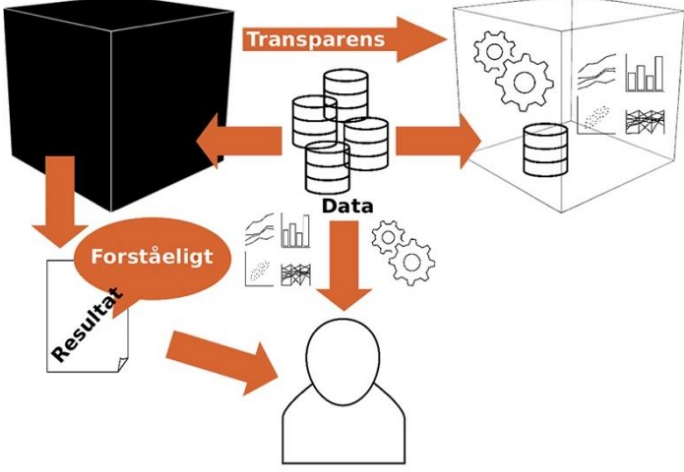


## Skema: Ansøgning om resultatkontraktmidler 2019-2020

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                            |                              |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------|-----|
| <b>Institut(ter):</b><br><br><b>Alexandra Instituttet</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | <b>Aktivitetsplan (titel):</b><br>Datadreven beslutningsstøtte – transparent og forklarlig dataanalyse til alle<br><br><b>Idéforslagstitel på bedreinnovation.dk:</b><br><br>Datadreven beslutningsstøtte – transparent og forklarlig dataanalyse til alle | <b>Aktivitetsplan nr.: 3</b> | FoU |
| <b>1) Manchettekst (kort resumé)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                            |                              |     |
| Dette projekt skaber tillid og tilgængelighed i datadrevne beslutningsstøtteværktøjer. Resultater gøres transparente samt lette at forklare via avanceret Machine Learning og visualiseringsværktøjer. Det skaber tryghed for brugere, der benytter dataanalyse.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                            |                              |     |
| <b>2) Aktiviteten kort (resumé)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                            |                              |     |
| <p>Der indsamles i disse år enorme mængder data - "Big Data" - i virksomheder inden for de fleste brancher, lige fra konventionelt manuelle erhverv i f.eks. plejesektoren og byggebranchen til højteknologiske brancher så som forsvarsindustrien og produktionssektoren. Data indsamles ofte ud fra deisen om, at der er viden og værdi i data – hvilket i høj grad er sandt – og dataindsamling sker ofte i regi af Big Data projekter. Men for at drage nytte af de indsamlede data, er det nødvendigt at analysere data med sofistikeret Machine Learning (ML) og visualiseringsværktøjer, således at virksomheder kan udnytte data til at træffe bedre beslutninger. Dette skaber en række udfordringer, som vi vil adressere i denne aktivitetsplan.</p> |                                                                                                                                                                                                                                                            |                              |     |
| <p>Udfordringerne består i et behov for <i>transparens</i> af ML og datadrevne analysemodeller samt et behov for <i>at forklare</i> deres resultater, især når de bliver uigennemskuelige for både den almene bruger og ekspertbrugeren. Dette er særligt nødvendigt, når data og beslutninger vedrører personer. Her kræver <a href="#">GDPR</a> nemlig, at beslutninger skal kunne forklares over for de parter, beslutningerne berører.</p>                                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                            |                              |     |
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                            |                              |     |
| <p>Øget transparens og forbedret forklaringsevne er derfor påkrævet for at kunne styrke tilliden til datadrevne beslutningsstøtteværktøjer og for at gøre ML og store datamængder mere tilgængelige for et bredere udvalg af brugere og derved udvide anvendelsesområdet af datadrevne modeller. Forøgelse af transparens og forklaringsevne i beslutningsstøtteværktøjer kan bl.a. indebære at visualisere data i værktøjerne samt at forklare, hvilke dele af dataene fører til et vist resultat.</p>                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                            |                              |     |

Transparens og evne til at forklare datadreven beslutningsstøtte er en del af ”explainable AI”, altså forklarlig kunstlig intelligens. Dette kan brydes ned i de følgende problemstillinger inspireret af Casimir Wierzynski, Senior Director hos Intel, der beskriver udfordringer for ML ([link](#)):

