



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2021 30 stp
Handelshøyskolen

Fra kundeavis til personlig kommunikasjon: en casestudie om datadrevet kundeinnsikt

From customer newspaper to personal
communication: a case study about data-driven
customer insight

Elise Løining & Tuva Marie Josefine Brynildsen
Master i økonomi og administrasjon – Business analytics

Forord

Denne oppgaven markerer avslutningen på vår master i økonomi og administrasjon med spesialiseringen business analytics ved Norges Miljø- og Biovitenskapelige Universitet (NMBU) våren 2021.

Covid 19 satte dessverre en stopper for mange sosiale og faglige aktiviteter vi hadde gledet oss til, men vi ser tilbake på to givende år hvor vi har lært enormt mye. Denne oppgaven har bidratt til å heve vår kompetanse på datadrevet innsikt, samt kunnskap innen customer relationship management. Oppgaven hadde aldri vært den samme uten hjelp fra Europris. Tusen takk for at dere har stilt opp til intervjuer, delt datasett med oss og alltid vært behjelpelige.

Mange andre har også vært involvert i denne prosessen og vi ønsker å takke alle sammen. Vi må gi en spesiell takk til vår veileder Mike Riess som har heiet på oss og vært til god hjelp gjennom hele perioden. Fellesveiledningene med Frode Alfnes har vært svært motiverende og har satt krav til kontinuerlig arbeid med oppgaven. Samboerne våre har også støttet oss i motgang og medgang, og det har ikke alltid vært en dans på roser med hjemmekontor på 60 kvadratmeter. Vi vil også takke våre arbeidsgivere som har gjort det mulig for oss å skrive denne oppgaven i kombinasjon med mye ansvar i våre deltidsjobber.

Avslutningsvis ønsker vi å takke hverandre for et godt samarbeid gjennom en rekke gruppeeksamener og nå til slutt masteroppgaven. Vi har til tider vært både frustrerte og lei, men vi har også hatt det mye morsomt og har utgjort et bra team.

God lesning!

Oslo, 01.06.2021

Elise Løining

Elise Løining

Tuva Brynildsen

Tuva Marie Josefine Brynildsen

Sammendrag

Datadrevet innsikt har for alvor blitt avgjørende for norske detaljforhandlere og skaper mange muligheter. Det store potensialet ligger nå i analyse av data og å skape verdi av innsikten.

Formålet med denne oppgaven er å utforske hvordan en norsk bedrift utnytter tilgangen de har på data, og hvilket potensial som ligger i datadrevet innsikt. Dette er en enkeltcasestudie av Europris ASA, hvor det har blitt gjennomført både kvalitative og kvantitative undersøkelser. Gjennom semistrukturerte intervjuer og analyse av transaksjonsdata indikerer resultatene at Europris er i startfasen av å benytte datadrevet innsikt til mer personlig kommunikasjon til kundene. Enkle automasjoner og testing av mer personaliserte nyhetsbrev med bakgrunn i historisk data er prosesser som er under utvikling.

Videre i studien illustreres mulighetene ved bruk av assosiasjonsregler for å skape et beslutningsgrunnlag som er basert på analyse av data. Ved benyttelse av assosiasjonsregler kan bedrifter få god innsikt i kundens kjøpsmønster. Dersom reglene kan predikere fremtidig atferd vil Europris kunne treffe kunden med mer relevant kommunikasjon. Assosiasjonsreglene kan også bidra til å løse mange av utfordringene som Europris har i forbindelse med personalisering av kommunikasjon. Avslutningsvis indikerer resultatene at det ligger et potensial for økt omsetning ved bruk av assosiasjonsregler.

Studiet konkluderer med at Europris har behov for å tilegne seg dypere kundeinnsikt, og at data mining kan være til god hjelp i denne prosessen. Videre kan assosiasjonsregler være en lønnsom metode for å løse flere av utfordringene som ble identifisert gjennom de kvalitative undersøkelsene.

Abstract

Data-driven insight is increasingly crucial for Norwegian retailers and creates many opportunities. The great potential now lies in the analysis of data and creating value out of the insight.

The purpose of this thesis is to study how one Norwegian company utilize their access to data and explore the potential value of analysing data. This study is a single-case study of Europris Norway, with both qualitative and quantitative data. Through semi-structured interviews and analysis of transactional data, the results indicates that Europris is in the initial phase of using data-driven insight for more personal communication to customers. They base their personal communication on a few simple automations and have started testing personalized newsletters based on historical data.

Furthermore, the study illustrates the possibility of using association rules to simplify decision-making based on analysis of data. By using association rules, companies can gain insight into the customer's buying pattern. If the rules can predict future behaviour, Europris will be able to meet customers with more relevant communication. The association rules can also help solve many of the challenges that Europris has in connection with personalization of communication. In conclusion, we see that there is great potential in using association rules to grow sales through the use of data-driven insight.

The study concludes that Europris needs to gain deeper insight to customer shopping patterns. The use of association rules can be a good first step for Europris regarding the challenges outlined in the qualitative surveys.

Innholdsfortegnelse

Forord	i
Sammendrag.....	ii
Abstract.....	iii
Figurliste	vii
Tabelliste	vii
1. Innledning.....	1
<i>1.1 Formål med oppgaven og problemstilling</i>	<i>2</i>
<i>1.2 Avgrensning.....</i>	<i>3</i>
<i>1.3 Struktur for oppgaven.....</i>	<i>4</i>
2. Europris	5
3. Teori og litteratur	7
<i>3.1 Customer relationship management.....</i>	<i>7</i>
3.1.1 Datadrevet markedsføring	7
3.1.2 Utdfordringer med datadrevet markedsføring	9
3.1.3 Personalisering.....	10
3.1.4 Kryss-salg	12
<i>3.2 Maskinlæring og personalisering.....</i>	<i>13</i>
3.2.1 Maskinlæring	14
3.2.2 Assosiasjonsregler	16
3.2.3 Klyngeanalyse	18
3.2.4 Collaborative filtering.....	20
4. Metode.....	21
<i>4.1 Forskningsdesign</i>	<i>21</i>

4.2 Datainnsamling	23
4.2.1 Litteratur	23
4.2.2 Primær data	24
4.2.3 Sekundær data	25
4.3 Datakilder	25
4.4 Datahåndtering og datahåndteringsplan	27
4.5 Dataanalyse av kvalitative intervjuer	27
4.6 Dataanalyse av transaksjonsdata - CRISP-DM	28
4.6.1 Forretningsforståelse	29
4.6.2 Dataforståelse	29
4.6.3 Dataforberedelser	30
4.6.4 Modellering	31
4.6.5 Evaluering	32
4.6.6 Deployment	32
4.7 Reliabilitet og validitet	33
5. Resultater	34
5.1 Personalisering og dataanalyse i Europris	34
5.1.1 Motivasjon for digitalisering	34
5.1.2 Nåværende bruk	35
5.1.3 Utfordringer	39
5.1.4 Fremtidig plan	40
5.2 Datadrevet innsikt	41
5.2.1 Modellutvelgelse	41
5.2.2 Assosiasjonsregler	43
5.2.3 Evaluering av regler	44
5.3 Forretningsverdi	48
5.3.1 Regler med høyest omsetning	48
5.3.2 Økning av consequent ved økning av antecedent	49

6. Diskusjon	52
6.1 <i>Bruk av datadrevet innsikt for personalisering.....</i>	52
6.2 <i>Utfordringer med datadrevet innsikt.....</i>	54
6.3 <i>Analyse av transaksjonsdata for personalisering</i>	55
6.3.1 <i>Modellvalg.....</i>	56
6.3.2 <i>Måling og evaluering av regler.....</i>	57
6.4 <i>Bruk av assosiasjonsregler i personalisering</i>	59
6.5 <i>Forretningsverdi.....</i>	62
7. Oppsummering.....	65
7.1 <i>Konklusjon.....</i>	65
7.2 <i>Videre forskning</i>	67
Litteraturliste	69
Vedlegg.....	74
<i>Vedlegg 1 – Intervjuguide</i>	74
<i>Vedlegg 2 – Oversikt over alle variabler i datasettet.....</i>	75
<i>Vedlegg 3 – Kundeavis for vår transaksjonsdataperiode</i>	76
<i>Vedlegg 4 – Informasjonsskriv NSD</i>	77
<i>Vedlegg 5 – Utrekninger</i>	80

Figurliste

Figur 1 - Strategikart Markedsføring (Peppers et al., 2016).....	11
Figur 2 - Utnyttelse av data ved bruk av kunstig intelligens (Huang og Rust, 2021).....	15
Figur 3 - Sannsynlighet for confidence, support og lift (Blattberg et al., 2008).....	17
Figur 4 - Faser for CRISP-DM modell (Chapman et al., 2000)	29
Figur 5 - Trappeløpsmodell	35
Figur 6 - Oversikt over automasjonsprosessen	37

Tabelliste

Tabell 1 - Strategisk rammeverk for KI i markedsføring (Huang og Rust, 2021).....	15
Tabell 2 - Oversikt over respondenter	26
Tabell 3 - Oversikt over informanter	27
Tabell 4 - Oversikt over aktiviteter for personalisert kommunikasjon	38
Tabell 5 - Oppsummering av modellutvalgelse	43
Tabell 6 - Topp 10 varer	44
Tabell 7 - Assosiasjoner med høyest support	45
Tabell 8 - Assosiasjonsregler med høyest confidence	46
Tabell 9 - Assosiasjoner med høyest lift.....	47
Tabell 10 - Kryss validering av regler	47
Tabell 11 - Varer med høyest omsetning.....	48
Tabell 12 - Gjennomsnittlig omsetning på assosiasjonsreglene	49
Tabell 13 - Økning av consequent ved økning av antecedent	50
Tabell 14 - Økning i omsetning ved bruk av confidence sammenlignet med expected confidence	50
Tabell 15 – Assosiasjonsregler av størst interesse	57

1. Innledning

Datadrevet innsikt får økende oppmerksomhet hos bedrifter rundt om i landet, og alle satser på digitalisering. I en rapport nylig publisert av Capgemini (2020) hevder de at datadrevet innsikt er avgjørende for at en bedrift skal lykkes, men de nordiske landene henger etter. Selskaper som behersker datadrevet innsikt, har store fordeler. Disse selskapene scorer høyere enn andre selskaper på kundetilfredshet, topplinjen, effektivitet og kostnadsbesparelser. Tillit til hvorvidt analyse av data kan føre til gode beslutninger er lavere i de nordiske landene, og dette resulterer i negativ påvirkning på bunnlinjen (Capgemini, 2020). Dersom nordiske selskaper ikke skaper verdi av dataene de innehar, vil internasjonale selskaper for alvor ta opp konkurransen i Norge.

Detaljhandelen har større tilgang til verdifull data enn tidligere. Det er helt nødvendig at denne blir benyttet til å øke bruken av datadrevet innsikt. Detaljhandelen har måttet tilpasse seg den digitaliserte verden, ved bare et tastetrykk har kunden produkter fra hele verden tilgjengelig. Dette setter helt nye krav til forhandlerne, og spesielt for de fysiske butikkene. En av hovedutfordringene er å få kunden til å komme tilbake, og fortsette å kjøpe varer av den samme forretningen (Zamil, Ahmad & Vasista, 2020). Denne relasjonsskapningen med kunden er noe av det viktigste for en bedrift i dag, og hvordan det kommuniseres med kundene blir derfor en sentral del av strategien til bedriftene (Anshari, Almunawar, Lim & Al-Mudimigh, 2019).

Innsikt i handlemønster kan være til stor hjelp for å forstå individuelle behov, og kan på den måten sørge for at kunden kommer tilbake til butikken. Samtidig må bedriftene håndtere en voksende mengde data og skape verdi ut av den. En måte å ta i bruk datadrevet innsikt, er å treffe kunder på en mer personalisert måte (Kumar, 2010). Analyse av data kan si noe om hvilke produkter som burde promoveres på et en-til-en nivå (Faridizadeh, Abdolvand & Rajae Harandi, 2018). Det har vist seg at å finne relevante produkter som skal være en del av kommunikasjon, kan være en kritisk utfordring for detaljforhandlere (Faridizadeh et al., 2018).

Datadrevet innsikt i forbindelse med personalisering av kommunikasjon i detaljhandel er derfor et spennende tema, og handler om å ta avgjørelser basert på dataanalyse og ikke på intuisjon (Provost & Fawcett, 2013). Denne oppgaven studerer hvordan Europris flytter fokuset fra massemarkedsføring til datadrevet innsikt og personlig kommunikasjon. I 2020 iverksatte de

prosjektet «personalisering». Prosjektet fokuserer på å treffe kunder med mer målrettet kommunikasjon basert på tidligere kjøp.

Europris, som mange andre detaljforhandlere i Norge, er i startfasen av å benytte seg av datadrevet innsikt i sin kommunikasjonsstrategi. Konkurransen i detaljhandel bransjen vokser, og kundene krever stadig mer av bedriftene. Det er ofte avgjørende å treffe kundene med personlig kommunikasjon og på denne måten å skape et kundeforhold. Personalisering blir i litteraturen introdusert som en mulighet for å øke andelen lojale kunder (Seng & Chen, 2010). Dagens muligheter er mange, teknologien er i stadig utvikling og tilgangen til data øker raskt. Dette resulterer også i utfordringer for bedriftene. Oppgavens praktiske relevans blir derfor å undersøke hvordan en detaljforhandler, ved bruk analyse av transaksjonsdata, kan skape større innsikt om kundene. På den måten kan de få et sterkere beslutningsgrunnlag sammenlignet med å ta valg på bakgrunn av intuisjon (Provost & Fawcett, 2013).

Denne studien er viktig av følgende grunner:

1. Den gir et overblikk over hvordan transaksjonsdata kan brukes for å skape innsikt om kundenes handlemønster.
2. Den gir bedrifter et beslutningsgrunnlag basert på analyse av data i stedet for intuisjon.
3. Den redegjør for hvilken forretningsverdi analyse av data kan ha for en bedrift.

1.1 Formål med oppgaven og problemstilling

Oppgavens formål er å studere hvordan en av de mest kjente detaljforhandlerne i Norge utnytter tilgangen de har på data. Studiet skal undersøke hvordan dataen kan skape kundeinnsikt ved å gjennomføre data mining, samt hvordan innsikten kan benyttes videre som virkemiddel i kommunikasjon med fokus på personalisering av nyhetsbrev og økt relevans når det gjelder produktpromotering.

Basert på oppgavens formål er det utformet tre forskningsspørsmål:

- *Hvordan jobber Europris med personalisering og datadrevet innsikt i dag, og hvilke utfordringer har de knyttet til dette?*
- *Hvordan kan analyse av transaksjonsdata hjelpe Europris å forbedre personalisering av kommunikasjonen til medlemmer av kundeklubben?*
- *Hva er den forventede forretningsverdien for Europris ved bruk av personalisering og datadrevet innsikt?*

Forskningsspørsmål 1 tar for seg hvordan Europris jobber med datadrevet innsikt og personalisering i dag og hvilke utfordringer de har rundt dette temaet. I forskningsspørsmål 2 benyttes det transaksjonsdata for å gjennomføre analyse av handlekurver. Her ligger fokuset på hvordan assosiasjonsregler kan gi innsikt i kundenes kjøpsmønster og hvordan Europris kan agere på denne. Det siste forskningsspørsmål omhandler hvilken verdi Europris kan forvente ved å agere på assosiasjonsregler, og hvordan Europris kan benytte seg av denne kunnskapen på mer rike datasett.

1.2 Avgrensning

Tid, ressurser og data var en avgjørende faktor for avgrensningene som ble gjort for studiets innhold. Oppgaven studerer kun en bedrift og fokuset er på én maskinlæringsmetode. Denne ble gjennomført på et mindre utvalg av data, da datakraften og ressursene som krevdes for å gjøre en mer dyptgående analyse over flere butikker og over en lenger periode ikke var tilgjengelig. Videre har personalisering blitt avgrenset til å bestå av kommunikasjonen i form av nyhetsbrev på epost til medlemmer av kundeklubben. Andre personaliseringstiltak som ikke er relevant for Europris i perioden de er i nå blir ikke lagt vekt på, men nevnt som fremtidige muligheter for bedriften.

Hovedfokus i denne oppgaven ligger ikke på kundens opplevelse av kommunikasjonen. Kunder har ikke blitt studert, og derfor er innsikten rundt hvilke tanker de har knyttet til mer personlige kommunikasjonen fra bedriften ikke et sentralt tema. Dette på grunn av tidsperspektivet på denne oppgaven, ved en større oppgave ville det vært interessant å undersøke.

1.3 Struktur for oppgaven

Dette studiet er delt inn i syv kapitler inkludert innledning. Kapittel 2 omhandler bedriften Europris. I dette kapitlet legges generell informasjon om Europris frem og deres digitaliseringsplaner basert på blant annet årsrapporter slik at leser skal få en oversikt over bedriften.

I kapittel 3 blir det presentert relevant litteratur og teori innenfor CRM, datadrevet innsikt, utfordringer, personalisering, kryss-salg og diverse maskinlæringsmodeller som resten av oppgaven bygger på. Videre blir metodene som er anvendt for å besvare våre forskningsspørsmål lagt frem i kapittel 4. I de to neste kapitlene vil først resultatene fra de kvalitative og kvantitative undersøkelsene legges frem i kapittel 5, for deretter og diskuterer funnene med bakgrunn i relevant teori og litteratur i kapittel 6. Til slutt legges konklusjon, implikasjoner og sterke og svake sider med studiet frem, før det introduseres forslag til videre forskning i kapittel 7.

2. Europris

Europris er et lavprisselskap med over 260 butikker rundt om i landet, og er dermed den største forhandleren innenfor sin kategori (Europris, 2020), og et av de mest kjente selskapene i Norge (Europris, u.å.). Selskapet tilbyr et stort utvalg av produkter innenfor et bredt kategorispekter, med alt fra matvarer og møbler til tur- og fritidsutstyr.

Selskapet ble stiftet i Stavanger i 1992 av Wiggo Erichsen. I ettertid har blant annet gründeren Terje Høili kjøpt seg inn i kjeden, hvor han var med på å gjøre merkevaren kjent for Norges befolkning (Europris, u.å.). Videre ble Europris notert på Oslo børs i 2015, og har i lang tid vokst raskere enn det gjennomsnittlige markedet (Europris, 2019). Koronapandemien som ble en realitet i Norge i mars 2020, har resultert i stor vekst for Europris (Lorvik, 2021). Totalsalg i 2020 endte på 8,388 millioner kroner med mer enn 36 millioner kundetransaksjoner, og ble dermed Europris sitt økonomiske beste år (Europris, 2020).

Samtidig øker konkurransen i markedet. I en rapport publisert av Menon economics legges det frem at Europris ikke bare konkurrerer mot lignende aktører, men også med ordinære dagligvare butikker på både matvarer og andre produkter (Wifstad, Jenssen, Eide, Grünfeld & Skogli, 2018). Europris må gjøre tiltak for å differensiere seg fra disse på både på pris og produkter (Europris, 2019). Dette gjenspeiles i Europris sine viktigste strategiske tiltak som omhandler å styrke pris- og kostnadsposisjonen, forbedre kundeopplevelsen og drive kundevekst (Europris, 2019).

Digitalisering og MER-kundeklubb

Endring i handlemønster hos dagens kunder er en av faktorene til at Europris har satt i gang store endringer når det gjelder digitalisering (Europris, 2019). I 2018 lanserte de sin e-handel-løsning med muligheter for både hjemlevering av produkter, og klikk og hent i butikk. De forklarer videre i sin årsrapport for 2019 at de har satt i gang en omfattende digitaliseringsprosess i alle deler av verdikjeden, men hovedfokuset skal være på kundeopplevelsen (Europris, 2019).

Europris sitt kundelojalitetsprogram “MER” er også i stadig vekst, og per 31. desember 2020 hadde kundeklubben 690.000 medlemmer (Europris, 2020). Ved at Europris stadig øker antall medlemmer av kundekubben vil dette også resultere i økt datamengde og nye CRM løsninger med

flere muligheter for personalisering gjennom kommunikasjon. Markedsføringen er i større grad enn tidligere et satsningsområde på digitale kanaler, og er en strategi for å tiltrekke flere kunder både til nettbutikk og fysiske butikker. Kommunikasjonen vil bli mer personalisert fremover, og allerede er det på plass et system som kan bidra til dette (Europris, 2020).

Ifølge Europris posisjonerer de seg som en omnichannel detaljforhandler (Europris, 2020). Omnichannel detaljforhandler kan forklares som “en integrert salgsopplevelse som kombinerer fordelene ved fysiske butikker med den informasjonsrike opplevelsen av online shopping” (Rigby, 2011). Europris ønsker at kundene skal kunne finne den informasjonen de trenger online, og deretter handle ut fra egne preferanser, enten det gjelder fysiske butikker eller via nettbutikk som i dag står for ca. 1% av totalt salg (Europris, 2020).

3. Teori og litteratur

For å kunne besvare forskningsspørsmålene er det hensiktsmessig å beskrive customer relationship management og hva dette innebærer. Innenfor customer relationship management vil det være datadrevet innsikt, personalisering og kryss-salg som er spesielt relevant for vår studie. Det vil også bli lagt frem ulike maskinlæringsmodeller som kan benyttes i forbindelse med personalisering og kryss-salg. Dette innebærer assosiasjonsregler, klyngeanalyse og collaborative filtering.

