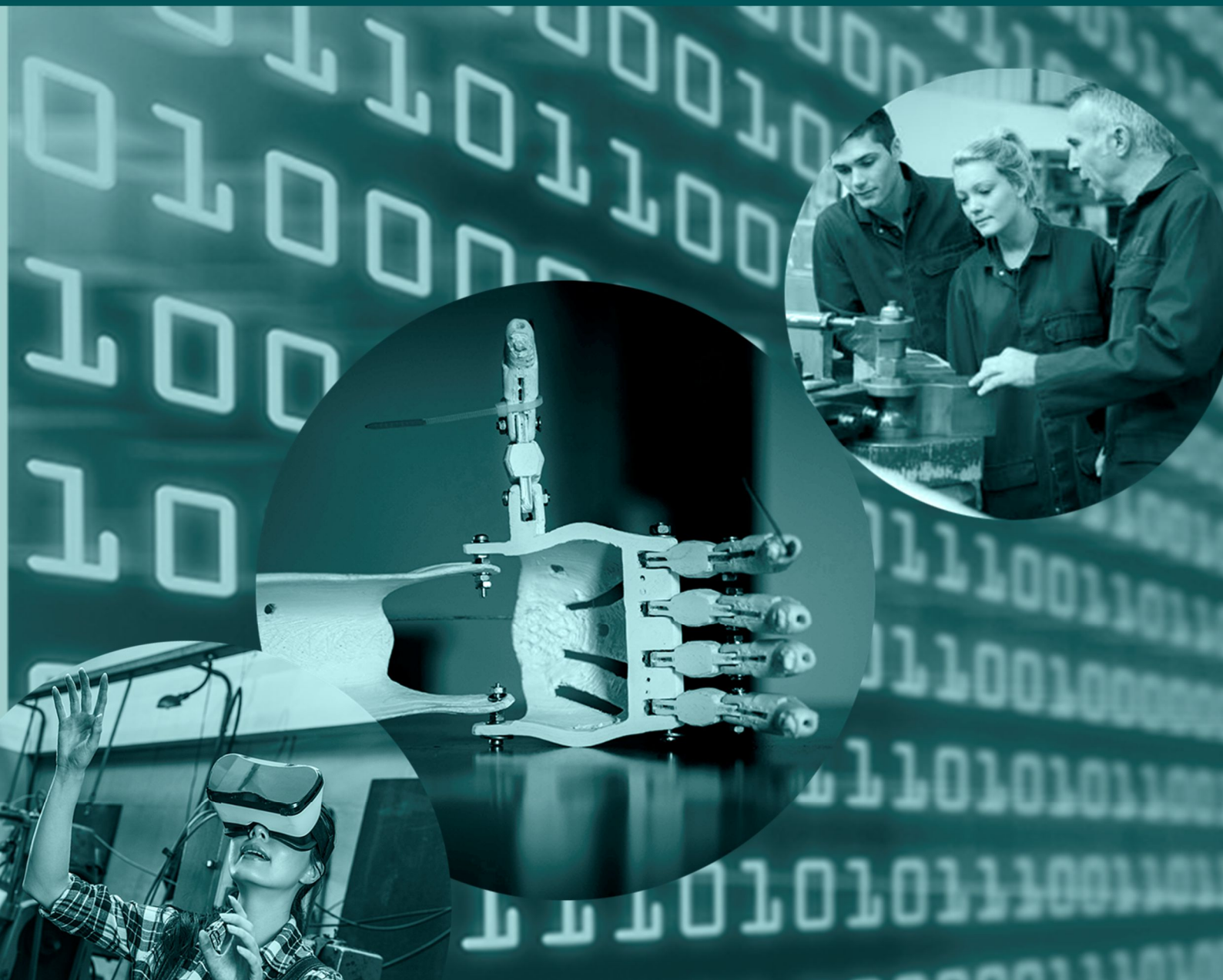
Big Data by Security

AFRAPPORTERING

INDUSTRIENS FOND FREMMER DANSK KONKURRENCEEVNE
The Danish Industry Foundation



Indledning

Big Data handler primært om at udnytte data til at foretage bedre beslutninger hvad enten disse foregår på det overordnede strategiske niveau eller er automatiserede operationelle mikrobeslutninger. Uden retvisende data er selv de bedste metoder utilstrækkelige, og her er udfordringen typisk, at de bedst mulige data er fortrolige enten fordi de er forretnings- eller personfølsomme. Det overordnede formål med Big Data by Security projektet er at bringe disse følsomme men værdifulde data i spil i data drevet it-systemer uden at gå på kompromis med fortroligheden.

