

Så skapas värde ur stora datamängder

– rätt hantering, avancerad analys och artificiell intelligens

Mars 2019

Linda Olofsson
Jeremy Jean-Jean
Mikael Ström



RISE Research Institutes of Sweden

RISE är Sveriges forskningsinstitut och innovationspartner. I internationell samverkan med företag, akademi och offentlig sektor bidrar vi till ett konkurrenskraftigt näringsliv och ett hållbart samhälle. Våra 2 700 medarbetare driver och stöder alla typer av innovationsprocesser. RISE är ett oberoende, statligt forskningsinstitut som erbjuder unik expertis och ett 100-tal test- och demonstrationsmiljöer för framtidssäkra teknologier, produkter och tjänster. www.ri.se

Från den 1 oktober 2018 är RISE Research Institutes of Sweden AB indirekt den enda aktieägaren i RISE IVF AB (före detta Swerea IVF AB).

RISE IVF AB
Box 104
431 22 Mölndal
Telefon 031-706 60 00
www.swerea.se

RISE IVF-rapport 19003

© RISE IVF AB

Förord

Denna rapport är resultatet av en förstudie finansierad av Swerea IVFs Intressentförening. Målsättningen har varit att ta fram en enkel guidebok som ger en introduktion till området stora datamängder, beskriver de möjliggörare i form av tekniska verktyg och system som finns samt beskriver hur dessa möjliggörare kan användas för att skapa värde i produktframtagningsskedjan i en organisation.

Innehållsförteckning

1	Introduktion till guideboken	5
2	Datainfrastruktur och verktyg för komplex analys	12
2.1	Datastruktur	12
2.2	Big data	13
2.3	Datainfrastruktur, från databas till datasjöar	14
2.4	Val av infrastruktur?	19
2.5	Moln eller lokalt?	20
2.6	Datainfrastruktur - Sammanfattning	21
2.7	Artificiell intelligens, maskininlärning, djupinlärning	22
2.8	Artificiell intelligens - Sammanfattning	29
3	Tillämpningar av stora datamängder för produktframtagningen	30
3.1	Data och analyser	30
3.2	Stora datamängder i produktion	31
3.3	Stora datamängder för produktutveckling	43
3.4	Värde för kunden	47
4	SWOT-analys av stora datamängder	50
4.1	Styrkor	51
4.2	Svagheter	52
4.3	Möjligheter	53
4.4	Hot	54
5	Slutsatser	55
6	Referenser	58

Sammanfattning

Stora datamängder har vuxit fram som den nya råvaran för att skapa värde för organisationen.

Denna rapport ger en introduktion till området genom att beskriva de tekniska verktyg och system som finns samt hur dessa möjliggörare kan användas för att skapa värde i produktframtagningsskedjan i en organisation.

Om en organisation lyckas dra nytta av den kraftfulla tekniken kommer det med säkerhet vara en stor konkurrensfördel medan det kommer att vara en stor nackdel om konkurrenter lyckas bättre. För att lyckas behöver organisationen både en analytisk mognad och en digital infrastruktur som möjliggör tillämpningen av de tekniska verktygen. Utöver att organisationen behöver förståelse för hur man kan använda sig av data för att skapa värde så måste kvaliteten och tillgången på råvaran data också säkerställas.

Summering av råd till organisationer som vill skapa värde av stora datamängder:

- Öka den analytiska förmågan i organisationen
- Identifiera och prioritera bland möjliga värden som tekniken kan erbjuda
- Säkerställ tillgången på rätt data av hög kvalitet, anpassa den digitala infrastrukturen utifrån de värden som ska skapas, inklusive skydd av värdekedjan mot intrång och störningar
- Ha koll på lagstiftning och standardisering.

Summary

Large amounts of data have emerged as the new raw material for creating value for the organization.

This report gives an introduction to the area by describing the technical tools and systems that exist and how these facilitators can be used to create value in the product realization process in an organization.

If an organization succeeds in taking advantage of the powerful technology, it will surely be a great competitive advantage while it will be a major drawback if competitors were to succeed better. To succeed, the organization needs both an analytical maturity and a digital infrastructure that enables the application of the technical tools. The organization needs not only to understand how to use data to create value, but also needs to ensure the quality and availability of the data.

Summary of advice for organizations that want to create value from large datasets:

- Increase the analytical ability of the organization
- Identify and prioritize among possible values that technology can offer
- Ensure the availability of the right high-quality data, adapt the digital infrastructure based on the values to be created, including protection of the value chain against intrusion and interference
- Keep track of legislation and standardization.

1 Introduktion till guideboken

Stora datamängder har vuxit fram som den nya råvaran för att skapa värde för organisationen.

Denna rapport är resultatet av en förstudie finansierad av Swerea IVFs Intressentförening. Målsättningen har varit att ta fram en enkel guidebok som ger en introduktion till området stora datamängder, beskriver de möjliggörare i form av tekniska verktyg och system som finns samt beskriver hur dessa möjliggörare kan användas för att skapa värde i produktframtagningsskedjan i er organisation.

I rapporten definierar vi stora datamängder, eller big data, som så stora mängder av information att det inte går att hantera på en vanlig dator eller via en vanlig databas. Det ställer krav på en infrastruktur. Det finns lite variation i litteraturen hur big data definieras. Ibland utgår man istället från de utmaningar som åtföljer hanteringen och värdeskapandet av stora datamängder, något vi återkommer till nedan, och låter detta big data definieras av dessa gemensamma utmaningar (Hu 2014). Resultatet blir i allt väsentligt detsamma – det behövs dedikerade verktyg och system för att kunna skapa värde av stora datamängder.

Hur kan stora datamängder skapa värde i en organisation? I en ofta citerad rapport från McKinsey (McKinsey 2011) pekas på det stora ekonomiska värde som finns i stora datamängder och rapporten identifierar ett antal sätt som stora datamängder kan skapa värde på för en organisation, bland annat:

- Data kan skapa transparens till exempel genom att dela data mellan olika avdelningar och enheter i organisationen vilket kan minska ledtider, förbättra kvalitet och öka överskådligheten
- Tillgång till data kan möjliggöra kontrollerade experiment för att analysera skillnader i prestation, visualisera variationer och påvisa behov
- Data kan möjliggöra segmentering för kund Anpassning av produkter och tjänster
- Data kan stödja och även ersätta beslutsfattande genom automatiserade algoritmer
- Data kan bidra till nya affärsmodeller, produkter och tjänster.

Hur ska man då tänka när man vill skapa värde ur stora datamängder? Det är viktigt att förstå vilket värde tekniken kan skapa, vilken mognadsgrad (både analytiskt och digitalt) organisationen har idag och var den behöver befinna sig för att kunna skapa det önskade värdet. Det vill säga:

- Förstå möjligheterna den nya tekniken skapar och vilket värde den skulle kunna bidra med i organisationen. Identifiera och prioritera värden som ska skapas. Vilken automatiserad information skulle skapa störst värde för organisationen? För kunderna?

-
- Analys, vilken analytisk och digital mognad har organisationen idag? I vilken utsträckning använder företaget idag dataanalys som underlag för sitt beslutsfattande och i vilken grad är denna dataanalys automatiserad. Vad krävs för att ta komma till nästa önskade nivå?

I denna guidebok vill vi bidra till insikt vad gäller båda ovanstående punkter. Vi kommer att fokusera främst på produktframtagningsprocessen. I kapitel 2 kommer vi att beskriva tekniken bakom stora datamängder för att visa hur kraftfull den är och vilka möjligheter den ger. I kapitel 3 kommer vi därefter att presentera hur denna teknik kan tillämpas i produktframtagningsprocessen i företaget. I det avslutande kapitlet (Slutsatser) kommer vi att knyta ihop och ge några konkreta råd hur organisationen kan gå vidare med arbetet.

Man kan definiera tre nivåer av analytisk mognad i organisationer (MIT Sloan review 2013):

- 1 **Analytiskt innovativa organisationer:** Detta är organisationer med en stark analytisk kultur och ett datadrivet beslutsfattande dvs de lutar sig på analys för strategiska beslut och innovation.
- 2 **Analytiskt utövande organisationer:** Detta är organisationer som har tillgång till data och jobbar på att bli mer datadrivna men använder data primärt för operationella förbättringar.
- 3 **Analytiskt utmanande organisationer:** Detta är organisationer som i hög grad förlitar sig på magkänsla i sitt beslutsfattande, som har svårt med access till data och saknar kunskap för att hantera data på ett värdeskapande sätt.

För att bli en analytiskt innovativ organisation krävs att man har tillgång till all relevant data och därtill möjlighet att utföra komplex analys på densamma. Data och analys är något som går över både avdelnings- och organisationsgränser. Företag som delar data internt får mer värde av sin analys. Det är också mer sannolikt att de företag som är mest innovativa och framgångsrika i sin analys delar sin data även utanför företaget. Så kallat datapartnership kan skapa stort affärsvärde (MIT Sloan review 2016) men skapar även frågetecken kring äganderätt, tillgång och andra rättigheter till den data som finns i samarbetet. Det ställer även krav på ett gemensamt språk; hur kategoriseras den delade informationen?

För att nå den fulla potentialen av data krävs således samarbete, att man delar data och att det finns åtkomst till all relevant data. För att kunna göra detta krävs en dedikerad infrastruktur kopplat till en digital mognad (kunskap om tekniken). Vi har valt att beskriva detta i fyra steg och använder dess tillämpning i produktion som exempel, se bild 1 nedan:

Steg 1: All relevant data är inte åtkomlig. Data är inte digitalt lagrad och/eller den är lagrad i "silo" till exempel data är finns i utrustningen eller finns lagrad i olika databaser. Tillgängliga data kan användas för viss databehandling och enklare beskrivande analys.

Steg 2: Data är digitalt lagrad med åtkomst över hela organisationen. Relevant utrustning är uppkopplad med data som lagras tillgängligt. Analys utförs kopplat till beskrivning till exempel av produktionsprocessen och diagnos av händelser.

Steg 3: All relevant data är samlad digitalt och används för komplex analys till exempel förutsägelse av produktionskapacitet.

Steg 4: Hela värdekedjan för data är automatiserad, från generering och lagring till komplex analys och åtgärd. Produktionen optimeras automatiskt.

I dessa olika steg ser beslutfattande och vem som vidtar åtgärder olika ut, se bild 2 nedan. I det första steget används data för beskrivning medan det är människan som utför diagnos, fattar beslut och vidtar åtgärder medan det i fjärde steget kan vara helt automatiserat.

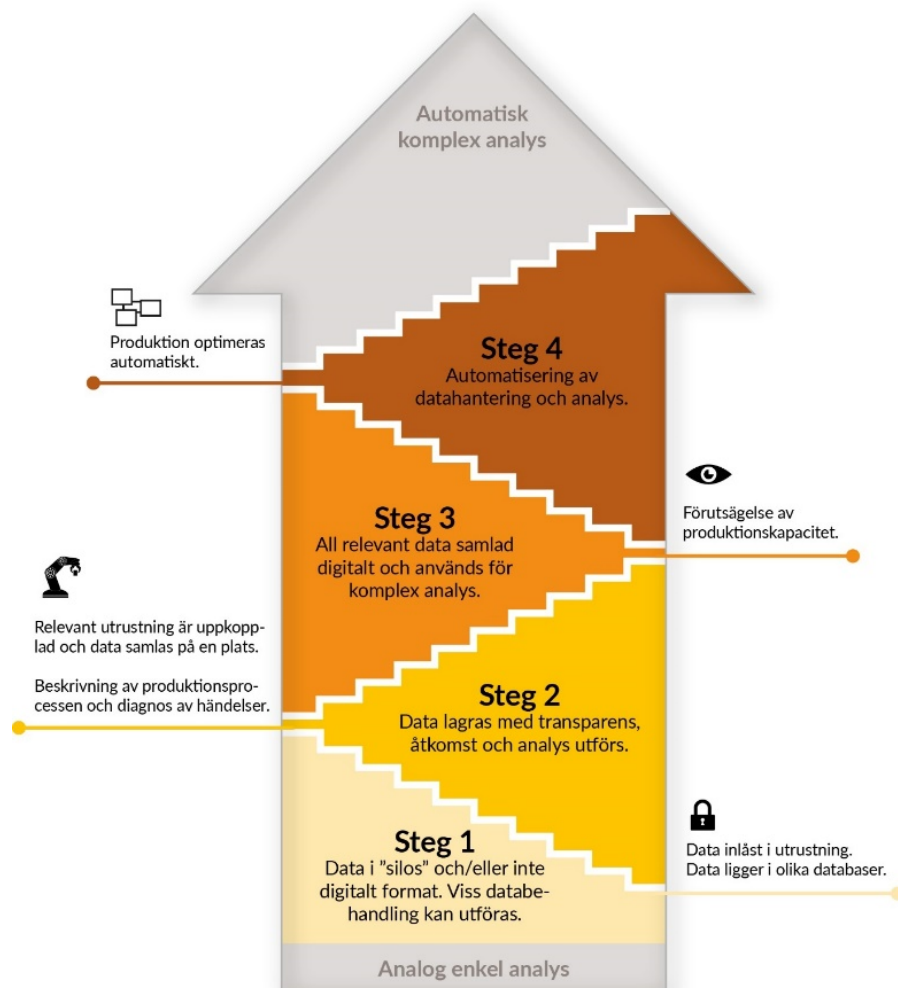


Bild 1 Digital infrastruktur och mognad indelad i fyra steg med exempel från produktion.

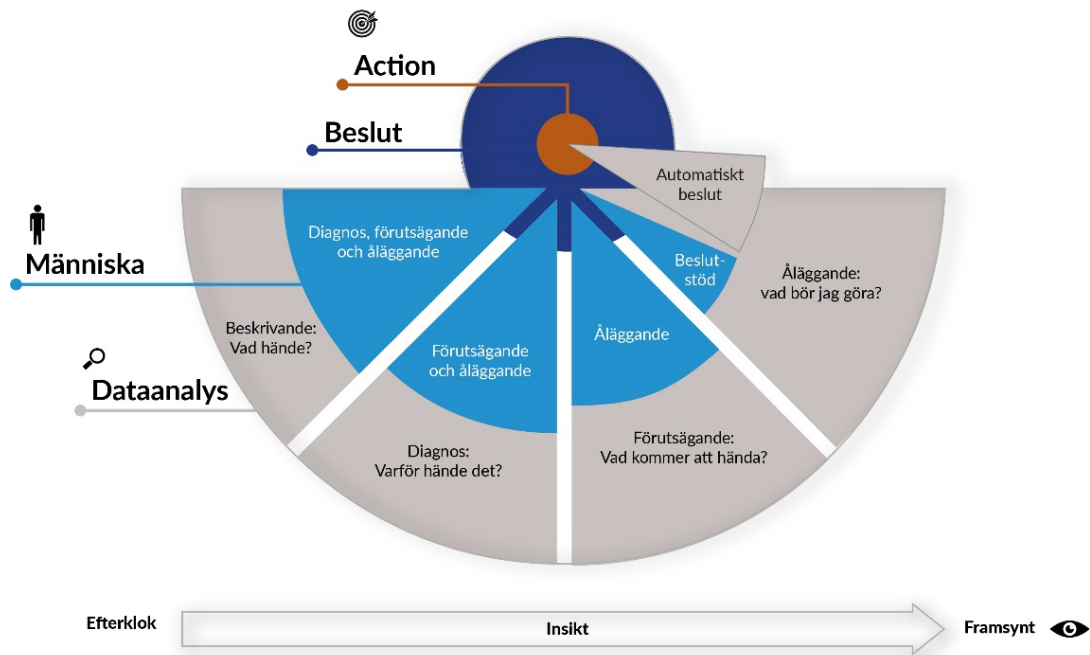


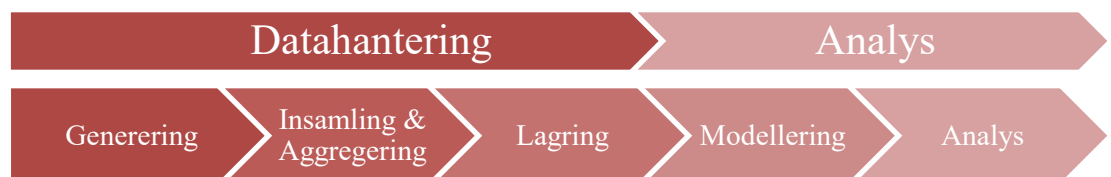
Bild 2 I det olika stegen automatiseras gradvis diagnos, beslutfattande och åtgärder. Något förenklat skulle man kunna säga att data går från att i steg ett användas till att förstå vad som hände (efterklok) till att i steg fyra förutse och föreskriva (framsynthet)

Faktaruta: Värdekedjan för stora datamängder

Värdekedjan för stora datamängder består av två delar uppdelade i fem underdelar:

- Datahantering
 - Generering
 - Insamling och Aggregering (inklusive sortering och preparering)
 - Lagring
- Analys
 - Modellering
 - Analys

För att kunna använda data i analys måste ett arbete läggas på att förbereda data. I synnerhet i de fall då data kommer osorterat från olika källor med olika format kan detta arbete vara ganska tids- och resurskrävande.



Var befinner sig organisationen just nu och vad krävs för att ta den dit där information från stora datamängder kan vara en hävstång för värde internt och för dess kunder? Vilken information skulle hjälpa organisationen om den fanns? Vad skulle man behöva förutse, optimera och automatisera? Vad skulle hjälpa dess kunder om de kunde göra, förutse etc?

