

UTFORDRINGER VED BRUK AV SELVBETJENTE ETTERRETNINGSVERKTØY F



TROND VEGARD JOHANNESSEN er førsteamanuensis ved Norges Handelshøyskole, hvor han også ble utdannet siviløkonom med doktorgrad. Johannessen forsker blant annet på presentasjon av data og IKT-baserte beslutningsstøttesystemer.



ANNA METTE FUGLSETH er professor emerita ved Norges Handelshøyskole. Hun har doktorgrad fra samme institusjon. Fuglseth er spesielt opptatt av å studere effekter av integrerte informasjonssystemer og beslutningsstøttesystemer på effektiviteten i forretningsorganisasjoner.

SAMMENDRAG

Artikkelen gir et bidrag til diskusjonen om fremtidens behov innenfor økonomisk utdanning. Utviklingen går i retning av at ledere og ansatte i høyere grad skal basere sine beslutninger på data, og at de selv skal søke etter og forberede data for analyse. Artikkelen presenterer resultater fra analyse av besvarelser på en hjemmeeksamen i datamodellering og

business intelligence for førsteårs bachelorstudenter. Resultatene viser at nyutviklede «brukervennlige» verktøy ikke kan erstatte brukernes kunnskaper om datamodellering, men også at studentene med innføring i grunnleggende datamodellering relativt raskt kom på et nivå hvor de kunne bruke verktøyene effektivt for beslutningsstøtte.

INNLEDNING

Formålet med denne artikkelen er å gi et bidrag til diskusjonen om fremtidens behov innenfor økonomisk utdanning sett fra arbeidsgivers perspektiv, jf. dette nummeret av *Magma*.

Vi er i en tid hvor begreper som *business intelligence* (BI), *business analytics* (BA) og *big data* stadig nevnes i forbindelse med beslutningsprosesser i organisasjoner. I det meste av litteraturen legges det vekt på hvilke muligheter utviklingen innenfor informasjons- og kommunikasjonsteknologi (IKT) gir for å ta mer effektive beslutninger. Det legges mindre vekt på å diskutere hvilke kunnskaper og ferdigheter beslutningstakerne må ha for å kunne bruke teknologien effektivt. I denne artikkelen vil vi spesielt peke på betydningen av grunnleggende datakunnskaper også hos fremtidens økonomer.

De tre begrepene BI, BA og *big data* er ikke veldefinerte. *Big data* refererer til den økende mengden data som er tilgjengelig i digital form, og som gir beslutningstakere muligheter til i høyere grad å basere sine beslutninger på analyse av data (McAfee & Brynjolfsson, 2012). BI er spesielt knyttet til søking etter og klargjøring av data for analyse, men omfatter også visse former for analyse som aggregering og sammenstilling av data. BA inkluderer avansert statistisk dataanalyse. Begrepene er imidlertid nært knyttet sammen (Chen, Chian, & Store, 2012). Nyttan av BA forutsetter at beslutningstakerne har relevante datasett å analysere. Det er imidlertid en tendens til å se BA som den *avanserte* delen av en datastøttet beslutningsprosess (Andersen & Wallace, 2015, s. 197). Dermed tones betydningen av datakunnskaper ned.

Samtidig utvikles «brukervennlige» verktøy for støtte av BI-prosesser, nemlig verktøy for *self-service BI* (SSBI), som vi har oversatt med selvbetjent etterretning. SSBI er en tilnærming til BI hvor beslutningstakere og deres assistenter innhenter og tilrettelegger data for analyse uten hjelp fra dataspesialister. SSBI støttes av verktøy som tillater brukerne å hente eksterne data, for eksempel fra internett, koble slike data til interne data og analysere datasettene uten at brukerne må vente på hjelp fra IKT-avdelingen.

En gjennomgang av litteraturen viser at søkelyset innenfor SSBI-verktøy er på utviklingen av ny funksjonalitet og brukervennlighet, se neste avsnitt. I en del av litteraturen uttrykkes bekymring for bruk av data som ikke er tilrettelagt av dataspesialister. Det foreslås forskjellige måter å begrense SSBI på, blant annet ved kun å tillate superbrukere å laste ned data fra eksterne kilder og ved å sette krav til datakvalitet (Alpar & Schulz, 2016, s. 153–154). Vårt syn er at det ikke er mulig å begrense lederes tilgang til data, og at en av fordelene med SSBI nettopp er at ledere som overvåker omgivelsene omhyggelig, kan ta flere data i betraktning i sine beslutningsprosesser. Vi bør heller utdanne økonomer som er i stand til å samle inn og analysere data, og som forstår utfordringene ved å tilrettelegge data selv, slik at de kan utnytte det potensialet som ligger i SSBI til å ta effektive beslutninger.

Vi har formulert to forskningsspørsmål knyttet til SSBI:

Er det en sammenheng mellom kunnskaper om datamodellering og effektiv bruk av SSBI-verktøy for beslutningsstøtte?

Hvis det er slik, hvilke utfordringer har brukere med begrensede kunnskaper om datamodellering når de bruker SSBI-verktøy i beslutningsprosesser?

Hensikten med det første spørsmålet er å finne ut om funksjonaliteten og brukervennligheten til SSBI-verktøy faktisk kan støtte sluttbrukere slik at tilrettelegging av data på egen hånd gir korrekte analyser. Hensikten med det andre spørsmålet er å finne ut hvilke problemer sluttbrukere eventuelt har med bruk av SSBI-verktøy.

For å svare på forskningsspørsmålene har vi laget en oppgave og analysert hvordan bachelorstudenter har løst en oppgave med bruk av SSBI-verktøy til å tilrettelegge og analysere data i forskjellige formater. En

vesentlig del av oppgaven var bruk av SSBI-verktøy for å analysere behovet for nye barnehager.

