikoscoren ikke bliver aldersafhængig på samme måde, som det var tilfældet i Pilot#1. Der vil, med denne nye metode, være den samme andel, der får tildelt en given risikoscore inden for hver aldersgruppe. Som et eksempel på udregningen af risikoscores betragtes de tiårige. For dette alderstrin er følgende cut-off værdi anvendt til omregningen mellem sandsynligheder og risikoscores:⁴

⁴De øvrige cut-off værdier der er benyttet til omregning fra prædikterede sandsynligheder til risikoscore for andre aldersgrupper, kan rekvireres ved henvendelse til forskergruppen.

Score = 1, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet [0; 0.00624]
Score = 2, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.00624; 0.00817]
Score = 3, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.00817; 0.01104]
Score = 4, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.01104; 0.01405]
Score = 5, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.01405; 0.01977]
Score = 6, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.01977; 0.02674]
Score = 7, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.02674; 0.03906]
Score = 8, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.03906; 0.05952]
Score = 9, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.05952; 0.10454]
Score = 10, hvis den præsikterede anbringelsessandsynlighed tilhører intervallet] 0.10454; 1]

Nødvendigheden af at bruge individdata

Af nøjagtig de samme årsager som argumenteret for i afsnit 3.2.1 og 6.1, er det en forudsætning for at kunne udvikle Beslutningsstøtten og den efterfølgende implementering i pilotkommunen, at der anvendes individdata i Pilot#2.

4. Risiko nr. 3: Fejl i modellen (algoritmen)

Hvis den algoritme, som Beslutningsstøtten baserer sig på, er fejlbehæftet, vil resultaterne fra Beslutningsstøtten (output) ligeledes være fejlbehæftede. Hvis Beslutningsstøttens algoritme er fejlbehæftet, er der risiko for usaglig påvirkning af socialrådgiverens sagsbehandling.

4.1 Foranstaltning nr. 1: Test og validering forud for afprøvning

Risikoen for fejl i modellen er reduceret ved følgende test og valideringer:

Forud for udviklingen af modellerne gennemførte VIA et kvalitativt studie af alle underretninger modtaget i Silkeborg Kommune og Hjørring Kommune i perioden januar til februar 2018. Formålet hermed var at konkretisere, hvilken beslutningsproces, Beslutningsstøtten skulle bistå med at skærpe. AU har forud for Pilot#2, som et led i valideringen af modellernes egnethed til at forudsige mistrivsel, desuden gennemført en række analyser med henblik på at sikre, at modellerne er følsomme over for andre saglige mål for mistrivsel, herunder f.eks. fremtidige sigtelser for kriminalitet og fremtidige somatiske diagnoser.

Beslutningsstøtten blev forud for Pilot #1 desuden testet på en række konkrete underretningssager i kommunerne med henblik på at sikre, at algoritmen inddrog de rigtige data og i øvrigt fungerede som tiltænkt. AU's databehandler, Indiesoft, foretog desuden et dataudtræk efter igangsættelsen af pilotafprøvningen, hvorefter AU tjekkede, at risikoen blev beregnet rigtigt. AU vil foretage samme tests forud for og under afprøvningen i Pilot #2. AU vil desuden forud for afprøvningen validere algoritmens performance på baggrund af registerdata fra Danmarks Statistik, Styrelsen for IT og Læring og Sundhedsdatastyrelsen.

AU vil desuden løbende tage stikprøver, hvor det undersøges, om Beslutningsstøtten beregner scoren rigtigt.

Det forhold, at modellerne er statiske, afhjælper desuden risikoen for, at der opstår utilsigtede fejl under afprøvningen af Beslutningsstøtten.

4.2 Foranstaltning nr. 2: Gentræning eller rekalkibrering af algoritmen

Risikoen for fejl er desuden reduceret ved gentræning og rekalkibrering af algoritmen.

VIA og AU valgte på baggrund af erfaringerne fra Pilot #1 således at reducere "mistrivsel" til ét enkelt outcome-mål samt at udvide antallet af variable. Dette med henblik på at gøre algoritmen i stand til at skabe mere præcise prædiktioner, der giver et mere dækkende helhedsbillede for de kommunale socialrådgivere. Algoritmen skal i overensstemmelse hermed trænes på baggrund af en væsentlig større datamængde, end det var tilfældet forud for Pilot #1.

VIA og AU vurderede, at det ikke var nødvendigt at foretage gentræning eller rekalkibrering af algoritmen under Pilot #1, idet denne afprøvning alene forløb over en periode på fire måneder. Det er ligeledes VIAs og AU's vurdering, at det ikke er nødvendigt at gennemføre gentræning eller rekalkibrering under Pilot #2, idet denne afprøvning alene vil forløbe over 1-3 måneder. Modellen vil dog blive gentrænet eller rekalkibreret forud for RTC/lodtrækningsforsøget, hvis afprøvningen i Pilot #2 viser et behov herfor. VIA og AU vil desuden løbende analysere og overveje, om modellen kan tilpasses/og eller forbedres i lyset af nye metoder og ny viden på området.

4.3 Foranstaltning nr. 3: Transparens omkring inputdata og vægtningheraf

VIA og AU vil, som beskrevet i de to forudgående afsnit, øge informationsniveauet til socialrådgiverne under Pilot #2, og herved give disse indsigt i hvilke oplysninger, Beslutningsstøtten har inddraget ved udregningen af risikoscoren samt hvordan disse oplysninger er blevet vægtet. Det er VIAs og AU's vurdering, at denne transparens vil øge sandsynligheden for, at fejl i algoritmen opdages, og derfor vil reducere risikoen for en usaglig påvirkning af socialrådgiverne.

Som tidligere beskrevet fik Socialrådgiverne i Pilot #1 ikke en nærmere forklaring på Beslutningsstøttens risikoscore. Omvendt blev Beslutningsstøtten først anvendt efter endt sagsbehandling, og Beslutningsstøtten indgik derved ikke som en aktiv del af sagsbehandlingen. Det synes derfor mindre risikobetonet, at socialrådgiverne i forbindelse med Pilot #1 ikke blev givet samme indsigt i Beslutningsstøttens datagrundlag mv.

4.4 Foranstaltning nr. 4: Kvalificeret personale

VIA og AU har og vil desuden fortsat reducere risikoen for fejl i algoritmen ved at anvende fagligt kvalificeret personale til udvikling af de statistiske modeller. VIA og AU har i denne forbindelse blandt andet sikret, at den primære udvikler har et indgående statistisk kendskab, hvilket både er opnået igennem uddannelse og erfaring. Derudover indgår der i projektet også andre medarbejdere med dybt statistisk kendskab. Endelig er risikoen reduceret ved at sende modelbeskrivelsen – den videnskabelige artikel – i eksternt review hos to anerkendte forskere på området.

