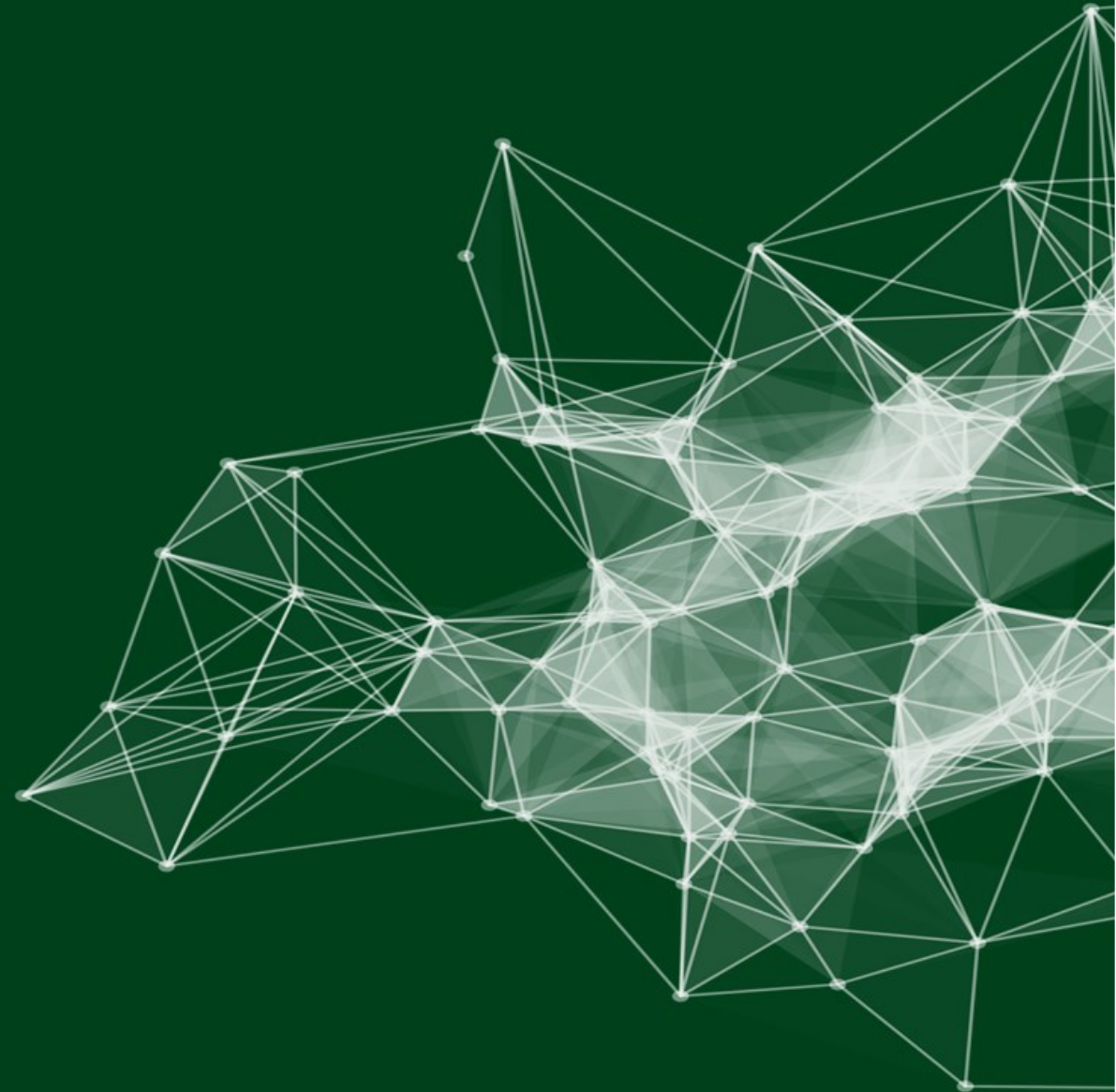


Kortlægning af analytics i staten

Oktober 2018



Forord og læsevejledning

Velkommen til Moderniseringsstyrelsens rapport om analytics i staten

Som led i Moderniseringsstyrelsens arbejde med at inspirere andre myndigheder til at arbejde mere datadrevet med deres kerneopgave, har vi i foråret 2018 gennemført en kortlægning af analytics i staten på baggrund af interviews med 25 institutioner fra 10 ministerområder. Vi har talt med direktører, kontorchefer og specialister, og besøgt diverse styrelser, departementer og andre typer institutioner på tværs af landet. Rapporten udgør vores udlægning af møderne, men størstedelen af institutionerne har bidraget med mødereferater og løbende opfølgninger, som grundlag for analysen. Stor tak til alle for deres hjælp og engagement!

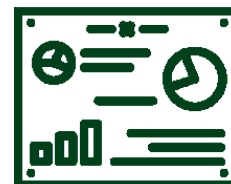
I denne rapport præsenterer vi de samlede resultater for undersøgelsen. Vi har opdelt rapporten efter tre centrale temaer/sektioner: *Analytics*, *Datamiljø* samt *Ledelse og kompetencer*. I hvert tema indgår to kapitler, og hvert kapitel behandler en række emner. I den første sektion definerer vi analytics og belyser statens erfaringer og potentialer inden for analytics. I den anden sektion beskriver vi datatyper, og hvordan data er organiseret i staten. I den tredje og sidste sektion undersøger vi den digitale lederrolle, projektstyring og kompetencer inden for analytics.

Rapporten kan læses fra ende til anden, men man kan også dykke ned i enkelte kapitler og læse dem uafhængigt af den samlede rapport. På næste side er en indholdsfortegnelse over hele rapporten. Vi indleder rapporten med et kapitel om formål og baggrund samt et sammenfattende ledelsesresumé.

Vi håber, at vi med rapporten kan inspirere myndigheder til at undersøge mulighederne for en bedre løsning af kerneopgaven gennem analytics.

Rigtig god læselyst!

Analytics

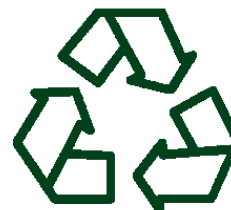


Hvad er analytics?



Hvilke erfaringer og potentialer har staten?

Datamiljø



Hvilke data har staten?



Hvordan organiserer og styrer staten data?

Ledelse og kompetencer



Hvilken rolle spiller statens ledelse?



Hvordan udvikles analytics?

Indholdsfortegnelse. Kortlægningen indeholder fire sektioner og otte kapitler



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

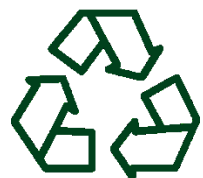
Definition af analytics og dets begrebsverden – hvad er analytics..... s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten..... s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten..... s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8

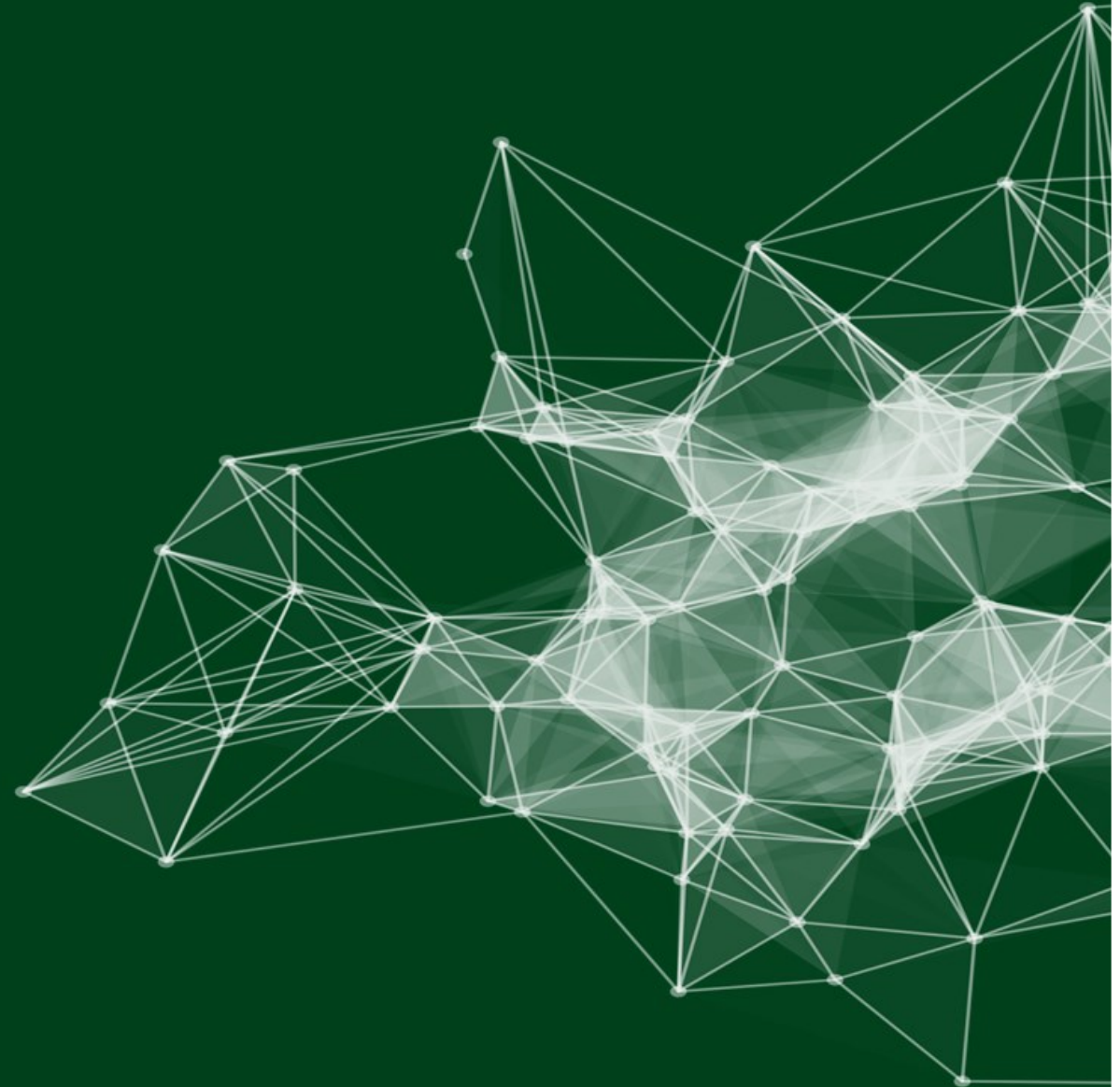


Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

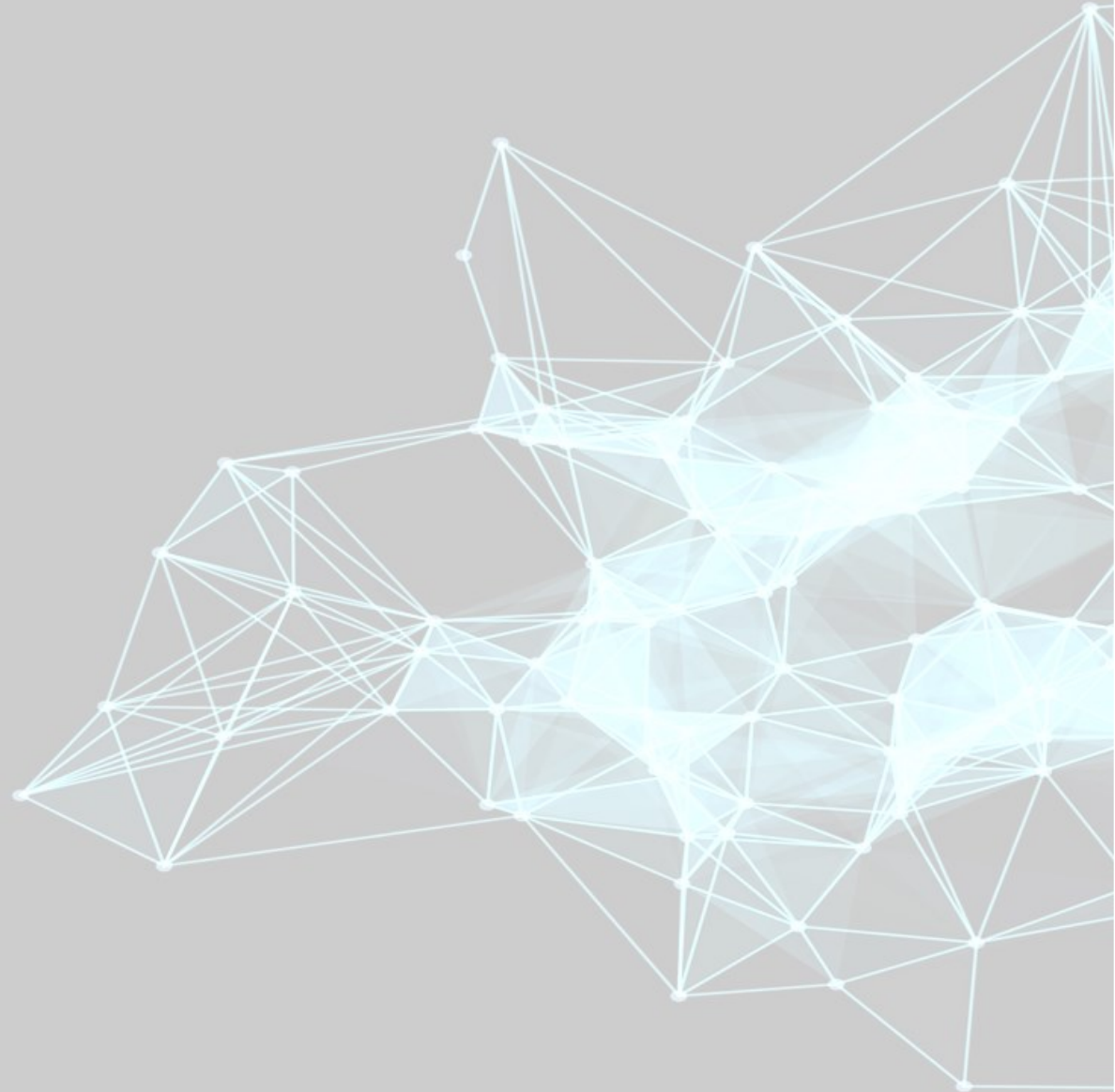
Sektion 1

Introduktion



Kapitel 1

Formål og baggrund



Moderniseringsstyrelsen sætter fokus på analytics i staten



Moderniseringsstyrelsens pejlemærke: Analytisk kraftcenter

I Moderniseringsstyrelsen har vi fokus på, at arbejde datadrevet i forhold til at udvikle den offentlige sektor. Vi anvender data og viden til analyser, nye indsigter og velfunderede beslutningsoplæg, men understøtter og inspirerer også andre myndigheder til selv at arbejde datainformeret med deres kerneopgaver. Som led i dette, sætter vi og Finansministeriet fokus på mulighederne og udfordringerne ved analytics.

Arbejdet er forankret i *Center for Modernisering og Analyse* under kontorchef Michael Sperling.



Kortlægning af analytics i staten

Hvad er machine learning? Hvordan adskiller det sig fra økonometri? Er det ikke bare business intelligence? Hvad er koblingen til big data? Spørgsmål som disse stiller flere, og der er generelt stor usikkerhed om, hvad analytics er, og ikke er, både inde i og uden for staten. Formålet med kortlægningen er derfor, at definere et fælles koncept for analytics, og på den baggrund rapportere statens *as-is* erfaringer og *to-be* potentialer inden for analytics.

Vores fire delmål

Kortlægning af analytics i staten



- Rammesætte en fælles forståelse af analytics i staten
- Skabe et samlet overblik over erfaringer og potentialer på tværs af staten
- Belyse centrale problemstillinger, udfordringer og løsninger inden for analytics
- Inspirere til udviklingen af eksisterende og nye analytics løsninger og tværgående samarbejde

Kortlægningen er gennemført i foråret 2018 med en bred inddragelse af 25 institutioner



Hvornår?

Kortlægningen er gennemført i foråret 2018, hvor interviews blev afholdt i perioden fra januar til marts, mens analysen blev udarbejdet fra april til juni.



Hvordan?

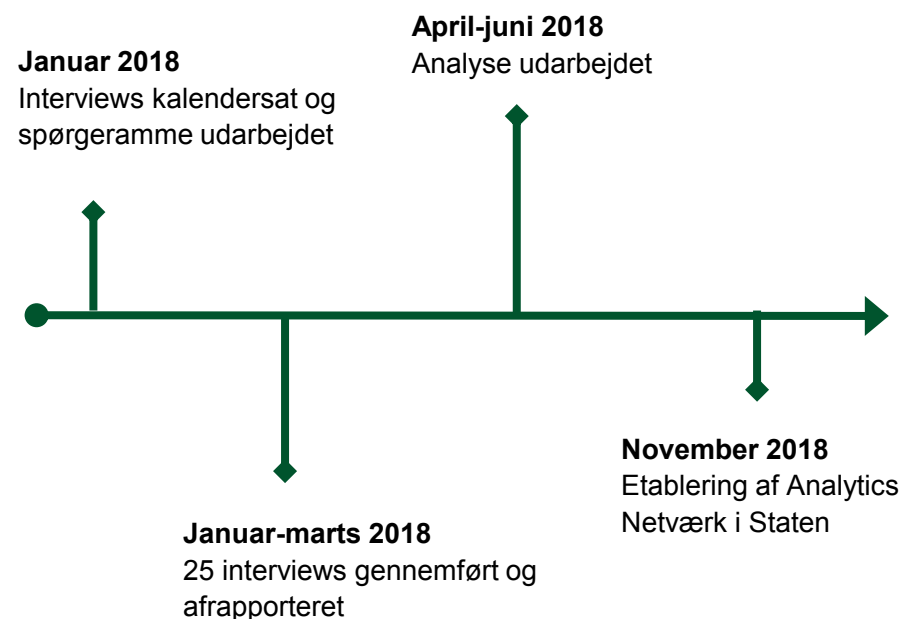
Kortlægningen har taget udgangspunkt i en eksplorativ og åben tilgang med henblik på at indfange institutionernes konkrete erfaringer og synspunkter inden for analytics. Der blev udarbejdet en semi-struktureret interviewguide med enkelte på forhånd identificeret temaer, men med stor grad af åbenhed over for at følge interviewpersonernes betragtninger og særlige opmærksomheder. Der blev identificeret 30-40 relevante institutioner/kontaktpersoner på baggrund af organisationsdiagrammer og tidligere erfaringer, hvoraf 25 stillede op til ca. 1-2 timers interview. Alle interviews er afrapporteret i notater i samarbejde med de fleste institutioner.



Hvem?

Kortlægningen er gennemført af Moderniseringsstyrelsens analysekontor, Center for Modernisering og Analyse.

Tidslinje



25 institutioner fra 10 ministerområder



Størstedelen af de deltagende institutioner er styrelser, men der er også enkelte departementer og særegne institutioner



Mangfoldig og varieret interviewgruppe

De interviewede institutioner adskiller sig på flere parametre, og sikrer derved et solidt og varieret datagrundlag for at kortlægge statens erfaringer og potentialer med analytics i staten. De 25 interviewede institutioner fordeler sig på 10 ministerområder, varierer fra 60 til +2000 medarbejdere fordelt over hele landet, og har meget forskellige kerneopgaver, der strækker sig fra vedligehold af det danske elnet over tilsyn af fødevarer til udvikling af beskæftigelsesindsatser.

Der har deltaget 20 styrelser, to departementer, ét selvstændigt statsligt selskab ved Energinet, én forsknings- og rådgivningsinstitution ved GEUS, og én selvejende institution ved Anti Doping Danmark.

Interviewpersonerne har været både direktører og kontorchef.

Overblik

Institutionstyper



Styrelser (20)

Arbejdsstilsynet, Styrelsens for arbejdsmarked og rekruttering, Fødevarestyrelsen, Miljøstyrelsen, Landbrugsstyrelsen, Konkurrence- og Forbrugerstyrelsen, Erhvervsstyrelsen, Sikkerhedsstyrelsen, Nævnenes Hus, Finansstilsynet, Søfartsstyrelsen, Lægemiddelstyrelsen, Styrelsen for Patientsikkerhed, Styrelsen for It og Læring, Styrelsen for Forskning og Innovation, Banedanmark, Trafik-, Bygge- og Boligstyrelsen, Styrelsen for Dataforsyning og Effektivisering, Slots- og Kulturstyrelsen og Rigsarkivet

Departementer (2)

Beskæftigelsesministeriet samt Børne- og Socialministeriet

Særegne institutioner (3)

Energinet, De Nationale og Geologiske Undersøgelser for Danmark og Grønland (GEUS) samt Anti Doping Danmark.



Kapitel 2

Ledelsesresumé

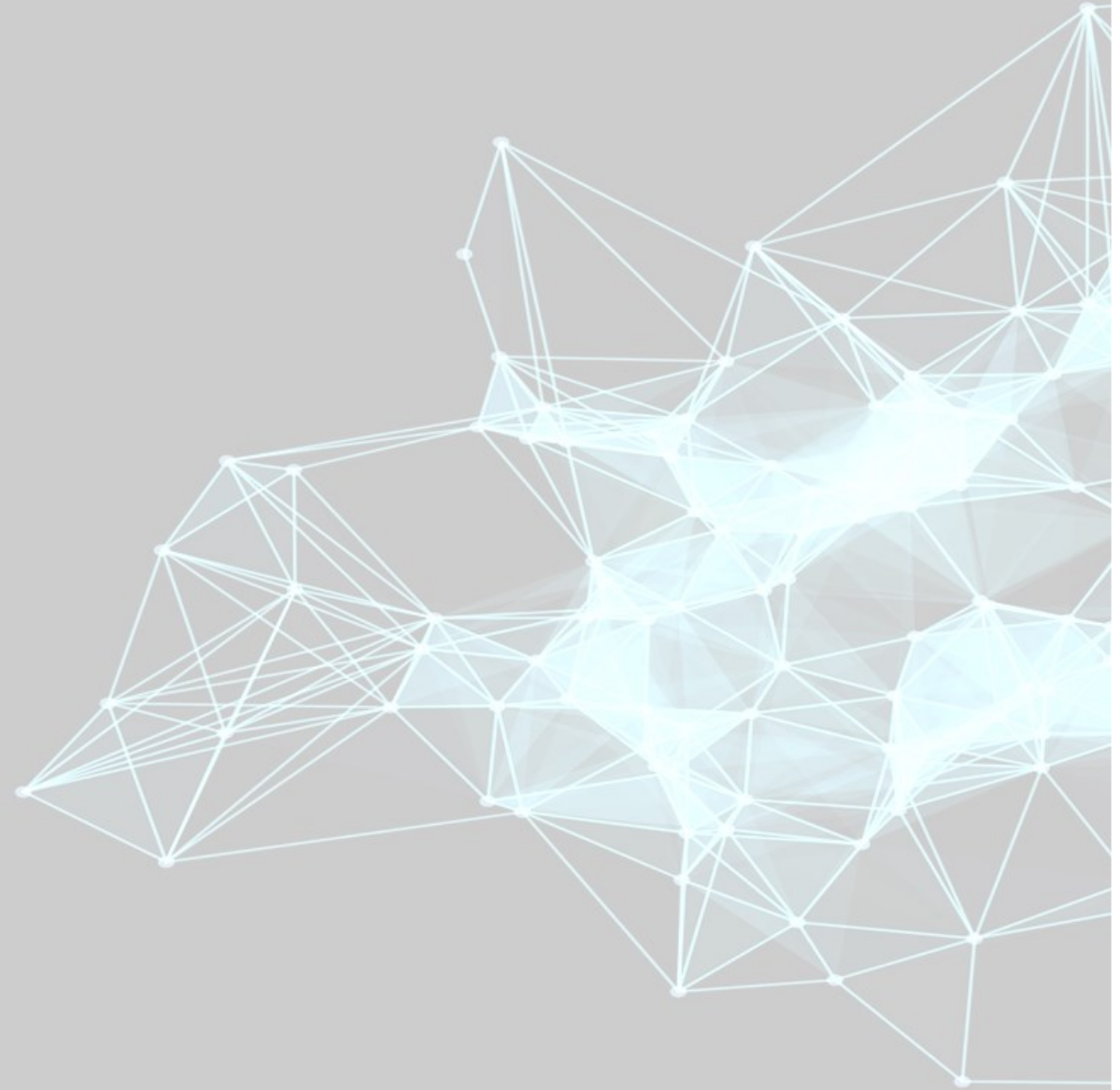
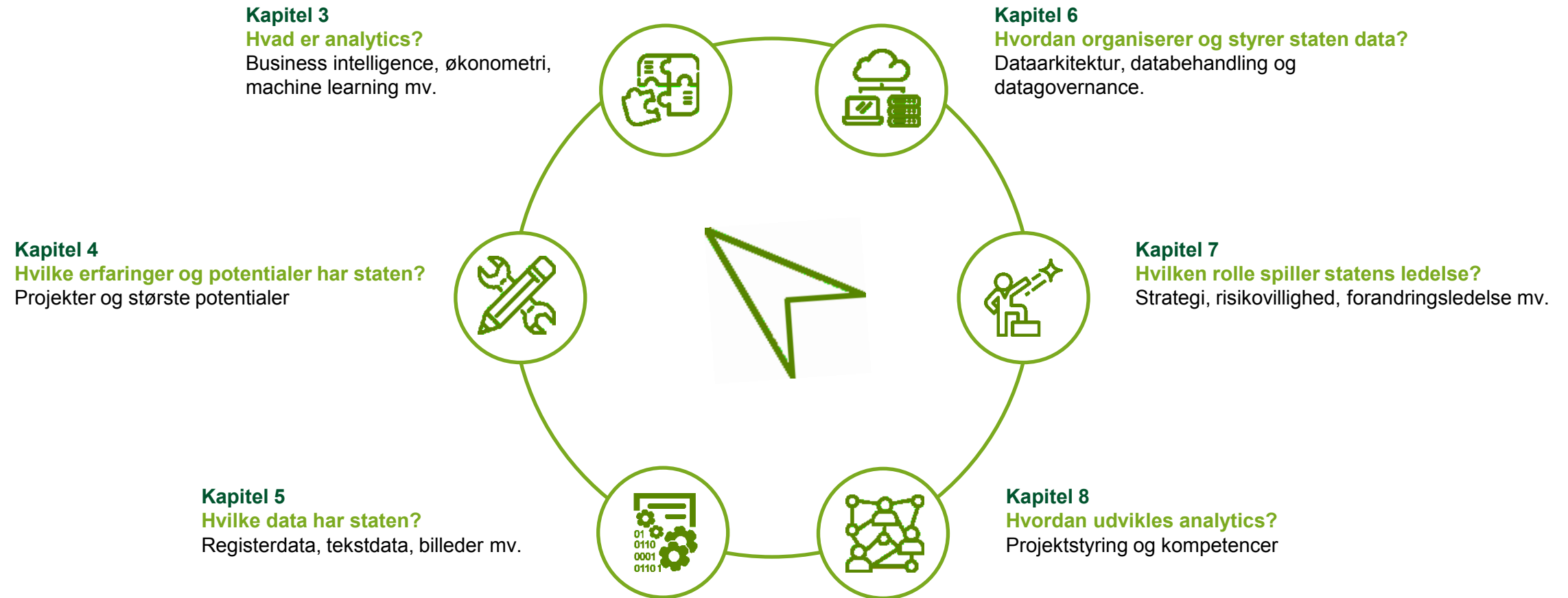


Illustration: Undersøgelsens kapitler og fokusområder



Hovedkonklusioner: Definition af analytics og dets begrebsverden



Hovedkonklusioner	Detaljer
Vi definerer analytics, som et begrebsfænomen, der dækker over fire analytiske hovedområder, som har til formål at forbedre et givent beslutningsgrundlag. Informationssystemer. Business intelligence er automatisk bearbejdelse og udstilling af data, der skal give ledere et overblik over centrale aspekter af forretningen. Opgavespecifik information giver omvendt medarbejdere adgang til og/eller udstiller information, der er relevant for udførelsen af en konkret opgave. Det kunne eksempelvis være tilsynsførende, der skal aflægge besøg hos en virksomhed.	s. 24-26
Økonometri. Via statistiske modeller søger økonometri at teste hypoteser om sammenhænge mellem uafhængige variable og én afhængig variabel. Det kunne eksempelvis være en hypotesetest af effekten på en afgiftsstigning på cigaretter: <i>Hæves prisen på cigaretter, har det så en negativ effekt på forbruget af cigaretter?</i> Machine learning og økonometri kan bygge på det samme matematiske grundlag, men machine learning fokuserer på outcome og mindre på samvariation og statistiske antagelser.	s. 27+29
Machine learning. Dette dækker over et hav af metoder og tilgange til data, men fælles for disse er, at de indeholder et læringselement i form af enten supervised eller unsupervised learning. Predictive modelling er eksempelvis en metode, der på baggrund af historiske data kan foretage supervised learning af en algoritme til kvalificeret forudsigelse af udfald. Det bliver blandt andet brugt af banker til kreditvurdering af lånsøgere til huskøb. Klyngeanalyser er omvendt en metode, der dækker over et bredt spektrum af algoritmer, der via unsupervised learning kan inddele data i klynger på baggrund af komplekse mønstre i data. Fredericia Bibliotek har eksempelvis brugt klyngeanalyse til at identificere særlige profiler på deres biblioteksgængere mhp. at målrette bibliotekets tilbud mod disse profiler. Tekst analytics kan derimod anvendes til grammatisk og semantisk kvantificering af tekstmateriale, der muliggøre automatisk bearbejdelse af tekst og efterfølgende analyser på datamateriale. Et eksempel på anvendelse af tekst analytics er identificering og frasortering af spammails.	s. 28-32
Matematisk optimering. Dette dækker over mange metoder til at optimere et bestemt output fx inden for vagtplanlægning og opgaveallokering. I shipping bliver matematisk optimering brugt til at optimere mængden af varer, der kan leveres over hele kloden og særligt i sejlads har det haft kæmpe indflydelse.	s. 33

Hovedkonklusioner: Erfaringer og potentialer inden for analytics

Kapitel 4
Hvilke erfaringer og
potentialer har
staten?



Hovedkonklusioner

Detaljer

Vi vurderer, at statens erfaringer med analytics er størst ved business intelligence og økonometri, og potentialet størst ved matematisk optimering og machine learning.