Den grundlæggende udfordring er at sikre, at data kan anvendes fortroligt og uden brug af tredjeparter med adgang til rådata. Hvor der findes gode metoder til at beskytte data på lager (at rest) og under transport (in transit), er den helt store udfordring at beskytte data mens de anvendes (in process). En af de mest lovende metoder er en særlig form for kryptografi kaldet "Secure Multiparty Computation" (SMC), som gør det muligt at foretage beregninger direkte på krypterede data. Kort fortalt foretages beregningerne i et netværk af computere og på en måde så den enkelte computer ingen viden har om de data der regnes på. SMC tilhører en klasse af såkaldte "bevisbar sikkerhedsløsninger", som betyder at udgangspunktet er et matematisk bevis for sikkerhedsegenskaberne. Med SMC er beviset fx at der er "zero knowledge" hos de computere der udfører beregningerne.

Som projekt har Big Data by Security vist hvordan SMC kan anvendes i en data-dreven virkelighed ved at skabe nye måder at samarbejde om brugen af data. På den ene side har projektet vist hvordan konkurrende virksomheder kan samarbejde om data uden at bryde med hverken GDPR eller konkurrenceregler. På den anden side har projektet skabt et konkret bud på en ny tilgang til fortrolig samstilling og brug af offentlige register og private datakilder. Sidstnævnte er konkret forankret i et sandbox projekt med Danmarks Statistik og Sundhedsdatastyrelsen.

Aktiviteter og Leverancer

Projektet har pivoteret omkring nedenstående to konkrete cases:

CASE 1) Denne case omhandler konkurrenceudsættelse af lån. Her anvendes SMC til at sambringe fortrolig information til brug for kreditvurdering og med det formål at gøre det nemmere at konkurrenceudsætte lån.

CASE 2) Denne case omhandler benchmarking af energiforbrug. Her anvendes SMC til at sambringe fortrolig information med det formål at udbredde best practice indenfor energieffektiv produktion uden at gå på kompromis med fortroligheden.

I begge tilfælde er der udviklet konkret software som har muliggjort en meget direkte kommunikation med interessenter samt gjort det muligt at forankre projektets resultater. Særligt case 2 har udviklet sig i projektets løbetid og været årsag til flere forlængelser af projektet. Årsagen hertil har været det underliggende proof-of-concept som har testet om hvorvidt SMC kan bringe data fra Danmarks Statistik et skridt tættere på en operationel men fortrolig data-dreven virkelighed.

Mere om Case 1, som omhandler forenkling af de processer der er involveret i konkurrenceudsættelse af lån til gavn for både låner og udlåner. Kreditvurdering og andre administrative procedure gør det omkostningsfyldt at indhente konkurrerende tilbud på lån. Case 1 viser hvordan disse omkostninger kan reduceres ved at samle fortrolige data til brug for fortrolig (krypteret) kreditvurdering samt benchmarking. Kreditvurderingen giver også mulighed for at anvende udlåners egen (krypteret) kreditscoringsfunktion. Endelig er kreditvurderingerne koblet med et auktionssystem som gør det muligt for kunderne at anvende analyserne til at konkurrenceudsætte lån nemt og billigt.

Case 1 er et eksempel på, hvorledes SMC kan bruges til skabe en troværdig koordinerende tredjepartsinstitution til forenkling (og ultimativt automatisering) af processer så som konkurrenceudsættelse af lån. Eller mere generelt hvordan SMC

kan skabe et grundlag for data- og analysesamarbejde mellem konkurrende virksomheder til fordel for de enkelte virksomheder samt deres kunder.

Mere om Case 2, som omhandler benchmarking af energiforbrug. Her anvendes SMC til at understøtte udbreddelsen af best practice ved fortrolig (krypteret) beregning af energibesparelses-potentialer for danske industrivirksomheder. State-of-the-art benchmarking analyser anvendes til at udpege konkrete forbedringspotentialer samt anonymiseret forbilleder for den enkelte virksomhed, som bruger af systemet. Ved gensidig accept mellem brugeren og de udpegede forbilleder, afsløres identiteten på de forbilleder der har det største læringspotentiale.

Case 2 er samtidigt Proof-of-Concept for en model til anvendelse af følsomme data fra Danmarks Statistik udenfor Danmarks Statistiks mure. Historisk har Danmark været førende indenfor kontrolleret adgang til person- og virksomhedsfølsomme informationer. Med denne case vises det hvorledes SMC kan fungere som en sikkerhedsmodel til at bringe fortrolige data i spil direkte i online statistiske applikationer. Hermed kommer case 2 med et bud på hvorledes Danmark kan fastholde om ikke øge sin styrkeposition i udnyttelsen af følsomme data (til fordel for både virksomheder og samfundet generelt), til trods for øget krav til databeskyttelse.