3.1 Customer relationship management

Det finnes en rekke definisjoner av customer relationship management (CRM). I denne oppgaven ses CRM på som «en forretningsstrategi som maksimerer lønnsomhet, inntekter og kundetilfredshet ved å organisere seg rundt kundesegmenter, fremme atferd som tilfredsstiller kunder og implementere kundesentriske prosesser.» (Buttle & Maklan, 2015, s. 4). Mange forbinder akronymet CRM med CRM-systemer. Det er fagområdet CRM som er av betydning i denne oppgaven, og ikke CRM-systemer.

Det å skape et forhold og håndtere relasjoner med sine kunder blir sett på som en kritisk suksessfaktor. Ifølge Buttle og Maklan (2015) vil bedriften oppnå bedre resultater om de klarer å drive god kunderelasjonsledelse og tilfredsstille kundene. Dermed vil de beholde sine mest lønnsomme kunder. Ved å beholde sine lønnsomme kunder, vil bedriften i det lange løp også få en større kundebase sammenlignet med de som kun fokuserer på å rekruttere nye kunder. Kumar (2010) trekker frem at lønnsom administrering av kunder på lang sikt er noe bedriften ikke kommer unna i dagens marked, både med tanke på konkurranse og økonomi. I tillegg er det kontinuerlige endringer i dagens marked spesielt med tanke på den overveldende tilgangen til blant annet data og teknologi.

3.1.1 Datadrevet markedsføring

Et viktig strategisk grep innen CRM de siste årene har vært lojalitetsprogrammer hvor bedrifter har fått svært god tilgang til data om kundene sine. Denne dataen i seg selv gir liten verdi, men det er generering av informasjon og innsikt på grunnlag av disse dataene som er avgjørende for å lykkes med å forstå kunden (Buttkus & Eberenz, 2019). I denne oppgaven brukes begrepet datadrevet innsikt om alle metoder og verktøy som kan bidra til å ta beslutninger basert på analyse

av data (Nazarov, 2019). Dette inkluderer blant annet statistiske teknikker, data mining, maskinlæring, kunstig intelligens og analyse av big data.

Det finnes mange ulike begreper som omfavner analyse av data i et kunderelasjonsperspektiv. Buttle og Maklan (2015) kaller denne prosessen for analytisk CRM og definerer det som «Analytisk CRM fokuserer på utvinning av kunderelaterte data for strategisk eller taktiske formål». Det vil si at analytisk CRM «samler, lagrer, henter ut, integrerer, prosesserer, forstår, distribuerer, bruker og rapporterer kundedata for både å øke kundenes og bedriftens verdi». Blattberg, Kim og Neslin (2008, s. 4) bruker begrepet «Database Marketing» og definerer det som «bruk av kundedatabaser for å øke markedsføringsproduktiviteten gjennom mer effektiv anskaffelse, oppbevaring og utvikling av kunder». Nazarov (2019, s. 13) bruker begrepet data drevet markedsføring som innebærer at beslutninger blir tatt med utgangspunkt i dataanalyse.

Videre har dataanalyse blitt en viktigere del av CRM-arbeid de seneste årene. Bedrifter jobber konstant med å innhente informasjon om atferden til kundene sine. Det kan være salgsdata, kjøpshistorikk, betalingshistorikk og markedsføringsdata. I tillegg kan det hentes inn eksterne datakilder som kan gi demografisk-, geografisk- og livstilinformasjon om kundene. Dette vil kunne være med på å identifisere ulike kundegrupper eller segmenter blant kundemassen, og gi en mer kompleks forståelse av kunden. Videre kan dette resultere i opparbeidelse av kunnskap om blant annet kundelojalitet (Buttle & Maklan, 2015). I tillegg kan datadrevet innsikt også resultere i økt kundetilfredshet og forbedret konkurranse og økonomisk ytelse (Ayyagari, 2021). Ved å ha denne kunnskapen vil bedriften kunne være mer målrettet i sin fremgangsmåte til salg, markedsføring og kommunikasjon, som vil øke sannsynligheten for at kunden benytter seg av tilbudet og dermed øke verdien av aktivitetene (Buttle & Maklan, 2015).

Blattberg et al. (2008) viser til noen fundamentale motivasjoner for hvorfor et firma skal bruke ressurser på datadrevet innsikt. Det trekkes frem økt markedsproduktivitet, skape og øke kunderelasjoner og skape bærekraftig konkurransefortrinn. Ifølge Liu (2015) er endringene i handlemønster en av hovedgrunnene til at bedrifter må endre til en mer analytisk tenkemåte når det gjelder CRM. Dagens datatilgang, muligheten til å innhente og benytte seg av denne, i tillegg til den stadige utviklingen innenfor teknologi gir dermed mange nye muligheter for å nå sine

kunder. Samtidig setter kundene større krav til bedriftene, og ønsker at bedriften er til stede på ulike enheter, med rett produkt, uansett tid og sted gjennom blant annet personalisering. Videre viser Blattberg et al. (2008) til motivasjonen om at datadrevet innsikt skaper og forbedrer kundeforhold. Å ha en kunderelasjonsstrategi er viktig av flere grunner, men strategien er lagt til grunn med troen på at det å beholde kunder er billigere enn å anskaffe nye, og å øke bevaringen av kunder er mer verdifullt enn å øke anskaffelsen av nye.

3.1.2 utfordringer med datadrevet markedsføring

Ragavi, Srinithi og Anitha Sofia (2018) har i sin studie listet opp en rekke utfordringer ved datadrevet innsikt for kunnskapsheving. Noen av utfordringene som nevnes er datastørrelse, data fra ulike datakilder, tolkning av resultater, presentasjon og visualiseringen av resultatene. Det menneskelige aspektet ved analysen med tanke på hvordan brukeren inkorporerer bakgrunnskunnskap inn i dataanalyse kan også være en utfordring. I tillegg til forståelsen og visualiseringen av resultatene av analysen. Dette punktet er avgjørende for å få verdi ut av analysen av data.

For å kunne benytte seg av datatilgangen vil verktøy for datadrevet markedsføring bli benyttet, dette kan skape problemer for ikke-tekniske beslutningstakere. Domene- og kommunikasjonsutfordringer mellom de ulike fagområdene er derfor en sentral del av utfordringene bedriften står ovenfor (Ayyagari, 2021). Ifølge Verhoef, Kooge og Walk (2016) er det også en fare for at bedrifter tenker for stort fra start når de skal begynne å utnytte dataen ved at kunnskapen rundt temaet ikke strekker til og resultatet av analyser dermed ikke blir verdiskapende. Det må foreligge en klar plan for prosessene før man setter i gang.

Videre Bradlow, Gangwar, Kopalle og Voleti (2017) peker på spesielt to statistiske hovedutfordringer tilknyttet big data-analyse innen detaljhandel. Den første omhandler datakomprimering som innebærer å gjøre datasettet mindre. Når volumet av data blir større, blir arbeidet med å analysere dette også større. Derfor vil det være nødvendig å komprimere dataen. Komprimering av data kan gjøres teknisk ved hjelp av komprimering av filformatet eller funksjonelt som komprimerer datavolumet ved hjelp av økonometriske verktøy. Den andre hovedutfordringen er Bayesian inference. Dette betyr at man har mye data, men at det trolig ikke

finnes fullkommen data på individnivå på alle kunder. Dette kan løses gjennom Bayesiske metoder som låner data fra andre kunde profiler som har rikelig informasjon for å kunne fullføre analyser som har behov for fullstendig informasjon.

Mange av de samme utfordringene oppleves når det skal implementere større og mer informative datasett fra flere kilder. Analyseprosessen av disse kildene består av mange steg, fra datainnsamling og rengjøring av data til datamodellering, tolkning og analyse (Nasser & Tariq, 2015). Disse Big data-utfordringene kan deles inn i tre hovedkategorier; data, prosess og ledelsesutfordringer. Disse kategoriene tar for seg dataens egenskaper, utfordringer i behandlingen av dataen og de etiske og juridiske spørsmålene rundt dataen og databehandlingen. Da det er et høyt volum data som blir lagret fra forskjellige datakilder vil disse komme i ulike typer og formater. Dette skaper igjen utfordringer når dataen skal renses og analyseres. Samtidig er det helt avgjørende at bedriften tenker på personvern og GDPR i behandlingen av dataen. Det er mange bedrifter i dag som har tilgang til store mengder data de ikke utnytter da dataen fortsatt er rå, ubehandlet og ustrukturert. Samtidig kan utnyttelsen av denne dataen skape verdier i form av kostnadsreduksjon, forbedring av beslutningstaking, produkter og tjenester (Nasser & Tariq, 2015).

Ayyagari (2021) legger også frem utfordringer knyttet til å sikre personvern og det juridiske ved innsamling og bruk av data. Ved blant annet transaksjonsdata blir ofte kundens kjøp registrert med produkt, tid, sted og kvantitet, dette kan igjen benyttes til å skape en kunde profil som kan brukes til personalisering av kommunikasjonen ut mot kunden. I denne fasen er det viktig å være bevisst på at dette kan være noe kunden ikke er klar over, og kan føle personaliseringen som inntrenging i privatlivet (Ayyagari, 2021).

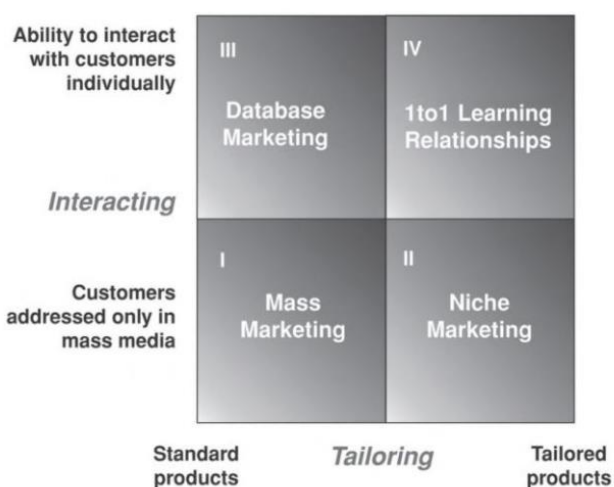
3.1.3 Personalisering

Det å benytte seg av kunde- og transaksjonsdata i beslutningsprosesser blir stadig mer populært. Wrzuszczak-Noga (2018), samt flere forskere og eksperter, mener det er mer gunstig å holde på de faste kundene og derfor introdusere personalisering på bakgrunn av den nåværende etterspørselen. Dette er et strategisk verktøy som kan benyttes for å differensiere seg selv fra stadig større konkurranse (Kwon & Kim, 2012). Videre forklarer Wedel og Kannan (2016) at ved å

benytte personalisering vil man bedre kunne treffe individuelle behov, som vil kunne resultere i høyere kundelojalitet og langsiktig lønnsomhet.

Personalisering som begrep har ulik betydning for forskjellige personer og fagområder. Det kan være vanskelig å sammenligne studier med hverandre på grunn av dette (Kwon & Kim, 2012). Fan og Poole (2006, s. 183) definerer personalisering som «en prosess som endrer funksjonalitet, grensesnitt, informasjonstilgang, innhold og særpreg til et system for å øke personlig relevans for et individ eller gruppe av individer». Personalisering kan også være det punktet i strategien hvor bedriften bestemmer seg for hvilken markedsføringsmiks som er best egnet for en enkelt kunde, ofte basert på innsamlet kundedata (Arora et al., 2008). Ved bruk av personalisering er det viktig å se på hvilke kunder som skal fokuseres på. Bedriften må derfor ha nok informasjon om kunden og kundens preferanser må komme tydelig frem i dataen (Kwon & Kim, 2012), for å treffe rett med personaliseringstiltakene.

For å illustrere ulike strategier for personalisering av kommunikasjon er figur 1 informativ. Dette er et strategikart som først ble presentert i 1997, men viser fortsatt en relevant oversikt over hvordan man kan tilnærme seg kunden.



Figur 1 - Strategikart Markedsføring (Peppers et al., 2016)

Mange bedrifter befinner seg i kvadranten nederst til venstre, hvor alle får den samme kommunikasjonen og de samme produktene. De siste årene har flere begynt å bevege seg mot det de kaller «Database marketing». I dette feltet er det i mye større grad interaksjon med kunden, men de tilbyr fortsatt standardiserte produkter. Mange ønsker å bevege seg helt til kvadranten øverst i høyre, hvor det er en en-til-en lærende relasjon mellom bedriften og kunden (Peppers, Rogers & Kotler, 2016).

3.1.4 Kryss-salg

I forrige kapitlet ble det presentert litteratur innenfor feltet personalisering. For å oppnå personalisering i form av å kommunisere mer relevante produkter til kunden, kan kryss-salg være et godt verktøy å benytte. Kryss-salg kan være en verdifull CRM-strategi for økt salg og kundelojalitet, samtidig som det er kostnadsbesparende å treffe målrettet i kommunikasjonen med de rette anbefalingene til de rette kundene, slik at kundenes interesse opprettholdes (Zhang, Priestley, DeMaio, Ni & Tian, 2021). Det er viktig for bedriften å opprettholde dette, og skape gode relasjoner med sine kunder da det koster fem ganger mindre å betjene eksisterende kunder enn å skaffe nye (Kamakura, 2008). Zhang et al. (2021) forklarer kryss-salg som praksisen med å oppdage eksisterende kunders kjøpemønster og engasjere de med tilleggsprodukter eller tjenester. Seng og Chen (2010) legger frem tre fordeler ved å benytte kryss-salg, hvor den første fordelen viser til økt salg og profitt ved å anbefale produkter til kunder som de har høyere sannsynlighet for å kjøpe. Den andre fordelen som nevnes i studiet er reduserte kostnader ved å ha en mer selektiv målretting. Den siste fordelen som nevnes er ved at kunden kjøper flere varer vil det forme en lojal kunde.

Videre foreslår Blattberg et al. (2008) to ulike kryss-salgs strategier: produkt- og kundesentrisk. Produktsentrisk vil si at man tar utgangspunkt i produktet, og målrettet treffer kunder som har høy sannsynlighet for å kjøpe produktet. Kundesentrisk tar utgangspunkt i kunden og hva den mest sannsynlig kommer til å kjøpe. Det finnes både ulemper og fordeler med begge, hvor den produktsentriske metoden muliggjør fullt fokus på varer med høy profitt og stordriftsfordeler. Kundesentriske kan bidra til enda bedre treffsikkerhet når det gjelder produktanbefalinger og kan derfor bidra til å maksimere salget.

Kryss-salg er som nevnt en god metode for å bidra til økt omsetning, men den har sine utfordringer. Kamakura (2008) har identifisert noen av disse utfordringer. I studiet nevnes det at det må legges en strategi for hvor mye og hvor ofte kunden skal treffes med kryss-salg for at interessen ikke skal bli nedskalert. I tillegg er det viktig med en database som innhenter informasjon om kundene og deres kjøp, som er tilgjengelige gjennom organisasjonen. Det er også en utfordring å vurdere hvilke produkter som passer i kryss-salg, og at ikke alle kunder ser på assosiasjonen mellom to eller flere produkter på samme måte (Hoanca & Mock, 2011). Et eksempel kan være ulikt syn mellom menn og kvinner i handlemønsteret.

Når det gjelder hvilke produkter som skal kryss-selges nevner også Blattberg et al. (2008) dette som en av hovedutfordringene med metoden og legger frem en prediktiv modell som kalles «Next product to buy». Hensikten er å predikere hva kunden mest sannsynlig kjøper neste gang. På den måten kan man prøve å kryss-selge dette produktet. En av utfordringene med denne metoden er at det er naturlig å tenke at dette ikke skaper verdi, ettersom kunden uansett ville kjøpt dette produktet. Poenget er å se på kunder med samme profil, og dermed kryss-selge til noen som har høy sannsynlighet for at også kjøper produktet.

Dagens tilnærming til kryss-salg baseres ofte på analyse av data. Ifølge Blattberg et al. (2008) finnes det flere ulike modeller som kan benyttes for å skaffe informasjonen som trengs. Boken nevner blant annet collaborative filtering, klynge analyse og assosiasjonsregler som mye brukte modeller. Disse vil bli beskrevet i punkt 3.2.

3.2 Maskinlæring og personalisering

Dybdeforståelse om kundenes preferanser er grunnlaget for enhver forretningsmodell. Analyse av data kan avdekke verdifull informasjon om preferansene til kundene (Buttkus & Eberenz, 2019). Med bakgrunn i historiske data kan maskinlæringsmodeller gi kunden kommunikasjon basert på hva modellen anbefaler. Anbefalingssystemer kan være en del av personalisering for kunden. Det er spesielt tre maskinlæringsmetoder som utpeker seg til bruk for å anbefale produkter: assosiasjonsregler, klyngeanalyse og collaborative filtering (Han & Kamber, 2006). Ifølge Blattberg et al. (2008) er det et stort potensial for å personalisere alle aspekter av alle markedsførings- og kommunikasjonstiltak ved å benytte disse modellene. I dette kapitlet vil først

maskinlæring bli definert og hvordan dette kan benyttes til personalisering av kommunikasjon, for deretter å gå mer i dybden på de tre nevnte metodene.

3.2.1 Maskinlæring

Maskinlæring kan defineres som automatiserte prosesser som henter ut mønster fra data (Kelleher, Mac Namee & D'Arcy, 2015). Datamaskiner blir programmert på en slik måte at de kan lære av inputen som er tilgjengelig (Shalev-Shwartz & Ben-David, 2014). Maskinlæringsmodeller kan deles inn i supervised og unsupervised learning. Supervised modeller, også kalt prediktive, har til hensikt å predikere et fremtidig utfall (Buttle & Maklan, 2015). Prediktive modeller består av en rekke predikatorer, som er variable faktorer som det er sannsynlig at vil påvirke fremtidig atferd eller resultater. Her ønsker man å si noe om «hva som vil skje i fremtiden». Eksempel på slike modeller er logistisk regresjon og klassifisering (Gangurde, Kumar & Gore, 2017). Unsupervised modeller har som mål å finne mønster og assosiasjoner blant dataen (Hastie, Tibshirani & Friedman, 2009). Oppgaven er å utforske dataen og finne relasjoner uten å predikere nye egenskaper (Gangurde et al., 2017). Disse kan også bli kalt deskriptive modeller og har til hensikt å beskrive og utforske dataen og besvarer spørsmålet «Hva skjedde?». Eksempler på slike modeller er klyngeanalyse og assosiasjonsregler (Buttkus & Eberenz, 2019).

Et steg videre fra deskriptiv og prediktive modeller, finnes preskriptive modeller. Kunstig intelligens blir ofte kalt preskriptiv (Buttkus & Eberenz, 2019). Her ønskes det å se på hvordan man kan få ting til å skje. Huang og Rust (2021) legger frem et rammeverk for utnyttelse av data ved bruk av kunstig intelligens (KI) i markedsføringskontekst. I Figur 2 deles kunstig intelligens inn i mekanisk, tenkende og følende. Mekanisk KI brukes for å automatisere oppgaver som er rutinepreget og repetitive (Huang & Rust, 2021). Prosessering av data som gir nye konklusjoner eller beslutninger kalles tenkende KI. Denne type KI gjør det mulig å identifisere mønster i dataen, som kan benyttes til blant annet anbefalingssystemer eller kryss-salg (Huang & Rust, 2021). Følende KI innebærer interaksjon med menneske for analyse av menneskelige følelser eller emosjoner. For eksempel sentiment analyse, naturlig språkbehandling og tekst-til-tale-teknologi (Huang & Rust, 2021). Følende KI gir fordeler når det kommer til personlige relasjoner da den reagerer på følelser.



Figur 2 - Utnyttelse av data ved bruk av kunstig intelligens (Huang og Rust, 2021)

Med bakgrunn i fordelingen over, har Huang og Rust (2021) lagt frem noen alternativer på hvordan dette kan brukes i dagens markedsføringsmiks. Dette er også gjentakende i andre studier, blant annet har Jarek og Mazurek (2019) hvor resultatene fra undersøkelsen som ble gjennomført for å avdekke hvilke implikasjoner KI hadde på markedsføringsmiksen viste at det hadde implikasjoner på både produkt, pris, promotering og plassering i form av automasjoner, personalisering av kommunikasjon og anbefalinger. Tabell 1 viser et strategisk rammeverk for KI i markedsføring.