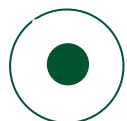
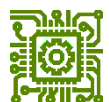
Informationssystemer. Business Intelligence er vidt udbredt i staten og bliver brugt til at træffe bedre beslutninger med udgangspunkt i data. Søfartsstyrelsen producerer fx BI-rapporter til chefer og direktionen mhp. på at træffe oplyste beslutninger. Opgavespecifik information bliver særligt brugt til at udstille information til tilsynsførere, hvilket muliggør et kvalificeret og effektivt tilsynsbesøg. Anti Doping Danmark udstiller eksempelvis information om fitnesscentrets historik, medlemmer mv. til kontrollører.

s. 39-41



Økonometri. Økonometri bliver i stort omfang brugt til evalueringer af indførte (ex post) og foreslåede (ex ante) politikker. Et eksempel på en ex post evaluering, er Beskæftigelsesministeriets evaluering af førtidspensionsreformens effekt for udsatte unge. Derudover bruges økonometri til at opnå større indsigt i et konkret fagområdes mekanismer eller sammenhænge, dvs. hvilke variable har betydning for et givent outcome. Styrelsen for It og Læring benytter fx økonometriske modeller til at analysere hvilke faktorer, der påvirker elevers sandsynlighed for at få en ungdomsuddannelse.

s. 42



Machine learning. Der er identificeret seks forretningsområder, hvor institutionerne har gjort sig erfaringer med eller ser potentiale i at anvende machine learning: 1) risikobaseret tilsyn, 2) anbefalingsværktøjer, 3) tendensanalyser, 4) kundesupport, 5) effektiv sagsbehandling og 6) datakvalitet. Der er størst erfaring/kendskab med de to første forretningsområder. Et eksempel herpå er risikobaseret tilsyn med data fra mange kilder, hvor en machine learning algoritme trænes til at vurdere risikoen for, at en virksomhed eller institution ikke overholder reglerne. Idéen er vidt udbredt, men erfaringerne få, på grund af juridiske barrierer og det såkaldte "årsagsforklaringsproblem". Arbejdstilsynet og Syddansk Universitet foretager et toårigt projekt, der undersøger mulighederne i at anvende økonometri og/eller machine learning til risikobaseret udvælgelse til tilsyn.

s. 43



Matematisk optimering. Dette er ikke særlig udbredt blandt de besøgte institutioner, på trods af et stort potentiale inden for ressourceoptimering med høj kombinatorik, eksempelvis vagtplanlægning og opgaveallokering. Opgaveallokering er et spørgsmål om, hvem der skal løse hvilken opgave. Dette bliver fx komplekst hos tilsynsinstitutioner når opgaver skal allokeres efter kompetencer, geografisk placering, estimeret besøgstid, risikovurdering, tidsfrister mv. I Sikkerhedsstyrelsen er dette en problematik, da opgaverne er spredt udover hele landet og kompetencekravene til opgaverne er meget forskellige.

s. 44-51

Erfaring Potentiale



Hovedkonklusioner: Datatyper



Hovedkonklusioner	Detaljer
Vi har identificeret en række forskellige data i staten, og klassificeret dem i følgende fem overordnede datatyper. Struktureret data. De fleste institutioner producerer eller anvender rigtig meget struktureret data, og det spænder fra registerdata (fx adresseregistre) over tekniske data (fx vandmiljøets tilstand) til administrative data (fx økonomi).	s. 58
Ustruktureret tekstdata. Ikke overraskende producerer, modtager og opbevarer institutioner en del ustruktureret tekstdata i word- og pdf-filer. Det kunne fx være klagesagsafgørelser (fx klager over sundhedsfaglig behandling), tilsynsrapporter (fx elinstallation) og produktgodkendelser (fx miljøkrav).	s. 59
Billeddata. Flere institutioner arbejder med forskellige billeddata. Det spænder fra fotos af planter over elledninger til fly- og satellitbilleder til produktion af fx landkort og kortlægning af oversvømmelsesrisiko.	s. 60
Lyddata. Enkelte institutioner arbejder med lyddata. Store dele af den moderne kulturarv er formidlet i lydformat i form af radio og video, og for at gøre materialet mere tilgængeligt er Rigsarkivet i gang med at "tagge" materialet digitalt. Lyddata kan også anvendes til at identificere vedligehold, hvor blandt andet Energinet vil undersøge, om lyde kan blive anvendt til at identificere vedligehold på transformatorstationer.	s. 61
Big data. Flere institutioner ser et analytisk potentiale i big data fra sociale medier, sensorer og andre datakilder. Der er fx perspektiver i at anvende unikke danske sundhedsregistre, sociale medier og real world data fra wearables mv. til at kunne gennemføre hurtigere godkendelser af nye lægemidler til livstruende sygdomme. Sociale medier vil også kunne bruges til at forudsige legetøjstrends eller til webscraping for at spore karteldannelser. Viden om kundernes trafik på hjemmesider kan forbedre kundesupporten, og mobildata vil kunne bruges til at estimere antal besøgende i fx slotsparker. Fremkomsten af big data og mulighederne ved analytics skaber et eksternt pres fra omverdenen på de offentlige myndigheders opgavevaretagelse. For det første en stigende forventning om, at sagsbehandlingen kan og skal gennemføres hurtigere. Miljøstyrelsen har eksempelvis afsøgt mulighederne for at anvende tekst analytics til at reducere sagsbehandlingstiden på produktgodkendelser ift. miljøkrav. For det andet medfører udviklingen af nye former for big data nye krav til, hvordan de offentlige myndigheder fører tilsyn. Derivathandler i den finansielle sektor foregår fx i et tempo og omfang, der kræver at Finanstilsynet anvender ny teknologi for kunne føre et effektivt tilsyn.	s. 62-64

Hovedkonklusioner: Datamanagement



Hovedkonklusioner

Detaljer

Vi har identificeret tre væsentlige forhold i organiseringen af data, der vedrører hhv. dataarkitektur, databehandling og datagovernance.



Dataarkitektur. Her er et væsentligt element etableringen af databaser, hvor data fra fagsystemer kan lagres sikkert og struktureret. Flere institutioner har en eller flere databaser lokalt, der typisk bliver udviklet og implementeret af konsulenter eller Danmarks Statistik (DST), men generelt driftes af institutionerne selv. Ansvar for driften af disse overdrages til Statens It frem til 2021. Enkelte institutioner bruger cloud computing til at opbevare data i skyen, herunder fx Erhvervsstyrelsen. Cloud computing er ifølge enkelte institutioner nødvendigt for at kunne sikre teknisk og forretningsmæssig fleksibilitet i anvendelsen af analytics, og bør prioriteres i forbindelse med samlingen af statens basale it-drift hos Statens It. En gruppe institutioner samler og tilgår deres og andres data hos DST via den såkaldte ministerieordning. Enkelte institutioner beskriver, at ministerieordningens aftalegrundlag kan være problematisk, idet hele institutionens adgang til data bliver lukket i en længere periode, hvis en medarbejder begår en fejl i strid med DSTs fortrolighedsregler.

s. 67-70



Databehandling. Dette vedrører den menneskelige behandling af data, snarere end dataarkitekturs mere tekniske fokus. Databehandling dækker m.a.o. over databearbejdelsen fra rådata til det endelige data, der skal bruges til analyser i organisationen. Flere institutioner anfører, at databehandling er en ressourcekrævende opgave, men at strukturering og rensning af data giver værdifuld indsigt i organisationens datalandskab, og dermed nye idéer til analyser. Det er uklart, præcist i hvilket omfang institutionerne bruger konsulenter til databehandling, men visse dele af disse opgaver bliver hos nogle institutioner foretaget af konsulenter eller af DST. Store dele af databehandlingsopgaverne er gentagende processer, hvor det samme rådata skal gennemgå samme behandling og formatering som tidligere rådata.

s. 71-73



Datagovernance. Dette vedrører regler og procedurer for forvaltning af data internt og eksternt i institutionen, hvor et væsentligt element er at behandle data i henhold til databeskyttelsesregler (GDPR), men også sikre ensartet og konsistente definitioner af data med henblik på at forbedre datagrundlaget, og herved øge mulighederne for at finde nye tværgående indsigter i data. Flere institutioner har udfordringer med at løfte organisationens datagovernance, med den konsekvens at potentialet af data ikke udnyttes optimalt. En mulig løsning kunne være indførelse af såkaldte "data stewards", der skal sikre, at styrelsens datastandarder bliver fulgt på tværs af organisationen. En anden mulig løsning er anvendelsen af robotautomatisering (RPA), der kan bruges til at sikre en ensartet registrering af data, der er fri for menneskelige tastefejl. Institutioner er påpasselige med at dele og modtage data på tværs af det offentlige på grund af usikkerhed om den juridiske ramme i lyset af databeskyttelsesforordningen, men også i lyset af den tiltagende samfundsdebat og bevågenhed om håndteringen af personoplysninger. Det står uklart for flere, hvilke data der må samkøres.

s. 74-78

Hovedkonklusioner: Digital ledelse



Hovedkonklusioner

Detaljer

De fleste institutioner har strategisk fokus på dele af analytics, men der er stor forskel på, hvor langt institutionerne er i deres visioner. Vi har identificeret en række centrale ledelsesperspektiver inden for analytics.



Strategisk fokus. Lederen skal være i stand til at se, hvordan "det digitale" understøtter organisationens kerneopgaver, men skal i lige så høj grad kunne se, hvordan det kan understøtte opgaver, som organisationen ikke løser i dag. Idéerne kommer nogle gange fra topledelsen, andre gange fra udviklingsenheder og i visse tilfælde fra enkelte medarbejdere. Flere institutioner ser potentialer ved machine learning og matematisk optimering, men afventer implementering til løsningerne er mere modne. Men har staten et ansvar som "first mover" med henblik på at modne teknologien? Endeligt er der enkelte institutioner, der har et særligt fokus på etiske dilemmaer i anvendelsen af data og kunstig intelligens, hvor debatten nåede helt nye højder efter Facebook skandalen med Cambridge Analytica blev afsløret i marts 2018.

s. 85-89



Brændende platform. Digitalisering har medført en række muligheder, men også udfordringer, som analytics kan bidrage til at håndtere. Flere institutioner beskriver en brændende platform. Erhvervsstyrelsen forklarer eksempelvis, at det er en udfordring at fastholde et effektivt tilsyn, når danske virksomheder kan oprette sig selv på fem minutter og dermed øger den daglige sagsmængde. Finanstilsynet har også udfordringer med at løfte deres tilsynsopgave på grund af omfanget og hastigheden af nye typer finansielle transaktioner.

s. 90



Risikovillighed. Flere institutioner fremhæver, at projekter inden for analytics kræver et særligt fokus fra topledelsen og en risikovillighed i projektinitieringen til at prioritere udvikling over produktion. Der er dog delte meninger herom. Nævnenes Hus forklarer, at de er villige til at prioritere deres udviklingsprojekter, herunder machine learning, over den daglige drift af kerneopgaven, idet gevinsten ved løsningen overstiger de midlertidige tab ved en mindre stabil drift. Landbrugsstyrelsen forklarer omvendt, at analytics projekter ikke må gå ud over kerneopgaven.

s. 91



Faglighed. Der er to elementer her. For det første beskriver enkelte institutioner, at ledere skal kunne drøfte grundlæggende metodiske og datamæssige valg for at kunne indgå i en faglig dialog om indholdet og retningen for analytics projekter. For det andet beskriver flere institutioner, at ledelsen skal kunne begrunde, hvorfor machine learning udpeger enkelte frem for andre – "årsagsforklaringsproblemet" – samt forsvare modellens risiko for fejl over for borgere og virksomheder.

s. 92-93



Forandringsledelse. Forandringer af medarbejderes arbejdsrutiner og kompetencekrav kræver en særlig ledelsesmæssig indsats med henblik på at sikre "buy in" og undgå modstand blandt medarbejdere, og dette fremhæves af enkelte institutioner som helt afgørende. Ledelse af forandring handler blandt andet om at vælge de rette tidspunkter for større ændringer, og her ser enkelte institutioner udflytning som et mulighedsrum.

s. 94-95





Hovedkonklusioner: Projektorganisering og kompetencer

Hovedkonklusioner

Detaljer

Vi har identificeret tre centrale elementer ift. projektstyring inden for analytics, og vurderet en række relevante kompetencer til udvikling af analytics.



Projektstyring. Institutioner fremhæver tre centrale elementer. 1) Der skal tages stilling til **projektets forankring og ressourcer**. Enkelte institutioner har valgt at forankre deres analytics projekter i særlige analytics teams, som fx Energinets team "Advanced Analytics". 2) Der skal tages stilling til projektets **metodiske tilgang**. Projektet kan fx følge en mere fokuseret "proof of concept"-tilgang med vægt på at afprøve idéer hurtigt og i lille skala mhp. at vurdere analytics løsningens realiserbarhed, økonomiske og kvalitative potentiale. Enkelte institutioner har positive erfaringer med at gennemføre "proof of concept", herunder særligt Erhvervsstyrelsen og Energinet. 3) Projektet skal undersøge **implementeringsudfordringer**. Flere institutioner ser idriftsættelse af analytics fra laboratorium til virkelighed som den sværeste opgave snarere end udviklingen af algoritmer mv., og beskriver en række tekniske og menneskelige udfordringer.

s. 99-102



Digital ledelseskompetencer. Ledelsen har brug for opkvalificering. Flere institutioner ser muligheder ved analytics, men har svært ved at konceptualisere deres visioner. Institutionernes ledelse har brug for at styrke deres evner til at indgå i en faglig dialog om indholdet og retningen af analytics projekter og diverse dataetiske problemstillinger. Institutionerne giver udtryk for, at det er svært at praktisere ledelsesmæssig risikovillighed i en offentlig kontekst, hvor udvikling af kerneydelserne ikke må ske på bekostning af stabil drift og service. Blandt institutionerne er der en lav grad af erfaring med forandringsledelse inden for analytics, da de fleste gennemførte projekter har været "proof of concept".

s. 103-104



Analytics kompetencer. De fleste institutioner har medarbejderprofiler inden for business intelligence, opgavespecifik information og/eller økonometri, der er i stand til at gennemføre deskriptive analyser af data såvel som grundlæggende regressionsanalyser ud fra en hypotetisk deduktiv tilgang. Det er typisk økonomer og erhvervsøkonomer eller fagspecialister, der er BI-autodidakte. De færreste institutioner har medarbejderprofiler inden for machine learning og matematisk optimering, som kan udvikle og træne algoritmer eller opbygge avanceret matematiske modeller til optimering af en konkret funktionsopgave. Enkelte institutioner har ansat "data scientists" (fysikere, ingeniører, mv.)

s. 105



Datamanagement kompetencer. Institutionerne har generelt få medarbejdere med kompetencer inden for dataarkitektur. Datavarehuse bliver typisk udviklet og implementeret af konsulenter, men driftes af institutionerne selv eller Statens It. Institutionernes kompetencer inden for databehandling er uklart, men bliver delvist løftet af konsulenter og Danmarks Statistik. Flere institutioner er ved at styrke deres datagovernance.

s. 106

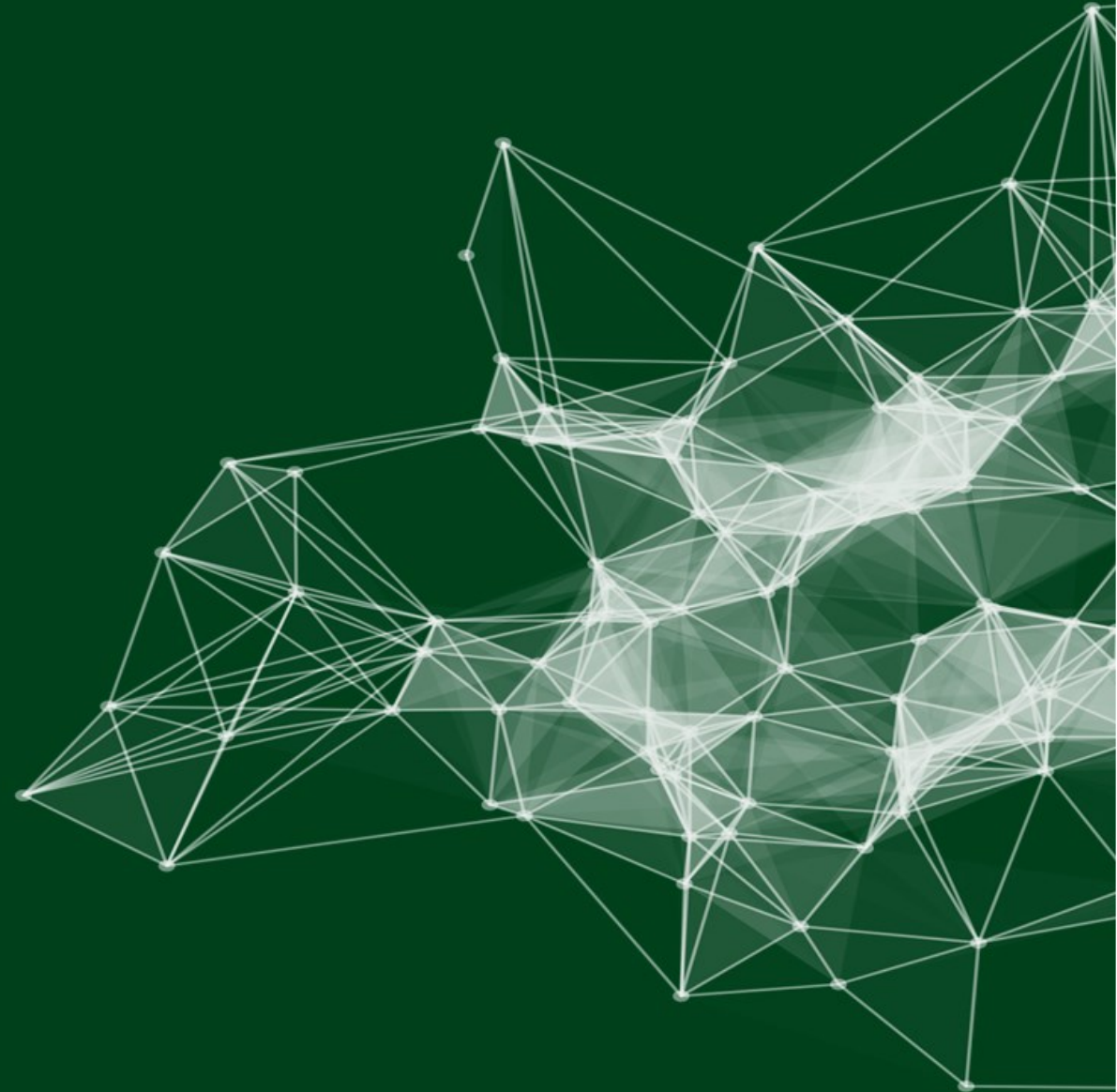


Forretningskendskab. Institutionernes medarbejdere er eksperter på deres fagområde, men skal oplæres ift. mulighederne ved analytics og projektarbejde.

s. 107

Sektion 2

Definition og erfaringer



Introduktion: I denne sektion definerer vi analytics og beskriver statens erfaringer og potentialer inden for analytics



Hvad er analytics? Hvilke erfaringer og potentialer er der i staten?

Analytics opfattes af mange som et tomt begreb, fordi det bliver brugt som et paraplybegreb for alle former for dataanvendelse og analysemetoder.

Vi definerer og rammesætter derfor begrebet i kapitel 3. Vi kommer ind på informationssystemer, økonometri, machine learning og matematisk optimering. Vi går i dybden med særligt machine learning, der for mange opleves som en "black box".

Der vil givetvis være forskellige indvendinger imod vores definition af analytics. Dem hilser vi velkommen. Hovedformålet med vores udlægning er at rammesætte statens fremtidige diskussioner om analytics, og anvende et fælles sprog om meget komplekse problemstillinger.

I kapitel 4 vender vi fokus mod statens status med erfaringer og potentialer inden for analytics. Det har ikke været muligt at præsentere alle 25 institutioners konkrete projekter, men vi har fremhævet de som passede bedst i vores rammesætning. Hovedvægten er på machine learning.

Kapitel 3

Hvad er analytics?



Definition. Vi definerer begrebet analytics og beskriver dets særlige kendetegner.

Kapitel 4

Hvilke erfaringer har staten?



Erfaringer og potentialer. Vi gennemgår statens erfaringer med analytics.

Baggrund: Det tomme begreb "analytics" og nyttiggørelse af data



Det tomme begreb

Hvad er analytics? Der findes mange svar på dette, og vi har mødt flere forskellige udlægninger fra de besøgte institutioner. Dækker det over machine learning metoder? Er business intelligence en del af analytics? Hvad er forskellen til økonometriske analyser? Og så videre...

Analytics er på mange måder et tomt begreb, fordi det anvendes som et paraplybegreb for dataanvendelse, og fordi der foretages implicitte afgrænsninger, afhængig af hvem man spørger.

Begrebsforvirringen forstærkes yderligere af it-industriens kommerialisering heraf, men også af at flere underliggende begreber går på tværs af flere analysemetoder, herunder fx lineær regression, der bruges både i økonometri og i machine learning. De bruges dog på to vidt forskellige måder, og disse vigtige og komplicerede nuancer bliver uklare i det store billede.



Nyttiggørelse af data

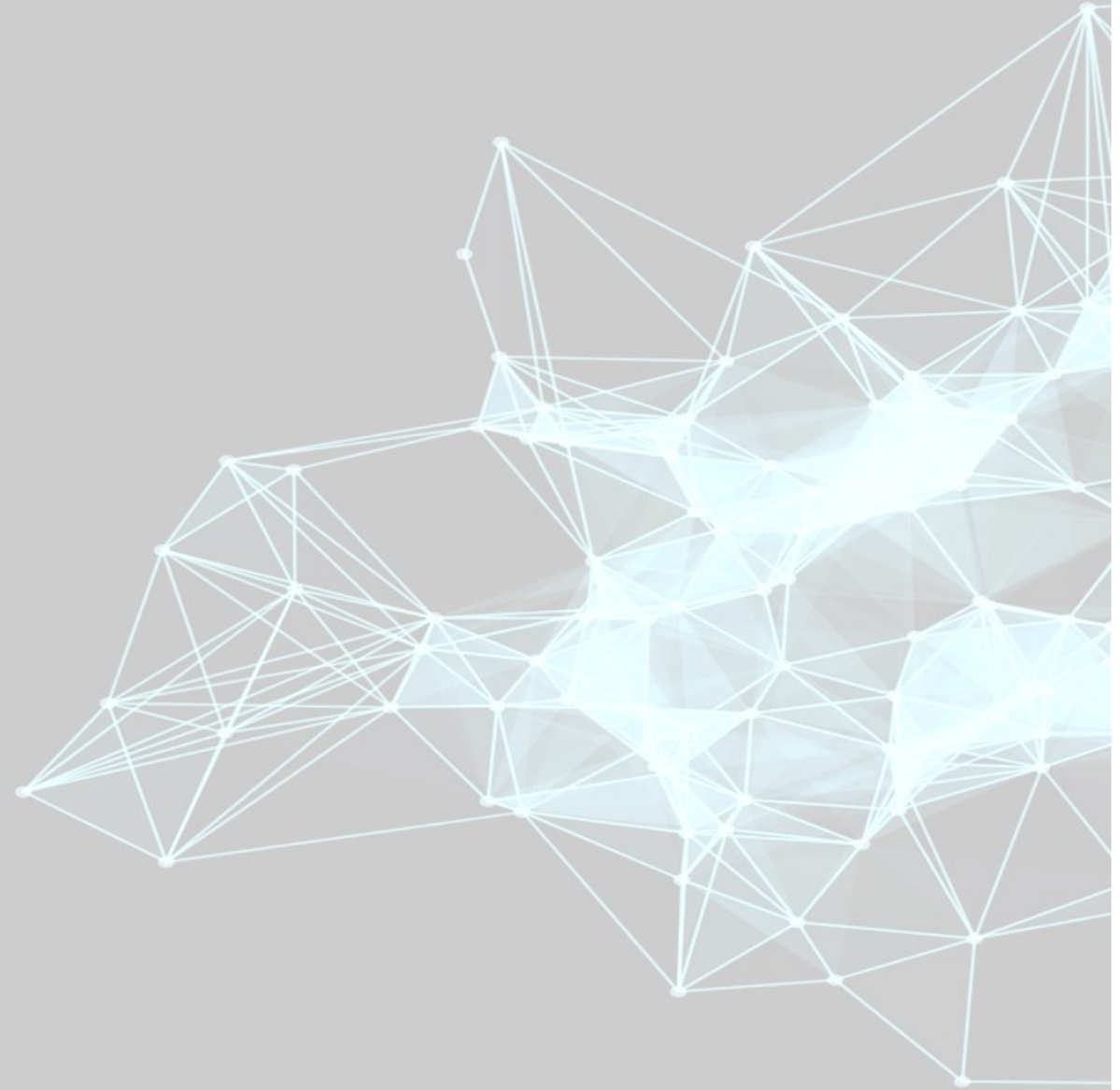
I 00'erne var der en bred erkendelse om, at data er "det nye guld". Siden har staten investeret i, og investerer fortsat i, it-arkitektur/systemer med henblik på at indsamle, opbevare og dele data.

Opbygningen af et statsligt datamiljø har og er fortsat en helt central opgave, men der er de seneste år kommet et øget fokus på at gøre bedre brug af data. Formålet med at nyttiggøre data, er at sikre, at ledere og medarbejdere i staten, træffer beslutninger på et oplyst grundlag, og derved kan forbedre opgavevaretagelsen. Der er en tendens til, at data alt for ofte ikke bliver anvendt optimalt - det "lever" ikke i institutionerne, men ligger "dødt" i systemer og databaser.

Men, hvordan kan data anvendes bedre? Den teknologiske udvikling, herunder kunstig intelligens, rummer mange muligheder for en bedre anvendelse og udstilling af data via eksempelvis business intelligence, regressionsanalyser, deep learning mv.

Kapitel 3

Definition af analytics



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Definition af analytics



Hvad er analytics?

Vi definerer og rammesætter analytics, og har følgende 10 konklusioner:

1. Vi definerer analytics til at omfatte informationssystemer, økonometri, machine learning og matematisk optimering
2. Business intelligence er automatisk bearbejdelse og udstilling af data, der skal give ledere et overblik over forretningen
3. Opgavespecifik information giver medarbejdere adgang til og/eller udstiller information, der er relevant for udførelsen af en konkret opgave
4. Økonometri tester hypoteser om, sammenhænge imellem uafhængige variable og en afhængig variabel ved datamodeller, med fokus på samvariation og estimationsbias
5. Økonometri og machine learning kan bygge på det samme matematiske grundlag, men machine learning fokuserer på outcome og mindre på samvariation og estimationsbias
6. Machine learning dækker over et hav af metoder og tilgange til data, fælles for disse er, at de indeholder et læringselement
 7. Predictive modelling kan på baggrund af historiske data foretage supervised learning af en algoritme til kvalificeret forudsigelse af udfald
 8. Klyngeanalyser dækker over et bredt spektrum af algoritmer, der via unsupervised learning kan inddele data i klynger på baggrund af komplekse mønstre i data
 9. Text analytics er en grammatisk og semantisk kvantificering af tekstmateriale, der muliggøre automatisk bearbejdelse af tekst og efterfølgende analyser på datamaterialet
10. Matematisk optimering dækker over mange metoder til at optimere et bestemt output, men der er tre elementer, der går på tværs af alle optimeringsproblemer

Vi definerer analytics til at omfatte informationssystemer, økonometri, machine learning og matematisk optimering



Informationssystemer

... kan inddeles i to grene:

1. Business intelligence: til ledelsesinformation.
2. Opgavespecifik information: til medarbejderen.

Business Intelligence er kendetegnet ved typisk at være en **bagudskuende**, deskriptiv udstilling af aggregerede forretningsdata, der danner et oplyst beslutningsgrundlag for ledere.

Opgavespecifik information er udstilling af data på individ-, virksomhed- eller institutionsniveau med **betydning for den opgave medarbejderen skal udføre**. På baggrund af informationerne kan en medarbejder træffe bedre beslutninger for at løse opgaven.

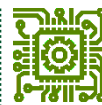


Økonometri

... er kendetegnet ved en **hypotetisk deduktiv** tilgang. Opstilling af hypoteser på baggrund af teoretiske forventninger om, hvilke uafhængige variable der påvirker den afhængige variabel. De teoretiske forventninger bliver styrende for analysen.

De enkelte variables samvariation med den afhængige variabel estimeres og giver indsigt i sammenhængene imellem de uafhængige og den afhængige variabel.

Økonometri er velegnet til at fremhæve generelle **tendenser** og **korrelation**, hvor resultaterne er udtryk for makrotendenser og gennemsnitsbetragtninger.



Machine learning

... er kendetegnet ved en **induktiv tilgang**, hvor forskellige algoritmer afprøves til at finde mønstre i data, der ikke beror på konkrete teoretiske forventninger. Data bliver styrende for analysen.

Machine learning dækker ikke over en enkelt metode, men over træningen og udviklingen af mange forskellige typer algoritmer. Algoritmerne kan angive konkrete forudsigelser og i visse tilfælde operationelle anbefalinger.

Machine learning er velegnet til at **identificere komplekse mønstre og systematikker** på baggrund af store mængder data, som en medarbejder ikke kan bearbejde.



Matematisk optimering

...omhandler maksimeringen eller minimeringen af en funktion. Øvelsen er at finde de optimale værdier, for de variable, der indgår i funktionen med henblik på enten at **maksimere eller minimere funktionen**.

I sin simpleste form kan dette være forholdsvis enkelt, hvis der kun indgår få variable i funktionen, men kompleksiteten stiger markant i takt med antallet af variable.

Matematisk optimering er velegnet til situationer, hvor en **konkret ressource skal udnyttes optimalt**. I funktionen kan det simuleres hvordan forskellige scenarier udspiller sig. Dette kan eksempelvis være vagtplansoptimering.

Business intelligence er automatisk bearbejdelse og udstilling af data, der skal give ledere et overblik over forretningen



Business intelligence er udstilling af udvalgte data i et brugervenligt format vha. en it-teknisk opsætning

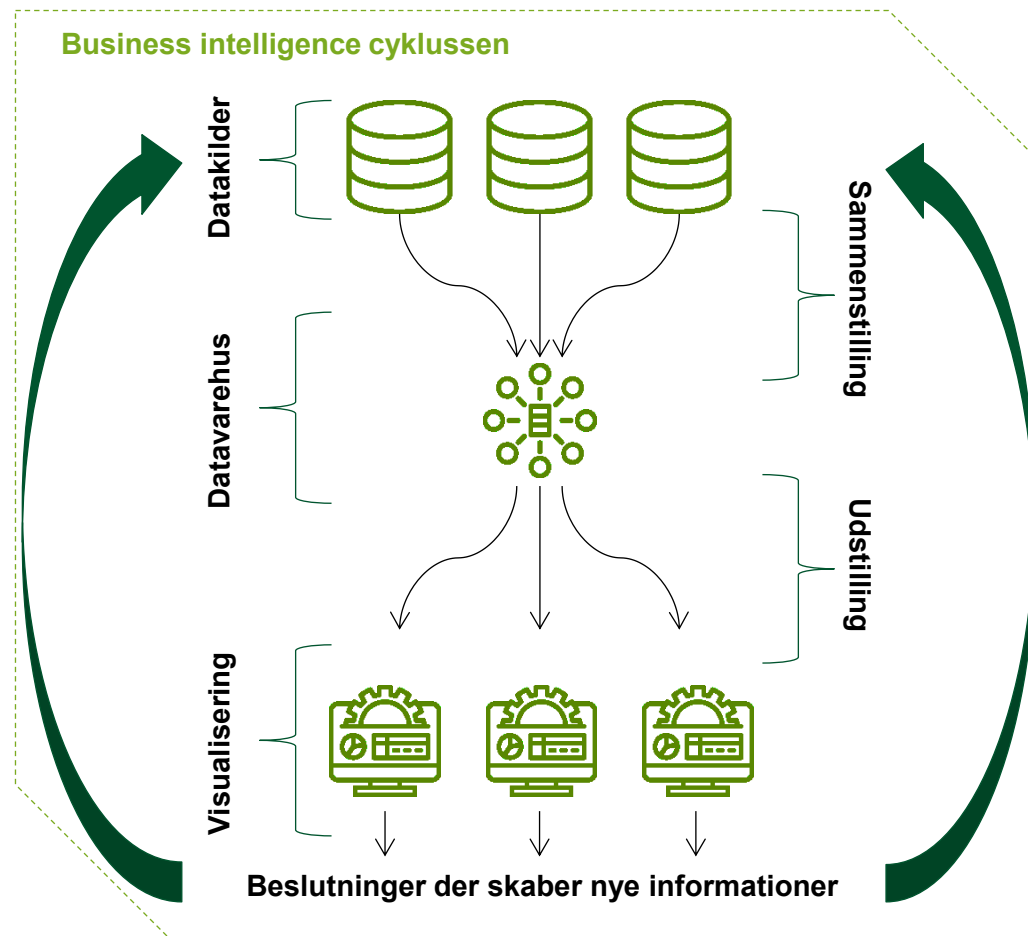
Formålet med Business intelligence (BI) er at bearbejde og udstille relevant forretningsdata på en visuel og enkel måde, for at give indsigt i forretningen og understøtte ledelsesbeslutninger.