Även med en hög digital mognad i organisationen finns utmaningar med stora datamängder som kräver kunskap och resurser. Organisationen måste bemästra hela värdekedjan för stora datamängder. Man talar om olika dimensioner när det gäller utmaningar för stora datamängder (Research now Gatner 2001) från början tre dimensioner vilka nu utvidgats till fem dimensioner, fem V:n på engelska. Dessa beskrivs ytterligare i kapitel 2:

- Volym (volume)
- Hastighet (velocity)
- Variation (variety)
- Tillförlitlighet (veracity)
- Värde (value)

För att bemästra värdekedjan måste man börja vid källan. Varifrån kommer data, hur genereras den? Är källan tillförlitlig? Vilka format är den på? Måste data behandlas (tvättas) genom sortering, filtrering, borttagning av felaktiga data och/eller aggregeras för att kunna lagras och analyseras? Den stora variationen på olika format gör det svårt att samla och integrera data från olika källor med skalbarhet. Nästa steg, att lagra data ställer krav på system som kan både samla och lagra stora datamängder samtidigt som systemet kan erbjuda avancerad analys. I kapitel 2 beskrivs tekniska lösningar som uppfyller dessa krav.

Faktaruta: Historik

Bakgrunden till området stora datamängder, eller big data, kan ses i den enorma utveckling som skett inom hårdvaruutvecklingen med miniatyrisering och kostnadsreduktion. Denna utveckling har inneburit att mängden data som går att spara till samma pris, har dubblats vartannat år (i enlighet med Intelgrundaren Gordon E. Moores förutsägelse). Samma utveckling har också bidragit till att andra komponenter som bygger på halvledarteknik samtidigt blivit ofantligt mycket billigare och bättre, till exempel kameror och andra typer av sensorer. Tillsammans med utvecklingen inom kommunikationsteknik och internet har det resulterat i en utveckling av tidigare aldrig skådat slag som nu möjliggör snabb generering och överföring av data i kombination med massiv digital lagring och komplex analys.

Detta utgör grunden för det som kallas Internet of Things (IoT) och ibland Internet of Everything (IoE). Det är därför vi nu kan koppla upp sensorer, maskiner och människor och samla in, spara, och analysera all den ofantliga mängd data som genereras. Stora tillgängliga datamängder möjliggör i sin tur möjlighet att skapa lärande algoritmer som kan tränas och lära sig på data för att kunna förutsäga och göra prognoser. Därmed kan artificiell intelligens skapas.

Data kan analyseras både i realtid och genom batchvis analys (Erl 2016).

Realtidsanalys ställer högre krav på infrastruktur och hastighet. Det betyder i princip att data måste analyseras i samma takt som den produceras. Om produktionstakten är hög krävs en snabb bedömning av data, med fördel automatiserad, behov av tvättning och tillförlitlighet måste snabbt bedömas.

Realtidsanalys, strömning, är en fördel i till exempel produktionsövervakning när man vill kunna avbryta eller korrigera då analys av data indikerar behov därav.

Batchvis analys kan vara tillräcklig vid analys av till exempel kundbehov vid produktutveckling eller skapandet av en ny tjänst. Då finns mer tid till städning av data och att skapa en bra struktur. Kapitel 2 och 3 belyser olika aspekter på vilken analys som kan och bör utföras.

Andra viktiga områden att ta hänsyn till när man ska skapa värde ur stora datamängder berör säkerhet och personlig integritet.

- Datasäkerhet (cyber security): hur skyddas organisationens data från stöld eller yttre manipulation? I skyddet av data bör hänsyn tas till hur pass värdefull och stöldbegärlig data kan vara för någon utomstående och graden av skydd designas därefter.
- General Data Protection Regulation (GDPR), på svenska Dataskyddsförordningen, är till för att skydda, stärka och harmonisera, enskildas grundläggande rättigheter och friheter, särskilt deras rätt till skydd

av personuppgifter. Dataskyddsförordningen gäller i hela EU och har till syfte att skapa en enhetlig och likvärdig nivå för skyddet av personuppgifter så att det fria flödet av uppgifter inom Europa inte hindras. Mycket i dataskyddsförordningen liknar de regler som fanns i personuppgiftslagen. (Datainspektionen 2019, Wikipedia - 1) För att uppfylla förordningen måste ett antal principer uppfyllas, för mer information se mer på datainspektionens webb: <https://www.datainspektionen.se/lagar-regler/dataskyddsförordningen/grundläggande-principer>

- Blockchain är en teknik som kan tillämpas när man vill säkerställa ursprunget till data och vara förvissad om att ingen manipulerat den. Det kan därför vara ett viktigt verktyg då det ställs höga krav på spårbarhet och tillit till ursprung hos data. Tekniken är förenklat beskrivet ett system av distribuerade huvudböcker/liggare som det inte går att ändra i (manipulera) eftersom alla i systemet har en egen kopia.
- Standardisering och lagstiftning: det pågår arbete kring standardisering av internet of things med bäring på stora datamängder och därtill kommer pågående arbete inom lagstiftning om framför allt ägande och tillgång till data. En utveckling om detta återfinns i kapitel 3.

2 Datainfrastruktur och verktyg för komplex analys

2.1 Datastruktur

Det är viktigt att förstå strukturen hos de uppgifter som kommer att samlas in och analyseras senare eftersom det kan påverka den typ av infrastruktur som behövs.

2.1.1 Strukturerade data

Strukturerade data är data som har organiserats i fasta fält och poster som en tabell, ett kalkylblad eller en databas (bild 3). De följer en definierad modell för hur de kommer att lagras och bearbetas. Uppgifterna kan ha olika typer: numerisk, valuta, datum, bokstäver etc.

first_name	last_name	order_id	order_total
Larsson	Erik	123456	12,34
Jonsson	Matilda	98765	98,76

Bild 3 Strukturerade data exempel

2.1.2 Halvstrukturerade data

Halvstrukturerade data är nivån mellan strukturerade och ostrukturerade data. Det följer en viss fördefinierad struktur men har fler frihetsgrader (bild 4). Taggar eller markörer används för att identifiera vissa element som kan hjälpa till med behandlingen av informationen. Vanliga exempel på halvstrukturerad data är metadata i ett e-postmeddelande (avsändare, mottagare, ämne, datum), en sångfil (artist, år, spår, album), loggarna från en maskin eller ett automatiserat system eller ett html-dokument (webbsidor).

```
[
  {
    first_name : "Larsson",
    last_name  : "Erik",
    order_id   : "123456",
    order_total : "12,34"
  },
  {
    first_name : "Jonsson",
    last_name  : "Matilda",
    order_id   : "098765",
    order_total : "98,76"
  }
]
```

Bild 4 Halvstrukturerade data exempel

2.1.3 *Ostrukturerade data*

Ostrukturerade data utgör cirka 80 % av de data som genereras varje dag. De kan inte enkelt läggas i en tabell eller databas och innehåller bilder, ljud, videor, pdf-filer, sociala nätverksmatningar, PowerPoint-presentationer, Word-dokument etc. Data kommer på väldigt olika format.

Eftersom data nu genereras i exponentiell takt och eftersom det mesta är ostrukturerat, måste företagen anpassa sig och kunna extrahera informationen för att få värde av den. Denna uppgift är en allt viktigare del av dagens verksamhet och kräver att specialkompetenser, verktyg och infrastrukturer finns.

2.2 **Big data**

Mängden data växer ständigt och det som idag räknas som en stor mängd data, kommer troligtvis att räknas som en genomsnittlig eller liten mängd data inom några år. Stora datamängder låter sig inte behandlas på en vanlig dator eller server. Experter talar om de fem V som vart och ett representerar en stor datautmaning.

- **Volume:** Lagring av data som har genererats och hur den bearbetas
- **Velocity:** Den hastighet som data skapas och behovet av realtidsanalys. Detta konstanta flöde av data utgör en stor teknisk utmaning som traditionell databehandling inte kan lösa.
- **Variety:** Blandningen av olika typer av data. Uppgifterna kan struktureras (text, siffror), halvstrukturerade (pdf, XML, e-post, sensordata) eller ostrukturerad (bilder, videoklipp, tweets)
- **Value:** Betydelsen, värdet eller funktionaliteten av data i organisation som konsumerar den.
- **Veracity:** Tillförlitlighet, trovärdighet och noggrannhet hos data som är beroende av källan, kräver mer behandling före analys. Ofta förekommer det i sociala medier innehåll som kan innehålla typsnitt, förkortningar och samtal och behöver extra bearbetning innan det kan vara meningsfullt för en dator.

E-handel, sociala medier och IoT (Internet of Things) är typiska datakällor som kommer att utmana organisationernas IT-infrastrukturer och processer och kräver specifika datalösningar.

Teknik för stora datamängder som datasjöar och algoritmer för maskininlärning används alltmer i alla branscher, eftersom de erbjuder många fördelar jämfört med mer traditionell lösning. Datasjöar (en grupp av datorer som lagrar och bearbetar all data från en organisation) kan hantera en nästan obegränsad mängd data och är den infrastruktur som stöder alla verktyg som senare används för analys. Artificiell intelligens och maskininlärning kan göra förutsägelser baserade på tidigare händelser och utföra diagnostik med bättre noggrannhet än människor, tack vare dess förmåga att förstå och upptäcka mönster i stora datamängder.

2.3 Datainfrastruktur, från databas till datasjöar

Datainfrastruktur är en kritisk del av dagens organisationer och hjälper till att utnyttja värdet från data. Behovet varierar beroende på typen av data, dess volym och antal källor – från en enda databas till en datasjö. Varje infrastruktur har sitt särskilda användningsområde och sina för- och nackdelar.

2.3.1 Databas

Databaser är det vanligaste infrastrukturen för lagring av data. Det är en organiserad samling data som oftast lagras i rad, kolumner och tabeller och som enkelt kan nå sammanställas och ändras med ett frågesystem. Det kan bara lagra data från en eller en begränsad mängd datakällor (bild 5).

Databaser används för att stödja organisationens interna verksamhet och kräver strukturerade data. Typiska databasapplikationer kan vara material- och produktionsstyrningssystem, system för kundrelationer, journaler på sjukhus, data från automatiserade system som väderstationer etc.

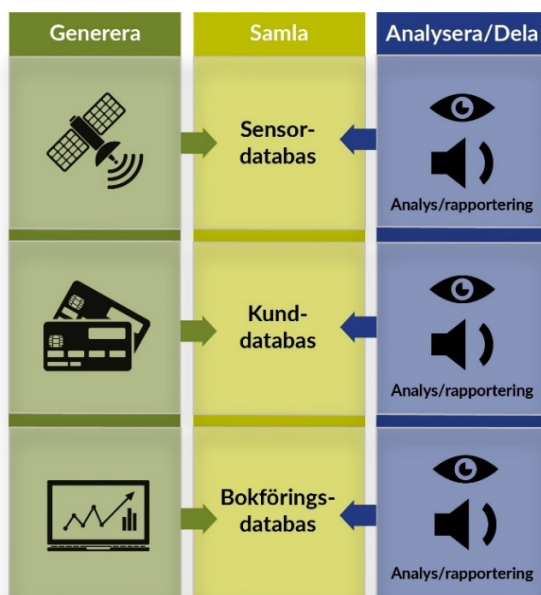


Bild 5 Gemensam arkitektur för en databas

Tabell: För- och nackdelar med databaser.

Fördelar	Nackdelar
<ul style="list-style-type: none"> ▪ Förmåga att lagra stor mängd information ▪ Bättre dataintegration ▪ Snabb åtkomst till information ▪ Minska dataredundans ▪ Informationssäkerhet 	<ul style="list-style-type: none"> ▪ Kräver mycket tid att designa ▪ Stora kostnader för uppstart av hårdvara och mjukvara ▪ Eventuella skador på databasen är en skada på alla program som använder den ▪ Dyrt underhåll ▪ Rapportering, visualisering och analys kan inte utföras över mycket stora uppsättningar av datakällor och dataströmmar ▪ Långsam när det gäller ostrukturerade data ▪ Kan leda till datasilos bildas där varje avdelning har sin egen databas som inte de andra kan komma åt, vilket gör det svårt att hitta nya datamönster.

När en databas blir alltför stor eller när en databassamling blir alltför svår att hantera kan ett datalager eller en datasjö användas.

2.3.2 Datalager

Datalager kan vara ett organisationsövergripande förvar för alla data som samlats in av organisationen och brukar användas för att lagra data från källor som Excel, ERP (Enterprise Resource Planning), CRM (Customer Relationship Management) eller finansiella databaser. Det är en relationsdatabas som lagras på en server eller i molnet och används för affärsunderrättelseaktiviteter, beslutsstöd och svara på användarnas förfrågningar. Det möjliggör snabb datahantering och analys, enkel integration i befintliga system och kan snabbt identifiera och korrigera fel i data. Datalager är bra lösning för att lagra historiska data för att avlasta en aktiv databas.

Den största fördelen med ett datalager är att det ger samma format för alla data av intresse oavsett källan. För att göra det måste rådata från olika källor förbehandlas genom att gå igenom ETL-fasen (extract, transform, load), innan de går in i lagret (bild 6). ETL är en kärnkomponent i ett datalagringsystem som:

- extraherar data från olika källor och omvandlar den till ett konsoliderat format redo för omvandling i nästa steg
- omvandlar data utifrån verksamhetsaspekter och krav på validering och lagring
- laddar data i datalagret

ETL-fasen är tidskrävande eftersom alla operationer måste definieras av en operatör och spåras grundligt över tid, men den är avgörande för att datalagrets funktion.

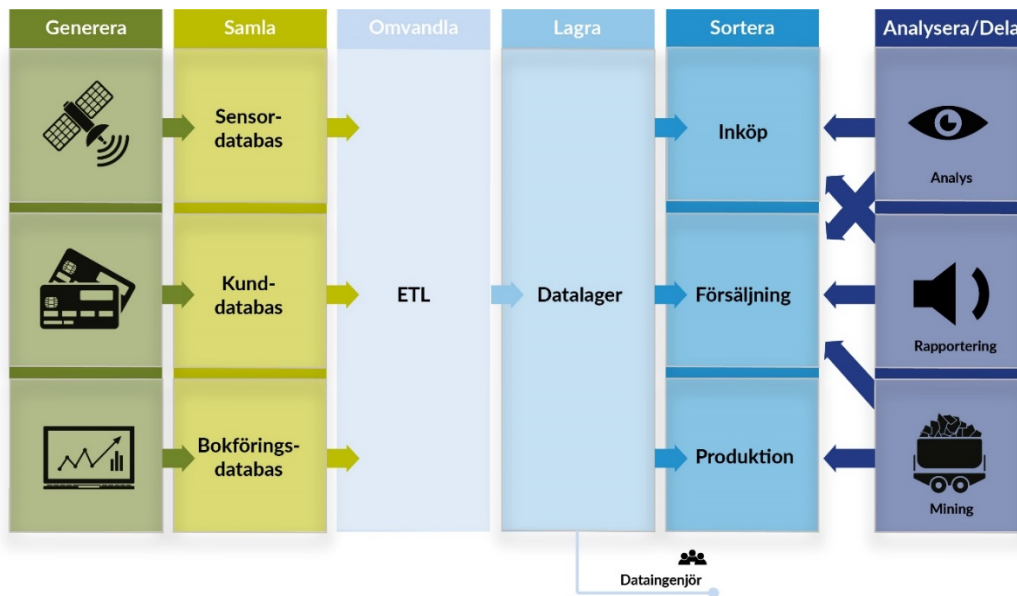


Bild 6 Schematisk bild av koppling mellan databas och datalager

Efter ETL-fasen placeras data i datalagret på en datahylla (specifika strukturer där data är redo att användas, på samma sätt som butikerna placerar produkten på hyllor till kunder) tillgängliga för användare genom verktyg, instrumentpaneler eller dylikt.

Tabell: För- och nackdelar med datalager

Fördelar	Nackdelar
<ul style="list-style-type: none"> Ger ett enda dataformat för alla data av intresse i en organisation Tillåter rapportering, visualisering och analys över mycket stora uppsättningar datakällor och strömmar Nästan varje anställd kan använda den Utmärkta sökmöjligheter Minskar datasiloeffekten Datamining för att hitta nytt mönster i data Säkerhet mot dataåtkomst 	<ul style="list-style-type: none"> Kräver mycket tid för att utforma ETL-scenen Datamodeller som är svåra att uppdatera eller ändra Jämförbart med enskilda databaser Tillgång till rådata saknas Dyrt underhåll Bearbeta rådata, ostrukturerade data eller komplexa data går långsamt

2.3.3 Datasjö

Eftersom datamängden ökar och så även dess komplexitet och behovet av hastighet, behövs mer lagrings- och processorkraft för att utföra dataanalys. En datasjö är en enda lagringsplats där alla data från en organisation kan lagras och bearbetas utan att struktureras först (dvs rådata). Uppgifterna delas och kopieras på ett datorkluster som tillåter varje maskin att analysera en del av data parallellt (bild 7). Detta ger viktiga vinster i prestation och gör det möjligt att lagra och bearbeta filer utan någon begränsning.

