

# BIG DATA

## BLIVER STORT I SEKTOREN

Store mængder data kan bruges til risikovurderinger, forhindre netbankssvindel og give automatiske kreditgodkendelser. Begrebet hedder Big Data og kan meget vel blive fremtidens it-revolution. Flere finansielle virksomheder bruger allerede elementer af Big Data, men teknologien er ikke i sig selv revolutionerende, så hvorfor er det blevet så populært?

**AF MILLE MORSING**

Du har måske prøvet at se på tøj i en webshop for derefter at se selvsamme tøj i en reklame på en anden hjemmeside. I stedet for standardiserede reklamer, der favner bredt, bruger virksomhederne nu dine data fra andre hjemmesider til at målrette deres markedsføring og appellere til for eksempel det tøj, der netop interesserer dig. Det kaldes Big Data og dækker over de store mængder struktureret og ustruktureret data, som du efterlader, når du køber en bog, tjekker vejret, synes godt om en opdatering på Facebook eller finder vej på et kort på nettet. De oplysninger kan virksomhederne bruge til at forbedre deres forretningsgange.

Og der produceres rigtig meget data. IBM vurderer, at der hver dag bliver produceret 2,5 milliarder gigabytes verden over, og 90 procent af verdens samlede datamængde er skabt alene over de sidste to år, og mængden forventes at blive fordoblet hvert tredje år. Derfor tales der i disse år rigtig meget om Big Data, og om hvordan man kan bruge de eksplosivt voksende data.

Men det er ikke noget nyt fænomen. Virksomheder har i årevis indsamlet og udnyttet data. Først talte alle om Data Warehouse, så var det Business Intelligence, og nu er storebroderen Big Data på alles læber.

Selvom kært barn har mange navne, og teknologien i grunden stadig er den samme, så er der alligevel sket et skift. Mængden af data er vokset, modeller er blevet mere komplicerede, og arbejdsgangene har ændret sig.

### Tilgængeligt data og billig hardware

Tidligere lå de forskellige data i hver deres silo, men nu arbejdes der på tværs af siloerne, og databasen er blevet én stor kasse med blandet viden, der skal sammensættes, så det giver mening. Disse data kan være interne som kundeinformation, opkald til callcentre, transaktioner, men også eksterne data som sociale medier og webcrawling, hvor man tager data fra andre hjemmesider til at optimere sine datamodeller.

Et eksempel er Amazon. Verdens måske førende webbutik bruger Big Data-elementer til at anbefale nye køb til kunder baseret på viden



**"Rådgiveren kan basere sin beslutning på 150 tidligere sager, mens en database kan have erfaring fra millioner af lignende sager, og det udvider grundlaget for beslutningen betragteligt"**

Christian Igel,  
professor ved Datalogisk Institut



## Hvad er Big Data?

Der findes mange forskellige definitioner på Big Data, og ikke alle er enige om, hvad begrebet omfatter. Nogle mener, at det er Big Data, når man har en enorm mængde data til rådighed, mens andre mener, at det er kompleksiteten af data, der afgør, om det er Big.

Big Data dækker over værktøjer, som bruges til at håndtere store komplekse mængder data. Disse data kan være både struktureret og ustruktureret og bruges til at optimere forretningsgange.

Big Data kan for eksempel bruges til:

- Kreditvurdering
- Afsløring af svig
- Målrettet marketing
- Konkursforudsigelser
- Kreditkortgodkendelse

om tidligere køb, kundens profil og andre kunder, som har købt lignende produkter.

Professor i datamining og machine learning ved Datalogisk Institut Christian Igel mener, at Big Data er blevet så stort, fordi hardware er blevet billigere og data mere tilgængeligt.

”Mængden af tilgængelig digital information vokser, og finansielle virksomheder indsamler og opbevarer data mere systematisk – og jo flere data man har tilgængelige, jo mere præcise kan systemerne blive”, siger Christian Igel og fortæller samtidig, at succeshistorier som Amazon og Google også har været med til at give andre virksomheder blod på tanden.