Tabell 1 - Strategisk rammeverk for KI i markedsføring (Huang og Rust, 2021)

Kunstig intelligens (KI)	Mekanisk KI	Tenkende KI	Følende KI
<i>Strategiske beslutninger</i>			
<i>Markedsundersøkelse</i>	Data innsamling	Markeds analyse	Kundeforståelse
<i>Markedsstrategi</i>	Segmentering	Målretting	Posisjonering
<i>Markedsaktiviteter (4P'er, 4C'er)</i>	Standardisering	Personalisering	Relasjonsbygging
- <i>Produkt/ Kunde</i>	Automatiserer prosesser og resultater av møte med kundens behov og ønsker	Personalisere produkter basert på kunde preferanser	Forstå og møte kundens følelsesmessig behov og ønsker
- <i>Pris/ Kostnad</i>	Automatiserer prosessen av prissetting og betaling	Personalisere priser basert på betalingsvillighet	Forhandle pris og justere kostnader interaktivt
- <i>Plassering /Lettvinhet</i>	Automatiserer kundetilgangen til produkter	Personalisering av interaksjon	Personalisere opplevelsen for kundeengasjement
- <i>Promotering/ Kommunikasjon</i>	Automatiserer kommunikasjonen med kunder	Tilpasset promotering av innhold for personlig kommunikasjon	Skreddersy kommunikasjon basert på kundens følelsesmessig preferanser og reaksjoner

I tillegg har Huang og Rust (2021) skissert hvordan disse tiltakene benyttes i dag og hvilke praksiser som er på vei. Her trekker forfatterne både frem markedsfører- og kundens side. Hvor noen av de fremvoksende praksisene er å treffe kunder mer målrettet på individnivå, benytte mer automasjoner for prosesser som innsamling av data og kombinasjon av datakilder og å danne mer personlige handleopplevelser Huang og Rust (2021).

3.2.2 Assosiasjonsregler

Assosiasjonsregler, også kalt market basket analyse (MBA), er en deskriptiv maskinlæringsmodell som kan benyttes til å analysere transaksjonsdata. Assosiasjonsregler kan bli brukt til å se hvilke produkter kunden ofte kjøper sammen. Algoritmen betegnes som en rask og nyttig modell for atferdsdeteksjon som kan benyttes til å komme med anbefalinger (Wrzuszczak-Noga, 2018). Hovedmålet er å søke etter mønster som tidligere ikke var kjent (Olson & Lauhoff, 2019). Modellen starter med å kategorisere kundens kjøpsadferd. Deretter identifiseres informasjon som kan gi fortjeneste. Til slutt brukes denne informasjonen og innsikten til å ta beslutninger (Olson & Lauhoff, 2019). Analysen passer godt til problemstillinger hvor det finnes mye data uten noen spesielle hypoteser å teste, og hvor det først og fremst ønskes å se på mønster (Blattberg et al., 2008). I tillegg til å se på hvilke produkter som ofte kjøpes sammen, kan det også kartlegges kundens kjøpsadferd ut fra tid på dagen, dersom en tidsvariabel er tilgjengelig (Gangurde et al., 2017). MBA blir sett på som en lett gjennomførbar og lett forståelig analyse for flere fagområder, teknikken er derfor mye brukt innen detaljhandel og av markedsføringsstrategier (Blattberg et al., 2008).

Ved gjennomføring av MBA benyttes ofte apriori algoritmen (Buttkus & Eberenz, 2019). Algoritmen danner regler, og det er ofte tilfellet at man ender opp med et høyt antall av disse, men man er kun interessert i de reglene som faktisk er interessante og gir god innsikt. Reglene blir lagt frem i form av «HVIS antecedent SÅ consequent», hvor både antecedent og consequent kan inneholde flere produkter (Zamil et al., 2020).

Videre blir det i litteraturen fremhevet tre kriterier for evalueringen av MBA: support, confidence og lift (Blattberg et al., 2008; Gangurde et al., 2017). Support måler hvor stor prosentandel en kombinasjon av varer utgjør i forhold til det totale antallet transaksjoner. Høy support betyr at

kombinasjonen av varer forekommer ofte. Dette sier noe om populariteten til produktene eller kombinasjonen av produktene, support kan derfor være med på å identifisere hva som kan skape trafikk til butikk eller nettside (Gangurde et al., 2017).

Det andre parametere er confidence som er sannsynligheten for at en handlekurv med antecedent på venstre side inneholder consequent på høyre side. Jo høyere confidence jo større avkastningsrate kan man forvente av en regel. Confidence kan benyttes til blant annet vareplassering eller kommunikasjon til kunden (Gangurde et al., 2017).

Til slutt er det lift som sier noe om styrken av tilknytningen mellom produktene på høyre og venstre side. Lift skal være med på å støtte confidence og support på deres svake sider. En lift høyere enn 1,0 indikerer at en handlekurv som inneholder produkt B, pleier å inneholde produkt A oftere enn handlekurver som ikke inneholder produkt B (Gangurde et al., 2017). Ved gjennomføring av assosiasjonsregler ønskes det at kriteriene oppnår så høye verdier som mulig, jo høyere verdier på support, confidence og lift, jo mer interessant blir regelen som er presentert (Blattberg et al., 2008). Figur 3 viser sannsynligheten for de tre ulike målene.

$$\begin{aligned}\text{Confidence} &= P(B|A) \\ \text{Support} &= P(BA) \\ \text{Lift} &= P(B|A)/P(B)\end{aligned}$$

Figur 3 - Sannsynlighet for confidence, support og lift (Blattberg et al., 2008)

Noen av utfordringene med Assosiasjonsregler er selve datasettet som benyttes. Transaksjonsdata kan ofte komme fra bedrifter som selger hundrevis av forskjellige varer, i tillegg til at det ofte er mange transaksjoner hver dag med lite i hver handlekurv (Olson & Lauhoff, 2019). Resultatet kan derfor bli store mengder regler å velge mellom etter analysen, og det kan være lønnsomt å differensiere analysen på eksempelvis ulike sesonger, ulike butikker og ulike kundeprofiler (Gangurde et al., 2017). Dette støttes av Zhang et al. (2021) som videre nevner at reglene kan bli

for generelle og derfor for lite relevant for viktige kundesegmenter. I de fleste tilfeller er det nødvendig å gjennomføre data-aggregering og plassere produkter i kategorier for at analysen skal gi mening. Dette kan skape videre problemer med at man mister detaljert innsikt (Blattberg et al., 2008).

Bedrifter er avhengig av å agere på innsikten de tilegner seg for å få verdi av analysen. Modellen er eksplorativ og kan kun produsere hypoteser. Derfor er det nødvendig å teste disse for å kunne si noe om verdien (Olson & Lauhoff, 2019). For å predikere fremtidige kjøp kan bedrifter videre benytte andre metoder for å teste reglene, som regresjon eller nevrale nettverk (Olson & Lauhoff, 2019).

Ifølge Wrzuszczak-Noga (2018) er assosiasjonsregler en rask og nyttig mekanisme for atferds deteksjon. Dette kan benyttes til blant annet anbefalinger. I tillegg kan analysen benyttes til produktplasseringen i butikk og designe ulike promoterings strategier i form av kryss-salg eller opp-salg (Blattberg et al., 2008). Olson og Lauhoff (2019) forklarer det som at MBA kan være avgjørende for valg av kampanjer og markedsføringsstrategier. Da analysen kan være med på å identifisere forbrukermønster som ikke tidligere har vært synlige for bedriften. Gangurde et al. (2017) legger til at den også kan klassifisere handleturene, og at modellen kan gi innblikk i hva som handles i de ulike butikkene på forskjellige tidspunkt. I tillegg til å identifisere korrelasjonen mellom produktene, kan datamingteknikken legge frem en forklaring på relasjonen, for å til slutt være et verktøy som kan bidra til økt profitt (Olson & Lauhoff, 2019). Som nevnt tidligere i kapittelet er MBA å regne som en deskriptiv modell. Men ifølge Gangurde et al. (2017) kan MBA-resultatene benyttes til å forutsi fremtidig oppførsel hos kunder, og dermed kunne kalle bruken av modellen som prediktiv.

3.2.3 Klyngeanalyse

Klyngeanalyse, også kalt cluster analyse, benytter seg av umerkede data og går ut på å segmentere eller gruppere lignende objekter, som kan være observasjoner eller hendelser, i klynger (Phorasim & Yu, 2017). Det vil si at dataen blir fordelt slik at relasjonen mellom objektene i en klynge har flere likhetstrekk enn objekter som er plassert i de andre klyngene (Hastie et al., 2009). Analysen

benyttes til å få en bedre forståelse for, og innsikt i karakteristikken til hver gruppe uten at klyngene er kjent på forhånd (Lu, 2018).

Algoritmen har blitt brukt på forskjellige problemstillinger, ved at den kan gjenkjenne mønster og benyttes til bildebehandling, statistisk dataanalyse og kunnskapsoppdagelse. Analysen er gunstig når det ikke finnes kunnskap om «target class eller group number» på forhånd (Phorasim & Yu, 2017).

I detaljhandel sammenheng kan klynge analyse brukes til å segmentere både produkter og kunder. Den kan identifisere hvilke produkter som ofte kjøpes sammen. Klyngene kan brukes til mersalg, eller koble klynger som underpresterer med klynger som overpresterer for å øke salget av varer som ikke selger godt. Det kan også brukes til produktplassering i butikken. Klyngeanalyse kan også brukes til å finne ulike kundegrupper. For eksempel kan det identifisere kunder som er lav verdi og høy verdi kunde (Buttkus & Eberenz, 2019).

Informasjonen som kommer frem av analysen kan bli kritisert for å ikke kunne forklare kundens kjøpsadferd da dataen er for aggregert og for at demografiske- og livsstils variabler ikke kan forklare heterogeniteten (Blattberg et al., 2008). Zhang et al. (2021) nevner at klynge analyse kan bli utfordrende på store datamengder på grunn av den beregningsmessige kompleksiteten til tradisjonelle klyngemetoder som for eksempel k-means. Modellen presterer også dårligere på personlig anbefalinger ifølge Blattberg et al. (2008), da algoritmen plasserer like objekter i klynger, og behandler alle innenfor samme gruppe likt.

Når algoritmen skal gjennomføres i sammenheng med produktanbefalinger og kryss-salg er det nødvendig å legge til transaksjonsdata fra alle kunder. Dette vil resultere i at modellen skaper mindre verdifulle løsninger, da mange forhandlere har tusenvis av produkter, tusenvis av kunder og utfordringen med at de fleste kunden bare kjøper noen få produkter per handletur (Zhang et al., 2021). Dette vil resultere i at prosessen blir svært tidskrevende.

3.2.4 Collaborative filtering

Collaborative filtering er en anbefalingsalgoritme som gir anbefalinger på bakgrunn av andre brukere. Det vil si at modellen benytter seg av rangeringer og handlinger fra andre eller lignende preferanser for å predikere. Det vil si at algoritmen legger til grunn at om brukere er enige om kvaliteten eller relevansen til noen varer, vil de sannsynligvis være enige når det kommer til andre varer også (Phorasim & Yu, 2017).

Collaborative filtering kan deles inn i to hovedformer: den ene er memory-based og den andre er model-based. Memory-based vil si “nærmeste nabo”, hvor anbefalingen på et målprodukt til en målkunde er basert på andre kunder med samme interesser. Model-based vil si at prediksjon om en målkundes preferanser er basert på om andre kunder som vanligvis liker samme produkter som den målkunden, også pleier å like målproduktet (Blattberg et al., 2008).

Modellen har flere utfordringer blant annet er den ofte avhengig av produktvurderinger fra kundene. Dersom dette ikke er tilgjengelig, vil det være vanskelig å anbefale produktet. I tillegg finnes det skalerbarhetsutfordringer og effektiviteten på anbefalingene er en utfordring (Neysiani, Soltani, Mofidi & Nadimi-Shahraki, 2019). Blattberg et al. (2008) viser også til at det er usikkerhet rundt hvordan modellen kan benytte seg av transaksjonsdata, da mye av litteraturen knyttet til modellen omhandler data om produktvurderinger og at denne er tilgjengelig. Modellen tar høyde for flere antesedenter på en systematisk måte, og kan derfor bli sett på som et steg videre fra MBA (Blattberg et al., 2008).

4. Metode

I dette kapitlet beskrives hvilken forskningsmetode og metodiske valg som er tatt underveis i prosessen for å best besvare forskningsspørsmålene. Kapitlet tar for seg forskningsstrategi og design, deretter datainnsamlingen og datakilder. Videre beskrives hvordan datahåndteringen av både de kvalitative og kvantitative dataene ble gjennomført, før reliabilitet og validitet knyttet til valgene som er gjort blir lagt frem til slutt. Arbeidet med denne masteroppgaven startet våren 2020 med idémyldring og samtaler med eksperter i bransjen. Prosessen fortsatte høsten 2020 i faget BUS345 som resulterte i en projektskisse og et skjelett som deretter har blitt videreutviklet utover våren 2021.

4.1 Forskningsdesign

Forskningsdesign er en overordnet plan og rammeverk for hvordan en skal besvare forskningsspørsmål (Saunders, Lewis & Thornhill, 2016, s. 348). Saunders et al. (2016) skiller blant annet mellom eksplorative og deskriptive studier. Ettersom vi ønsker å få mer kunnskap om datadrevet innsikt og personalisering av kommunikasjon faller dette studiet innenfor kategorien eksplorativ undersøkelse. Her er hovedfokus å oppdage og få innsikt i fenomenet, og øke vår forståelse for problemområdet basert på våre forskningsspørsmål.

Ved innsamling av data finnes ulike strategier å velge mellom og en av disse er casestudie. Denne oppgave er å betrakte som en enkeltcasestudie ettersom det er valgt én bedrift som studeres. Ifølge Jacobsen (2005) er en casestudie nyttig dersom det ønskes å få god innsikt og en virkelighetsnær beskrivelse av et fenomen. I denne oppgaven benyttes det flere ulike datainnsamlingsmetoder for å sikre god kunnskap og forståelse, og for å kunne besvare forskningsspørsmålene. Strategien sies også å svare godt på spørsmål som omhandler «hvorfor», «hva» eller «hvordan» (Saunders et al., 2016). Dette er kjennetegn som passer dette studiet godt. En av svakheten ved enkeltcasestudie er begrensingene rundt generaliserbarhet, ettersom studien er omhandler en bedrift. Det vil likevel være metoder og teknikker i denne oppgaven som kan brukes i andre casestudier for å etterprøve resultatene.

Videre er det valgt en kombinasjon av både kvalitativ- og kvantitativ analyse. Kvalitative analyser ble brukt for å få informasjon om bedriften. Dette skapte dybde- og detaljforståelse, og en helhetlig

oppfatning av det studerte fenomenet. Kvantitativ metode ble brukt for å analysere store mengder numerisk data (Jacobsen, 2005). Kombinasjonen kan ifølge Tashakkori og Creswell (2007) defineres som blandet metode og er «forskning hvor forsker samler inn og analyserer data, integrerer resultatene og trekker slutninger ved hjelp av både kvalitative og kvantitative metoder i en enkel case». Det å kombinere kvalitative og kvantitative data i en undersøkelse blir sett på som verdifullt da det kan hentes ut styrker av begge metodene uten begrensninger (Östlund, Kidd, Wengström & Rowa-Dewar, 2011). Dette studiet kan derfor sies å ha en pragmatisk tilnærming til designet (Saunders et al., 2016).

For å opparbeide en dypere forståelse for fenomenet, har undersøkelsesprosessen bestått av å bevege seg frem og tilbake mellom teori og empiri. Denne metoden kalles for abduktiv forskningstilnærming og er ofte svært nyttig å benytte i studier for å få mer forståelse for fenomenet (Saunders et al., 2016). Dubois og Gadde (2002) legger til at fremgangsmåten er godt egnet til casestudier.

4.2.1 Valg av case

I starten av denne masteroppgaven var det et ønske om å lære mer om data mining og maskinlæring. Gjennom vårt nettverk ble det våren 2020 opprettet kontakt med Europris, som var i startfasen av å benytte datadrevet innsikt. De fortalte at de hadde igangsatt et prosjekt kalt «personalisering» tilknyttet CRM avdelingen. De ønsket å benytte sine store mengder data til å bli bedre kjent med sine kunder, og skape målrettet kommunikasjon på et mer individuelt plan. De hadde gjort noen manuelle analyser av kundedata og ønsket å utforske dette mer.

Høsten 2020 ble brukt til å forstå temaet customer relationship management. Med kunnskapen opparbeidet i denne perioden, ble det ønskelig å tilegne mer kunnskap om Europris, og utfordringene de hadde rundt utnyttelse av data. I tillegg ble gjennomføring av analyser som kunne hjelpe Europris videre mot mer datadrevet innsikt om sine kunder aktuelt. Via lederen for prosjektet personalisering, ble det opprettet kontakt med resterende informanter som kunne tilføre oppgaven verdi, i tillegg til historisk anonymisert transaksjonsdata som var et godt datasett for dette studiet.

4.2 Datainnsamling

Denne oppgaven er en enkeltcasestudie og inkluderer i hovedsak én undersøkelsesenheter. Enheten som undersøkes er CRM-avdelingen i Europris Norge. For å sikre personvern tilknyttet innhenting av data ble det søkt godkjenning av prosjektet hos NSD. Det ble lagt frem hvordan innsamling av data til våre kvalitative undersøkelser gjennom intervjuerprosesser skulle bli gjennomført med tanke på anonymisering og sletting av data. Videre ble det presentert at våre kvantitative transaksjonsdata ble tilgjengelig fra Europris og informasjon om hva dette datasettet inneholdt.

Datainnsamlingen ble delt inn i tre kategorier: litteratur-, primær- og sekundær data. Med litteratur menes artikler og fagbøker som inneholder teori og rammeverk som reflekterer og beriker oppgavens tema. Litteraturen ble benyttet til å få kunnskap om tema, i tillegg til å diskutere funnene våre fra resultatdelen. Primær data er data samlet for å kunne besvare forskningsspørsmål én «Hvordan jobber Europris med personalisering og datadrevet innsikt i dag og, hvilke utfordringer har de knyttet til dette?» og dermed danne et bilde for analysearbeidet og modelleringen som kunne skape verdi for Europris. Det var hensiktsmessig å benytte kvalitative intervjuer i denne prosessen. Det ble derfor gjennomførte intervjuer av ansatte i Europris for å få en oversikt over dagens situasjon. I tillegg ble en spesialist innenfor feltet datadrevet innsikt intervjuet for å få et bilde på hvordan situasjonen i det norske detaljhandelmarkedet er i dag. Til slutt er sekundærdata samlet inn, hvor det ble benyttet kvantitative data i form av transaksjonsdata fra anonymiserte kunder og gjennomførte analyser på dataen for å besvare resterende forskningsspørsmål.

4.2.1 Litteratur

Når det gjelder innsamling av litteratur har det blitt benyttet ulike plattformer. Artikler er i hovedsak innhentet gjennom søkeplattformene Google Scholar og bibliotekbasen Oria via NMBU's nettside. I begynnelsen ble søkeord som «Customer relationship management» og «CRM retail» benyttet. Dette resulterte i mange artikler om CRM-systemer noe som ikke var relevant for oppgaven. «Analytical CRM» i kombinasjon med ord som «retail», «supermarkets» og «grocery store» ble også brukt. «Analytical CRM» ble oppdaget at ikke var uttrykket som var mest brukt i tidligere forskning. Søkene ble derfor kombinerte med støtteordene «AI», «database marketing» og «analytical marketing». Dette viste seg å gi bedre resultater, og flere artikler som var mer rettet mot vårt tema.

Videre har det blitt benyttet artikler og litteratur fra tidligere semester på NMBU. Blant annet har litteratur fra emnene “INN350 Digitalisation and digital business models” og “INN355 Machine learning for business process optimisation” vært svært verdifulle. I tillegg til artikler og bøker som har blitt foreslått av våre veiledere.

Spesielt innen datadrevet innsikt er matematikken og statistikken uforandret de siste 25 årene. Vår erfaring er at den nyere forskningen inneholder kombinasjoner av ulike teknikker for å finne de mest presise metodene. Litteratur fra mellom 2000 og 2010 har inneholdt mye relevant informasjon for vårt formål. Litteraturen mellom 2010 og 2020 har ofte gått i detaljnivå på hvordan de ulike modellene er programmert, noe denne studien ikke har til hensikt å gå i dybden på.

4.2.2 Primær data

For å kunne besvare forskningsspørsmålene er det nødvendig med innsikt i hvordan Europris jobber med CRM og datadrevet innsikt i dag. Alle prosesser og initiativer er ikke alltid dokumentert godt nok og det var derfor hensiktsmessig å intervjuer de ansatte hos Europris. Det ble benytte semi-strukturerte en-til-en intervjuer (Saunders et al., 2016). Saunders et al. (2016) forklarer at semi-strukturerte intervjuer ikke er standardiserte. Dette betyr at gjennomføring av intervjuet blir gjort med bakgrunn i ulike tema og ikke faste spørsmål. Spørsmålene kan også komme i ulik rekkefølge ut fra hva som er naturlig i samtalen. På denne måten kunne oppfølgingsspørsmål bli stilt ut ifra hva kandidaten svarte på de forhåndsbestemte temaene (Jacobsen, 2005, s. 149). Intervjuene ble gjennomført på denne måten for at informanten skulle snakke mest mulig fritt, slik at de ikke ble ledet inn på klare svar på bakgrunn av intervjuguiden.

Gjennomføringen skulle i hovedsak bli gjort ansikt til ansikt, da dette bygger tillit, åpenhet og en god flyt i samtalen (Jacobsen, 2005). På grunn av situasjonen med Covid-19, måtte andre metoder vurderes for å få gjennomført intervjuene. Alternativet ble å benytte Teams som møtekanal. Dette begrenset muligheten til å tyde kroppsspråk og annen ikke-verbal kommunikasjon noe.