5. Risiko nr. 4: Ulovlig forskelsbehandling på grund af indlejrede skævheder og fordomme (bias)

Hvis Beslutningsstøttens datasæt – både under udvikling og drift – er præget af historisk (utilsigtet) skævhed, vil systemet med høj sandsynlighed også medføre utilsigtet direkte eller indirekte ulovlig forskelsbehandling. Selvom rådata (inputdata) registreres korrekt, kan datasættet muligvis ikke være fuldt repræsentativt, eller analysen af dataene kan indeholde skjulte skævheder. Ulovlig forskelsbehandling kan også opstå som følge af en forkert udvikling og programmering af algoritmer.

5.1 Foranstaltning nr. 1: Gennemgående træning af Beslutningsstøtten

Risikoen er søgt reduceret ved at træne Beslutningsstøtten på store mængder repræsentative registerdata fra Danmarks Statistik, inden Beslutningsstøtten er taget i anvendelse i praksis, herunder på et stort antal underretningssager.

Forud for afprøvningen og implementeringen af Beslutningsstøtten i Pilot #1 indledte VIA og AU desuden et samarbejde med kommunernes datavarehuse for at sikre, at de udvalgte variable var defineret på samme måde som de data, AU har haft tilgængelige via Danmarks Statistiks registre. AU vil ligeledes validere de udvalgte variable forud for afprøvningen i Pilot #2.

5.2 Foranstaltning nr. 2: Test og fravalg af variable

Risikoen er desuden søgt reduceret ved at sikre, at modellerne ikke giver anledning til prædiktioner baseret på bestemte parametre. Det er for Pilot#2 f.eks. besluttet, at modellen ikke skal medtage oplysninger om køn, etnicitet og socioøkonomisk status for at forhindre, at modellen viderefører en uhensigtsmæssig forskelsbehandling på baggrund af disse parametre.

VIA og AU vil ligeledes sikre, at ovenstående parametre ikke inkluderes i Pilot #2. VIA og AU vil desuden under udviklingen af modellen samt i evalueringen af afprøvningsfasen teste, om modellen uagtet bevæger sig i en retning, som kan tyde på uhensigtsmæssig forskelsbehandling. Idet Beslutningsstøttens bagvedliggende model er statistisk, er det ikke nødvendigt at teste modellen hyppigere, idet den ikke ændrer sig i lyset af evt. nye mønstre i de data, den behandler i praksis.

5.3 Foranstaltning nr. 3: Kvalificeret personale

VIA og AU har og vil som anført i afsnit 4.4 fortsat anvende personale med et indgående statistisk kendskab til udvikling af modellerne, hvilket bistår til at reducere risikoen for at forkert udvikling og programmering fører til ulovlig forskelsbehandling.

6. Risiko nr. 5: Behandling af for mange eller ikke nødvendige oplysninger

Projektet indebærer behandling af blandt andet følsomme personoplysninger samt CPR-numre. Hvis Beslutningsstøtten udvikles, uden at der sker en tilstrækkelig vurdering af, om behandlingen af personoplysningerne til udvikling eller anvendelse af systemet er begrænset til det nødvendige, relevante og tilstrækkelige, medfører dette en risiko for overtrædelse af princippet om dataminimering, og dermed at der i projektet behandles for mange eller ikke nødvendige oplysninger. I dette afsnit adresserer vi derfor proportionaliteten i databehandlingen. I afsnittet fremlægger vi endvidere vores overvejelser omkring, hvorfor brug af fiktive, anonymiserede og aggregerede oplysninger ikke er tilstrækkelige til oplæring, modeludvikling og test. Endvidere fastlægges retningslinjer for den periodiske evaluering af proportionaliteten i projektet.

I konsekvensanalysen er der foretaget en vurdering af behandlingens nødvendighed og proportionalitet, som fastslår, at der er behov for at behandle personoplysningerne (Poul Schmith, VIA University College og Aarhus Universitet, 2021). I notatet *Forvaltningsretlig legalitetskontrol af Beslutningsstøtten* adresserer Poul Schmith desuden variablenes saglige begrundelse (Poul Schmith, 2021). Deres vurdering er, at der i projektet alene inddrages saglige variable.

6.1 Foranstaltning nr. 1: Dokumentation for vurdering af proportionaliteten i databehandlingen

AU har alene og vil forsat kun inddrage personoplysninger af betydning for projektets formål. Vi skelner mellem to forskellige formål ift. udviklingen af den statistiske model, der har været afprøvet i Pilot#1 og som skal afprøves i Pilot#2:

- 1) Modeludvikling og
- 2) Test af modellens prædiktive egenskaber

6.1.1 Modeludvikling

I udviklingen af de statistiske modeller i Pilot#1 og Pilot#2 anvendes alene data, som socialrådgiverne har adgang til inden for egen forvaltning og som bidrager til at forudsige anbringelse og andre mål for mistrivsel. Modellerne er i begge tilfælde udviklet på en forskningsserver hos Danmarks Statistik med pseudonymiserede registerdata.

I estimationen af de statistiske modeller anvendes machine learning teknikker. Princippet i machine learning er, at algoritmen identificerer hvilke variable, der har en betydning for det udfald, vi er interesserede i; i dette tilfælde børn og unges anbringelse uden for hjemmet⁵. I projektet anvender vi derfor de data, som

1. socialrådgiverne har til rådighed, og
2. som er vurderet som lovlige og saglige at behandle i forbindelse med vurdering af underretninger (Poul Schmith, 2021), og
3. som ikke bidrager til at videreføre diskrimination i behandlingen af underretninger (se evt. afsnit 3.2.2 om fravalg af variable).

Der er således en række variable, som er fravalgt i arbejdet med udviklingen af modellerne.

Alle variable anvendt til udvikling af modellerne i Pilot#1 og Pilot#2 er listet i Tabel 2 nedenfor. Af oversigten fremgår det, at vi i Pilot#2 har anvendt flere variable og registre end i Pilot#1. Det skyldes, at vi har ønsket at forbedre modellens prædiktive evner ved at inkludere flere variable. Ved at bruge flere variable i kombination med en ny type af machine learning model har vi således udviklet en mere præcis model, som dermed er bedre til at identificere, hvilke børn der er i risiko for at blive anbragt – og dermed for at mistrives (se Trygfondens Børneforskningscenter, 2021b). Formålet er at udvikle den mest præcise algoritme, baseret på de informationer, der er tilgængelige og lovlige for socialrådgiveren at anvende, og som ikke viderefører historiske forskelle i afgørelser mellem bestemte undergrupper (især køn, alder og etnicitet).