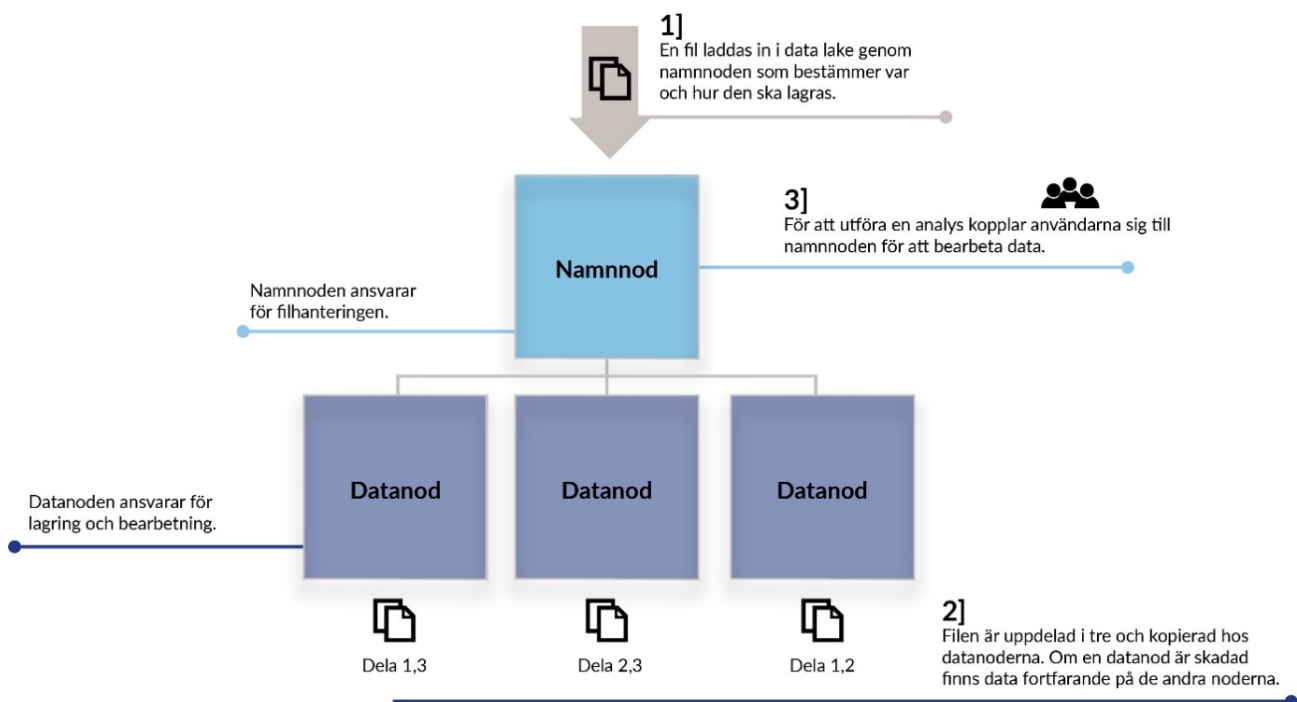


Bild 7 Arkitektur för en datasjö

Till skillnad från datalager kan datasjöar lagra rådata utan att gå igenom ETL-fasen, vilket innebär att rådata är tillgänglig för dataspecialister (bild 8). Detta ökar dramatiskt möjligheterna att hitta nya mönster genom att tillåta användning av algoritmer för artificiella intelligens.

Datasjöarnas arkitektur erbjuder fördelarna med att kombinera lagringsmöjligheter med datakraft till en relativt låg kostnad. Den är också mycket mer flexibel än ett datalager när det gäller vilken typ av data den kan ta emot och analysera och är lätt att uppgradera om mer lagring eller datakraft behövs.

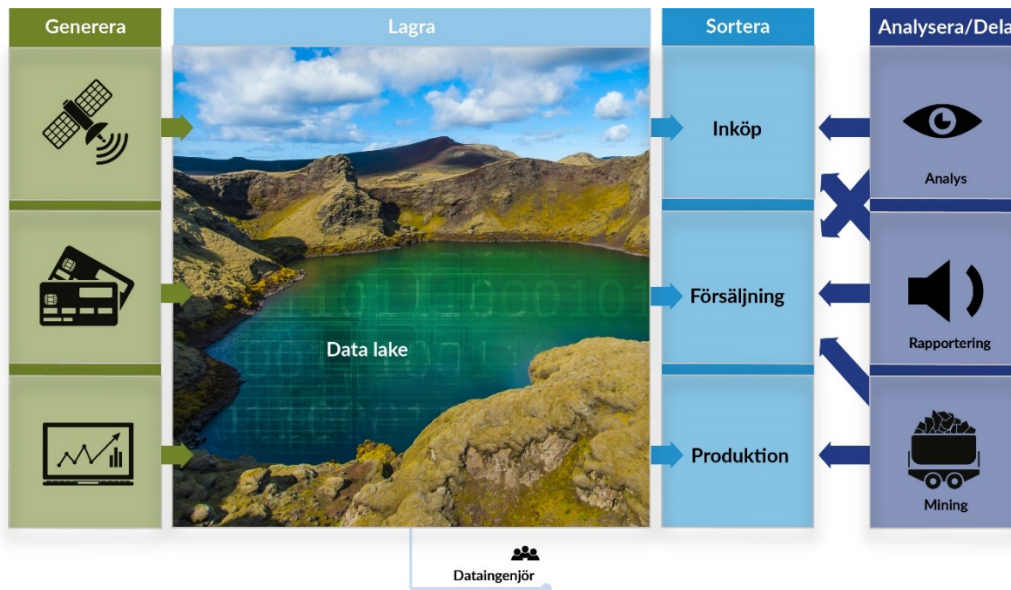


Bild 8 Schematisk bild av en datasjö

Ett av de största problemen med stora datamängder är dess volym och komplexitet. Datasjö-lösningen tar bort behovet av att flytta data från en plats till en annan för att utföra analys. Information kan analyseras även i realtid och det finns ett brett utbud av analytiska verktyg (visualisering, maskininlärning, metadataanalys etc.). En datasjö kan användas av flera användare samtidigt med olika säkerhets- och åtkomstnivåer.

Tabell: För- och nackdelar med en datasjö

Fördelar	Nackdelar
<ul style="list-style-type: none"> ■ Kan lagra en obegränsad mängd data ■ Billig hårdvara ■ Tillgång till rådata för att hitta nytt mönster ■ Minskar datasiloeffekten ■ Snabbt att uppdatera datamodell ■ Tillåter rapportering, visualisering och analys av mycket stor uppsättning datakällor och strömmar ■ Flexibel med vilken typ av data den kan lagra och mycket bra för att bearbeta dem alla (strukturerad, ostrukturerad, halvstrukturerad) 	<ul style="list-style-type: none"> ■ Kräver mycket tid att installera ■ Kan bli ett datatrask om denna inte underhålls korrekt ■ Kräver dataspecialist för att förbereda data för analys

2.4 Val av infrastruktur?

Som tidigare nämnts har databas, datalager och datasjö olika särdrag. De används allmänt för att lagra data men har olika syften. Vilken som ska väljas beror främst på typen av data och dess användning.

2.4.1 *Volym att analysera*

Datasjöar och datalager är gjorda för stora datamängder. Organisationer föredrar ofta att börja med databaser eftersom de är billigare än andra lösningar, kan behandla mycket information och är relativt lätta att förstå och navigera. Flytten från databaser till datalager eller datasjö kommer från behovet av att diversifiera antalet källor att analysera, den begränsade kapaciteten att utföra realtidsanalys och mängden data som ska lagras.

2.4.2 *Rådata eller strukturerade data*

Huvudskillnaden mellan en datasjö och ett datalager eller en databas är den typ av datastruktur de håller. Datasjöar lagrar rå obehandlad data medan datalager och databaser lagrar raffinerade och bearbetade data. Denna skillnad bör utgöra huvudargument till vilket infrastruktur man väljer.

Rådata

Eftersom en datasjö lagrar rådata kräver den mycket större lagringskapacitet än en databas eller ett datalager. Data som strömmar in i datasjön behöver inte ha något särskilt syfte, utan lagras för eventuell senare användning. Dess data är formbart och kan snabbt bearbetas för något specifikt syfte och är idealiskt för artificiell intelligens att hitta nya mönster i. Den är också mycket flexibel eftersom det inte finns någon fördefinierad struktur. Dock krävs att specialister förfinar uppgifterna. Om detta inte görs tillräckligt väl, finns risk att datasjön förvandlas till ett "datatrask" där det blir svårt att återfinna och analysera data.

Bearbetade data

Datalager och databaser lagrar endast bearbetad data som har ett visst syfte. Det innebär att all data i ett datalager används för ett fast mål. Detta bidrar till att spara lagringsutrymme och göra data lätt att förstå för de flesta anställda, men gör det svårt och dyrt att tilldela nytt syfte för data.

2.4.3 *Applikationer*

Databaser, datalager och datasjöar existerar ofta inom samma organisation. Datasjöar används i stor utsträckning för att bearbeta strukturerade och ostrukturerade data med maskininlärningsalgoritmer medan datalager används för analytiskt syfte.

Vanligen används datasjöar av sjukvårdsindustrin, eftersom mycket av uppgifterna är ostrukturerade (bilder från patientjournaler eller kliniska data) medan finansindustrin gynnas mest av datalager eftersom det mest handlar om strukturerade data.

2.5 Moln eller lokalt?

Att lagra och bearbeta data i en molnlösning blir alltmer populär bland eftersom det innebär många fördelar. Innan val av en molnlösning bör ett antal faktorer, som beskrivs nedan, beaktas.

2.5.1 *Implementering*

I en lokal miljö är resurser dedikerade till IT-infrastrukturen, underhåll och implementering av program. Företagen ansvarar för att upprätthålla lösningarna och hela dess process samt nödvändiga hårdvaruuppgraderingar och underhåll. De behöver kompetenser internt för att hantera allt, men det ger extra flexibilitet när det gäller vilken typ av programvara som används och kan utvecklas.

Med en molnlösning är molnleverantören värd för resurserna och företag kan få tillgång till dessa resurser så mycket som de vill och närhelst de vill. Programvarulösningarna är begränsade av mjukvaruleverantörsprogrammet, vilket möjliggör mindre flexibilitet men är ofta bättre uppdaterad.

2.5.2 *Kostnad*

I en lokal miljö måste organisationen initialt investera i en IT-infrastruktur. Organisationen ansvarar sedan för driften av systemet, såsom hårdvara, kraftförsörjning, lokal och underhåll.

Med en molnlösning betalar företag bara för de resurser som de använder, utan drifts- eller underhållskostnad.

2.5.3 *Kontroll*

I en lokal miljö behåller företagen alla sina uppgifter och har full kontroll över dem. I högreglerade industrier med extra reglering av privata data är detta av stor vikt.

Med en molnlösning ligger kontrollen av data hos molnleverantören. Även om tillgänglighet och säkerhet ligger i fokus hos leverantörer, finns ändå risk att till exempel inte nå sin data vid ett driftstopp.

2.5.4 *Säkerhet*

I en lokal miljö är företagen fullt ansvariga för säkerheten i sin IT-infrastruktur och data. Industrier med känsliga uppgifter väljer ofta lagring lokalt.

Med en molnlösning ligger säkerheten för data i molnleverantörens hand. Majoriteten av leverantören erbjuder krypteringsfunktioner och robusta brandväggar. Även om tillgänglighet och säkerhet är i fokus hos leverantörer, har det inträffat överträdelser.

Fler och fler företag använder hybridmoln, där känsliga data lagras lokalt medan övrig data lagras i molnet.

2.6 Datainfrastruktur - Sammanfattning

Vald datainfrastrukturen kommer att påverka förutsättningarna för vilken typ av dataanalys ett företag kan utföra. Det är viktigt att tänka igenom det och planera på lång sikt. Valet beror mest på vilken typ av data som kommer att hanteras och dess användning.

Datastruktur

Det finns tre typer av data: strukturerade (tabeller, databaser), halvstrukturerade (maskinloggar, e-postmetadata) och ostrukturerad (bilder, video, tweet). Ostrukturerade data uppgår idag till 80 % av genererad data.

Big data

Stora datamängder är när data inte längre kan hanteras av en enda dator. De kombinerar vanligtvis följande egenskaper: hög volym, hög hastighet och stor variation och kräver en särskild infrastruktur.

Datainfrastruktur

Databaser är den enklaste och billigaste infrastrukturen att bygga. Det gör det möjligt att lagra strukturerade data från maskiner, kunder eller transaktioner på ett effektivt sätt. Det är enkelt att söka information och de flesta är användarvänliga i sin utformning. Emellertid hanteras bara strukturerade data och det erbjuds begränsade möjligheter när det gäller dataintegration, eftersom det är svårt att ansluta databaser till varandra.

Datalager är en stor mängd datainfrastruktur som fungerar som ett enda förvar för alla data i ett företag. Liksom databaser tillåter det att lagra strukturerade data. Det är dyrt och tidskrävande att bygga men erbjuder mycket bra analysmöjligheter. Det är enkelt att fråga och användarvänligt.

Datasjö är en stor mängd datainfrastruktur som fungerar som ett enda förvar för alla data i ett företag. Till skillnad från databaser och datalager har den en mycket flexibel arkitektur som kan lagra data direkt från källan. Den kan lagra och bearbeta alla typer av data (strukturerad, halvstrukturerad och ostrukturerad), den erbjuder de bästa möjligheterna när det gäller datautvinning och maskininlärning och köras på billig hårdvara. En stor del av lagrad data lagras utan syfte och en data specialist behöver förbereda den innan den kan användas.

I lokal miljö eller en molntjänst

Valet mellan en molntjänst eller lokal lagring beror oftast på datas känslighet (säkerhet och kontroll), intern IT-kompetens och tillgänglighet av programvara. Idag går allt fler över till hybridmoln. Ett alternativ där känslig data lagras lokalt och övrig data i en molntjänst.

2.7 Artificiell intelligens, maskininläring, djupinläring

Ända sedan antiken har skapandet av maskiner som kan tänka, känna och agera varit föremål för vår fantasi. Men det var först 1950, då Alan Turing publicerade rapporten ”Computer Machine Intelligence”, som området började tas på allvar. Frågeställningen i rapporten var om maskiner kan tänka och en metod för att testa det föreslogs, det så kallade Turing-testet. Några år senare, 1956, användes termen ”Artificiell Intelligence” för första gången av den amerikanska datavetaren John McCarthy och detta blev början på en intensiv forskningsperiod inom området. Dagens möjligheter att hantera stora datamängder och dataanalyser har inneburit en snabb tillväxt för AI och möjliggjort användandet i industriella applikationer. AI är fortfarande långt ifrån att kunna uppträda som människor. Den AI som utvecklas idag är "smal AI" eller "svag AI" som endast kan imitera intelligent mänskligt beteende av mycket specialiserade och specifika uppgifter. Dessa inkluderar till exempel datorseende, NLP (naturlig språkbehandling), robotteknik eller maskininläring (ML) där datorer fungerar lika bra eller bättre än människor (bild 9).

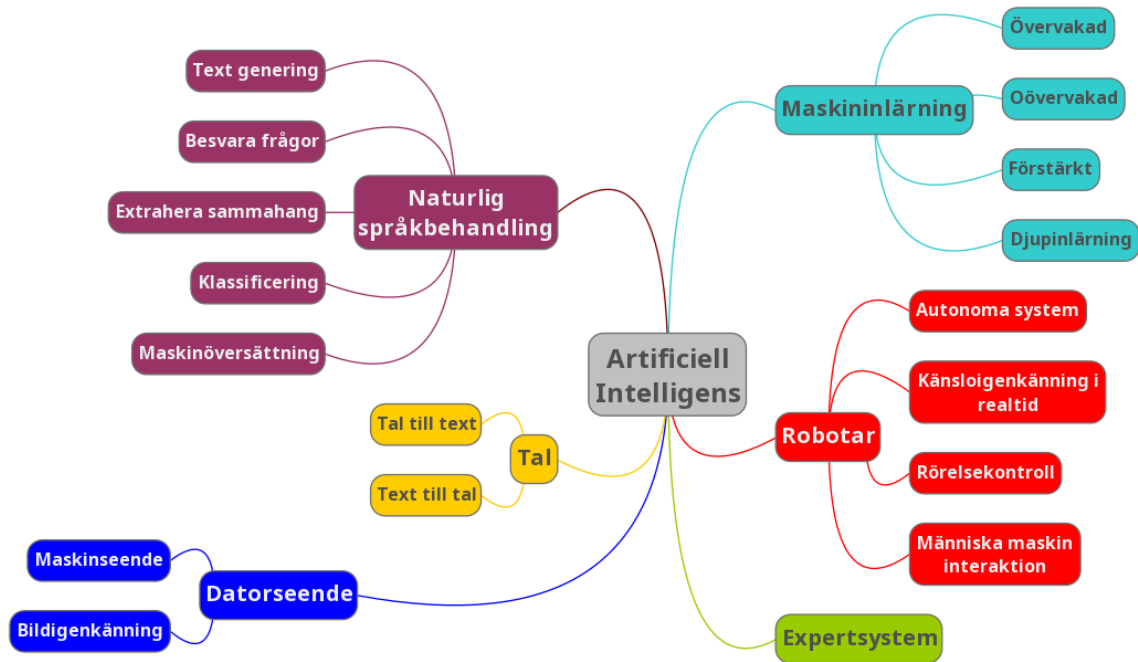


Bild 9 Områden i AI-forskning

2.7.1 Maskininlärning

Maskininlärning (ML) är en delmängd av AI och föddes ur teorin om att datorer kan lära sig utan att programmeras för att utföra specifika uppgifter. ML använder algoritmer för att analysera data och upptäcker dolda mönster mellan olika variabler. En maskin kan lära sig att göra förutsägelser på samma sätt som människan lär sig genom att observera världen runt sig och dra slutsatser utifrån tidigare erfarenheter. Ju mer erfarenhet, desto bättre förutsägelser. Denna förmåga kan användas för ett brett spektrum av uppgifter, såsom ansiktsgenkänning, medicinsk diagnostik, bedrägeribekämpning eller köprekommendationer online (S. Finlay, 2017).

Det finns många ML-modeller och algoritmer att använda för att hitta mönster i data. Den prediktionsmodell som följer av maskininlärningsprocessen används för att generera nya förutsägelser som organisationer ska fatta beslut om.