Resten av artikkelen er organisert som følger: I neste avsnitt går vi gjennom litteraturen om SSBI. Deretter presenterer vi forskningsmodellen og forklarer hvordan vi har samlet inn og analysert data. I det følgende avsnittet beskriver vi resultatene. I siste avsnitt diskuterer resultatene og begrensninger ved undersøkelsen.

LITTERATURGJENNOMGANG

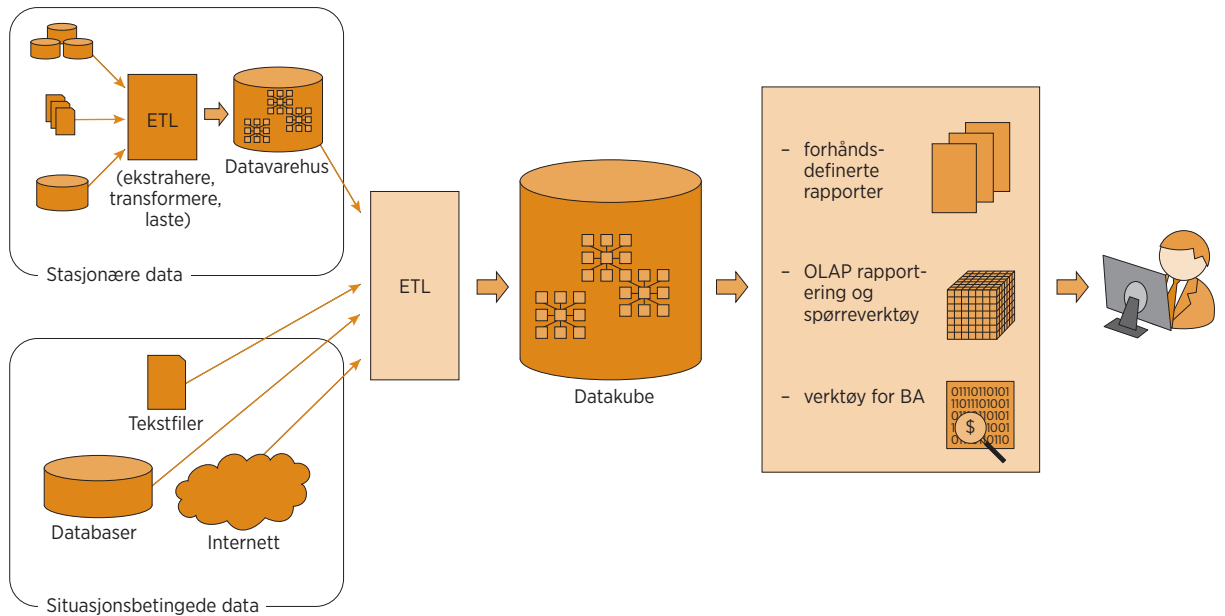
I dette avsnittet forklarer vi først vesentlige begreper med vekt på begrepene *business intelligence* (BI) og selvbetjent BI (SSBI), og vi fokuserer på forskjellen mellom de to begrepene. Deretter gjennomgår vi litteraturen om SSBI for å posisjonere vår forskning.

BUSINESS INTELLIGENCE (BI) OG SELVBETJENT BUSINESS INTELLIGENCE (SSBI)

BI defineres på flere måter (se Presthus, 2015 for en oversikt). Felles for definisjonene er at formålet med BI er å støtte sluttbrukere, det vil si både ledere og ansatte, slik at de kan ta «bedre» beslutninger (Watson, 2009), altså mer målrettede beslutninger (Pfeffer & Salancik, 1978, s. 11), eller «bedre og raskere» beslutninger (Chaudhuri, Dayal, & Narasayya, 2011). Noen definisjoner legger vekt på teknologien. Andre definisjoner omfatter både teknologien og bruken av teknologien. Vi er enige med Ask (2013) i at man bør skille mellom teknologi, arkitektur og bruk av teknologi. For denne artikkelen finner vi det hensiktsmessig med en snever definisjon av begrepet IKT, slik at vi bruker begrepet for maskinvare, programvare og datanettverk, altså utstyr som er ment å øke menneskers arbeids- og informasjonsbehandlingskapasitet (Zuboff, 1989; se også Orlikowski, 1992). Ordet arkitektur kommer fra latin *architectura*, som betyr bygningskunst (Hastrup, 1983). Begrepet betegner her sammensetning av IKT-komponentene.

Med utgangspunkt i Simons modell for beslutningstaking som en prosess med fasene *intelligence*, utforming av alternativer, valg, utførelse og oppfølging (Simon, 1960, s. 2–3), definerer vi BI som søking etter og analyse av data i eksterne og interne kilder etter forhold som indikerer behov for beslutninger. Simon hadde lånt betegnelsen *intelligence* fra militæret, hvor betegnelsen brukes om innsamling og analyse av spesielt fienders planer og handlinger, det vil si etterretningsvirksomhet.

FIGUR 1 Forenklet SSBI-arkitektur.



SSBI defineres også på flere måter. Schlesinger og Rahman (2015) legger vekt på at brukerne må forstå hvordan organisasjonen og omgivelsene er representert i datavarehuset, slik at de er mindre avhengige av dataspesialister når de bruker datavarehuset. Et datavarehus er en database med interne og eksterne data som spesielt skal støtte lederes beslutningstaking. Datavarehuset organiseres og kvalitetssikres av dataspesialister (Devlin & Murphy, 1988). Abelló et al. (2013) legger vekt på at hensikten med SSBI er at brukere som ikke er dataspesialister, skal være i stand til å inkludere data som ikke finnes i datavarehuset, i sine analyser. Alpar og Schulz (2016) skjelner mellom tre nivåer av SSBI. Det laveste nivået går på bruk av data i predefinerte rapporter. Det neste nivået omfatter i tillegg tilgang til data på detaljert nivå i datavarehuset. Det tredje nivået svarer til beskrivelsen hos Abelló et al. (2013).