Datagrundlaget for BI er primært strukturerede data. Hovedsageligt bruges forretningsdata såsom timeregistrering, produktion, økonomi, salg mv. til BI. I visse tilfælde bruges også ekstern data, såsom markedsudviklinger, samfundsøkonomi etc. De konkrete data til BI kan variere meget afhængigt af organisationens kerneopgave.

Automatisk generering af BI kræver en it-teknisk opsætning, eksempelvis i form af et datavarehus til sammenstilling og opbevaring af data. Online analytical processing (OLAP) og dashboards bruges ofte til udstilling og visualisering af data, og til sidst skal der være systemer til jævnlig opdatering af BI-rapporterne.

BI er typisk en bagudskuende udstilling af, hvordan organisationen klarer sig. I visse tilfælde indgår forecasting til at give informationer om centrale aspekter for, hvordan organisationen vil klare sig fremadrettet. Dette kan eksempelvis være forecasting af ferieafvikling.

BI er altid endestationen for dataanvendelsen. I BI skal ledelsen kunne aflæse, hvordan nye redskaber og initiativer påvirker organisationen og den generelle performance i organisationen. På den baggrund kan ledelsen træffe mere oplyste beslutninger om ændringer og justeringer.



Opgavespecifik information giver medarbejdere adgang til og/eller udstiller information, der er relevant for udførelsen af en konkret opgave



Opgavespecifik information er en deskriptiv udstilling af konkret data til at understøtte medarbejderes opgavevaretagelse

Der er mange opgaver, hvor det er relevant for medarbejderen at have historiske data om opgaven. Dette kan være oplysninger om en elev ved uddannelsesvejledning, oplysninger om en borgers offentlige sager til job- eller socialindsatsen osv. Kendetegnende for disse opgaver er, at både uddannelsesvejlederen, tilsynsføreren og sagsbehandleren kan træffe bedre beslutninger i det daglige, hvis de har adgang til de bagvedliggende informationer.

Datagrundlaget for opgavespecifik information er ofte det samme som for business intelligence, men forskellen fra business intelligence er, at det er konkret information helt ned på individ-, virksomheds- og institutionsniveau, der giver medarbejderen informationer om den foreliggende opgave.

Teknisk forudsætter det tre ting. 1) Informationerne skal være samlet i et datavarehus. 2) Medarbejderen skal have mulighed for automatisk at udtrække eller tilgå de konkrete informationer fra datavarehuset. 3) Informationerne skal struktureres og udstilles på en overskuelig måde, som medarbejderen hurtigt kan orientere sig i.

Opgavespecifik information indeholder ikke nødvendigvis avancerede analyser. Medarbejderen foretager ofte selv tolkningen af data, men man vil potentielt via kunstig intelligens kunne udarbejde anbefalinger til, hvilke handlinger medarbejderen bør foretage på data.

Eksempel

Oplysninger til tilsynsførerne om virksomheden



Når en tilsynsfører skal aflægge besøg hos en virksomhed, er det relevant at få generelle informationer om virksomheden samt de historiske data fra tidligere besøg.

De generelle informationer kan være antal medarbejdere, muligvis fordelt på stillingsbetegnelse, omsætning, branchekategori mv. Historiske data fra tidligere kontrolbesøg kan være hvilke påbud og krav, der før er blevet stillet, særlige opmærksomhedspunkter hos virksomheden mv. For at kvalificere tilsynsbesøget skal tilsynsføreren have mulighed for at udtrække disse data i en rapport, eller tilgå informationen i et online system ude på ruten.

Dette muliggør, at tilsynsføreren kan forberede sig på besøget og udlede hvilke særlige opmærksomhedspunkter, der gør sig gældende for virksomheden.

Økonometri tester hypoteser om sammenhænge mellem uafhængige variable og en afhængig variabel



Økonometri tager udgangspunkt i en hypotese, der testes ved hjælp af data, statistik og en matematisk model, som menes at beskrive virkeligheden

Inden for økonometri er det overordnede formål at teste en hypotese. Der er fokus på én eller flere af de forklarende variables ("x-variable") effekt på outcome ("y-variablen"). Samvariation og estimationsbias er fundamentale aspekter i økonometriske analyser, da de forklarende variable er omdrejningspunktet. Selve udfaldet af y-variablen spiller en mindre rolle.

Samvariation beskriver, hvordan både den afhængige variabel er korreleret med de uafhængige variable, samt hvordan de uafhængige variable er korreleret med hinanden. For at underbygge samvariationen og modelresultaterne anvendes teoretiske argumenter ofte i økonometriske analyser. Det store fokus på samvariation i økonometri betyder, at man opnår stor indsigt i, hvordan de forskellige variable forholder sig til hinanden.

Estimationsbias beskriver, om x's effekt på y over/undervurderes, hvilket undersøges ved forudsætningstests, der giver indsigt i datagrundlaget for modellen, fx fordelingsanalyser. Datagrundlaget for økonometriske analyser er typisk mindre datasæt på aggregeret niveau eller en stikprøve af en population. På baggrund af stikprøven foretages statistiske analyser på om, resultaterne i stikprøven også gælder for populationen. De mindre datasæt betyder, at de statistiske analyser er sårbare overfor egenskaberne i datasættet, fx outliers, stikprøveudvælgelse, variation mv.

Eksempel

Effekten af en afgiftsstigning på cigaretter

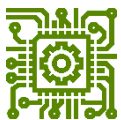


Afgifter er et udbredt instrument til at adfærdsregulere forbrugsvaner hos borgerne. Afgifter på cigaretter er en af de mest velkendte forbrugsafgifter, og effekten af en afgift på cigaretter kan estimeres i en økonometrisk analyse.

Y-variablen er forbruget af cigaretter. Den primære x-variabel er afgiftsstigningen. Den teoretiske forventning er at hæves prisen på cigaretter, har det en negativ effekt på forbruget af cigaretter. For at have styr på samvariationen mellem afgiftsstigningen og forbruget, skal der indgå kontrolvariable, såsom køn, geografi, socioøkonomisk status og alder, da der er en teoretisk forventning om, at de påvirker forbruget af cigaretter. Fx forventes det, at effekten af en afgiftsstigning vil være mindre for ældre end yngre rygere. Herefter kan det estimeres, at fx for hver krone prisen stiger, falder forbruget med én cigaret pr. ryger.

Økonometri muliggør således at tydeliggøre generelle tendenser for, hvordan en afgiftsstigning påvirker cigaretforbruget.

Machine learning dækker over et hav af metoder og tilgange til data, men fælles for disse er, at de indeholder et læringselement



Machine learning foretages enten ved supervised eller unsupervised learning

Machine learning er i manges forståelse én metode, men machine learning er et paraplybegreb for mange underliggende metoder, med yderligere begreber, der knytter sig til metoderne, hvilket gør begrebet endnu mere komplekst. Kendetegnende for alle ML-metoder er, at de indeholder et element af læring. Der er to hovedformer for læring supervised og unsupervised learning.

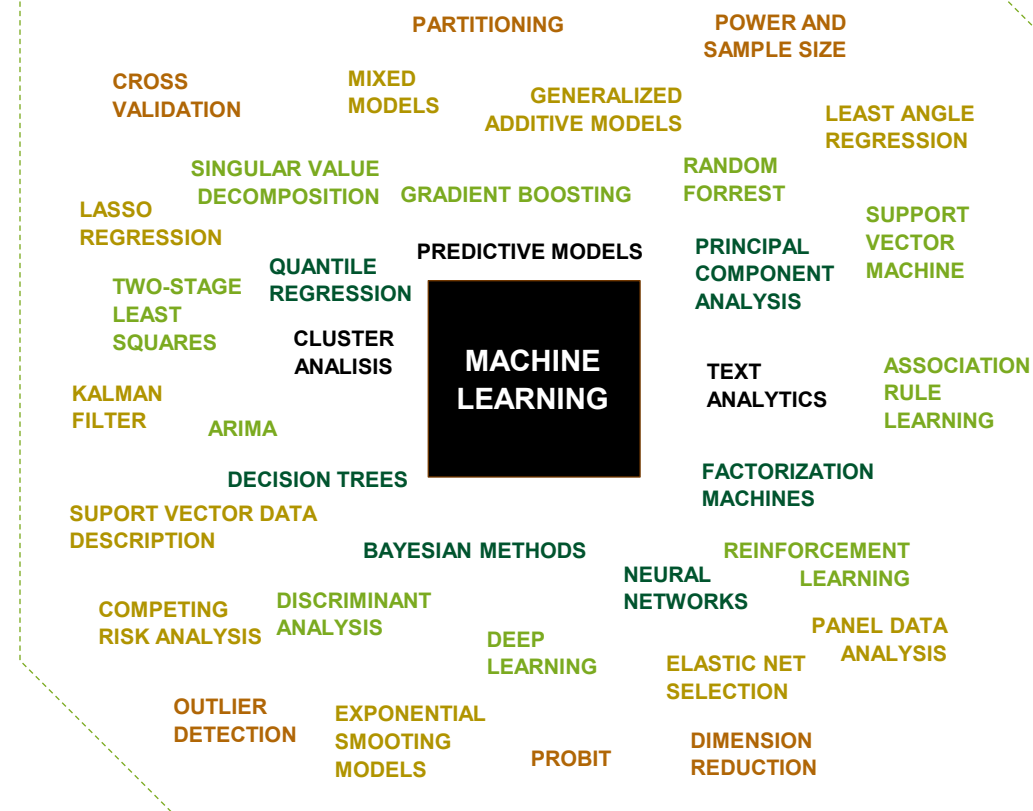
I supervised learning benyttes et træningsdatasæt med labelled data, dvs. hver observation kan kategoriseres efter den variabel man er særlig interesseret i (outcome) og et testdatasæt. Træningsdatasættene bruges til at træne algoritmen i at finde sammenhænge mellem input og outcome, mens testdatasættet benyttes til at vurdere algoritmens præcision. Ofte er det vanskeligt at forklare algoritmen, men præcisionen kan måles.

I unsupervised learning arbejdes der med unlabelled data, hvilket betyder at outcome-variablen ikke er kendt. Manglen på en outcome-variabel betyder, at der ikke er et endegyldigt resultat i unsupervised learning, hvorfor funktionens præcision ikke kan vurderes. I stedet trænes en algoritme i at finde mønstre og systematikker i input data, der bruges til at kategorisere og opdele data. Dette anvendes fx til at kategorisere tekstdokumenter, efter hvilke ord de indeholder.

Vi vender nu fokus mod tre centrale metoder inden for machine learning predictive modelling, klyngeanalyse og tekst analytics.

Illustration

Machine learnings black box



Økonometri og machine learning kan bygge på det samme matematiske grundlag, men machine learning fokuserer på outcome og mindre på samvariation og estimationsbias



Machine learning adskiller sig særligt fra økonometri i forhold til øget fokus på outcome, samt følsomhed overfor korrektheden i outcome-variablen

Modellerne i økonometri og machine learning (ML) kan begge basere sig på lineær og logistisk regression. Det matematiske grundlag kan således være det samme, men det bruges på to vidt forskellige måder. I økonometri giver det store fokus på estimationsbias og samvariation indsigt i, hvordan og hvorfor resultaterne forekommer. I ML nedtones denne del, og der er et stort fokus på, hvad outcome er – eller snarere hvad forudsigelsen af outcome er.

Outcom'et af ML er oftest et meget operationelt resultat, der konkret kan måles i forhold til dets korrekthed i virkeligheden. Derfor er præcisionen af forudsigelsen af outcome essentiel, snarere end hvordan og hvorfor forudsigelsen er fremkommet. Algoritmen trænes og justeres i forhold til om outcome er korrekt. Ved denne induktive tilgang justeres modellen på baggrund af algoritmens præcision i forhold til virkelighedens hændelser og ikke på baggrund af teoretiske overvejelser.

ML er følsom overfor fejl på outcome-variablen. I træningen af en algoritme er det afgørende, at outcome-variablen ikke indeholder bias eller fejl, da det er styrende for algoritmen. (Illustreret i eksemplet)

Datakvaliteten i de uafhængige variable er mindre afgørende i ML. Ofte anvendes machine learning på mikro-niveau med mange observationer. Dette medfører, at støj i input-variablene, såsom outliers, datakvalitet mv. udjævnes af datamængden og i lavere grad influerer outcome. Derudover kan flere ML algoritmer håndtere (usystematisk) manglende værdier for input variablene.

Eksempel

Er kautionsudmålingsalgoritmen i USA racistisk?



I USA er der pladsmangel i fængslerne. Derfor har retssystemet indført en algoritme, der kan forudsige sandsynligheden for, om den tiltalte vil udføre kriminalitet igen. Algoritmen bruges som beslutningsstøtte for dommere til at fastsætte, om tiltalte er for farlig til at blive løsladt mod kaution. Problemet er, at algoritmen er blevet kritiseret for at være racistisk.

Dette skyldes, at afroamerikanere typisk får tildelt en højere risikoscore alene på grund af deres hudfarve. Kritikere argumenterer, at afroamerikanere vurderes hårdere, da de er mere udsatte for at blive arresteret igen, grundet øget politiindsatser i boligområder med mange afroamerikanere og grundet racisme i politistyrken. Dette afspejler sig i at afroamerikanere oftere bliver arresteret igen, hvilket er styrende for algoritmens læring og kautionsudmålingen.

Det argumenteres således, at de underliggende racistiske strukturer er indlejret i algoritmen.



Predictive modelling kan på baggrund af historiske data foretage supervised learning af en algoritme til kvalificeret forudsigelse af udfald



Predictive modelling er en proces, hvor trends anvendes til at designe og udvikle en model, der kan bruges til forudsigelser på individniveau

Predictive modelling estimerer sandsynligheden for et outcome baseret på karakteristika på input-variable. Ved at træne en algoritme kan den identificere systematikker i store mængder data for, hvilke karakteristika i input variable, der medfører et bestemt outcome. Algoritmen kan sammenholde de generelle tendenser med den enkelte enhed og give en konkret sandsynlighedsvurdering for outcome'et.

Predictive modelling er et eksempel på supervised learning, hvor en algoritme trænes til at lave forudsigelser baseret på flere træningsdatasæt. I træningen af algoritmen opdeles data typisk i trænings-, validerings- og testdatasæt. Trænings- og valideringsdatasættet bruges til parallelt at træne algoritmen og samtidig validere model og variabelselektion. Testdatasættet bruges til at sikre generaliserbarheden af modellen.

I supervised learning er det afgørende, at outcome-variablen ikke har manglende værdier, at datakvaliteten er høj og at den ikke indeholder bias. Hvis data på outcome ikke er retvisende indlæres det i modellen, da det er styrende for funktionen mellem input og outcome. I kraft af at outcome-variablen til en vis grad er styrende for algoritmen, arbejdes, der ofte med flere outcome-variable, der giver et mere nuanceret billede af forholdet imellem input og outcome.

Eksempel

Kreditvurdering af lånsøgere



Et af de klassiske eksempler på predictive modelling er kreditvurderinger af lånsøgere.

Når man ønsker at låne penge til huskøb, har bankerne behov for at vide om man kan betale pengene tilbage med renter, så det ikke bliver en dårlig forretning for banken. Derfor laver de en kreditvurdering af låntagere.

Kreditvurderingen tager udgangspunkt i input-variable om økonomi, fx tidligere lån, forbrugsvaner, uddannelse, aktiver mv. Disse variable bliver bearbejdet af algoritmen, som sammenholder det med erfaringerne fra tusindvis af tidligere låntagere. Dette udmunder i en konkret forudsigelse - en vurdering - af låntager baseret på tidligere erfaringer.

Klyngeanalyser dækker over et bredt spektrum af algoritmer, der via unsupervised learning kan inddele data i klynger på baggrund af komplekse mønstre i data



Klyngeanalyse er en velegnet måde til at klassificere store mængder data i grupper og opnå indsigt i karakteristika ved data

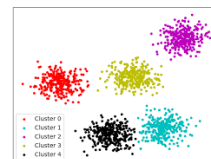
Klyngeanalyser består i sin grundform af at gruppere observationer ud fra princippet om, at observationerne i klyngerne har fælles ligheder, og at de adskiller sig fra andre klynger. Klyngeanalyser bygger på en eksplorativ tilgang til data, hvor man ønsker at opnå større indsigt i strukturerne i data.

Klyngeanalyser dækker over et bredt spektrum af algoritmer, der kan identificere mønstre i data. Dette skyldes, at det er svært at definere, hvad der konstituerer en klynge, da det afhænger af de underliggende systematikker i data. Resultaterne fra én algoritme kan være vidt forskelligt i forhold til resultaterne fra en anden algoritme. Hvilken algoritme, der er mest anvendelig, baserer sig primært på strukturerne i data. På denne baggrund har databehandleren en vigtig rolle i at vurdere hvilken metode, der er mest anvendelig i forhold til data og til dels formålet. Anvendeligheden af en algoritme kan måles, da der er statistiske mål for, hvor stor en andel af data, som algoritmen kan forklare.

Til klyngeanalyser anvendes unsupervised learning. Den meget eksplorative tilgang betyder, at klyngeanalyser kan have mange iterationer før man opnår et resultat, der beskriver systematikkerne i data og før man har fundet den algoritme, der beskriver data bedst. Klyngeanalyser anvendes ofte som det første skridt i analyser med store mængder data for at få et overblik over strukturerne i data, hvilket bidrager til at kvalificere, hvordan data kan anvendes fremadrettet.

Eksempel

Segmentering af biblioteksgængere



Fredericia Bibliotek har haft et konsulentfirma til at afdække de profiler, der anvender biblioteket. I undersøgelsen udsendte de et spørgeskema til brugerne, der gav baggrundsinformationer på brugerne. I en klyngeanalyse af baggrundsdata identificerede man fire profiler på biblioteks-gængere: børneforælderen, senioren, funktionæren og den uddannelsessøgende. Funktionæren var eksempelvis karakteriseret ved at være ml. 50-59, kvinde, ansat som lærer, pædagog eller sygeplejerske og boede i en mindre by i oplandet. Analysen blev brugt strategisk til at målrette bibliotekets tilbud mod disse profiler.

Dette er en af de simpleste eksempler på en klyngeanalyse. Øges datamængden og baggrundsinformationerne kan der identificeres meget komplekse mønstre, der karakteriserer enhederne i datasættet. Dette kan anvendes i opgavevaretagelsen eller til videre analyse.

Tekst analytics er en grammatisk og semantisk kvantificering af tekstmateriale, der muliggøre automatisk bearbejdelse af tekst og efterfølgende analyser på datamaterialet



Tekst analytics kan udvikle algoritmer, der kan bearbejde tekstmateriale og udlede meningsfulde konklusioner af tekst

Det overordnede formål med tekst analytics er at identificere nøgleord eller konstruktioner af særlig interesse i større tekstbaserede datasæt, oversætte det til et "computational" sprog og udlede mening af det ved forskellige analytics metoder.

Det første trin i tekst analytics er at omdanne store mængder ustrukturerede tekstdata til strukturerede data. Dette sker ved, at en machine learning algoritme (fx klyngeanalyse) tæller, grupperer og kategoriserer ord samt bearbejder den grammatiske struktur for at danne et grundlæggende datasæt. Denne proces sker hele tiden i en vekselvirkning imellem algoritmens bearbejdelse og justeringer fra databehandleren.

Dernæst foretages semantiske analyser på det grundlæggende datasæt, hvor natural language processing (NLP) bearbejder tekstmaterialet for synonymer, kontekstforståelse og koncepter, så der ikke er misforståelser af tekstmaterialet. For eksempel, når ordet "kører" indgår i samme sætning som trafik, referer det til bilkører og ikke dyrene.

Når ovenstående proces har fundet sted kan det bruges på flere forskellige måder. Algoritmen, der er blevet udviklet til at bearbejde data, kan bruges i fx sagsbehandlingen til at bearbejde lignende tekstdokumenter. Derudover kan det data, der er blevet udledt af algoritmen indgå i en anden algoritme fx et træningsdatasæt til predictive modellering.

Eksempel

Text analytics til spamfiltrering

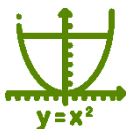


En væsentlig problemstilling for virksomheder, offentlige institutioner og private personer er spam-mails, der indeholder malware.

Denne problemstilling bliver i visse tilfælde løst af tekst analytics. Det sker vha. at klyngeanalyser og NLP ved unsupervised learning anvendes til at nedbryde de grammatiske strukturer og de semantiske konstruktioner, der ofte gør sig gældende i spammails. Dette er eksempelvis, at spam mails har en "madding", der skal få læseren til at trykke på mailen, såsom en pengegevinst, krav om hurtig respons etc. Denne type information kan bruges til at danne et labelled træningsdatasæt, hvor input og output er kendt, hvorefter en tekst analytics algoritme kan trænes til at identificere disse mønstre.

Algoritmen kan herefter anvendes til at frasortere de mails, der følger mønstrene, som blev identificeret som kendetegnende for spam mails.

Matematisk optimering dækker over mange metoder til at optimere et bestemt output, men der er tre elementer, der går på tværs af alle optimeringsproblemer



Matematisk optimering beregner, hvilke værdier på input variablene i en funktion, der giver det bedste output.

Matematisk optimering er maksimering eller minimering af en funktion, ved at fastsætte de optimale input for variablene i funktionen. Overordnet indeholder optimeringsproblemer tre elementer.

1. *Et numerisk output man ønsker at minimere eller maksimere.*
Dette kan være antallet af stemmer i en valgkamp, produktionsomkostninger, kørselstid mv.
2. *Et sæt variable, der kan manipuleres for at optimere målet.*
Dette er variable man kan justere på i forhold til at opnå et bedre output. I en valgkamp kan det være hvor mange ressourcer, der skal bruges på fundraising, udgifterne til eksponering, antallet af byer, der skal besøges osv. Komplexiteten i en optimeringsfunktion stiger markant i takt med antallet af variable i funktionen.
3. *Begrænsninger på variable, der kan manipuleres.*
Nogle variable kan være underlagt begrænsninger, der kan have væsentlig indflydelse på funktionen og komplicere funktionen yderligere. Dette kunne eksempelvis være, at der er en øvre grænse for, hvor mange penge der må bruges i en valgkamp, eksponeringstid, der er til rådighed på TV osv.

Matematisk optimering dækker over mange forskellige tilgange, såsom lineær-, stokastisk-, non-lineær-, kombinatorisk- og netværksoptimering. Alle disse kan bruges til forskellige problemer og har særegne styrker, svagheder og kompleksitetsgrad.

Eksempel

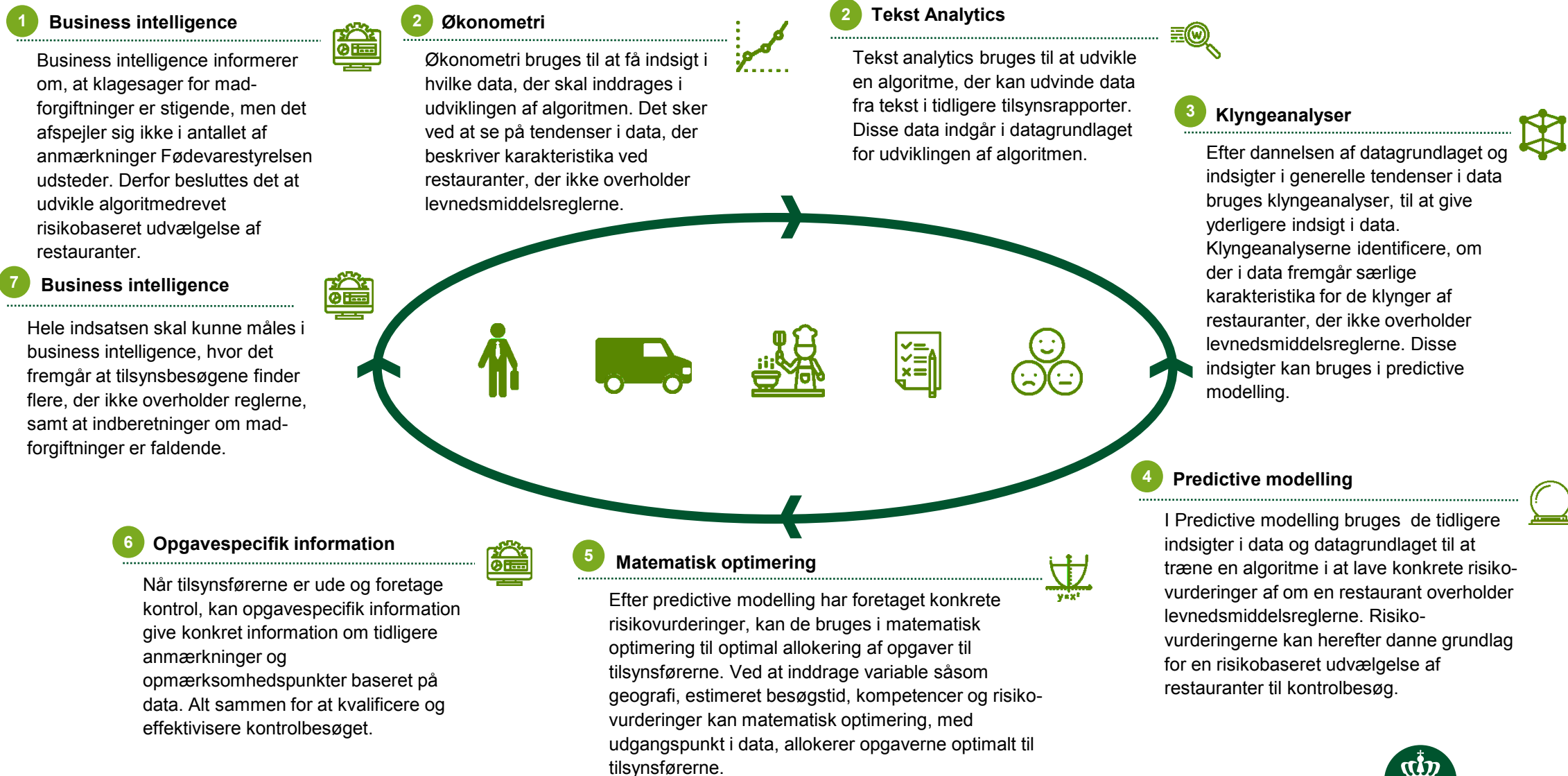
Matematisk optimering af shippingruter



I shipping bliver matematisk optimering brugt til at optimere mængden af varer, der kan leveres over hele kloden og særligt i sejlads har det haft kæmpe indflydelse.

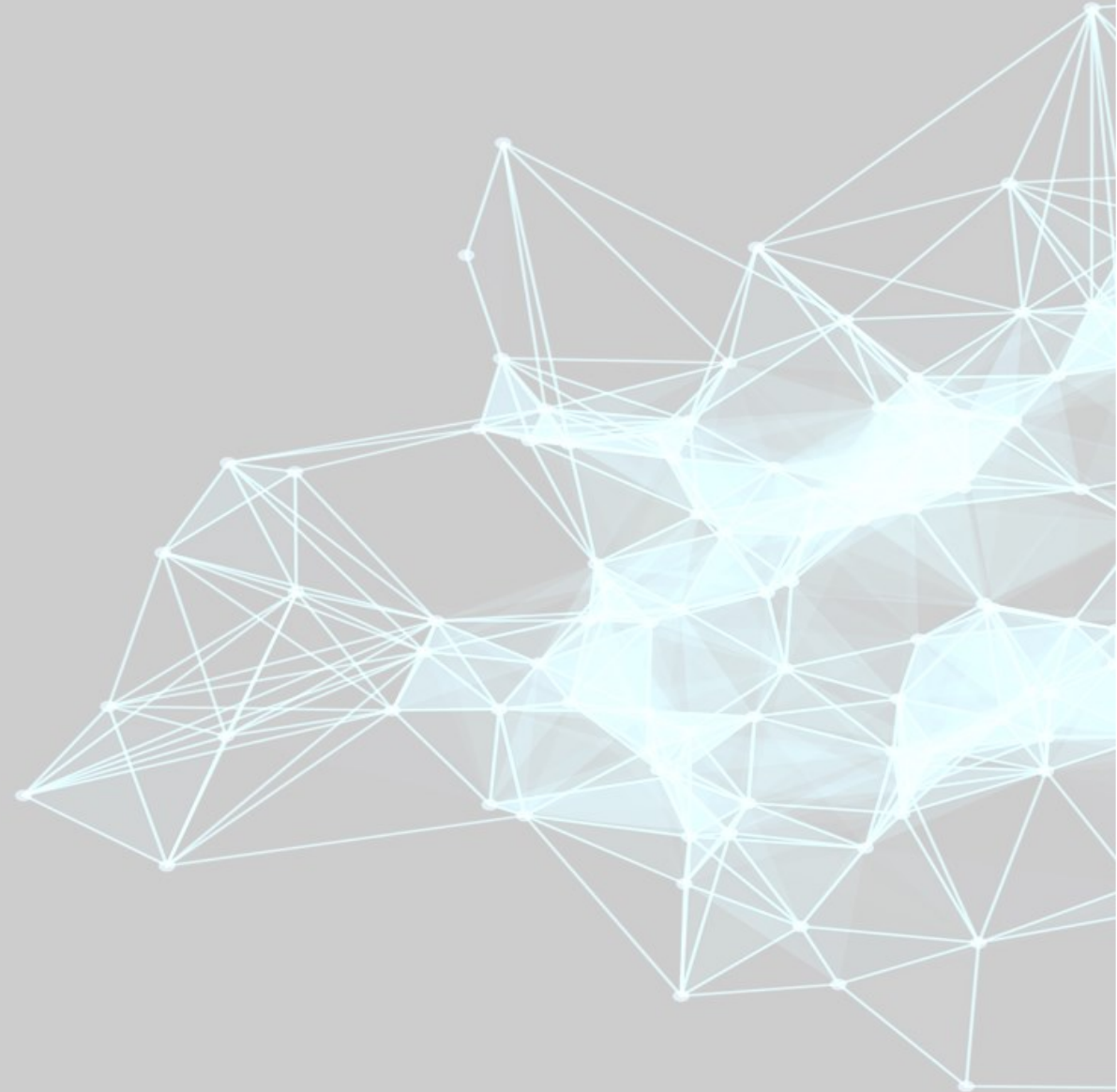
Med kæmpe containerskibe, der fragter tusindvis af varer i én sejlads, er selv de mindste faktorer betydningsfulde. Derfor bliver ruteplanlægningen minutiøst fastlagt for at minimere tiden, der bliver brugt på sejlads. I dette indgår variable, såsom geografi, vejrforhold, skibstype, gast-flaskehalse, brændstofforbrug, af- og pålæsning, andre skibes ruter mv. Alle disse faktorer kan have en markant indflydelse på sejladstiden. Fx er brændstof en variabel med begrænsninger, da der ikke bare kan bruges uanede mængder brændstof, både fordi det ikke er muligt at tanke op medmindre man lægger til havn, og fordi omkostningerne for brændstofforbruget ved at holde en høj hastighed ikke svarer til gevinsterne ved en tidsbesparelse.