Inlärningsmetoden kan delas in i tre kategorier: övervakat lärande, oövervakat lärande och förstärkt lärande.

2.7.2 Övervakat lärande

Övervakad maskininlärning utgör för närvarande det mesta av ML som används över hela världen. Det är särskilt bra för klassificeringsproblem och regression (bild 10) som bedrägeri, förebyggande underhåll eller optimering av lager.

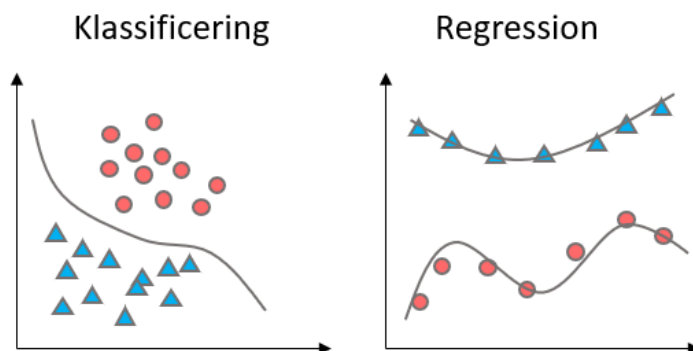


Bild 10 Övervakat lärande, hitta den bästa passformen mellan funktionerna (klassificering) eller den bästa passformen för varje funktion (regression)

Med denna träningsmetod ges en viss inlärningsalgoritm bestämda värden på ingångar, utgångar, och scenarier. En uppsättning träningsdata (ingångar) tillsammans med de förväntade resultaten (utgångar) tillhandahålls av en handledare. Algoritmen måste hitta relationen som länkar inmatningen till utgången och lär sig genom att utvärdera mot de redan kända resultaten. Beroende på felet som hittades mellan beräkning och det förväntade resultatet, modifierar den sin modell i enlighet med detta. Väl tränad kan ML-algoritmen användas för att göra förutsägelser och hjälpa till att ta affärsbeslut genom att analysera nya rådata (bild 11).

Ett framgångsrikt övervakat maskininlärningsprojekt bygger mycket på det data som används i träningen. Som med vilken statistisk metod som helst, är det viktigt att träningsdatasetet är opartiskt, innehåller tillräckligt med exempel på varje målutfall, har en tillräcklig mängd data och tillräcklig precision. Även om det är tidskrävande är det viktigt att prioritera att bygga en bra modell för maskininläringen(<https://cloud.google.com/blog/products/gcp/preparing-and-curating-your-data-for-machine-learning>).

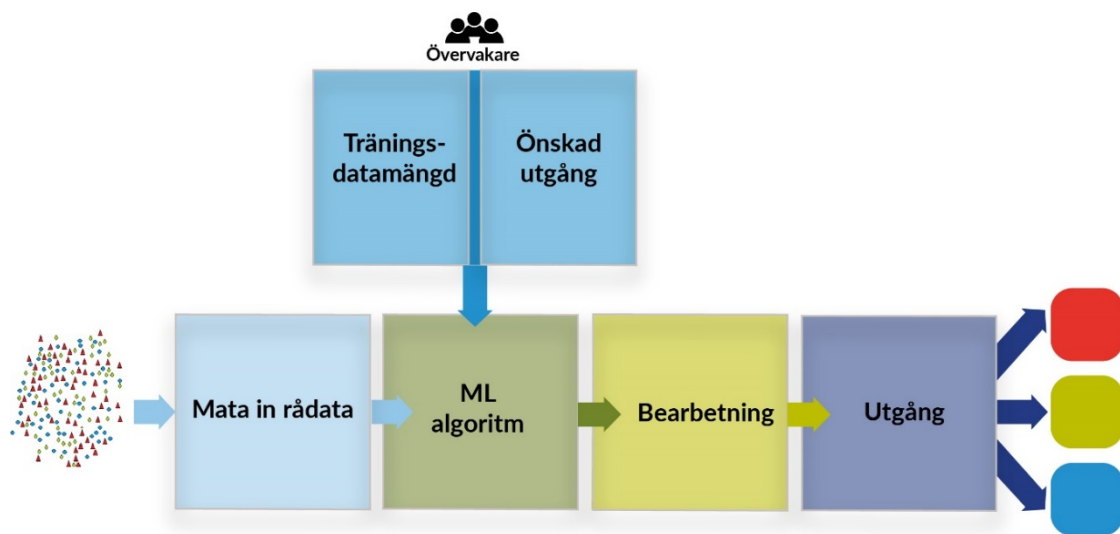


Bild 11 Övervakad maskininläring

I fallet med t ex bedrägeribekämpning måste data för träning innehålla exempel på varje enskilt fall (legitim eller bedräglig transaktion), vara representativ för hela populationen och ha tillräckligt med exempel för att modellen ska vara säker på sina förutsägelser.

För regressionsfall måste data för träning ha tillräckligt med historiska data och tillräcklig tidsprecision för förutsägelse. För att försöka förutse glassförsäljningen per timme under sommaren, skulle data innehålla minst tre somrar och en eller fler observationer per timme.

2.7.3 Oövervakat lärande

I oövervakat lärande finns inga bestämda ingångar och utgångar. Algoritmen måste själv hitta en lösning. Målet är att utforska data och hitta någon struktur i den (bild 12). Det fungerar bäst för gruppindelningar, till exempel att hitta de viktigaste attributen för kundsegmentering eller gruppering av aktiemarknader som uppför sig på liknande sätt.

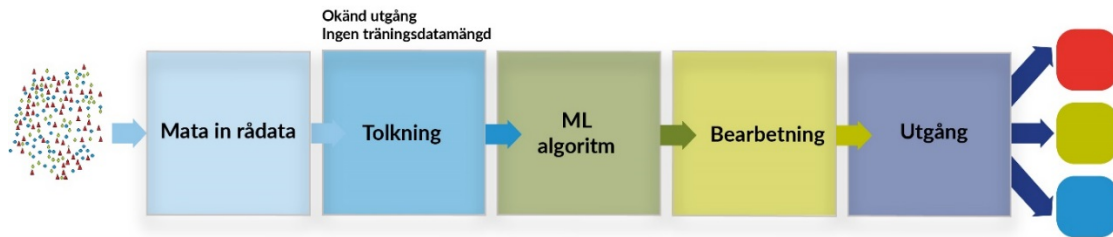


Bild 12 Övervakad maskininlärning

När det gäller kundsegmentering med hjälp av oövervakad inlärning, skulle ingående data kunna vara all inköphistorik från kunder utan att ge någon annan information (till exempel kön, ålder, plats). Målet för algoritmen är att endast skapa grupper av kunder beroende på deras inköphistorik. Detta kan exempelvis användas för att tillhandahålla online-rekommendationer till en viss kundgrupp. Ett annat exempel på gruppering är att skilja blommor från samma art uteslutande beroende på deras egenskaper (bild 13).

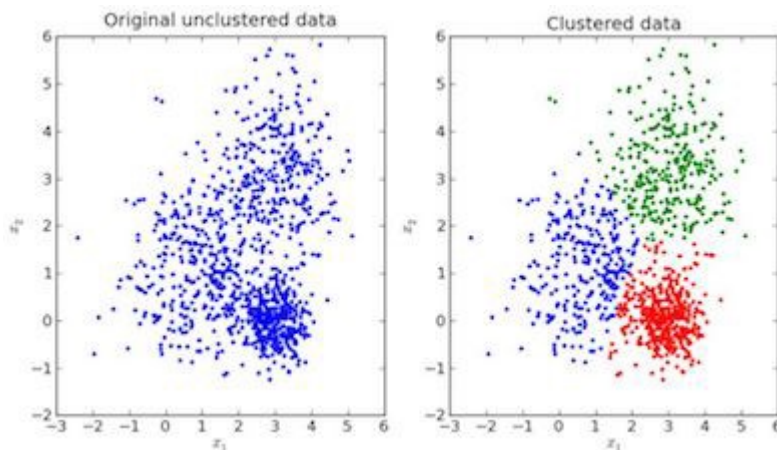


Bild 13 Övervakad maskininlärning av Iris-dataset (R. A. Fisher, *The use of multiple measurement in taxonomic problems*, 1936)

Oövervakad maskininlärning används fortfarande sällan men det är troligt att det kommer att öka under de kommande åren på grund av dess utforskande förmåga.

2.7.4 Förstärkt lärande

Förstärkt lärande handlar om att vidta lämpliga åtgärder för att maximera en belöning i en viss situation. Det används ofta för att låta en maskin lära sig en specifik uppgift som att gå, spela ett spel eller landa en raket. Till skillnad från övervakat eller oövervakat lärande finns det ingen dataset i träningen. Agenten (algoritmen) sätts i en speciell situation eller miljö med ett mål att nå. Den kan endast utföra en begränsad uppsättning åtgärder (muskelkontroll, göra godkända rörelser, raketkontroll) och få belöningar och återkoppling (avsluta en tävling, vinna spelet eller landa raketerna samtidigt som bränsleförbrukningen minimeras)

beroende på resultatet av dess åtgärder. På så sätt kan agenten lära av sina handlingar och förbättra sig (bild 14).

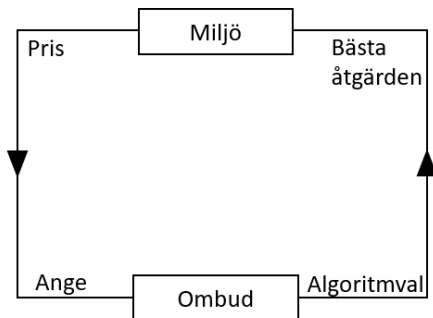


Bild 14 Förstärkt lärande

Denna typ av lärande används ofta i robotar, självkörande bilar eller för att utmana människor på vissa spel (Go (D. Silver et al., 2017), Chess (D. Silver et al., 2018), Starcraft II (<https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>)). Det yttersta målet är att skapa ett allmänt AI som kan lära sig att lösa alla slags problem på kort tid.

2.7.5 *Artificiellt neuralt nätverk och djupinlärning*

Artificiellt neuralt nätverk (ANN) och djupinlärning är två termer som ofta används inom AI. ANN är ett av de viktigaste verktygen som används vid maskininlärning. Djupinlärning är en delmängd av maskininlärning (bild 9) och använder omfattande ANN-arkitekturer för att utföra analys.

2.7.6 *Artificial Neural Network (ANN)*

ANN är ett ramverk för många olika maskininlärningsalgoritmer och efterliknar det biologiska neurala nätverket i en hjärna. Det kan användas med någon av de tidigare beskrivna inlärningsmetoderna och är kärnkomponenten för djupinlärning. Den ursprungliga tanken bakom ANN är att lösa problemet på samma sätt som mänskliga hjärnor gör.

ANN består av minst tre lager av neuroner där varje skikt använder utgången från föregående lager som ingång (bild 15). Liksom synapser i hjärnan, sänder kopplingen mellan artificiella neuroner en signal som kommer att behandlas av neuroner och vidarebefordras till nästa lager.

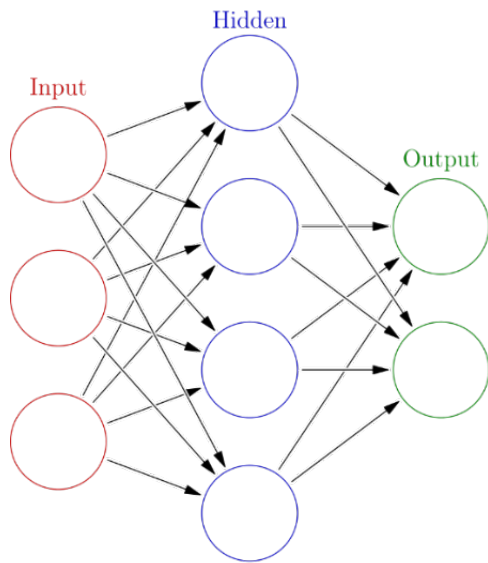


Bild 15 Artificiellt neuralt nätverk och dess neuroner som färgade cirklar

Varje artificiellt neuron beräknar signalerna som kommer från alla ingångar x med en matematisk funktion f och en särskild vikt w . Den summerar sedan alla resultat och skickar denna utgång y till nästa lager av neuroner om ett visst tröskelvärde (eller bias) θ har nåtts. När den nått utgångsläget jämförs resultatet med det förväntade resultatet och ett fel beräknas. Genom att justera vikterna och förspänna varje iteration kan nätverket minska detta fel och förbättra dess effektivitet (bild 16).

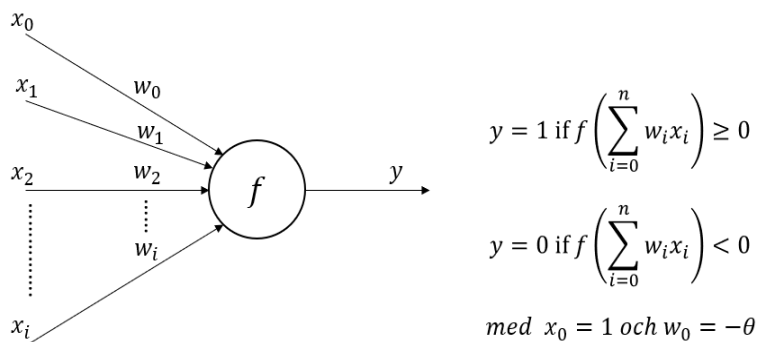


Bild 16 Artificiellt neuron

Nätverket kommer att få en djupare förståelse för data och dess komplexitet genom att lägga till lager och artificiella neuroner till den. Det kallas djupinlärning. Classic ANN har vanligtvis cirka två dolda lager av neuroner medan djupt lärande nätverk kan ha upp till 150.

2.7.7 *Djupinlärning*

Djupinlärning är en delmängd av maskininlärning, men den stora skillnaden mellan de två är att den djupa inlärningsalgoritmen kan bestämma sig själv om en förutsägelse är korrekt eller inte, medan maskinlärande når en plåtå och behöver därefter en ingenjör för att gå in och göra justeringar. Eftersom djupinlärning är en ANN, kan den bestämma sig om dess förutsägelser är korrekta eller inte och har överlag bättre förutsägelser än maskininlärning. Det är den teknik som kommer närmast mänsklig intelligens och används för olika uppgifter som dataseende, taligenkänning, social nätverksfiltrering, medicinsk diagnos, brädspel eller generering av ny design (<https://www.mathworks.com/discovery/deep-learning.html>).

Det enklaste sättet att bygga ett djupinlärande nätverk är att använda ett ramverk, en verktygslåda, som erbjuder alla byggstenar för att designa, träna och validera neurala nätverk. De mest populära är Tensor Flow, Pytorch, Caffe2, Matlab, Keras eller Chainer.

ANN och djupinlärning är beräkningsmässigt intensiva och kräver att minst en GPU(Graphic Processing Unit) används för att få resultat inom rimlig tid. Denna hårdvara är särskilt effektiv vid beräkning av matrismultiplicering och faltning, såsom ANN medan en CPU (Central Processing Unit) ger sämre resultat. En GPU tar bara några minuter att köra en djup inlärningsalgoritm medan en CPU tar några timmar (J. Lawrence, 2017). Det rekommenderas därför starkt att endast använda GPU för djupinlärande.

Förutom hårdvarukravet, är en stor nackdel med denna teknik brist på transparens kring inlärningsprocessen och vad precis det neurala nätverket gör för att få resultatet. Omfattande forskning pågår för att göra djupinlärning mer transparent så att ingenjörer kan vara säkra på vad det lär sig och logiken bakom förutsägelsen.

2.8 Artificiell intelligens - Sammanfattning

Artificiell intelligens och maskininlärning revolutionerar snabbt vårt samhälle tack vare den mängd data vi genererar. När fler och fler företag inför den i sin produktion eller i sina tjänster är det viktigt att förstå de konkurrens fördelar den ger.

Maskininlärning

Maskininlärning är olika metoder och algoritmer som analyserar data och hittar mönster i informationen. Den lär sig hur data beter sig, kan göra förutsägelser, klassificera data och kan lära sig nya förmågor. Det finns tre huvudmetoder för inlärningen:

Övervakat lärande: En operatör ger ingångsdata och de förväntade resultaten till en dator som försöker hitta hur man länkar ingångsdata till resultaten. När maskinen väl är utbildad kommer den kunna bedöma nya data korrekt. Det är den mest använda inlärningstekniken och är mycket effektiv för klassificerings- eller regressionsproblem såsom bedrägeribekämpning, trendanalys eller prediktivt underhåll.

Oövervakad inlärning: En operatör ger ingångsdata till maskinen men inga förväntade resultat eller mål. Datorn behöver själv hitta ett mönster i data. Det är särskilt användbart för datautforskning och gruppering, som till exempel att definiera olika konsumentbeteende.

Förstärkt lärande: Ingen ingångsdata används, utan maskinen är placerad i en speciell situation och kan utföra vissa åtgärder för att nå ett mål. Beroende på dess handlingar ges positiv eller negativ återkoppling. Metoden används i robotik för att lära en maskin hur man går, spelar ett spel eller kör en bil.

Artificiella neurala nätverk och djupinlärning

Artificiella neurala nätverk (ANN) efterliknar mänsklig intelligens genom att använda samma neurala struktur. Det består av lager av neuroner som sänder en signal vidare till nästa neuronlager, för att leverera ett resultat i det sista skiktet. Tack vare sin struktur kan de vara självlärande.