Vi bygger på beskrivelsen hos Abelló et al. (2013). Dagens forretningsomgivelser er karakterisert ved global konkurranse, raske teknologiske og økonomiske endringer og press på kostnader. For å håndtere slike omgivelser må ledere være i stand til å oppdage data som kan reflektere uventede og svake endringer i omgivelsene, og vurdere slike endringer kritisk i sine

beslutningsprosesser: «Er dette signaler som vi må reagere på?»

Figur 1 viser en forenklet versjon av en SSBI-arkitektur, basert på beskrivelsen hos Abelló et al. (2013).

Figuren viser en beslutningstaker som søker etter og analyserer data. Organisasjonens datavarehus vises øverst til venstre. Datavarehuset inneholder historiske og aggregerte data som anses relevante for ledelsesformål. Slike data kalles stasjonære data (Abelló et al., 2013). Dataspesialister trekker ut, organiserer og laster data fra interne og eksterne kilder inn i datavarehuset (ETL-prosesser). Figuren illustrerer at SSBI-arkitekturen er en utvidelse av datavarehuset, og at datavarehuset er en vesentlig datakilde for SSBI. I tillegg til stasjonære data omfatter SSBI-arkitekturen situasjonsbetingede data, det vil si data som kan være relevante for en beslutningsprosess, men som ikke inngår i stasjonære data (Löser, Hueske, & Markl, 2008). SSBI-arkitekturen settes opp og administreres av brukeren, det vil si at beslutningstakerne håndterer ETL-prosessene selv.

Til høyre i figuren vises eksempler på BI- og SSBI-verktøy. Predefinerte rapporter er knyttet til det første brukernivået hos Alpar og Schulz (2016) nevnt ovenfor.

For SSBI er spesielt OLAP-verktøyene vesentlige. OLAP står for *on-line analytical processing* og brukes om analyseverktøy med kraftfulle operasjoner. Som figuren indikerer, knyttes OLAP-verktøy ofte til en såkalt datakub. Kuben vises vanligvis tredimensjonalt fordi tre dimensjoner er enkelt å illustrere, men en datakub kan omfatte flere dimensjoner – den er altså n-dimensjonal. Eksempler på dimensjoner er tid, sted, kundesegment og produktsegment. Med OLAP-verktøy kan brukerne raskt presentere data langs hver dimensjon på skjermen, og de kan vise utsnitt av kub (slice and dice). I tillegg kan brukerne bore seg ned i detaljer (drill down), og de kan aggregere data (drill up) (Kimball & Ross, 2013).

Den vesentligste forskjellen mellom «tradisjonell» BI og SSBI er fleksibiliteten i SSBI-arkitekturer som gir brukerne mulighet til å inkludere data som de mener er relevante for å vurdere trusler og muligheter i omgivelsene og organisasjonen uten å være avhengig av dataspesialister. Data i datakuben kan være relevant for utforskning av ett problem eller én mulighet, eller data kan reflektere databehov som ikke er tatt inn i organisasjonens datavarehus ennå.

POSISJONERING AV ARTIKKELEN

Det meste av litteraturen om SSBI er knyttet til diskusjon av den teknologiske utviklingen som har gjort det mulig å utvikle SSBI-arkitekturer og analytiske verktøy for brukere som ikke er eksperter på datateknologi. Noen forfattere diskuterer utfordringer med å utvikle datakuber, og de foreslår forskjellige måter som teknologien kan støtte beslutningstakere på når de utvikler datakuber (Abelló et al., 2013; Varga, Romero, Pedersen, & Thomsen, 2014). Schlesinger og Rahman (2015) diskuterer krav til analytiske verktøy og til beskrivelsen av data (metadata) for at brukerne skal få støtte.

Andre forfattere diskuterer risikoer knyttet til SSBI og foreslår flere måter å begrense brukernes tilgang til data på. Alpar og Schulz (2016, s. 153) foreslår for eksempel at man bør dele brukerne opp i superbrukere og ikke-superbrukere og kontrollere aksess til data slik at kun superbrukerne har lov til å hente inn data fra andre kilder enn datavarehuset. Disse forfatterne foreslår også at det må settes visse minimumskrav til datakvalitet.

Vi mener det ikke er praktisk mulig å begrense tilgangen til data for ledere på strategisk og taktisk nivå. En av fordelene med SSBI er nettopp at teknologien gir ledere

som overvåker omgivelsene omhyggelig, mulighet til å ta hensyn til data som kan reflektere svake signaler, i sine beslutningsprosesser.

Vi tror at SSBI har et potensial til å forbedre beslutningsprosesser. I flere undersøkelser har vi erfart at lederes datasøkings- og analyseprosesser har økt effektiviteten av deres beslutninger. Et eksempel er rederen som kontinuerlig overvåket markedene for svake signaler om endringer i etterspørselen etter transporttjenester. Han samlet selv inn data og analyserte dem ved hjelp av regneark og statistikkpakker. På grunn av denne overvåkingen kunne han reagere raskt og høste *first-mover*-fordeler ved flere anledninger (Fuglseth, 1989; Fuglseth, 2005, s. 293). Han hadde også en grundig forståelse av kvaliteten på data som han samlet inn, og han visste hvilke datakilder som var mest troverdige (Fuglseth, 1989). Rederen var økonom med kunnskaper om databehandling. Han brukte imidlertid mye tid på å legge inn data manuelt. En av fordelene ved den teknologiske utviklingen er at flere data er tilgjengelige digitalt (McAfee & Brynjolfsson, 2012), og at det er blitt betraktelig enklere å integrere og analysere data.