- P1. **Evne til at forklare resultater:** Kan vi bygge modellen, så som den kan forklare resultater til enhver person på en forståelig måde? Dermed vil datadreven beslutningsstøtte være mere tilgængelig, og dette kunne også kræves af GDPR.
- P2. **Evne til at forklare grundlæggende mekanismer:** Kan vi forklare grundlæggende mekanismer af problemet ved hjælp af en ML-model, dvs. ikke kun give et resultat men også forklare, hvilke dele af data har påvirket resultatet mest? Dermed vil man kunne forstå både problemet og løsningen bedre.
- P3. **Transparens af modellen:** Kan vi skabe indblik i modellens processer ved hjælp af værktøjer? Dette vil gøre det nemmere for enhver person at forstå ML-modellens virkninger og dermed skabe en lettere adgang til datadreven beslutningsstøtte.
- P4. **Bias og fairness:** Kan vi identificere bias i eksisterende systemer? Hvilken indflydelse har bias på datadreven beslutningsstøtte? Kan vi verificere, at beslutninger er truffet på retfærdig vis, dvs. undgå at en gruppe diskrimineres i forhold til anden, [eksempelvis sørge for, at alle har samme mulighed for at få et lån](#). Dette er primært også et samfundsmæssigt behov, da vi garanterer, at beslutninger er retfærdige over for alle, der berøres af dette, og dermed kan tilliden til ML-systemer øges.

For at adressere P1-4 vil vi udvikle nye algoritme- og visualiseringsværktøjer der gøre det muligt for domæneeksperter at udforske og kontrollere de automatiske analysemetoders resultater og virkninger uden teknisk indsigt i ML-teknologien.

Derudover vil vi udvikle ydelser, der hjælper virksomheder med at identificere hvilke state-of-the-art ML-modeller, der sammen med de udviklede værktøjer kan løse deres udfordringer. Det vil øge antallet af mulige anvendelsesområder og effektiviteten af datadrevne modeller i praksis.

For at øge fairness i datadrevne modeller (P4) vil vi udvikle en rådgivningsydelse til at informere om fairness i data og datadrevne modeller. Desuden vil vi opbygge en vidensbase om årsager og virkninger af bias i datadrevne modeller. Disse kompetencer kan bruges til at yde konsultation til virksomheder om fairness af deres modeller og give retningslinjer for god praksis for brug af datadreven beslutningsstøtte.

### 3) Markedsbehov, erhvervs- og samfundsmæssige potentialer

#### Behov for datadreven beslutningsstøtte

Gartner vurderer medio 2017, at frem mod 2021 vil *Data and Analytics* blive centralt og drivende for markedsledende virksomheders beslutningsprocesser [1], samt at konventionelle Business Intelligence systemer vil være utilstrækkelige i den udvikling. Det betyder, at virksomheder skal omstille deres beslutningsprocesser til at være datadrevne for at kunne følge trop.

At dette er en nært forestående udvikling, samt at der er et bredt behov for datadreven beslutningsstøtte i dansk erhvervsliv, kan desuden ses ud fra kommentarerne fra aktivitetsplanens opslag på Bedreinnovation.dk [2]. Her har både SMV'er samt store virksomheder fra en lang række brancher udtrykt interesse: forsvarsindustrien (Terma), forsyningsbranchen (Aarhus Vand, Vitani), plejesektoren (Sekoia, Schultz Pleje og Hjemmeservice), produktionssektoren (Grundfos, Jeppe Lützhøft, FieldSense), byggebranchen (Exigo), undervisningssektoren (Gyldendal) og transport- og logistiksektoren (Saint Gobain, Vejdirektoratet). De forskelligartede brancher gør det til et oplagt aktivitetsforslag for Alexandra Institut, da man har erfaring i at levere til virksomheder i mange forskellige brancher.

Konkrete behov for transparent og forklarlig beslutningsstøtte ses ligeledes ud fra kommentarer aktivitetsplanens opslag på Bedreinnovation.dk. Eksempelvis skriver Martin Løkke, Vice President hos Terma:

*'Nogle beslutninger kan tages automatisk, mens andre kræver at der er "Man-in-the-loop" i beslutningen. Her vil en visualisering af de ofte store og komplekse mængder data kunne forbedre beslutningsprocessen.'*

Det viser, at man hos en af dansk erhvervslivs mastodonter ser et klart behov for at muliggøre, at brugeren kan være aktiv medspiller i en beslutningsproces og ikke blot er aftager af et resultat. Små virksomheder har samme behov for at kombinere Visual Analytics med ML, som Henrik Pedersen fra Exigo kommenterer:

*'Ved at parre machine learning og visual analytics med vores domæneviden inden for byggebranchen er det vores håb, at vi kan hjælpe med at skubbe byggebranchen i en retning, hvor beslutninger ikke længere alene baseres på projektlederens eller entreprisederens mavefornemmelser men i langt højere grad lader beslutninger være baseret på tilgængelige data.'*

Denne kommentar eksemplificerer, at små virksomheder, der har dybt domænekendskab inden for en branche (byggebranchen i Exigos tilfælde), har behov for assistance til at indføre datadreven beslutningsstøtte.