I Pilot#1 blev der anvendt en post-estimationsmetode (Post-LASSO) til at udvælge de ni variable (blandt de variable listet i Tabel 2), der forklarede den største del af variationen i udfaldet. Variable med begrænset forklaringskraft blev således fravalgt for at sikre, at datagrundlaget ikke var større end nødvendigt. Algoritmen udvalgte følgende ni variable: Barnets alder, tre typer af underretninger (BAGGRUND 2, 7 og 9⁶), antal underretninger om samme barn inden for de seneste hhv. 90 og 180 dage, antal forebyggende foranstaltninger inden for de seneste 180 dage samt antal tidligere anbringelser af barnet inden for hhv. det seneste år og seneste 5 år.⁷ Der er således udvalgt et meget begrænset sæt af variable til den endelige model, som blev afprøvet i forbindelse med Pilot#1. De konkrete variable, som indgik i modellen, var således udvalgt af en statistisk algoritme som værende de – blandt de tilgængelige – som forklarede det meste af variationen i den afhængige

⁵ I Pilot#1 blev der anvendt et andet udfaldsmål. Dette blev adresseret i afsnit 3.2.1.

⁶ BAGGRUND 2 Kriminalitet hos barn/ung

BAGGRUND 7 Overgreb mod barn/ung f.eks. seksuelt eller voldeligt

BAGGRUND 9 Misbrug hos forældre

⁷ Den præcise algoritme fremgår af TrygFondens Børneforskningscenter (2020), internt notat, samt af afsnit 3.2.1 i indeværende notat.

variabel. Den statistiske model har samtidigt fravalgt variable, som ikke har haft nogen stor samvariation med anbringelse.

I Pilot#2 anvender vi en anden og mere kompliceret estimationsmetode, XGBoost. Med denne metode anvendes også et større sæt af variable. Når denne metode alligevel er valgt, skyldes det, at den er bedre til at forudsige børn og unges anbringelse sammenlignet med Post-LASSO metoden.

I forbindelse med eventuelle fremtidige opdateringer af den statistiske model vil vi fortsat tilføje eller fjerne variable til modellen baseret på deres evne til at forudsige børns fremtidige anbringelse og/eller mistrivsel.

Tabel 2: Register og variable anvendt til udvikling af modellerne i Pilot#1 og Pilot#2

Register og variable anvendt til udvikling	Indgår i udvikling Pilot#1	Indgår i udvikling Pilot#2
Register 1a: Underret_tryg1415 (svarer til det nuværende ”Børn og unge – Underretninger”) a. PNR (Personnummer) b. BAGGRUND1-BAGGRUND20 (Baggrund 1-20 for underretningen) c. UNDRETDATO (Underretningsdato) d. UNDRETTYPE (Underretningstype)	Ja	Ja
Register 1b: BUU (Børn og unge – Underretninger) a. PNR (Personnummer) b. BAGGRUND1-BAGGRUND20 (Baggrund 1-16 for underretningen) c. KOMNR (Kommunennummer) d. UNDRETDATO (Underretningsdato) e. UNDRETTYPE (Underretningstype)	Nej	Ja
Register 2: BEF (Befolkningen) a. PNR (Personnummer) b. PNRM (Personnummer på mor) c. PNRF (Personnummer på far) d. FOED_DAG (Fødselsdato) e. KOEN (Køn) – Kun Pilot#1 f. KOM (Kommunekode) – Kun Pilot#1 g. IE_TYPE (Indvandrere, efterkommere, personer med dansk oprindelse) – Kun Pilot#1 h. FM_MARK (Forældremarkering) – Kun Pilot#2 i. CIVST (Civilstand) – Kun Pilot#2	Ja	Ja

Registere og variable anvendt til udvikling	Indgår i udvikling Pilot#1	Indgår i udvikling Pilot#2
Register 3: BUFO (Børn og unge forebyggende foranstaltninger) a. PNR (Personnummer) b. PGF (Paragrafkode) c. SAG_VFRA (Startdato for hændelsen) d. SAG_VTIL (Slutdato for hændelsen)	Ja	Ja
Register 4: BUAH (Børn og unge anbragte hændelsesregister) a. PNR (Personnummer) b. EFFEKTDATO (Effektueringsdato) c. HAENDELSE (Hændelse i anbringelsessag) d. NYKOM (Handlekommune)	Ja	Ja
Register 5: HUST (Husstandsoplysninger) a. PNR (Personnummer) b. BOPIKOM (Adressen i kommunen) c. YEAR (Året hvor den pågældende information er registeret)	Ja	Nej
Register 6: DOD (Døde i Danmark) a. PNR (Personnummer) b. DODDATO (Dato for dødsfald)	Nej	Ja
Register 7: FLYT (Flytninger i Danmark) a. PNR (Personnummer) b. HAEND_DATO (Dato for flytningen)	Nej	Ja

6.1.2 Test af modellens prædiktive egenskaber

I forbindelse med udviklingen af den statistiske model til Pilot#2 testes modellens prædiktive egenskaber gennem en ekstern validering af modellen⁸. Her valideres algoritmens egenskaber på baggrund af analyser gennemført med pseudynomiserede registerdata stillet til rådighed under Danmarks Statistiks forskerordning af Danmarks Statistik, Styrelsen for IT og Læring, og Sundhedsdatastyrelsen. Alle analyser er således udført på en sikker forskerserver hos Danmarks Statistik. Variablene og registrene fremgår i Tabel 3 nedenfor. Variablene i tabellen er alene inddraget med det formål at evaluere den statistiske models egenskaber i en ekstern validering, og de listede variable er således *ikke* anvendt til at udvikle modellen, lige som de ikke indgår som input i den algoritme, der testes i Pilot#2.

Evalueringen af modellens prædiktive egenskaber er vigtig, da vi med modellen ønsker at kunne identificere børn og unge, der er risiko for at komme til at mistrives i en sådan grad, at de har behov for særlig støtte. Børn og unges udsathed eller mistrivsel er imidlertid et komplekst mål, som ikke direkte kan observeres i de kommunale systemer eller i nationale registre. Vi har derfor været nødt til at støtte os til en indikator, der kan hjælpe os med at identificere børn og unge, der er i risiko for at mistrives. Den indikator vi har valgt, er som ovenfor nævnt *anbringelse*. Beslutningsstøtten er således trænet til at kunne forudsige barnets/den unges risiko for at blive anbragt inden for det næste år efter modtagelse af en underretning.

Børn og unges udsathed eller mistrivsel er dog et spørgsmål om meget mere end hvorvidt et barn eller en ung anbringes uden for hjemmet. Vi har derfor undersøgt sammenhængen mellem Beslutningsstøttens beregnede risikoscores og en række andre mål for børn og unges mistrivsel og udsathed. Denne analyse foretages med henblik på at undersøge, om vi med den statistiske model rent faktisk prædikerer mistrivsel blandt børn og unge. Denne eksterne validering af Beslutningsstøtten er afgørende for at kunne fastlægge hvorvidt Beslutningsstøtten, som er trænet til at forudsige anbringelse, også er følsom over for andre mål for udsathed og mistrivsel. Til det har vi behov for at trække på data i andre nationale registre, der indeholder variable, der afspejler andre former for mistrivsel f.eks. hvorvidt barnet/den unge har været offer for en straffelovsforbrydelse eller selv er blevet sigtet for en, har ulovligt skolefravær m.v. Den fulde liste af variable og registre er oplistet i Tabel 3 nedenfor.