Djupinlärning är en delmängd av maskininlärning och använder alla tre ovan beskrivna inlärningsmetoderna. Den använder omfattande ANN och är den teknik som kommer närmast mänsklig intelligens. Djupinlärning används i självkörande bilar, utför realtidsöversättning, skapar musik eller målar, kan användas förprediktivt underhåll eller vid identitetskontroll med hjälp av ansiktsigenkänning.

ANN är beräkningsmässigt intensivt och kräver att särskild maskinvara (GPU) används för att få resultat inom rimlig tid.

3 Tillämpningar av stora datamängder för produktframtagningen

I följande kapitel kommer olika infallsvinklar på användning av stora datamängder i produktframtagningsprocessen att beskrivas.

Data som samlas in kan ge värdefull information om vår omvärld. Trender och förändrade beteenden hos befolkningen kan betyda att marknadens behov har eller kommer att förändras. Demografi, konjunktur, kultur, råvarutillgång och miljöfaktorer påverkar våra konsumtionsbeteenden. Förändringar kan påverka specifikationer och kalkyler för nuvarande och framtida produkter. Information kan även komma från produkter som är i användning. Ökade reklamationer, förändrat användande, förändrat beteende hos användare eller ett ökat användande i allmänhet av företagets produkter kan vara värdefull information för att styra både utveckling och produktion.

3.1 Data och analyser

Data som samlas in kan användas på olika sätt med olika ambitioner. Nedan beskrivs den principiella skillnaden mellan olika analyser som kan göras. Dessa är:

- Beskrivande analyser
- Diagnostiserande analyser
- Förutsägande analyser
- Föreskrivande analyser

3.1.1 *Beskrivande analyser*

Beskrivande analyser ger en bild av något som har hänt i det förflutna. Exempelvis kan en beskrivande analys svara på frågan hur många reklamationer har vi fått in under den senaste månaden och på vilka av företagets produkter. Data kan exempelvis visas som ett stapeldiagram där antalet reklamationer visas som staplar per dag eller vecka.

3.1.2 *Diagnostiserande analyser*

Denna typ av analys försöker svara på orsaken till att något hänt i förfluten tid. Ett exempel kan vara att svara på varför antalet reklamationer har ökat eller minskat. Diagnostiserande analyser kräver ofta sammanställning av data från flera olika källor för att hitta samband. Verktygen för diagnostiserande analyser erbjuder ofta en interaktiv dialog med användaren där fokus kan läggas på olika delar av sammanställda data. De frågor som ställs är ofta mer komplicerade jämfört med beskrivande analyser. Målet är ofta att hitta mönster, samband och trender som kan peka ut orsaken till att något har hänt.

3.1.3 Förutsägande analyser

Förutsägande analyser pekar ut ett troligt utfall från ett tänkt framtida läge. Förutsägelsen baserar sig dock på historiska data. Ett exempel kan vara att förutsäga risken för att få reklamationer på en produkt utifrån några typiska parametrar som beskriver produkten och dess användning. Genom att samla in data från alla reklamationer som kommit in till företaget samt annan data som kan ha haft en påverkan på om produkten reklamerats, kan samband och mönster hittas. Denna typ av analyser kräver att man har avancerade verktyg för analysen som kan hitta mönster och samband i historiska data och använda dessa mönster och samband för att bedöma utfallet i en ny kombination av dessa data. Det är viktigt att förstå att det kan finnas ytterligare samband som härrör från data som inte samlats in. Ändras förhållanden som påverkar dessa data så kan tillförlitligheten i förutsägelseerna minska. Ett sätt att skapa modeller för att göra förutsägelser är att använda maskinlärande algoritmer som tränas med historiska data. Modellerna används sedan med ”nya” data och kan då förutsäga ett utfall inom modellens duglighetsområde. Förutsägelsen kan vara av olika typ. Den kan ange en sannolikhet att en reklamation ska inträffa eller vara ett absolut belopp som hur många reklamationer vi kommer att få.

3.1.4 Föreskrivande analyser

Föreskrivande analyser pekar ut en åtgärd som bör genomföras. Föreskrivande analyser baserar sig ofta på förutsägelser men kompletteras med en trolig handling för att till exempel minska antalet reklamationer. Föreskrivande analyser har ofta ett högt värde för användaren men kräver också mest kunskap för att göra.

3.2 Stora datamängder i produktion

I många industriella sammanhang behöver man fatta beslut som berör både människor och maskiner. Vi vill inte skada oss själva eller andra. Arbetsskador och skador av bristfälliga produkter kan skapa stort lidande och dessutom vara kostsam. På samma sätt är det med maskiner och de produkter som maskinerna tillverkar. Är de bristfälliga kan det bli kostsam och leda till stor skada. I alla beslut finns det en risk att beslutet kan vara fel och leda till just den typ av skada som beskrivs ovan. Risken uppstår eftersom det finns en osäkerhet som i sin tur beror på att något är okänt. Vi saknar information eller kunskap eller ibland båda delarna. Vi kan naturligtvis försöka att uppskatta risken och arbeta med planer för att mildra risken. Detta leder till att risken fortfarande finns men att vi har ett sätt att förhålla oss till den. Ett annat sätt att se på saken är att undanröja den osäkerhet som finns genom att inhämta mer information och kunskap om det som orsakar risken och med den informationen och kunskapen undanröja risken.

Om det föreligger en risk att en utrustning i en produktionsmiljö fungerar dåligt på grund av vibrationer så kan man uppskatta hur stor denna risk är och även hur stor risken är att det uppstår vibrationer som är skadliga. Ett annat sätt är att kontinuerligt mäta vibrationerna som maskinen utsätts för och samtidigt mäta produktionsutfallet för att se om det finns vibrationer och om det ser ut att påverka

produktionsutfallet. Här är det viktigt att komma ihåg att kausalitet och korrelation inte alltid följs åt. Hittar man korrelationen så kan man dock anta att kausaliteten finns och söka vidare får att få den kunskap som behövs för att förstå risken och eventuellt undanröja den stället för att hantera den.

Med hjälp av en kombination av kunskap och information kan vi fatta beslut baserat på fakta. Utan fakta så fattar vi beslut på magkänsla, gissningar och åsikter och hoppas att vi får rätt (Bjerring 2019).

Den tekniska utvecklingen går allt snabbare och kapaciteten i datorer och nätverk har historiskt sett ökat hela tiden, samtidigt som priset på hårdvara sjunker. Förutsättningarna för att samla in och lagra data i stora mängder blir allt bättre. Därmed ökar möjligheterna att få mer och bättre informationsunderlag för både kunskapsutveckling och för att fatta bättre beslut.

3.2.1 Vad vill vi fatta beslut om – vad vill vi förstå?

Vi vill förstå det som har hänt och fatta beslut om framtiden. Vi vill ofta styra produktionen mot uppsatta mål utifrån ett nuläge. Vi måste veta var vi är för att staka ut vägen framåt. I begreppet nuläge måste vi rimligtvis även inkludera det som hänt i närtid, och längre tillbaka i tiden om det kan anses kunna vara en del av nuläget. Känner vi inte nuläget så vet vi inte effekten av våra åtgärder. Det vill säga att vi kan se att en åtgärd har lett till en förändring men vi vet inte fullt ut förändringens innebörd.

Vi vill förstå nuläge och uppkomna förändringar i parametrar som påverkar:

- Kvalitet/ kassationer
- Störningar
- Ledtider
- Leveransförmåga
- Effektivitet
- Säkerhet
- Inre och yttre miljö

Vi vill fatta beslut om bland annat:

- Långsiktiga förbättringar
- Kortsiktiga förbättringar
- Investeringar
- Produktionsvolymer

Vi vill förstå effekten av gjorda åtgärder utifrån en faktabaserad bild av världen omkring oss samt se trender som kan förvarna om något. Andra saker som kan vara intressanta att använda intelligenta system till kan vara att:

- Kontrollera utformning och montering av komponenter
- Förutse haverier
- Förutse köbildningar
- Optimera produktionen

-
- Planera förebyggande underhåll
 - Bygga modeller för planering, prognostisering och produktionsutfall

Det är också väsentligt att poängtera att nyttan med stora datamängder för arbete med produktion inte enbart handlar om operativa driften av produktion, utan att ge underlag och förståelse för beslut över hela produktionsutrustningarnas/-systemets livscykel från anskaffning, utformning, utveckling, installation, drift, förändring, och avveckling av produktion. Dessutom kan man resonera utifrån vem som har nytta av dessa datamängder och då urskilja:

- olika organisationer med olika roller: olika tillverkare, eventuellt i flera led i en värdekedja, leverantörer av produktionsutrustning och produktionssystem samt kunder och kunders kunder
- olika roller internt hos dessa, till exempel hos tillverkande företag kan finnas produktutveckling, beredning, produktion, underhåll, logistik, försäljning.

En regel är att beslut ska fattas på lägsta möjliga nivå i en organisation för att vara så nära källan och effekten av beslutet. Den som fattar beslutet ska vara nära den verklighet som besluten berör och så direkt som möjligt se effekten av beslutet. Här är tron att det finns en analogi att närhet till data, information och observation är viktigt för beslutens kvalitet. Denna närhet antas vidare kunna delvis skapas genom att samla in och sammanställa data till information och sprida denna till alla behövande på företaget. På detta vis antas beslut på alla nivåer bli mer välgrundade och effekten bli synlig i de dataströmmar som samlas in efter en åtgärd som följer på beslutet. Här kan man fundera på skillnaden mellan data och information. En skillnad kan vara att data saknar sammanhang för den som beskådar data medan information är data i ett sammanhang. En text på ett totalt främmande språk kan databehandlas av oss. Vi kan räkna antalet förekomster av olika bokstäver, räkna antalet ord och räkna ut medellängden på orden, men det är fortfarande bara data. Om vi får motsvarande text på ett språk vi förstår så blir det information förutsatt att texten är meningsfull.

3.2.2 Tolka och agera på data

När vi som människor ska betrakta ett komplext system, som en produktionsmiljö ofta är, så betraktar vi något där antalet parametrar ofta är större än antalet mätpunkter. Vi samlar förmodligen in fakta men långt mindre än vad som behövs för att ge en entydig bild av det vi ser. Vi försöker ändå tolka bilden och fatta beslut utifrån våra tolkningar. Vi gör detta med den hastighet vi förmår som människor. Modern teknik för att mäta, som inkluderar flera typer av sensorer, kameror som filmar och tar stillbilder, mätningar av variationen i strömförbrukning i olika utrustningar, manuella inmatningar i olika system mm, kan ge en enorm mängd data som samlas in med höghastighet. Mängden data och mängden parametrar som mäts blir snabbt oöverskådlig för oss människor. Vi behöver ta hjälp av datorer för att hinna registrera, lagra, sortera och tolka data. Datorer kan på ett uthålligt sätt bygga modeller av data som till stor del motsvarar de variationer som finns. Genom att använda artificiell intelligens kan data tolkas,

mönster och trender kan ses och datorn kan agera utifrån detta betydligt snabbare än människan agerar med korrigerande åtgärder om så behövs. Även här blir närheten till mätpunkterna och systemet som mäts av betydelse. Lokala datorer kan samla in data lokalt och med hjälp av datorbaserade modeller skicka ut korrigerande signaler i realtid. Data kan också samlas från flera produktionsmaskiner i större mängder under längre tid för analys och åtgärder som kan tillåta eller kräver en längre tidcykel.

3.2.3 *Det komplexa tekniska systemet*

Produktionssystem blir allt mer komplexa att utveckla, installera, driva, felsöka och förändra. Detta drivs av

- Teknisk utveckling, mycket beroende på digitalisering och möjligheter att koppla samman system.
 - Mekaniska system blir integrerade system med mekanik, elektronik, datorsystem.
 - Isolerade system blir integrerade med andra maskiner, datorsystem, m m.
- Hårda krav på verksamheten att snabbt och billigt leverera produkter med hög kvalitet och motsvarande krav på produktionssystem.
 - Detta ökar behov av snabbhet i organisation och integration av delsystem, samtidigt som marginalerna kring personal, material, ekonomi generellt sett minskar.

Att produktionssystemen blir allt mer uppkopplade kan tolkas på olika sätt, allt från att en enhet har en egen adress på Internet (Internet of things) till att enheten kan ge och ta emot data via ett lokalt nätverk. En effekt av att enheter blir mer uppkopplade är att vi kan förse dem med sensorer för att läsa av mätvärden från enheten och ge signaler till manöverorgan som kan ställa om något i enheten till exempel spänning, ström, läge eller något annat som kan styras via signaler. Mätvärden från sensorerna kan tas upp som data genom att läsa av dem med viss periodicitet. Här gör det faktum att enheten är uppkopplad att data kan samlas på en dator (server) någonstans i det nätverk som enheten tillhör. Här finns ett val att göra om data ska lagras lokalt hos enheten eller lagras på en central server.

Faktaruta: Internet of things

Sakernas internet (från engelskans Internet of Things, IoT) är vardagsföremål som hushållsapparater, kläder och accessoarer, men även maskiner, fordon och byggnader, med inbyggd elektronik och internetuppkoppling, vilket gör att de kan styras eller utbyta data över nätet. Nätverket gör att sakerna kan kontrolleras och delge information från andra platser vilket underlättar för den fysiska världen att integreras närmare med datasystem och kan resultera i högre effektivitet, exakthet och ekonomiska fördelar. När sakernas internet innehåller sensorer och ställdon blir det ett exempel på cyberfysiska system, dit också intelligenta hus, smarta elnät, digitala assistenter och intelligenta transportsystem hör. (Wikipedia – 2)

Sensorer, manöverorgan, datorer och det sammanlänkande nätverket utgör kärnan i ett uppkopplat system. Till nätverket räkas nätverksnav av olika typer.

Uppkopplingen av enheter i produktionsapparaten kan vara trådbunden eller trådlös. Detta kan påverka pålitligheten hos uppkopplingen vilket man bör tänka på. Ofta är trådbunden uppkoppling pålitligare än trådlös uppkoppling.

Data måste lagras så att dess sammanhang dokumenteras. Om man mäter ett mätvärde i en utrustning så måste man veta vilken utrustning det var, när mätningen gjordes, vilken typ av produktion som kördes mm. Utan sammanhang så minskar det informativa värdet av data.

3.2.4 Arkitektur på det uppkopplade systemet

Ett uppkopplat system kan utformas med en arkitektur som innehåller olika nivåer. Här kan närheten till mätpunkterna och systemet som mäts vara av betydelse. Lokala datorer kan samla in data och lagra och behandla dessa lokalt. Data kan också samlas från flera produktionsmaskiner i större mängder under längre tid för analys och åtgärder där data lagras och behandlas centralt. Man kan med andra ord säga att det kan finnas behov av lokala snabba system för respons i realtid samt övergripande system där analyser och åtgärder görs utifrån mer data och sammanställda data. De olika nivåerna är ofta förbundna med företagets lokala nätverk.

Nedan följer en beskrivning på arkitektur för uppkopplade system (Sethi and Sarangi 2017).

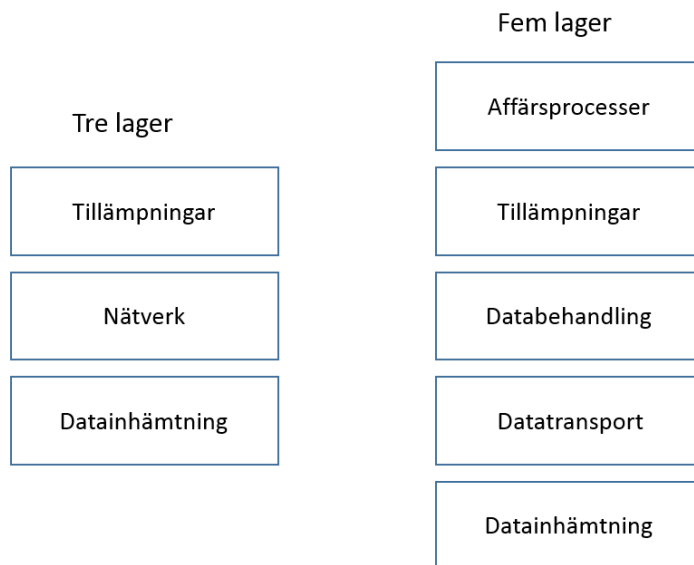


Bild 17 Här visas de lager som man brukar dela in ett företags IT-system

Arkitekturen kan beskrivas antingen i tre lager:

- Tillämpningar
- Nätverk
- Datainhämtning

eller i termer av fem lager

- Affärsprocesser
- Tillämpningar
- Databehandling
- Datatransport
- Datainhämtning

Lagren i trelagersarkitekturen beskrivs nedan:

3.2.5 *Datainhämtning*

Detta lager innehåller sensorer som samlar data från en omgivning. I detta lager mäter vi intressanta parametrar eller identifierar objekt som finns i omgivningen. Sensorerna läses av och sparas som data.

3.2.6 *Nätverk*

Nätverket kopplar samman alla enheter i ett system med serverdatorer. Detta lager kan även innehålla funktionalitet för att sända och databehandla sensordata.

3.2.7 *Tillämpningar*

Detta lager innehåller tillämpningar (datorprogram) som förser användarna med viss funktionalitet. Det kan vara ett MPS-system, system för att hantera avvikelser, system för dataanalys eller annat system som utför någon funktion för en användare.

Denna beskrivning är ofta väl grov och för att ytterligare fånga aspekter på det uppkopplade systemet så används ibland den ovan nämnda beskrivningen med fem lager. De ytterligare lagren förutom ”datainhämtning” och ”tillämpningar” beskrivs nedan.

3.2.8 *Transport*

Detta lager förmedlar data från lagret för datainhämtning till lagret för databehandling och tvärt om. Här används olika typer av medel för detta. Exempel är RFID och NFC för mycket korta avstånd, Bluetooth för avstånd upp till runt tio meter, wifi och zigbee för medellånga avstånd och trådbundna LAN för längre avstånd.