Ud over kommentarer fra virksomheder på Bedreinnovation.dk har vi kontakt til en række virksomheder, der allerede nu er interesserede i at engagere sig i en aktivitet om Datadreven Beslutningsstøtte. Disse virksomheder er et udpluk af virksomheder, som Alexandra Instituttet har kontakt til ud fra vores brede berøringsflade i gennem workshops, seminarer, konferencer, forskningsprojekter, kommercielle ydelser, tidligere resultatkontrakter mv. Specifikt kan nævnes den igangværende resultatkontrakt *Teknologier og værktøj til udnyttelse af Big Data*, vores deltagelse i forskningsprojektet *DABAI*, samt vores engagement i innovationsprojektet *Innovativ brug af Big Data*. Følgende virksomheder har udtrykt interesse for at deltage i en aktivitetsplan om datadreven beslutningsstøtte ud fra konkrete behov, og disse vil kunne fungere som mulige case-samarbejdspartnere:

- *EnviDan, der leverer rådgivning til forsyninger og kommuner inden for vandforsyning, afløbssystemer og renseanlæg:* EnviDan udforsker muligheder for at integrere ML i deres produkter og er i den forbindelse også interesseret i øget transparens og evne til at forklare datadrevne modeller. Dette vil både hjælpe dem til at forstå, hvordan ML kan bedst anvendes i deres produkter men også skabe større tillid hos deres kunder til datadreven beslutningsstøtte.
- *Conscius, der blandt andet leverer software til el-leverandører:* Conscius er interesseret i at undersøge mulighederne for at anvende datadreven beslutningsstøtte til yderlige ydelser for deres kunder, som vil få gavn af øget transparens i Conscius' software og evne til at forklare resultater.

Som følge af EnviDans og Conscius' interesse vurderer vi, at der er et stort potentiale inden for forsyningsbranchen for at anvende de ydelser, der udvikles under aktivitetens forløb. Der vil være fokus på virksomheder, der leverer produkter, ydelser og rådgivning til f.eks. el-leverandør, forsyningselskaber eller kommuner.

Denne interesse fra virksomheder, der leverer produkter, ydelser og rådgivning gør sig også gældende for virksomheder inden for andre brancher. På aktivitetsplanens opslag på Bedreinnovation.dk har Exigo, Vitani og Gyldendal udtrykt interesse for at kunne udbyde datadreven beslutningsstøtte til deres kunder.

Inden for plejesektoren har både Schultz Pleje og Hjemmeservice samt Sekoia udtrykt interesse for aktivitetsplanens opslag på Bedreinnovation.dk, hvorfor vi ser gode muligheder for en case sammen med en virksomhed i plejesektoren. Interessant ved netop disse to virksomheder er, at de agerer inden for samme branche – plejesektoren – men på forskellige måder. Hvor Schultz Pleje og Hjemmeservice er udførende inden for branchen, leverer Sekoia ydelser til virksomheder i branchen.

I disse relationer er den primære målgruppe for denne aktivitetsplan rådgivningsvirksomheder, der leverer produkter, ydelser og rådgivning til virksomheder inden for specifikke brancher. Den sekundære målgruppe er virksomheder, der er udførende inden for de specifikke brancher. F.eks. entreprenører der

køber risikostyringsværktøjer hos Exigo, forsyningsselskaber der køber rådgivning hos EnviDan eller produktionsvirksomheder der køber energiadministrationsydelser hos Vitani.

Desuden er der også udtrykt interesse fra større produktvirksomheder, der selv har interesse i datadreven beslutningsstøtte. Her kan nævnes Terma, Grundfos, Saint Gobain og Aarhus Vand, som alle har udtrykt deres interesse på aktivitetsplanens opslag på Bedreinnovation.dk.

Ud over gevinst for virksomheder er der også et samfundsmæssigt behov for **transparente og fair** modeller. Det er kritisk, at ML-modeller, der benyttes i en beslutningsproces, ikke indeholder en *bias*, der diskriminerer en bestemt gruppe mennesker. Eksempler fra USA har vist, at ML-modeller med bias er brugt til strafudmåling i retssager [3], uden at resultaterne kunne forklare. Dette medfører tab af tillid til ML-metoder og dermed til langsom udvikling og udbredelse af datadrevne modeller. Derfor er det vigtigt at støtte ansvarlig brug af modellerne samt øge transparensen og evnen til at forklare modellernes resultater. Dette er desuden en forudsætning for ibrugtagning af ML i borgerrettede beslutningsprocesser i kommuner som følge af krav til algoritmisk forklarlighed, hvor grundlaget for beslutninger, der vedrører borgere, skal kunne forklare over for borgeren.

### Tilgængeliggørelse af data

Datadreven beslutningsstøtte er en essentiel komponent i at omdanne data til viden og dermed beslutningsgrundlaget for at skabe vækst. Forskere fra CBS beskriver i deres omfattende model for DataProfit [4], at opsamling, opbevaring og tilgængelighed af data er en grundlæggende og forudsættende præmis for at kunne skabe datadreven vækst. De seneste års fremskridt inden for Big Data har genereret tekniske løsninger, der er velegnede til at opsamle og opbevare store datamængder. Tilgængeliggørelsen af data, derimod, er en langt mere kontekstuel udfordring, da domæneeksperterers behov kan variere kraftigt på tværs af både virksomheder og brancher.