⁸ Der er ikke gennemført en ekstern validering af modellens performance i forbindelse med Pilot#1.

Tabel 3: Registere og variable anvendt til ekstern validering af modellen i Pilot#2

Registere og variable anvendt til validering i Pilot#2	
Register A: KROF (Ofre for straffelovsforbydelser)	e. PNR (Personnummer) f. YEAR (Året hvor den pågældende information er registeret) g. OFR_GERFRADT (Gerningens startdato) h. OFR_GER7 (Offeret gerningsindholdskode DS)
Register B: KRSI (Kriminalstatistik sigtelser)	a. PNR (Personnummer) b. YEAR (Året hvor den pågældende information er registeret) c. SIG_SIGTDTO (Sigtelsesdato) d. SIG_GER1DTO (Gerningsdag, startdato)
Register C: LPRADMDIAG (Landspatientregisteret)	a. PNR (Personnummer) b. YEAR (Året hvor den pågældende information er registeret) c. C_DIAG
Register D: PSYK_ADM (Landspatientregistret psykiatri - administrative oplysninger)	a. PNR (Personnummer) b. RECNUM (LPR-identnummer) c. D_HENDTO (Henvisningsdato) d. D_INDDTO (Indlæggelsesdato)
Register E: PSYK_DIAG (Landspatientregistret psykiatri - diagnoser)	a. RECNUM (LPR-identnummer) b. C_DIAG (Diagnosekode)
Register F: Trivsel_2015_2018 (De nationale trivselsmålinger, data fra Styrelsen for IT og Læring)	a. PNR (Personnummer) b. AAR (Året hvor den pågældende information er registeret) c. Q1 - Q40

Registre og variable anvendt til validering i Pilot#2

Register G: Fravaer_2011_2018 (Skolefravær, data fra Styrelsen for IT og Læring)

- a. PNR (Personnummer)
- b. DageUlovFra (antal dage med ulovligt fravær)
- c. Skoleaar (skoleår)
- d. Maaned (måned for registrering af fravær)
- e. DageAktiv (Antal skoledage i måneden)

Register H: T_SCOR_1995_2018 (Tandlægedata, fra Sundhedsdatastyrelsen)

- a. PNR (Personnummer)
- b. DATO (Dato for undersøgelse)
- c. FLADE_3_3 (Antal temporære og permanente tandflader med kode 3 – traume)
- d. GINGIV (tandkødsbetændelse)
- e. SVAERZON (højeste værdi af permzon/tempzon)
- f. I ALT_TD (Summen af antallet af tilstedeværende temporære tænder og antallet af frembrudte permanentetænder)
- g. I ALT_FL (Summen af antallet af frembrudte permanente flader og antallet af tilstedeværende temporære flader)
- h. DEF_T_3 (antal temporære og permanente tænder med kode 1,2,4,5,6 indberettet DeMFT-index)

6.1.3 Overvejelser om brug af individdata til modeludvikling og eksternvalidering

Som det fremgår af indledningen til dette afsnit, behandler vi i forskningsprojektet blandt andet følsomme personoplysninger samt CPR-numre. Det er tilfældet både i udviklingen af de statistiske modeller (følsomme personoplysninger), i den eksterne validering af modellen i Pilot#2 (følsomme personoplysninger) samt i pilotafprøvningsne sammen med deltagerkommunerne (følsomme personoplysninger og CPR numre). En måde at minimere risici for de registrerede er at lade være med at anvende individdata, og et naturligt spørgsmål er derfor, hvorfor brug af fiktive, anonymiserede og aggregerede oplysninger ikke er tilstrækkelige til formålet?

I projektet er det fuldstændigt afgørende, at vi har adgang til historiske data, data med et unikt ID og til data på individniveau. Vi anvender historiske data fra *alle* danske underretningssager med henblik på at identificere, hvilke typer af information, der historisk har kunnet bidrage til at forklare, hvorfor nogle børn og unge bliver anbragt uden for hjemmet efter kommunen har modtaget en underretning; og hvorfor andre ikke gør. Ved brug af fiktive data ville vi ikke have et empirisk grundlag at identificere disse typer af information ud fra, og vi ville således hverken kunne undersøge, hvilke informationer, der kan prædiktere en fremtidig anbringelse/mistrivsel, ligesom vi med fiktive data heller ikke ville kunne undersøge, hvorvidt vores model rent faktisk kan forudsige fremtidig mistrivsel. Det vil med andre ord ikke være muligt for os at konstruere en prædiktiv risikomodel til vurdering af underretninger med et fiktivt datasæt. Vi har desuden valgt at anvende populationsdata

til udvikling af modellerne, hvilket er med til at reducere den statistiske usikkerhed, der kan være ved alene at basere modeludviklingen på en mindre stikprøve.

Som det fremgår af ovenstående Tabel 2 og 3 anvender vi en række forskellige registre til udvikling og validering af de statistiske modeller. Hver observation i registrene har et unikt ID (pseudonymiseret). Uden dette unikke ID ville det ikke være muligt at koble data på tværs af registrene, og det ville dermed ikke være muligt at udvikle eller evaluere modellerne med informationer fra flere registre. Dette er ganske enkelt ikke muligt. Dataene i registrene er effektivt pseudonymiserede (se afsnit 6.3), og det er således ikke muligt for forskerne at identificere fysiske personer i datasættene (ligesom det heller ikke er formålet).

Det er endvidere ikke tilstrækkeligt at anvende data i aggregeret form, da vi i udviklingen af modellen har behov for den variation, der er i data på individniveau. Hele pointen med redskabet er, at vi gerne vil være i stand til at identificere, hvilke børn og unge der er i risiko for at mistrives og blive anbragt, og hvilke der ikke er. Vi har med andre ord behov for så præcist som muligt at kunne fastslå, hvor på risikoskalaen, den enkelte underretning ligger. Det kræver, at vi kender hele skalaen, når vi skal estimere modellen, og ikke kun skalaens gennemsnit, som vi ville få, hvis vi kun anvendte aggregerede data.

6.2 Foranstaltning nr. 2: Dokumentation for modelvalg

Modelvalg er et andet element, der har betydning for, hvor mange variable, der er nødvendige. Vi vil i relation til Pilot#2 derfor undersøge hvilken statistisk model, der giver den bedst mulige prædiktion baseret på færrest data. Dette arbejde gør vi rede for i dette afsnit.