Utvelgelsesprosessen av intervjuobjekter var formålsstyrt og alle hadde derfor tilknytning til data drevet innsikt (Jacobsen, 2005). Tre ansatte i Europris ble intervjuet og alle hadde en tilknytning til «personaliseringsprosjektet». Disse hadde alle forskjellige funksjoner knyttet til prosjektet og

arbeidet rundt CRM og kundedata. Intervjuobjektene viser bredde når det gjelder både kjønn, fagbakgrunn og alder. For å sikre personvern av kandidatene, ble det benyttet et informasjonsskriv fra NSD. Dette informerte kandidaten om at de kunne bli gjenkjent i oppgaven ved stillingstittel eller kunnskapen som ble gitt, men at navn og andre kjennetrekke ble anonymisert i tillegg til at oppgaven er konfidensiell.

4.2.3 Sekundær data

I tillegg til primærdata ble det også benyttet sekundærdata for å besvare forskningsspørsmålene. Sekundærdata kan defineres som data som er samlet inn tidligere for andre årsaker (Saunders et al., 2016). Denne type data blir ofte benyttet i case studier, og fordelen med å utnytte allerede innsamlet datamateriale er blant annet tidsperspektivet og datakvaliteten (Saunders et al., 2016). Samtidig kan det være utfordrende å få tilgang på denne dataen og det er mulig at dataen kommer i en form som ikke lar seg tilpasse til studiets formål (Saunders et al., 2016).

Sekundærdataen vil i dette studiet først og fremst være transaksjonsdata som er rådata hentet fra administrative systemer innad i Europris. Rådataene benyttes til å teste en maskinlæringsmodell for å gi et eksempel på hvordan Europris kunne benytte dataen de har tilgang til, på måter de ikke gjør i dag. Informasjon om at dette var data som kunne bli brukt i dette studiet ble kommunisert tidlig i prosessen, slik at planleggingen rundt mulighetene for utnyttelse av denne dataen kunne bli undersøkt parallelt med de kvalitative undersøkelsene.

I tillegg til transaksjonsdataen som sekundærkilde, ble tidligere kundeaviser og årsrapporten til Europris benyttet for å få et bedre innblikk i hvordan bedriften arbeider i dag og hvilke mål de ønsker å oppnå. Kundeavisen viste til ukestilbudene Europris gjennomførte i perioden vår transaksjonsdata stammer fra. Dette resulterte i at en tydeligere kunne se hva som lå bak reglene fra analysen.

4.3 Datakilder

I dette studiet har det som nevnt blitt benyttet både primær- og sekundærdata. I dette kapitlet vil datakildene bli lagt frem.

Respondenter

For å samle relevante respondenter som kunne gi god innsikt til å besvare våre forskningsspørsmål, ble det opprettet kontakt med Europris for å få et innblikk i hvem som arbeidet med CRM og datadrevet innsikt i bedriften. Det viste seg at det var flere avdelinger involvert i prosjektet “personalisering” som resulterte i at intervjuene ble gjennomført med flere ansatte med ulike bakgrunn og daglige arbeidsoppgaver. På denne måten ble det innhentet informasjon fra flere ulike syn når det kom til temaene som skulle undersøkes. CRM-sjef ble intervjuet to ganger da det var nødvendig å innhente ytterligere informasjon om prosjektet.

Tabell 2 - Oversikt over respondenter

<i>Hvem</i>	<i>Hva</i>	<i>Når</i>
<i>CRM-sjef</i>	Systemetablering, utviklingen og etablering av kundeklubben til Europris. Har hovedansvaret for personaliseringsprosjektet og de ansatte involvert.	1: 24.11.20 2: 09.03.21
<i>IT</i>	Ansvar for det tekniske i håndteringen av kundedata, CRM og kampanjer. Jobber mest med å samle inn data, både kundedata og kjøpsdata og koble det sammen.	04.03.21
<i>Controller</i>	Jobber nå hovedsakelig med e-handel, det digitale teamet og mye med IT om hvordan dataen skal flyte i tillegg til rapportering.	02.03.21

Alle respondentene er tett på personaliseringsprosjektet, men har ulik kunnskap om temaet vi tar opp. Tilnærmingen til intervjuene med de forskjellige blir derfor lagt opp på ulike måter ved at spørsmålene i intervjuet ble spisset mot den rollen de hadde i prosjektet. På grunn av anonymisering er intervjuobjektene tilfeldig tildelt prosjektdeltaker 1, 2 og 3 i resultatkapittelet.

Informant

Det var ønskelig med mer kunnskap om temaene «datadrevet innsikt» og «personalisering» innenfor detaljhandel. Av den grunn ble det gjennomført intervju med en ekspert på området som ga oss innsikt og ideer rundt hvordan dette kunne benyttes innenfor detaljhandel fremover. Dette ga et overblikk over hvilke utfordringer informanten kunne se for seg. Kunnskapen ble benyttet til å få et overblikk over tema.

Tabell 3 - Oversikt over informanter

Hvem	Hva	Når
Professor	Ekspertområder innenfor Big data, CRM, dataanalyse, digitalisering.	23.02.21

4.4 Datahåndtering og datahåndteringsplan

Ifølge (Jacobsen, 2005) er det tre grunnleggende krav knyttet til forholdet mellom de som forsker og de som blir forsket på. Disse kravene er informert samtykke, krav på privatliv og krav på å bli korrekt gjengitt. I begynnelsen av hvert intervju ble intervjuobjektene informert om at deltakelse var frivillig. I tillegg ble det på forhånd sendt ut en oversikt over temaer som skulle undersøkes, dermed ble intervjuobjektene forberedt på hva intervjuet skulle omhandle (Jacobsen, 2005)

Deretter ble det gjort en evaluering av hvor følsom informasjonen som skulle samles inn var. Det ble valgt å anonymisere intervjuobjektene og det var derfor nødvendig å ta ut data som ville gjøre det lett å identifisere dem (Jacobsen, 2005). I tillegg til anonymiseringen var det et ønske fra Europris om at oppgaven skulle være konfidensiell i to år, og oppgaven vil dermed ikke bli publisert på NMBU's sider før denne tiden har gått.

Samtykkeskjema ble utsendt til intervjuobjektene på forhånd slik at det ble godkjent å gjennomføre intervjuene over videoverktøyet Teams. Ved starten av hvert intervju ble det spurt om godkjenning for å ta opptak og video av samtalen. Videoen ble benyttet til transkribering, før opptaket ble slettet. I tillegg til data fra intervjuene ble det laget en plan for behandling av transaksjonsdata tilsendt fra Europris. Dataen ble beholdt på trygge servere gjennom NMBU, i tillegg til at dataen ble behandlet i Dataiku og analysene gjennomført i SAS Enterprise Miner. Datahåndteringsplanen ble meldt inn til NSD og ble godkjent.

4.5 Dataanalyse av kvalitative intervjuer

Ved analyse av dataen ble det benyttet fire steg som er presentert av Jacobsen (2005): dokumentere, utforske, systematisere og kategorisere. Det første som ble gjort etter gjennomføring av intervjuene var å transkribere lydopptakene (dokumentere). Dette bidro til at det ble fanget opp informasjon som ble oversatt i selve intervjuet. Videre ble transkriberingene lest igjennom for å få

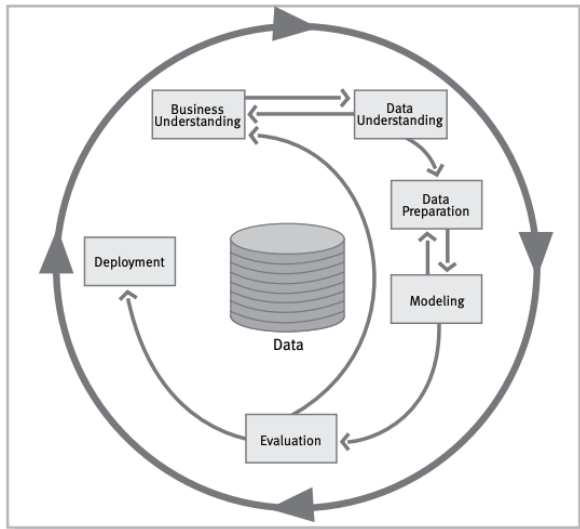
et bedre helhetsbilde av prosessene som foregikk i dag. På denne måten ble elementer som skilte seg ut identifisert (utforske). For å få oversikt over dataen ble teksten delt inn i kategorier og temaer hvor uttalelser fra intervjuobjektene ble samlet (systematisere og kategorisere). Kategoriene bygde på temaene fra intervjuguiden og utsagn fra intervjuobjektene. Til slutt ble svarene sammenlignet i de ulike kategoriene for å se om det var likheter eller ulikheter i uttalelsene (sammenbinde).

På denne måten ble det dannet et bilde av hvordan prosessen og arbeidsoppgavene til de ulike ble gjennomført. Dette resulterte i en forståelse for hvordan Europris som bedrift arbeidet med datadrevet innsikt, hvor langt de hadde kommet, hvilken datatilgang de hadde, hva de ønsket å oppnå med å sette i gang et personaliseringsprosjekt og hvilke utfordringer som de mente var de mest relevante. Dette var funn som skapte et godt grunnlag for videre analyse av transaksjonsdata og for diskusjonskapittelet.

Hensikten med de kvalitative intervjuene var å få innblikk i hvordan Europris jobber med datadrevet innsikt og hvilke utfordringer de møter. Derfor er det i resultatkapittelet i stor grad gjengitt informasjon fra intervjuene gjennom å beskrive hvordan Europris gjør det i dag, og ikke lagt vekt på alle nyanser mellom de ulike svarene til intervjuobjektene. Nyansene på svarene er også av interesse, men ikke relevant for denne studien i like stor grad.

4.6 Dataanalyse av transaksjonsdata - CRISP-DM

Metoder som ofte blir brukt i data mining prosjekter er KDD, CRISP-DM og SEMMA. CRISP-DM og SEMMA er mer forretningsorientert enn KDD. CRISP-DM er også mer detaljert enn SEMMA (Shafique & Qaiser, 2014). Valget av metode falt derfor på CRISP-DM. CRISP-DM er utarbeidet med bakgrunn i praktiske og virkelighetsnære data mining prosjekter (Chapman et al., 2000). Den deles inn i seks faser som vist i figur 4.



Figur 4 - Faser for CRISP-DM modell (Chapman et al., 2000)

4.6.1 Forretningsforståelse

For å skaffe forretningsforståelse ble det i hovedsak brukt kvalitative intervjuer som er beskrevet tidligere i dette kapittelet. Gjennom denne fasen ble det identifiserte noen av Europris sine business problemer.

4.6.2 Dataforståelse

I den andre fasen handlet det om å forstå dataen. Dataen må samles inn, beskrives, utforskes og kvalitetssikres (Chapman et al., 2000). Gjennom dialog med Europris ble det bestemt at det skulle hentes ut data fra første kvartal 2020. Ettersom Europris er børsnotert ønsket de at dataen som skulle benyttes i denne oppgaven ikke skulle inneholde upubliserte opplysninger om omsetning. Første kvartal 2020 ble derfor valgt, og datasettet inneholdt transaksjonsdata for alle Europrisbutikker i perioden. Datasettet ble deretter lastet opp i Dataiku som er en plattform for håndtering av data og bygging av kunstig intelligens modeller (Dataiku, u.å.).

Gjennom funksjonen «explore» i dataiku kunne dataen utforskes. Formatet på dataen var at hver rad inneholdt opplysninger om hver handlekurv. Det vil si at hver rad hadde en receipt_id som er unik for hver handel. Deretter inneholdt dataen 41 ulike kolonner med ulik informasjon om hver handlekurv. For eksempel customer_id, item, currency og tax_lines. Datasettet var delt opp i ulike filer, og det totale datasettet var for stort for å åpne på våre datamaskiner.

4.6.3 Dataforberedelser

Data forberedelse er den tredje fasen og innebærer å selektere, rense, konstruere, integrere og formatere data (Chapman et al., 2000). I denne fasen ble det nødvendig å minimere datasettet til et utvalg slik at arbeidet med analysen og modelleringen kunne foregå på en effektiv måte. Det ble raskt klart at datasettet på 37,3 GB var for omfattende for vår analyse. For å få et datasett som var håndterlig ble det vurdert som hensiktsmessig å velge ut en enkel butikk. Butikken ble tilfeldig valgt, i tillegg til at det kun ble beholdt transaksjoner som inneholdt kunde_id. På denne måten ble kun medlemmer av kundeklubben studert. Ved å velge kun én butikk og kun kundeklubbmedlemmer blir det vanskelig å generalisere funnene våre, men hensikten er å kunne illustrere hvordan analysen kan gjøres på et rikere datasett.

Variablene som ble valgt var customer_id, receipt_id, timestamp, product_id, productname, netprice og grossprice. Kunde_id er unik og representerer at kunden er medlem av kundeklubben. Receipt_id inneholdt unike kvitteringsnummer, på denne måten kunne handlekurvene skilles fra hverandre. Timestamp var tiden handelen ble gjennomført, variabelen inneholdt både dato og klokkeslett. Productname og product_id identifiserer hvilke varer som har blitt kjøpt. Til slutt gir netprice og grossprice informasjon om prisen på produktet.

Ved videre utforskning av data ble det tydelig at produktene som gikk mest igjen i datasettet var bærepose og pant, disse ble derfor fjernet i sin helhet da dette ikke var relevant for vår analyse. Videre var det flere ulike varer innenfor samme produktkategori, noe som ville påvirke svarene fra en kommende modellering. Flere varer ble derfor samlet under samme produktnavn. Eksempler på dette var hvis produktet hadde flere ulike smaker eller lukter ble disse samlet til å kun hete produktnavnet. For eksempel «pumpesåpe orkide» og «pumpesåpe eple» ble hetende «pumpesåpe». Alle ulike sjokolader ble hetende kun sjokolade. Aggregeringen ble gjort i dataiku sin funksjon kalt «variable clustering». Programmet foreslår hva som burde slås sammen ut ifra lignende ord i datasettet. Det var også en del som denne noden ikke klarte å identifisere som dermed måtte gjennomføres manuelt. For eksempel alle smaker av Dent og Fishermansfriend ble endret til Pastiller. Etter dataforberedelsene var det endelige datasettet redusert til et utvalg med totalt 602 produkter.

4.6.4 Modellering

Den fjerde fasen kalles modellering. I denne fasen velges modellteknikk og test-design i tillegg til at det bygges og velges modeller (Chapman et al., 2000). Med bakgrunn i de funnene fra forretningsforståelses-, dataforståelse- og dataforberedelse steget, var assosiasjonsregler den modelleringsteknikken som ble valgt. Dypere beskrivelse av Market basket analyse og andre mulige modeller finnes i teorikapittelet, og diskusjon rundt modellvalg finnes i diskusjonskapittelet.

Det finnes mange ulike data mining softwares hvor det kan gjennomføres market basket analyse (Blattberg et al., 2008). To av de største leverandørene av datamining verktøy er SAS og SPSS (Buttle & Maklan, 2015). Det har blitt benyttet SAS Enterprise Miner (SAS EM) i dette studiet, da lisensen for programmet er tildelt gjennom NMBU og med bakgrunn i erfaringen som er opparbeidet ved verktøy fra tidligere bruk.

Datasettet ble lastet opp i SAS EM, før assosiasjonsregel noden ble benyttet. Denne noden krever en ID variabel og en target variabel (SAS, 2017). Variablene Receipt_ID ble brukt som ID, og Productname ble brukt som target variabel. På grunn av tid og ressurser ble ikke timestamp og kunde_id benyttet. Prisen på varene ble brukt i etterkant for å regne på lønnsomhet av reglene i excel. Se vedlegg 5 for utregning. For å kunne bruke assosiasjonsregel noden måtte selve dataoppsettet endres når det kom til hvordan produktene var presentert. I det opprinnelige datasettet lå alle produktene i samme linje per transaksjon. Dette måtte endres slik at hvert produkt fikk én linje hver.

I assosiasjonsregel noden ble det gjort endringer på noen av de forhåndsinnstilte egenskapene. Det ble testet ulike support levels, men endte til slutt på 5%. Dette nivået ga oss interessante regler i tillegg til at de hadde et visst nivå av support. Det ble også gjort begrensninger på antall relasjoner til maks to for at ikke resultatene skulle bli for komplekse for denne studien. Etter at noden ble kjørt, presenterte SAS EM de 25 mest interessante reglene basert på de objektive målene support, confidence og lift. Disse tre målene ble i hovedsak brukt for å evaluere resultatet av modellen. Til tross for at det ikke er vanlig å dele unsupervised modeller inn i trening og test datasettet, ble dette gjort for å forsøke å validere resultatene på alternative måter. 60% av datasettet

ble tilfeldig valgt til treningsdatasett, og de resterende 40% utgjorde test-datasettet. Dette ble gjort i Dataiku. Resultatet av dette finnes i kapittel 5.2.3.

4.6.5 Evaluering

Evaluering er den femte fasen og innebærer evaluering av resultat, gjennomgang av prosessen og kartlegge videre steg (Chapman et al., 2000). Når resultatene ble evaluert, ble outputen vurdert til å kunne svare på forretningsmålene, som ble identifiserte gjennom de kvalitative undersøkelsene. Våre resultater er hentet fra et utvalg av datasettet og kan derfor ikke regnes som signifikante for hele Europris. Modellen som benyttes er derimot passende for å skape innsikt og identifisere mønster i dataen som Europris deretter kan agere på i form av personalisering. Samtidig ble det gjort vurderinger tilknyttet faktorer som kan ha innvirkning på reglene som sesongvariasjon, butikkvariasjon og effekten av kampanjer. Dette er faktorer som må vurderes ved eventuell implementering av regler.

Videre ble prosessen og kvaliteten av data mining prosjektet vurdert. Dataen som er anvendt i modelleringen har vært tilsendt fra Europris, og derfor tilgjengelig for bedriften. Gjennom prosessen har modelleringen blitt dokumentert og valg som er tatt slik at analysen kan gjentas i fremtiden. Når neste steg i prosessen skulle vurderes, ble det lagt mye fokus på hvordan Europris kunne utnytte resultatene, og eventuelle ressurser dette krevde. I tillegg til hvordan reglene kunne bidra i personaliseringen og mulige forretningsverdier. I Europris sitt tilfelle ville en analyse av et rikere datasett være fordelaktig før en eventuell implementering for å skape bedre innsikt i kundenes handlemønster.

4.6.6 Deployment

Det siste punktet i CRISP-DM er deployment. I denne fasen skal det legges en plan for deployment, implementering, vedlikehold, sluttrapport og gjennomgang av prosjektet (Chapman et al., 2000). I denne fasen blir det lagt frem hvordan analysen kan benyttes og videreutvikles av Europris. Da vår analyse bygger på et mindre utvalg blir det metoden og mulighetene som ligger til grunn ved assosiasjonsregler som Europris kan benytte seg av i sitt personaliseringsprosjekt, og hvordan analysens regler kan benyttes til å treffe kunden mer personlig med de rette produktene.

4.7 Reliabilitet og validitet

Reliabilitet refererer til i hvilken grad datainnsamlingsteknikker og analyseprosesser vil gi konsistente funn (Saunders et al., 2016). Det vil si om resultatene vil bli de samme ved senere undersøkelser ved bruk av et lignende forskningsdesign (Saunders et al., 2016). Det er flere variabler som kan ha påvirket reliabiliteten i denne oppgaven. I den kvalitative delen av oppgaven tar forskningsspørsmålet for seg hvordan Europris arbeider med datadrevet innsikt og personalisering i dag. For å svare på dette er ansatte i Europris intervjuet. Her kan svar avhenge av egne synspunkter og meninger, i tillegg til at informasjon kan ha blitt tilbakeholdt da Europris er et børsnotert selskap. Til tross for dette, oppleves det som at de ansatte er kunnskapsrike og villig til å dele informasjon under intervjuprosessen. Mye av grunnen for at de velvillig delte informasjon kan være den utsatte publisering på 2 år som sikrer at forretningshemmeligheter ikke blir lekket. Videre ble ordrett transkribering av opptakene som ble gjort under intervjuene benyttet, for å være sikre på at ordlyd og informasjonen ble tolket rett. Det ble også gjennomført et ekstra intervju med leder hvor informasjonen fra intervjurundene ble gjennomgått og bekreftet. Fokuset for oppgaven er hvordan Europris arbeider med personalisering til dags dato, og funnene vil da kunne være annerledes ved et senere tidspunkt fordi prosjektet i stadig utvikling.

Når det gjelder vår kvantitative undersøkelse har datasettet blitt tilsendt fra bedriften. Videre har dette blitt rensert og gjort klar for videre modellering og analyse. Gjennom tidligere underkapitler i metode kapitlet har det blitt informert om alle endringer og aggregeringer som er gjort, og program som er brukt. I tillegg til valg som har blitt tatt i selve modelleringen. Dette gjør det lett for andre som ønsker å gjennomføre samme analyse, og få det samme resultatet.