Udfordringen med tilgængeliggørelse af data er en opgave som nogle virksomheder er i stand til at løfte på egen hånd ved at ansætte professionelle analytikere eller i store virksomheders tilfælde etablere hele afdelinger. For mindre virksomheder kan det derimod være en uoverkommelig udfordring, hvis eksisterende Business Intelligence eller Artificial Intelligence hyldevarer ikke er tilstrækkelige. Sådanne virksomheder risikerer at sakke bagud ift. nyskabende analysemetoder, simpelthen fordi innovationshøjden er for høj for dem selv at bære. Det betyder, at danske virksomheders vækst potentielt hæmmes, hvis de ikke kan løfte innovationsbyrden selv.

Selv om større virksomheder selv har stærke tekniske kompetencer, er det også en udfordring for dem at tilegne sig viden og kompetencer, der muliggør, at de kan benytte nyskabende teknologier. Større virksomheder vil derfor kunne drage nytte af at kunne modtage rådgivning, der er på forkant med den teknologiske udvikling. Hvor mindre virksomheder vil kunne drage direkte nytte af en dedikeret indsats inden for udvikling af kompetencer for datadrevne beslutningsstøtteværktøjer, vil det også være til gavn for større virksomheder f.eks. inden for produktionssektoren eller forsvarsindustrien.

Nye udfordringer med datadreven beslutningsstøtte handler om **transparens af modellen**, der bliver brugt til at finde sammenhænge i data og **evne til at forklare** modellens resultater på en forståelig måde. Ønsket for transparens kan have flere årsager:

- *Virksomheden arbejder med persondata.* Her kræver [GDPR](#), at brug af persondata er transparent, herunder specielt, hvis disse bliver brugt til datadreven beslutningsstøtte. Dette forudsætter enten transparens af modellen, evne til at forklare resultatet, eller en kombination af begge. Persondata-dreven beslutningsstøtte kunne f.eks. bruges i plejesektoren, forsyningsbranchen, undervisningssektoren, sundhedsvæsenet eller finanssektoren med henblik på at kunne tilbyde bedre og mere individuelle produkter eller ydelser til borgerne/kunderne.

- *Klæde domæneeksperter og andre aftagere på til at forstå, agere og tage beslutninger ud fra beregnede analyseresultater.* Her er det essentielt at øge transparensen og at gøre det muligt at fortolke datadrevne analyseresultater. Målet er at fremme datadreven beslutningstagning yderligere ved at gøre det lettilgængeligt for brugere på alle niveauer i virksomheder og organisationer, der ofte ikke er ML- eller dataanalysespecialister, men som skal omsætte og tage konsekvenserne af beslutningerne.
- *Der kræves en "man-in-the-loop" i processen.* Her er det vigtigt, at denne person kan forstå modellens resultater. Dette kræves ofte i service- eller sikkerhedskritiske processer eller systemer, der forekommer f.eks. i forsvarsindustrien eller sundhedsvæsen.

Mange ML-modeller som f.eks. neurale netværk kan i dag typisk ikke leve op til disse krav, da de er uigennemskuelige og vanskelige at fortolke for selv domæneeksperter på grund af "black box"-karakteren af modellerne samt mængden og kompleksiteten af de bagvedliggende data.

Med denne aktivitetsplan vil vi fokusere på at tilgængeliggøre transparent og forklarlig dataanalyse – primært med fokus på behovene hos SMV'erne, men metoderne vil også være anvendelige i større virksomheder. Dette gælder både for virksomheder, der vil være direkte brugere af datadreven beslutningsstøtte samt virksomheder, der udvikler og udbyder rådgivningsydelser med datadreven beslutningsstøtte til virksomheder inden for specifikke brancher. Resultaterne af aktiviteten vil give virksomhederne en markedsfordel i form af et udvidet anvendelsesområde og øget produktkvalitet. Desuden vil slutbrugerne profitere af, at produkter er udviklet individuelt til deres behov.

#### **Behov for datadreven beslutningsstøtte hos SMV'er**

Et groft estimat på antallet af små og mellemstore virksomheder, der vil kunne drage nytte af en denne aktivitet kan udregnes ved hjælp af Danmarks Statistiks ERHV4 tabel over arbejdsstørrelse fordelt på DB07 med 127 NACE branchekoder fra 2016.

Ved at filtrere på branchekoder som stemmer overens med virksomheder, som har kommenteret på aktivitetsplanens opslag på Bedreinnovation.dk, giver dette et totalt antal virksomheder på 40.130, hvis man kun inkluderer virksomheder med arbejdsstedsstørrelse fra 1 ansat til og med 99 ansatte. Inkluderer man i stedet kun virksomheder med en arbejdsstedsstørrelse fra 2 og op til 99 ansatte, udgør tallet 22.025 virksomheder.