6.2.1 Modelbeskrivelser

Vi har estimeret fire forskellige modeller til at prædiktere et barn eller en ungs risiko for at mistrives. De fire modeller varierer i graden af kompleksitet og antallet af inputvariable. De fire testede modeller er en lineær sandsynlighedsmodel, en logistisk regressionsmodel med variabeludvælgelse (LASSO), en random forest model samt en XGBoost model.

Fælles for alle fire modeltyper er, at de forsøger at forklare, hvorvidt en underretning fører til en (gen)anbringelse uden for hjemmet i løbet af det kommende år. Det vil med andre ord sige, at den afhængige variabel i den statistiske model er den samme for alle fire modeltyper. Den er givet ved en binær indikatorvariabel, som er lig med et, for de børn som bliver anbragt inden for et år efter modtagelsen af en underretning, og lig med nul for alle andre underretningssager. Som nævnt indledningsvist i dette afsnit, er modellens formål at identificere børn og unge som er i risiko for mistrivsel. Da mistrivsel ikke er direkte observerbart i de administrative registre vi har til rådighed, har vi valgt at bruge anbringelser uden for hjemmet som en såkaldt *proxyvariabel* for mistrivsel. Der er således ikke en 1-1 sammenhæng mellem mistrivsel og anbringelser, men der vil typisk

være en høj grad af mistrivsel blandt de børn, som ender med at blive anbragt. Dette er rationalet for brugen af anbringelser som vores observerbare mistrivselssindikator.

Den lineære sandsynlighedsmodel er den simpleste af de fire metoder, da dette blot er standard lineær regressionsmodel. I dette tilfælde kan alle modellens parametre nemt udregnes ved hjælp af mindste kvadraters metode, og disse parametre vil være nemme at fortolke på. Svaghederne ved denne type model er, at den bygger på antagelse om en lineær sammenhæng mellem alle inputvariable og outcomevariablen, og at der ikke er indbygget variabeludvælgelse i denne metode. Det vil sige, at der ikke automatisk bliver taget højde for mulige interaktionseffekter samt at alle inputvariable vil bidrage til modellens output, selvom nogle af inputvariablene potentielt set ikke er gode prædiktorer for mistrivsel. Endvidere har modellen en række statistiske svagheder, som f.eks. at den principielt kan prædiktere sandsynligheder større end 1 eller mindre end 0.

Den logistiske regressionsmodel i kombination med LASSO muliggør variabeludvælgelse som en del af estimationsproceduren. I dette tilfælde tager man udgangspunkt i en relativt velkendt statistisk model – den logistiske regressionsmodel – og tilføjer en mekanisme til modellen (LASSO), der gør at irrelevante inputvariable ekskluderes fra modellen. Det vil med andre ord sige, at denne models output typisk vil være baseret på færre prædiktorer end det er tilfældet for den lineære sandsynlighedsmodel.

De to sidste modeller, random forest og XGBoost, adskiller sig væsentligt fra de to forrige metoder, da de er baseret på en maskinlæringsmetode kaldet *beslutningstræer*. Ved anvendelse af beslutningstræer lægges der mindre struktur ned over sammenhængen mellem outcome- og inputvariable. Det vil sige, at disse modeller bruger en mere datadrevet tilgang til at bestemme sammenspillet mellem outcome og inputs. Specielt er disse metoder velegnede til at finde komplekse ikke-lineære relationer i data. Den primære forskel mellem random forest og XGBoost metoden består i, hvorledes beslutningstræer anvendes. Random forest er en ensemble-metode, hvor modellens endelige output er givet ved et gennemsnit af outputtet fra en række *uafhængige* beslutningstræer. XGBoost metoden er også en ensemble-metode, men i dette tilfælde er modellens output baseret på en række *afhængige* beslutningstræer. Afhængigheden består i, at for hvert skridt i algoritmen forsøger man at gro et beslutningstræ, som retter op på de fejl, det forrige beslutningstræ lavede. Fælles for begge metoder er, at der til at gro et givet beslutningstræ kun anvendes en delmængde af de tilgængelige inputvariable. Det betyder, at der er en variabeludvælgelsesmekanisme indbygget i disse algoritmer, men at den varierer fra skridt til skridt i algoritmerne.

6.2.2 Modellernes egenskaber

Vi har implementeret og testet de fire modeltyper på baggrund af et datasæt bestående af underretninger modtaget i tidsrummet fra og med april 2016 til og med december 2017. Det rå output fra de statistiske modeller er en prædikeret sandsynlighed for at underretning fører til en anbringelse uden for hjemmet i løbet af det

kommende år. For at vurdere hvilken model der giver de bedste prædiktioner, udregnes AUC (eng: Area Under the Curve) værdier for hver model. En AUC-værdi er et tal mellem 0,5 og 1, hvor højere værdier angiver bedre prædiktioner. En værdi på 1 betyder at modellen ikke har lavet nogle fejlprædiktioner, og altid vil kunne vurdere korrekt. Mere formelt, kan en AUC-værdi fortolkes som sandsynligheden for at en model vil tildele en højere prædikeret sandsynlighed til en underretning som fører til en anbringelse inden for et år, end til en underretning som ikke fører til en anbringelse inden for samme tidsrum. Det vil sige, at jo højere AUC værdi, jo bedre vil modellen være til at vurdere hvilke underretninger som kræver ekstra handling fra myndighedernes side.

Tabel 4: AUC-værdier for de fire testede modeltyper

Metode	AUC (%)	95% konfidensinterval
Lineær sandsynligheds model	82,53	81,86 – 83,21
Logistic regression (LASSO)	82,26	81,59 – 82,93
Random Forest	83,36	82,72 – 84,01
XGBoost	84,08	83,45 – 84,71

Noter: Denne tabel angiver AUC-værdier for fire modeltyper. Outcome: Anbringelse uden for hjemmet inden for et år. Modellerne er udviklet på baggrund af 120.395 underretninger (svarende til 63.303 forskellige børn) modtaget mellem april 2016 og december 2017. Modellernes prædiktive egenskaber er evalueret på baggrund af 52.649 underretninger (svarende til 27.341 forskellige børn) modtaget mellem april 2016 og december 2017.

Det fremgår af Tabel 4, at modellerne opnår AUC-værdier mellem 82,26% og 84,08% på evalueringsdatasættet som består af 52.649 forskellige underretninger. De to træbaserede modeller, random forest og XGBoost, opnår de højeste AUC-værdier, hvilket indikerer, at disse metoder giver de mest nøjagtige forudsigelser vedrørende de analyserede underretninger. Disse to metoder er også de to metoder med størst kompleksitet, og som kræver den største grad af fine-tuning fra en statistisk synsvinkel. Kompleksiteten består blandt andet i, at det ikke umiddelbart er muligt at gennemskue hvorledes prædiktionerne fra modellerne er udregnet, da de som tidligere nævnt er baseret på adskillige forskellige beslutningstræer, hvis individuelle bidrag er vanskelige at kvantificere.