### Computeren giver et lån

Flere finansielle virksomheder er også begyndt at se på mulighederne ved Big Data, som for eksempel vil kunne bruges til at vurdere, om kunden skal have et ekstra lån, og om en hacker prøver at bryde ind i netbanken. I fremtiden kunne man også forestille sig, at hvis du søgte efter biler på internettet og skrev statusopdateringer om, at du var træt af at tage bussen, så kunne der dumpe en mail ind med et tilbud fra banken om et billån.

Et andet eksempel er, når kunder skal kreditvurderes af bankrådgivere.

Det vil være en subjektiv vurdering baseret på rådgiverens uddannelse og erfaring, men med Big Data vil beslutningen i høj grad ligge hos computeren, der baserer resultatet på objektiv statistik.

”Rådgiveren kan basere sin beslutning på 150 tidligere sager, mens en database kan have erfaring fra millioner af lignende sager, og det udvider grundlaget for beslutningen betragteligt. Kombineret med menneskelig erfaring vil det derfor føre til bedre beslutninger”, fortæller Christian Igel, som netop har lavet et forskningsprojekt med Danske Bank, hvor de udelukkende brugte en database til at foretage kreditvurdering.

Tanken bag projektet var, at i stedet for en formaliseret kreditvurdering, der blev krydstjekket med data efterfølgende, så skulle processen være komplet datastyret med data fra tidligere lånesager. Selvom Danske Bank i flere år har brugt modeller til at assistere risiko- og kreditvurdering, så bekræftede projektet med Datalogisk Institut, at teknologien

kan hjælpe med udviklingen af modeller.

”Vi har længe kendt til den måde at lave modeller på, som kaldes machine learning, men med projektet fik vi bekræftet, at vi kan bruge en ny teknologiplatform til formålet. Jeg ser en masse potentiale i Big Data, men de konkrete fordele ligger længere ude i fremtiden”, siger chef for Danske Banks Risk and Credit Models, Karsten Bo Malten, og fortæller samtidig, at Danske Bank er i gang med at opbygge infrastrukturen til at håndtere de mange ustrukturerede data.

Projektet var en succes, men selvom computeren kunne træffe beslutningerne uden menneskelig indgriben, så mener Christian Igel ikke, at vi i fremtiden helt skal forlade os på data og analysemodeller.

”Computere kan træffe meget præcise beslutninger og finde komplekse mønstre, som mennesker ville overse, men i visse tilfælde vil den rigtige beslutning kræve information, der ikke kan trækkes ud af databasen. Her kommer mennesket ind i billedet, fordi vi kan reagere ud fra sund fornuft og intuition”.

### Store kræfter på spil

Nordea er også en af de banker, som har taget Big Data til sig. I Nordea bruger de teknologien til risikovurdering, blandt andet til Credit Value Adjustment, som er modpartsrisikoen ved market-forretninger,

og Value at Risk, som er et estimat udviklet til at forudsige det maksimale tænkelige tab. Her laver de tusindvis af simulationer på forskellige markedsudfald, men med så store mængder data kræver det også stærke computerkræfter at holde styr på dem, fortæller Charles Tye, der er manager for Nordeas Fixed Income, Derivatives and Risk Systems, IT Capital Markets.

”Udfordringen kommer, når vi er nødt til at opdatere risici i realtid med nye transaktioner eller undersøge resultaterne nærmere for at forstå dem. Resultaterne er nemlig ikke lineære, og en ny transaktion kan enten forhøje eller mindske den samlede risiko for hele porteføljen, hvilket betyder, at vi skal behandle en potentielt stor mængde transaktioner, hver gang vi laver en beregning”, fortæller Charles Tye.

For at kunne beregne de store mængder data bruger Nordea in-memory-servere. Normalt har man data liggende på

“Jeg ser en masse potentiale i Big Data, men de konkrete fordele ligger længere ude i fremtiden”.

Karsten Bo Malten,  
chef for Danske Banks  
Risk and Credit Models

diske og lader så databasesoftwarens læse og skrive til diske, efterhånden som der bliver behov for det, men med in-memory-teknologi kører databaserne automatisk i ram, hvilket går hundrede til tusinde gange hurtigere.