Pointen er ikke det totale antal virksomheder, da en filtrering af virksomheder på kun 127 branchekoder givetvist er en grovkornet søgning, og det reelle antal aftagervirksomheder vil være lavere end både 40.130 og 22.025 virksomheder. Dog giver filtreringen et fingerpeg om, at der findes mange virksomheder inden for brancher, hvor datadreven beslutningsstøtte vil være relevant. Og vurderet ud fra aktivitetsplanens kommentarer på Bedreinnovation.dk, er det højaktuelt netop nu.

#### **4) Videnspredning og inddragelse**

- **Inddragelse af danske virksomheder:** Vi vil lave en bred behovsafdækning på tværs af brancher gennem brugerstudier og/eller workshops for at identificere anvendelsesområder og udfordringer med transparent og forklarlig dataanalyse. Der vil desuden etableres en følgegruppe af 4-6 virksomheder i målgruppen, der møder 1 gang om året. Vi vil desuden gennemføre mindst 2 cases med danske virksomheder i forskellige brancher, forventeligt i forsyningsbranchen og plejesektoren (se også 3). Case-forløbene og de løbende resultater af samarbejdet vil vi formidle i f.eks. case-udgivelser, short-stories og via input til arrangementer.
- **Eksisterende initiativer:** Alexandra Instituttet er tilknyttet til en række konsortier og netværk inden for IT og Big Data, som f.eks. InfinIT, dansk.ai og DABAI. Vi vil bruge disse til at organisere workshops og rekruttering af relevante virksomheder for case-projekter men vil også kunne hjælpe med udfordringer med implementering af state-of-the-art metoder og identifikation af behov for dybdegående, ny forskning og udvikling af nye metoder.

- **Samarbejde med vidensinstitutioner:** Alexandra Instituttet er forankret blandt datalogiske institutter på de danske universiteter og har derfor stærke relationer til bl.a. Datalogisk Institut ved Aarhus Universitet og Københavns Universitet samt DTU Compute og IT-Universitetet. Disse relationer vil vi kapitalisere på ved at inddrage ekspertviden fra forskere i vores arbejde i projektet – f.eks. i specifikke cases eller ved publikation af artikler.
- **Formidling af aktiviteter og resultater:**  
Vi vil afholde en række vidensspredningsarrangementer som har til formål dels at informere interesserede virksomheder om eksisterende muligheder og udfordringer og dels vil vi afholde workshops målrettet til slutbrugere med henblik på at skabe mere tillid til datadreven beslutningsstøtte i befolkningen. I løbet af aktivitetsperioden forventer vi at inddrage mindst 20 forskellige virksomheder og slutbrugere i vidensspredningsaktiviteterne  
Vidensspredningsarrangementerne suppleres af formidling via nyhedsbreve, PR-kampagner, sociale og webmedier, branchemedier, fagmedier og nationale medier. Formidlingen af resultater og aktiviteter vil være content-baseret og anvendelsesorienteret, og der vil så vidt muligt benyttes eksempler, trends/tendenser og cases til at vise udbyttet af arbejdet i aktivitetsplanen. Derudover prioriteres den personlige, opsøgende dialog med virksomheder, brancheorganisationer, væksthuse og netværk. Vi forventer at vække interesse hos min. 80 danske virksomheder i forbindelse med formidling.  
Publikationer: Vi forventer desuden at udgive mindst en conferenceartikel eller forskningsartikel i andet år af aktivitetsplanens og at formidle resultaterne gennem 3-4 indlæg i fagmedier eller foredrag i relevante fora.

## 5) Konkrete aktiviteter

De centrale aktiviteter vil omhandle (relation til problemstillinger fra afsnit 2 er betegnet f.eks. P1):

### A1: State-of-the-art afdækning

- P1-P2: Indledningsvist gennemføres en afdækning af state-of-the-art værktøjer, der benyttes til at forklare datadrevne modellers resultater og grundlæggende mekanismer.
- P3: Der vil ligeledes blive lavet en opsamling og analyse af state-of-the-art metoder, der har til hensigt at fremvise ML-modeller og data mere transparent.
- P4: Opbygning af en videnbase om årsager og effekter af bias i datadrevne modeller.

### A2: Kompetenceopbygning og teknologiudvikling

- P1-P3: Opbyg viden om implementering og anvendelsesområde af eksisterende state-of-the-art metoder (som f.eks. SHapley Additive exPlanations, [SHAP](#), der præsenterer en uniformeret tilgang til at forklare resultater fra ML-modeller), så at de rigtige værktøjer kan vælges afhængig af problemstilling og branche.
- P1-P2: Udvikling af nye algoritmiske værktøjer, der giver indsigt i, hvorfor en ML-model har truffet en beslutning. For P2 vil dette primært fokusere på analyse af, hvilke faktorer er mest afgørende for beslutningen.
- P3: Udvikling af nye teknikker til at gøre det muligt for domæneeksperter at udforske og kontrollere de automatiske analysemetoder og deres resultater uden teknisk indsigt i ML-teknologien.
- P4: Udvikling af rådgivningsydelser til at informere om bias i data og fairness af datadrevne modeller.
- Deltagelse i 1-2 relevante forsknings- og industrikonferencer.