Validitet går på om man faktisk måler det man ønsker å måle (Saunders et al., 2016). Reliabilitet og validitet henger nært sammen, ved høy reliabilitet er også validiteten høy (Ringdal, 2013). Dermed vil utfordringene nevnt over kunne påvirke validiteten, og menneskelige målefeil kan være en faktor for å undergrave validiteten til arbeidet som har blitt gjort. Tolkninger av dataen har derfor blitt bekreftet underveis i undersøkelsesprosessen. Måleinstrumentene for valideringen av den kvalitative delen er primærdataen, i form av intervjuene som er gjennomført. Sekundærdataen er på samme måte måleinstrumentet for den kvantitative delen av oppgaven. I tillegg benyttes teori og metode for å gi mer utfyllende beskrivelser av caset.

5. Resultater

I dette kapittelet legges de kvalitative og kvantitative funnene frem. Det er anvendt både intervjuer og dataanalyse for å besvare forskningsspørsmålene. Resultatkapittelet består av tre deler som har til hensikt å gjenspeile forskningsspørsmålene. Den første delen tar for seg funnene fra intervjuene med de ansatte i Europris. I del to presenteres resultater fra analyse av transaksjonsdata. Del tre omhandler funn hva gjelder forventet forretningsverdi for Europris ved bruk av datadrevet innsikt.

Fremstillingen av resultatene vil variere etter hvilken type data som blir beskrevet. Resultater fra intervjuene vil i hovedsak bli fremstilt gjennom gjengivelse. Det vil noen steder bli brukt direkte sitat for å understreke hovedpoenger. I dataanalysen blir det presentert utklipp og innsikt fra analyseverktøyet med forklaringer til hva disse resultatene betyr. Avslutningsvis blir det lagt frem utregninger basert på identifiserte assosiasjonsregler fra dataanalysen.

5.1 Personalisering og dataanalyse i Europris

Det første forskningsspørsmålet som skal besvares er: *«Hvordan jobber Europris med personalisering og datadrevet innsikt i dag, og hvilke utfordringer har de knyttet til dette?»*. Europris sitt prosjekt som kalles «personalisering» er det området som i hovedsak har blitt studert, da dette er et prosjekt som foregår parallelt med denne oppgaven. Resultatene i dette kapittelet blir lagt frem i tre ulike temaer som ble identifisert på tvers av intervjuene. Temaene er motivasjon, utfordringer og fremtidig arbeid.

5.1.1 Motivasjon for digitalisering

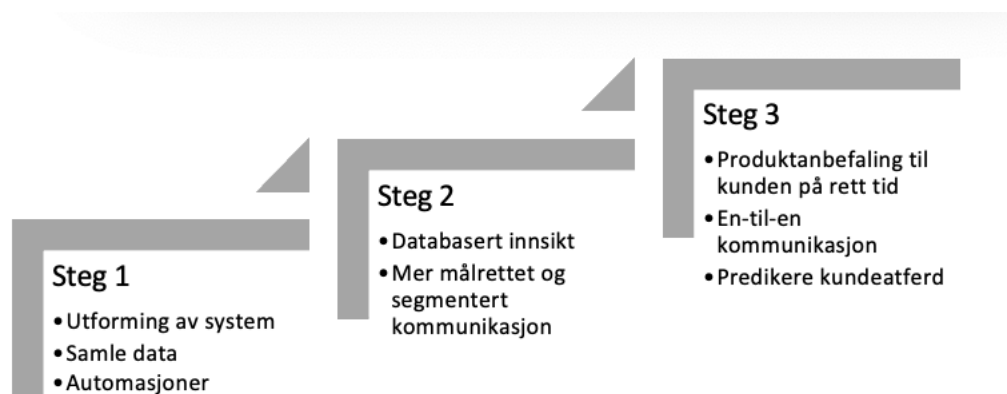
For å oppnå en forståelse av hvordan Europris har kommet dit de er i dag var det sentralt å få innsikt i motivasjonen bak digitalisering og personaliseringsprosjektet. Europris har i mange år vært styrt av katalog og ukestilbud hvor alle kunder mottar identisk markedsføring og tilbud. Våre funn indikerer at hele verdikjeden er bygget med bakgrunn i denne tankegangen. Eksempelvis gjør innkjøpsavdelingen innkjøp basert på ukestilbud. Med andre ord er det bedriften som styrer hva kunden har behov for eller skal få tilbud på i en gitt uke. Dette tror Europris kommer til å endre seg i fremtiden og at verdien ligger i at kunden får tilbudet når behovet oppstår. Til tross for stort fokus på digitalisering, ble det identifisert at 95% av markedsføringsbudsjettet til Europris fortsatt

brukes på kundeavisen. De tror at mye av grunnen til dette er nettopp at hele verdikjeden til Europris er bygget rundt ukestilbud, og ikke rundt kundebehovet. Europris har de siste årene hatt økt fokus på innsamling av kundedata og anvendelse av teknologi for å forstå kundens behov bedre. Organisasjonen og spesielt ledelsen oppleves som svært mottakelige for digitalisering, dersom et initiativ kan bidra til økt omsetning.

I 2020 ble prosjektet «personalisering» påbegynt. Dette innebærer at alle medlemmer av kundeklubben MER ikke lenger skal motta det samme nyhetsbrevet eller tilbudet, men at kommunikasjonen blir mer personlig. I den anledning nevnes det at en av fordelene med at kundeklubben ble etablert relativt sent, var at mange bedrifter før dem hadde banet vei for hvordan en skulle ta hensyn til personvern ved innsamling av data. Det er enighet blant alle tre respondentene at hovedmotivasjonen for prosjektet er å øke kundefrekvensen. Ved å få alle kundeklubbmedlemmene til å handle en ekstra gang i året kan omsetningen ifølge en av objektene potensielt øke med 80 millioner.

5.1.2 Nåværende bruk

Under første intervju med Europris ble det skissert et trappeløp som forklarer hvor i utviklingsløpet Europris er med tanke på personalisering og datadrevet innsikt. Ifølge Europris befinner de seg på det første steget i denne prosessen. Figur 5 illustrerer denne modellen.



Figur 5 - Trappeløpsmodell

Steg 1” omhandler i hovedsak innsamling og utnyttelse av data til automasjoner i kommunikasjonen basert på kundeaktiviteter. Prosessen forklares som et prosjekt som har tatt mye

tid. Arbeidet startet for rundt 8 år siden, hvor første aktivitet var opprettelse av en ny kundemaster. Dataen som har blitt samlet inn over en lenger periode har vært benyttet til å følge kunden på et anonymt nivå. Det er i de senere månedene at dataen har kunnet bli anvendt på et mer individuelt og personlig nivå ved at systemer for dette har kommet på plass i Europris. Det har gjort det mulig å både spore og analysere transaksjonsdata samt koble dataen til blant annet kundeID.

Gjennom kundeklubben MER kan dataen og informasjonen benyttes til å dele kunder inn i ulike kundegrupper og lister. Prosjektdeltaker 2 nevner at de nå har mulighet til å følge kundens aktiviteter gjennom hvordan de reagerer på utsendte eposter:

“Vi kan nå se hvordan en kunde reagerer på eposter de får tilsendt. Vi kan se om de klikker på den, hva de klikker på, om det blir konvertert til salg i butikk eller på nett, og om de kjøper det som var med i nyhetsbrevet eller noe annet.”- Prosjektdeltaker 2

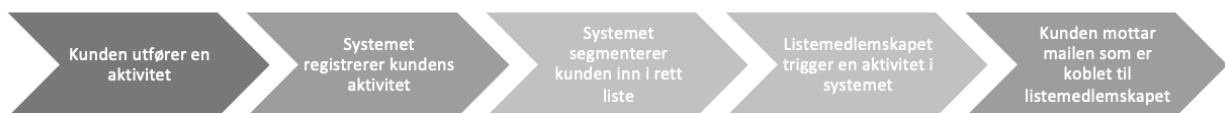
Videre har Europris bygget en teknisk rigg for datainnsamling ved utsendelse av nyhetsbrev. Dette gir de mye bedre innsikt enn de har hatt tidligere. Ifølge Europris har dette tatt lang tid og har på mange måter vært en milepæl. Gjennom kundeklubben får de også innhentet en del informasjon om kunden. Kunden kan selv legge inn egne interesser og favorittbutikk på sin profil. I tillegg kan kontonummer registreres når kunden er medlem i MER-kundeklubben. Dermed blir alle kjøp tilknyttet kundens kontonummer registrert på kundens profil, og Europris får god oversikt over kundens kjøpshistorikk. Europris har også tilgang på tredjepartsinformasjon. Bisnode kan blir koblet på for å berike dataen med det som ligger i det offentlige registeret som adresse og fødselsdato. Europris kan dermed få tilgang til data som gir informasjon om sannsynligheten for livssituasjon, type bil, type bolig, årsinntekt og lignende informasjon.

Analyse av transaksjonsdata har foregått på et veldig manuelt nivå frem til nå. Eksempelvis har det blitt sett på hva det blir handlet mest av i ulike perioder, hvilke produkter ulike aldersgruppe og kjønn handler mest av og hvor ofte kundene er innom butikken. Gjennom intervjuene ble det avdekket at Europris i stor grad har brukt disse forenklede analysene av data for å få kundeinnsikt, uten å videre utnytte informasjonen. Grunnen til dette er at det foreløpig ville blitt en veldig manuell prosess å segmentere, lage kundelister og tilpasse produktanbefalingene.

Europris har også den siste tiden initiert en rekke automasjoner i forbindelse med målet om mer personlig kommunikasjon. Prosjektdeltaker 3 beskriver det selv som enkle automasjoner:

“Dersom en kunde kjøper noe bærekraftig får de en epost med takk for at man kjøpte bærekraftig. Kvitteringen fra handelen blir generert og går inn i systemet i sanntid og blir prosessert. På den kvitteringen ligger et bærekraftig produkt, da går kunden inn i en kundeliste som trigger en e-post.” – Prosjektdeltaker 3

De jobber stadig med videreutvikling av automasjonene og utforsker hvordan systemet kan analysere kunden i sanntid og koble kjøpemønstrene mot de rette tiltakene. Dagens automasjoner bygger altså på kundens handlinger i sanntid. Dette innebærer at når en kunde utfører en aktivitet som er koblet til en automasjon, vil systemet oppfattet dette og legge kunden til i en unik liste som automatisk genererer eposten som er koblet til listen. Et annet eksempel som ble belyst, var deres automasjon for klippekort på hundemat. Det vil si at når en kunde som er medlem av MER-kundeclubb kjøper en pakke med hundemat vil de bli lagt til i en liste hvor det trigges en mail om at de har mottatt et klippekort på hundemat på bakgrunn av kjøpet de akkurat gjennomførte. Figur 6 viser en oversikt over dagens automasjonsprosess.



Figur 6 - Oversikt over automasjonsprosessen

Europris benytter også selvregistrerte interesser fra deres MER-kunder for å tilpasse kommunikasjon i nyhetsbrevutsendelser. Det vil si at alle kunder, i sin profil på nettsiden, kan velge mellom 6 ulike interesser under kategorien de har kalt for “personalisering”, og hake av for de som er av interesse. Valgene man har er student, interiør og trender, hund, garn, julekalender og fikse selv. På bakgrunn av de valgte interessene vil kunden motta kommunikasjon rettet til sin kategori.

Neste steg i trappeløpet omhandler prosessene Europris er i gang med å teste i dag. Prosjekt «personalisering» startet tidlig i 2020, og er et stort satsningsområde som inngår i CRM-arbeidet til Europris. Prosjektet er i startfasen, men stadig i utvikling og parallelt med denne oppgaven har Europris startet å teste utsendelser av kommunikasjon med bakgrunn i informasjon fra tidligere kjøp, for å promotere mer relevante produkter til kunden. Den første testen handlet i hovedsak om å kartlegge hvorvidt systemet fungerte slik det skulle, og om kundene som mottok personaliserte nyhetsbrev responderte. Det ble anbefalt produkter på bakgrunn av hvilken kategori kunden har handlet innenfor tidligere, men ikke de faktiske produktene de har kjøpt mest av. Dette ble utført ved å segmentere kundene, etterfulgt av en utvelgelse av produkter innenfor kategorien Europris ønsket å annonsere i nyhetsbrevet.

“Hvis en kunde har gått inn og handlet vaskemiddel eller tøymykner så går man inn og ser på topplisten for produkter av tøymykner, så anbefaler vi innenfor den kategorien til deg. Det går ikke på andre kunde profiler, men mer volum innenfor kategorien på det du allerede har kjøpt.” – Prosjektdeltaker 2

I utgangspunktet er det kun data tilknyttet Europris sin kundeklubb MER som er benyttet til å utføre analyser og tester. Kundene som har blitt plukket ut til prosjektet vil motta ulike nyhetsbrev med ulike produkter annonsert. Tabell 4 viser en oversikt over de aktivitetene og tiltakene som anvendes av Europris i dag, som et første steg mot mer personlig kommunikasjon.

Tabell 4 - Oversikt over aktiviteter for personalisert kommunikasjon

<i>Nåværende bruk av personalisering</i>	<i>Beskrivelse</i>	<i>Kommentar</i>
<i>Velkomstilbud</i>	I det samme øyeblikket kunden blir medlem av kundeklubben vil personen motta en verdikupong på kr 100,- ved kjøp over kr 500,-.	Trigges på kundeaktivitet
<i>Klippekort (lojalitetskonsept)</i>	Dersom kunden har registrert sitt kredittkort i kundeklubben, vil hen kunne motta klippekort på noen bestemte varer.	Trigges på kundeaktivitet
<i>Takk for at du handler miljøvennlig</i>	Ved kjøp av et miljøvennlig produkt vil kunden automatisk få en e-post som forteller at du har kjøpt et miljøvennlig produkt.	Trigges på kundeaktivitet

<i>Nyhetsbrev med selvvalgte interesser</i>	I kundeklubben kan kunden velge ut hvilke interesser hen har, og deretter får hen en del i nyhetsbrevet som inneholder varer fra denne kategorien.	Trigges av hva kunden har satt som interesser på sin profil
<i>Nyhetsbrev med bakgrunn i tidligere kjøp</i>	Europris undersøker varene en kunde har kjøpt mest av, deretter blir kunden anbefalt forhåndsplukkede produkter innenfor samme kategori.	Kun på relevans, ingen andre insentiver som tilbud.

5.1.3 utfordringer

Bevegelsen fra masse-markedsføring til mer personlig kommunikasjon har innvirkninger på flere deler av organisasjonen. Gjennom intervjuprosessen ble fem hovedutfordringer tilknyttet dagens oppgaver identifisert. Disse fem er henholdsvis:

1. *Domeneutfordring*
2. *Utvelgelse av data*
3. *Regler for produktanbefaling*
4. *Utviklingshastighet, integrasjoner og ressurser*
5. *Hvor skal man starte?*

Digitale initiativer, i dette tilfellet personaliseringsprosjektet, krever at flere fagområder samarbeider. Europris nevner at det i noen tilfeller kan være utfordrende med ulike domenebakgrunn. Kommunikasjon- og fagutfordringer mellom avdelinger, spesielt mellom IT-avdelingen og markedsføringsavdelingen blir nevnt som en større utfordring for Europris. I intervjuene blir det beskrevet at de ulike fagområdene snakker ulike fagspråk, og at de derfor har en person som fungerer som en brobygger mellom de ulike domenene.

Bedriften har stor tilgang til data fra både interne og eksterne datakilder. På bakgrunn av dette ble det identifisert to utfordringer. Hvordan utnytte dataen og hvilken data som skal utnyttes for å skape innsikt, og hvilke produkter som skal kommuniseres ut til kundene. Med den første utfordringen menes også integreringen av tredjepartsdata, samt hvordan og når denne skulle benyttes som støtte i kommunikasjonen. Når det gjelder den andre utfordringen forklares det at Europris opplever vanskeligheter med å vite hvilke produkter som skal kommuniseres til kunden.

Dette er utfordrende blant annet med tanke på varesortimentet Europris har og variasjon i varesortimentet gjennom året. Denne utvelgelsen gjøres i dag manuelt, da flere av produktene ikke kan promoteres ut til kundene. Og man kan dermed ikke automatisk kommunisere med bakgrunn i hva kunden tidligere har kjøpt.

I likhet med mange av de andre utviklingsprosjektene hos Europris byr også personaliseringsprosjektet på utfordringer knyttet til utviklingshastighet, integrasjoner og ressurser. I dag er det ingen som jobber kun med prosjektet, det kommer i tillegg til de ansatte sine ordinære arbeidsoppgaver. Følgelig oppstår det utfordringer knyttet til å frigjøre ressurser, samtidig som det resulterer i at prosjektet ikke har den progresjonshastigheten Europris skulle ønske. En siste utfordring som ble identifisert var knyttet til prioritering av initiativer. De opplever det som utfordrende å identifisere hva som er mest lønnsomt å starte med i forbindelse med personlig kommunikasjon.

5.1.4 Fremtidig plan

Et gjentakende tema i intervjuene med Europris var hva de ønsket å oppnå fremover. Da det gir et innblikk i hvilke prosesser og aktiviteter de ønsker å implementere i sin strategi fremover er det hensiktsmessig å videre utforske dette. Som nevnt har Europris tilgang til flere typer data, både data hentet fra kundeklubben, i tillegg til anonym transaksjonsdata og tredjepartsdata.

I første omgang ønsker Europris å analysere dataen for å få bedre kundeinnsikt, for deretter å benytte innsikten i dagens allerede etablerte aktiviteter og automasjoner i alle kanaler. Videre vil bedriften benytte seg av dataen til å anbefale de rette produktene på et individuelt nivå basert på blant annet tidligere kjøp og livssituasjon. Det vil si at de har til hensikt å benytte seg av Big data som kan gi et tydeligere bilde av kundens livssituasjon. I tillegg er det et ønske om at denne innsikten skal kunne anvendes i flere deler av verdikjeden, ikke bare innenfor markedsføring og kommunikasjon, men også med tanke på innkjøpsstrategi og fysisk plassering i butikk.

«Strategien videre er å øke lojaliteten til Europris. Jeg har egentlig ikke så stort fokus på å øke handlekurven, men å øke besøksfrekvensen, at de blir utsatt for kommunikasjon som vi ser er relevante sånn at de besøker oss oftere enn de gjør i dag.» - Prosjektdeltaker 2

Ønsket om større utnyttelse av tilgjengelig data og økt datadrevet innsikt bygger på den mulige gevinsten Europris ser at de kan oppnå med personalisering. Bedriften har derfor satt seg et tentativt styringsmål om å ha 50 automasjoner innenfor personalisering i løpet av 2021. Dette innebærer både små og store automasjoner. Med utgangspunkt i alle de rundt 700 000 kundeklubbmedlemmene er 400 000 valgt ut for å være et statisk segment som kontinuerlig skal måles.

De ønsker å måle fire ulike kundesegmenter:

- 1. Personalisering - kunden får personlige produktanbefalinger.*
- 2. Automasjoner - dersom en kunde handler et spesifikt produkt vil den få en spesifikk kommunikasjon som er automatisert.*
- 3. Kundegruppen får både personalisering og automasjoner.*
- 4. Kundegruppe som ikke får noen av kommunikasjonsformene over.*

Europris vil måle disse fire segmentene gjennom 2021 og analysere hvordan de utvikler seg basert på de ulike tiltakene. De blir målt på et overordnet nivå og på aktivitetsnivå, som vil si at de blir målt på blant annet kunde verdi totalt, totalen på handlekurven, antall varer i handlekurven og frekvens til butikk. Det vil også bli sendt ut spørreundersøkelser, for å se hva kunden synes om kommunikasjonen de blir utsatt for.

5.2 Datadrevet innsikt

5.2.1 Modellutvelgelse

Forskningsspørsmål nummer to er «*Hvordan kan analyse av transaksjonsdata hjelpe Europris forbedre personalisering av kommunikasjonen til medlemmer av kundeklubben?*». For å svare på dette forskningsspørsmålet ble det først gjennomført en evaluering av hvilken metode for analyse av transaksjonsdata som ville være fordelaktig å ta i bruk. Utvelgelsen beror i hovedsak på informasjon vedrørende Europris sin nåværende utnyttelse av transaksjonsdata. Et ønske var å kunne illustrere hvilke muligheter de har for å benytte datadrevet innsikt i større grad enn de gjør i dag. Det vil være lite hensiktsmessig å illustrere en metode som er langt i fra det Europris har kompetanse, ressurser og muligheter til å implementere. Våren 2021 begynte Europris å benytte

historiske data for å treffe sine kunder med mer relevant kommunikasjon på sine nyhetsbrev. Det vil derfor lønne seg å velge en modell som kan bidra til nettopp dette.

Datasettet som undersøkes, består av et utvalg fra anonymisert transaksjonsdata fra én butikk over én måned. Resultatene fra modellen må passe studiets formål, det vil si at modellen må kunne finne skjulte mønster i dataen som kan skape beslutningsstøtte og et grunnlag for kommunikasjonsstrategier som kryss-salg. Før resultatene fra den valgte modellen blir presentert, legges det frem en forklaring på hvorfor modellen ble valgt. I teorikapittelet ble det redegjort for tre ulike maskinlæringsmodeller: assosiasjonsregler, klyngeanalyse og collaborative filtering. Collaborative filtering er en anbefalingsalgoritme, som ofte benyttes på nettsider for å anbefale produkter til kundene og dataen som benyttes er i utgangspunktet produkthanbefalinger (Blattberg et al., 2008). Ettersom studiet i hovedsak skal ta for seg nyhetsbrev og datasettet ikke inneholder informasjon om produktvurderinger, vil ikke denne metoden benyttes for å illustrere hvordan Europris kan bruke datadrevet innsikt.