Faktaruta RFID

RFID står för Radio frequency identification och är en teknik för att överföra information på avstånd från transpondrar och minnen. Dessa kallas för taggar och innehåller oftast information om det objekt de sitter på. De enklaste varianterna av RFID har en påtagligt enkel uppbyggnad och de består endast av ett unikt nummer som de kan sända ut några decimeter. Detta är den oftast förekommande typen som används idag. Med denna typ av RFID-transponder har man all information lagrad i en separat databas. Posten där informationen finns lagrad är bunden till det unika id-nummer som taggen innehåller. Denna enklare typ motsvarar i praktiken vanliga streckkoder. Nästa prisnivå har taggar som är mer avancerade och har inbyggt minne som det går att skriva till flera gånger, dock är minnet ganska begränsat. Avläsaren består av ett oscillerande magnetfält som inducerar en tillräcklig spänning i antennen för att taggen ska kunna sända sitt innehåll. Taggen kan vara liten nog för att passa in i en vanlig prislapp, sätts in under huden på ett djur eller människa för identifiering med hjälp av radiovågor. (Wikipedia- 3)

Faktaruta: NFC

NFC står för Near Field Communication och är en metod för kontaktlös överföring av data över korta sträckor. Namnet kommer av att antennens längd avsiktligt anpassas så att den är en multipel (1/2, 1/4, 1/8 osv) av våglängden och därför inte tillåter en stående våg, vilket begränsar räckvidden till området nära (near field) antennen som vanligtvis är ca 10 cm, och därigenom blir ytterst svår att avlyssna på avstånd. Syftet med tekniken är att skapa en säker, intuitiv och enkel kommunikationskanal mellan olika elektroniska enheter, primärt inriktat mot mobiltelefoner. Exempelvis kan en mobiltelefon med inbyggt NFC användas som betalmedel, populärt kallat "mobil plånbok", i en butik eller kollektivtrafik. (Wikipedia – 4)

Faktaruta Bluetooth

Bluetooth LE (engelska för blåttand) är en standard för trådlös, kortväga överföring mellan till exempel headset och mobiltelefon eller mellan tangentbord och dator. Bluetooth 1.0 släpptes år 1999. År 2010 kom det första stödet för lågeffektvarianten Bluetooth LE, som snabbt blev populär. Bluetooth 5 som släpptes år 2016 har fokus på låg effekt. Bluetooth LE är inte bakåtkompatibel med tidigare versioner av Bluetooth. Bluetooth 4.0 tillåter implementation av både klassisk Bluetooth och Bluetooth LE som kan använda samma hårdvara och samma radiofrekvens på 2,4 GHz. (Wikipedia -5)

Faktaruta Zigbee

Zigbee är en standard för trådlös styrning och monitorering av utrustning i till exempel hem, fastigheter och industrier – i miljöer där behov finns. Plattformen har stöd för olika typer av förgreningsstruktur. Tekniken är utvecklad för att vara energisnål. Zigbee bygger på radiostandarden IEEE 802.15.4. Egenskaper som framhålls är: energisnål, trådlös, återkoppling att kommandon har utförts, låg investeringströskel för att komma igång i liten skala samt enkel installation. Många stora teknikföretag uppger stödja denna standard. (Wikipedia – 6)

3.2.9 Databehandling

Detta lager är ett mellanlager som samlar, lagrar och processar data. Detta lager kan förse angränsande lager med olika typer av funktioner. Här används databaser, molntjänster och databehandlingsmetoder hämtade från tekniker med stora datamängder.

3.2.10 Affärsprocesser

På denna nivå finns tillämpningar, säkerhetssystem samt system för att hantera hela det uppkopplade systemet.

3.2.11 Processer över och under transportlagret

De beskrivna arkitekturerna sträcker sig från datainhämtning som ofta görs nära en fysisk verklighet med produktionsmaskiner och produkter hela vägen upp till en användare som använder verktyg på en högre nivå som sammanställer data, analyserar data och kan användas för att exempelvis sända ut produktionsorder till produktionsapparaten. Det kan dock finnas behov av att agera snabbt nära de uppkopplade enheterna i produktionsmiljön. I dessa fall kan man förstärka de lägre nivåerna under lagret för datatransport med funktionalitet för lagring och analys av data med möjligheter att direkt ge styrsignaler till manöverorgan utan att passera upp till högre nivåer ovanför lagret för datatransport. För att skilja på processer som sker under lagret för datatransport och processer som sker över detta lager så talar man om "fog computing" (eller dimman på svenska, se bild 18) under transportlagret och "cloud computing" (eller molnet på svenska) ovanför transportlagret. Processer som sker i "dimman" det vill säga nära produktionsutrustningen är de som kräver snabba svarstider och kan hanteras med begränsade resurser som finns under transportlagret. Ibland berikas de undre lagren med mer datorkraft i form av smarta sensorer, datorer och smarta nätverksnav. Smarta nätverksnav kan sköta lokala processer som att samla, spara och analysera data lokalt och ge respons till lokala manöverdon i realtid. Processer som sker ovanför transportlagret är de som kräver mer resurser och som inte är lika tidskritiska.

I de fall man berikar de lägra lagren så kan detta innebära att man använder sig av produktionsmaskiner med egen datorkapacitet och artificiell intelligens.

3.2.12 Sensorer och manöverorgan

I den så kallade dimman finns sensorer och manöverorgan. Sensorer kan vara av en rad olika typer. De kan mäta allt från ström, spänning, ljus och ljud till acceleration och infraröd strålning. De kan även vara RFID-sensorer som är sensorer som märker upp objekt och kan läsas av beröringsfritt. Andra mer avancerade sensorer är till exempel kameror, GPS-sensorer, magnetometrar, kemiska sensorer och sensorer som placeras på människokroppen. Manöverorgan kan vara hydrauliska, pneumatiska och elektriska ställdon. Även motorer, elektromagneter, sugklockor, bromsar och signallampor kan räknas hit. Sensorerna mäter storheter i en omgivning som blir till signaler för behandling,

analys och eventuellt beslut. Manöverorganen omsätter besluten i form av styrsignaler till fysisk påverkan på produktionssystemet.

3.2.13 *Kommunikation för trådlösa sensorer*

Trådlösa sensorer har ofta begränsad tillgång till elektrisk kraft och minne och behöver därför bland annat enklare tekniker för kommunikation jämfört med traditionella trådlösa nätverk som använder IP (Internet Protocole). Det finns en rad alternativ för dessa enheter. Några exempel är IEEE 802.15, 6LoWPAN, RFID, NFC, SigFox och LoraWan.

Faktaruta IEEE 802.15.4

IEEE 802.15.4 är en standard för trådlösa så kallade personnära nätverk med låg hastighet. Trådlöst personnära nätverk (WPAN = Wire less Personal Area network) är skapade för trådlös energisparande kommunikation inom ett personnära arbetsområde. Det personnära nätverket kan vara en brygga i kommunikationen med mer vidsträckt nätverk på "högre nivå". Standarden definierar det fysiska lagret och kräver att någon definierar de överliggande lagren. Standarden är basen för många enklare och energisnåla varianter av nätverkskommunikation som exempelvis Zigbee. (Wikipedia – 7)

Faktaruta 6LoWPAN

6LoWPAN står för energisparande nätverkskommunikation i personnära nätverk enligt internetprotokollet version 6 (IPv6). IPv6 är bland annat skapat för att lösa bristen på internetadresser som tidigare standarder inte hanterar. IPv6 är utformat för att datapaket kan sändas i nätverk som baseras på IEEE 802.15.4. 6LoWPAN är tänkt att det ska möjliggöra för enheter med begränsningar, när det gäller energitillgång och processorstorlek, att vara en del av "sakernas Internet". (Wikipedia – 8)

Faktaruta Sigfox

Sigfox är en nätverksoperatör som sätter samman trådlösa nätverk för att ansluta objekt med låg effekt som elmätare och smarta klockor som måste vara påslagna över dygnet och hela tiden avge små mängder data. (/Wikipedia- 9)

Faktaruta LoRaWAN

LoRaWAN är utvecklat för att definiera de högre lagren i ett nätverk och finns i huvudsak till för att hantera kommunikation mellan olika personnära nätverk och enheter i noder på dessa. Kommunikation från lokala enheter skickas till en central nod som bland annat gör skärhetskroller och styr över nätverket. (Wikipedia – 10)

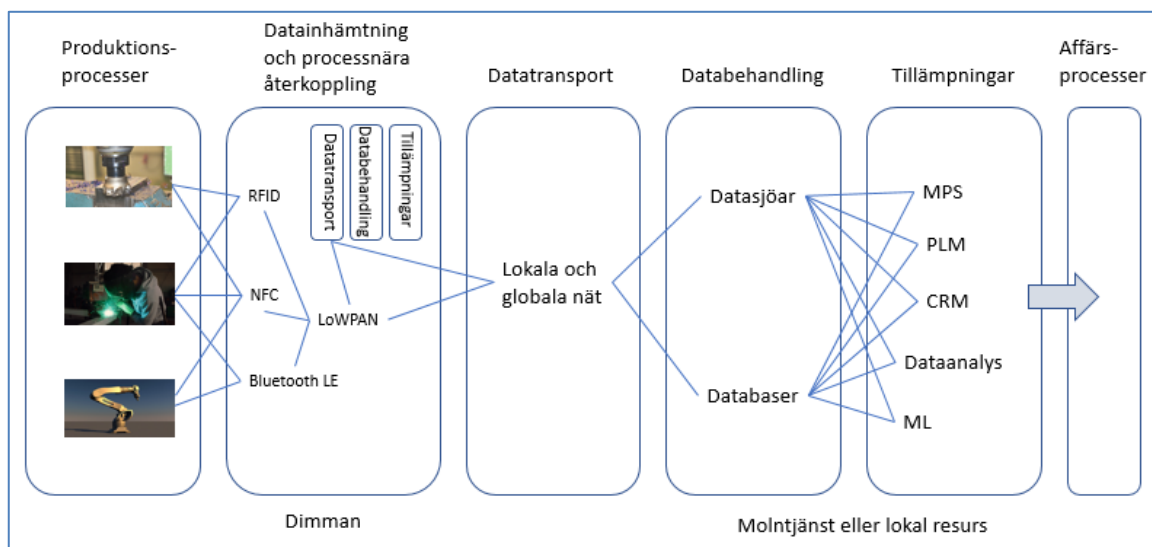


Bild 18 Principiell skiss över de områden som behandlas i texten. Jämför med bild 17

3.2.14 Integrationsplattformar

Det finns kommersiella integrationsplattformar som har en ambition att koppla ett helhetsgrepp på användningen av data i industriell miljö. Tydliga exempel finns från leverantörer som erbjuder automationsutrustning som industrirobotar och PLC (maskinnära digital styrutrustning) samt tillämpningsprogram som CAD, CAM och olika typer av analysprogram. En integrationsplattform av denna typ kan täcka in flera av de nivåer som visas i bild 17. Integrationsplattformen kan innehålla funktioner för att koppla upp utrustningar, hantera korta realtidsloopar i maskinnära kommunikation och samtidigt sända data till molntjänster för vidare analys och kombinerande med annan data. Det finns även mindre integrationsprogram som täcker in några av dessa nivåer. Exempel finns på integrationsprogram som täcker in de lägre nivåerna under transportlagret. Marknadssituationen för denna typ av mjukvaruprodukter kan vara osäker och det är därför klokt att göra en noggrann marknadsanalys innan man fattar beslut att använda en viss plattform. Många standarder för kommunikation, bristande kompatibilitet och tredjepartsprodukter i plattformarna kan påföra begränsningar i både interna och externa informationsflöden.

3.2.15 IT-landskapet

Den miljö där data samlas in kan bestå av många olika tillämpningar av IT-system. Ofta är ett produktionssystem som finns på ett företag ett pussel av olika separata informationssystem. Exempelvis kan man ha ett MPS-system, ett tidsregistreringssystem, system för avvikelshantering, system för programmering av utrustningar och hantering av programmen, webbaserade handelssystem m m.

System har olika sätt att lagra data och erbjuder olika och varierande möjligheter att läsa in data och exportera data. Bristande kompatibilitet kan begränsa effektiviteten i arbetet eftersom det kan vara svårt att göra analyser på den totala mängden data. Ska data flyttas med manuella metoder så kan det ge inmatningsfel som kan sänka kvaliteten på data. När en ny källa till data ska tas i bruk så kan det innebära att det skapas ytterligare ett system som är svårt att integrera med övriga system.

Detta gör att det behövs insikter i problematiken att integrera olika system och det är viktigt att detta inte förbises.

3.2.16 Datamodeller

Data måste lagras så att dess sammanhang dokumenteras. Om man mäter strömmen i en motor så måste man kunna lagra identiteten på motorn, när i tiden mätningen gjordes, vilken typ av produktion som kördes m m. Utan sammanhang så minskar det informativa värdet av data. Data kan lagras lokalt nära en uppkopplad enhet eller centralt på en server. När man lägger upp en strategi för att samla data är det klokt att fundera på hur data hänger ihop. Loggade data från en maskin kan relateras till annan data. Om man loggar till exempel vibrationer i en maskin så kan det finnas skäl att relatera dessa data till liknade data i maskinens omgivning. Vi måste då ha ett gemensamt begrepp i den ena datamängden som kan relateras till ett motsvarande begrepp i den andra datamängden. Ett sådant begrepp kan vara tid. Ett annat någon form av rumsligt begrepp. För produkter och deras ingående delar brukar till exempel artikelnummer vara ett begrepp som data kan kopplas till. Det gäller dock att artikelnumret pekar ut en komponent på ett unikt sätt genom att kombineras med till exempel versionsnummer eller serienummer eftersom flera komponenter då kan ha samma artikelnummer men kan skiljas ut med versionsnummer eller serienummer eller ibland båda. De begrepp som unikt identifierar en mängd data om exempelvis en komponent, kallas nyckelfält (Date 1986). Genom att definiera och identifiera nyckelbegrepp för de företeelser som insamlade data beskriver så kan en mer eller mindre sammanhållen datamodell skapas. Hur data semantiskt hänger ihop är viktigt att veta för att data ska kunna behandlas och analyseras.

3.2.17 Osäkerheten är stor

En del försök att samla data, från en produktionsmiljö, för att se olika fenomen har resulterat i ändlösa strömmar av data utan att visa något signifikant mönster. I andra fall har villkor för att detektera en förändring i en dataström genererat enorma mängder av avvikelser som har haft naturliga förklaringar och inte pekat

ut något av betydelse. Andra förhållanden som är svåra att bedöma är vilka formella standarder och informella standarder som kommer att gälla. Utvecklingen av molntjänster sker snabbt och det är oklart vilken standard och vilken prestanda man kan förvänta sig i framtiden när det gäller lagring av stora datamängder (Sugandhi M, Gaganjot K, 2017). Exempel finns i det förflutna när de som satsade på en internationell standard för uppkopplade system i industriell miljö var de ända som satsade på just denna standard, vilket gav enorma problem längre fram. Andra företag satsar idag på en enda leverantör av system för datahantering och analys och tvingas då ställa krav på leverantörer av produktionsutrustning att utrustningen är kompatibel med det valda systemet för databehandling. Rätt eller fel vet ingen idag.

3.3 Stora datamängder för produktutveckling

Stora datamängder och Internet ger oss nya möjligheter inom produktutveckling. Tre principer har formulerats i spåren av denna utveckling (Tan and Zhan 2016). Dessa är: fånga kundvärdet, samarbeta med andra och agera autonomt.

3.3.1 Viktigt att fånga kundvärdet

För företag som tjänat sina pengar på produkter som de utvecklat, är det viktigt att säkerställa att produkterna motsvarar användarnas förväntningar. Om en användare förväntar sig en viss funktionalitet och denna inte finns eller uteblir av någon anledning så skapar det besvikelse. Moderna sätt att kommunicera över Internet via sociala media, forum mm gör att förväntningar och idéer på funktionalitet hos produkter sprids och därmed ökar kraven på de som utvecklar produkterna. Produktutvecklare måste känna till de senaste strömningarna på marknaden. Att ha en direkt koppling till marknaden och användarna av företagets produkter är allt viktigare. Den som inte tar till vara de möjligheter som moderna medel för kommunikation ger kan bli avskärmade från marknadens signaler och blir därmed lätt frånsprungen av konkurrenterna. Det är viktigt att snabbt uppfatta nya behov samt att få återkoppling på hur befintliga och nya produkter tas emot. Som tidigare nämnts är det viktigt att tidigt se förändringar och trender i data som skvallrar om förändringar i demografi, konjunktur, livsstilar mm. Från dessa ord kan vi dra slutsatsen att det är viktigt att vara uppkopplad mot marknaden för att fånga kundvärdet.

3.3.2 Samarbeta med andra

Med en stor skara användare som lätt kan få kontakt över sociala media så kan man räkna med att det kommer att uppstå diskussioner om hur bra en produkt är. Som konsument kan man lätt göra en sökning på nätet för att hitta omdömen om en produkt. Vill man veta vad slags problem som kan uppstå kan man lätt skriva in produktens namn samt ordet problem i sökfönstret på en webbläddrare och göra en sökning. Är produkten behäftad med fel så kan man, i vissa fall, få en lång lista med träffar. Inte helt ovanligt så uppstår det användargrupper och forum där produkterna diskuteras. I många fall så kan man hitta modifieringar av en produkt eller hur den kan användas ihop med andra produkter för att få "bästa effekt". Det

är med andra ord många hjärnor som föder tankar om en produkts användbarhet och hur den bäst ska vara utformad. Mot denna bakgrund framstår det inte helt osannolikt att den som kläcker den bästa idén om en produktändring eller en ny typ av produkt kommer från användarkollektivet och inte från utvecklingsavdelningen på det företag som tagit fram produkten. Här måste man vara ödmjuk och snabbt lära sig att äga andras tankar och idéer. Framför allt innan konkurrenterna gör det.