### A3: Behovsafdækning

- Parallelt med state-of-the-art (A1) gennemføres behovsafdækning gennem brugerstudier mhp. at identificere udfordringer og behov med eksisterende ML metoder samt mulige anvendelsesområder og problemstillinger ift. de konkrete cases som der arbejdes på (A4).
- Der etableres også en følgegruppe med deltagelse af 4-6 virksomheder fra målgruppen, som deltager i 1 årligt møde, hvor formålet er at drøfte resultater og status på udviklingen af teknologiske serviceydelser, og hvorledes disse matcher målgruppens behov. Der opnås herunder en afklaring af de involverede Alexandra Instituttets rolle ift. private rådgivere og leverandører mhp. at optimere synergi

og komplementaritet og undgå konkurrenceforvridning. Den opnåede fælles forståelse sammenfattes skriftligt.

#### **A4: Case-forløb**

- Under case-forløbet vil der iterativt udvikles prototyper på værktøjer, der kombinerer ML-funktionalitet algoritmiske værktøjer og visualisering. Case projekterne udvælges efter nyhedsværdi og innovationshøjde således at de både indeholder et relevant anvendelsesperspektiv og giver mulighed for implementering af state-of-the-art metoder (A1) samt udvikling af nye metoder baseret på virksomhedernes behov (A3).
- Vi forventer mindst 2 specifikt behovsafdækkende og kompetenceopbyggende prototypeforløb med virksomheder inden for ca. 2 brancher. Nogle case-forløb kan eventuelt suppleres med InnoBooster-forløb med virksomhederne eller som en funktionalitetstilføjelse af eksisterende projekter.
- Gennemførelsen af case-forløbene fordrer hvervning af virksomheder til case-forløbene. Dette er dog udfordringer, som Alexandra Instituttet er godt positioneret til at håndtere gennem vores brede berøringsflade, og vi ser derfor hvervning af virksomheder til case-forløb som en realiserbar udfordring.

#### **A5: Vidensspredning**

- P1-P3: Workshops til at informere virksomheder om trade-off mellem nøjagtighed, transparens og evne til at forklare resultaterne i datadrevne modeller og gevinsten ved at bruge transparente modeller over for "black-box" modeller.
- P4: Workshops til at informere om bias i data og modeller og om hvilke farer, der kan være.
- Formidling af resultater fra A2 vil blive publiceret på konferencer eller som forskningsartikler til at bidrage til state-of-the-art metoder.

Den største udfordring i denne aktivitetsplan ligger i, at vi foreslår en generel applikation af værktøjer i forbindelse med Machine Learning inden for forskelligartede brancher. Det er en udfordring, da der kan være stor forskel i branchernes datamæssige modenhed, selv om de overordnede teknologier er de samme. Det kan f.eks. komme til udtryk i form af utilstrækkelig historisk dataindsamling i virksomheden eller manglende omstillingsparathed. Dog er indholdet af aktivitetsplanen forholdsvis markedsnært, og hvis det ikke allerede fremgår som en del af virksomhedernes strategi, vil det gøre det inden for 5 år. Af samme grund forventer vi ikke, at det bliver en udfordring at finde virksomheder til case-samarbejde.

### **6) Nyhedsværdi og ambitionsniveau**

Beslutninger baseret på dataanalyse har fået stor opmærksomhed i de seneste år, og der anvendes en bred vifte af beslutningsstøtteværktøjer på tværs af domæner. Udviklingen af værktøjer sker på baggrund af øget regnekraft og tilgængelighed af store datamængder. Deres beslutninger baserer sig ofte på automatiske analyser af Machine Learning (ML)-modeller, hvilket gør det vanskeligt for mennesker at forstå. Den øgede interesse og efterspørgsel skaber således nye udfordringer, når datadreven beslutningsstøtte skal udbredes til flere erhverv og organisationer. Disse udfordringer består i et behov for transparens af ML og datadrevne analysemodeller samt et behov for at forklare deres resultater.

De fleste ML-modeller er ikke transparente i sig selv trods deres brede anvendelse og modenhed. Derfor kræves der nye algoritme- og visualiseringsværktøjer, der kan kombineres for at gøre ML-modeller og store datamængder lettere at fortolke. Emnet får stor opmærksomhed på vigtige Machine Learning konferencer [5-7], men opmærksomheden på disse problemer er stadig forholdsvis ny. Derfor er der behov for at udvikle nye metoder for at gøre ML-modeller mere transparente og forståelige samt at gøre eksisterende metoder lettere at anvende i praksis.

Som GTS-institut har vi de nødvendige kompetencer inden for forskning og anvendelse til at modne eksisterende state-of-the-art metoder på markedet for danske virksomheder. Desuden har Alexandra

Instituttet igennem den tidligere resultatkontrakt om *Teknologier og værktøj til udnyttelse af Big Data* og DABAI opbygget stærke state-of-the-art kompetencer inden for både ML og Visual Analytics. Dette giver et solidt grundlag for at udvikle nye ydelser og teknologier, der understøtter transparent og forklarlig datadreven beslutningsstøtte og yderligere bidrager til den internationale forskning på området.