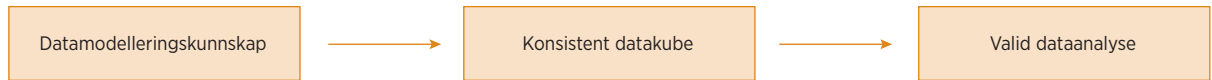
Alpar & Schulz (2016, s. 153) hevder imidlertid at flere sluttbrukere ikke er i stand til å utnytte potensialet som SSBI har til å integrere stasjonære og situasjonsbetingede data, uten hjelp av dataspesialister (se også Stodder, 2015). Som nevnt ovenfor diskuterer flere forskere risikoene ved SSBI. Risikoene er relatert både til utviklingen av en konsistent datakub og til relevant bruk av analyseverktøy. Løsningen på problemene synes å være knyttet mer til restriksjoner og kontroll enn til utdanningen av sluttbrukerne. Alpar og Schulz (2016, s. 152) mener at brukere med «de rette ferdighetene og tilgangsrettigheter» («appropriate skills and access rights») burde få adgang til å hente inn og utforske data. Forfatterne forklarer imidlertid ikke hva de mener med «rette ferdigheter», og vi har ikke funnet diskusjoner av hvilke ferdigheter brukere bør ha for å kunne bruke SSBI-verktøy effektivt. Formålet med denne artikkelen er å gi et bidrag til diskusjonen om utdanningen av fremtidige økonomer i datafag.

METODE

BEGREPSMODELL

Basert på ovenstående diskusjon utviklet vi begrepsmodellen som er vist i figur 2:

FIGUR 2 Begrepsmodell.



Vi antar at validiteten av dataanalysene med SSBI-verktøy påvirkes av hvordan data er organisert i datakuben. Med validitet mener vi at analysene gir korrekte resultater. I vår undersøkelse hadde vi spesifisert visse analyser som skulle legges ved besvarelsen, slik at vi kunne kontrollere om operasjonene i datakuben ga korrekte resultater. Med en konsistent datakube menes at data logisk kan sammenstilles, og at nødvendige koblinger mellom data er spesifisert korrekt. Dessuten antar vi at brukernes kunnskaper om datamodellering påvirker deres evne til å sette opp en konsistent datakube.

OPPGAVE

For å finne ut om kunnskaper om datamodellering påvirker bruken av SSBI-verktøy, utarbeidet vi hjemmeeksamen i et obligatorisk kurs i databehandling som også var egnet til å besvare forskningsspørsmålene. Studentene ble bedt om å hjelpe en leder i en barnehagekjede med å identifisere behovet for nye barnehager. Studentene var i sitt første år på bachelorprogrammet, og det var 401 studenter på kurset. Besvarelsene ble levert inn i grupper på tre–fire studenter, og studentene hadde 8 1/2 dag til å besvare oppgaven med en rapport på maksimalt 30 sider.

Hjemmeeksamenen bestod av tre deloppgaver. I den første deloppgaven skulle studentene utvikle en normalisert datamodell som tilfredstilte visse krav til databasen for drift av barnehagekjeden. I den andre oppgaven skulle studentene lage en dimensjonal datamodell basert på data fra fem datakilder. Datamodellen skulle implementeres ved hjelp av SSBI-verktøyene MS Power Query og MS Power Pivot. Disse verktøyene er nå integrert i MS Excel 2016. Power Query skulle vesentlig brukes som et ETL-verktøy til å hente ut, tilpasse og laste data inn i Power Pivot. I den tredje oppgaven skulle studentene utføre spesifikke analyser og komme med en anbefaling om nye barnehager til lederen for barnehagekjeden.

Fire av de fem datakildene var tilgjengelige som en del av oppgaven:

- en tabell med data om antall barn i hver type barnehage for perioden 2002–2013 (private, kommunale og fylkeskommunale/statlige barnehager)
- en tabell med data om Norges kommuner per 2013
- en tekstfil med data om Norges befolkning per kommune for perioden 1986–2014
- en tekstfil med data om antall barnehager per barnehagetype for perioden 2002–2013

Den femte kilden var et html-dokument med data om norske fylker som studentene måtte hente fra Wikipedia.

Før utlevering av oppgaven hadde studentene hatt to forelesninger om datamodellering og to forelesninger om Power Query og Power Pivot. Forelesningene om datamodellering gikk på grunnleggende prinsipper for utvikling av normaliserte datamodeller og en innledning til dimensjonal datamodellering. En datamodell er en modell for organisering av data i en database. Relasjonsdatabaser er den mest brukte formen for databaser. I en relasjonsdatabase er data organisert i tabeller (relasjoner) som kan kobles sammen ved hjelp av entydige verdier i kolonner som kalles nøkler. Hver tabell representerer en entitetstype, det vil si en gruppe fenomener vi ønsker å lagre data om, for eksempel kunder, produkter, ordrer. Hver rad i tabellen representerer en entitet, for eksempel en kunde. Hver kolonne i tabellen representerer en egenskap ved entiteten, for eksempel navn og adresse. En egenskap eller et sett egenskaper som velges til å identifisere hver entitet i tabellen entydig, kalles en primærnøkkel. En fremmednøkkel er en egenskap eller et sett egenskaper som inneholder primærnøkkel fra en annen tabell. Fremmednøkler brukes for å koble tabeller.

Studentene lærte å dokumentere datamodeller i form av ER-diagram og tabelldefinisjoner. Et ER-diagram viser entitetstyper og koblinger mellom disse. En tabelldefinisjon beskriver navnet på en tabell (entitetstypen) og hvilke egenskaper ved entitetstypene vi ønsker å lagre opplysninger om. Dessuten inneholder

tabelldefinisjoner opplysninger om primærnøkler og fremmednøkler (Fuglseth, Håtuft, & Johannessen, 2010).