Förutom användarna till företagets produkter kan återförsäljare, serviceverkstäder och återvinnare bidra med konstruktiva åsikter som kan förbättra företagets produkter och vara grund till nya produkter.

Genom att styra upp och kanalisera alla synpunkter så kan ett innovationskollektiv skapas som arbetar för att förbättra och förnya ett företags produkter långt mer effektivt än om man tänkt själv.

Det gäller med andra ord att skapa och vara en del av ett innovationskollektiv runt företagets produkter och studera dataströmmar från dessa kollektiv.

3.3.3 Agera autonomt

Tiden till marknaden är ofta viktig. Det är även viktigt att kunna omsätta alla de signaler som en uppkoppling till marknaden ger och som ett innovationskollektiv kan erbjuda. Att snabbt ge ut nya eller uppdaterade produkter och utnyttja den återkoppling som ges är av största vikt i många fall. För att detta ska vara möjligt måste kommunikationsvägarna vara korta och byråkratiska. En grupp produktutvecklare som ansvarar för en produkt måste självständigt kunna ta in och analysera all data som kommer in. Därefter måste de kunna och få agera på utfallet av analysen och omsätta det i förbättrade eller nya produkter. Korta och byråkratiska kommunikationsvägar och möjlighet att agera självständigt gör att en produktutvecklingsgrupp måste vara tvärfunktionell för att ha de kompetenser som behövs och den behöver även ha ett delegerat ansvar.

Från det ovanstående kan vi dra slutsatsen att det är viktigt att vara **autonom**.

3.3.4 Att vara uppkopplad mot marknaden - sociala media

Genom att följa diskussioner på sociala media och olika forum så kan användarnas synpunkter fångas upp. Tydliga exempel finns från turist- och resebranschen där man kan läsa om i stort sätt allt från hotell, utflykter och restauranger. När en gäst lämnat ett boende, ber bokningstjänsten om ett omdöme av hur boendet har varit på ett antal bedömningspunkter. Dessa omdömen summeras och visas som ett medeltal i bokningstjänsten. Ett annat konkret exempel är ett köp av en större fritidsbåt där köparen var av åsikten att båtens skrov hade inre skiktbristningar mellan laminaten. En diskussion uppstod på ett forum för fritidsbåtar. Uppmärksamheten runt affären växte och till slut blev den så pass besvärande för leverantören att de tvingades till en uppgörelse med köparen. Ytterligare ett exempel är en leverantör av GPS-navigatorer (Pitta & Fowler 2005) som ville etablera sig på den amerikanska marknaden. De

annonserade efter betatestare på en webbsida som besöktes av målgruppen. Genom ett antal frågor till de som anmält sig, utsågs en testgrupp med de rätta förutsättningarna. Deltagarna fick en gratisversion av produkten och uppmanades att rapportera sina erfarenheter på det forum där de först rekryterades från och även andra liknade forum. Leverantören fick in värdefull återkoppling om produktens duglighet samtidigt som de fick publicitet runt sin produkt. Det bör här påpekas att detta kan få både positiv och negativ påverkan på produktens möjligheter att nå framgång på marknaden. Allt för negativ återkoppling kan skada att varumärke mer än det hjälper (Pitta & Fowler 2005).

Det finns en rad diskussionsforum om olika ämnen. Exempel är hobbyer som sportfiske, veteranbilar, fritidsbåtar men även familjeangelägenheter som förlorade familjemedlemmar och sjukdomar kan diskuteras. Deltagarna kanske inte har en motpart med samma intresse i sin närmaste omgivning men kan via nätet hitta likasinnade. De deltar ofta anonymt på forumet med ett alias men måste ge en validerad e-postadress till forumet så att de kan få notifieringar om aktiviteter på forumet. Ett forums fokus är ofta nischat och kunskapsutbytet kan vara substantiellt. Det börjar med ett delat intresse men så småningom byggs relationer mellan deltagarna när förtroendet dem emellan ökar med tiden. Att hitta vägar att utnyttja den information om människor med olika intressen och olika livssituation kan vara en ovärderlig källa för att förstå hur en produkt kan leverera värde till människor.

Det rekommenderas att utvecklare av produkter som vill dra nytta av sociala media och diskussionsforum inte ska ta en aktiv roll på forumen utan låta andra "sköta snacket" och aldrig fälla negativa kommentarer om konkurrenters produkter. Här finns en svår balansgång att bemästra.



Bild 19 Det kan finnas många kanaler för informationsdelning för de som är intressenter till en produkt

3.3.5 Rapportering från den egna produkten

Det finns många exempel på produkter som samlar data som rapporteras in till originaltillverkaren av produkten. Exempel finns inom fordonsbranschen där såväl bilar som tyngre fordon är uppkopplade via Internet och kan rapportera data och ta emot data. Frågeställningen är ofta hur produkten kan leverera mer värde till kunden. Värdet kan brytas ner till olika företeelser som påverkar värdet. Dessa kan ha både positiv och negativ påverkan på levererat värde. Exempel kan vara:

- Reparation
- Service
- Förebyggande underhåll
- Kortare leveranstid
- Driftsförhållanden
- Kortare utvecklingstid

Genom att studera historiska data och data i närtid, kan exempelvis förebyggande underhåll optimeras för att minska stillestånd på grund av reparation. Genom att studera driftsförhållanden kan leasingavtal optimeras mot lägre risk för både uthyrare och kund. Driftsförhållanden kan även ge utvecklingsavdelningen information så att de kan utveckla produkter med bättre egenskaper för att passa olika behov. Produkternas position kan ge information om när och var resurser för uppdatering, service och reparation bör finnas för att optimera kundvärdet.

Har man ambitioner i denna riktning så är det viktigt att ställa de frågor som man vill ha svar på tidigt och utgå från domänkunskap för att kunna planera för att tillfredsställa behovet av insamlade data och resurser för att analysera data. Eftersom tekniken att samla stora datamängder och analysera dessa så finns det ofta ett utbildningsbehov i en organisation som vill gå denna väg.

Med en stor del insamlade data från den egna produkten, kan möjligheter till värdeskapande analys öka om man kan dra nytta av annan data som finns i vår omgivning. Exempel är data som skvallrar om förändringar i vår livsstil, konsumentbeteenden, ekonomisk utveckling, befolkningstillväxt m m. Användare av ett företags produkter kan få ett ökat kundvärde genom ett tillämpningsprogram till mobiltelefon som även, med kundens medgivande, rapporterar in data om kundens beteende på områden som inte direkt berör produkten men som kan hjälpa företaget att förstå användarnas behov nu och i framtiden.

3.4 Värde för kunden

Nedan kommer beskrivningar om hur modeller och analys av resultat från insamlade data kan användas.

3.4.1 Modeller för att utforska en lösningsrymd

Genom att systematiskt samla testdata från utvecklingsfasen och även samla data från produkter i användning, kan modeller av produkten och delsystem i denna tas fram som kan användas för att utvärdera olika delar av en lösningsrymd eller en funktionsrymd. När förutsättningar för användandet ändras eller om produkten förändras, kan man utifrån dessa modeller lättare förstå om produkten är i ett dugligt område i lösningsrymden. För komplexa produkter som bygger på flera teknologier blir ibland orsak och verkan svåra att överblicka och förstå. Datorn blir då ett nödvändigt verktyg för att tolka insamlade data. Simuleringsmodeller kan byggas genom att använda lärande algoritmer (se bild 20 och 21). Simuleringsmodellerna kan sedan kombineras med andra modeller som tillsammans bildar en digital testbänk för delvis nya koncept.

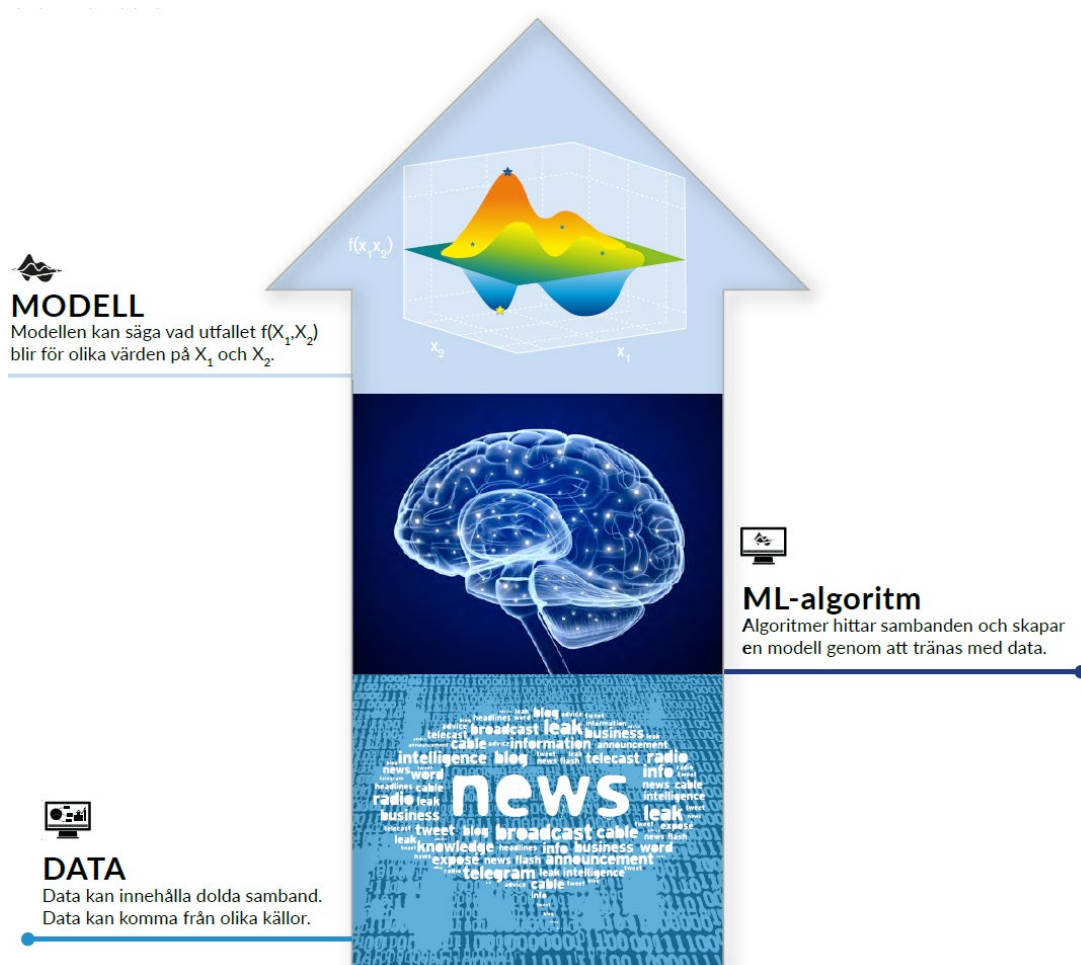


Bild 20 Vi går från data till modeller som hjälper oss att simulera en produkts beteende

3.4.2 Modeller för att ge produkten egenskaper

Produkten kan med sina egna sensorer och medel för kommunikation samla data som lagras i produkten. Data kan användas för att med maskinlärande algoritmer skapa modeller för reglering av olika egenskaper hos produkten (se bild 20 och 21). Vi märker ibland knappt hur webbsidor, mobiltelefoner och handdatorer ändrar beteende när vi använder dem. Fordon kan tala om för oss som förare när vi börjar bli trötta och behöver en paus i körningen.

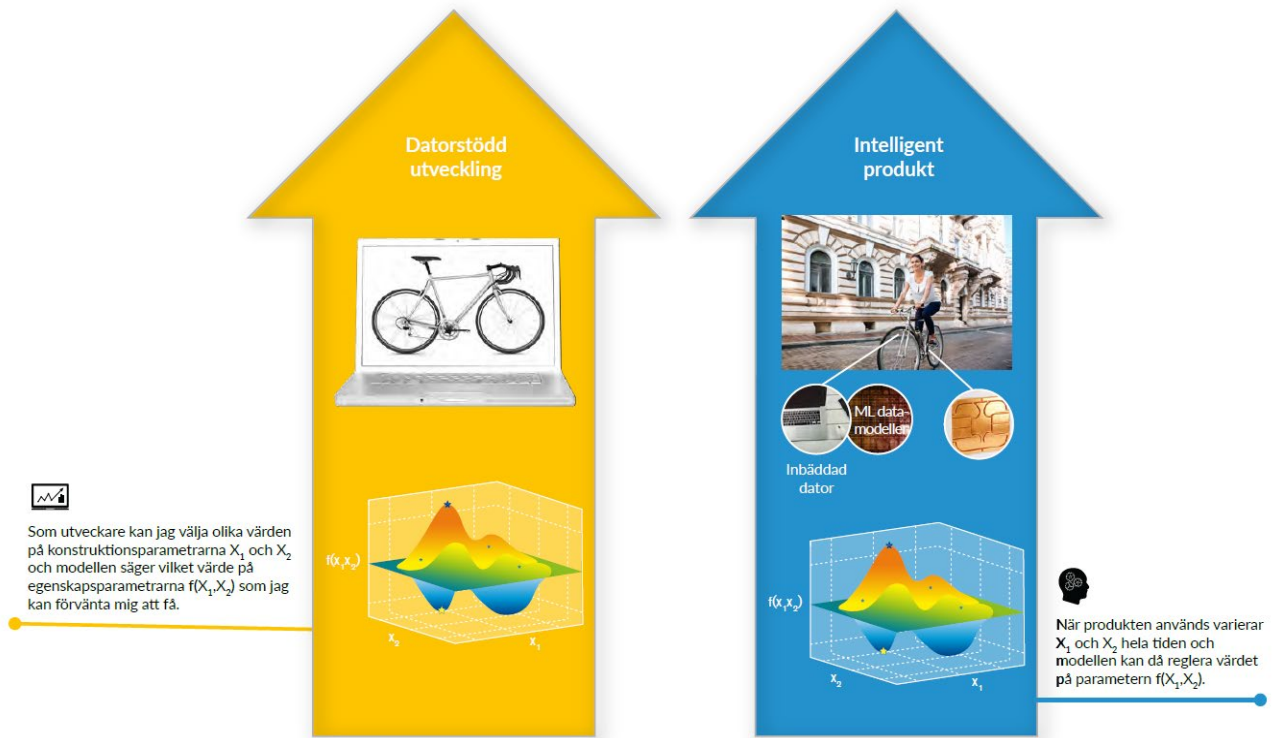


Bild 21 Modeller kan användas i de verktyg som används för att utveckla produkten eller kan användas i produkten för att ge den egenskaper.

4 SWOT-analys av stora datamängder

Utifrån den litteratur vi tagit del av under arbete med denna guidebok har vi identifierat styrkor, svagheter, möjligheter och hot med tekniken och tillämpningarna för stora datamängder som kan vara värda att begrunda då man ska ge sig i kast med att skapa värde utifrån data. Nedanstående tabell summerar dessa och i styckena nedan diskuterar vi generella drag hos dessa och vad man bör tänka på.

Styrkor (Strengths)	Svagheter (Weaknesses)
<ul style="list-style-type: none"> ■ Beskrivning, diagnos, förutsägelser, åtgärd, optimering ■ Datadrivna beslut, faktabaserade beslut ■ Automatiska beslut, automatisering ■ Transparens ■ Bra underlag för att göra datamodeller ■ Bra teknik för att göra datamodeller ■ Förutsäga problem ■ Se trender tidigt ■ Bättre överblick över resurser 	<ul style="list-style-type: none"> ■ Tillförlitligheten i data ■ Bias i data (pga aggregering, tvättning) ■ Integritet ■ Svårighet att identifiera kritiska data ■ Det är svårt att få tillgång till data ■ Man kan drunkna i data ■ Dåligt säkerhetsskydd ■ Bristande digital mognad bland personalen ■ Inlåst data i föråldrade system ■ Värdet av mänsklig erfarenhet och förmåga minskar ■ Kräver kompetens i BD och ML
Möjligheter (Opportunities)	Hot (Threats)
<ul style="list-style-type: none"> ■ Kvalitet, effektivisering och kostnadsreduktion i produktionen ■ Arbetsbesparing ■ Skapa värde för organisationen och kunderna ■ Planera underhåll bättre ■ Riskminimering ■ Bättre flexibilitet ■ Leverantörer kan få information om utrustningar för förbättring av dessa ■ Transparenta data kan ge nya affärer ■ Gemensam situationsuppfattning och samsyn 	<ul style="list-style-type: none"> ■ Konkurrenterna är bättre på att använda tekniken ■ Säkerhet: intrång, manipulation, stöld ■ Säkerhet: känslighet i systemen, strömavbrott etc ■ Extern datakvalitet ■ Lagar och regleringar, till exempel GDPR ■ Övertro på resultatet och tekniken ■ Transparenta data kan äventyra affären om till exempel kostnadsdata blir kända ■ Äganderätten till data är oklart ■ Föråldrade system används i värdekedjan

4.1 Styrkor

Styrkor (Strengths)

- Beskrivning, diagnos, förutsägelser, åtgärd, optimering
- Datadrivna beslut, faktabaserade beslut
- Automatiska beslut, automatisering
- Transparens
- Bra underlag för att göra datamodeller
- Bra teknik för att göra datamodeller
- Förutsäga problem
- Se trender tidigt
- Bättre överblick över resurser

Många av ovanstående styrkor är kopplade till att data ger fakta om kunder, produkten, organisationen och produktionen som underlag till beslut snarare än att beslut fattas baserat på magkänsla. En djupare analys och förståelse om hur allt hänger samman kan erbjudas genom data och detta bidrar till transparens, överblick och bättre beslutsunderlag.