De enkelte metoder kan føde ind i hinanden og være tæt forbundet fx kan alle metoder blive anvendt til at kvalificere tilsynsopgaven



Kapitel 4

Erfaringer og potentialer med analytics



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Erfaringer og potentialer inden for analytics (del 1)



Hvilke erfaringer og potentialer har staten?

Vi gennemgår statens erfaringer og potentialer inden for analytics, og har følgende konklusioner:

1. Statens erfaringer med analytics er størst ved business intelligence og økonometri og potentialet er størst i matematisk optimering og machine learning
2. Beslutningsstøtte via business intelligence er relativt udbredt i staten, og bliver brugt til at træffe bedre beslutninger med udgangspunkt i data
3. Opgavespecifik information bliver brugt til at udstille information til tilsynsførende, hvilket muliggøre et kvalificeret og effektivt tilsynsbesøg
4. Økonometri bruges bredt i staten til policyevalueringer, samt at opnå større indsigt i generelle tendenser i data
5. Matematisk optimering er ikke særlig udbredt blandt de besøgte institutioner, på trods af et stort potentiale inden for ressourceoptimering med høj kombinatorik
6. Der er identificeret seks forretningsområder, hvor institutionerne har gjort sig erfaringer med eller ser potentiale i at anvende machine learning

... fortsætter på næste slide...

Opsummerende resultater: Erfaringer og potentialer inden for analytics (del 2)



Hvilke erfaringer og potentialer har staten?

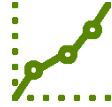
Vi gennemgår statens erfaringer og potentialer inden for analytics, og har følgende konklusioner:

6. Algoritmedrevet risikobaseret tilsyn med data fra mange kilder, er vidt udbredt som idé, men hjemmelsspørgsmålet og årsagsforklaringsproblemet hæmmer erfaringerne
7. Algoritmedrevet risikobaseret tilsyn med meget data fra få kilder, er ikke vidt udbredt, men implementeringen er større, da hjemmelsproblemet ikke gør sig gældende
8. Anbefalingsværktøjer kan gøre opgavespecifik information intelligent, hvilket gør sig gældende i otte projekter, men kun ét er implementeret
9. Tendensanalyser til strategisk indsats er i mindre omfang udbredt på trods af, at hjemmelsproblemet delvis er overflødigt
10. Machine learning bliver i mindre omfang brugt til kundesupport og brugermålretning, men flere institutioner ser muligheder i at bruge det til forbedring af deres hjemmesider
11. Kvalitetssikring af data ved machine learning rummer et potentiale for at øge datakvaliteten i staten, men der er kun enkelte institutioner med konkrete erfaringer
12. Effektiv sagsbehandling ved machine learning rummer et enormt potentiale i takt med at de danske sprogkompetencer inden for kunstig intelligens udvikles

Statens erfaringer med analytics er størst ved business intelligence og økonometri, og potentialet er størst ved matematisk optimering og machine learning



Informationssystemer



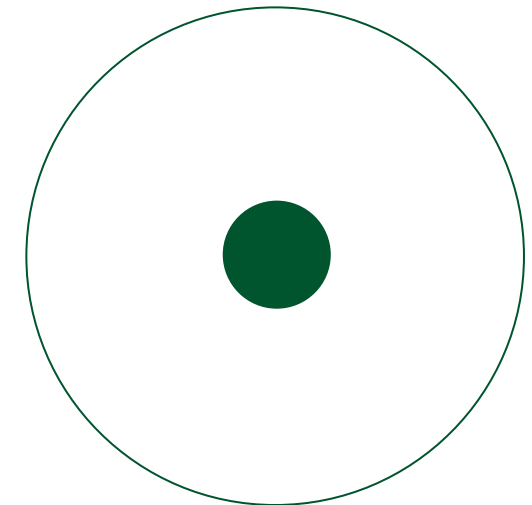
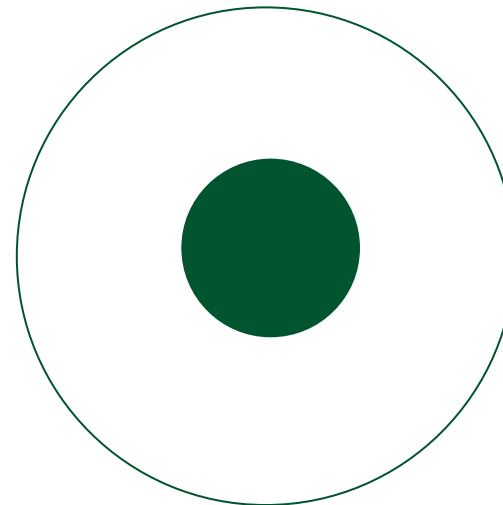
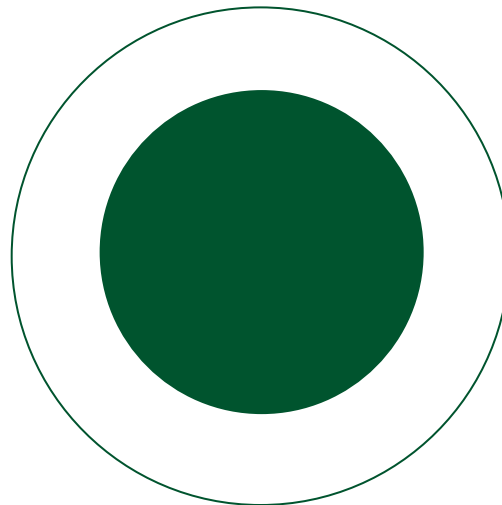
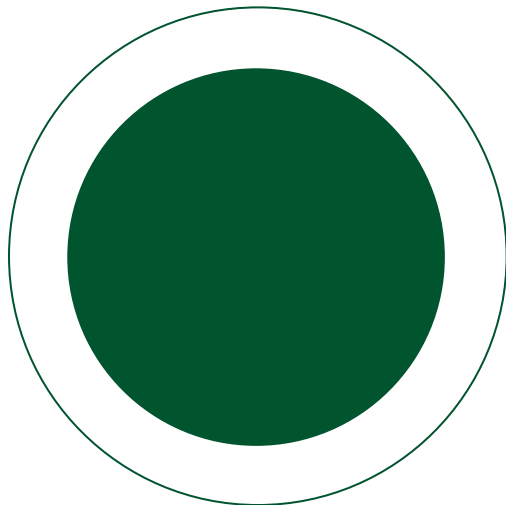
Økonometri



Machine learning



Matematisk optimering



Erfaring Potentiale



Business Intelligence er relativt udbredt i staten og bliver brugt til at træffe bedre beslutninger med udgangspunkt i data



Hovedparten af de besøgte institutioner anvender business intelligence til ledelsesinformation

Institutionerne bruger business intelligence (BI) til at bearbejde og udstille forretningsdata på en visuel og enkel måde for at give indsigt i forretningen for ledere, og derved understøtte ledelsesbeslutninger. BI bliver således et fundament for strategiske beslutninger.

Datagrundlaget til BI bygger primært på institutionernes interne data. Nogle interne data til BI er ens på tværs af flere institutioner, mens andre er kontekstspecifikke for institutionen. Data som går på tværs er fx ESDH-data, tidsregistrering, trivsel mv. De kontekstspecifikke data er fx data om, hvor meget Styrelsen for Dataforsyning og Effektivitets egne data bliver anvendt af andre og hvem der anvender det. Ekstern data er i begrænset omfang anvendt af institutionerne til BI.

Graden af automatisk generering og opdatering af BI varierer på tværs af institutionerne. Hovedparten har data i et datavarehus, hvor visse institutioner har implementeret OLAP/kuber ovenpå datavarehuset, der automatisk bearbejder data og udstiller det i dashboards. Dette muliggør en hurtig opdateringsfrekvens, i visse tilfælde "real time". Andre institutioner bearbejder og visualiserer data manuelt, hvorfor opdateringsfrekvensen er mindre hyppig.

Enkelte institutioners BI er interaktivt, hvor lederne kan "drill down" i data, fx hvis BI viser, at trivslen er faldet i styrelsen, kan en leder udfolde data ved at trykke på de enkelte afdelinger. Hvorved det kan identificeres om det gør sig bredt gældende eller om det er isoleret til enkelte afdelinger. Herefter kan data fx fordeles på jobkategori for at se, om det er de fuldmægtige, der ikke trives osv.

Caseeksempel

Moderniseringsstyrelsens BI



I Moderniseringsstyrelsen modtager chefer og direktionen dashboards indeholdende økonomi-, HR- og faglige nøgletal. Dette er fx en vurdering af målopfyldelsen i arbejdsprogrammet udtrykt i smileys, hvilke aktiviteter, der bliver brugt ressourcer på og hvordan sygefraværet er. Dashboardet er interaktivt, så chefer kan drill-down på interessante perspektiver. Disse dashboards anvendes i ledelsesøjemed til at lave bedre ressourceprioriteringer, imødekomme potentielle problemer og holde fokus på målene.



Opgavespecifik information bliver brugt til at udstille information til tilsynsførende, hvilket muliggøre et kvalificeret og effektivt tilsynsbesøg



Mange tilsynsførere kan via en tablet eller telefon tilgå opgavespecifik information ude på ruten til orientering om særlige opmærksomhedspunkter for tilsynsbesøget

Formålet med opgavespecifik information er at skabe et mere kvalificeret og effektivt tilsynsbesøg, ved at give tilsynsførerne adgang til institutionens systemer ude på ruten. Ved ekstern adgang til systemerne kan tilsynsførerne orientere sig i konkrete informationer om virksomheden eller institutionen inden besøget, og på denne baggrund identificere opmærksomhedspunkter. Derudover kan de indtaste resultaterne fra tilsynsbesøget umiddelbart efter besøget.

Datagrundlaget for opgavespecifik information baserer sig på organisationsspecifikke informationer som tilsynsinstitutionerne har samlet i deres datavarehus eller database(r). Tilsynsførerne fra arbejdstilsynet har eksempelvis adgang til information, om virksomheden har tidligere påbud, virksomhedsstørrelse, indberettet arbejdsskader mv. Datagrundlaget er således det samme som til BI, forskellen er, at BI aggregerer data om alle tilsynsbesøg, hvorimod opgavespecifik information er mikro-data på det enkelte tilsynsbesøg.

Teknisk kan tilsynsførerne tilgå institutionens datavarehus eller databaser eksternt fra fx en tablet eller telefon, for at orientere sig i de konkrete informationer (se eksempel). Typisk er det via en platform, der er designet til at de tilsynsførerne hurtigt kan orientere sig i data.

Caseeksempel

ADD udstiller relevant information til kontrollører ude på ruten



Anti Doping Danmark (ADD) aflægger kontrolbesøg i motions- og fitnesscentre og foretager dopingkontroller på udvalgte medlemmer.

Til disse besøg har kontrollanter adgang til konkret information om fitnesscenteret fra ADDs efterretningsdatabase IBASE. IBASE indeholder information om fitnesscenterets historik og dets medlemmer. Informationen om centerets medlemmer dækker bl.a. over resultater fra social network analysis af medlemmers vennelister på sociale medier. Hvis et medlem eksempelvis har kontakter på de sociale medier, der tidligere er blevet taget i motionsdoping, kan denne person være relevant at udtage til dopingkontrol.

Kontrolløren får således udstillet konkret information, der bidrager til en mere kvalificeret og effektiv kontrolindsats.

Økonometri bruges bredt i staten til evalueringer, samt til at opnå større indsigt i generelle tendenser i data



Økonometri bliver i stort omfang brugt til evalueringer af indførte og foreslåede politikker

Evalueringer bliver både foretaget på politikker, der er blevet indført (ex post), samt i politikudviklingen af hvilke effekter forskellige nye tiltag vil have (ex ante).

I disse analyser estimeres det, hvilken effekt den pågældende politik har haft eller vil få kontrolleret for relevante kontrolvariable. Denne type analyser foretages både i styrelser og departementer, men der er en tendens til, at departementer i højere grad udfører ex ante evalueringer, mens styrelser udfører ex post evalueringer.

Økonometri bruges i opgavevaretagelsen til at opnå større indsigt i mekanismerne på området

Økonometri er særligt anvendeligt til at afdække kausalitet. Dette sker ved at opstille hypoteser for hvilke uafhængige variable, der påvirker en afhængig variabel. Ved denne metode opnås en større indsigt i hvilke variable, der har betydning for et givent outcome, og på hvilken måde det har betydning for outcome.

Denne indsigt er anvendelig for politikudviklingen, da det kan bidrage til at danne grundlag for hvilke indsatsområder, der skal understøttes for at opnå det ønskede resultat.

Caseeksempel

Policyevaluering af førtidspensionsreformen ift. udsatte unge



Beskæftigelsesministeriet har foretaget en policy-evaluering af effekten af førtidspensionsreformen for udsatte unge. Evalueringen viste, at de udsatte unge i højere grad kom i beskæftigelse eller startede på en uddannelse. På baggrund af evalueringen konkluderes det, at de første tidlige effekter af førtidspensionsreformen var som ønskede.

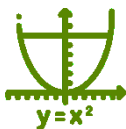
Caseeksempel

Kausalstudier af variable, der påvirker elevers uddannelse



Styrelsen for It og Læring laver økonometriske modeller over hvilke variable, der påvirker elevers sandsynlighed for at få en ungdomsuddannelse. Dette kunne eksempelvis være forældres uddannelse, elevens trivsel, fritidsaktiviteter mv. Identificering af, hvilke variable, der er afgørende for elevers uddannelse, kan bidrage til at fokusere indsatsområder på de vigtigste aspekter.

Matematisk optimering er ikke særlig udbredt blandt de besøgte institutioner, på trods af et stort potentiale inden for ressourceoptimering med høj kombinatorik



Kompleksiteten i opgaveallokering og vagtplanlægning sandsynliggør suboptimale løsninger, som kan løses ved matematisk optimering, men erfaringerne er sparsomme

Flere institutioner står over for problemstillingen, at kompleksiteten i ressourceallokeringen gør det besværligt at finde den optimale løsning. Af de interviewede institutioner gør problemstillingen sig særligt gældende hos institutioner, hvor kompleksiteten i vagtplanlægningen og opgaveallokeringen er problematisk fx tilsynsinstitutioner.

Vagtplanlægning er et spørgsmål om, hvem der skal være på arbejde hvornår. Denne opgave bliver kompleks når bestemte medarbejder skal være tilstede for at undgå flaskehalse. Kombineret med en stor medarbejdervolumen, arbejdstidsregler, afspadsering, ferie mv. er kompleksiteten markant. Problemstillingen gør sig bl.a. gældende i kødkontrollen, da der skal være kødkontrollanter tilstede visse steder i slagterprocessen.

Opgaveallokering er et spørgsmål om, hvem der skal løse hvilken opgave. Dette bliver fx komplekst hos tilsynsinstitutioner når opgaver skal allokeres efter kompetencer, geografisk placering, estimeret besøgstid, risikovurdering, tidsfrister mv. I Sikkerhedsstyrelsen er dette en problematik, da opgaverne er spredt udover hele landet, og kompetencekravene til opgaverne er meget forskellige.

Matematisk optimering kan bearbejde alle disse variable og derved overkomme kompleksiteten for at finde den optimale ressourceallokering.

Perspektivering

Vagtplanlægning på sygehuse



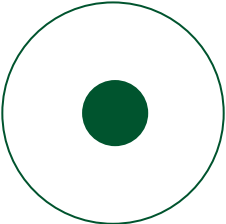
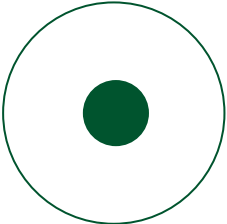
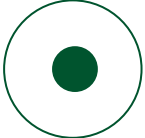
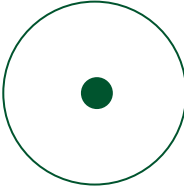
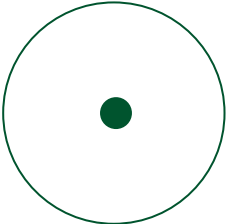
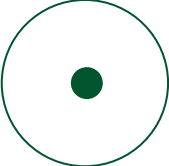
Rigsrevisionen kritiserede i 2015 hospitalernes brug af personaleressourcer. Efter kritikken identificerede regionerne vagtplanlægning som et gunstigt område for bedre anvendelse af personaleressourcerne.

Ved at anvende data i en matematisk optimeringsmodel indeholdende overenskomst regler, kompetencekrav, aktiviteter, medarbejder ønsker/krav, faggrupper, vagtdimensioner, såsom dag/aften/nattevagter, team mv., har visse sygehuse optimeret vagtplanlægningen, så der på vagterne er de rette kompetencer, faggrupper, antal medarbejdere til stede for at løse opgaverne. Vagtplanlægningen overholder samtidig overenskomstregerne, ferieafviklingen, fair fordeling af vagter og i et vist omfang medarbejdernes ønsker. Dette har markant forbedret ressourceanvendelsen.

Enkelte politikredse har ligeledes indført det samme type værktøj for vagtplanlægningen.



Der er identificeret seks forretningsområder, hvor institutionerne har gjort sig erfaringer med eller ser potentiale i at anvende machine learning

Forretningsområde	Risikobaseret tilsyn	Anbefalings-værktøjer	Tendensanalyser	Kundesupport	Effektiv sagsbehandling	Datakvalitet
Beskrivelse	Machine learning rummer et stort potentiale for at lave risikovurderinger, der kan danne grundlag for udvælgelse til risikobaseret tilsyn.	Machine learning kan indarbejdes i værktøjer, som giver konkrete anbefalinger til medarbejderen om den specifikke opgave, der udføres.	Machine learning har et potentiale for at identificere komplekse tendenser, der kan bruges i en strategisk indsats.	Telefonisk kundesupport og hjemmeside-optimering kan understøttes og forbedres med machine learnings-algoritmer.	Sagsbehandling af gentagende tekst-tunge sager rummer et stort potentiale for at blive bearbejdet helt eller delvist af machine learning	Machine learning kan bruges i et metadataperspektiv til at sikre, at de data, der bliver genereret, har en høj datakvalitet.
Machine learnings-metoder	Predictive modelling Text analytics	Predictive modelling	Klyngeanalyser Predictive modelling	Predictive modelling	Text analytics Predictive modelling	Predictive modelling Klyngeanalyser
Erfaringer og potentiale						

Erfaringer Potentiale



Algoritmedrevet risikobaseret tilsyn med data fra mange kilder, er vidt udbredt som idé, men hjemmelsspørgsmålet og årsagsforklaringsproblemet hæmmer erfaringerne



Ved at inddrage data fra mange datakilder kan en machine learning algoritme trænes til at vurdere risikoen for, at en virksomhed eller institution ikke overholder reglerne

Formål. Denne tilgang bruger predictive modelling til at lave præcise risikovurderinger af virksomheder/institutioner. Dette sker ved at inddrage mange variable fra flere forskellige datakilder for at lave komplekse profiler af de virksomheder/institutioner, der ikke overholder reglerne. De inddragede variable behøver ikke i sig selv at have en stor forklaringskraft, da det er helheden, der er relevant.

Udbredelse. Denne tilgang er vidt udbredt på idébasis, men konkrete erfaringer er begrænsede. Hele seks institutioner udtrykker, at de har en konkret idé til at anvende algoritmer i deres risikobaserede tilsyn og de er i gang med at afsøge mulighederne. Kun to institutioner har igangsat et projekt for at undersøge anvendeligheden og udfordringerne.

Lessons learned. Der er identificeret to hovedudfordringer ved denne tilgang. 1) Den væsentligste udfordring er hjemmelsspørgsmålet. Det er svært at skaffe lovhjemmel både til at indsamle data på tværs af forskellige institutioner og samkøre data til administrativt brug. 2) Den anden væsentlige udfordring er årsagsforklaringsproblemet. Institutioner har behov for at kunne begrunde overfor virksomheder, institutioner eller borgere årsagen til den pågældende risikovurdering, og dette er besværligt ved machine learning, da kompleksiteten i sammenhængende imellem variablene gør det vanskeligt at forklare begrundelsen for risikovurderingen.

Caseeksempel

Sammenligning af økonometriske og machine learnings-metoder til risikovurdering af virksomheders arbejdsmiljø



Arbejdstilsynet (AT) og Syddansk Universitet (SDU) foretager et toårigt projekt, der undersøger mulighederne i at anvende økonometri og/eller machine learning (ML) til risikobaseret udvælgelse til tilsyn samt afvejningen imellem de to tilgange.

Et centralt aspekt i projektet er afvejningen mellem præcise risikovurderinger og årsagsforklaringer. Ved at samkøre ATs data (fx tilsynsdata, data om arbejdsulykker mv.), med data fra STAR, SDU og Danmarks Statistik kan ML øge præcisionen i risikovurderinger af, om virksomheder overholder reglerne, men ML er udfordret på årsagsforklaringer. Økonometri har lavere præcision i risikovurderinger, men giver derimod dyb indsigt i årsagsforklaringer, der kan bruges i forbyggende indsatser.

Projektet skal oplyse AT om fordele og ulemper ved de to tilgange, men tilgangene anses ikke nødvendigvis som gensidigt udelukkende, de kan også være komplementære.



Kendskabet til algoritmedrevet risikobaseret tilsyn med meget data fra få kilder er ikke vidt udbredt, men implementeringen er større, da hjemmelsproblemet bliver irrelevant



Med store mængder data på et afgrænset område kan en machine learning algoritme finde mønstre og anomalier i data, der indikerer om en virksomhed eller institution ikke overholder reglerne

Formål. Denne tilgang bruger machine learning-algoritmer til at finde mønstre i store mængder data fra få kilder med henblik på at finde "normalen" i data. Algoritmen kan identificere, hvornår data afviger fra normalen. De unormale mønstre kan danne grundlag for en risikobaseret udvælgelse til kontrol, samt indikere konkrete opmærksomhedspunkter for kontrollanter.

Udbredelse. Denne tilgang med få datakilder er mindre udbredt som koncept end tilgangen med mange datakilder, men hos dem, der har indført det, er erfaringerne væsentligt større. Erhvervsstyrelsen og Landbrugsstyrelsen har hver idriftsat en løsning med machine learning, og Finanstilsynet har foretaget "proof of concept" på to løsninger.

Lessons learned. Årsagen til de større erfaringer er, at tilgangen hovedsageligt baserer sig på institutionernes egne data, hvorfor hjemmelsspørgsmålet bliver irrelevant. Den væsentligste udfordring er datakvalitet. Tilgangen kræver et afgrænset område, der indeholder store mængder data med en tilstrækkelig datakvalitet. Hos Landbrugsstyrelsen er det satellitbilleder på marker, hos Finanstilsynet er det teksttunge Own Risk and Solvency assessment rapporter og flere hundredtusind datapunkter for derivathandler.

Caseeksempel

Billedanalyser på satellitbilleder til at kontrollere tilskudsberettigelse



Landbrugsstyrelsen fører tilsyn med, om landmænd er berettigede til forskellige former for landbrugsstøtte. Et af kravene for berettigelsen til den grønne støtte er, at en andel af markarealerne er belagt med efterafgrøder for at mindske mængden af kvælstof i jorden.

For at kvalificere den risikobaserede udvælgelse til tilsyn anvendes machine learning i kombination med andre metoder til at foretage billedanalyser på satellitbilleder fra EU-satellitten Copernicus. På billederne kan det med en forholdsvis stor sandsynlighed bestemmes, om markarealerne er belagt med efterafgrøder, hvilket kan indgå i risikovurderingen af landmænd. Risikovurderingen kan således anvendes til at målrette tilsynsbesøgene mod de landmænd, der forventes ikke at overholde kravene.

Projektet er idriftsat, og Landbrugsstyrelsen ser et potentiale i at udvide metoden til at blive anvendt på andre områder.

Anbefalingsværktøjer kan gøre opgavespecifik information intelligent, hvilket gør sig gældende i otte projekter, men kun ét er implementeret



Machine learning kan indarbejdes i konkrete værktøjer, der kan understøtte og give anbefalinger til medarbejderne

Formål. Formålet med denne anvendelse af machine learning er, at en algoritme kan bearbejde store mængder historiske data og herefter give konkrete anbefalinger eller analytiske informationer til en medarbejder om en bestemt opgave. Tilgangen kan være en overbygning på opgavespecifik information, jf. s. 26 og 41, hvor en algoritme udleder anbefalinger om en konkret opgave på baggrund af data fra tilsvarende opgaver. Forskellen imellem algoritmedrevet anbefalingsværktøjer og opgavespecifik information er, at ved algoritmedrevet anbefalingsværktøjer udleder en algoritme anbefalinger på baggrund af alle data om opgavetypen, hvorimod i opgavespecifik information udleder medarbejderen selv anbefalinger på baggrund af konkret information om opgaven og egne erfaringer.

Udbredelse. Algoritmedrevet anbefalingsværktøjer rummer et stort potentiale hos institutionerne og de første erfaringer er gjort. Fire institutioner har erfaringer med denne anvendelse af machine learning fordelt på otte forskellige projekter.

Lessons learned. Kendetegnende for denne type værktøjer er, at de giver konkrete informationer eller anbefalinger om en opgave. Det skaber etiske problemer, når et analytics-værktøj giver anbefalinger om mennesker (se eksempel). På områder, der ikke omhandler mennesker, er denne type værktøjer meget operationelle og understøttende for medarbejderne.

Caseeksempel

Profileringsværktøj af elevers uddannelsesmuligheder



UNDERVISNINGS
MINISTERIET
STYRELSEN
FOR IT OG LÆRING

STIL har foretaget en "proof of concept" på en predictive model for forudsigelse af, om en elev har størst sandsynlighed for at gennemføre en erhvervs- eller gymnasial-uddannelse. Modellen viste, at det med 85 procent sikkerhed kan forudsiges om en elev vil gennemføre enten en erhvervs- eller gymnasial-uddannelse og det forventes, at modellen vil være væsentligt mere sikker, hvis det var muligt at anvende mere data fra fx Danmarks Statistik.

Profileringsværktøjet forventes ikke at blive implementeret foreløbigt, da der er et etisk spørgsmål om forudbestemt skæbne, hvor eleverne bliver fastlåst i nogle bestemte kasser, de ikke kan bryde ud af. I sin ideal anvendelse er værktøjet understøttende for ungdomsvejlederen, så eleven ikke bliver vejledt ned af en sti som mange andre elever er faret vild på. Men i den værste anvendelse af værktøjet, bliver det en fordelingsnøgle, hvor elever aldrig kan bryde ud af modellens forudsigelser.



Tendensanalyser til strategisk indsats er i mindre omfang udbredt på trods af, at hjemmelsproblemet delvist er irrelevant



Machine learning kan anvendes til at identificere makrotendenser, der kan være brugbart i det strategiske arbejde

Formål. Formålet med denne anvendelse af machine learning er at identificere nye tendenser, der kan bruges i en strategisk indsats. En væsentlig fordel ved machine learning, og særligt klyngeanalyser, er deres evne til at finde mønstre i data, som ikke ville træde frem i økonometriske metoder. I machine learning anvendes en induktiv tilgang, hvor data er styrende for analysen, hvilket kan give indsigt i nogle nye tendenser analytikeren ikke havde overvejet. I økonometri anvendes derimod en deduktiv tilgang, hvor teoretiske forventninger er styrende for analysen.

Udbredelse. Denne anvendelse af machine learning er i begrænset omfang udbredt. Styrelsen for Arbejdsmarked og Rekruttering har foretaget to POC på løsninger, der giver en unik indsigt i bevægelser og tendenser på arbejdsmarkedet og Sikkerhedsstyrelsen har haft et konsulentfirma til at foretage en analyse af mønstre på legetøjsmarkedet.

Lessons learned. Denne anvendelse er også udfordret af hjemmelsproblemet, da det kræver data på tværs af flere kilder og samkøring af disse data, men der er alligevel en væsentlig forskel fra de øvrige tilgange. Denne anvendelse har typisk ikke et administrativt formål, men et statistisk formål, hvorfor der er bredere beføjelser for, hvilke data der må samkøres jf. slide 76. Hjemmelsproblemet gør sig dog fortsat gældende for indhentning af data fra andre institutioner.

Caseeksempel

Netværksanalyse på beskæftigedes bevægelser på arbejdsmarkedet til identifikation af disruptionstrends



For at få indsigt i bevægelserne på arbejdsmarkedet har STAR indgået et samarbejde med forskere om udførelse af en netværksanalyse.

Analysen baserer sig på data om, hvilke job beskæftigede skifter fra, og hvilke job de skifter til. Netværksanalysen finder systematikker i, om der er nogle jobfunktioner, der oplever en særlig høj afstrømning eller tilstrømning. Analysen kan også identificere, om det fx er personer med en lang videregående uddannelse, der strømmer mod jobs, der typisk varetages af personer med en mellemlang videregående uddannelse.

Analysen kan bruges til at identificere disruptionstrends. Fx hvis en bestemt type job hurtigt oplever en stor afstrømning kan det skyldes at jobfunktionen er ved at blive disruptet.

Machine learning bliver i mindre omfang brugt til kundesupport og brugermålretning, men flere institutioner ser muligheder i at bruge det til forbedring af deres hjemmesider



Machine learning kan anvendes til at forbedre kundesupporten og/eller hjemmesider ved at forudsige kunders og brugers behov

Formål. Formålet med denne anvendelse er, at kunne lave "real time" forudsigelser af, hvad kunder eller brugere efterspørger. Dette søger at mindske den mængde tid, det tager for en kunde eller bruger at indhente information samt begrænser den mængde tid, en institution skal bruge på at vejlede en kunde eller bruger.

Udbredelse. Erfaringerne med denne anvendelse af machine learning er begrænsede. Der er identificeret tre institutioner, der har gjort sig erfaringer med denne type anvendelse af machine learning. Lægemiddelsstyrelsen har idriftsat en algoritme til at optimere forslagene på lægemiddelsbeskrivelser på produktresume.dk. Erhvervsstyrelsen er fortsat i projektudviklingsfasen (se eksempel), og Styrelsen for Arbejdsmarked og Rekruttering påtænker at igangsætte et projekt i 2018. Flere institutioner ser dog et potentiale i at bruge det til at brugermålrette deres hjemmesider, men konkrete erfaringer og projekter er fraværende.

Lessons learned. På trods af, at denne type anvendelse af machine learning er vidt udbredt i det private erhvervsliv, er der overraskende få institutioner, der gør brug af denne anvendelse. Dette kan skyldes, at flere statslige institutioner har begrænset kunde- og brugerkontakt, hvorfor businesscasen sjældent er til stede for denne type løsninger.

Caseeksempel

Erhvervsstyrelsen kundesupport på VIRK



Erhvervsstyrelsen er i gang med at gennemføre et Proof of Concept på at optimere deres kundesupport til virksomhedsportalen virk.dk.

Erhvervsstyrelsen har installeret en såkaldt "tæller" på virk.dk. Tælleren registrerer kundens bevægelser på hjemmesiden, og sender data videre til lagring i deres datavarehus. Hvis kunden efterfølgende ringer til kundesupport, videresendes informationen til en machine learnings algoritme, der sandsynliggør, hvilket problem kunden ringer ind for at få hjælp til. Dette bliver udstillet til supporteren, som derved hurtigt kan identificere, hvilken problematik kunden ønsker support til. Løsningen er fortsat under udvikling og trænes hele tiden til at blive mere og mere præcis.

Løsningen kan både effektivisere kundesupporten betydeligt ved at minimere opkaldslængden samt højne kvaliteten i supporten, da kunden får hurtig og præcis vejledning på deres spørgsmål.