Konkrete serviceydelser er (relation til problemstillinger fra afsnit 2 er betegnet f.eks. P1):

**S1. Værktøjskasse til transparent Machine Learning:**

- A. Analyseværktøjer, der kan forklare, hvilke features i data, der bidrager i hvilket omfang til et givet resultat. Dette indeholder praktisk anvendelse af state-of-the-art metoder, dvs. efter virksomheders specifikke behov. Men også udvikling af nye værktøjer efter behov (P1, P2).
- B. Præsentation af analyseresultater så de kan forstås af domæneeksperter uden dataanalytisk baggrund (P1, P2).
- C. Analyse og visualisering af indre virkemåder af neurale netværk for at øge transparens heraf (P3).

**S2. Rådgivningskompetencer til udvikling af transparente beslutningsstøtteværktøjer (P1-P3):**

- A. Hvilke teknologier passer til hvilke typer af opgaver? Hvilke teknologier er modne, hvilke er hyldevarer, hvilke kræver tilpasning før anvendelse?
- B. Hvad er trade-off mellem nøjagtighed og transparens? Hvilke metoder (sammen med vores værktøjskasse) skaber både tilstrækkelig transparens og tilfredsstillende resultater?
- C. Hvilke succesfulde case-eksempler findes der på transparens og mulighederne for at forklare resultater? Hvordan kan man øge transparensen på en eksisterende ML-model, der allerede bliver brugt i produktionen?

**S3. Vidensbase om bias og fairness af datadreven beslutningsstøtte (P4):**

- A. Udarbejdelse af E-bøger, artikler eller workshops med henblik på at informere om bias og fairness og hvilke faldgruber, der kan opstå, og hvordan man kan undgå dem.
- B. Udarbejdelse af et white paper, der kan fungere som reference for forholdsregler omkring bias og fairness ved brug af ML til beslutninger.

Det efterfølgende afsnits relation til serviceydelser betegnes f.eks. S1.

Vi forventer, at vi vil udvikle mindst 2 prototyper gennem case-projekter, der benytter sig af både state-of-the-art metoder og nyudviklede værktøjer. Den samlede erfaring og de udviklede værktøjer vil blive en del af værktøjskassen (S1) og rådgivningsydelserne (S2) og derefter kunne anvendes i nye projekter.

Både rådgivningskompetencer og vidensbase om bias og fairness vil supplere Alexandra Instituttets eksisterende kompetencer inden for Big Data og Artificial Intelligence.

## 7) Vidensamarbejde og –hjemtagning

Alexandra Instituttet forventer samarbejde med følgende vidensinstitutioner under aktivitetens forløb:

- Danish National Research Foundation's Centre of Excellence for International Courts (iCourts) i forbindelse med deres forskningsprojekt PACTA, som netop handler om "impact assessment" af datadreven beslutningsstøtte, og hvordan man kan forklare "black-box" modeller (befinder sig i ansøgningsrunde til DFF).
- DHI: I forbindelse med DHIs ansøgning om "Innovativt produkt design for danske virksomheders adgang til det globale marked" er der potentiale for samarbejde. Her ønskes en undersøgelse af potentialet for at anvende ML-teknologier på data, der beskriver sundheds- og miljøbelastning af kemikalier i forskellige produkter. Dette kan kræve transparente løsninger, da beslutninger har kritisk betydning både for producenten og brugeren af produktet.
- Danske universiteter - primært Datalogisk Institut fra Københavns Universitet, DTU og Aarhus Universitet - som vi allerede har eksisterende samarbejder med.
- Relevante innovationsnetværk inddrages i videnspredning og rekruttering af virksomhedsdeltagere i aktiviteterne, og der trækkes på forskere og virksomhedserfaringer herfra, f.eks. DABAI og InfinIT.



- Vidensudveksling ved deltagelse i relevante fora og konferencer og publikationer i tidsskrifter (se også afsnit 5 og 9).

## 8) Sammenhæng med instituttets strategi og afsæt i instituttets ressourcer

Udvikling af nye kompetencer, ydelser og produkter inden for Big Data og Kunstig Intelligens er et punkt i Alexandra Instituttets strategiske plan for de kommende 2 år. Dette ses bl.a. ved Dansk Center for Anvendt Kunstig Intelligens, som for nyligt blev etableret af Alexandra Instituttet. Datadreven beslutningsstøtte ligger i naturlig forlængelse heraf, da det benytter ML og visualiseringsværktøjer til at gøre Big Data, tilgængelige for brugere. Samtidig vil aktivitetsplanen give nye spidskompetencer, der supplerer vores viden inden for ML og visualisering af data. Sammen vil vi få en stærkere portefølje af værktøjer og ydelser inden for Big Data til at hjælpe virksomheder med at bruge deres data.