De to cases er udviklet i tæt samarbejde med interessenter samt offentlig organisationer herunder Danmark Statistik samt Forbruger og Konkurrencestyrelsen. Det tætte samarbejde med såvel private og offentlige virksomheder og organisationer, har været afgørende for anvendeligheden samt den reelle forankring af projektets resultater.

Effekt

Projektet har konkretiseret og formidlet et alternativ til traditionel samkøring og opsamling af data i stadig større datasiloer. Ved brug af SMC er det muligt at anvende data som hvis de var samkørt på traditionel vis men uden at data reelt deles eller samles et sted.

Denne virtuelle tilgang til brug af data understøtter en data-dreven virkelighed hvor data er under decentral kontrol. Det vil sige at vi reelt kan stoppe yderligere opbygning af datasiloer men stadig anvende og samstille data på tværs af offentlige og private datakilder. Afhængig af hvordan teknologien anvendes, kan denne udvikling medføre betydelige ændringer på tværs af udbudskæder og industrier. Det kan give den enkelte borger kontrol over egne data og skabe nye samarbejder mellem virksomheder og organisationer. I begge tilfælde vil det potentielt påvirke relationen mellem data subjektet og de mange store og små serviceudbydere som anvender data. Hvornår og hvordan denne ændring vil indfinde sig er svært at sige og det er også stadig en påstand at der vil komme en sådan ændring. Der er dog en stigende tendens i denne retning.

Projektets to cases har primært fungeret som trædestene til reel værdiskabelse. Hvor case 1 angiver nye muligheder for samarbejde mellem konkurrende banker og øget konkurrence på lån, afhænger den reelle effekt af bankernes lyst til at samarbejde. I projektet har der været god dialog med en lang række banker, men om de reelt er interesseret i at udnytte projektets resultater er uklart. Uanset bankernes appetit på samarbejde, er resultatet et meget konkret eksempel på en ny metode til fortrolig samarbejde på tværs af juridiske og konkurrencemæssige skel som vil kunne overføres direkte til andre brancher.

I case 2 er effekten mere håndgribelig. På baggrund af projektets Proof-of-Concept udført indenfor rammerne af Danmarks Statistiks forskerservice, blev det besluttet at udvide testen af løsningen i et reelt distribueret setup på tværs af danske registre. Via en udvidelse af projektet er der derfor etableret et såkaldt “sandbox projekt” hvor Danmarks Statistik og Sundhedsdatastyrelsen udgør grundlaget for en “two-party SMC” model til virtuel samkøring og anvendelse af data på tværs af de to registre.

Forankring og Formidling

Den mest konkrete forankring af projektets resultater er det omtalte sandbox projekt, hvor Danmarks Statistik og Sundhedsdatastyrelsen tester projektets resultater i en udvidet såkaldt Virtuel Platform. Sandbox projektet er igangsat i regi af Big Data by Security projektet hvor den tekniske del er gjort klar til reel brug. Afslutningen af sandbox projektet og test med rigtige data vil foregå udenfor rammerne af Big Data by Security projektet og forventets gennemført i 2019.

Projektet har via et tæt samarbejde med blandt andet organisationerne Data for Good Foundation og Datafair samt en række andre interessenter indenfor anvendelse af sundhedsdata, igangsat en række initiativer som vil videreføre projektets resultater og gøre brug af den virtuelle platform. Projektet har ligeledes fungeret som trædesten til nye projektansøgninger i både Danmark og EU, som hvis de bliver finansieret vil bygge videre på projektets resultater.

Projektdeltagerne Alexandra Instituttet og Partisia, har bidraget til udvikling af software i regi af open source projektet FRESCO (https://fresco.readthedocs.io/en/latest/). FRESCO udgør grundlaget i den udviklede software og projektets bidrag har skabt øget interesse for FRESCO og vil være med til at udbrede SMC teknologien via dette open source community.

Projektet har også indirekte bidraget med input til et kommisionsarbejde i USA via Danmarks Statistiks deltagelse i forarbejdet med "Commision of Evidence-based policymaking". I den afsluttende rapport nævnes SMC som en lovende teknologi til virtuel brug af registre og som følge heraf bedre datagrundlag for politiske beslutninger. https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf.

Endelig er projektets resultater løbende blevet delt med de involverede interessenter og via projektdeltagernes netværk samt projektets hjemmeside (https://bigdatabysecurity.dk/da). Derudover er der afholdt to større workshops med bred deltagelse fra en række virksomheder og offentlige organisationer.

PROJEKTNAVN:
Big Data by Security

BEVILINGSMODTAGER:
Københavns Universitet

PROJEKTANSVARLIG:
Kurt Nielsen

MAIL:
kun@ifro.ku.dk

TELEFONNUMMER:
26 18 19 71

HOVEDNUMMER:
35 33 68 00

INDUSTRIENS FOND
FREMMER DANSK KONKURRENCEEVNE
The Danish Industry Foundation