Klyngeanalyse er også beskrevet i teorikapittelet. Ifølge Blattberg et al. (2008) er ikke klyngeanalyse like godt egnet til personaliserte anbefalinger, som andre lignende modeller. Modellen sier heller ikke noe om forholdet mellom kombinasjonen av produktene og årsakssammenhengen (Buttkus & Eberenz, 2019). Enkeltprodukter som kjøpes ofte vil derfor dominere resultatene. Med tanke på Europris sitt pågående prosjekt og mål om personalisert kommunikasjon på individnivå vil ikke denne modellen illustreres i dette studiet.

Assosiasjonsregler kan være en god modell for å minimere utfordringene ved klyngeanalyse, da den blant annet estimerer sannsynligheten av å kjøpe et produkt basert på kjøp av et annet (Buttkus & Eberenz, 2019). Assosiasjonsregler er spesielt passende til store datasett, som for eksempel transaksjonsdata fra et supermarked. I tillegg er det en data mining-teknikk som er både enkel å forstå og gjennomføre, og resultatet er enkelt å overføre til markedsførings- og kommunikasjonsstrategier (Blattberg et al., 2008). Modellen kan videre bidra med å identifisere hvilke kunder som er passende som mottakere av mailkommunikasjon (Anderson, Jolly & Fairhurst, 2007).

I tabell 5 finnes en oppsummering av hvilke utvelgelseskriterier som er benyttet i valg av modell. Tabellen viser at assosiasjonsregler er den metoden som er best egnet for vårt formål med bakgrunn i hvor langt Europris har kommet med sine analyser og hva de ønsker å oppnå fremover.

Tabell 5 - Oppsummering av modellutvelgelse

	<i>Assosiasjonsregler</i>	<i>Klyngeanalyse</i>	<i>Samarbeidssystemer</i>
<i>Kompleksitet</i>	Lav	Lav	Medium/Høy
<i>Krav til CRM-system og analytisk infrastruktur</i>	Lav	Lav	Medium/Høy
<i>Godt egnet for transaksjonsdata</i>	Ja	Ja	Nei
<i>Personalisering av direkte e-post</i>	Ja	Ja	Nei
<i>Betinget sannsynlighet</i>	Ja	Nei	Ja

5.2.2 Assosiasjonsregler

På bakgrunn av modellutvelgelsesprosessen har assosiasjonsregler blitt gjennomført, også kalt market basket analyse. Det er viktig å understreke at målet med denne analysen å illustrere hvordan dette kan gjennomføres, og resultatene vil i liten grad være signifikant ettersom datautvalget er lite i størrelse og over en relativt kort tidsperiode. Fokuset blir derfor å se på hvilke implikasjoner assosiasjonsregler har og hvordan analysen kan bidra i produktutvelgelse som beslutningstøtte.

For å skape forståelse av de videre analysene vurderes det som hensiktsmessig å gjøre en kort beskrivelse av den utvalgte dataen. Utvalget består av transaksjonsdata fra en Europrisbutikk i perioden januar 2020. Den prosesserte dataen består av 1021 handlekurver og i gjennomsnitt 6,3 (6523/1021) varer per handlekurv. Antall ulike produkter er 1162 og de 10 mest solgte varene står for 36% av antall varer solgt i utvalget. Tabell 6 viser en oversikt over de 10 mest solgte varene. Det er solgt 581 antall «Pumpesåpe», og er den varen som er solgt mest i perioden. «Pumpesåpe» var i 179 av handlekurvene og har en expected confidence på 17,5%. Med utgangspunkt i vårt datasett, kan man si at det er 17,5 % sannsynlighet at en handlekurv inneholder «Pumpesåpe».

Tabell 6 - Topp 10 varer

Topp 10 varer av 1162	Antall	Antall i %	Expected confidence
<i>Pumpesåpe</i>	581	8,9%	17,5 %
<i>Sjokolade</i>	450	6,9%	12,14%
<i>Opptenningsbriketter</i>	227	3,5%	10,57%
<i>Fryseposer</i>	196	3%	7,54%
<i>Toalettpapir</i>	179	2,8%	12,24%
<i>Garn</i>	157	2,4%	2,74%
<i>Lys</i>	155	2,4%	6,17%
<i>Kjeks</i>	143	2,2%	7,25%
<i>Kaffe</i>	124	1,9%	4,51%
<i>Snack pot</i>	122	1,9%	5,48%

For å identifisere regler benyttes SAS Enterprise Miner. Dette er en brukervennlig programvare som har mange forhåndsinnstilte verdier som gjør det enkelt å kjøre ulike maskinlæringsmodeller. Ved bruk av assosiasjonsregel noden er verdien for support level av stor interesse. Denne brukes for å sette et minimum nivå for «support», det vil si at én av varene i regelen må ha en support verdi på over 5% for å hevde at varene assosierer. Den referer til andelen av én vare med høyest frekvens (SAS, 2017). Ulike nivåer ble testet, men en support level på 5% ble valgt, da dette resulterte i regler som kunne brukes for å illustrere hvordan utvelgelse av regler kunne gjennomføres. Noen andre forhåndsinnstilte verdier som er verdt å merke seg er antall relasjoner. Denne omhandler antall varer som er med i reglene, og ble satt til maks 2. Dette for å minske kompleksiteten i forklaringer og illustrasjoner av reglene. Regler med fler enn 2 relasjoner er også interessante å se på, men medfører økt kompleksitet og det blir vanskeligere å finne interessante regler (Blattberg et al., 2008).

5.2.3 Evaluering av regler

Assosiasjonsregler havner under kategorien unsupervised learning. Fremgangsmåten for hvordan man sammenligner og evaluerer modellene er ulike for supervised og unsupervised learning. Ved supervised learning finnes det en rekke måleverktøy som sier noe om hvor godt modellen presterer. Ved unsupervised learning må man gjøre en totalvurdering som ofte beror på heuristiske argumenter (Hastie et al., 2009). I market basket analyse brukes ofte målene confidence, support

og lift. Med bakgrunn i disse scorene kan interessante regler velges ut. I det følgende kapitlet oppsummeres hvilke regler som scorer best på disse tre målene.

Support

Ifølge analysen er den hyppigste varekombinasjonen «Toalettpapir & Tørkerull». Det vil si at 2,94% av alle handlekurvene i utvalget inneholder disse to varene. Ofte velges en lav support verdi for å få frem de mindre åpenbare sammenhengene (Han & Kamber, 2006). Alle reglene i vårt datasett har relativt lav support verdi. Dette er fordi det er veldig mange ulike produkter i datasettet. I tabell 7 finnes de tre assosiasjonene med høyest support.

Tabell 7 - Assosiasjoner med høyest support

<i>Assosiasjoner</i>	<i>Transaksjonsantall</i>	<i>Support %</i>
<i>Toalettpapir, Tørkerull</i>	30	2,94 %
<i>Oppvaskbørste, Fryseposer</i>	11	1,08 %
<i>SNACK POT, BRUS</i>	11	1,08 %

Fra resultatene av analysen kommer det frem at Tørkerull og Toalettpapir ble kjøpt sammen i 30 handlekurver. Det kan være utfordrende å forstå om regelen er relevant. Tørkerull og Toalettpapir har en samlet support på 2,94 %, som vil si at 2,94% av alle handlekurvene inneholder Tørkerull og Toalettpapir. Dersom dette kan indikere at den samme prosentandelen av alle fremtidige kunder handler denne kombinasjonen, vil det tilsvare relativt mange kunder og regelen kan derfor være interessant. I benyttet datasett har Toalettpapir en support på 12,24%. På grunn av den høye supporten til Toalettpapir, kan Toalettpapir og en annen tilfeldig vare også oppnå høy samlet support. I tillegg er support er symmetrisk. Det vil si at samlet sannsynligheten er lik for både «Hvis Tørkerull så Toalettpapir», som «Hvis Toalettpapir så Tørkerull». Verdien gir derfor ikke informasjon om hvordan varene kan påvirke hverandre. For å oppnå denne informasjonen er det nødvendig å se på Confidence.

Confidence

I tabell 8 vises de tre assosiasjonene med høyest confidence. I motsetning til Support er Confidence assymetrisk det vil si at den sier noe om forholdet mellom varene. Resultatet fra vår analyse viser

at «hvis balsam så shampoo» måles til den høyeste verdien av confidence blant de utvalgte reglene. Confidence indikerer at det er 52,94 % sannsynlighet for at en handlekurv som inneholder balsam også inneholder shampoo. Til sammenligning viste resultatene at det er 26,47 % sannsynlighet for at en handlekurv som inneholder shampoo også inneholder balsam. Det vil si at dersom en kunde kjøper balsam, så er det større sannsynlighet for at kunden også kjøper shampoo enn ved motsatt rekkefølge. Verdien på Expected confidence viser hva som er sannsynligheten for at en tilfeldig handlekurv inneholder consequent, med andre ord vare B. I regelen over betyr det at det er 3,33% sannsynlighet for at en tilfeldig handlekurv inneholder shampoo.

Tabell 8 - Assosiasjonsregler med høyest confidence

<i>Assosiasjoner</i>	<i>Confidence</i>	<i>Expected confidence</i>
<i>BALSAM ==> SHAMPOO</i>	52,94%	3,33%
<i>TØRKERULL ==> TOALETTPAPIR</i>	46,88%	12,24 %
<i>LYSPÆRE ==> OPPTENNINGSBRIKETTER</i>	28,57 %	10,58 %

Confidence er en god metode for å velge regler, men ikke uten ulemper. I utgangspunktet ville «Hvis Tørkerull så Toalettpapir» vært en interessant regel, men ved å se på expected confidence for Toalettpapir er denne relativt høy med 12,24%. Det kan tyde på at sjansen for å kjøpe Toalettpapir alene er høy, og derfor kan regelen være misvisende.

Lift

Lift er et annet evalueringsparameter som prøver å overkomme utfordringene som support og lift har (Blattberg et al., 2008). Lift måler forskjellen mellom confidence og expected confidence. Parameteret er i likhet med support symmetrisk, og gir ingen informasjon om hvordan produktene påvirker hverandre. I vårt utvalg viser Lift at dersom en kunde kjøper shampoo er det 15,9 ganger mer sannsynlig at kunden også kjøper balsam og motsatt, enn kunder som ikke kjøper shampoo.

Tabell 9 - Assosiasjoner med høyest lift

<i>Assosiasjoner</i>	<i>Lift</i>
<i>BALSAM, SHAMPOO</i>	15,9
<i>SHAMPOO, DUSJSÅPE</i>	6,86
<i>SØPPELPOSER, OPPVASKBØRSTE</i>	4,54

Kryss-validering

For å evaluere modellen, utover support, confidence og lift, er det mulig å gjennomføre kryss-validering. Dette er en metode som er mye benyttet ved prediktive modeller, men det er også gjennomførbart ved deskriptive modeller som assosiasjonsregel. Datasettet kan deles inn i trening- og test sett. For deretter å sammenligne hvordan reglene som er indentifisert i trening settet presterer i test settet. Dersom samme regel blir indentifisert i de to datasettene, kan for eksempel målene på confidence sammenlignes (Van Der Aalst, 2016).

Dette ble også gjennomført på våre analyser. Datasettet ble delt inn i treningssett som bestod av 60% av utvalget, og et testsett som bestod av 40% av utvalget. Treningssettet indentifiserte nesten ingen av de samme reglene som det totale datasettet. Testsettet indentifiserte mange av de samme reglene med generell høyere confidence. Det kan være mange grunner til akkurat disse resultatene. Dette er ikke en optimal metode for å evaluere slike modeller. En av annen metode for å evaluere om utvalgte regler er interessante er gjennom å teste ulike regler i nyhetsbrev og måle grad av relevans gjennom hvor mange som klikker seg inn på nyhetsbrevet (Blattberg et al., 2008). I dette tilfelle vil denne metoden være bedre enn kryss validering. Generaliserbarhet er en utfordring ettersom Europris er en kjede med stort varesortiment, kontinuelig tilbud på ulike varer og varevariasjon mellom butikker. Tolkning av regler og ytterligere diskusjon om valg av regler og dens reliabilitet finnes i diskusjonskapittel punkt 6.3.2.

Tabell 10 - Kryss validering av regler

<i>Regel</i>	<i>Confidence 100% av utvalget</i>	<i>Confidence Train 60% av utvalget</i>	<i>Confidence Test 40% av utvalget</i>
<i>BALSAM ==> SHAMPOO</i>	52,94	62,5	-
<i>SHAMPOO ==> BALSAM</i>	26,47	38,46	-
<i>SHAMPOO ==> DUSJSÅPE</i>	23,53	-	33,33

<i>DUSJSÅPE ==> SHAMPOO</i>	22,86	-	31,82
<i>SØPPELPOSER ==> OPPVASKBØRSTE</i>	22,22	-	27,27
<i>OPPVASKBØRSTE ==> SØPPELPOSER</i>	16	-	19,35

Assosiasjonsregler er en populær metode innen detaljhandel og kan bidra som støtte i beslutningstaking. Varekombinasjonene «balsam, shampoo», «shampoo, dusjsåpe» og «søppelposer, oppvaskbørste» har i vår analyse høyest mål på både confidence og lift. Med utgangspunkt i utvalget vil disse reglene sies å ha høyest prediktiv effekt.

5.3 Forretningsverdi

I forrige delkapittel ble det identifiserte tre assosiasjonsregler basert på transaksjonsdata fra en Europrisbutikk i januar 2020. Disse ble evaluert med bakgrunn i målene support, confidence og lift. For å sikre at reglene gir verdi for Europris vil det være hensiktsmessig å se på varepris og vareantall i tillegg. Forsknings spørsmål nummer tre er «*Hva er den forventede forretningsverdien for Europris ved bruk av dataadrevet innsikt?*». I dette kapitlet illustreres hvilke regler som representerer størst verdi i norske kroner. Det er de viktigste resultatene fra utregningene som blir presentert. Se vedlegg 5 for fullstendig utregning.

5.3.1 Regler med høyest omsetning

I utvelgelsesprosessen av regler kan det være formålstjenlig å se på hvilken omsetning de ulike reglene har. Dersom en regel har tilsynelatende høy support, confidence og lift, kan det likevel være en mulighet for at varene i reglen har lav omsetning og derfor ikke er av stor interesse. For å skape et referansepunkt på hvor omsetningsnivået for hver vare ligger presenteres topp 3 varer med høyest omsetning:

Tabell 11 - Varer med høyest omsetning

<i>Plassering</i>	<i>Vare</i>	<i>Antall</i>	<i>Omsetning i NOK</i>
1	SJOKOLADE	450	20.658,-
2	TOALETTPAPIR	182	9.868,-
3	OPPTENNINGSBRIKETTER	227	9.649,-

Sjokolade er den varen med høyest omsetning og som er solgt flest ganger. Resterende varer ligger under 10.000 kroner i omsetning for den utvalgte perioden. Ingen av varene fra reglene i forrige kapittel er i denne listen, men de ligger på henholdsvis plass nummer 15 (Shampoo), 17 (Dusjsåpe), 55 (Balsam), 58 (Søppelposer) og 65 (Oppvaskbørste) av 1162 varer.

Tabell 12 viser hvilke regler som har høyest gjennomsnittlig omsetning av de tre utvalgte reglene. Kombinasjonen shampoo og dusjsåpe er den regelen som utgjøre størst verdi i omsetning. Hovedgrunnen for at «shampoo, Dusjsåpe» har høyere omsetning enn «balsam, shampoo» er at det er solgt flere dusjsåper enn balsam. Enhetsprisene er omtrent de samme på disse varene.

Tabell 12 - Gjennomsnittlig omsetning på assosiasjonsreglene

Rangering	Assosiasjoner	Gjennomsnittlig omsetning på regelen
1	SHAMPOO, DUSJSÅPE	2785,-
2	BALSAM, SHAMPOO	2278,-
3	SØPPELPOSER, OPPVASKBØRSTE	788,-

Hovedpoenget med å se på omsetningen på de ulike reglene er for å sammenligne beløpene med hverandre. Ved valg mellom de tre foreslåtte reglene ville det være mer naturlig å velge «Shampoo, Dusjsåpe» og «Balsam, Shampoo» sammenlignet med «Søppelposer, Oppvaskbørste» som har relativt mye lavere omsetning.

5.3.2 Økning av consequent ved økning av antecedent

I videre utvelgelse av regler kan det være av interesse å se på hvordan salget av antecedent kan påvirke salget av consequent. I tabell 13 vises en oversikt over hvor mye 15% økt salg av antecedent, vil øke salget av consequent(B). Poenget er også å sammenligne reglene med hverandre. 15% økning ble valgt da produktene i reglene er lavverdi varer, og det var ønskelig å se en tydelig økning i consequent for å illustrere forretningsverdien.

Tabell 13 - Økning av consequent ved økning av antecedent

Dersom antall solgte	BALSAM	øker med 15%, så vil antall solgte	SHAMPOO	øke med	4,0 %
Dersom antall solgte	SHAMPOO	øker med 15%, så vil antall solgte	BALSAM	øke med	7,9 %
Dersom antall solgte	SHAMPOO	øker med 15%, så vil antall solgte	DUSJSÅPE	øke med	3,4 %
Dersom antall solgte	DUSJSÅPE	øker med 15%, så vil antall solgte	SHAMPOO	øke med	3,5 %
Dersom antall solgte	SØPPELPOSER	øker med 15%, så vil antall solgte	OPPVASKBØRSTE	øke med	4,6 %
Dersom antall solgte	OPPVASKBØRSTE	øker med 15%, så vil antall solgte	SØPPELPOSE	øke med	1,7 %

«Hvis shampoo så balsam» er den kombinasjonen som potensielt kan øke salget av consequent mest. Dersom antall solgte shampoo øker med 15%, vil potensielt antall solgte balsam øke med 7,9%. «Hvis oppvaskbørste så søppelpose» er den regelen hvor consequent øker minst med henholdsvis 1,7%.

Tidligere ble det sett på hvilken regel som ga høyest omsetning med utgangspunkt i gjennomsnittlig varepris og antall solgte varer. Det kan også være interessant å se på hvilken regel som gir høyest omsetning gitt at salg av antecedent kan si noe om sannsynligheten for salg av consequent. I tabell 14 vises økning i omsetning gitt at salget av antecedent øker med 15%, og consequent øker med den betingede sannsynligheten(confidence). For eksempel indikerer våre utregninger at regelen «Hvis balsam så shampoo» gir en økt omsetning på 154,20 kr. Til sammenligning blir økningen i omsetning mye lavere dersom man tar utgangspunkt i at antecedent øker med 15%, og at consequent øker med sannsynligheten for at en tilfeldig handlekurv inneholder consequent (expected confidence). Våre utregninger viser en betydelig lavere omsetning på regelen uten bruk av den betingede sannsynligheten.

Tabell 14 - Økning i omsetning ved bruk av confidence sammenlignet med expected confidence

<i>Antecedent(A)</i>	<i>Consequent(B)</i>	<i>Økning i omsetning A->B (Regel)</i>	<i>Økning i omsetning A + B (expected confidence)</i>
<i>BALSAM</i>	SHAMPOO	154,20	90,95
<i>SHAMPOO</i>	BALSAM	300,90	257,90
<i>SHAMPOO</i>	DUSJSÅPE	292,20	260,42
<i>DUSJSÅPE</i>	SHAMPOO	222,76	171,49
<i>SØPPELPOSER</i>	OPPVASK-BØRSTE	88,33	77,94
<i>OPPVASK-BØRSTE</i>	SØPPELPOSER	51,84	45,11

Reglene er ikke nødvendigvis relevant for Europris, basert på at datasettet er lite i størrelse og over en relativt kort tidsperiode. Det er likevel ønskelig presentere hvilken verdi det kan ha for Europris å implementere ulike personaliseringstiltak i kommunikasjonsstrategien ved å benytte assosieringsregler som beslutningsstøtte. Eksempelvis er å målrettet kommunikasjonen av shampoo til kunder som kun kjøper balsam, dette er et eksempel på å ta i bruk kryss-salg som kommunikasjonsstrategi. En annen mulighet kan være å promotere bare balsam til kunder som kjøper begge slik at salget på shampoo automatisk øker i takt med salg av balsam. En siste mulighet kan være å kommunisere begge produkter samlet.

Ved å ta ut utgangspunkt i assosiasjonsregler kan våre utregninger vise at det er potensiale for å øke omsetning sammenlignet med å ikke bruke reglene. Det er i hovedsak kryss-salgsstrategier som benyttes med assosiasjonsregler og ulike muligheter for hvordan Europris kan benytte seg av dette for å øke forretningsverdi diskuteres i kapittel 6.4 og 6.5.

6. Diskusjon

I dette kapittelet diskuteres funnene fra resultatkapittelet opp mot teori og litteratur. I de to første kapitlene drøftes det hvordan Europris arbeider med personalisering og datadrevet innsikt i dag, og hvilke utfordringer de har knyttet til dette. Videre diskuteres resultatene fra analyse av transaksjonsdata, valgene som er tatt underveis og hvilke muligheter Europris har fremover. Avslutningsvis diskuteres det hvordan personlig kommunikasjon ved bruk av assosiasjonsregler kan påvirke salget og omsetningen til Europris.