Men det er ikke kun danske banker, der flirter med Big Data. Den islandske bank Landsbankinn har også haft succes med at bruge mønstergenkendelse i et Big Data-system til at overvåge sine kunders transaktioner for at forhindre netbankssvindel. Banken bruger loginmønstre fra 250 millioner brugere fra 8.000 banker verden over til at opdage, hvis der foregår noget usædvanligt på kundernes bankkonti. Det kan være, hvis kunder overfører et beløb, på et tidspunkt hvor de normalt aldrig logger på, eller overfører mere, end de plejer. Systemet kan opdage de mistænkelige transaktioner og dermed ringe eller sende en sms til kunder, som verificerer, at det er dem, som foretager transaktionen.

### Big Data er kommet for at blive

Selvom virksomheder og it-professionelle taler meget om fordelene ved at indsamle en masse data, mener Christian Igel, at det er vigtigt, at virksomhederne først finder ud af, hvad de skal bruge al den information til, før de bruger penge og kræfter på at indsamle det.

”Man er nødt til at vende det på hovedet og starte med at finde behovet. Hvad vil jeg tilbyde mine kunder? Hvordan vil jeg forbedre min forretning? Derefter skal man se på, hvordan de nødvendige data kan indsamles. Hvis man bare indsamler data, fordi alle andre gør det, så vinder man ingenting”, siger Christian Igel og påpeger, at virksomheder ikke behøver at bruge enorme ressourcer på at høste ustruktureret data fra sociale medier for at udnytte Big Data.

”Internt kundedata er allerede værdifuld viden, så hvis virksomhederne gør deres hjemmearbejde, holder sig opdateret og bringer deres system op til dataming standarder, er de allerede godt placeret i forretningslandskabet”.

Big Data er talk of the year, men det er ikke bare en trend, der vil blive glemt om et par år. Det er en udvikling, mener Christian Igel, og han ser ingen grund til, at virksomheder stopper med at indsamle data.

”Jeg forventer en øget grad af dataindsamling, og når virksomheder begynder at opbevare mere data, vil vi se en lang række nye anvendelser. Fra et forskningsperspektiv er der mange spændende udfordringer i at udvikle metoder til at trække nyttig viden ud af kæmpe databaser. På uddannelsesområdet ser vi også, at studerende lærer mere dataanalyse end for bare få år siden”. ■

## Data prissætter ejendomme

I en ikke så fjern fremtid kan det meget vel tænkes, at computermodeller præcist kan forudsige, hvad en ejendom er værd, hvis den skulle handles nu. Virksomheden e-nettet har udviklet en analysemodel baseret på elementer fra Big Data, som kan værdisætte fast ejendom. Modellen prissætter ejendommene med udgangspunkt i data fra offentlige registre, demografisk data og udbudsdata fra ejendomsmæglerne. Modellen anvendes i dag til vurderinger, hvor der ikke er krav om fysisk fremmøde på ejendomme. Der er dog stadig et stykke vej, før man helt kan undvære at være fysisk til stede på ejendommen.

”Modellen kommer til kort, når det gælder de subjektive vurderinger af ejendommens tilstand. Hvor gammelt er køkkenet, eller er der hul i taget? Det kan man jo kun se, når man står derude. Alle de her ustrukturerede data vil vi gerne have fat i, fordi det kan løfte modellen endnu mere”, siger e-nettets forretningsudvikler Anders Hvid, som gerne vil skabe samarbejdsalliancer på tværs af sektoren for at dele disse data.

”Det er jo ikke et konkurrenceparameter at værdisætte ejendomme, og lidt forenklet findes der kun én sandhed. Derfor burde sektoren arbejde sammen om både datadeling og modeller. I dag vurderes en ejendom først af ejendomsmæglerne, derefter af pengeinstitutter og til sidst af SKAT. Det er spild af ressourcer”.

Modellen fungerer i dag, som et støtteværktøj for professionelle aktører, og ifølge Anders Hvid vil den inden længe kunne afløse en del af de fysiske besigtigelser.

”Jo flere data modellen medtager, jo stærkere bliver den. Om vi på sigt vil opleve modelbaserede udbudspriser, er svært at forudsige, men ikke umuligt. I sidste ende er den rigtige pris dog fortsat, hvad en given køber vil betale, på det tidspunkt ejendommen er udbudt”. ■