I den kommende strategi er der desuden planlagt en væsentlig indsats i forhold til at udbygge samarbejde og relationer til flere universiteter (f.eks. AU, KU, DTU, ITU). Vi har specifikt afsat konkrete aktiviteter og milepæle om at samarbejde med universiteter under aktivitetens forløb.

## 9) Tidsplan og milepæle

### MILEPÆLE ÅR 1:

#### Vidensamarbejde, -hjemtagning- og kompetenceopbygning

- P1-P4: Igangsættelse af mindst 1 universitetssamarbejde omkring at gøre det muligt at forklare ML-modellernes resultater og gøre dem mere transparente og deres resultater mere retfærdige.
- Deltagelse i mindst 1 konference om Machine Learning (ML) eller Artificial Intelligence, hvis muligt med fokus på en eller flere udfordringer (P1-P4).
- Identifikation af mindst 1 virksomhed og opstart af case-forløb med denne baseret på state-of-the-art metoder (P1-P4). Case-samarbejdspartnere kan være en virksomhed og/eller en kommune/region.

#### Udvikling af teknologisk service

- Primo år 1: første version af værktøjskasse af state-of-the metoder og værktøjer (S1), der kan:
  - P1 og P2: Forklare resultaterne af ML-modeller.
  - P3: Gøre datadreven beslutningsstøtte mere transparente.
- Påbegyndt udvikling af rådgivningsydelse S2 baseret på input og feedback fra brugerstudier, følgegruppen og første erfaringer fra case-forløb.
- P4: Første version af vidensbase (S3) over fairness i data og datadrevne beslutningsmodeller.
- Ultimo år 1: Opdateret version af værktøjskassen (S1) i forhold til brugbarhed af state-of-the-art metoder baseret på case-forløbet.

#### Inddragelse og videnspredning

- Opstart af følgegruppe bestående af relevante målgruppevirksomheder, inklusive 1. møde i følgegruppen.
- Brugerstudier hos 3-4 virksomheder mhp. at samle op på behov og udfordringer.
- 1 workshop som har til formål at identificere fælles problemer og udfordringer – også i forhold til state-of-the-art metoder. Resultater af dette vil give input til rådgivningsydelse S2 og S3.
- Første erfaringer ift. problemstillingerne distribueres i mindst 1-2 fagrelevante medier eller fora.

### MILEPÆLE ÅR 2:

#### Vidensamarbejde, -hjemtagning- og kompetenceopbygning

- Deltagelse i mindst 1 relevant konference eller fora.
- Opstart af yderligere mindst 1 case-forløb, såfremt den ikke er opstartet første år, som benytter sig af state-of-the-art metoder eller de nye udviklede metoder. De markedsnære teknologiske ydelser i år 2 inddrages og afprøves løbende i case-forløbene med henblik på tilpasning.

#### Udvikling af teknologisk service

- P1-P3: Videreudvikling af værktøjskassen (S1) med nye metoder, der supplerer state-of-the-art metoder (baseret på erfaringer fra 1 case-forløb, workshops og følgegruppen i år 1 og 2).

- P4: Opdateret vidensbase (S3) i forhold til bias suppleres af erfaringer fra workshops og case-forløb.
- Udvikling af mindst 1 ny teknologi (metode, teknik, værktøj) under problemstilling P1-P3 gennem case-samarbejder (S1).
- Videreudvikling af rådgivningsydelse S2.

#### **Inddragelse og videnspredning**

- Identifikation af yderligere mindst 1 virksomheder til case-forløb og opstart af denne.
- Årligt følgegruppen møde.
- Opdaterede erfaringer distribueres i 1-2 fagrelevant medium eller som foredrag i relevante fora.
- Der udgives mindst 1 forskningsartikel gennem en konference eller tidsskrift.
- Minimum 1 arrangement gennem et innovationsnetværk eller som gå-hjem-møde for at informere om erfaringerne samlet i år 1 og nyudviklede metoder.

#### **Referencer**

- [1] Gartner, 100 Data and Analytics Predictions Through 2021 <https://www.gartner.com/doc/3746424>
- [2] Datadreven Beslutningsstøtte, Bedreinnovation.dk, <https://bedreinnovation.dk/datadreven-beslutningsst%C3%B8tte-%E2%80%93-transparent-og-forklarlig-dataanalyse-til-alle>
- [3] <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>, ProPublica, May 2016
- [4] T. Ritter, et al., DataProfit: Kompetencekort for Datadreven Vækst, CBS, april 2017, ISBN: 978-87-93226-19-7
- [5] Workshop on Explainable Artificial Intelligence (<http://home.earthlink.net/~dwaha/research/meetings/faim18-xai/>)
- [6] Fairness, Accountability and Transparency in Machine Learning (<https://www.fatml.org/>)
- [7] Workshop on Human Interpretability in Machine Learning (<https://sites.google.com/view/whi2018/home>)