Normaliserte datamodeller er vanlig for transaksjonsdatabaser, det vil si databaser som støtter prosesser på operasjonelt nivå i en organisasjon, for eksempel salgsprosessene og innkjøpsprosessene. Normaliseringsprosessen består i å sette opp tabellene med minst mulig overlapp i data. På kurset lærte studentene å analysere en oppgave og foreta normaliseringsprosessen med spesifisering av entitetstyper, egenskaper og nødvendige koblinger mellom entitetstypene.

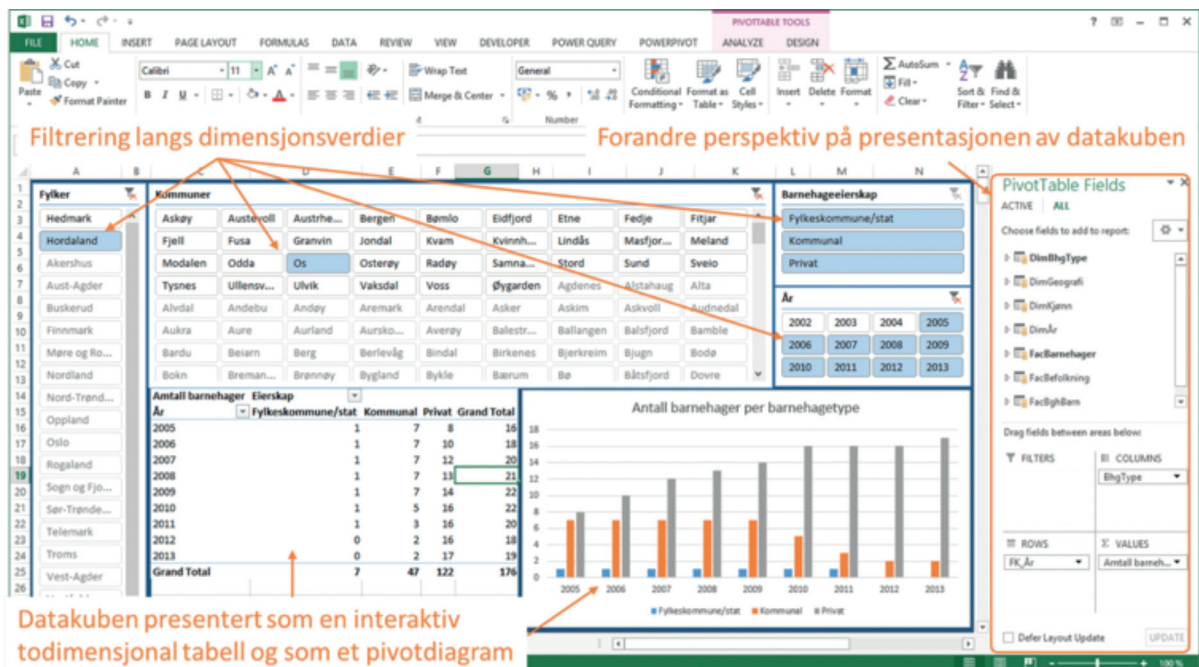
Formålet med dimensjonale modeller er å utforme databaser som gir et ledelsesperspektiv på data, altså legges det mindre vekt på transaksjoner og mer vekt på analyse av utviklingen i aggregerte data over tid. Data organiseres i fakta- og dimensjonstabeller. Prinsippene med å organisere data i tabeller og koblinger mellom tabeller ved hjelp av nøkler er imidlertid som ved normaliserte datamodeller, men normaliseringsprosessen går ikke like langt. Resultatet er færre tabeller og enklere spørringer mot databasen (Kimball & Ross, 2010). Pedagogisk kan det være hensiktsmessig å lære studentene forskjellene mellom normaliserte datamodeller og dimensjonale modeller ved først å utforme en

normalisert datamodell og deretter «denormalisere» modellen til en dimensjonal modell.

Oppgave 2 og 3 om utvikling og analyse av en datakube krevde at studentene skulle:

1. analysere oppgaven om barnehagekjeden, og finne ut hvilke data de trengte og langs hvilke dimensjoner data skulle analyseres,
2. vurdere innholdet i hver datakilde og finne ut hvordan dataseriene måtte sammenstilles logisk,
3. bruke Power Query til å
 - a. ekstrahere relevante data fra datakildene,
 - b. identifisere og/eller opprette primær- og fremmednøkler,
 - c. omforme data til et hensiktsmessig format, for eksempel konvertere data fra tekststrenger til numeriske verdier,
 - d. kombinere egenskaper fra flere kilder før data ble lastet inn i Power Pivot, og
 - e. finne hensiktsmessige navn til tabeller og kolonner, herunder tydelig angi nøkler i kuben,
4. presentere data i en datakube i Excel som vist i figur 3, og
5. utføre og dokumentere analysene i datakuben som spesifisert i oppgaven.

FIGUR 3 Eksempel på presentasjon av en datakube i Excel.



DATAINNSAMLING OG -ANALYSE

Innsamlingen av data ble påvirket av at oppgavene var knyttet til en hjemmeeksamen i et obligatorisk kurs, og at ikke alle sensorene hadde samme opplegg for å vurdere deloppgavene. Totalt 120 besvarelser ble innlevert. Sensorene for 70 besvarelser hadde brukt karakterene A–F for hver deloppgave, slik at karakterer for 70 besvarelser var relevante for vår undersøkelse. Av disse 70 hadde vi tilgang til 20 besvarelser med løsninger og argumentasjon.

Vi utførte først en Spearman-korrelasjonsanalyse av karakterene for oppgave 1 og 2. Vi brukte karakterene for oppgave 1 som et anslag for studentenes kunnskaper om datamodellering og karakterene for oppgave 2 som et anslag for om studentene kunne utvikle en konsistent datakubee.