Kvalitetssikring af data ved machine learning rummer et potentiale for at øge datakvaliteten i staten, men der er kun enkelte institutioner med konkrete erfaringer



Machine learning kan anvendes til at sikre datakvaliteten, og/eller påpege opmærksomhedspunkter i data

Formål. Formålet med denne anvendelse er, at machine learning kan bruges i et metadataperspektiv til at finde mønstre og systematikker i hvornår datakvaliteten er lav. Algoritmen bliver således et redskab til at rense data og højne datakvaliteten samt til at få en indsigt i hvor store problematikker, der er med datakvaliteten.

Udbredelse. Denne anvendelse af machine learning begrænser sig til tre institutioner. Miljøstyrelsen har idriftsat en løsning på kvalitetssikring af vandmiljøprøver, Fødevarestyrelsen har foretaget POC på et mindre og selvudviklet projekt til kvalitetssikring af kontrolscores i kontrolrapporter, og GEUS er i forberedelsesfasen til at igangsætte et forsøg med at anvende machine learning til at kvalitetssikre visse dele af deres store mængder data.

Lessons learned. Denne anvendelse er forholdsvis tilgængelig for institutioner, da det er kvalitetssikring på egne data. Derudover ville denne type anvendelse af machine learning være anvendelig hos mange institutioner. Dette skyldes, at flere institutioner giver udtryk for at manglende datakvalitet er en væsentlig stopklods for aktiv anvendelse af data i opgavevaretagelsen.

Caseeksempel

Miljøstyrelsen kvalitetssikring af vandmiljøprøver



Miljøstyrelsen (MST) foretog i efteråret 2017 et Proof of Concept i samarbejde med eksterne konsulenter på kvalitetssikring af vandmiljøprøver.

MST overvåger vandets miljø og økologiske tilstand ved at foretage laboratorietests af vandmiljøprøver fra målestationer placeret i åer og vandløb rundt omkring i Danmark. Problemet er, at målestationerne engang imellem er behæftet med fejl, hvilket kontaminerer prøverne. For at disse prøver ikke indgår i vandmiljødata, og at de omkostningsfulde processer vedrørende indhentning af prøver, laboratorietests osv. ikke igangsættes, har MST udviklet en machine learning-algoritme, der kan lave en risikovurdering af, hvornår en vandmiljøprøve er kontamineret. Dette sker ved, at målestationerne indsender data via sensorer til algoritmen, der analyserer data og giver en konkret risikovurdering.

Projektet forventes at blive idriftsat i løbet af 2018.

Effektiv sagsbehandling ved machine learning rummer et enormt potentiale i takt med, at de danske sprogkompetencer inden for kunstig intelligens udvikles



Machine learning, særligt i form af tekst analytics, kan anvendes til effektiv sagsbehandling

Formål. Formålet med denne anvendelse af machine learning er, at metoderne bruges til at understøtte sagsbehandlingen ved at foretage dele af sagsbehandlingen automatisk. Den machine learnings-metode, der oftest bliver brugt i denne type anvendelse, er tekst analytics. Tekst analytics kan bearbejde store mængder tekstmateriale, der ville være meget tidskrævende for en medarbejder at gennemlæse og sagsbehandle. Tekst analytics er ikke en erstatning af sagsbehandleren, men et redskab, der kan foretage den indledende sagsbehandling og fokusere sagsbehandlerens arbejde.

Udbredelse. To institutioner har gjort sig erfaringer med machine learning til effektiv sagsbehandling. Nævnenes Hus har igangsat et POC på en machine learnings algoritme til præcedensdannelse. Miljøstyrelsen har nedlagt et tekst analytics projekt til sagsbehandling af produktgodkendelsesansøgninger.

Lessons learned. Udfordringen for effektiv sagsbehandling med tekst analytics er de manglende danske sprogkunderskaber inden for kunstig intelligens. Det er en kæmpe opgave at træne algoritmer i danske sprogkunderskaber både grammatisk og semantisk. Dette forstærkes af, at det fundament inden for danske sprogkunderskaber i kunstig intelligens, man kan bygge videre på i et projekt, er sparsomt. På grund af for dårlige sprogkunderskaber i kunstig intelligens, i forhold til hvad der blev krævet af sagsbehandlingen, nedlagde Miljøstyrelsen et POC efter tre uger som ellers skulle vare seks uger.

Caseeksempel

Nævnenes Hus udtræk af sager til præcedens



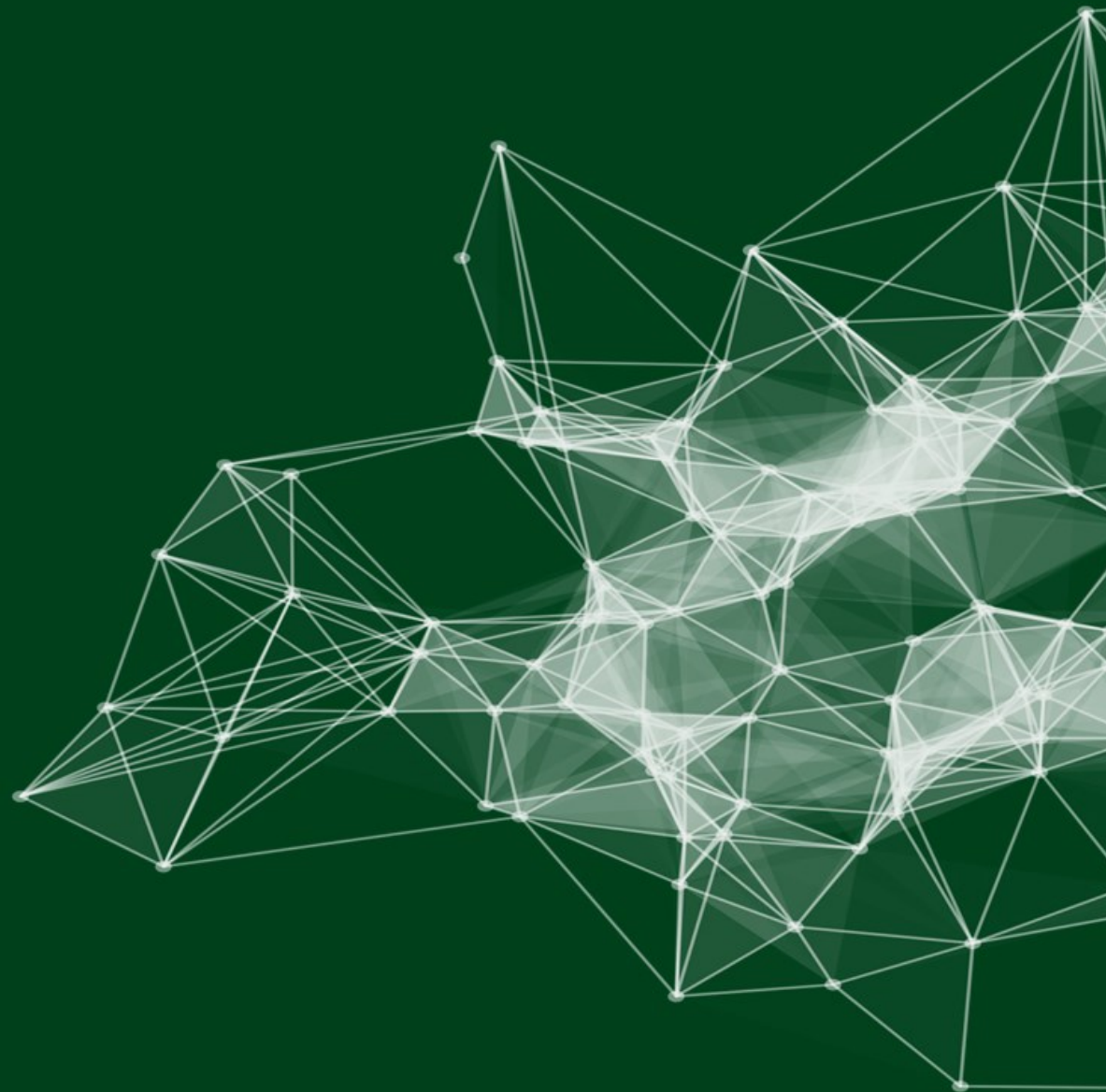
Nævnenes Hus er anden instans på klagesager og sagsbehandler mange klagesager. For at forbedre kvaliteten i afgørelserne og mindske behandlingstiden vil Nævnenes Hus udvikle en machine learning-algoritme, der kan aflæse indkomne klagesager, og derefter automatisk fremsøge lignende sager og afgørelser i deres fagsystemer. Dette giver sagsbehandleren mulighed for en mere effektiv sagsbehandling samt sikrer høj kvalitet i afgørelserne.

På sigt ønsker Nævnenes Hus, hvis muligt, at stille algoritmen til rådighed for kommuner, statslige institutioner og regioner, så effektiviseringen af sagsbehandlingen og kvalificeringen af afgørelserne kan udbredes til hele den offentlige sektor. På længere sigt vil algoritmen også kunne give en konkret vurdering af sandsynligheden for medhold i klagen til borgeren i klageøjeblikket.

Projektet vil potentielt kvalificere og effektivisere klageprocessen for anden instans, første instans og borgeren.

Sektion 3

Datamiljø



Introduktion: Vi gennemgår i denne sektion statens datamiljø



Hvilke data har institutionerne, og hvordan bliver de forvaltet?

Digitalisering og nye teknologier har medført, at der produceres store mængder af data fra forskellige kilder. De fleste har kendskab til registersamlinger, men hvilke andre data gemmer der sig i institutionernes databaser?

I forlængelse af dette undersøger vi, hvordan institutionerne har organiseret deres datamanagement ift. dataarkitektur, databehandling og datagovernance. Med øgede krav til databeskyttelse og potentialer ved udnyttelsen af nyeste teknologiske muligheder, er der behov for at styrke fundamentet for smidig og effektiv organisering og styring af data.

I kapitel 5 beskriver vi institutionernes forskellige datatyper.

I kapitel 6 undersøger vi institutionernes datamanagement.

Kapitel 5

Hvilke data har staten?



Datatyper. Vi beskriver, hvilke strukturerede og ustrukturerede data institutionerne har.

Kapitel 6

Hvordan organiserer og styrer staten data?



Datamanagement. Vi viser, hvordan institutionerne har organiseret deres dataarkitektur, databehandling og datagovernance.

Baggrund: Digitaliseringen af bureaukratiet fra 1960 og frem



Strøm til papir, PC'en og internettet

Digitaliseringen af bureaukratiet startede i 1960erne med de første computere til elektronisk databehandling, og oprettelsen af "Statens Datacentral" i 1959. De første datatyper i staten var løn- og pensionsberegninger, men også andre regel- og rutineprægede administrative opgaver.

Den digitale forandring tog for alvor fart med lanceringen af pc'en i midten af 1980erne. Det gjorde det muligt for institutionerne at administrere de politisk fastlagte økonomiske rammer i tråd med moderniseringsprogrammet fra 1983, herunder fx indkøb, løbende budgetopfølgning, registrering af sygedage mv.

Det næste store digitale skridt blev taget med fremkomsten af internettet i 1990erne. Nu kunne institutionerne præsentere deres fagområder og kerneopgaver via hjemmesider, men også digitalisere borger- og virksomhedsrettede serviceydelser og indberetninger via elektroniske blanketter, formularer mv.



Digital koordinering

Den nationale it-strategi fra 1999 anbefalede at skabe overblik over it-udviklingen samt en øget koordineret indsats i lyset af blandt andet sagen om de tidligere amters udvikling af 14 forskellige elektroniske patientjournaler med forskellige tekniske standarder, der ikke kunne kommunikere med hinanden.

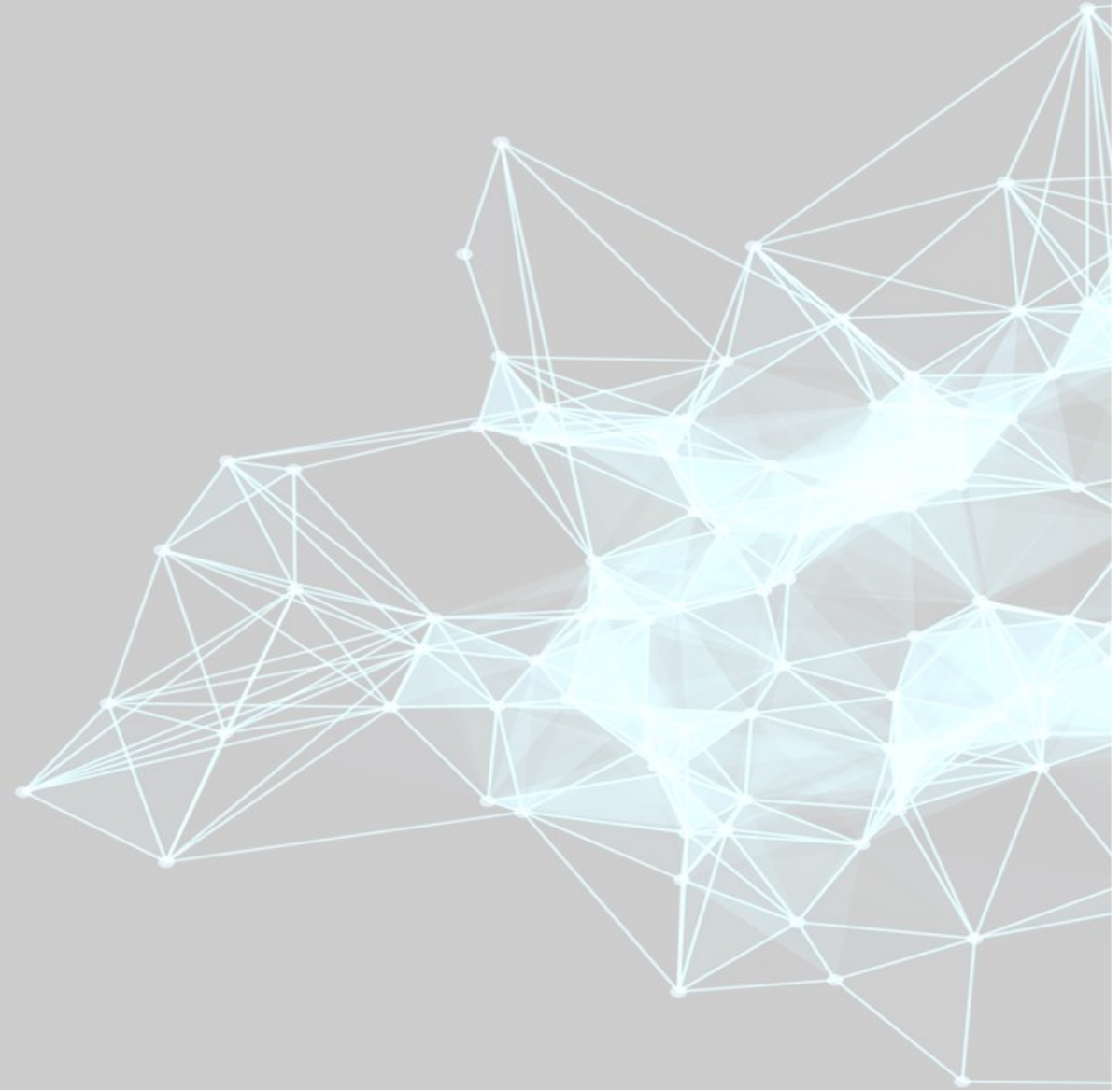
For at styrke den digitale koordination og styring udviklede "Den Digitale Taskforce" den første fællesoffentlige digitaliseringsstrategi 2001-2004 og er senere blevet fulgt op af nye strategier, herunder senest den fællesoffentlige digitaliseringsstrategi 2016-2020 og strategi for it-styring i staten 2017.

Et af de store indsatsområder var standardiseringen og sammenkoblingen af en række "øer af data" med henblik på at kunne udveksle data på tværs af myndigheder. Det kunne fx være digitaliseringen af CPR-registret. Dette arbejde taler særligt ind i datagovernance, som vi berører under datamanagement.



Kapitel 5

Datatypes



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

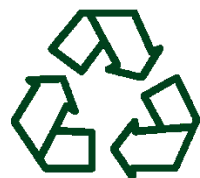
Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Datatyper

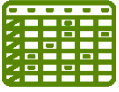


Hvilke typer data har staten?

Vi beskriver institutionernes strukturerede og ustrukturerede data, og har følgende konklusioner:

1. De fleste institutioner producerer eller anvender rigtig mange strukturerede data, og det spænder fra adresseregistre til vandmiljø over smileyordninger
2. Institutioner producerer, modtager og opbevarer ustruktureret tekstdata i word- og pdf-filer, men de færreste institutioner gennemfører analyser på deres tekstdata
3. Flere institutioner arbejder med billeddata – fx fotos af planter, elledninger og fly- og satellitbilleder til landkort og oversvømmelser
4. Enkelte institutioner arbejder med lyddata i form af fx radiomateriale og lydgener i toget
5. Flere institutioner ser et analytisk potentiale i big data fra sociale medier, sensorer og andre datakilder
6. Fremkomsten af big data og nye teknologier skaber et eksternt pres fra omverden på institutionernes måde at løse kerneopgaven

De fleste institutioner producerer eller anvender rigtig mange strukturerede data, og det spænder fra registerdata over tekniske data til administrative data



De fleste institutioner har forskellige strukturerede data, som de enten gennemfører analyser på eller stiller til rådighed for virksomheder, borgere og forskere

Strukturerede data refererer til dataoplysninger som har en høj grad af organisering med forudbestemte sæt af regler, der definerer datahierarkiet og forbindelsen imellem dem.

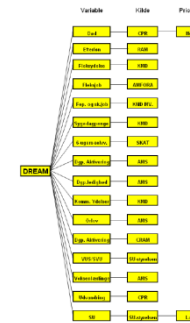
Der findes **registerdata**, der vedrører data indsamlet med et administrativt formål af myndigheder, og oplysninger der registreres på person-, virksomheds- eller objektniveau. Det kunne eksempelvis være Danmarks Adresseregister (SDFE), smileyordning over virksomheders arbejdsmiljø (Arbejdstilsynet), registrering af besætninger og husdyr (Fødevarestyrelsen), forløbsdata om personers beskæftigelse i databasen DREAM (STAR) mv. Flere institutioner beriger i øvrigt deres data med Danmarks Statistiks registersamlinger.

Derudover er der **tekniske data**, der vedrører myndighedernes fagtekniske oplysninger. Det kunne eksempelvis være data om vandmiljøets tilstand (Miljøstyrelsen), og data om det danske eltransmissionsnet (Energinet).

Der findes også **administrative data**, der vedrører myndighedens interne forhold om HR, økonomi, aktiviteter mv. Disse data kan fx bruges til ledelsesinformation med henblik på prioritering mellem opgaver og ressourcer.

Caseeksempel

Forløbsdata om personers beskæftigelse



Styrelsen for Arbejdsmarked og Rekruttering bruger DREAM databasen til at lave forløbsanalyser over tid, hvor der er fokus på skift mellem ordninger og varigheden af givne ydelsesperioder med henblik på at identificere eventuelle marginaliserede persongrupper, effekten af særlige indsatser, estimere ledighedsancienniteten mv.

DREAM databasen indeholder oplysninger om etnisk herkomst, bopælskommune, dødsfald, overførselsindkomst mv.

DREAM bygger på flere myndigheders registre, herunder data fra SKAT, Beskæftigelsesministeriet og Uddannelse- og forskningsministeriet.

Institutioner producerer, modtager og opbevarer ustruktureret tekstdata i word- og pdf-filer, men de færreste institutioner gennemfører analyser på deres tekstdata



Institutioner producerer, modtager og opbevarer rigtig meget ustruktureret tekstdata

Ustruktureret data er passive aktiver/oplysninger, der har en lav grad af organisering, som fx tekstindholdet i tilsynsrapporter. Der bliver produceret og journaliseret meget skriftligt materiale i de statslige institutioner, men institutioner modtager også mange dokumenter fra virksomheder og borgere.

Der findes **klagesagsafgørelser**, herunder fx i forbindelse med klager på miljø- og fødevarerområdet (Nævnenes Hus) eller klager over sundhedsfaglig behandling (Styrelsen for Patientsikkerhed).

Derudover findes der rigtig mange forskellige **tilsynsrapporter**, herunder fx på områder vedrørende elinstallation (Sikkerhedsstyrelsen), regnskab (Erhvervsstyrelsen, Finanstilsynet), miljøbeskyttelse (Miljøstyrelsen), jordbrug og planteavl (Landbrugsstyrelsen).

Der findes **produktgodkendelser**, fx godkendelse af produkter i henhold til en række miljøkrav (Miljøstyrelsen), men også sikkerhedskrav (Sikkerhedsstyrelsen). Så er der også en helt særlig opgave, der vedrører opbevaring af **kulturhistoriske tekster** mv. (Rigsarkivet).

De fleste institutioner analyserer ikke på deres tekstdata, men der er enkelte institutioner som fx Miljøstyrelsen, der har testet tekst analytics til at bearbejde produktgodkendelser i henhold til miljølovgivningen. Projektet blev lukket før tid på grund af analytics løsningens utilstrækkelige dansk kundskaber.

Caseeksempel Eleftersynsrapporter



Eleftersyn er en del af huseftersynsordningen, hvis hovedformål er at beskytte køber og sælger mod problemer med mangler ved fast ejendom i forbindelse med køb og salg.

Huseftersynet indeholder både en tilstandsrapport og en elinstallationsrapport.

Elinstallationsrapporten er en skriftlig rapport, der indeholder resultater af gennemgangen af bygningens elinstallationer, herunder i hvilket omfang elinstallationerne er funktionsdygtige og lovlige.

Rapporten bliver udarbejdet af private elinstallatører, men det er Sikkerhedsstyrelsen, der skal føre tilsyn hermed.

Flere institutioner arbejder med billeddata – fx fotos af planter, elledninger og fly- og satellitbilleder til landkort og oversvømmelser



Institutioner tager mange forskellige fotos, og bruger dagligt fly- og satellitbilleder i udførelsen af deres kerneopgave

Institutioner tager billeder i forskellige sammenhænge. Miljøstyrelsen tager fx billeder af dyr, planter og svampe af den danske natur, og Energinet tager dronebilleder af elledninger med henblik på vedligehold.

For institutioner, der arbejder med at kortlægge Danmark, forvalte og kontrollere miljø- og naturforhold eller studere klimaforandringer, er satellitbilleder og luftfotos en afgørende kilde til nøjagtig viden. Det kan eksempelvis være satellitbilleder til produktion af nye landkort over Grønland og Danmarks Højdemodel til at kortlægge vandets indtrængen, oversvømmelsesrisiko og udbredelse i by, kyster og vandløb (Styrelsen for Dataforsyning og Effektivisering). Eller satellitbilleder fra EU's Copernicus Program til overvågning af landbrugsarealer (Landbrugsstyrelsen).

Caseeksempel Copernicus



Det europæiske jordobservationsprogram Copernicus er verdens største satellitprogram, der via sine Sentinel-satellitter leverer frie data og tjenester til brug for både erhvervslivet, myndigheder og til private brugere. Detaljeringsgraden, kvaliteten og tilgængeligheden af data indebærer en markant forøgelse af mulighederne for at anvende og nyttiggøre satellitdata.

Satellitbilleder af Danmark bliver produceret, når den enkelte satellit passerer Danmark på sin vej omkring jordkloden. Nogle satellitter passerer Danmark hver tredje dag, mens andre passerer en gang om måneden. Optiske satellitter indsamler kun data for de arealer på jorden, der er skyfri på det tidspunkt, hvor satellitten passerer, mens radarsatellitter kan indsamle data gennem skydækket. Det er forventningen, at opmåling ved brug af satellitter fremover vil afløse flere af de traditionelle metoder som bl.a. foto optaget fra fly.

Enkelte institutioner arbejder med lyddata i form af fx radiomateriale og lydgener i toget



Store dele af den moderne kulturarv er formidlet i lydformat i form af radio og video

Rigsarkivet opbevarer den danske kulturarv. Dette indbefatter også de enorme mængder radio- og videomateriale, der er blevet produceret op igennem 1900 tallet og i det 21. århundrede. Denne type data indeholder vigtige historiske informationer og kan bruges både i forskning og underholdningshenseender. For at gøre materialet mere tilgængeligt er Rigsarkivet i gang med at foretage "tagging" på deres materiale. Dette betyder, at de sætter digitale "tags" på hvad materialet indeholder fx hvilke emner radioavisen i 1947 behandlede.

Lyddata kan også anvendes til at identificere vedligehold

Når en mekaniker, reparerer en bil bruger han flere sanser til at diagnosticere, hvad der er i vejen med bilen. Synet til at identificere rustskader, lugtesansen til at identificere filterfejl ved udstødningen og hørelsen til at høre om gearkassen knasser.

Disse metoder, der bliver brugt af faglærte personer med stor praktisk erfaring, kan i visse tilfælde automatiseres og overføres til et computerprogram. Dette gør sig blandt andet gældende for Banedanmark, der ser et potentiale i at installere mikrofoner i togene, som kan bidrage til fx at identificere hvornår tærsklerne for lydgener bliver overskredet i togene. Energinet har også igangsat et projekt, for at teste om lyde kan blive anvendt til at identificere vedligehold på transformatorstationer.

Caseeksempel

Den digitale stationsvagt for transformatorstationer

ENERGINET

I gamle dage undersøgte en stationsvagt transformatorstationernes tilstand ved en fysisk besigtigelse af stationerne. Dette foregik ved at stationsvagten kiggede, mærkede og lyttede til om transformatorstationerne trængte til vedligeholdelse.

Med inspiration i den gamle stationsvagt igangsatte Energinet et projekt, hvor man installerede mikrofoner på enkelte transformatorstationer. Den lyddata man indsamlede skulle indgå i træningen af en machine learning algoritme, der kan identificere mønstre i data for, hvordan en transformatorstation lyder lige inden den går i stykker.

Med denne viden ville det være muligt at foretage præventive vedligehold på transformatorstationer og metoden kan potentielt udbredes til andre områder.

Denne metode er en del af den store gren inden for predictive modellering, der vedrører præventivt vedligehold.

Flere institutioner ser et analytisk potentiale i big data fra sociale medier, sensorer og andre datakilder



Nye teknologier genererer nye mangfoldige data med en hidtil uset computerkraft, der igen medfører stigende datamængder. Flere institutioner ser muligheder ved dette.

Big data karakteriseres som regel ved de såkaldte tre V'er. Det er kendetegnet ved de eksponentielt stigende mængder data (volume), mangfoldigheden i data som genereres på sociale medier, IoT og andre teknologier (variety) og hastigheden i datagenereringen på grund af øget computerkraft (velocity).

Flere institutioner ser potentialer i big data på deres fagområde. Der er fx perspektiver i at anvende unikke danske sundhedsregistre, sociale medier og real world data fra fx wearables mv. til at kunne gennemføre godkendelser af nye lovende lægemidler til livstruende sygdomme hurtigere (Lægemiddelstyrelsen). Sociale medier vil også kunne bruges til at forudsige legetøjstrends (Sikkerhedsstyrelsen) eller til webscrapping for at spore karteldannelser (Konkurrence- og Forbrugerstyrelsen). Viden om kundernes trafik på hjemmesider kan forbedre kundesupporten (Erhvervsstyrelsen), og det undersøges om mobildata vil kunne bruges til at tælle antal besøgende i fx slotsparker (Slots- og Kulturstyrelsen).

I forhold til sensorer vil fx signalsystemerne på hele jernbanenettet udskiftes frem mod 2030 og vil kunne opsamle flere og mere præcise informationer om, hvor togene er, ankomsttidspunkt og afvigelser i forhold til køreplanen (Banedanmark). Dette gælder også på energiområdet, hvor intelligente hjem kan styre forbrug og produktion mere præcist og smart (Energinet).

Caseeksempel

Data fra sociale medier til lægemiddelsgodkendelse



LÆGEMIDDELSTYRELSEN

Lægemiddelstyrelsen overvåger bivirkningerne af lægemiddel på baggrund af data på indberettede bivirkninger, men lægemiddelstyrelsen ser et potentiale i at udvide datagrundlaget og få større indsigt i de langsigtede bivirkninger af forskellige lægemidler.

Til at få større indsigt i de langsigtede bivirkninger ser Lægemiddelstyrelsen et stort potentiale i at anvende "Real World Data". Disse data indbefatter data, der bliver genereret efter et lægemiddel er blevet markedsgodkendt. Dette kan fx være data fra flere kliniske databaser, såsom Landspatient-registret, dødsårsagsregistret mv. Derudover kan det også være de enorme mængder data, der bliver generet fra forskellige sundhedsure (applewatch og fitbit), registreringer i sundheds-apps om kost, motion, søvn mv. Alle disse data indeholder enorme informationer, der kan være afgørende for indsigten i et lægemiddels virkninger.



Fremkomsten af big data og nye teknologier skaber et eksternt pres fra omverdenen på institutionernes måde at løse kerneopgaven



Der bliver genereret mere og mere data, og den private sektor udnytter mulighederne gennem nye teknologier. Det medfører i visse tilfælde at opgaveløsningen i det offentlige skal gentænkes

Institutionerne har givet udtryk for to områder, hvor der opstår et pres fra omverdenen.

For det første er der en stigende samfundspolitisk forventning om, at sagsbehandlingen kan og skal gennemføres hurtigere. Miljøstyrelsen har eksempelvis, som nævnt tidligere, afsøgt mulighederne for at anvende tekst analytics til at reducere sagsbehandlingstiden på produktgodkendelser ift. miljøkrav. Fødevarestyrelsen afsøger også muligheden for at anvende en blockchain-baseret teknologi til at strømline processen for tilbagekaldelse af produkter, der ikke overholder fødevarelovgivningen.

For det andet medfører udviklingen af nye former for big data nye krav til, hvordan de offentlige myndigheder fører tilsyn. Udviklingen af derivathandler i den finansielle sektor foregår fx i et tempo og omfang, der gør det umuligt for Finanstilsynet at føre et effektivt tilsyn af disse uden anvendelse af machine learning-metoder. Dette gælder også for Erhvervsstyrelsen, hvor digitaliseringen af virksomhedsregistreringer gør, at flere data bliver mere tilgængelige, hvilket har stor betydning for behovet, tilrettelæggelse og omfanget af tilsyn og kontrol på deres område.

