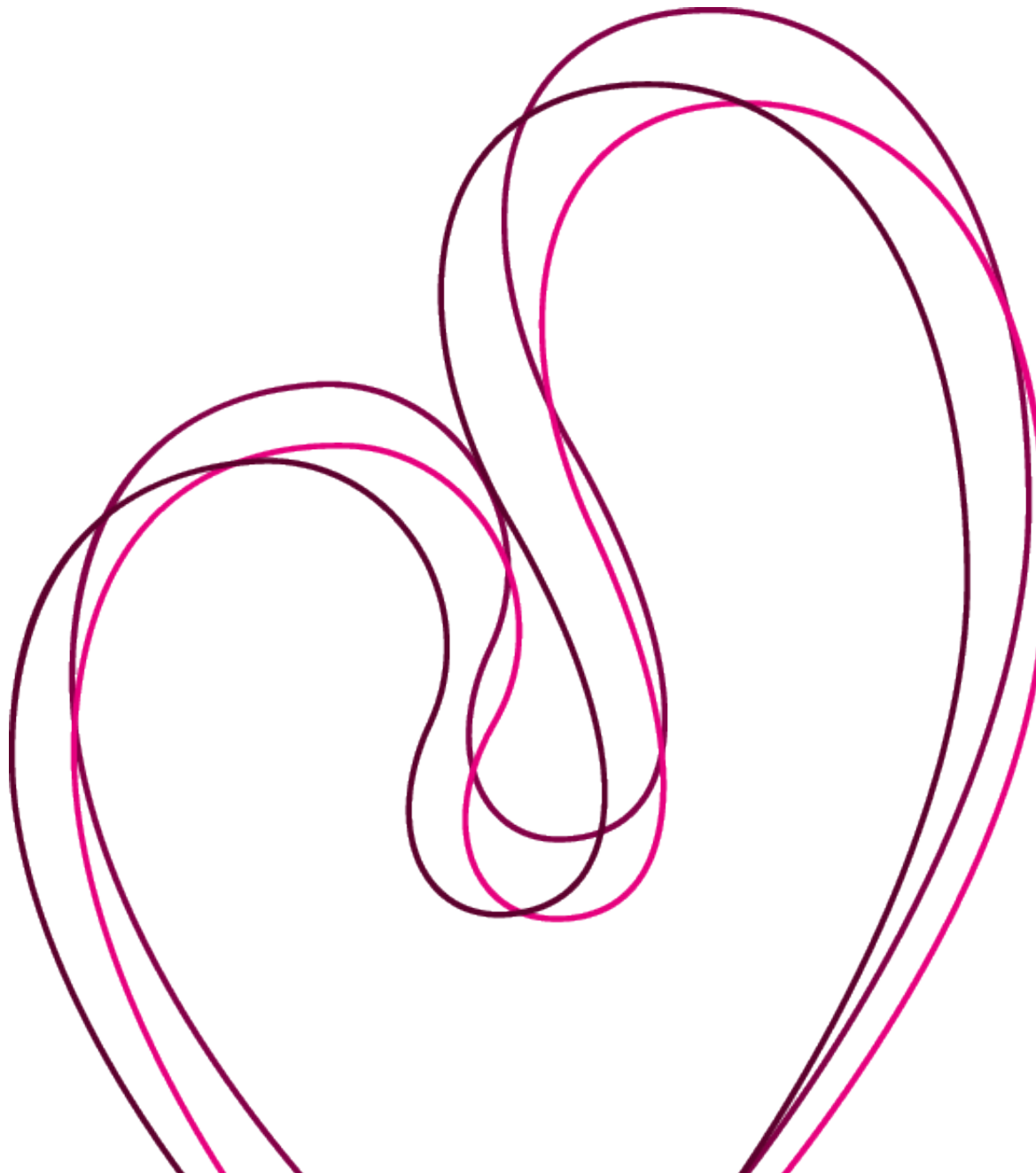


Mitigerende foranstaltninger

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

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

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

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

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).

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

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 velbegrunderet 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 oplyste variable 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.