Deretter analyserte vi de 20 besvarelsene inngående. Vi sorterte besvarelsene etter karakterer for oppgave 1 og undersøkte i hvilken grad svakheter ved besvarelsen av oppgaven kunne følges til besvarelsen av oppgave 2, det vil si utviklingen av datakuben. Til sist sjekket vi hvordan svakheter ved utviklingen av datakuben påvirket svarene på studentenes dataanalyser og deres anbefalinger til lederen av barnehagekjeden.

RESULTATER

FORSKNINGSPØRSMÅL 1

Vi fant en signifikant korrelasjon mellom karakterene for oppgave 1 og 2, som representerer henholdsvis kunnskaper om datamodellering og konsistens i datakuben ($R = 0,380, p < 0,001$). Tallene støtter antakelsen i begrepsmodellen (figur 2) om en sammenheng mellom studentenes kunnskaper om datamodellering og i hvilken grad de er i stand til å utvikle en konsistent datakubee.

FORSKNINGSPØRSMÅL 2

Av de 20 besvarelsene som vi analyserte i detalj, hadde tre besvarelser fått karakteren D+ eller lavere på datamodelleringsoppgaven. Disse studentgruppene hadde også fått karakteren D eller lavere på oppgaven som omhandlet utvikling av datakuben. De tre gruppene greide å utvikle en datakubee, og presentasjonen av analysene var tilnærmet lik eksemplet i figur 3. Resultatene av analysene var imidlertid feil, og studentene hadde ikke oppdaget det.

De resterende 17 besvarelsene ble brukt for å studere i hvilken grad misforståelser i datamodelleringsoppga-

ven kunne spores tilbake til utviklingen av datakuben. Vi fant at de fleste av studentene hadde den kunnskapen om datamodellering som var nødvendig for å kunne løse oppgaven.

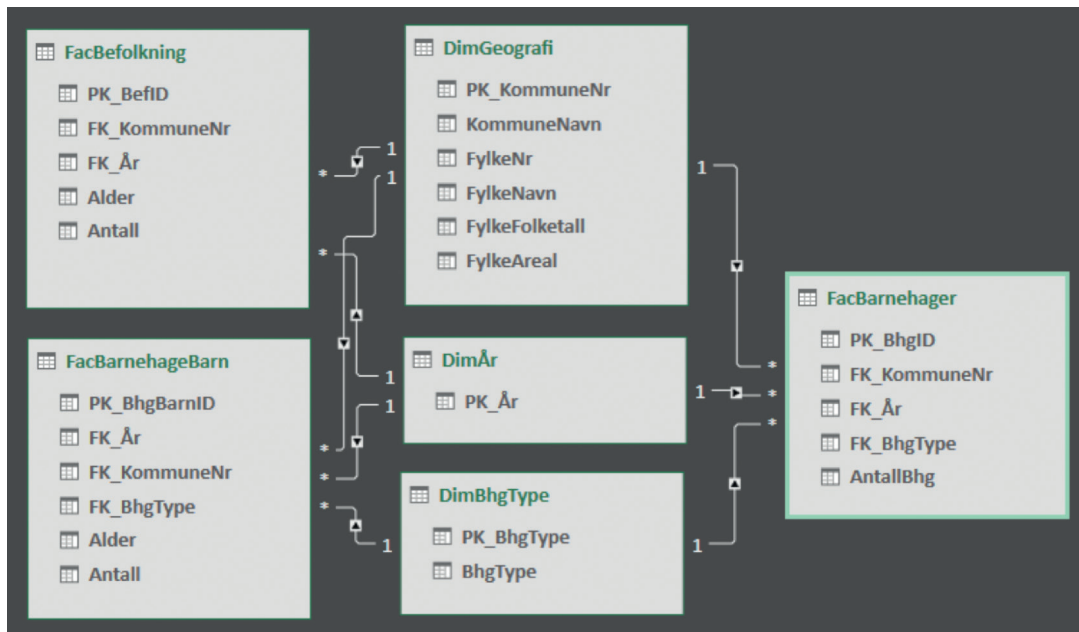
Studentenes problemer knyttet til løsning av oppgave 1 dreide seg i stor grad om at det ikke var samsvar mellom ER-diagrammet og tabelldefinisjonene, for eksempel ved at en kobling i ER-diagrammet ikke var representert ved en kobling av primær- og fremmednøkler i de tilsvarende tabelldefinisjonene. Flere feil av denne typen i samme besvarelse tyder på at studentene ikke helt har forstått prinsippene for datamodellering.

Elleve av de 70 studentgruppene, det vil si cirka 16 prosent, fikk karakteren A i oppgave 2. Av de 20 besvarelsene vi hadde tilgang til, var det imidlertid kun én gruppe som fikk A, se figur 4.

I det følgende vil vi bruke listen som ble presentert i metodeavsnittet til å illustrere utfordringer som de 16 studentgruppene som fikk karakterene B og C, hadde da de løste oppgave 2 og 3.

1. Få grupper hadde klart å presentere data i dimensjoner og faktatabeller, selv om prinsipper og eksempler på dimensjonal modellering hadde vært gjennomgått på kurset. De fleste studentene hadde likevel klart å definere primær- og fremmednøkler i tabellene, det vil si koblet tabellene slik at analyser utført på datakuben ga konsistente svar.
2. Andre utfordringer var knyttet til tolkningen av entitetene med tanke på å sammenligne data fra ulike kilder. I datakildene var for eksempel data om antall barn i barnehage samlet inn 31. desember hvert år, mens generelle befolkningsdata var samlet inn 1. januar. Studenter som ikke hadde oppdaget forskjellen i innsamlingstidspunkt eller tolket forskjellen på en hensiktsmessig måte, sammenlignet data for etterfølgende år i stedet for samme år. Resultatet var at analysene ga et feilaktig bilde av barnehagedekningen, det vil si antall barn i barnehage / antall barn.
3. De fleste studentene kunne bruke Power Query til å ekstrahere relevante data fra datakildene for å utføre de analysene de ble bedt om. Besvarelsene avslørte imidlertid noen tekniske utfordringer knyttet til klargjøring av dataene før de ble lastet inn i Power Pivot, for eksempel:
 - splitte en tekststreng og velge ut deler av en streng
 - konvertere en tekststreng til en numerisk verdi

FIGUR 4 Eksempel på en dimensjonal modell for en besvarelse som fikk god karakter.