6.2.3 Diskussion af modelvalg

Som det ses af ovenstående, har de fire testede modeller ikke samme prædiktive egenskaber. Dette er til trods for, at de alle fire anvender den samme mængde af information som input. Fra et pragmatisk synspunkt kan man argumentere for at man da bør vælge den mest præcise prædiktive model, da man ellers forspilder chancen for at gøre bedst mulig brug af de tilgængelige data i sagsbehandlingen af underretninger. Dog kan det også være fordelagtigt at undersøge, om man kan opnå den tilnærmelsesvis samme forudsigelseskraft ved brugen af færre data, da det fra en dataetisk synsvinkel er vigtigt at ikke-relevant information ikke indgår i behandlingen af underretninger. I den forbindelse er det blevet undersøgt, om man først kan udvælge de mest relevante variable baseret på den logistiske LASSO regressionsmodel, og derefter anvende den reducerede mængde af variable i en XGBoost model. Analysen viser, at denne to-trins procedure kun forringer AUC-værdien ganske lidt, idet AUC værdien kun forringes med 0,25% ved denne fremgangsmåde. Denne reduktion er ganske lille taget i betragtning af at antallet af variable bliver kraftigt reduceret ved denne procedure. Dog har vi ikke valgt at benytte denne fremgangsmåde, da det kan resultere i en model hvor der eksempelvis bliver taget højde for at et barn bor alene med enlig far, men omvendt ikke tages højde for om et barn bor med en enlig mor. Dette vurderes ikke at være hensigtsmæssigt i praksis, at modellen på denne måde kan blive "asymmetrisk" i forhold til typen af information den har kendskab til.

6.3 Foranstaltning nr. 3: Retningslinjer for pseudonymisering

Som ovenfor nævnt har vi udarbejdet og implementeret retningslinjer for, hvornår data skal pseudonymiseres. I dette afsnit gør vi rede for disse retningslinjer.

Al data indsamlet af AU i projektet kan og vil blive pseudonymiseret senest tre måneder efter vi har iagttaget vores oplysningspligt over for de registrerede. Pseudonymiseringen af data følger to spor, alt efter om data opbevares på AU's forskerserver hos Danmarks Statistik eller på AU's projektfolder på O-drevet⁹.

6.3.1 Data, der uploades på Danmarks Statistiks forskerserver

Alle data der i forbindelse med projektet uploades i projektfolderen på Danmarks Statistiks forskerserver slettes fra AU's lokale projektfolder på O-drevet, så snart data er uploaded på serveren og det er bekræftet, at data er korrekt uploaded. Hvis der er behov for at gemme data lokalt, f.eks. fordi der i forbindelse med afprøvningen er indsamlet tekst, der ikke kan uploades på Danmarks Statistiks forskerserver, pseudonymiseres data efter de retningslinjer, der er angivet i nedenfor.

⁹O-drevet er et internt drev på AU's servere. Adgangen til drevet er begrænset til de mest nødvendige medarbejdere, og der foretages logging. Dette er der redegjort for bl.a. i afsnit 8.1 og 8.2 nedenfor.

Data uploaded til Danmarks Statistiks forskerserver bliver pseudonymiseret gennem følgende retningslinjer (for flere detaljer, se Danmarks Statistik, 2019):

Det fremgår af Danmarks Statistik (2019), at:

”Den enkelte borger og virksomhed har krav på, at data behandles fortroligt. Data afgivet til Danmarks Statistik enten direkte fra borger/virksomhed eller indirekte fra en anden offentlig myndighed må kun anvendes til statistik og analyse, aldrig til administrativ sagsbehandling. Deraf følger også, at Danmarks Statistik ikke kan udlevere oplysninger som direkte gennem personnummer, virksomhedsnummer, adresse mv. kan henføres til en person eller en virksomhed. Reglerne er fastlagt dels i databeskyttelsesforordningen og dels i Danmarks Statistiks datasikkerhedspolitik. Findes på følgende LINK: <http://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang.aspx> Det samme princip følger Danmarks Statistik for såkaldt pseudonymiserede data, dvs. data hvor personnummer, virksomhedsnummer eller anden identifikationsnøgle er helt fjernet eller erstattet med et tilfældigt løbenr. Årsagen er, at data som altovervejende hovedregel er meget detaljerede, og at en identifikation ikke kan udelukkes ved at fjerne personnummeret eller andre nøgler.”

Alle forskere der tilgår data via Danmarks Statistiks forskerplaceringsordning følger Danmarks Statistiks regelsæt vedr. behandling af persondata, herunder at alle persondata er pseudoanonymiseret. [Du kan læse mere om Danmarks Statistiks regelsæt her](#). Vi har derudover foretaget risikovurdering af Danmarks Statistiks forskerplaceringsordning for forskernes behandling af data under Aarhus Universitets dataansvar og har desuden implementeret yderligere organisatoriske og tekniske sikkerhedsforanstaltninger. Disse inkluderer blandt andet pseudoanonymisering, sikkerhedskursus for nye forskere, awareness i GDPR, supervision og kontroller der sikrer at de enkelte forskere behandler data inden for rammerne af gældende tilladelser. For mere detaljeret information henvises til ECON/CIRRAUS fælles retningslinjer. [Du kan læse ECON/CIRRAUS fælles retningslinjer her](#).

For yderligere detaljer kan henvises til Danmarks Statistik (2020).

6.3.2 Data, der opbevares i AU's projektfolder på O-drevet

Alle data, der opbevares i AU's projektfolder på O-drevet pseudonymiseres efter følgende procedure:

- 1) Data sorteres i en tilfældig rækkefølge.
- 2) Observationerne tildeles et løbenummer fra 1 og frem.
- 3) Alle CPR-numre slettes
- 4) Datasættes gennemgås med henblik på at sikre, at ikke andre variable indeholder informationer, der gør det muligt at identificere individer. I så fald slettes disse også.

6.4 Foranstaltning nr. 4: Retningslinjer for evaluering af proportionaliteten i databehandlingen

Som ovenfor nævnt har AU desuden udarbejdet og implementeret retningslinjer for, hvor ofte model og data skal gennemgås med henblik på at vurdere, om data stadig er nødvendige i lyset af formålet. Disse retningslinjer vil vi redegøre for i dette afsnit.

I forbindelse med eventuelle fremtidige opdateringer af den statistiske model vil vi fortsat tilføje eller fjerne variable til modellen baseret på deres evne til at forudsige børns fremtidige anbringelse og/eller mistrivsel og herunder afprøve, om en statistisk variabeludvælgelse kan gennemføres, uden at modellens prædiktioner forringes. Formålet med denne foranstaltning er at sikre, at modellen til enhver tid alene bygger på det nødvendige datagrundlag.