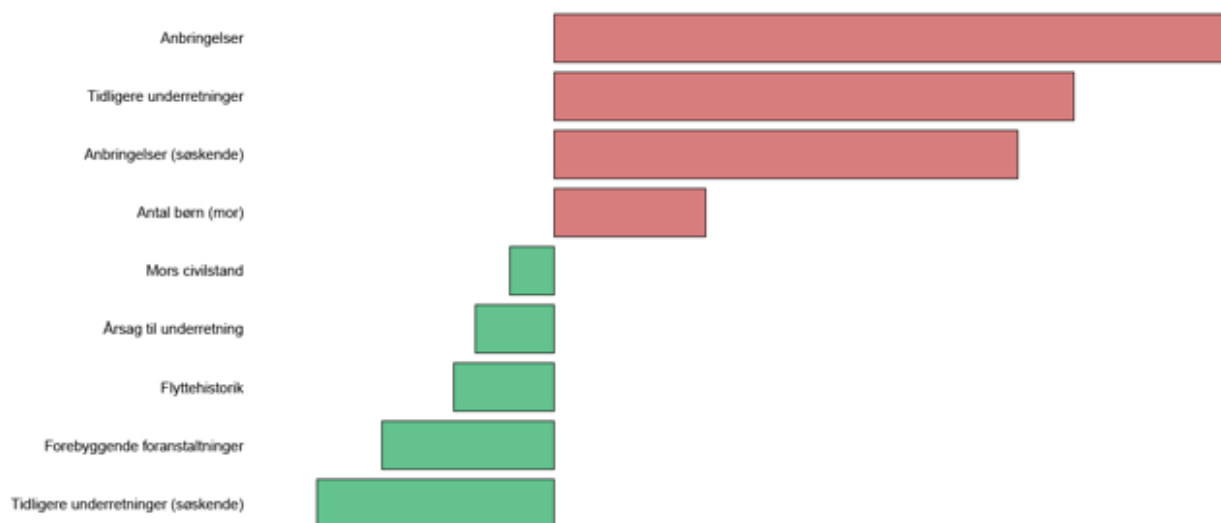
For at sikre at Beslutningsstøtten er så "fair" som muligt, er det valgt, at modellen ikke skal have kendskab til den etniske oprindelse af de børn og de familier, som værktøjet bruges på. Ved at udelade denne type information fra modellen, sikres det, at en given risikovurdering fra Beslutningsstøttens 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 prædiktions. 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 prædiktions 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 modelprædiction.

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

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

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

VIA og AU valgte på baggrund af erfaringerne fra Pilot #1 således at reducere ”*mistrivsel*” til ét enkelt udcømemål 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 recalibrering 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 recalibrering under Pilot #2, idet denne afprøvning alene vil forløbe over 1-3 måneder. Modellen vil dog blive gentrænet eller recalibreret 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 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.

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 variabelenes 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 børns 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: Registere og variable anvendt til udvikling af modellerne i Pilot#1 og Pilot#2

Registere 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 (Forældremærkning) – Kun Pilot#2 i. CIVST (Civilstand) – Kun Pilot#2	Ja	Ja

Registre og variable anvendt til udvikling	Indgår i udvikling Pilot#1	Indgår i udvikling Pilot#2
Register 3: BUFO (Børn og unge forebyggende foranstaltninger) <ul style="list-style-type: none"> 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) <ul style="list-style-type: none"> a. PNR (Personnummer) b. EFFEKTDATO (Effektueringsdato) c. HAENDELSE (Hændelse i anbringelsessag) d. NYKOM (Handlekommune) 	Ja	Ja
Register 5: HUST (Husstandsoplysninger) <ul style="list-style-type: none"> 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) <ul style="list-style-type: none"> a. PNR (Personnummer) b. DODDATO (Dato for dødsfald) 	Nej	Ja
Register 7: FLYT (Flytninger i Danmark) <ul style="list-style-type: none"> 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 pseudodynamiserede 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 prædikterer 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: Registre og variable anvendt til ekstern validering af modellen i Pilot#2

Registre og variable anvendt til validering i Pilot#2
<p>Register A: KROF (Ofre for straffelovsforbydelser)</p> <ul style="list-style-type: none"> e. PNR (Personnummer) f. YEAR (Året hvor den pågældende information er registeret) g. OFR_GERFRADT (Gerningens startdato) h. OFR_GER7 (Offeret gerningsindholdskode DS)
<p>Register B: KRSI (Kriminalstatistik sigtelser)</p> <ul style="list-style-type: none"> a. PNR (Personnummer) b. YEAR (Året hvor den pågældende information er registeret) c. SIG_SIGTDTO (Sigtelsesdato) d. SIG_GER1DTO (Gerningsdag, startdato)
<p>Register C: LPRADMDIAG (Landspatientregisteret)</p> <ul style="list-style-type: none"> a. PNR (Personnummer) b. YEAR (Året hvor den pågældende information er registeret) c. C_DIAG
<p>Register D: PSYK_ADM (Landspatientregistret psykiatri - administrative oplysninger)</p> <ul style="list-style-type: none"> a. PNR (Personnummer) b. RECNUM (LPR-identnummer) c. D_HENDTO (Henvisningsdato) d. D_INDDTO (Indlæggelsesdato)
<p>Register E: PSYK_DIAG (Landspatientregistret psykiatri - diagnoser)</p> <ul style="list-style-type: none"> a. RECNUM (LPR-identnummer) b. C_DIAG (Diagnosekode)
<p>Register F: Trivsel_2015_2018 (De nationale trivselsmålinger, data fra Styrelsen for IT og Læring)</p> <ul style="list-style-type: none"> 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 permanente tænder)
- 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 ekstern validering

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 mistriivsel blandt de børn, som ender med at blive anbragt. Dette er rationalet for brugen af anbringelser som vores observerbare mistriivselssindikator.

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 mistriivsel. 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 ensemblemetode, hvor modellens endelige output er givet ved et gennemsnit af outputtet fra en række *uafhængige* beslutningstræer. XGBoost metoden er også en ensemblemetode, 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 uploadet på serveren og det er bekræftet, at data er korrekt uploadet. 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.

⁹ O-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 krænkelser 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: Brugere 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 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. 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 lokaladministratorrettigheder. 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](#).*