- identifisere og/eller opprette primær- og fremmednøkler

Befolkningstall per fylke ble importert fra Wikipedia som en tekststreng, for eksempel '3&505&287198&287 198'. Her representerer tegnene 287198 befolkningstallet. Dette tallet finnes mellom det andre og tredje ampersand-symbolet (&). Disse tegnene måtte trekkes ut fra tekststrengen og konverteres til en tallverdi.

Enkelte studenter klarte ikke å opprette unike primær- og fremmednøkler slik at de kunne koble tabellene. Data om kommuner og fylker var for eksempel plassert i to forskjellige datakilder og måtte således kobles sammen i en geografidimensjon. Denne operasjonen krevde at studentene først definerte en felles egenskap i de to tabellene. Konsekvensen av ikke å koble de to tabellene var at studentene fikk problemer med å aggregere data per fylke i sine analyser.

Mange studenter hadde ikke navngitt tabeller og kolonner konsistent. Dermed fremgikk ikke nøklene tydelig når tabellene ble lastet inn i Power Pivot, og noen studenter fikk derfor problemer når de skulle opprette koblinger mellom tabellene. Dessuten ble

det uklart hvilke egenskaper de data som inngikk i datakuben, faktisk representerte.

4. De fleste studentene kunne lage en presentasjon tilsvarende eksemplet i figur 3. Tekniske utfordringer diskutert ovenfor gjorde det imidlertid vanskelig å utnytte det fulle potensialet i SSBI-verktøyene til å analysere data.
5. De alvorligste feilene var knyttet til kolonne- og radoperasjoner i Power Pivot. Få studenter gjorde slike feil, men feilene hadde omfattende konsekvenser for analysene. Videre så det ikke ut til at studentene oppdaget denne typen feil. Studentene kommenterte resultater som ikke virket logiske, som at det var flere barn i barnehager enn det var barn i kommunen. I stedet for å sjekke om det var feil i datakuben de hadde utviklet, eller i måten de brukte de analytiske verktøyene på, forsøkte de å finne argumenter som støttet resultatene av analysen. Et eksempel på slike argumenter var at enkelte barnehager sannsynligvis hadde mange barn fra andre kommuner. Studentene hadde tilsynelatende ikke sjekket sine resultater mot en delmengde av datasettet, for eksempel ved å sammenligne resultater fra Power Pivot med resultater fra velkjente funksjoner i Excel.

DISKUSJON OG KONKLUSJON

I vår undersøkelse har bachelorstudenter brukt SSBI-verktøyene Power Query og Power Pivot til å utvikle en datakube og utføre spesifikke analyser på datakuben.

Undersøkelsen viser at studentene ikke hadde problemer med å bruke Power Query til å ekstrahere data. Noen få studenter hadde tekniske utfordringer med å forberede data for Power Pivot, spesielt knyttet til håndtering av tekststrenger. De fleste studenter kunne bruke Power Pivot og Excel, det vil si at selv studenter uten grunnleggende kunnskaper i datamodellering fikk til en presentasjon i Excel som ligner på figur 3.

Vår undersøkelse har vist at sluttbrukere trenger grunnleggende kunnskaper om datamodellering for å kunne bruke SSBI-verktøy til å utvikle en datakube som kan danne grunnlag for valide analyser. Studenter som ikke hadde forstått hvordan man skal definere primær- og fremmednøkler mellom tabeller, fikk ikke etablert en konsistent datakube, og deres dataanalyser ga ikke korrekte resultater. Undersøkelsen støtter dermed argumentene av Stodder (2015) og Alpar og Schulz (2016) når det gjelder risikoene knyttet til bruk av SSBI-verktøy, men i stedet for å legge restriksjoner på brukerne mener vi at vi bør utdanne brukerne.

Undersøkelsen viser også at selv om få studenter klarte å organisere data i hensiktsmessige dimensjoner og faktatabeller, hadde de fleste av studentene ervervet kunnskaper om grunnleggende datamodellering, slik at de var i stand til å forberede data med nødvendige primær- og fremmednøkler før de lastet data inn i Power Pivot. Disse studentene fikk også valide analyser av behovet for nye barnehager.

Flere studentgrupper med feil i aggregeringene av og beregninger på data oppdaget tilsynelatende ikke feilene. Noen av gruppene kommenterte at resultatene virket ulogiske, men de forsøkte å finne argumenter som støttet deres resultater fremfor å teste om resultatene faktisk var valide.

Undersøkelsen har implikasjoner for undervisning i datafag. Selv om SSBI-verktøy som Power Query og Power Pivot er enkle å bruke rent teknisk, støtter verktøyene ikke brukerne med å utvikle en konsistent datakube. Uten kunnskaper om hvordan man skal sette opp en datakube, kan de kraftfulle operasjonene i kubene gi resultater som er forskjellige fra de resultatene som brukerne intenderte, og bidra til ineffektive beslutninger. Utviklingen av «brukervennlige» SSBI-arkitekturer og

-verktøy kan ikke erstatte undervisning i prinsipper for datamodellering og trening i anvendelse av prinsippene på utfordrende oppgaver.

I tillegg til undervisning i prinsipper for datamodellering bør undervisningen omfatte hvordan studentene skal analysere en oppgave *før* de begynner å utvikle datakuben, altså grunnleggende problemløsning. Dessuten bør studentene lære hvordan de sjekker at operasjonene på data gir resultater som intendert, for eksempel ved å trekke ut deler av datakuben og kontrollere aggregeringene/beregningene i Excel.