6.1 Bruk av datadrevet innsikt for personalisering

Ifølge Nazarov (2019) har datadrevet innsikt blitt stadig viktigere de siste årene. Funnene fra forrige kapittel indikerer at Europris allerede har satt i gang prosesser for å benytte seg av kundedataen de har tilgjengelig. Dette gjøres i form av automatisk segmentering av kunder inn i ulike lister basert på gitte kundeaktiviteter. Automasjoner er også benyttet i forbindelse med innhold i nyhetsbrev. Disse har til nå vært basert på egenvalgte interesser på medlemsprofilen i MER-kundeklubben. Videre forklarer Europris at de er i starten av prosjektet personalisering hvor de anbefaler produkter til kunder basert på hvilken produktkategori de handler innenfor. På en side kan dette tyde på at Europris er godt i gang med å utforske mulighetene med datadrevet innsikt. På en annen side ble det forklart at Europris benytter 95% av markedsbudsjettet på kundeavisen. Dette kan tyde på at en stor del av ressursene i CRM-avdelingen ikke benyttes til digitaliseringstiltak.

I rammeverket til Huang og Rust (2021) kan aktivitetene Europris gjennomfører sammenlignes med det de kaller «mekanisk KI», som er det enkleste nivået av kunstig intelligens og innebærer datainnsamling, standardisering og segmentering. Europris har i mange år arbeidet med å samle inn data på en systematisk måte og det har gjennom intervjuene blitt identifisert at de har gode rutiner og prosesser for datainnsamling. Våre funn indikerer videre at Europris har standardisert en rekke markedsaktiviteter gjennom enkle automasjoner av kommunikasjon. Funnene antyder derfor at de er på nivå med «mekanisk KI» med tanke på datainnsamling og standardisering. De er derimot kun i startfasen av å segmentere kunder ved å identifisere preferanser. Dette hentyder til at segmentering er et område som må videreutvikles og dette er noe som vil bli videre diskutert i punkt 6.4.

På bakgrunn av rammeverket til Huang og Rust (2021) og våre funn har Europris potensialet for å utvikle seg mer mot «tenkende KI». Dette innebærer i stor grad personalisering av alle områdene i markedsføringsmiksen. Eksempelvis personalisering av produkter basert på kundens preferanser, personalisert pris, personalisert tilgang på produkter, og avslutningsvis personalisert kommunikasjon. Det sistnevnte er hovedtema for denne oppgaven og viser at Europris ønsker å personalisere mer. I utgangspunktet er det kun kommunikasjonen de har fokus på, men funnene viser at de i fremtiden også ønsker å personalisere flere områder av markedsføringsmiksen.

Personalisering av kommunikasjon i seg selv kan på mange måter være omfattende å studere, da personalisering er et stort og bredt tema. Dette kan eksempelvis innebære at firma bruker kundens navn i emnefeltet på e-poster de sender ut. Det kan også være bruk av kunstig intelligens som benytter kundens følelsesmessige preferanser i kommunikasjonen (Huang & Rust, 2021). Med andre ord kan det strekke seg fra enkle automasjoner til mer komplekse systemer. Ifølge våre funn er personalisering et satsningsområde for Europris. Hensikten er å være en aktuell detaljforhandler ved stadig økende konkurranse og konvertere flere lojale kunder. I litteraturen ser vi at Peppers et al. (2016) allerede i 1997 foreslo et strategikart for hvordan bedrifter bør arbeide med kommunikasjon til sine kunder. Denne modellen kan anses som like aktuell i dag ettersom det er rimelig å anta at mange norske detaljforhandlere befinner seg på samme sted som Europris med tanke på arbeidet med personalisering av kommunikasjon. Våre funn indikerer at Europris nå beveger seg mot datadrevet innsikt, noe som i strategikartet blir kalt database markedsføring. Det vil si at interaksjonene med kunden blir tilpasset individet. Ifølge PWC (2015) er en av årsakene til at nordiske bedrifter ikke praktiserer en-til-en kommunikasjon har vært mangel på kunnskap rundt hvordan data skal benyttes eller analyseres. I tillegg oppleves det som utfordrende å strukturere ulike datakilder. Europris har ifølge våre funn vært bevisst på dette de siste 7-8 årene, noe som kan ha gjort overgangen til å faktisk agere på dataen til en mindre krevende prosess.

Videre har Europris i sine årsrapporter vist at de er en bedrift med stadig omsetningsøkning. Funnene våre tilsier at dette har skapt usikkerhet rundt nødvendigheten for integreringen av mer datadrevet markedsføring- og kommunikasjonstiltak. Samtidig har ledelsen vært åpen for endringer så lenge CRM-avdelingen kan vise til positive resultater på sine gjennomføringer. På en

side har dette skapt en god fremgangsmåte ved at tiltak må planlegges godt, testes og evalueres før de går videre med andre tiltak. Dette er ifølge Blattberg et al. (2008) en god fremgangsmåte hvor tiltak testes gjennom eksperimenter og på den måten får innsikt i kundens preferanser. På en annen side kan usikkerheten nevnt over, skape mulige forsinkelser for personaliserings initiativer. I de kvalitative undersøkelsene ble utviklingshastighet identifisert som en gjentakende utfordring for Europris. Dette kan tyde på at det fortsatt er krevende for CRM-avdelingen å sette i gang tiltak basert på datadrevet innsikt.

6.2 utfordringer med datadrevet innsikt

I litteraturen til Han og Kamber (2006) er en av hovedutfordringene ved datadrevet innsikt de ulike datakildene. I likhet med litteraturen ble dette identifisert som en utfordring som har vært krevende også for Europris. Resultatene viser at bedriften har brukt flere år på å få samlet datakildene, slik at disse kan anvendes og benyttes til å skape verdi. Utnyttelsen av dataen har først startet det siste året. Derimot opplever bedriften fortsatt utfordringer knyttet til hvordan alle datakildene skal utnyttes til det fulle, slik at kommunikasjonen til kundene blir så relevant som mulig. Her beskrives blant annet utnyttelse av tredjepartsdata som indikerer livssituasjonen til en kunde som en fremtidig mulighet for relevansøking av produktanbefalingene i nyhetsbrev.

Videre kan datadrevet innsikt ifølge Ayyagari (2021) være en utfordrende prosess, spesielt for ikke-tekniske beslutningstakere. Et av intervjuobjektene fra Europris forteller at ulike domenekunnskap hadde vært utfordrende i personaliseringsprosjektet. Spesielt var det utfordrende med tanke på kommunikasjonen mellom de som utviklet løsningene i IT-avdelingen og de som hadde ekspertise på markedskommunikasjon. Av den grunn fungerte en av prosjektdeltakerne til Europris som en brobygger mellom IT-avdelingen og markedsavdelingen. Ifølge Han og Kamber (2006) kan domenekunnskap om fagområdet hvor prosjektet utvikles gi økt fokus og hastighet på prosessen. Det er også enklere å se interessante mønstre fra resultatene av analysene dersom en har domenekunnskap. Wang og Oppenheim (2003) trekker også frem at dersom et datadrevet innsikt skal bli vellykket må det gi informasjon raskt, være lett tilgjengelig og brukervennlig. Dette kan i stor grad dreie seg om at de ulike fagområdene må tilpasse seg hverandre. Dersom Europris vil dra nytte av implementeringen kan et viktig tiltak være å ha fokus på at de ulike fagområdene

for å få til et godt samarbeid. Det kan ses på som lite hensiktsmessig å benytte datadrevet innsikt dersom få personer kan dra nytte av informasjonen.

Utviklingshastigheten, integrasjoner og ressurser ble også lagt frem som utfordringer under intervjuprosessen. Europris nevnte at dette var utfordringer de ikke kom unna, uansett prosjekt. Da de ansatte blir tildelt oppgaver knyttet til nystartede prosjekter i tillegg til oppgavene de sitter med til vanlig. Dette bidrar til at prosjektenes utviklingshastighet ikke skjer i ønsket tempo. I sitt studie nevner Ayyagari (2021) at integrering av datadrevet innsikt kan være med på å frigjøre ressurser ved at gjentakende operasjoner blir automatisert, og den menneskelige arbeidskraften blir frigitt til å utføre andre oppgaver som igjen kan hjelpe med utviklingshastigheten på videre prosjekter. Ved at Europris er i stadig utvikling av både automasjoner, datadrevet innsikt og utnyttelse av Big data vil de ifølge studien kunne frigjøre ressurser underveis i utviklingen.

Til slutt blir det i flere av studiene i teorikapittelet nevnt at bruken av mer data, førte med seg utfordringer tilknyttet personvern og det juridiske rundt dette temaet. Nasser og Tariq (2015) fremlegger personvern som en av hovedutfordringene ved bruk av større mengder og mer personlig data. Imidlertid er dette er en utfordring som, ifølge våre funn, Europris har jobbet godt med fra start. I intervjuene ble det forklart at bedriften kom sent i gang med innsamling av gjenkjennbar data, og hadde derfor GDPR på plass fra start. Det vil si at de informerer kunden om hvilken data de henter inn og hvordan denne skal brukes, og dette er noe alle kunder må godkjenne som en del av Europris sine betingelser når man blir medlem i kundeklubben. Ayyagari (2021) påpeker i sin studie at utfordringene knyttet til bruken av data, og hvordan kunden reagerer på personalisert kommunikasjon er noe som må tas hensyn til. Funnene fra intervjuene tyder på at dette er noe Europris har tatt i betraktning i sin kommende kommunikasjonsplan. Da det er planlagt utsendelse av spørreundersøkelse til kunden for å få en tilbakemelding på hva de synes om denne typen kommunikasjon.

6.3 Analyse av transaksjonsdata for personalisering

Europris er som tidligere omtalt i startfasen av hva de kaller personalisering. Dette innebærer for bedriften at kundene vil motta mer personlige og relevant kommunikasjon i form av nyhetsbrev. For å kunne undersøke hvordan Europris kan forbedre personalisering av kommunikasjon har det

blitt tatt et valg om å ta utgangspunkt i hvordan Europris definerer personalisering, dagens prosesser, utfordringer knyttet til disse og hvilken data som skal benyttes i analysen. Europris har nylig begynt med utnyttelse av kjøpshistorikken til kundene. Hensikten er i hovedsak å øke relevans på nyhetsbrevne. I den forbindelse har transaksjonsdata blitt analysert for å se om dette kan hjelpe Europris å forbedre personalisering og relevans av kommunikasjon. I kapitlet diskuteres det kort hvorfor assosiasjonsregel-analyse blir gjennomført. Deretter diskuteres resultatene fra analysen.

6.3.1 Modellvalg

Analyse av transaksjonsdata kan utføres på flere måter. Basert på modellutvelgelsen i resultatkapitlet ble assosiasjonsregler valgt som modell. Dette er en av de mest brukte metoden innen detaljhandel (Blattberg et al., 2008), og er også egnet for transaksjonsdata (Olson & Lauhoff, 2019). Analysen kan fungere som beslutningstøtte i valgprosesser når produkter blant annet skal kryss-selges eller bli promotert sammen. I kombinasjon med fokus på personalisering kan kunden dermed oppleve å få mer relevant kommunikasjon med utgangspunkt i hva de har handlet tidligere.

Assosiasjonsregler i seg selv er en deskriptiv modell. Modellen kan skape hypoteser om hvordan virkeligheten ser ut, deretter må hypotesen testes for å si noe om verdien (Olson & Lauhoff, 2019). Ifølge våre funn gjenspeiler dette måten Europris arbeider på, hvor hypotesetesting står sentralt. Det vil si en steg for steg fremgangsmåte for å se effekten av hvert tiltak som settes inn. Dermed kan assosiasjonsregler være en passende metode for Europris ettersom de har erfaring med denne måten å jobbe på.

Videre er det rimelig å anta at transaksjonsdataen som er nødvendig for å gjennomføre analysen er lett tilgjengelig for Europris. Funnene i resultatkapitlet tilsier at dette er data de allerede har gjennomført manuelle analyser på, uten å agere noe mer på denne. Ved å benytte assosiasjonsregler kan de derfor nå finne mønster som identifiseres av modellen, og teste reglene i sin personaliserte markeds kommunikasjon med beslutningsstøtte oppnådd gjennom analysen.

6.3.2 Måling og evaluering av regler

Tabell 15 viser en oversikt over de reglene som er vurdert som mest interessante. En utfordring knyttet til de identifiserte reglene er i hovedsak at datasettet kun er et utvalg, og reglene derfor ikke direkte kan overføres til Europris sin kommunikasjonsstrategi. Hensikten er å belyse hvordan dette kan gjennomføres slik at Europris kan hente inspirasjon for hvordan de kan benytte datadrevet innsikt for mer personalisert kommunikasjon.

Tabell 15 – Assosiasjonsregler av størst interesse

Rule	Expected	Conf	Confidence	Support	Lift	Transaction Count
BALSAM ==> SHAMPOO	3,33	52,94	0,88	15,9	9	
SHAMPOO ==> BALSAM	1,67	26,47	0,88	15,9	9	
SHAMPOO ==> DUSJSÅPE	3,43	23,53	0,78	6,86	8	
DUSJSÅPE ==> SHAMPOO	3,33	22,86	0,78	6,86	8	
SØPPELPOSER ==> OPPVASKB	4,9	22,22	0,78	4,54	8	
OPPVASKBØRSTE ==> SØPPEL	3,53	16	0,78	4,54	8	

For å finne interessante regler kan det gjøres både objektive og subjektive målinger (Dahbi, Jabri, Ballouki & Gadi, 2017). Objektiv måling og evaluering av regler handler om hvor statistisk sterk regelen er. Eksempelvis kan dette være support, confidence og lift som er de tre objektive målene som har blitt anvendt i dette studie for å velge interessante regler. Disse tre målingen er utbredte metoder, men det finnes mange andre objektive mål som prøver å gjøre utvelgelsen av interessante regler enklere. I en studie gjennomført av Tan, Kumar og Srivastava (2002) ble det foreslått en rekke alternative målinger. I studiet konkluderte de med at målingene har ulike egenskaper og at ingen av de alternative målemetodene er konsekvent bedre enn andre. Videre blir det foreslått at en domeneekspert ser på egenskapene til alle målingene og deretter gjør en beslutning på hvilket mål som passer som beslutningsgrunnlag for utvelgelse av regler.

Videre er ikke alltid alle kombinasjoner like interessante til tross for at modellen statistisk sett sier at de er det (Olson & Lauhoff, 2019). Europris påpeker at det er visse varer de ikke ønsker at man skal få tilbud på, som for eksempel tobakk. Dette fører oss inn i det som kan kalles subjektive evalueringer, som kan være hvor uventet og handlingsbar regelen er (Dahbi et al., 2017). Det er vanskelig for en statistisk modell å forstå at eksempelvis øl og bleier er en uventet kombinasjon å kjøpe. Modellen viser kun hvor ofte de kjøpes sammen, men på den andre siden kan et menneske vurdere at dette er en uventet og interessant regel. Modellen kan også i liten grad oppfatte hvor

handlingsbar reglen er, med andre ord om det er mulig å benytte reglen i for eksempel markedsføringstiltak. Her kan man ta hensyn til forretningsmålene til bedriften, og hvordan reglene kan benyttes til å nå disse (Olson & Lauhoff, 2019). Når Europris skal velge ut interessante regler bør de derfor både gjøre objektive og subjektive evalueringer av reglene.

Assosiasjonsreglene er et resultat av historiske data. Derfor bør en være bevisst på faktorer som kan påvirke utfallet av analysen. En mulig påvirkningsfaktor kan være produkter som var på tilbud den gitte perioden eller ulike sesongvarer gjennom året. (Gangurde et al., 2017) foreslår blant annet å differensiere analysen på sesongvarer, og ulike salgsperioder. Det kan se ut til at dette også er tilfellet ved vår analyse. Kundeavisene for perioden som vårt datautvalg er hentet fra, viser at det var tilbud på de varene som er utvalgt som interessante (Se vedlegg 3). Ifølge Buttkus og Eberenz (2019) vil assosiasjonsregler, sammenlignet med klyngeanalyse, være mindre påvirket av varer som har vært på tilbud. Grunnen til dette er at ved å velge en lavere support terskel verdi vil algoritmen gi kombinasjoner som ikke har høyest salgsvolum. Det kan derfor være fordelaktig for Europris å benytte assosiasjonsregler. Assosiasjonsregler vil likevel være påvirket av varer som har vært på tilbud og en annen mulig løsning er å koble på tidsserie data som kan identifisere ulike trender og sesongmønster (Buttkus & Eberenz, 2019). Europris burde derfor være påpasselig med å ta spesielt hensyn til disse faktorene da de hver uke kommer ut med en kundeavis med ulike tilbud, i tillegg til at flere av bedriftens varer er sesongbasert og varebeholdningen kan variere fra butikk til butikk. Dette er alle variabler som vil være med på å påvirke utfallet av assosiasjonsreglene.

Videre kan det oppstå flere utfordringer ved å benytte assosiasjonsregler. En av de største utfordringene tilknyttet detaljhandel, er størrelsen på varesortimentet (Blattberg et al., 2008; Olson & Lauhoff, 2019). Dette ble en tydelig utfordring i gjennomføringen av vårt studie. Datasettet ble for kompleks og inkluderte for mange ulike produkter, noe som medførte at dataen måtte aggregeres slik at produktene ble plassert innenfor færre kategorier. Olson og Lauhoff (2019) foreslår å sortere alle varene inn i 30 kategorier. De mener at antall kategorier ikke må være for lite slik at analysen ikke gir noen verdifull informasjon. På den annen side kan det heller ikke være for mange kategorier, da det kan bli umulig å se hvilke kategorier som assosierer. Europris har et enormt varesortiment og en mulighet kan være å benytte informasjonen fra studiet referert til over

og på den måten fordele produktene i omkring 30 kategorier for å finne interessante regler. Valget som blir gjort avhenger også av hvilke produkter Europris ser på som verdifulle å undersøke, og bedriften bør derfor selv gjøre en vurdering og eventuelle tester for å tilpasse disse kategoriene.

Til slutt kan regler velges ut med bakgrunn i hva bedriften ønsker å oppnå, en mulighet kan være kryss-salg. Hoanca og Mock (2011) foreslår tre kriterier for at to varer skal kunne kryss-selges. Først og fremst må vare A være solgt flere ganger enn vare B. I vårt tilfelle har minst en av reglene blitt valgt uten hensyn til dette kriteriet. Dermed kan det være en mulighet for at denne regelen ikke er like relevant som andre kunne vært. Kriterium nummer to anbefaler at de to varene må være komplementære og over «saturation level». Dette betyr at det kan være utfordrende å kryss-selge to varer dersom de som mottar kryss-salget ikke kan bruke varen. Det siste kriteriet foreslår at de to varene må være innenfor prissensitivitetsrangen til kunden, med andre ord vil det være lite hensiktsmessig å kryss-selge vare til de som handler vare A, dersom vare B er mye dyrere.

6.4 Bruk av assosiasjonsregler i personalisering

I dette kapittelet diskuteres det hvordan Europris kan benytte assosiasjonsregler i sin personalisering av kommunikasjon. De identifiserte reglene fra vår analyse har kun vært eksplorativt og hensikten har vært å demonstrere hvordan utvelgelsen av regler kan foregå. En måte Europris kan benytte reglene på er eksempelvis til beskrivende analyse av tidligere kundeatferd og på den måten skape god innsikt. Dersom Europris velger å utforske datadrevet innsikt i større grad finnes det mange andre muligheter. En vanlig måte er å benytte seg av resultatene fra assosiasjonsregler til å predikere kundenes fremtidige atferd, og dermed ha mulighet til å påvirke denne. Selve bruken av resultatene kan derfor sies å være prediktiv (Gangurde et al., 2017).

En mulig måte å anvende assosiasjonsregler på, som ifølge våre funn er mest nærliggende der de er i dag, er som beslutningsstøtte med hensyn til hvilke produkter som skal promoveres i nyhetsbrev. Kostnaden ved å sende et identisk nyhetsbrev per e-post til alle kundene i kundeklubben er sannsynligvis ikke stor. Derfor kan man tenke at det ikke har noen hensikt å forbedre denne prosessen. På den andre siden er alle varene i nyhetsbrevet ikke alltid like relevant for alle kundene. Dette kan resultere i at kundene får et negativt forhold til e-postene og legger de

i søppelposten (Wu et al., 2020), og kan derfor være en god begrunnelse for hvorfor man burde utforske mulighetene innenfor datadrevet innsikt for å opprettholde kundens interesse for bedriften.

Dersom Europris velger å benytte assosiasjonsregler til utvelgelse av varer i nyhetsbrevet, blir det neste steget å finne ut av hvordan utvelgelsen skal foregå. Som presentert tidligere kan kryss-salg være et passende verktøy for å utnytte informasjonen fra analysen. Blattberg et al. (2008) mener at det finnes to ulike inngangsvinkler. Den første tar for seg varer som har høyest sannsynlighet for å bli kjøpt på tvers av kunder. Dette kalles produktsentrert kryss-salg og kan gi Europris muligheten til å velge ut lønnsomme produkter fra de identifiserte assosiasjonsreglene for videre å promotere disse. Videre kan de velge å fokusere på hver enkelt kunde og hvilke produkter som det er mest sannsynlig at kunden kjøper. Dette kalles kundesentrert og går ut på at kunden mottar et nyhetsbrev med varer som er målrettet for hen. Hvilken av disse strategiene Europris bør velge beror på om de som organisasjonen er produkt- eller kundesentrert. Det er rimelig å anta at ved økt fokus på personalisering, beveger de seg fra produkt- til kundefokus for å skape mer lojale kunder.