Efter som Pilot#2 kun forventes at have en varighed på en til tre måneder, og lodtrækningsforsøget tre til seks måneder, vil vi ikke gen-evaluere proportionaliteten, mens afprøvningen af Pilot#2 pågår eller mens lodtrækningsforsøget pågår, da det er vigtigt for at kunne evaluere modellen, at det er den samme model, der har været anvendt. Modellen vil blive genestimeret med inklusion af nyeste data for underretninger mellem Pilot#2 og lodtrækningsforsøget. I forbindelse med denne gen-estimation vil vi genevaluere proportionaliteten i databehandlingen.

Hvis modellen efterfølgende skal anvendes i daglig drift vil AU vejlede den ansvarlige organisation om at fastsætte retningslinjer for løbende at evaluere proportionaliteten i databehandlingen.

7. Risiko nr. 6: Dataophobning

VIA og AU behandler personoplysningerne i projektet med henblik på at udvikle et beslutningsstøttende redskab, der er egnet til at skærpe socialrådgivernes beslutningsgrundlag ved underretninger om børn og unges mistrivsel. Hvis VIA og AU fortsætter med at opbevare oplysningerne efter formålets ophør, udgør dette en risiko for krænkelse af de registreredes rettigheder.

7.1 Foranstaltning nr. 1: Overordnede retningslinjer for sletning

Risikoen er reduceret ved, at både VIA og AU sikrer effektiv sletning i overensstemmelse med de retningslinjer og procedurer for sletning, der er angivet i konsekvensanalysens afsnit 2.6 (Poul Schmith, VIA University College og Aarhus Universitet, 2021). VIA og AU vil således senest slette data 5 år efter indgåelse af aftale om publicering, således som det er foreskrevet hos flere videnskabelige tidsskrifter i forbindelse med publicering og eventuel verificering.

De implicerede systemer, hvor personoplysningerne opbevares og behandles, understøtter en effektiv og uigenkaldelig sletning.

AU har i databehandleraftalerne med henholdsvis Indiesoft og Danmarks Statistik desuden betinget sig sletning efter AU's nærmere instruks. AU vil betinge sig samme vilkår under Pilot #2.

8. Risiko nr. 7: Uautoriseret adgang til oplysningerne

I forbindelse med både Pilot #1 og Pilot #2 kan der være en risiko for, at uvedkommende får adgang til de data, der indsamles via Beslutningsstøtten. Dette kan medføre et brud på persondatasikkerheden med potentielle krænkelse af de registreredes rettigheder og frihedsrettigheder til følge. Uvedkommende udgøres i denne henseende både af udefrakommende, der tvinger sig adgang til oplysningerne, samt VIAs og AU's egne medarbejdere, Indiesofts medarbejdere og kommunernes medarbejdere, når disse ikke har et nødvendigt arbejdsbetinget behov for at have adgang til oplysningerne.

Det vil desuden udgøre en risiko, hvis uvedkommende under Pilot #2 får adgang til de statistiske modeller, da en påvirkning af disse kan medføre fejl i algoritmen og derved skabe risiko for, at Beslutningsstøtten påvirker sagsbehandlingen på usaglig vis.

Uautoriseret adgang til oplysningerne kan være et resultat af såvel et utilstrækkeligt eksternt som internt sikkerhedsniveau. Uautoriseret adgang dækker tillige over tilfælde, hvor Indiesoft uretmæssigt anvender oplysningerne til egne formål eller uretmæssigt sender oplysningerne til underleverandører, herunder til tredjelande.

8.1 Foranstaltning nr. 1: Bruger og adgangsstyring

Beslutningsstøtten anvender den seneste udgave af Microsofts løsning til autorisation og adgangskontrol. Brugere oprettes ved, at der sendes et engangslink til en side, hvor de skal oprette et komplekst kodeord. Linket i mailen udløber efter kort tid. Gennem brug af roller er systemet opbygget således, at det er muligt at styre, hvilke brugere, der må oprettes og ændre personoplysninger, og hvilke brugere, der kan oprette socialrådgivere og trække rapporter.

AU havde i databehandleraftalen med Indiesoft vedrørende Pilot #1 desuden instrueret Indiesoft i at begrænse adgangen til personoplysningerne til de medarbejdere, for hvem det var nødvendigt at have adgang for at opfylde Indiesofts forpligtelser over for AU. Indiesoft var desuden instrueret om at behandle oplysningerne i projektet fortroligt og ikke uberettiget at udnytte, videregive eller på anden måde videreformidle de fortrolige oplysninger.

De data, der blev videregivet til AU efter Pilot #1, opbevares på et sikkert drev, hvor kun projektgruppens AU-ansatte medlemmer samt René Karberg fra Indiesoft¹⁰ har adgang. Michael Rosholm (data owner) og Sanne

¹⁰ René Karbergs/Indiesofts adgang til det sikrede drev er alene åbent i forbindelse med afprøvningerne, hvor René Karberg/Indiesoft skal kunne levere udtræk af data. I mellemliggende perioder mellem afprøvninger, lukkes René Karbergs/Indiesofts adgang til drevet.

Dalgaard Toft (data manager) kan tildele og fjerne personers adgang til mappe samt styre læse- og skriveretigheder. Adgangen til det sikre drev opdateres løbende.

AU's medarbejdere skal desuden benytte VPN med to-faktor autentifikation ved fjernadgang (fx hjemmearbejdsplads). Den ene faktor er brugernavn og password, mens den anden faktor er afhængig af, hvem medarbejderen er, og hvorvidt medarbejderen benytter udstyr, der er privat eller ejet af AU.

AU's medarbejdere kan kopiere data til eget eller eksternt udstyr fra det sikre drev, men denne overførsel logges, hvilket gør AU i stand til følge op, hvis der er mistanke om, at en medarbejder eller konsulent misbruger dennes adgang, jf. afsnit 8.2.

I VIA opbevares data enten på et fælles drev med adgang for flere eller på et drev udelukkende tilknyttet den enkelte medarbejder. Hvis flere skal have adgang, og data lægges på et fælles forskningsdrev, er det kun forskergruppens medlemmer fra VIA, der har adgang. VIAs medarbejdere vil i relation til projektet alene anvende forsknings-pc'er. At arbejde på en forsker-pc indebærer bl.a., at der ikke kan opnås lokaladministratorretigheder. Adgangen gives af VIA service, hvilket vil sige, at adgangskontrollen er centralt administreret.

Det er alene udvalgte socialrådgivere, der har haft adgang til Beslutningsstøtten i Pilot #1. De udvalgte socialrådgivere har ikke haft mulighed for at se hinandens brug af Beslutningsstøtten, og efter endt workflow har den enkelte sagsbehandler heller ikke haft mulighed for at genbesøge vurderingerne i systemet.