Treningen bør omfatte oppgaver som utfordrer studentene på logiske aspekter ved datakildene, for eksempel å finne ut hvordan data må tilpasses for å bli kompatible i kubene, jf. eksemplet med innsamlingsdatoer i hjemmeeksamen. Hertil kommer tekniske utfordringer som å konvertere tekstdata til numeriske verdier, plukke ut deler av en tekststreng og lignende. Hver trenings sesjon må følges opp av en debrifing, det vil si presentasjon av løsning av oppgaven, eventuelt alternative løsninger, og en diskusjon av studentenes erfaringer fra treningen, for å fremme læring (Savery, 2015).

Besvarelsene av hjemmeeksamen viser at kun få studenter hadde forstått hvordan de kan utvikle dimensjonale datamodeller. Vi ser at det er behov for å utvide undervisningen og øvingene her.

Undersøkelsen har begrensninger. Funnene er basert på detaljerte analyser av kun 20 besvarelser, og analysene er basert på *resultatene* av studentenes arbeid. Vi har ikke studert *hvordan* studentene har arbeidet med hjemmeeksamen, altså prosessene. Dessuten ga hjemmeeksamen studentene tilgang til nødvendige data, og alle datakilder var utførlig beskrevet i oppgaven. Oppgaven omfatter dermed ikke søking etter data og vurderinger knyttet til kvalitetssikring av data. Oppgaven omfattet heller ikke såkalt ustrukturerte data, det vil si data som ikke passer i en definert struktur.

Hovedinntrykket av undersøkelsen er at «brukervennlige» SSBI-verktøy ikke kan erstatte brukernes kunnskaper om datamodellering, men også at 16 timers undervisning og trening fikk de fleste av studentene på et nivå hvor de kunne bruke verktøyene effektivt for beslutningsstøtte. Forhåpentligvis har undervisningen også gikk studentene innsikt i egne begrensninger, slik at de vet når de bør søke hjelp hos dataspesialister for å tilrettelegge sitt datamateriale. M

REFERANSELISTE

- Abelló, A., Darmont, J., Etcheverry, L., Golfarelli, M., Mazón, J.-N., Naumann, F., ... Vossen, G. (2013). Fusion Cubes: Towards Self-Service Business Intelligence. *International Journal of Data Warehousing and Mining*, 9(2), 66–88.
- Alpar, P., & Schulz, M. (2016). Self-Service Business Intelligence. *Business & Information Systems Engineering*, 58(2), 151–156.
- Ask, U. (2013). Business Intelligence Practices: Adding Evidence from Organizations in the Nordic Countries. *International Journal of Business Intelligence Research*, 4(2), 1–18.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An Overview of Business Intelligence Technology. *Communications of the ACM*, 54(8), 88–98. doi: 10.1145/1978542.1978562.
- Chen, H., Chiang, R.H.L., & Storey, V.C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Devlin, B.A., & Murphy, P.T. (1988). An architecture for a business and information system. *IBM Systems Journal*, 27(1), 60–80.
- Fuglseth, A.M. (1989). *Beslutningsstøtte: metode for diagnose av lederes informasjons- og situasjonsoppfatninger*. Bergen: Norges Handelshøyskole.
- Fuglseth, A.M. (2005). Information systems and strategic resources. I A.M. Fuglseth & I.A. Kleppe (red.), *Anthology for Kjell Grønhaug in celebration of his 70th birthday* (s. 279–307). Bergen: Fagbokforlaget.
- Fuglseth, A.M., Håtuft, J.V., & Johannessen, T.V. (2010). *PC-bruk 2 for høyskoler og universiteter*. Bergen: Fagbokforlaget.
- Hastrup, T. (1983). *Latin-dansk ordbog*. København: Gyldendal.
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3. Utg.). Indianapolis: Wiley.
- Löser, A., Hueske, F., & Volker, M. (2008). Situational Business Intelligence. I M. Castellanos, U. Dayal, & T. Sellis (red.), *Business Intelligence for the Real-Time Enterprise*, 1–11. Berlin: Springer.
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution (cover story). *Harvard Business Review*, 90(10), 60–68.
- Orlikowski, W.J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, 3(3), 398–427.
- Pfeffer, J., & Salancik, G.R. (1978). *The External Control of Organizations*. New York: Harper & Row.
- Presthus, W. (2015). *Business Intelligence Utilisation through Bootstrapping and Adaptation*. Göteborg: University of Gothenburg.
- Savery, J.R. (2015). Overview of Problem-Based Learning: Definitions and Distinctions. I A. Walker, H. Leary, C.E. Hmelo-Silver, & P.A. Ertmer (red.), *Essential Readings in Problem-Based Learning: Exploring and Extending the Legacy of Howard S. Barrows* (s. 5–18). West Lafayette: Purdue University Press.
- Schlesinger, P.A., & Rahman, N. (2015). Self-Service Business Intelligence Resulting in Disruptive Technology. *Journal of Computer Information Systems*, 56(1), 11–21.
- Simon, H.A. (1960). *The new science of management decision*. New York: Harper.
- Stodder, D. (2015). *Visual Analytics for Making Smarter Decisions Faster*. TDWI Best Practices Report, 1–39.
- Varga, J., Romero, O., Pedersen, T.B., & Thomsen, C. (2014). Towards Next Generation BI Systems: The Analytical Metadata Challenge. I L. Bellatreche & M.K. Mohania (red.), *Data warehousing and knowledge discovery* (s. 89–101). Heidelberg: Springer.
- Watson, H.J. (2009). Tutorial: Business intelligence. – Past, present, and future. *Communications of the Association for Information Systems*, 25, 487–510.
- Zuboff, S. (1988). *In the age of the smart machine: The future of work and power*. New York: Basic Books.