Caseeksempel

Derivathandler sætter nye standarder



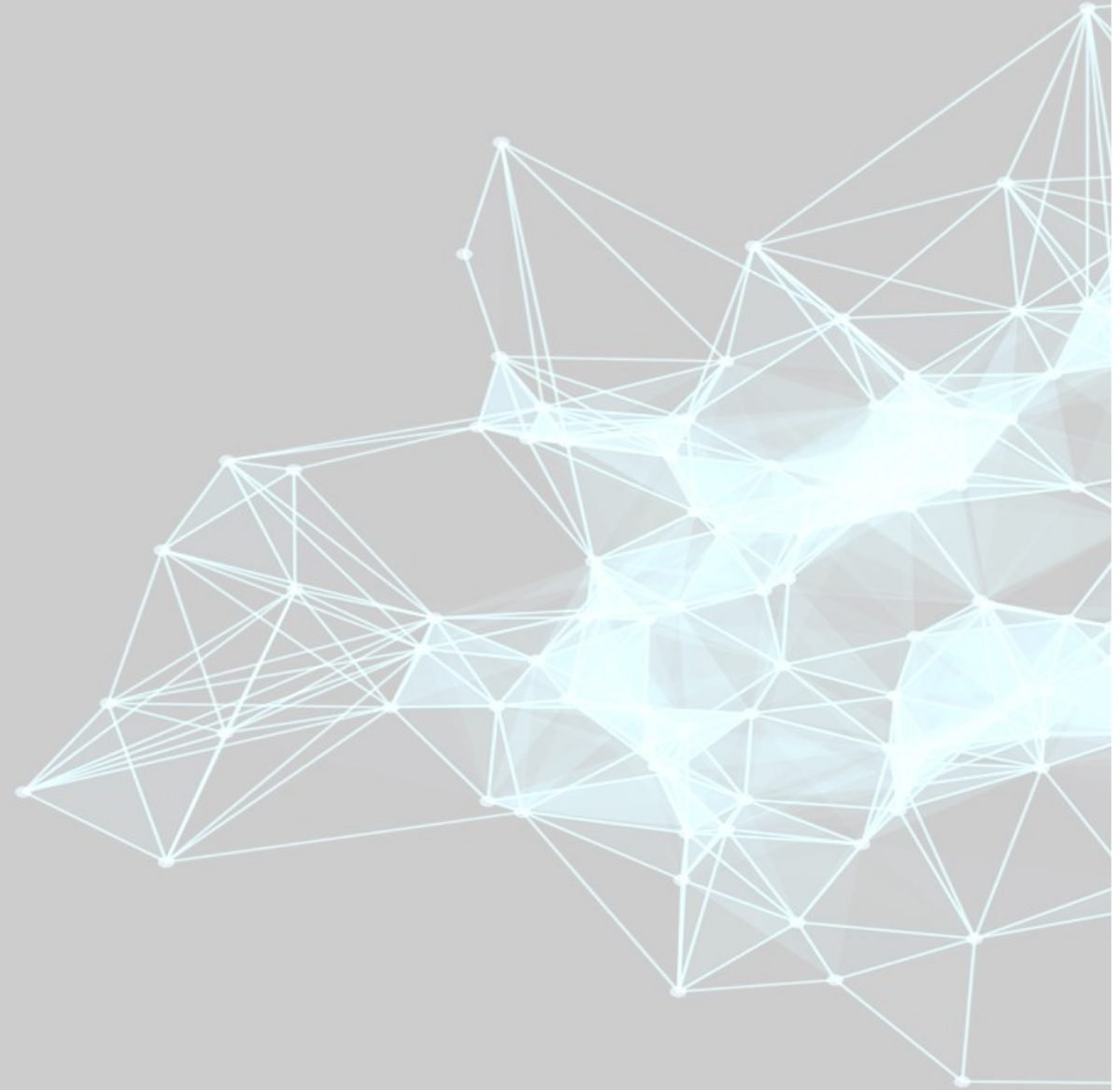
Der bliver hver dag indgået tusindvis af derivathandler. Et derivat er et værdipapir, der bygger ovenpå et underliggende produkt fx en aftale om, at man vil sælge en bestemt aktie eller mængde råvare på en bestemt dato i fremtiden. Markedet for derivathandler har eksploderet over flere årtier og det var en af de væsentlige faktorer bag finanskrisen.

Opmærksomheden på at derivathandler kan og har haft markante indflydelser på finansmarkedet, har skabt et øget fokus på tilsyn med derivathandlerne. Problemet er at mængden og hastigheden i hvor hurtigt handlerne bliver indgået, foregår i et tempo, der sætter store krav til finanstilsynets systemer. Dette har medført, at de er begyndt at afsøge mulighederne for at bruge machine learning til at bearbejde data, da disse metoder er blandt de få, der kan håndtere datamængden og kompleksiteten i data.

Projektet er foreløbigt implementeret på prøvebasis som supplement til deres normale systemer, for at medarbejderne kan blive fortrolige med værktøjerne.

Kapitel 6

Datamanagement



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

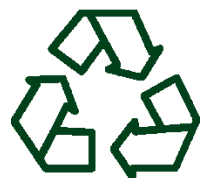
Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Datamanagement



Hvordan organiserer og styrer de statslige institutioner deres data?

I dette kapitel behandler vi de forskellige komponenter af datamanagement, samt juridiske udfordringer ved datadeling og samkøring.

Vi præsenterer følgende 11 hovedkonklusioner:

1. Der er tre væsentlige forhold i organiseringen og styringen af data: Dataarkitektur, databehandling og datagovernance
2. Flere institutioner har databaser på deres egen lokation, der typisk bliver udviklet og implementeret af konsulenter/Danmarks Statistik, og enkelte bruger cloud computing
3. Enkelte institutioner fremhæver, at cloud computing sikrer teknisk og forretningsmæssig fleksibilitet i anvendelsen af analytics, og bør prioriteres af Statens It
4. Flere institutioner samler og tilgår deres og andres data hos Danmarks Statistik
5. Institutioner fremhæver at databehandling er en ressourcetung opgave, men at strukturering og rensning af data giver værdifuld indsigt i potentialer inden for analytics
6. Enkelte institutioner bruger konsulenter/Danmarks Statistik til dele af databehandlingen, men det er uklart i hvilket omfang
7. Enkelte institutioner anvender Danmarks Statistik til at opgøre og formidle statistik på deres fagområde
8. Flere institutioner har udfordringer med organisationens datagovernance med den konsekvens at potentialet af data ikke udnyttes optimalt
9. Flere institutioner vil løse problemer med datagovernance ved anvendelse af "data stewards" og robotautomatisering
10. På grund af juridiske hensyn etablerer flere institutioner adskilte databaser til henholdsvis administrative og statistiske formål
11. Enkelte institutioner har juridiske udfordringer med at dele og modtage relevante data fra andre myndigheder og virksomheder
12. Flere institutioner er usikre på juraen for styringen af data, og er i gang med at afdække databeskyttelsesforordningens konsekvenser i forhold til datasamkøring

Der er tre væsentlige forhold i organiseringen og styringen af data: Dataarkitektur, databehandling og datagovernance



Dataarkitektur

... dækker over **opbygningen, implementeringen og driften af it-systemer**. Dette omfatter fx netværk, servere, storage og standarder for disse, så det er muligt at systemerne kan "tale" sammen. Andre centrale aspekter af dataarkitekturen er driften af operativsystemer og middle-ware, servicedesk mv., og vedligehold og drift af applikationer og fagsystemer.

Et væsentligt element er etableringen af databaser, hvor data fra fagsystemer kan lagres sikkert og struktureret. Det vedrører også at kunne stille analytiske værktøjer, såsom statistik- og visualiseringsprogrammer til rådighed, og i det hele taget sikre et it-miljø, der følger med den teknologiske udvikling.



Databehandling

... er den menneskelige behandling af data, snarere end dataarkitekturens mere tekniske fokus. Databehandling er **struktureringen, rensningen, berigelsen og valideringen af data til det videre analysearbejde**, fx opdeling af institutioner i meningsfulde kategorier til et analytisk formål, rensning af data for dubletter, overblik over manglende datapunkter mv.

Et væsentligt element i databehandlingen er dataudtræk (SQL, ETL, data wrangling mv.). Rådata, der bliver udtrukket fra en datakilde, skal altid formateres og systematiseres til lagring i en database, og dataudtræk fra en database skal næsten altid kategoriseres og bearbejdes i forhold til analyseformålet.



Datagovernance

... vedrører **regler og procedurer for forvaltning af data internt og eksternt i institutionen**, som fx information om data (metadata), politik for adgang, deling og genbrug (ejerskab) mv.

Et væsentligt element her er at behandle data i henhold til databeskyttelsesregler, men også sikre ensartet og konsistente definitioner af data med henblik på at forbedre datagrundlaget, og herved øge mulighederne for at finde nye tværgående indsigter i data.

Flere institutioner har databaser på deres egen lokation, der typisk bliver udviklet og implementeret af konsulenter/Danmarks Statistik, og enkelte bruger cloud computing



Flere institutioner har deres data opbevaret på egen lokation, mens enkelte bruger cloud computing.

Institutionernes dataarkitektur organiserer og opbevare data ved tre forskellige databasetyper.

Størstedelen af institutionerne har et datavarehus, som deres primære database, hvor data samles og tilgås. Disse bliver typisk udviklet og implementeret af konsulenter, men driftes af institutionerne selv, bortset fra enkelte sikkerhedsopdateringer. Finanstilsynet, Styrelsen for Patientsikkerhed, Landbrugsstyrelsen og GEUS er eksempler på institutioner, hvis primære database er et datavarehus.

Visse institutioner samarbejder med Danmarks Statistik i oprettelsen af et datavarehus. Uddannelses- og Forskningsministeriet har etableret et datavarehus i samarbejde med Danmarks Statistik, med data om de selvejende uddannelsesinstitutioner. Desuden indsamler Børne- og Socialministeriet data på socialområdet i samarbejde med Danmarks Statistik, og opbevarer data på Danmarks Statistiks ministerieordning.

Enkelte institutioner bruger i stedet cloud computing – se definition på næste slide – til at opbevare data. Dette gør sig gældende for Erhvervsstyrelsen som i samarbejde med konsulenter er overgået til at have en cloud baseret løsning.

Perspektivering

Samling af basal it-drift hos Statens It



Enkelte institutioner nævner, at ansvaret for driften af deres datavarehus skal overdrages til Statens It. Statens It varetager blandt andet kontrakt- og leverandørstyring af outsourcet it-drift, og konkurrenceudsætter denne i rammeaftaler.

Statens It har i december 2015 udarbejdet en strategi for selektiv sourcing, hvor der ved tilgang af nye eller større opgaver vurderes, om opgaven skal outsources helt, delvist eller løftes selv (egenproduktion). Strategien er i tråd med den generelle trend på markedet, hvor institutioner outsourcer udvalgte dele af deres it-systemer, samtidig med at de bibeholder eget datacenter.

Eksempler på dette er Nordea og SKAT, der har valgt at hjemtage dele af tidligere outsourcete it-driftsopgaver med henblik på at kunne sikre større kontrol og opretholdelse af nødvendige it-kompetencer in-house.

Enkelte institutioner fremhæver, at cloud computing sikrer teknisk og forretningsmæssig fleksibilitet i anvendelsen af analytics, og bør prioriteres af Statens It



Cloud computing giver ifølge enkelte institutioner bedre muligheder for at udnytte potentialerne ved analytics, herunder særligt machine learning

Den teknologiske udvikling forventes i de kommende år at ændre it-driftsopgavens karakter markant, og den generelle tendens på markedet for it-drift er mod virtualiseret serverdrift og cloud computing.

Med virtualiseret serverdrift fjernes afhængigheden mellem it-systemer (applikationer) og de underlæggende fysiske servere ved brug af virtualiseringssoftware. Ved virtualiseret serverdrift er det muligt at drifte flere it-systemer på det samme antal fysiske servere. En af fordelene ved anvendelse af denne teknologi er, at omkostningerne til anskaffelse og drift af fysiske servere (særligt strømforbruget) reduceres markant.

Cloud computing er en forretningsmodel, hvor man køber standardiserede it-produkter (software, hardware eller tjenesteydelser) som en service over internettet. Kunden skal dermed ikke installere og vedligeholde it-produkterne selv 'in-house'. Cloud computing giver mulighed for hurtig op- og nedskalering af kapaciteten efter forbrug og behov, hvilket eksempelvis kan reducere kødannelse, når mange borgere på samme tid ønsker at logge på en given tjeneste. Det kan give øget computerkraft og hukommelse til analyse af større datasæt, og virtuel adgang til algoritmer via skyen. Cloud computing indgår som initiativ i den fællesoffentlige digitaliseringsstrategi for 2016-2020.

Perspektivering

Cloud computing i Statens It



Enkelte institutioner fremhæver, at Statens It skal sikre hurtig og agil integration til private leverandørers it-løsninger, hvis potentialet inden for analytics skal udnyttes.

Statens It arbejder i retningen af en såkaldt *hybrid leverancemodel*, hvor dele af it-driftsopgaven vil blive varetaget i Statens It's eget datacenter eller egen cloud, mens andre dele af opgaven vil blive varetaget hos én eller flere cloud-leverandører.

På længere sigt er det forventningen, at Statens It kommer til at anvende og styre flere forskellige cloud-udbydere til it-drift. Dels for at sikre konkurrence og for at undgå afhængighed af én cloud-udbyder, og dels et hensyn til at kunne understøtte kundernes differentierede behov og krav.

Flere institutioner samler og tilgår deres og andres data hos Danmarks Statistik



Flere institutioner beriger egne data med Danmarks Statistiks registersamlinger via ministerieordningen

Danmarks Statistik (DST) er den centrale myndighed for dansk statistik. De indsamler, bearbejder og offentliggøre statistiske oplysninger om det danske samfund. DST har omfattende registersamlinger, der delvist består af data produceret af offentlige myndigheder, og delvist data produceret af DST. Som udgangspunkt indhentes så mange administrative registre som muligt hos fx Erhvervsstyrelsen og Skat, men i en række tilfælde er DST afhængig af, at virksomheder og borgere indberetter data.

Institutioner kan således via DST få adgang til en række registerdata på tværs af fagområder, og til det formål har DST oprettet en lignende version af forskerordningen, som hedder ministerieordningen. Her bliver data stillet til rådighed på en mere fleksibel måde, der imødekommer ministerområders og institutioners behov for ad hoc analyser med korte tidsfrister. Der er tale om ikke-personhenførbare mikrodata, dvs. data på individ eller virksomhedsniveau, hvor personnummer, virksomhedsnummer og andre identifikationsnøgler er erstattet med et tilfældigt løbenummer. Output fra analyserne kan bruges offentligt, men mikrodata må ikke forlade DST's servere.

Enkelte institutioner beskriver, at ministerieordningens aftalegrundlag kan være problematisk...

...idet hele institutionens adgang til data bliver lukket i en længere periode, hvis én enkelt medarbejder har begået en fejl, der er i strid med DST's fortrolighedsregler. Dette kan være en udfordring på særligt evidensstunge ministerområder, hvor statistik er helt centralt for kerneopgaven. For at minimere konsekvenserne ved lukket dataadgang, indgår ministerområder ministerieordninger med DST for hver enkelt institution i stedet for én samlet ministerieordning. Dette medfører betydelige meromkostninger, og giver incitament til at udvikle og drifte egne datacentre udenom DST.

Institutioner fremhæver at databehandling er en ressourcetung opgave, men at strukturering og rensning af data giver værdifuld indsigt i potentialer inden for analytics



Databehandling består typisk af fire delprocesser, hvor rådata klargøres endeligt til brug for analyser

Databehandling dækker over databearbejdelsen fra rådata til det endelige data, der skal bruges til analyser i organisationen. Der er fire delprocesser, der gør sig særligt gældende for databehandling: strukturering, rensning, berigelse og validering.

Strukturering omhandler struktureringen af rådata på en analytisk hensigtsmæssig måde. I strukturering indgår fx tildeling af id-numre til enhederne, oprettelse af tidsvariable og aggregering af data i meningsfulde kategorier, såsom inddeling af indkomstdata i indkomstkategorier.

Rensning indbefatter at tjekke data for manglende værdier, indflydelsesrige outliers, rette tekstlabels, så data er konsistent og kvalitetssikret.

Berigelse bygger ovenpå de to foregående processer. Det vedrør at hæve blikket fra de enkelte datapunkter til at se på data som en helhed og overveje om data kan beriges med andre data eller om der kan foretages analyser på den eksisterende data, der kan danne nye relevante variable.

Validering omhandler at sikre sig, at de processer, der er blevet foretaget er konsistente på tværs af hele datasættet og at datasættet i sin helhed er meningsfuldt.

Flere institutioner mener, at databehandlingsopgaven giver en værdifuld indsigt i data, som er afgørende i det videre analysearbejde.

Flere institutioner nævner, at databehandling er en ressourcekrævende opgave, der skal prioriteres, idet det giver værdifuld indsigt i institutionens data og potentialer inden for analytics.

Landbrugsstyrelsen, Børne- og Socialministeriet og Styrelsen for Forskning og Uddannelse ser det som helt afgørende med indgående kendskab til institutionens data, for at kunne gennemføre stærke analyser og i øvrigt få idéer til mulige analyser. Dette kendskab opbygges særligt i struktureringen og rensningen af data, hvor medarbejdere arbejder i rådata.

Berigelse og validering af data giver et samlet overblik over data og dets validitet. Dette er mindre ressource- og tidskrævende end strukturering og rensning.

Enkelte institutioner bruger konsulenter/Danmarks Statistik til dele af databehandlingen, men det er uklart i hvilket omfang



Det er uklart, præcist i hvilket omfang institutionerne selv udfører databehandlingen, men visse dele af disse delprocesser bliver hos nogle institutioner foretaget af konsulenter/Danmarks Statistik

Store dele af databehandlingsopgaverne er gentagende processer, hvor det samme rådata skal gennemgå samme behandling og formatering som tidligere rådata. I tilfælde hvor der bliver tilføjet nye data, skal der opsættes processer for behandlingen af den nye type rådata. Visse dele af disse processer bliver hos nogle institutioner foretaget af konsulenter eller Danmarks Statistik (DST).

Hos Slots- og Kulturstyrelsen, Styrelsen for Forskning og Uddannelse og Børne- og Socialministeriet foretages dele af databehandlingen i samarbejde med DST, hvor data også efterfølgende bliver opbevaret og kan tilgås af institutionerne selv og andre brugere af ministerordningen og forskerservice.

Hos flere andre institutioner bliver konsulenter brugt til at foretage databehandlingen af data, på trods af at visse dele af databehandlingen er præget af repetitive opgaver. Styrelsen for Arbejdsmarked og Rekruttering har eksempelvis et Datateam af konsulenter tilknyttet til at foretage store dele af deres databehandling.

Enkelte institutioner anvender Danmarks Statistik til at opgøre og formidle statistik på deres fagområde



Danmarks Statistik varetager opgørelser og formidling af statistik på vegne af enkelte institutioner

Det kunne eksempelvis være Slots- og Kulturstyrelsen, hvor indsamling, bearbejdning og produktion af statistiske data varetages af Danmarks Statistik. Det vedrører fx biblioteksdata fra folk- og forskningsbiblioteker, museernes besøgstal mv. Formidlingen af data foregår på Danmarks Statistiks hjemmeside.

Alternativt kan institutioner formidle statistik via deres egen hjemmeside ved at købe en personlig version af Statistikbanken, der kaldes StatHost løsning. Herved kan institutioner præsentere egne data i en velkendt opsætning med visning af data både i tabeller, figurer og kort samt mulighed for at eksportere data til excel-formater m.fl.

Erhvervsstyrelsen har eksempelvis i samarbejde med Danmarks Statistik udviklet en regional statistikbank, hvor der samlet er en række indikatorer, som måler de danske regioners præstationer og rammevilkår inden for uddannelse, iværksætteri, innovation, forskning og udvikling.

Flere institutioner har udfordringer med organisationens datagovernance med den konsekvens at potentialet af data ikke udnyttes optimalt



Forvaltning og styring af data går som regel under betegnelsen datagovernance, der vedrører regler og procedurer for datastyring på tværs af institutionens databaser og andre datakilder

Flere institutioner beretter, at de ikke har en fælles datagovernance, der går på tværs af organisationen og dets enkelte kontorer. Det medfører et fravær af ensartet, konsistente datastandarder og -definitioner med den konsekvens at kvaliteten af data forringes, og det analytiske potentiale i data ikke udnyttes til at finde nye indsigter mv.

Den manglende datagovernance skyldes blandt andet en historisk betinget organisatorisk indretning i siloer, hvor afdelinger og kontorer har opbygget deres egne regler og procedurer for håndtering af de specifikke data de forvalter.

Det kunne eksempelvis være Trafik-, Bygge- og Boligstyrelsen, der modtager mange eksterne data vedrørende jernbane, byggeri, luftfart, erhvervstransport mv. Med henblik på at udnytte det analytiske potentiale har de igangsat et arbejde, der skal sikre en fælles datagovernance. Dette gælder også for Banedanmark.

Derudover har særligt institutioner, der løbende er sammensat på baggrund ressortændringer, udfordringer med datagovernance. Miljøstyrelsen og Nævnenes Hus forklarer eksempelvis, at de arbejder på at integrere de forskellige dataarkitekturer (databaser, fagsystemer mv.) og håndterer de forskellige data, der er et produkt af tidligere institutioners datagovernance.

Tænkt caseeksempel

Ensartet anonymisering af data på tværs af en organisation som følge af persondataforordningen



Når GDPR træder i kraft er det ikke tilladt at arbejde med personhenførbare data. I et hypotetisk eksempel kan dette betyde, at datasæts skal anonymiseres, hvis id-numrene for enhederne er angivet ved CPR-numre. En måde at anonymisere data er at tildele enhederne løbenumre.

Problemet opstår hvis denne proces ikke bliver koordineret på tværs af organisationen. Hvis økonomiafdelingen tildeler løbenumre til enhederne i deres datasæt over lønudbetaling, mens HR-afdelingen tildeler andre numre til datasættet for trivsel. Uden ens id-numre kan datasættene ikke længere samkøres, da det er umuligt, at sammenkoble data i de to datasæt til de samme enheder.

Problemstillingen virker indlysende, men i praksis opstår denne type lavpraktiske problemer konstant, hvilket nødvendiggør en tværgående datagovernance for hele organisationen.

Flere institutioner vil løse problemer med datagovernance ved anvendelse af "data stewards" og robotautomatisering



Flere institutioner forklarer, at det er svært at implementere datagovernance, men at det kan lykkes vha. "data stewards"

Ifølge flere institutioner viser erfaringen, at indførelsen af en fælles datagovernance kræver en særlig implementeringsindsats, der dels bistår og overbeviser kontorer og medarbejdere om behovet og fordelene ved ændringerne, og dels, løbende inddrager kontorer og medarbejdere med henblik på at justere regler og procedurer ud fra praktiske erfaringer, hensigtsmæssighed mv. Til at løse denne opgave fremhæves "data stewards" som et virkemiddel, der kan skabe en smidig og effektiv implementering af datagovernance på tværs af organisationen.

Flere institutioner fremhæver også at robotautomatisering kan sikre ensartet dataregistrering fri for menneskelige fejl

"Robot proces automatisering" (RPA) er automatisering af ensartet og gentagende processer, fx hvis en medarbejder gentagende gange skal indtaste scores fra tilsynsrapporter i en database. Denne proces er udsat for menneskelige fejl i form af fejlindtastninger og afvigelser i praksis, hvis flere medarbejdere foretager samme proces. Derudover er det ressource- og tidskrævende arbejde. Denne type opgave kan varetages af en "robot", der kan programmeres til at foretage processen fejlfrit, ensartet og døgnet rundt. På denne måde kan RPA anvendes til at sikre ensartet data på tværs af hele organisationen.

Caseeksempel

Fødevarerstyrelsen overvejer "data stewards"



Miljø- og Fødevarerministeriet
Fødevarerstyrelsen

Fødevarerstyrelsen overvejer at indføre en ordning, hvor alle relevante kontorer evt. skal udnævne en data steward, der skal sikre, at styrelsens standarder for datagovernance bliver fulgt på tværs af organisationen. Udover at sikre compliance med databeskyttelsesregler, vil data stewards også kunne sikre bedre kendskab til data og formidle data videre til fx analytiske formål. Det er i overvejelserne, men altså ikke udmøntet.

Caseeksempel

Flere institutioner vil forbedre datagovernance med RPA



Sikkerhedsstyrelsen, Landbrugsstyrelsen og Finanstilsynet har tilkendegivet at, de afsøger mulighederne for at implementere RPA på en række manuelle, repetitive og driftstunge opgaver. RPA påpeges af alle, som et middel til at kunne højne datakvaliteten, ved at standardisere ensartet processer. Derudover vil RPA frigive ressourcer til kerneopgaven ved at automatisere ensformige og trættende opgaver.



På grund af juridiske hensyn etablerer flere institutioner adskilte databaser til henholdsvis administrative og statistiske formål



Der etableres adskilte databaser for at sikre at data indsamlet til et statistisk formål ikke bruges til administrative analyser

Når myndigheder indsamler data fra borgere og virksomheder sker det i henhold til særlige administrative eller statistiske formål. En række data er formålsbestemt til alene at måtte benyttes til statistiske analyser, og det kræver derfor en udtrykkelig hjemmel, hvis disse data skal indgå i administrative analyser.

Dette er eksempelvis tilfældet på sundhedsområdet, hvor Sundhedsdatastyrelsen har to datavarehuse til henholdsvis statistiske formål (fx sundhedsregistre) og administrative formål (fx tilsynsrapporter, ligsynsrapporter).

I enkelte tilfælde opstår der også den modsatte situation, hvor data der er indsamlet med et administrativt formål ikke må anvendes til et statistisk formål.

Samkøring af data i den administrative database er som udgangspunkt ikke tilladt, men skal foregå i den statistiske database

Oplysninger i administrative databaser indeholder som regel data på individniveau med CPR-nøgler, der gør det muligt at tilgå forskellige data om samme person eller virksomhed på tværs af fagsystemer. Det er som udgangspunkt ikke tilladt at samkøre disse data, men det kunne fx være relevant i forhold til afsløring om ulovlige bopælsforhold eller socialt bedrageri. Det er omvendt tilladt at samkøre data i den statistiske database, men her er data anonymiseret i præsentationen af resultaterne.

STAR nævner, at fordelene ved samkøring af data på beskæftigelsesområdet i et statistisk datavarehus bl.a. er, at det muliggør avancerede statistiske registeranalyser, der kan give mange nye indsigter. Deres statistiske database er derfor helt afgørende for at finde mønstre på tværs af data.

Enkelte institutioner har juridiske udfordringer med at dele og modtage relevante data fra andre myndigheder og virksomheder



Institutioner er påpasselige med at dele data på tværs på grund af uklarheder og usikkerhed om juraen på området

Ministerier og institutioner står over for øgede krav om evidensbaseret politikformulering og datadrevet opgaveløsning på fx tilsynsområdet, og undersøger derfor mulighederne for at få adgang til relevant data fra andre myndigheder, som kan indgå i deres analyser. Enkelte institutioner har adgang til flere forskellige data på tværs af fagområder, mens andre ikke har adgang.

Det har eksempelvis ikke været muligt for Arbejdstilsynet at sikre en lovhjemmel til udlevering af virksomheders lønoplysninger fra SKAT, med henblik på at undersøge, om løn er en indikator for dårligt arbejdsmiljø. Anti Doping Danmark forklarer også, at de oplever hindringer i adgangen til relevant data fra andre offentlige myndigheder.

Det fremhæves af enkelte institutioner, at den manglende deling af data på tværs af institutioner, er et udtryk for tvivl om og forskellige fortolkninger af lovgivningen, i forhold til under hvilke omstændigheder data må deles, og hvilke data der må videregives. På denne baggrund er praksis blandt hovedparten af institutionerne at afvise anmodninger, hvis der er tvivl om lovgrundlaget.

I lyset af denne problemstilling har Erhvervsstyrelsen fremsat et lovforslag, der giver dem adgang til andre myndigheders data, hvis det er nødvendigt for deres opgavevaretagelse.

Caseeksempel

Indsamling og samkøring af data fra andre myndigheder



Erhvervsministeriet har fået gennemført et lovforslag (L149) med ikrafttrædelse d. 1. juli 2018 om Erhvervsstyrelsens behandling af data.

Loven giver Erhvervsstyrelsen lovhjemmel til at:
"indsamle og behandle, herunder foretage samkøring af, oplysninger fra andre offentlige myndigheder og fra offentlige tilgængelige kilder, når det er nødvendigt af hensyn til udførelsen af styrelsens opgaver"

Dette betyder, at Erhvervsstyrelsen har mulighed for at indsamle og samkøre data fra andre myndigheder, hvis det er nødvendigt for udførelsen af bl.a. deres kontrol og tilsynsindsatser. I tilfælde af at Erhvervsstyrelsen finder oplysninger, der er nødvendige for andre myndigheders kontrol og tilsynsindsats har de også lovhjemmel til at videregive disse oplysninger til andre myndigheder.

Flere institutioner er usikre på juraen for styringen af data, og er i gang med at afdække databeskyttelsesforordningens konsekvenser i forhold til datasamkøring



Flere institutioner er i tvivl om, hvilke data der må samkøres, og hvilke konkrete krav og betingelser databeskyttelsesforordningen stiller til behandlingen af personoplysninger

Der hersker usikkerhed blandt flere institutioner om den juridiske ramme i lyset af databeskyttelsesforordningen, men også i lyset af den tiltagende samfundsdebat og bevågenhed om håndteringen af personoplysninger.

Styrelsen for It og Læring har fx nedsat en arbejdsgruppe af jurister med henblik på at udrede databeskyttelsesforordningens konsekvenser i forhold til samkøring af data mv. Dette kompliceres yderligere af den aktuelle samfundspolitiske debat om trivselsmålinger af elever og generelt etikken og it-sikkerheden i dataanvendelse. Styrelsen for Patientsikkerhed oplever også en stor politisk bevågenhed i beskyttelsen af sundhedsdata, som er med til at flytte grænser for, hvilke data der kan eller bør samkøres.

Enkelte styrelser påpeger, at Finansministeriets tværgående rolle er et oplagt afsæt til at bidrage med juridisk rådgivning om samkøring af data.

Perspektivering

Persondataforordningen og samkøring af persondata

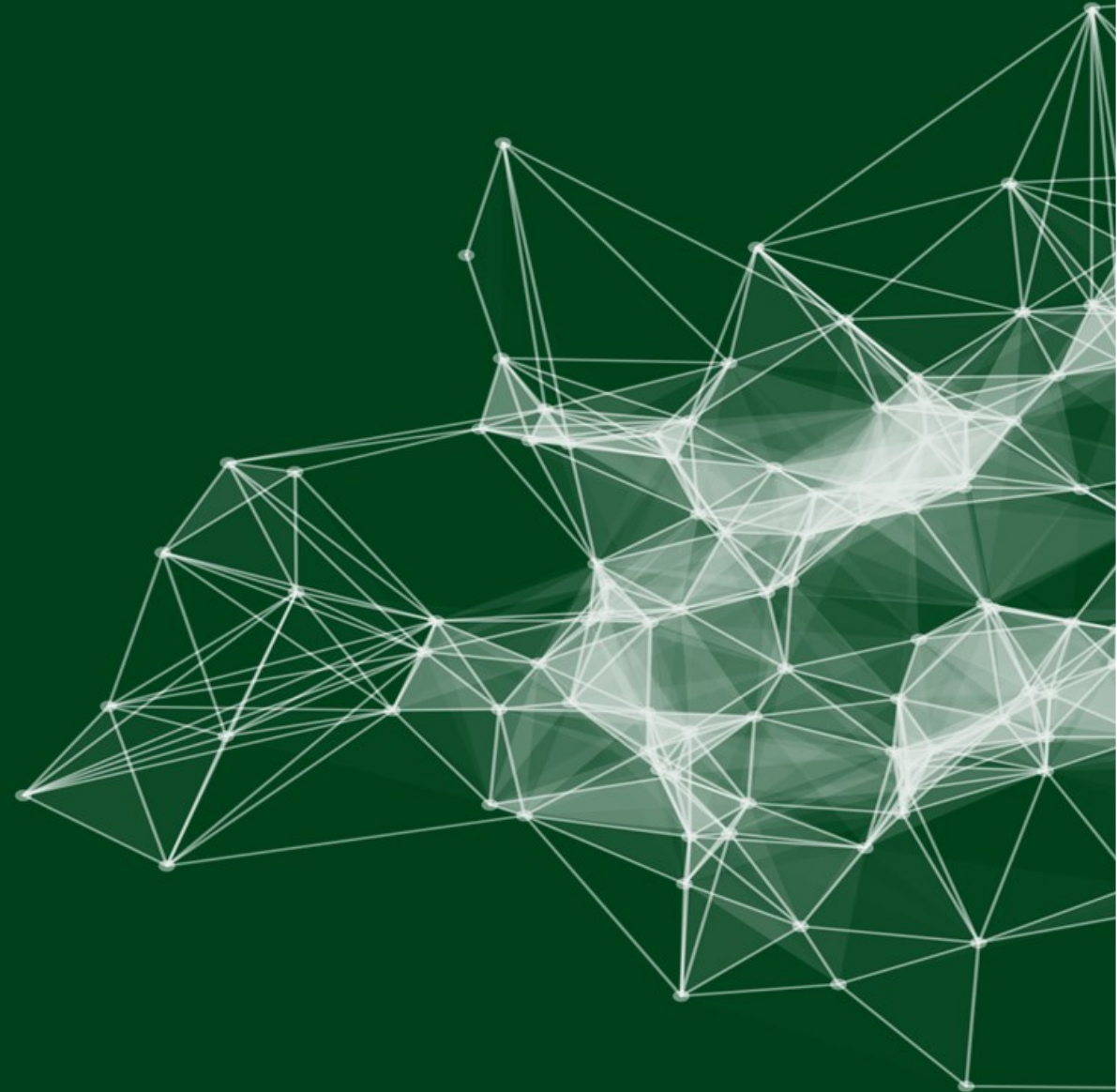


I forbindelse med indførelsen af EU's persondataforordning har Justitsministeriet fremlagt deres forslag til supplerende bestemmelser. Herunder har bestemmelserne for samkøring af persondata været meget omdiskuteret.