De samme forhold som beskrevet ovenfor vil fortsat gøre sig gældende i forhold til Pilot #2. Den enkelte sagsbehandler vil dog kunne genbesøge vurderingerne i systemet, idet der som nævnt i konsekvensanalysens afsnit 2.2.2.2 (Poul Schmith, VIA University College og Aarhus Universitet, 2021) vil være mulighed for at lade Beslutningsstøtten følge sagen.

8.2 Foranstaltning nr. 2: Logning

I AU logges al databehandling på det sikre drev. Data, som undtagelsesvist opbevares på en ekstern enhed (fx krypteret USB eller SD-kort), logges ikke. Dog logges det, at dataene er blevet overført til den eksterne enhed. Loggen krypteres efter ét år og opbevares herefter af AU i 20 år, jf. AU's generelle retningslinjer om logning.

Indiesoft blev under Pilot #1 instrueret i at foretage fuldstændig logning på enhver behandling af personoplysningerne i systemet, hvorved kommunernes færden i systemet ligeledes blev logget. Loggen fra Pilot #1 blev slettet af Indiesoft, da pilotprojektet blev lukket ned, og backup blev slettet efter 70 dage. Indiesoft overførte dog loggen til AU. AU har efterfølgende slettet denne log. Indiesoft overførte samtidigt dataudtrækket fra afprøvningen til AU og VIA, der opbevarer denne i overensstemmelse med retningslinjerne anført i konsekvensanalysens afsnit 2.6 (Poul Schmith, VIA University College og Aarhus Universitet, 2021).

I VIA logges alle transaktioner til og fra forskningsserverne. Logningen opbevares i et år.

VIA og AU kan med logningen føre kontrol med hvem, der har tilgået data, og hvornår de har gjort dette. Logningen forhindrer ikke direkte uautoriseret adgang til personoplysninger, men den gør VIA og AU i stand til følge op, hvis der er mistanke om, at en medarbejder eller konsulent misbruger vedkommendes adgang udover det arbejdsbetingede behov. Derudover kan logningen have en præventiv effekt, hvis medarbejdere gøres bevidste om, at deres handlinger på de sikrede drev og servere bliver loggede og potentielt gennemgået med henblik på kontrol.

VIA og AU vil følge samme retningslinjer som angivet ovenfor i Pilot #2.

8.3 Foranstaltning nr. 3: Kryptering

Der anvendes anerkendt kryptering mellem AU og AU's eksterne leverandører. AU's servere overholder i øvrigt alle sikkerhedskrav.

VIA's forskningsserverne er BitLocker-krypterede, og adgang til disse kræver både en forsker-pc og VPN-adgang. VIA's forsker-pc'er kan alene anvende et USB-drev, hvis det er krypteret.

I forbindelse med afprøvningen i Pilot #1 blev dataene opbevaret på henholdsvis Hjørring Kommunes hardware i Hjørring og Indiesofts hardware hos Amazon i Irland. Alle dataene blev krypterede under transporten via Internettet, og navn samt CPR-nummer blev opbevaret krypteret i databaserne. AU vil i Pilot #2 ligeledes stille krav om kryptering til sin(e) databehandler(e).

I det omfang, det bliver nødvendigt at udveksle personhenførbare data mellem AU og VIA i Pilot #2, vil data sendes krypteret med sikker post med digital signatur eller via krypteret USB/SD-kort (f.eks. krypteret med 7-zip eller BitLocker).

Disse foranstaltninger bidrager samlet set til at reducere risikoen for, at uvedkommende eksterne aktører får uberettiget adgang til personoplysningerne, herunder f.eks. med henblik på misbrug eller forvanskning af data.

9. Risiko nr. 8: Manglende fastlæggelse af dataansvar, herunder mangler i aftalekomplekset

Det er af afgørende betydning, at VIA og AU har et klart overblik over rækkevidden af deres dataansvar, når de behandler personoplysningerne i projektet. Dette gælder både i forhold til databehandlere, men også i forhold til andre tredjeparter såsom Danmarks Statistik, samarbejdskommunerne mv. Manglende overblik over parternes dataansvar kan medføre risici for de registrerede, da det risikeres, at hverken VIA og AU påtager sig ansvar for en del af behandlingen, herunder at VIA og AU ikke indgår de aftaler, som er nødvendige i henhold til databeskyttelsesretten.

9.1 Foranstaltning nr. 1: Fastlæggelse af dataansvar

Som redegjort for i konsekvensanalysens afsnit 1.2.1 (Poul Schmith, VIA University College og Aarhus Universitet, 2021) er VIA og AU selvstændige dataansvarlige og vil fortsat agere med denne ansvarsfordeling. VIAs og AU's dataansvar er herved fastlagt for begge piloter.

9.2 Foranstaltning nr. 2: Indgåelse af de nødvendige aftaler

AU behandler, herunder opbevarer, fortsat de oplysninger, som blev indsamlet i forbindelse med Pilot #1. VIA har slettet de data, der blev delt i forbindelse med Pilot #1. VIA og AU vil sikre, at de i pilot #2 indgår en videregivelsesaftale i overensstemmelse med videregivelsesbekendtgørelsen, inden de deler data mellem dem.

AU vil i henhold til Pilot #2 desuden indgå en videregivelsesaftale med den valgte samarbejdskommune.

VIA og AU vil ligeledes sikre, at hver part indgår de(n) nødvendige databehandleraftale(r), jf. databeskyttelsesforordningens artikel 28, når det endeligt er afgjort hvilken/hvilke databehandler(e), de vil gøre brug af under pilotforsøget.

10. Litteratur

Danmarks Statistik (2018). Næsten hvert andet barn med gentagne og grove underretninger i 2015 fik ikke støtteforanstaltning. [Artiklen kan læses her.](#)

Danmarks Statistik (2019). Introduktion til forskningsservice i Danmarks Statistik. [Artiklen kan læses her.](#)

Danmarks Statistik (2020). Datafortrolighedspolitik. [Artiklen kan læses her.](#)

Poul Schmith, VIA University College og Aarhus Universitet (2021). Konsekvensanalyse vedrørende databeskyttelse. Kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#).

Poul Schmith (2021). Forvaltningsretlig legalitetskontrol af Beslutningsstøtten. Kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#).

Lundberg, Scott M., and Su-In Lee (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems.

TrygFondens Børneforskningscenter (2020). Data og modelbeskrivelse vedrørende projekt Beslutningsstøtte ved underretninger. Internt notat.

TrygFondens Børneforskningscenter (2021a). Kort projektbeskrivelse. Kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#).

TrygFondens Børneforskningscenter (2021b). Egenskaber ved den statistiske model i forskningsprojektet Underretninger i fokus. Kan tilgås på projektsiderne på [TrygFondens Børneforskningscenter](#) eller [UC Viden](#).