Den nya utvecklingen inom komplex analys (artificiell intelligens) erbjuder dessutom möjligheter till automatisering och optimering som tidigare inte var möjligt. Utvecklingen går mycket fort och möjligheterna exploderar. Om en organisation lyckas dra nytta av de kraftfulla verktyg som utvecklas kommer det med säkerhet vara en stor konkurrensfördel. Det kräver tillgång till data och en bra hantering av data, dvs att organisationen kan bemästra hela värdekedjan för stora datamängder.

4.2 Svagheter

Svagheter (Weaknesses)

- Tillförlitligheten i data
- Bias i data (pga aggregering, tvättning)
- Integritet
- Svårighet att identifiera kritiska data
- Det är svårt att få tillgång till data
- Man kan drunkna i data
- Dåligt säkerhetsskydd
- Bristande digital mognad bland personalen
- Inlåst data i föråldrade system
- Värdet av mänsklig erfarenhet och förmåga minskar
- Kräver kompetens i BD och ML

Ovanstående svagheter är till stor del kopplade till tillgången och hanteringen av data. För att kunna träna algoritmer i den komplexa analysen krävs tillgång till stora mängder data. Då en stor del av arbetet med att skapa värde ur stora datamängder består av en korrekt datahantering behövs kompetens. Utmaningen blir att dels få tag på relevanta data från tillförlitliga källor samtidigt som bias undviks genom att inte sortera data på ett skevt sätt. Korrekt statistisk analys i datahanteringen behövs.

Undermålig eller felaktigt hanterad data kan få allvarliga konsekvenser i underlag till beslut. Styrning av en process utifrån felaktiga eller data med bias säger sig självt kan bli väldigt fel. Detsamma gäller naturligtvis om data blivit manipulerad och därför är det viktigt att skydda data med tillräcklig säkerhet i systemen. I synnerhet data som är intressant för tredje part måste skyddas. Det finns också lagstiftning som kräver att organisationen ska skydda viss typ av data, till exempel GDPR. Det är viktigt att det finns en datainfrastruktur inom organisationen som möjliggör delande av data mellan avdelningar och enheter för att kunna dra full nytta tekniken.

4.3 Möjligheter

Möjligheter (Opportunities)

- Kvalitet, effektivisering och kostnadsreduktion i produktionen
- Arbetsbesparing
- Skapa värde för organisationen och kunderna
- Planera underhåll bättre
- Riskminimering
- Bättre flexibilitet
- Leverantörer kan få information om utrustningar för förbättring av dessa
- Transparenta data kan ge nya affärer
- Gemensam situationsuppfattning och samsyn

Rätt använd kan de tekniska verktygen bidra till organisationens konkurrenskraft på i princip alla områden. Optimera energianvändning, spara arbete, skapa kundvärde, minimera risk och skapa flexibilitet, är bara några av de områden som kan nämnas.

Sammanställda och tolkade data kan ge en gemensam situationsuppfattning där beslut blir mer samstämmiga. Besluten baserar sig på bättre underlag varför risker minimeras eller undanröjes och man därmed undviker arbetskrävande omtag.

En automatiserad insamling, sammanställning och tolkning av data innebär en enorm arbetsbesparing jämfört med att detta ska göras med manuella metoder.

Eftersom tekniken erbjuder så många möjligheter blir det viktiga för organisationen att skilja ut vad som är mest värdeskapande i den situation organisationen befinner sig.

Att ta fram en långsiktig strategi där det första steget identifieras och de efterföljande stegen bygger på de föregående, är att rekommendera. Organisationen behöver gissningsvis både investera i att bygga den tekniska infrastrukturen för att hantera data på ett bra sätt samtidigt som man bygger upp sin analytiska förmåga och förståelsen för teknikens möjligheter.

4.4 Hot

Hot (Threats)

- Konkurrenterna är bättre på att använda tekniken
- Säkerhet: intrång, manipulation, stöld
- Säkerhet: känslighet i systemen, strömavbrott etc
- Extern datakvalitet
- Lagar och regleringar, till exempel GDPR
- Övertro på resultatet och tekniken
- Transparenta data kan äventyra affären om till exempel kostnadsdata blir kända
- Äganderätten till data är oklart
- Föråldrade system används i värdekedjan

De externa hot som finns med tekniken är dels relaterade till säkerhet och dels till konkurrenters förmåga att utnyttja tekniken till sin fördel. När det gäller säkerhet så finns det två aspekter – dels datasäkerhet (manipulation, stöld och tillförlitlighet) och dels systemsäkerhet (känsligheten hos systemen för strömavbrott, internet krascher etc.).

Datahantering har på senare år fått mycket uppmärksamhet och det driver fram ny lagstiftning på området, till exempel GDPR. Lagstiftningen framöver kommer bland annat att hårdare reglera äganderätten till data men också själva datahanteringen. Massövervakningslagar som Freedom Act (2015) (tidigare Patriot Act) och Cloud Act (2018) i USA kan också störa affärsverksamheten genom att ge amerikanska myndigheter möjlighet att kräva att amerikanska molnleverantörer ger tillgång till data lagrade på servrar även utanför USA. I det avseendet rekommenderade eSam (eSamverkansprogrammet) att undvika att spara känslig eller klassificerad information hos US-baserade molnleverantör och uppmuntrade användningen av stark kryptering.

För molnbaserade lösningar är utmaningen att konceptet för dessa fortfarande är under stor utveckling. Utvecklingen av molntjänster sker snabbt och det är oklart vilken standard och vilken prestanda man kan förvänta sig i framtiden när det gäller lagring av stora datamängder (Sugandhi M, Gaganjot K, 2017).

5 Slutsatser

En organisation behöver både en analytisk mognad och digital infrastruktur för att kunna dra full nytta av den nya kraftfulla tekniken. Om en organisation lyckas dra nytta av de kraftfulla verktyg som utvecklas kommer det med säkerhet vara en stor konkurrensfördel, medan det kommer att vara en stor nackdel om konkurrenter lyckas bättre. Utöver att organisationen behöver förståelse för hur man kan använda sig av data för att skapa värde så måste kvaliteten och tillgången på råvaran data säkerställas. Det är också viktigt att komma ihåg att grundläggande i datahanteringen är statistisk kompetens och analytisk mognad. All filtrering/tvättning och aggregering av data måste göras på ett sådant sätt att man undviker bias, som i värsta fall kan leda till felaktiga slutsatser. Säkerhet för att skydda värdekedjan för dataintrång bör anpassas efter situationen. En bedömning av hur stor risken för intrång (stöld/manipulation) och störningar (till exempel elavbrott) bör göras och ligga till grund för de åtgärder som vidtas.

De organisationer som tillhör de analytiskt utövande eller de analytiskt utmanande, bör lägga strategiskt arbete på hur utveckling och komplettering av den analytiska mognaden kan åstadkommas.

I formulerandet av målsättning och strategi för sitt arbete med stora datamängder bör en analys utifrån värdeskapande utgöra grunden. Tekniken erbjuder många möjligheter och organisationen måste därför göra ett arbete med att definiera hur den kan skapa störst värde och prioritera mellan de möjligheter som finns utifrån de resurser som är tillgängliga.

Då området haft en explosiv utveckling har lagstiftning och standardisering inte hängt med. Emellertid har framför allt lagstiftare vaknat och det är att vänta att ny lagstiftning kommer att ha stor påverkan framöver speciellt inom ägande av data.

Summering av råd till organisationer som vill skapa värde av stora datamängder:

- Lyft den analytiska förmågan i organisationen
- Identifiera och prioritera bland möjliga värden som tekniken kan erbjuda
- Säkerställ tillgången på rätt data av hög kvalitet, anpassa den digitala infrastrukturen utifrån de värden som ska skapas, inklusive skydd av värdekedjan mot intrång och störningar
- Ha koll på lagstiftning och standardisering

Datainfrastrukturen definierar vilken typ av dataanalys ett företag ska kunna utföra. Det är viktigt att tänka igenom det och planera på lång sikt. Valet beror mest på vilken typ av data som kommer att hanteras, hur mycket data det finns och dess användning. Ett stort antal organisationer använder sig numera av databaser och vill förbättra sin datakapacitet för att ligga före konkurrenterna. Datasjöar och datalager är infrastrukturer som kan samla, lagra och analysera data från en hel organisation. De kan göra data tillgänglig över organisationsgränser och undvika negativa effekter av stängda datasilor samt erbjuda möjlighet till omfattande dataanalys. Datasjöar är speciellt lämpade för ostrukturerade data, men kräver en dataingenjör för att behandla data innan företagsanvändaren kan

använda den. Datalager är effektivare med strukturerade data och är lättare att använda för företagsanvändare.

Valet mellan molnlösning eller lokal lagring beror för det mesta på datas känslighet (säkerhet och kontroll), intern IT-kompetens och tillgängligheten av programvara. Nuförtiden är det mer populära alternativet att gå till ett hybrid där känsliga data lagras lokalt men resten lagras i ett moln.

Förbättrad datainfrastruktur ger möjlighet att implementera AI-algoritmer för att automatisera processer eller införa andra lösningar baserade på dataanalys. Artificiell intelligens och maskininlärning revolutionerar snabbt vårt samhälle tack vare den mängd data vi genererar. När fler och fler företag använder det i sin produktion eller i sina tjänster är det viktigt att förstå vilken konkurrensfördel det kan ge.

Maskininlärning (ML) är olika metoder och algoritmer som går igenom data för att analysera, skapa förståelse och dra slutsatser av dess mönster. ML-algoritmen lär sig hur data beter sig och kan göra förutsägelser, klassificera eller lära sig nya förmågor. Resultaten som ges av en AI är starkt beroende av sin inlärningsmetod. Det mesta av lärandet idag görs genom övervakat lärande där en operatör hjälper maskinen att lära sig. Men oövervakat lärande och förstärkt lärande kan leverera extremt insiktsfulla resultat och har stor potential.

Inom AI-området finns även artificiella neurala nätverk (ANN), som försöker reproducera hjärnans neuroner, vilka används alltmer i det som benämns djupinlärning. Djupinlärning är en delmängd av maskininlärning och tillämpar de tre tidigare beskrivna inlärningsmetoderna ovan. Metoden använder omfattande ANN och är den teknik som ger resultat närmast mänsklig intelligens. Djupinlärning används i självkörande bilar, kan utföra realtidsöversättning, skapa musik eller måla, förutsäga underhåll eller utföra ansiktsgenkänning. Det är en beräkningsmässigt intensiv metod som kräver särskild maskinvara (GPU) för att kunna köras inom rimlig tid.

Summering av tekniska råd till organisationer som vill skapa värde av stora datamängder:

- Lär känna din data. Undersök vilka typ av data ((semi-) strukturerad eller ostrukturerad) som genereras inom din organisation.
- Kartlägg den befintliga infrastrukturen, databaser och datakällor som redan finns samt den interna kompetens (IT, analytiker etc) ni har som hanterar data.
- Definiera kortsiktiga, halvlånga och långsiktiga mål som du vill uppnå när det gäller användning av data och automatisering. Dessa kan vara:
 - Kortsiktigt: samla in och analysera data från en maskin
 - Halvlångt: använd AI för att utföra kvalitetskontroll med hjälp av insamlade data
 - Långsiktigt: samla in data från hela tillverkningskedjan och optimera produktionen

-
- Uppgradera din IT-infrastruktur och IT-kapacitet i enlighet med dina mål

När det gäller att tillämpa stora datamängder för att skapa värde i produktframtagningsskedjan finns ett par slutsatser och råd att ta till sig:

- Inhämta kunskap inom området för att fatta rätt beslut om hur tekniken ska användas
- Ämnesområdet är brett och kräver kunskaper i matematik, programmering, maskinlärande, artificiell intelligens, informationsmodellering och IT-system
- Det kommer troligen att krävas en del testande av explorativ art innan man hittar rätt tillämpning som ger eftersökt värde för det egna företaget och dess kunder.

6 Referenser

- Hu, Han, et al, Toward Scalable systems for big data analytics: a technology tutorial, IEEE, vol 2, pp 652-687, (2014)
- McKinsey global report –Big data: the next frontier for innovation, competition and productivity, (2011)
- MIT Sloan Management review, From value to vision; Reimagining the possible with data analytics (2013)
- MIT Sloan Management review, Lessons from becoming a data-driven organization, (2016)
- Research now Gartner: 3D data management: controlling data volume, velocity and variety (2001)
- Erl, Thomas, et al, Big data fundamentals, concepts, drivers & techniques (2016)
- <https://www.datainspektionen.se/lagar--regler/dataskyddsförordningen/> (2019) Dataskyddsförordningen. (2018, december 10).
- Wikipedia- 1, *Wikipedia*, Hämtad 07.46, april 5, 2019 från [//sv.wikipedia.org/w/index.php?title=Dataskyddsf%C3%B6rordningen&oldid=43964047](https://sv.wikipedia.org/w/index.php?title=Dataskyddsf%C3%B6rordningen&oldid=43964047).
- [Lawrence, J, Malmsten, J, Comparing TensorFlow Deep Learning Performance Using CPUs , GPUs , Local PCs and Cloud \(2017\)](#)
- Finlay, S., Artificial Intelligence and machine learning for business; Relativistic Books (2017)
- D. Silver & Al., Mastering the game of Go without human knowledge, Nature 550, pp. 354-359 (2017)
- D. Silver & Al., A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play, Science vol. 362, Issue 6419, pp. 1140-1144 (2018)
- <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>
- <https://cloud.google.com/blog/products/gcp/preparing-and-curating-your-data-for-machine-learning>
- <https://www.mathworks.com/discovery/deep-learning.html>
- <http://www.esamverka.se/nyheter/nyheter/2018-11-12-esam-ser-risker-med-molntjanster-i-offentlig-sektor.html>
- H.R.2048 - USA FREEDOM Act of 2015, Public Law No: 114-23 (2015)
- S.2383 - CLOUD Act (2018)
- Sethi P and Sarangi R, Internet of things: Architectures, protocols and applications, Journal of Electrical and Computer Engineering (2017)
- Tan K H, Zhan Y, Improving new product development using big data: a case study of an electronics company, R & D Management (2017)
- Pitta D A, Fowler D, Online consumer communities and their value to new product developers, Journal of Product & Brand Management (2005)

-
- Date C. J., *An Introduction to Database Systems*, Vol 1, 4th Ed. Addison-Wesley Publishing Company (1986)
 - Sugandhi M, Gaganjot K, *Cloud Deep Down – SWOT Analysis*, 2017 2nd International Conference on Telecommunication and Networks (2017)
 - Bjärring M., *Increasing the Value of Data in Production Systems*, Chalmers University of Technology, Gothenburg, Sweden (2019)
 - Wikipedia – 2, Sakernas internet. (2018, november 15). *Wikipedia*, . Hämtad 14.26, april 1, 2019 från [//sv.wikipedia.org/w/index.php?title=Sakernas_internet&oldid=43861106](https://sv.wikipedia.org/w/index.php?title=Sakernas_internet&oldid=43861106)
 - Wikipedia – 3, Radio Frequency Identification. (2019, februari 18). *Wikipedia*, . Hämtad 14.29, april 1, 2019 från [//sv.wikipedia.org/w/index.php?title=Radio_Frequency_Identification&oldid=45055209](https://sv.wikipedia.org/w/index.php?title=Radio_Frequency_Identification&oldid=45055209)
 - Wikipedia – 4, Närfältskommunikation. (2019, januari 22). *Wikipedia*, . Hämtad 14.31, april 1, 2019 från [//sv.wikipedia.org/w/index.php?title=N%C3%A4rf%C3%A4ltskommunikation&oldid=44377434](https://sv.wikipedia.org/w/index.php?title=N%C3%A4rf%C3%A4ltskommunikation&oldid=44377434)
 - Wikipedia – 5, Bluetooth. (2019, april 1). *Wikipedia*, . Hämtad 14.35, april 1, 2019 från [//sv.wikipedia.org/w/index.php?title=Bluetooth&oldid=45319556](https://sv.wikipedia.org/w/index.php?title=Bluetooth&oldid=45319556)
 - Wikipedia – 6, Zigbee. (2017, april 3). *Wikipedia*, . Hämtad 14.36, april 1, 2019 från [//sv.wikipedia.org/w/index.php?title=Zigbee&oldid=39457568](https://sv.wikipedia.org/w/index.php?title=Zigbee&oldid=39457568)
 - Wikipedia – 7, Wikipedia contributors. (2019, March 26). IEEE 802.15. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:40, April 1, 2019, from https://en.wikipedia.org/w/index.php?title=IEEE_802.15&oldid=889558290
 - Wikipedia – 8, Wikipedia contributors. (2019, March 15). 6LoWPAN. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:41, April 1, 2019, from <https://en.wikipedia.org/w/index.php?title=6LoWPAN&oldid=887884134>
 - Wikipedia – 9, Wikipedia contributors. (2019, March 5). Sigfox. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:42, April 1, 2019, from <https://en.wikipedia.org/w/index.php?title=Sigfox&oldid=886299006>
 - Wikipedia – 10, Wikipedia contributors. (2019, March 26). LoRa. In *Wikipedia, The Free Encyclopedia*. Retrieved 14:43, April 1, 2019, from <https://en.wikipedia.org/w/index.php?title=LoRa&oldid=889550784>