I det oprindelige forslag skulle ressortministeren kunne give tilladelse til at persondata, der er indsamlet med forskellige formål, måtte samkøres. Efter flere diskussioner i Folketinget og i offentligheden er der indgået en aftale imellem regeringen, Dansk Folkeparti og Socialdemokratiet om, at retsudvalget og et andet relevant fagudvalg kan give institutioner lov til at samkøre persondata til nye formål. Derudover skal datatilsynet og relevante eksterne organisationer høres.

Sektion 4

Ledelse og kompetencer inden for analytics



Introduktion: Vi beskriver i denne sektion ledelse og kompetencer inden for analytics



Hvilken rolle spiller ledelse, hvilke særlige kendetegn er der inden for projektstyring, og hvilke kompetencer er der behov for inden for analytics?

Digitalisering er i dag en gennemgribende del af institutioners daglige praksis, der stiller nye krav til digital ledelse på tværs af afdelinger og på tværs af ledelseshierarkier.

Alle faglige ledere har i dag ansvar for, at digitalisering og bedre udnyttelse af data bidrager til at kvalificere og effektivisere kerneopgaven, men hvad skal der til for at lykkes med at indfri institutionens potentiale inden for analytics?

I kapitel 7 stiller vi skarpt på en række ledelsesmæssige perspektiver inden for analytics.

I kapitel 8 gennemgår vi projektorganisering og kompetencer inden for analytics.

Kapitel 7

Hvilken rolle spiller statens ledelse?



Digital ledelse. Vi beskriver en række ledelsesmæssige perspektiver inden for analytics.

Kapitel 8

Hvordan udvikles analytics?



Projektorganisering og kompetencer. Vi beskriver organisering og styring af analytics projekter, samt hvilke kompetencer der er behov for.

Baggrund: Ledelse af digital organisering fra kælder til kvist og på tværs af organisationen



Nye krav til digital ledelse

Forståelse for "det digitale" og en udvidet forståelse for mulighederne ved analytics inden for ens forvaltningsområde er i dag et stigende krav til lederen i en politisk styret organisation.

Digitalisering og brug af data er ikke længere alene it-chefens domæne, men i lige så høj grad de øvrige faglige ledere og i særdeleshed organisationens topledelse. Dette skifte sker omkring 00'erne, hvor digitalisering bliver en del af flere organisationers strategiske arbejde i kølvandet af den Digitale Taskforce, jf. s. 54.

Men hvad er de nye krav? Digital ledelse går ud på, at forstå hvordan teknologi og organisation spiller sammen og forandrer hinanden gensidigt. Nogle af de nye krav er:

- Aktiv tilgang til ny teknologi: Ledelsesopgaven handler om at se organisatoriske muligheder i digitalisering og analytics.
- Tidlig involvering i projekter: Ledelsen skal øve indflydelse på grundlæggende valg af data, teknik mv., der betinger den endelige løsning og beslutninger.
- Blik for forandringsledelse: Digitalisering medfører nye arbejdsgange, roller og kompetencekrav, som igen betyder at ledelsen skal kunne lede og udvikle den organisatoriske kultur.



Nye krav til tværorganisatorisk organisering

Der har i mange årtier været en tendens til at organisere sig efter fagområder i hierarkisk adskilte siloer. Digitale forandringer og projekter kan dog ofte ikke placeres i én silo, fordi de som regel er tæt forbundet med andre fag- eller forvaltningsområder.

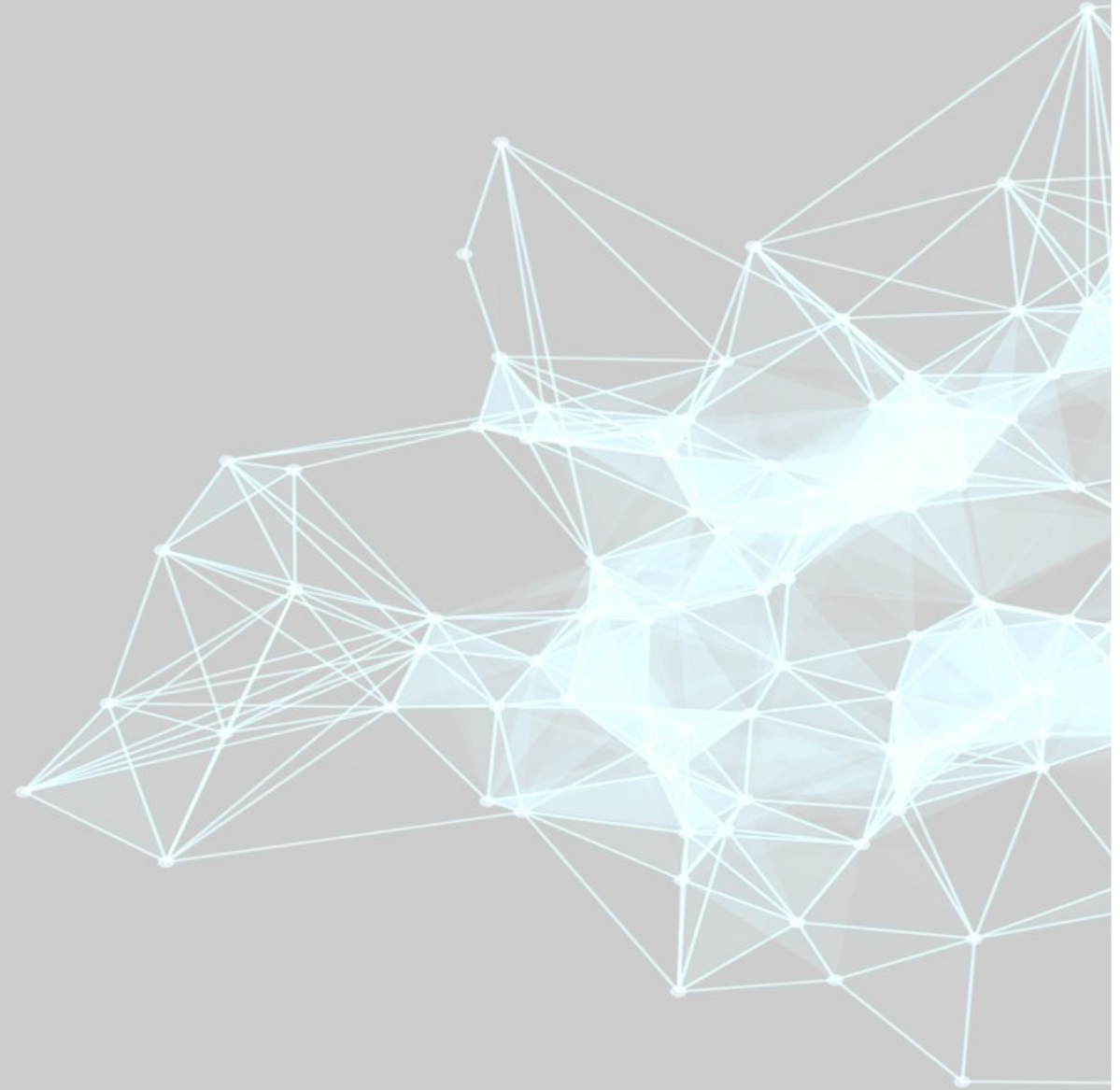
Dette stiller nye krav til tværorganisatorisk samarbejde, men netop dét er en udfordring for organisationen, fordi koordinering på tværs er svær at praktisere. I det lys bliver det en ledelsesopgave at sikre et frugtbart og effektivt samarbejde på tværs af institutionens kontorer og fagligheder.

Den stadige tilstrømning af nye teknologier gør, at den digitale ledelsesopgave ikke kan indhegnes i en episodisk tilgang, der snart går over eller afsluttes, men det giver i højere grad mening at se det digitale som en kontinuerlig strøm af mulighedsrum.



Kapitel 7

Digital ledelse



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Digital ledelse



Hvilken rolle spiller ledelse?

Vi beskriver en række ledelsesmæssige perspektiver inden for analytics, og har følgende konklusioner:

1. De fleste institutioner har strategisk fokus på dele af analytics, men der er stor forskel på, hvor langt institutionerne er kommet med at implementere
2. Flere institutioner arbejder på at kvalitetssikre data, men der er forskel på, hvordan ledelsen løser opgaven
3. Hvor kommer idéerne til analytics løsninger fra?
4. Har staten et ansvar som "first mover" med henblik på at modne teknologien?
5. Enkelte institutioner har fokus på en række etiske dilemmaer i anvendelsen af data og kunstig intelligens i den offentlige sektor
6. Flere institutioner beskriver en brændende platform og opfatter analytics som løsningen
7. Flere institutioner fremhæver, at topledelsens risikovillighed i opstartsfasen er helt afgørende for succesfuld anvendelse af analytics
8. Enkelte institutioner beskriver, at ledelsen skal være i stand til at deltage i en faglig dialog med dataeksperter om indholdet og retningen for analytics projekter
9. Flere institutioner beskriver, at ledelsen skal kunne begrunde, hvorfor enkelte udpeges frem for andre, samt forsvare modellens risiko for fejl
10. Enkelte institutioner understreger, at god forandringsledelse er afgørende for at medarbejdere accepterer og anvender analytics løsninger i hverdagen
11. Enkelte institutioner bruger udflytning som vindue til at gennemføre større digitale udviklingsprojekter

De fleste institutioner har strategisk fokus på dele af analytics, men der er stor forskel på hvor langt institutionerne er i deres visioner for de næste 5-10 år



Hvilke udfordringer står vi overfor? Hvordan kan analytics bidrage? Hvad er vores plan?

Flere og flere topledere stiller spørgsmål som disse i forhold til analytics, og der er et stigende strategisk fokus på, hvordan analytics kan bidrage til at løse nogle af institutionens udfordringer.

På grund af digitaliseringens accelererende udvikling er institutioner under konstant forandringspres, og topledelsens ansvar er at styre forretningsudviklingen. Lederen skal have blik for såvel potentialer og løsninger, som risici og forventelige udfordringer. Lederen skal være i stand til at se, hvordan "det digitale" kan understøtte de aktuelle kerneopgaver, hvordan det kan understøtte kerneopgaver, som institutionen ikke løser i dag, samt lede i organisatoriske forandringer, der følger med. Dette kræver en visionær ledelse, der løbende trykprøver status quo, og er på forkant med udviklingen, idet fx indsamling af flere eller nye data er tidskrævende. Samtidig skal ledelsen have et teknisk kendskab for at kunne forstå, hvilke udviklingsprojekter der er mere udfordrende end andre.

Størstedelen af institutionerne har eller er ved at formulere et strategisk fokus på dele af analytics. Der er dog stor forskel på, hvor langt institutionerne er. Rigsarkivet er eksempelvis ved at udarbejde en digitaliseringsstrategi med fokus på at digitalisere kulturarven og herved blandt andet opnå større indsigt i hvem, hvordan og hvad kulturobjekterne bruges til. Arbejdstilsynet, Miljøstyrelsen, Finanstilsynet m.fl. har digitaliseringsstrategier, der fokuserer på at transformere organisationen til at blive datadrevet, herunder bedre brug af data og machine learning i kontroløjemed.

Caseeksempel

Energinets strategi for 2018-2020: Energi over grænser



Energinet har i deres strategi, fokus på nye veje til digitalisering og brug af data, og identificerer en tendens inden for big data og kunstig intelligens:

"Digitalisering er en megatrend, der spreder sig over alle sektorer i hele verden. En del af denne megatrend er de hastigt forøgede teknologiske muligheder for at indsamle meget store mængder af data, som med stadigt mere intelligent software kan underkastes analyser med hidtil usete muligheder."

Energinet tror på smartere drift med digitale løsninger:

"Der er potentialer i øget brug af databehandling og digitale teknologier til drift og udvikling af Energinets infrastruktur. Det gælder fx omstilling til digitale stationer samt droner og avanceret billedanalyse til at identificere vedligeholdelsesbehov."

Flere institutioner arbejder på at kvalitetssikre data, men der er forskel på hvordan ledelsen løser opgaven



En række institutioner mangler overblik over data og datakvaliteten er ofte tvivlsom

Flere institutioner, som eksempelvis Banedanmark, Trafik-, Bygge- og Boligstyrelsen, Sikkerhedsstyrelsen, Miljøstyrelsen m.fl. har igangsat et datastrategisk arbejde, der undersøger institutionens datamanagement, herunder indsamling og styring af data, med henblik på at øge datakvaliteten.

Der er dog forskel på, hvordan institutionerne kvalitetssikrer data. Enkelte institutioner gennemgår hele institutionens datamanagement og ser på data i dets helhed. Andre institutioner kaster sig i højere grad over enkelte dataområder, hvor datakvaliteten er tvivlsom, og arbejder med en løbende forbedring af data.

Miljøstyrelsen bruger eksempelvis machine learning til at kvalitetssikre vandmiljøprøver, jf. s. 50.

Men hvad er vejen til god datakvalitet, og hvornår er kvaliteten tiltrækkelig høj til at kunne bruges? Dette spørgsmål stiller flere institutioner.

Hvad er vejen til god datakvalitet?

Helt grundlæggende er analysen ikke stærkere end de data, den bygger på – m.a.o. er data fundamentet under et hus. Der er dog forskel på, hvilken tilgang ledelsen vælger for at sikre en højere datakvalitet.

En gruppe ledere kan karakteriseres som risikoaverse, og vælger derfor en samlet revision af institutionens datamanagement ud fra det hensyn at reducere sandsynligheden for fejl. Der stilles meget høje krav til pålideligheden af data, og hvornår data kan anvendes i opgaveløsningen. Dette kendetegner den typiske tilgang i offentlig forvaltning, hvor embedsmænd skal leve op til sandhedspigten, jf. Kodeks VII.

En anden gruppe ledere er mere risikovillige, og vælger derfor at målrette revisionen af institutionens datamanagement ud fra det hensyn at øge sandsynligheden for, at data kan berige opgaveløsningen. Der stilles snarere minimumskrav til pålideligheden af data, og man anvender data i opgaveløsningen, så længe fejlkilderne kan forsvares og data forbedrer analysen og beslutningsgrundlaget. Dette kræver en stærkere metodisk faglighed, blandt andet for at kunne begrunde, hvorfor ovenstående tilgang kan siges at sikre ”den oplevede sandhed”, mens den mere pragmatiske tilgang reelt forbedrer sandheden ud fra den betragtning, at god datakvalitet sikres ved aktiv og iterativ brug af data i den daglige opgavevaretagelse.

Hvor kommer idéerne til analytics løsninger fra?



Idéerne kommer nogle gange fra topledelsen, andre gange fra udviklingsenheder og i visse tilfælde fra enkelte medarbejdere

Der er forskel på, hvor idéerne kommer fra i institutionerne.

Det kan komme fra topledelsen, hvor direktionen aktivt er i dialog med forskellige aktører som fx konsulenthuse og universiteter, men idéerne kan også opstå som et pres fra omgivelserne på ens fag- eller forvaltningsområde. Det oplever eksempelvis Finanstilsynet, hvor den finansielle sektor i høj grad driver udviklingen inden for analytics.

Idéerne kommer også fra særligt nedsatte teams inden for digitalisering og analytics. De arbejder dedikeret med at udvikle og teste idéer for hurtigt at kunne afprøve løsningens potentiale. Dette vender vi tilbage til i næste kapitel.

Idéerne kan også drives af medarbejdere, der har en særlig forretningsforståelse og erfaring med de daglige aktiviteter og processer i institutionen. Det gør sig eksempelvis gældende i STAR og Fødevarestyrelsen. Derudover oplever fx Beskæftigelsesministeriet en øget efterspørgsel blandt de yngre medarbejdere med økonomibaggrund på kurser i machine learning.



Har staten et ansvar som "first mover" med henblik på at modne teknologien?



Flere institutioner ser potentialer ved machine learning og matematisk optimering, men afventer implementering til løsningerne er mere modne

Nye teknologier og analysemetoder udvikles og testes som regel først i den private sektor før den offentlige sektor afprøver potentialer og anvendelighed i en kommunal, regional eller statslig kontekst. Det kan der være flere fordele ved både økonomisk og kvalitetsmæssigt.

Der er dog samtidig en erkendelse blandt flere institutioner, at anvendelsen af machine learning og matematisk optimering i staten kræver komplekse tilpasninger og ledelsesmæssige indsatser.

I forhold til matematisk optimering afventer flere institutioner, at løsningen modnes yderligere på grund af kompleksitet og udfordringer med forandringsledelse til en mere effektiv vagt- eller opgaveallokering, jf. forrige slide.

I forhold til tekstanalyse til effektivisering af tekstunge sagsbehandlingsopgaver, afventer flere institutioner udviklingen i markedet af flere dansksprogede løsninger. Ifølge Finanstilsynet og Miljøstyrelsen vil det være for omkostningstungt selv at træne og udvikle danske sprogfærdigheder inden for kunstig intelligens. Andre stiller dog spørgsmålstegn ved, om der er kritisk masse i markedet til at fremme udviklingen af danske tekstanalyse løsninger. Det efterlader spørgsmålet: Er der behov for "first movers" i staten? Er der behov for fælles offentlige investeringer?

Caseeksempel

Citater



Miljø- og Fødevareministeriet
Miljøstyrelsen

"Det kan ikke betale sig at anvende tekstmining, før der er opbygget tiltrækkelige dansk kundskaber i tekst mining løsninger på markedet"



Uddannelses- og
Forskningsministeriet

—
Styrelsen for Forskning og Innovation

"Der er stadig et stort potentiale i mere klassiske økonometriske analyser. Vi ser ikke et akut behov for at skulle anvende fx machine learning for at få nye indsigter"



LÆGEMIDDELSTYRELSEN

"Vi har implementeret Watson med betydelig succes, men det er en økonomisk tung byrde at bære som styrelse alene"

Enkelte institutioner har fokus på en række etiske dilemmaer i anvendelsen af data og kunstig intelligens i den offentlige sektor



Enkelte institutioner berører dataetik som afgørende for at kunne sikre tillidsfuld anvendelse af machine learning og øvrig kunstig intelligens

Der hersker generelt en samfundsmæssig nervøsitet om hvad offentlige og private institutioner dokumenterer, hvilke digitale spor borgere og forbrugere efterlader, hvad sporene anvendes til og hvilke konsekvenser det vil have. Debatten nåede helt nye højder efter Facebook skandalen med Cambridge Analytica blev afsløret i marts 2018.

De etiske dilemmaer opstår ved, at digitaliseringen og analytics på den ene side giver muligheder for nye indsigter og nye samfundsgavnige indretninger, men på den anden side udfordrer grundlæggende frihedsrettigheder, fx privatlivets fred og retten til at blive glemt, og hvem der har ansvaret for at forklare algoritmernes resultater og anbefalinger.

Enkelte institutioner antyder, at der er behov for at få generelle og fagpolitiske retningslinjer for, hvordan dataetiske dilemmaer og hensyn kan vægtes og besvares. Er det fx rimeligt at identificere studerendes risiko for frafald? Er det rimeligt at måle elevens trivsel? Er det rimeligt at bruge mobildata til at identificere antal besøgende i slotsparker? Er det rimeligt at lagre individuelle sundhedsdata og finde interessante mønstre?

Et andet centralt spørgsmål man kan stille er omvendt, om vi med de teknologier og datamængder vi har til rådighed i dag, kan forsvare at vi ikke anvender data?

Perspektivering

Disruptionsrådet har fokus på dataetik



"Digitalisering og dataanvendelse rummer et enormt potentiale for vækst og innovation og giver mulighed for at udvikle nye tjenester og produkter, der ikke tidligere er set. Samtidig stiller brugen af ny teknologi og data os over for en række etiske dilemmaer i forhold til den måde vi indretter vores forretning og samfund på, fx i forhold til privatlivets fred, gennemsigtighed og ansvar ved brug af algoritmer. Hvis potentialet ved digitalisering og ny teknologi skal indfries, forudsætter det, at der i samfundet er tillid til, at de nye teknologier og data anvendes på en ansvarlig måde. Dataetik er ikke alene en udfordring for det private erhvervsliv, men også for den offentlige sektor, f.eks. i forbindelse med trivselsundersøgelser i folkeskolen. Der er dataetisk en forskel på at være borger og forbruger: En forbruger kan vælge et andet produkt og dermed andre rammer for datahåndtering, mens borgeren indgår i forpligtende fællesskaber, som man ikke kan fravælge (f.eks. skole). (1)"

Flere institutioner beskriver en brændende platform og opfatter analytics som løsningen



Digitalisering har medført en række muligheder, men også udfordringer, som analytics kan bidrage til at håndtere.

Digitalisering har som udgangspunkt givet institutionerne bedre muligheder for at effektivisere sagsgange, administrere økonomiske rammer, kommunikere med omverden mv., jf. s. 54.

Der følger dog også en række udfordringer.

Erhvervsstyrelsen forklarer eksempelvis, at det er en udfordring at fastholde et effektivt tilsyn, når danske virksomheder kan oprette sig selv på fem minutter og dermed øger den daglige sagsmængde. Finanstilsynet har også udfordringer med at løfte deres tilsynsopgave på grund af omfanget (stor volumen) og hastigheden (høj velocity) af nye typer finansielle transaktioner – to af v'erne fra definitionen af big data. Og Lægemiddelstyrelsen forventer, at myndigheder i løbet af ganske få år vil se sig konfronteret med data af en kompleksitet, der vil rejse en række spørgsmål, hvis besvarelse er afgørende for lægemiddelsikkerhed og –vækst, men også for tilliden til myndighedernes evne til at levere svar, jf. debatten om bioetik.

En del af løsningen på disse udfordringer er at skabe overblik over data, identificere mønstre og sikre en intelligent prioritering – og dette kan dele af analytics.

Caseeksempel

Citater



”Digitaliseringen af registreringen gør, at flere data bliver mere tilgængelige, hvilket har stor betydning for behovet, tilrettelæggelse og omfanget af tilsyn og kontrol på området.”



LÆGEMIDDELSTYRELSEN

”Udviklingen i datakompleksiteten løber utrolig hurtigt, og det er afgørende at lovgivere og myndigheder følger med”



”Vi har været belastet af store datamængder og flere af dem vi fører tilsyn med er dem, der rykker mest. Vi bliver derfor også nødt til at rykke og lave et kompetenceløft for at kunne følge med sektoren”

Flere institutioner fremhæver, at topledelsens risikovillighed i opstartsfasen er helt afgørende for succesfuld anvendelse af analytics



Projekter inden for analytics kræver et særligt fokus fra topledelsen og en risikovillighed i projektinitieringen til at prioritere udvikling over produktion

Erfaringen viser, at flere institutioner har udfordringer med at gennemføre succesfulde analytics projekter, der som regel er kendetegnet ved en høj grad af kompleksitet og involverer flere medarbejdere med forskellige fagligheder fra forskellige kontorer.

Flere institutioner fremhæver, at en af årsagerne hertil kan være manglende mod og risikovillighed fra topledelsen til i opstartsfasen at prioritere udvikling over hverdagens drift af kerneopgaven, herunder for en kortere periode at dedikere relevante medarbejdere væk fra driften til udvikling på fuld tid.

Der er dog delte meninger herom.

Nævnenes Hus forklarer, at de er villige til at prioritere deres udviklingsprojekter, herunder machine learning, over den daglige drift af kerneopgaven, idet gevinsten ved løsningen overstiger de midlertidige tab ved en mindre stabil drift.

Landbrugsstyrelsen forklarer omvendt, at analytics projekter ikke må gå ud over kerneopgaven.

Caseeksempel

Citater



"Projekterne om opbygning af vores datavarehus og udvikling af machine learning til at understøtte sagsbehandling prioriteres over produktion"



Miljø- og Fødevareministeriet
Landbrugsstyrelsen

"Projekter inden for analytics må ikke gå udover vores kerneopgave – at udbetale landbrugsstøtte til erhvervet"

Enkelte institutioner beskriver, at ledelsen skal være i stand til at deltage i en faglig dialog med dataeksperter om indholdet og retningen for analytics projekter



Ledere skal kunne drøfte grundlæggende metodiske og datamæssige valg for at kunne indgå i en faglig dialog om indholdet og retningen for analytics projekter

Kan vi stole på vores data? Måler vi det, vi tror vi måler? Hvilke konsekvenser har vores definition og valg af data på vores beslutninger? Kan jeg anvende dette analyseværktøj i denne kontekst?

Spørgsmål som disse skal institutioners ledere i højere grad være i stand til at drøfte, for at kunne rammesætte og navigere retningen for, hvad det specifikke analytics værktøj helt konkret skal løse, men også for at kunne fortolke resultaterne og forstå dets begrænsninger.

Ledere skal eksempelvis have blik for, hvordan trivsel og sygefravær er målt og illustreret i deres dashboard, når de træffer beslutninger. De skal kende til p-værdi og signifikansniveau, hvis de bruger økonometriske analyser, som grundlag for nye policy tiltag. De skal kunne forstå ROC kurven, når de diskuterer algoritmens træfsikkerhed. Og de skal forstå mekanismerne og kompleksiteten ved matematisk optimering, herunder hvornår man bør sætte begrænsninger på det matematisk optimale af hensyn til fx kultur og menneskelige værdier.

Datagrundlaget betinger de beslutninger, der bliver truffet på baggrund af værktøjerne, og det er derfor vigtigt, at ledelsen involverer sig tidligt i analytics projekter og er inde over grundlæggende definitioner og data (fra)valg. Beslutninger bliver ikke kun truffet på baggrund af data, men også implicit i udvælgelsen af data, jf. perspektiveringen til højre.

Perspektivering

Forskningsprojekt: Hvordan forandrer data den offentlige sektor?



IT-Universitetet undersøger i forskningsprojektet *Data as Relation: Governance in the Age of Big Data* hvordan digitalisering forandrer det offentliges arbejdsvilkår, og hvilke muligheder og faldgruber der er ved at anvende data som styringsværktøj.

Èt af de områder, der undersøges vedrører betydningen af valg af data. Ifølge Brit Ross Wintherik, der leder projektet, bør vi forstå brugen af big data og digitalisering i den offentlige sektor som en fundamental ændring af beslutningsprocesserne, ikke blot som en opgradering af gamle infrastrukturer.

Hun siger:

”Vi vil gerne udvide ideen om, hvornår en beslutning bliver taget. I det øjeblik man begynder at udvælge, hvilke data man vil indsamle om borgerne, er man faktisk allerede i gang med at træffe nogle politisk betingede beslutninger. Beslutningsprocesser er ikke kun noget, der foregår, efter man har indsamlet data” (1).



Flere institutioner beskriver, at ledelsen skal kunne begrunde, hvorfor machine learning udpeger enkelte frem for andre, og forsvarer modellens risiko for fejl



Det offentlige skal være i stand til at forklare, hvordan machine learning algoritmen fungerer, og hvorfor enkelte borgere eller virksomheder bliver udpeget frem for andre

Som beskrevet tidligere, jf. s. 24 og 28-32, er machine learning et værktøj til effektivt at give indsigt i sammenhænge i store mængder data, der er uoverskuelige for mennesket. Men hvem forklarer algoritmen, når algoritmen bliver for kompleks? Og hvilke særlige hensyn er der i den offentlige sektor?

Flere institutioner fremhæver, at udfordringen ved anvendelse af machine learning i den offentlige sektor er, at institutionen som helhed, og de enkelte ledere og medarbejdere specifikt, skal kunne begrunde, hvorfor algoritmen lige netop udpeger fx én elev frem for en anden i klassen eller vurderer den nylediges risiko for langtidsledighed.

Machine learning rejser derfor en vigtig debat om, i hvilket omfang det offentlige kan gøre brug af algoritmer og kunstig intelligens, og i hvilken grad det offentlige skal kunne forklare algoritmens resultater eller anbefalinger. Der ligger et dilemma i dette, idet komplicerede algoritmer kan være vanskelige om ikke umulige at forklare, men modsat til tider giver højere præcision. I hvilke tilfælde er præcisionen afgørende, og i hvilke bliver hensynet til forklaringen tungestvejende.

Machine learning introducerer en kendt usikkerhed, som ledere skal kunne forsvare over for borgere og virksomheder

Der følger med machine learning også en kendt usikkerhed, idet modellen sandsynliggør risikoen for fejl, og dermed synliggør, hvilke potentielle konsekvenser løsningen kan have for enkelte virksomheder eller borgere. En synlighed af fejl, der er målbar og mere eksplicit end menneskelige fejl ved manuelle arbejdsrutiner, en sagsbehandlers daglige beslutninger mv.

Det kunne eksempelvis være anvendelsen af machine learning til ejendomsvurderinger, hvor modellens evne til at "gætte" værdien kan eksplicit kvantificeres.

Anvendelsen af machine learning medfører således en accept af den kendte usikkerhed med negative konsekvenser for enkelte virksomheder eller borgere, og denne fejlmargen skal myndigheden og lederen være i stand til at forsvare.

Det kræver en grundlæggende teknisk og metodisk forståelse for modellens funktion og databehandling, men eventuelt også et beredskab, der har opgørelser over antallet af menneskelige fejl versus modellens fejlmargen. Samtidig kræver det en forståelse for, hvordan modellen løbende bliver trænet og over tid får en mindre og mindre fejlrate. Derudover kan man overveje at foretage yderligere valideringsforløb eller manuelle korrektioner, hvis man ikke kan acceptere modellen som det endelige beslutningsværktøj.

Enkelte institutioner understreger, at god forandringsledelse er afgørende for at medarbejdere accepterer og anvender analytics løsninger i hverdagen



Forandringer af medarbejderes arbejdsrutiner og kompetencekrav kræver en særlig ledelsesmæssig indsats med henblik på at sikre "buy in" og undgå modstand blandt medarbejdere

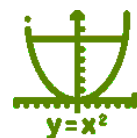
Ledelse af forandringer spiller en væsentlig rolle i omstillingen fra "det gamle" til "det nye" i enhver organisation, og stiger i takt med antallet af berørte medarbejdere og teknologiens kompleksitet. Forandringsledelse er svært, og ifølge undersøgelser kører op mod 70 pct. af større forandringer af sporet (1).

Der er tre centrale perspektiver på arbejdet med at skabe forandring (2):

- Meningsskabelse – ledere skal skabe legitimitet omkring nye løsninger ved at skabe individuel og fælles mening
- Oversættelsesprocesser – ledere skal oversætte ideer, koncepter, redskaber mv. med henblik på at viden om løsningerne rejser fra én kontekst til en anden
- Læring – ledere skal have fokus på varetagelsen af et læringsrum med henblik på at medarbejdere oplever, at de kan mestre de nye løsninger

Caseeksempel

Forandringsledelse i forhold til matematisk optimering



Som beskrevet tidligere, jf. s. 33 og 43, kan matematisk optimering anvendes til at finde en optimal fordeling af opgaver, herunder fx fordeling af kontrolsteder til tilsynsførende.

Den ledelsesmæssige udfordring er her at sikre legitimitet fra de tilsynsførende til at allokere opgaver/kontrolsteder ud fra det mest optimale hensyn for institutionen (lavest antal kørte kilometer), og ikke ud fra suboptimale hensyn for enkelte tilsynsførende (tættest placerede kontrolsteder). Dette skal lederen kunne formidle og oversætte til de tilsynsførende med øje for deres kontekst og konkrete arbejde. Samtidig skal ledelsen sikre læring af de tilsynsførende til at forstå modellens grundlæggende principper for allokering af kontrolsteder.

Hvis medarbejder ikke deler ledelsens synspunkter, viser erfaringen, at der opstår modstand. Ledere skal derfor overveje om der er behov for at indsætte begrænsninger i modellen, der tilgodeser de enkelte medarbejderes suboptimale rationaler.

Anm.: (1) Beer & Nohria 2000 i "Ledelse af forandring i politisk styrede organisationer" af Peter Hagedorn-Rasmussen i Ledelse i politisk styrede organisationer 2017.

(2) Afsnittet om forandringer af medarbejdere bygger på "Ledelse af forandring i politisk styrede organisationer" af Peter Hagedorn-Rasmussen i Ledelse i politisk styrede organisationer, 2017



Enkelte institutioner bruger udflytning som vindue til at gennemføre større digitale udviklingsprojekter



Ledelse af forandring handler også om at vælge de rette tidspunkter for større ændringer, og her ser enkelte institutioner udflytning som et mulighedsrum.

Institutioners organisatoriske og digitale opbygning er historisk betinget, og der er flere institutioner, der i dag ikke har vedligeholdt og udviklet deres datamiljø tilstrækkeligt til at kunne anvende analytics løsninger optimalt.

Det viser sig, at enkelte institutioner bruger udflytningen som vindue til at igangsætte større digitale udviklingsprojekter med henblik på at opdatere institutionens datamiljø til 2018.

Nogle af de fordele, der blive nævnt ved udflytning er fx:

- Ledelsesmæssig momentum til at gentænke institutionens ofte fragmenterede datamiljø på grund af diverse ressortomlægninger over tid
- Tilgang af mange nye medarbejdere gør det alt andet lige lettere at bryde gamle arbejdsrutiner og skabe en ny organisationskultur

Caseeksempel

Etablering af Nævnenes Hus



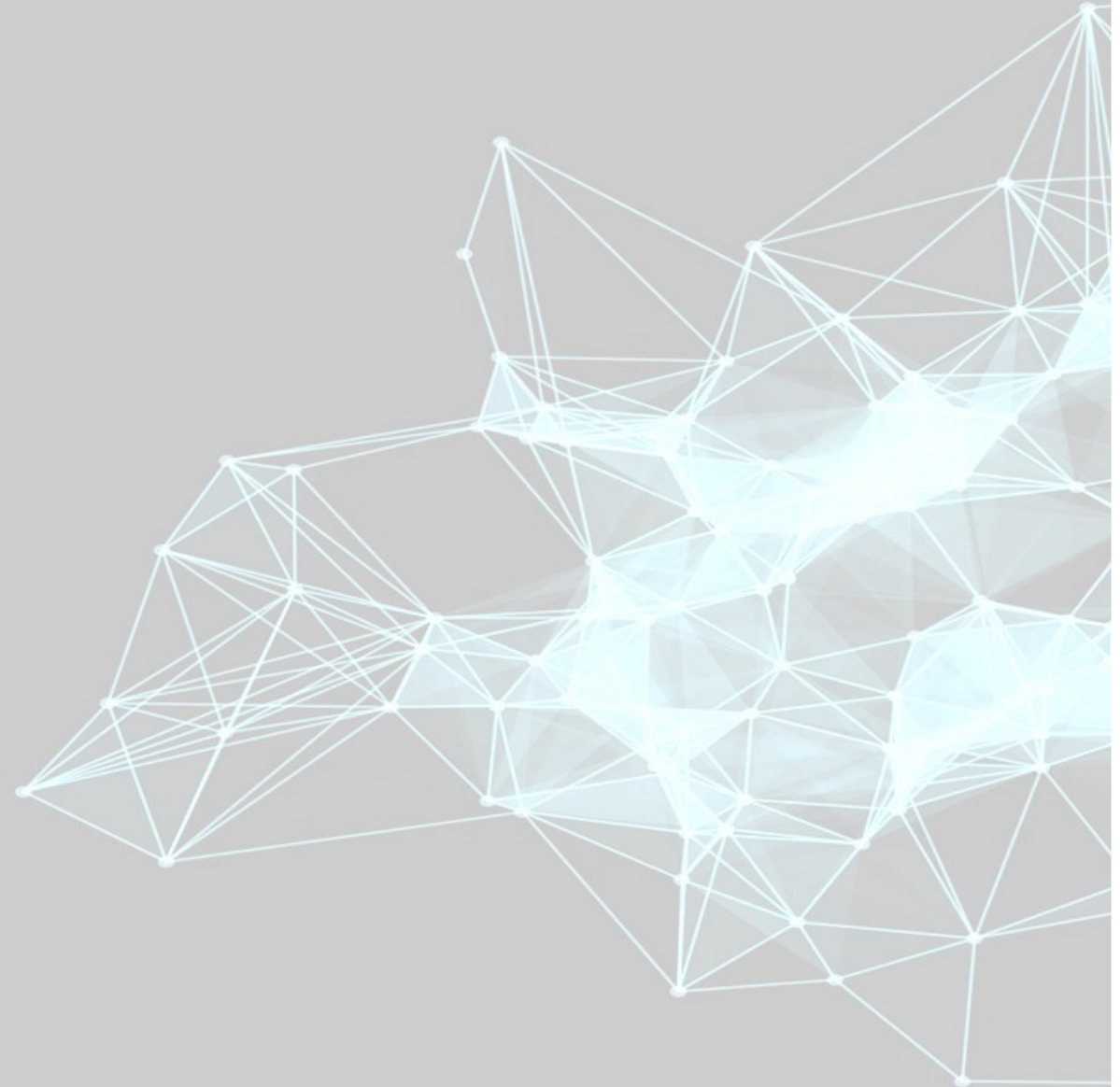
Som led i Bedre Balance I har regeringen samlet 16 nævn under ét tag i Nævnenes Hus i Viborg med en fælles administrativ ledelse, anvendelse af samme it-systemer mv.

I den forbindelse har ledelsen besluttet, at alle de enkelte nævns data skal samles i ét fælles datavarehus, og at potentialerne ved at anvende machine learning som understøttende værktøj til sagsbehandlingen skal afprøves.

Nævnenes Hus oplever stor opbakning fra alle medarbejdere, og forklarer selv, at det skyldes at organisationen er ny og hovedparten af medarbejderne kun har været i organisationen i 2 år.

Kapitel 8

Projektorganisering og kompetencer



Indholdsfortegnelse.



Introduktion

Vi introducerer undersøgelsens formål og baggrund, og giver et ledelsesresume.

1



Formål og baggrund

Introduktion til undersøgelsen og dens grundlag..... s. 04

2



Ledelsesresume

Overblik og sammenfatning..... s. 10



Analytics

Vi definerer analytics og belyser statens erfaringer og potentialer inden for analytics.

3



Definition af analytics

Definition af analytics og dets begrebsverden – hvad er analytics.....s. 18

4



Erfaringer og potentialer inden for analytics

Projekter og muligheder inden for analytics – hvor står staten.....s. 35



Datamiljø

Vi beskriver datatyper, og hvordan de er organiseret.

5



Datatyper

Strukturerede og ustrukturerede data – hvilke data har staten.....s. 52

6



Datamanagement

Organisering og styring af data – hvordan forvalter staten data..... s. 64



Ledelse og kompetencer

Vi undersøger den digitale lederrolle, projektstyring og kompetencer inden for analytics.

7



Digital ledelse

Ledelsesmæssige perspektiver på "det digitale" og analytics – hvilken rolle spiller ledelse s. 79

8



Projektorganisering og kompetencer

Projektstyring og kompetencer inden for analytics projekter – hvordan udvikles analytics..... s. 96

Opsummerende resultater: Projektorganisering og kompetencer



Hvordan udvikles og testes nye analytics løsninger?

Vi beskriver nu organisering og styring inden for analytics, samt hvilke kompetencer der er behov for, med følgende konklusioner:

1. Hvordan organiseres og styres analytics projekter? – forankring, metode og implementering.
2. Der skal tages stilling til projektets forankring og ressourcer – enkelte institutioner har valgt at etablere særlige analytics teams
3. Der skal tages stilling til projektets metodiske tilgang – enkelte institutioner har positive erfaringer med at afprøve idéer hurtigt og i lille skala
4. Projektet skal undersøge mulige implementeringsudfordringer – flere institutioner ser idriftsættelse af analytics fra laboratorium til virkelighed som den sværeste opgave
5. Analytics projekter involverer en række forskelligartet kompetencer på forskellige tidspunkter fra idéudvikling til projektudførelse over implementering
6. Kompetencer inden for digital ledelse – ledelsen har brug for opkvalificering
7. Kompetencer inden for analytics – kun få institutioner har medarbejdere inden for machine learning og matematisk optimering
8. Kompetencer inden for datamanagement – institutioner henter konsulentbistand til databehandling, men er det hensigtsmæssigt?
9. Kompetencer inden for forretningskendskab – institutionernes medarbejdere er eksperter på deres fagområde

Hvordan organiseres og styres analytics projekter?



Der skal tages stilling til projektets forankring og ressourcer...

Der er flere vigtige forhold i organiseringen og styringen af analytics projekter, men følgende tre er helt centrale.

For det første skal der tages stilling til projektets forankring og ressourcer, dvs. hvem eller hvilket kontor har hovedansvaret for at drive projektet, og hvilke ressourcer er der behov for.

Det er vigtigt at placere det ledelsesmæssige ansvar entydigt for at undgå, at projektet lander mellem to stole eller går tabt i den daglige drift.

Der skal også tages stilling til, om der er behov for indhentning af særlige kompetencer.



...metodiske tilgang...

For det andet skal der tages stilling til projektets metode, dvs. hvordan analytics projektet skal gennemføres.

Der synes generelt at være to tilgange i hver sin ende af et kontinuum.

På den ene side kan projektet følge en mere helhedsorienteret og systematisk tilgang med vægt på at projektet skal tænkes ind ift. institutionens øvrige datamiljø, strategiske arbejde, projekter mv.

På den anden side kan projektet følge en mere fokuseret og agil tilgang med vægt på at afprøve små som store idéer hurtigt og i lille skala.



... og mulige implementeringsudfordringer

For det tredje skal mulige implementeringsudfordringer adresseres.

Det er vigtigt, at analytics projekter skitserer implementeringsforløb og forholder sig til mulige udfordringer i implementeringen af specifikke løsninger, idet de bør indgå i beslutningsgrundlaget for om den pågældende løsning kan realiseres i praksis.

Er der nogle særlige opmærksomhedspunkter /risici? Hvilke forebyggende eller afbødende indsatser er der behov for?

Der skal tages stilling til projektets forankring og ressourcer – enkelte institutioner har valgt at etablere særlige analytics teams



Hvem er ansvarlig?

Analytics projekter involverer som regel flere aktører inden for institutionen, men ofte også på tværs af ministerområder, herunder fx i forbindelse med dataindsamling, relevante snitflader mv. Det er derfor vigtigt, at det ledelsesmæssige ansvar for projektet bliver placeret entydigt for at sikre fremdrift, og for at minimere risikoen for at projektet går tabt mellem to stole.

Enkelte institutioner har valgt at forankre deres analytics projekter i særlige analytics teams i deres digitaliseringsafdeling. Det kunne eksempelvis være Miljøstyrelsen, der har etableret et såkaldt team ”Jump”, eller Energinets team ”Advanced Analytics”, som begge har til opgave, at drive og innovere institutionens analytics dagsorden.

Energinets analytics team har hovedansvaret for at udvikle og teste konkrete idéer samt opstille businesscases inden for fire uger, hvorefter det er op til det berørte fagkontor, at vurdere, om de ønsker at sætte løsningen i drift og finansiere løsningen via deres budget. Herved er der et klart ejerskab i løbet af processen imellem de involverede parter.

Andre institutioner har omvendt valgt, at projekterne skal forankres i de enkelte fagkontorer. Det kunne eksempelvis være Finanstilsynet eller Landbrugsstyrelsen, som i højere grad placerer ansvaret hos alle fagkontorer med henblik på at få den faglige viden om den konkrete opgaveløsning i spil i udviklingen af mulige analytics løsninger.

Hvilke ressourcer og kompetencer er der behov for?

I forlængelse af projektets forankring, skal der tages stilling til projektets ressourcer for yderligere at minimere risikoen for at projektet går tabt i den daglige drift.

Det er flere institutioners erfaring, at det er svært at løfte analytics projekter parallelt med den øvrige drift og opgaveløsning. Af samme grund har Energinet og Miljøstyrelsen netop valgt at etablere særlige analytics teams dedikeret til at drive dette arbejde, hvor der er fuldtidsansat 3-5 medarbejdere med kompetencer inden for data science.

Det er vigtigt, at projektgruppen har adgang til de rette data science kompetencer, og det skal derfor afklares, om det er behov for at institutionerne opbygger disse internt, herunder fx opkvalificere eller rekruttere nye medarbejdere, eller indhente ekstern bistand, herunder fx konsulenter, samarbejde med universiteter eller ansættelse af erhvervs ph.d.

Endeligt er det ligeså vigtigt, at projektgruppen har adgang til kompetencer inden for det berørte fag- eller forretningsområde, hvor analytics løsningen evt. skal anvendes. Fagpersonel bør inddrages for at skabe synergi mellem forretningsforståelsen og den teknologiske viden, og sikre, at løsningen er realiserbar i virkeligheden.

Vi vender tilbage til de forskellige kompetenceprofiler senere i kapitlet.

Der skal tages stilling til projektets metodiske tilgang – enkelte institutioner har positive erfaringer med at afprøve idéer hurtigt og i lille skala



Hvordan skal projektet gennemføres?

Der synes generelt at være to tilgange i hver sin ende af et kontinuum.

På den ene side kan projektet følge en mere helhedsorienteret og systematisk tilgang med vægt på en trinvis modningsproces, samtænkning ift. institutionens øvrige datamiljø, strategiske arbejde, projekter mv. Det kunne eksempelvis være, at der skal være styr på institutionens samlede datagovernance før analytics projekter igangsættes. Ledelsen prioriterer som udgangspunkt sikker drift af kerneopgaven over udviklingsprojekter. Tilgangen flugter med de mere risikoaverse ledere, jf. s. 86.

Et eksempel på ovenstående kunne være Landbrugsstyrelsen, hvor det er de enkelte fag- eller forretningsområder, der skal udvikle og teste idéer på baggrund af en række trinvis niveauer i modenhedstrappen.

På den anden side kan projektet følge en mere fokuseret tilgang med vægt på at afprøve små som store idéer hurtigt (agil og kort tidsramme) og i lille skala (afgrænset opgave) for at kunne vurdere om en analytics løsning er realiserbar, og har økonomisk og kvalitativt potentiale (value for money). Herved bliver effektiviseringspotentialet mere synligt, og dermed kan beslutningstagere investere i implementeringen af analytics løsninger på et mere oplyst grundlag. Ledelsen ser det som et formål i sig selv at be- eller afkræfte løsningernes potentiale, idet fejl gør institutionen klogere på ens forretning og kerneopgaver og bringer den tættere på succes (fail fast). Tilgangen betegnes ofte "proof of concept" metoden, og flugter i højere grad med de mere risikovillige ledere.

Det kunne eksempelvis være Erhvervsstyrelsen og Energinet.

Caseeksempel

Citater



"Idéer til analytics skal afprøves ud fra "Proof of concept"-metoden, der giver mulighed for hurtigt og agilt at teste om løsningen har "value for money"

ENERGINET

"Det er vigtigt, at projekterne blive lavet small-scale først, så man kan teste løsningen og vurdere om der er en businesscase. Denne metode minimerer risikoen for store udviklingsomkostninger betydeligt og bidrager til, at man finder frem til de gode løsninger"

Projektet skal undersøge mulige implementeringsudfordringer – flere institutioner ser idriftsættelse af analytics fra laboratorium til virkelighed som den sværeste opgave



Hvorfor undersøge implementeringsudfordringer?

Flere institutioner fremhæver, at de største udfordringer ved analytics projekter som regel ligger i forbindelse med implementeringen eller idriftsættelse af løsningerne snarere end udviklingen af algoritmer mv.

Som beskrevet i slide 94 under forrige kapitel kræver anvendelsen af nye digitale løsninger et særligt ledelsesmæssigt fokus med blik for meningsskabelse, oversættelsesprocesser og læring. Hvis ledere ikke prioriterer ”at lede forandringen”, opstår der misforståelser og i værste fald modstand fra de medarbejdere eller brugere, der skal anvende løsningen.

Projektgruppen skal derfor foretage en række implementeringsovervejelser, der skitserer implementeringsforløbet, og identificerer mulige udfordringer og risici, der kan hæmme en optimal anvendelse af den pågældende analytics løsning, og i værste fald en pessimal anvendelse, der kan medføre en forringet opgaveløsning.

Ledelsen skal derfor have indsigt og overblik over det potentielle udfordringsbillede samt mulige aftagende/forebyggende indsatser, og på den baggrund vurdere om løsningen er holdbar uden for laboratoriets testmiljø og realiserbar i virkeligheden.

Hvilke mulige udfordringer er der i implementeringen?

Overordnet set kan man opdele udfordringsbilledet i tekniske og menneskelige udfordringer.

De tekniske udfordringer vedrører datamiljøet, hvor det skal afklares om analytics løsningen kan integreres med institutionens nuværende dataarkitektur, databehandling og datagovernance, jf. s. 67. Det kunne eksempelvis være spørgsmål som disse:

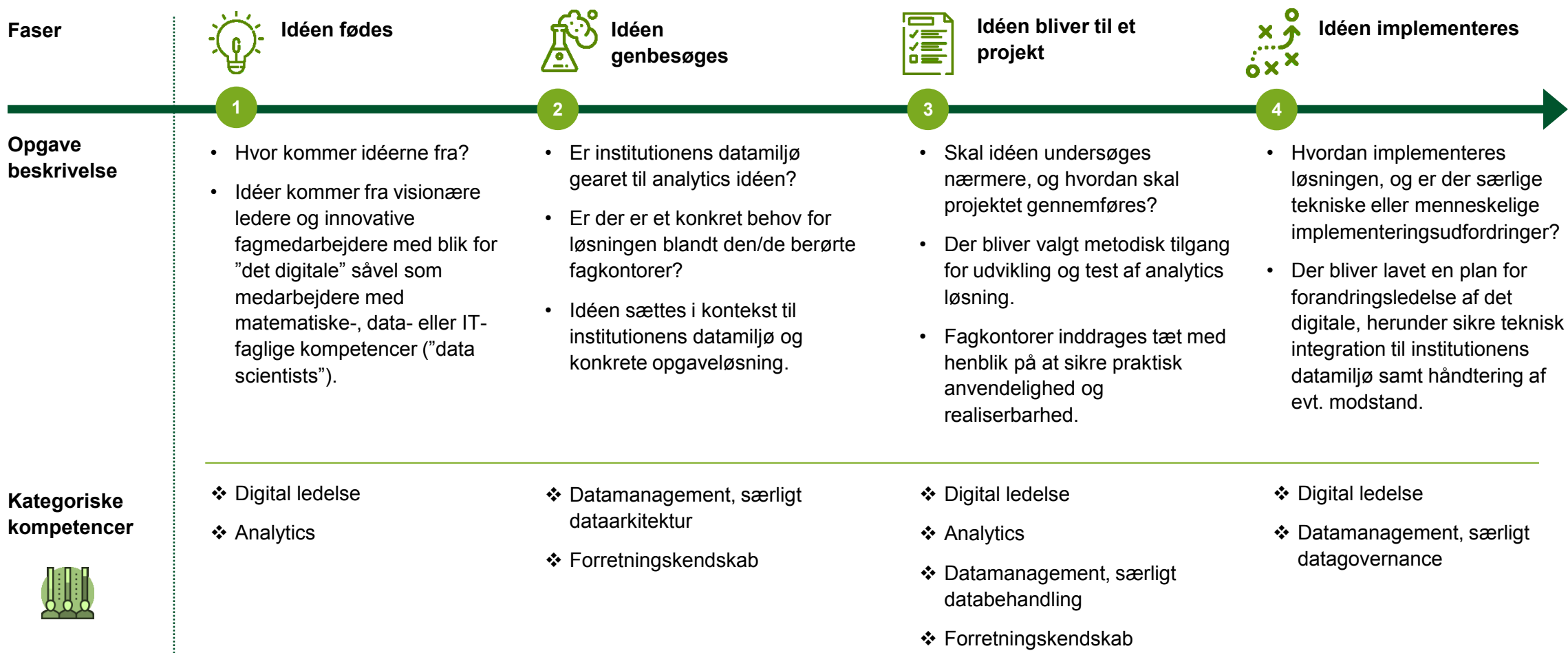
- Er der særlige tekniske udfordringer, der hindrer at det nye BI-system trækker og opdaterer data automatisk fra institutionens fagsystemer, og hvordan skal vi vedligeholde BI-systemet de næste to-tre år?
- Hvis medarbejdere skal anvende tablets med opgavespecifik information i landområder – er der tilstrækkelig internetdækning her?

De menneskelige udfordringer vedrører brugernes omstilling fra ”det gamle” til ”det nye”, og kompetenceudvikling, som fx:

- Hvor stor modstand kan vi forvente fra medarbejdere, og er der særlige persongrupper, der vil være kritiske?
- Har medarbejdere tilstrækkelige digitale kompetencer, eller er der behov for at opkvalificere?

God forandringsledelse skal forbygge modstand, men også sikre, at medarbejderes erfaringer fra virkeligheden med løsningen bliver hørt og brugt til at foretage evt. tilpasninger.

Analytics projekter involverer en række forskelligartet kompetencer på forskellige tidspunkter fra idéudvikling til projektudførelse over implementering



Kompetencer inden for digital ledelse – ledelsen har brug for opkvalificering



Fase Kompetencer inden for digital ledelse

Flere institutioner ser muligheder ved analytics, men har svært ved at konceptualisere deres visioner

De fleste institutioner har strategisk fokus på dele af analytics, men der er stor forskel på hvor langt institutionerne er i deres visioner, jf. s. 85. En mindre gruppe institutioner, fx Energinet og Erhvervsstyrelsen, er nået relativt langt, men flertallet af institutioner efterspørger et fælles koncept og begrebsapparat inden for analytics dagsorden, som kan understøtte idéudviklingen mv.

Der indhentes generelt eksternt bistand til inspiration, mens vidensdeling på tværs af staten er mangelfuld.

Institutionernes ledelse har brug for et digitalt kompetenceløft

Der synes, at være en mangel på ledere som evner at indgå i en faglig dialog om indholdet og retningen for analytics projekter, herunder analytics værktøjernes grundlæggende funktion og metode, jf. s. 92-93. Ledere skal i stigende grad kunne forklare diverse dataetiske problemstillinger, samt forstå mulige konsekvenser bag analytics metoderne, herunder fx hvorfor machine learning algoritmen udpeger enkelte borgere eller virksomheder frem for andre. Derudover synes det generelt at være svært at praktisere ledelsesmæssig risikovillighed i analytics projekter i en offentlig kontekst, hvor udvikling af kerneydelserne som udgangspunkt ikke må ske på bekostning af stabil drift og service, jf. s. 91.

Flertallet af institutionerne har ikke erfaring med ”at lede” større implementeringer af analytics projekter

Forandringer af medarbejderes arbejdsrutiner og kompetencekrav kræver en særlig ledelsesmæssig indsats med henblik på at sikre ”buy in” og undgå modstand blandt medarbejdere, jf. s. 94-95.

Størstedelen af analytics projekter i staten har været på ”proof of concept” niveau, hvor kun enkelte projekter er blevet implementeret, herunder fx Energinet, Erhvervsstyrelsen og Landbrugsstyrelsen. Der er generelt en lav grad af erfaring med forandringsledelsen af analytics.

Kompetencer inden for analytics – kun få institutioner har medarbejdere inden for machine learning og matematisk optimering



Fase Kompetencer inden for analytics



+



Institutionerne har generelt medarbejdere med kompetencer inden for informationssystemer og økonometri, men ikke machine learning og matematisk optimering

De fleste institutioner har medarbejderprofiler inden for business intelligence, opgavespecifik information og/eller økonometri. De er som minimum i stand til gennemføre deskriptive analyser af data såvel som grundlæggende regressionsanalyser ud fra en hypotetisk deduktiv tilgang. Det er typisk økonomer og erhvervsøkonomer, men i enkelte institutioner også fagspecialister, der er BI-autodidakte. Fødevarestyrelsen har eksempelvis ansat dyrlæger, farmaceuter mv. med selv lærte BI-kompetencer med henblik på at sikre en stærk faglig kobling til forretningsområdet.

De færreste institutioner har medarbejderprofiler inden for machine learning og matematisk optimering, som er i stand til at udvikle og træne algoritmer eller opbygge avanceret matematiske modeller til optimering af en konkret funktionsopgave. Enkelte institutioner har ansat såkaldte "data scientists", herunder Energinet, Erhvervsstyrelsen og Miljøstyrelsen i form af bl.a. fysikere, ingeniører og matematikere. Derudover forsøger enkelte institutioner at oplære deres økonomer/statistikere i machine learning, matematisk optimering mv., herunder fx Arbejdstilsynet, STIL og STAR. STIL har gode erfaringer hermed, mens en anden institution har fremhævet, at det er begrænset, hvor avanceret teknologier økonomer/statistikere kan lære, idet de som regel ikke kan programmere.

En gruppe institutioner oplever, at det er svært at rekruttere og fastholde "data scientists" inden for machine learning eller matematisk optimering på grund af en begrænset lønramme, og en manglende risikovillighed til løbende prioritering af udviklingsprojekter. Data scientists er i høj kurs i den private sektor, og det er derfor svært for de offentlige institutioner at konkurrere på løn med de private institutioner. Desuden bliver det fremhævet, at det er svært at fastholde data scientists, hvis deres udviklingsprojekter forsinkes eller nedprioriteres pga. proces og drift. Flere institutioner indhenter generelt ekstern bistand i form af konsulenter, samarbejde med universiteter eller ansættelse af erhvervs ph.d.

En anden gruppe institutioner er omvendt lykkedes med at rekruttere og fastholde både erfarne og nye data scientists. De fremhæver, at det væsentligste er at skabe et dedikeret arbejdsmiljø i institutionen, hvor der er "like-minded" ingeniører, matematikere mv. De finder det attraktivt at få mulighed for at arbejde med de mange unikke og forskelligartede offentlige data, og motiveres af at kunne løse og forbedre samfundsmæssige problemstillinger. Derudover følger der en række attraktive arbejdsvilkår i det offentlige, som tiltrækker nyuddannede data scientists, herunder fx arbejdstid. Det handler i høj grad om at skabe opmærksomhed omkring ens dagsorden på Hackathon og lignende.

Kompetence-vurdering



Informationssystemer og økonometri



Machine learning og mat. opt.

Kompetencer inden for datamanagement – institutioner henter konsulentbistand til databehandling, men er det hensigtsmæssigt?



Fase Kompetencer inden for datamanagement



Institutionerne har generelt få medarbejdere med kompetencer inden for dataarkitektur. Datavarehuse bliver typisk udviklet og implementeret af konsulenter, men driftes af institutionerne selv eller Statens It.



Dataarkitektur dækker over opbygningen, implementeringen og driften af it-systemer, hvor et væsentligt element ift. analytics er etableringen af databaser og sikre et it-miljø, der følger med den teknologiske udvikling, jf. s. 67. Størstedelen af institutionerne indhenter konsulenter eller benytter Danmarks Statistik til hhv. udvikling og implementering af datavarehuse, herunder fx Nævnenes Hus, Erhvervsstyrelsen, Arbejdstilsynet, Uddannelses- og Forskningsministeriet og Børne- og Socialministeriet. Driften varetages som udgangspunkt af institutionerne selv eller Statens It, og der synes derfor generelt at være medarbejdere ansat, som kan vurdere om institutionens datamiljø er gearet til den pågældende analytics løsning. Alle institutioner uden for et større it-fællesskab skal overgå til Statens It frem mod 2021, og det bør derfor afklares nærmere, i hvilket omfang Statens It bidrager eller kan bidrage som teknisk rådgiver ift. datamiljø og analytics, herunder fx cloud computing.



Institutionernes kompetencer inden for databehandling er uklart, men bliver delvist løftet af konsulenter og Danmarks Statistik



Databehandling vedrører strukturering, rensning, berigelse og validering af data til det videre analysearbejde, og det væsentlige ift. analytics er, at foretage dataudtræk der kategoriseres og bearbejdes ift. det pågældende analyseformål, jf. s. 67. Institutionernes kompetencelandskab ift. databehandling er uklart, men visse dele af disse processer bliver hos nogle institutioner foretaget af konsulenter/Danmarks Statistik, herunder fx STAR, Styrelsen for Forskning og Uddannelse, Børne- og Socialministeriet. Store dele af databehandlingsopgaverne er gentagende processer, hvor det samme rådata skal gennemgå samme behandling og formatering som tidligere rådata, og det bør derfor undersøges nærmere, om indhentning af konsulentbistand til disse opgaver er hensigtsmæssigt.



Flere institutioner er ved at styrke deres kompetencer inden for datagovernance



Datagovernance vedrører regler og procedurer for forvaltning af data internt og eksternt i institutionen, hvor et væsentligt element ift. analytics er at sikre ensartet og konsistente definitioner af data mv. samt sikre overholdelse af databeskyttelsesregler, jf. s. 67. Flere institutioner efterspørger medarbejdere med juridiske og datamæssige kompetencer inden for datagovernance, som kan sikre overholdelse af databeskyttelsesloven, og opbygge en effektiv og transparent styring og udveksling af data i organisationen, herunder fx Trafik-, Bygge- og Boligstyrelsen.

Kompetencer inden for forretningskendskab – institutionernes medarbejdere er eksperter på deres fagområde



Fase Kompetencer inden for forretningskendskab



+



Institutionernes medarbejdere er eksperter på deres fagområde, men skal oplæres ift. mulighederne ved analytics og projektarbejde

Forretningskendskab vedrører den faglige indsigt i det pågældende fag- eller forretningsområde, hvor en analytics løsning evt. skal tages i brug, og institutionerne har givetvis ekspertviden på deres respektive områder. Institutionerne har derfor som udgangspunkt medarbejdere med forretningskendskab, som kan kvalitetssikre analytics idéen og behovet ift. deres konkrete opgaveløsning, men også sikre at løsningen rent faktisk er praktisk anvendelig og realiserbar.

Ikke desto mindre, synes der at være behov for at organisere inddragelsen af fagmedarbejdere i analytics projekter, herunder oplæring ift. analytics og dets muligheder/udfordringer, men også projektarbejde.

Kompetence-
vurdering