Som tidligere nevnt finnes det flere måter å anvende assosiasjonsreglene på, en annen løsning for Europris kan være å innføre et anbefalingssystem. Systemet kan brukes til å predikere hvilke varer fra det seneste nyhetsbrevet kunden sannsynligvis vil kjøpe, eksempelvis den neste måneden. På denne måten kan et slikt system personalisere nyhetsbrev til kunden og øke relevansen. Det er lettere for kunden å finne varer som interesserer dem (Wu, Chiang & Wu, 2020), og utvelgelsen av varer til nyhetsbrev med direkte e-post kan planlegges i forkant av utsendelse. En annen måte Europris kan dra nytte av assosiasjonsregler på, er anbefalingssystemer i sanntid (Blattberg et al., 2008). Eksempelvis gjennom å gi kunder anbefalinger på produkter som assosierer med tidligere kjøp, basert på hva andre kunder har kjøpt. Dette kan implementeres i flere kanaler som nettsiden og sosiale medier, men også gjennom automatiske utsendelse av SMS. Derimot vil sanntidsdata sette andre krav til datakilder og hvilken modell som er gjennomførbar (Blattberg et al., 2008). Europris har i dag noen automasjoner i sanntid, og det er derfor rimelig å anta at håndtering av sanntidsdata er noe de behersker.

Videre foreslår Gangurde et al. (2017) å benytte assosiasjonsregler til avdekke kundens formål med handleturen. Kunden kan for eksempel ha som formål å handle frokostartikler, hagemøbler eller tekstiler. Dersom Europris identifiserer formålet med handleturen så kan kommunikasjonen tilpasses deretter. Dette kan være til hjelp når man skal beslutte når kunden skal treffes med kommunikasjon, og hvilke produkter som er passende på ulike tidspunkt.

Ved å anvende assosiasjonsregler til personalisering er det, som presentert tidligere, mange valg bedriften må ta. Funn fra resultatkapittelet indikerer at mange av de potensielle utfordringene knyttet til dette, allerede er utfordringer Europris er bevist på. For det første er hvilke produkter som skal kommuniseres en av utfordringen som tidligere har blitt identifisert. Ved assosiasjonsregler må Europris også ta hensyn til tidligere kjøp for å skape relevant kommunikasjon. Wu et al. (2020) foreslår å løse denne utfordringen med å bruke data på hvor ofte kunden handler, og definerer hva som regnes som sjeldent og ofte. I algoritmen øker effekten av tidligere atferd ved at kunden handler ofte og synker når kunden handler sjeldnere. For å skape rikere innsikt foreslår Gangurde et al. (2017) å differensiere assosiasjonsregler basert på ukesalg versus helgesalg, månedstart versus månedslutt, differensiere ulike sesonger av året, differensiere mellom butikker og ulike kundeprofiler.

For det andre kan det være utfordrende å avgjøre hvilke kunder som skal motta mer personlig kommunikasjon. Ifølge funnene våre kan det være krevende å fokusere på hele kundebasen ut fra hvor langt Europris har kommet i sin utvikling av databasert innsikt. Ifølge Paretos lov står 20% av kundene for 80% av omsetningen (Hardy, 2010). Med hensyn til dette kan en løsning for Europris være å segmentere sine kunder inn i høy-verdi og lav-verdi ved hjelp av data mining, for deretter å benytte assosiasjonsregler til å finne skjulte mønster hos de ulike segmentene. Videre kan man predikere hva de ulike segmentene kommer til å kjøpe, og fordele markedsførings- og kommunikasjons ressursene deretter. Dessuten ble det avdekket i et studie at rundt 80% av de mest lønnsomme kundene var de samme hver måned (Wu et al., 2020). Det kan derfor virke fornuftig å fokusere på de kundene som handler ofte. På en annen side er det viktig å vurdere om disse kundene handler uansett og dermed vurdere om fokuset skal være på de lojale kundene, eller de som ikke handler like ofte.

For det tredje benyttes det i dette studie historisk data i form av transaksjonsdata, dette kan skape noen utfordringer for modelleringen og reglene som dannes. Ved at miljøforandringer som nye konkurrenter eller produkter entrer markedet, kan resultatet av modelleringen få mindre verdi da den bygger på historisk informasjon (Buttkus & Eberenz, 2019). En måte å løse dette på er ved å utføre små eksperimenter enten i butikk eller ved deler av produktporteføljen, for å blant annet identifisere utfallet av promoteringen eller den personlige kommunikasjonen. Disse eksperimentene kan benyttes til å justere gjeldene antakelser og til kryss-validering (Buttkus & Eberenz, 2019).

Gjennom dette kapittelet har det blitt diskutert hvordan Europris kan benytte seg av informasjonen fra assosiasjonsregel-analysen til å personalisere kommunikasjonen i form av nyhetsbrev, i tillegg til å dra fordel av kryss-salg som verktøy. Europris nevner i sin fremtidige plan at personaliseringen skal implementeres i alle kanaler, dette åpner opp for mange muligheter i tiden fremover. Europris kan eksempelvis utnytte assosiasjonsregler til å implementere nye funksjoner på nettside og sosiale medier. Plattformene kan brukes til å eksponere kunden for andre relevante produkter i tillegg til de som kunden selv undersøker. Videre kan det legges til prisinsentiver og tilbud til i kommunikasjonen til en gruppe kunder som kjøper varer fra relevante regler, for å se om dette har innvirkning på salget, frekvens til butikk, og antall varer i handlekurv (Griva, Bardaki, Pramadari & Papakiriakopoulos, 2018).

6.5 Forretningsverdi

I dette kapittelet diskuteres forretningsverdien Europris kan oppnå ved bruk av assosiasjonsregler, hvor mulige utfall kan være økt frekvens, større handlekurver og dermed høyere omsetning. For å oppnå verdi kan Europris, som nevnt i kapittelet over, sette i gang flere tiltak på bakgrunn av informasjonen fra assosiasjonsreglene. Disse kan blant annet benyttes til vareplassering i butikk, promotering av varer, opp-salg, mersalg og kryss-salg. Hvor kryss-salg er en av de vanligste verktøyene som kombineres med assosiasjonsregler (Blattberg et al., 2008), og det verktøyet som det har blitt fokusert på i denne oppgaven. Som tidligere nevnt betyr kryss-salg å engasjere kunden med tilleggsprodukter basert på eksisterende handlemønster (Zhang et al. (2021). Seng og Chen (2010) antyder at et mulig utfall ved å benytte seg av kryss-salg er økt salg og profitt ved at kunden tilbys produkter de har større sannsynlighet for å kjøpe. Funnene i den kvalitative delen av studiet

tilsier at Europris ønsker å tilby sine kunder nyhetsbrev med mer relevante produkter, og kryss-salg kan dermed være en god strategi for å oppnå dette. I tillegg legger Seng og Chen (2010) frem at målrettet og personlig kommunikasjon kan bidra til verdi i form av å skape fler lojale kunder. Ved å øke antallet lojale kunder vil det være rimelig å anta at dette vil resultere i økt omsetning, da kundene velger Europris fremfor andre forhandlere, legger flere varer i handlekurven og er tilbakevendende.

På en annen side nevner Zhang et al. (2021) at assosiasjonsregler kan produsere for generelle regler, som videre i kryss-salg kontekst ikke gir nok verdi og innsikt i hvilke produkter det vil lønne seg å kommunisere ut til de ulike kunden. En mulig utfordring kan også være regler som ikke gir like mye verdi for alle kunder. Hoanca og Mock (2011) legger frem et eksempel på at menn og kvinners handlemønstre og produktkombinasjoner ikke alltid er like. Videre hevder Kamakura (2008) at en av utfordringene med verktøyet kan være å overselge til noen kunder slik at de lærer å ikke respondere på kommunikasjonen som de blir eksponert for. Det er rimelig å anta at alle utfordringen over kan bidra til mindre verdiskapning, da kunder som ikke mottar kommunikasjon som inneholder relevante produkter, sannsynligvis ikke gjennomfører kjøp på bakgrunn av denne kommunikasjonen. Utvelgelse av interessante regler kan derfor være helt avgjørende for at metoden skal være effektiv. Europris bør være bevisst alle de nevnte utfordringen for å kunne skape forretningsverdi av innsikten.

Europris er en lavpriskjede og ved å undersøke transaksjonsdataen som er anvendt i dette studie kommer det frem at de fleste handlekurvene inneholder småvarer med lav pris. Ifølge Kamakura (2008) er det viktig å gjøre grundige vurderinger rundt hvilke produkter som skal promoteres for å få mest mulig nytte av regelen. En mulig innfallsvinkel er å velge ut de reglene som inneholder varer som gir høy inntekt (Kamakura, 2008). På den andre siden er det ikke nødvendigvis slik at våre resultater fra analysen gjenspeiler virkeligheten. Ved et datasett som inneholder transaksjonsdata fra flere bedrifter og lengre perioder kan regler med varer av høyere verdi være viktige å vurdere, eventuelt varer med høy margin fremfor varer med lavmargin.

For å kunne få et innblikk i mulig omsetningsøkning har det i resultatkapittelet blitt gjennomført utregninger med fokus på å øke confidence på reglene. Det vil si å øke handlekurver som

inneholder antecedent på venstre side (produkt A) og consequent på høyre side (produkt B). For å kunne ha et sammenligningsgrunnlag, ble salget på antecedent økt med 15% og omsetningen regnet ut basert på baseline, med andre ord uten bruk av regelen (expected confidence), videre ble det regnet ut økt omsetning basert på regelen. Våre utregninger indikerer at kommunikasjon som baseres på reglene, vil kunne resultere i en høyere omsetningsvekst enn ved at bedriften tar kommunikasjonsvalg på bakgrunn av intuisjon. På en annen side kan det være ressurskrevende å implementere et data mining-prosjekt, både i form av systemer, menneskelig kompetanse og personvern (Ragavi et al. (2018)). Det kan derfor være hensiktsmessig å se det totale bildet og det rimelig å anta at det er mindre krevende å gjennomføre deskriptive analyser for å opparbeide seg dypere kundeinnsikt, enn å implementere store prediktive modeller.

Ovenfor har det blitt nevnt ulike verdier Europris kan oppnå ved å bruke datadrevet innsikt i form av assosiasjonsregler, blant annet økt profitt ved bruk av kryss-salg (Seng & Chen, 2010). Det er derimot også essensielt med en god implementeringsplan for å kunne utnytte resultatene av analysen optimalt og oppnå denne forretningsverdien (Chapman et al., 2000). Gjennom oppgaven har det blitt lagt vekt på at assosiasjonsregler skal kunne implementeres i dagens prosesser, systemer og aktiviteter. Ved å gjøre det på denne måten vil det være enkelt for Europris å benytte reglene uten å måtte gjøre mange nye investeringer. I tillegg kan det være viktig å ta i betraktning hvordan kunden opplever mer personlig kommunikasjon. Dersom kommunikasjonen ikke oppleves som mer relevant for kunden kan dette få negative konsekvenser for Europris. Det kan derfor være nødvendig å gjennomføre denne type kundeundersøkelser som første steg i implementeringsplanen. Derimot bør det presiseres at implementeringen ifølge Chapman et al. (2000) ikke er siste steg i prosessen, men at det er viktig med videre arbeid for hele tiden å forbedre bruken av datadrevet innsikt. Da læringspunkter fra prosessen og fra den løsningen som har blitt implementert kan utløse nye og mer konkrete forretnings spørsmål (Chapman et al., 2000).

7. Oppsummering

7.1 Konklusjon

Formålet med denne oppgaven har vært å undersøke hvordan Europris utnytter tilgangen de har på kundedata og hvilket potensial som ligger i datadrevet innsikt. Oppgaven er begrenset til personalisering av kommunikasjon, og hvordan analyse av transaksjonsdata kan benyttes på dette området. For å sikre at formålet med oppgaven blir ivaretatt ble det utarbeidet tre forskningsspørsmål som i det neste vil bli besvart:

Hvordan jobber Europris med personalisering og datadrevet innsikt i dag, og hvilke utfordringer har de knyttet til dette?

Gjennom kvalitative undersøkelser ble det identifisert at Europris benytter automasjoner i sitt arbeid med personalisering. Automasjonene bygger på konkrete aktiviteter kunden utfører som trigger utsending av kommunikasjon i form av e-post. Videre ble det avdekket at Europris hadde vanskeligheter knyttet til hvordan denne kundedataen kunne anvendes for å skape verdi. Det er igangsatt en prosess for å benytte historiske transaksjonsdata som et tiltak for å tilby mer relevant personalisert kommunikasjon. Funnene indikerte at målene fremover var å kommunisere med kunden en-til-en, bidra til økt frekvens, skape lojale kunder og øke antall varer i handlekurvene.

I forbindelse med personaliseringsprosjektet har denne studien identifisert noen hovedutfordringer. Domeneutfordringer, utnyttelse av data, produktutvelgelse til kommunikasjon i tillegg til utviklingshastighet, integrasjoner og ressurser blir utpekt som de mest krevende utfordringene. Oppsummert kan funnene tyde på at Europris har liten innsikt om kjøpsatferden til kundene sine og hvordan de eventuelt skal agere på denne innsikten.

Hvordan kan analyse av transaksjonsdata hjelpe Europris å forbedre personalisering av kommunikasjonen til medlemmer av kundeklubben?

Analyse av transaksjonsdata kan gjennomføres på mange ulike måter med ulike formål. I denne oppgaven ble det gjort en modellutvelgelsesprosess som tok i betraktning funnene fra den kvalitative undersøkelsen, innholdet i datautvalget og formålet med oppgaven. Av de modellene som ble lagt frem som alternativer, var det assosiasjonsregler som ble identifisert som mest

passende. Ved bruk av assosiasjonsregler kan det være en utfordring å finne interessante regler. I resultatdelen ble de objektive målene support, confidence og lift, brukt for å finne de reglene som hadde høy prediktiv verdi. Videre ble det drøftet betydningen av å benytte subjektive mål, i tillegg for å sikre at reglene kan benyttes i personalisering av kommunikasjon. Reglene inneholdt varekombinasjoner som kan bidra som støtte i beslutningstaking ved utvalg av produkter for mer personaliserte nyhetsbrev.

Hva er den forventede forretningsverdien for Europris ved bruk av personalisering og datadrevet innsikt?

Denne oppgaven fokuserer på hvordan relevant kommunikasjon kan bidra til økt omsetning og på den måten bidra til forretningsverdi for Europris. Det foreslås i denne oppgaven å se på mulighetene for å anvende assosiasjonsregler i kryss-salgsstrategier og på den måten skape mer personlig kommunikasjon. Funnene illustrerer en mulig omsetningsøkning dersom reglene har prediktiv verdi. Det ble gjennomført utregninger bakgrunn av varepris og antall solgte enheter. Basert på funnene kan det være hensiktsmessig for Europris å fokusere på assosiasjonsregler i personalisering av nyhetsbrev, fremfor tilfeldig valgte varer. Da dette potensielt vil gi større omsetning, økt relevans i nyhetsbrev og flere lojale kunder.

Implikasjoner

Ulike rammeverk og litteratur har blitt benyttet for å kunne diskutere resultatene i oppgaven. Studiet finner støtte for funnene som har blitt identifisert i både de kvalitative og kvantitative undersøkelsene. Blant annet er alle utfordringene som ble avdekket i forbindelse med personaliseringsprosjektet velkjente utfordringer gjengitt i teori og litteratur. Videre blir det tydelig gjennom studie at forskningen identifiserer flere muligheter for utnyttelsen av datadrevet innsikt enn Europris har uttrykt, men dette kan ses på som naturlig da Europris enda er i startfasen av utviklingen. Med andre ord kan case-studiet ha bidratt til å bekrefte at mye av litteraturen på feltet er relevant i dag, og at Europris kan ha nytte av denne informasjonen fremover.

Funnene indikerer i praksis at Europris har en god grunnmur for prosessen videre mot en personalisert en-til-en kommunikasjonsstrategi, basert på datadrevet innsikt. Ved å benytte funnene fra dette studiet som utgangspunkt og beslutningsstøtte i disse prosessene, vil bedriften

også kunne treffe kundene med mer relevant informasjon. Samtidig som de kommuniserer produkter kunden ikke kjøper i dag og slik kan påvirke både frekvensen og handlekurven. Europris kan også bygge videre på våre analyser ved å lage prediktive modeller. Eksempelvis et anbefalingssystem som forteller kunden hva andre kunder som handlet det samme, også kjøpte. På den måten blir handleprosessen mer personlig for kunden.

Styrker og svakheter

I denne studien har det kun blitt gjennomført undersøkelser på en bedrift, Europris. Dette kan resultere i svakheter når det gjelder generalisering av funnene. Valg har blitt tatt på bakgrunn av hvor langt bedriften har kommet i å benytte datadrevet innsikt i personalisering og hvilke utfordringer og mål de har satt for fremtiden. Det er dermed ikke sagt at andre tilsvarende bedrifter har samme tilgang på data og systemer som Europris benytter i dag. På en annen måte har informasjonen innhentet fra intervjuer med de ansatte i Europris gitt nyttig innsikt om temaet, da intervjuene ble gjennomført med ansatte fra ulike fagområder i Europris. Dette styrker dermed oppgaven da ulike perspektiver, kompetanse og kunnskap preger svarene.

Studiet har også i første omgang handlet om hvordan Europris kan benytte seg av datainnsikt i personalisering av kommunikasjon. Forbrukerens perspektiv på dette har ikke vært hovedfokus og kan være med på å begrense funnene, da kunnskapen om hvordan kunder av Europris reagerer på mer personlig kommunikasjon ikke er undersøkt. Ved å utvide tidsperioden kunne denne informasjonen vært innhentet i form av eksempelvis spørreundersøkelser til et utvalg kunder, og dermed kunne dette vært med på å styrke oppgavens resultater.

Selv om studiet konkluderer med at Europris kan ha nytte av å bruke assosiasjonsregler i personaliseringsprosessen, er ikke dette nødvendigvis tilfellet i realiteten da det ikke har vært mulighet for å teste antakelsene grunnet begrenset tid. Hadde prosessen vært fulgt over en lenger periode ville man med større sannsynlighet sagt noe om utfallet av implementeringen.

7.2 Videre forskning

Dette studiet har tatt for seg hvordan en detaljforhandler i Norge arbeider med datadrevet innsikt og personalisering i dag, hvor en metode for utnyttelse av data og hvordan denne prosessen kan

utvikles har blitt lagt frem. Da studiet tar for seg en enkel bedrift er et naturlig neste steg i forskningen å studere flere aktører innenfor samme bransje for å utvide litteraturen på området. Det kunne derfor være interessant å få innblikk i større selskaper for å danne et sammenligningsgrunnlag for norske bedrifter.

Studiet har synliggjort utfordringene ved å benytte mer data og agere på denne. I tillegg til utfordringer som kan løses ved å implementere datadrevet innsikt. Det kunne derfor vært interessant å etterprøve på et senere tidspunkt om Europris har opplevd beslutningsstøtte ved å benytte assosiasjonsregler i utvelgelsen av produkter til kommunikasjon. Videre hadde det vært interessant å se på resultatet av implementeringen av mer personaliserte tiltak både hos Europris og i andre norske bedrifter. Ved å undersøke hvordan forbrukerne reagerer og om personaliseringen har en positiv effekt på omsetning, salg og kostander.

Dette studiet har i hovedsak fokusert på én data mining-modell, det kunne derfor vært interessant å utvide informasjonen i datasettet og undersøkt om andre modeller kunne gitt annen verdifull innsikt. For videre å utforske hvilken innvirkning denne innsikten kunne hatt på personaliseringen. Studiet er også avgrenset til å omhandle personalisering innenfor kommunikasjon i form av nyhetsbrev. Personalisering er et stort tema, og det kunne derfor vært spennende å undersøke videre personalisering ved datadrevet innsikt i andre deler av markedsføringsmiksen og via andre kanaler.

Litteraturliste

- Anderson, J. L., Jolly, L. D. & Fairhurst, A. E. (2007). Customer relationship management in retailing: A content analysis of retail trade journals. *Journal of Retailing and Consumer Services*, 14(6), 394-399.
- Anshari, M., Almunawar, M. N., Lim, S. A. & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, 15(2), 94-101.
- Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., ... Neslin, S. (2008). Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters*, 19(3), 305-321.
- Ayyagari, M. R. (2021). A framework for analytical CRM assessments challenges and recommendations. *Int. J. Bus. Soc. Sci*, 10(5).
- Blattberg, R. C., Kim, B.-D. & Neslin, S. A. (2008). *Database Marketing: Analyzing and Managing Customers* Springer.
- Bradlow, E. T., Gangwar, M., Kopalle, P. & Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1), 79-95.
- Buttkus, M. & Eberenz, R. (2019). *Performance Management in Retail and the Consumer Goods Industry : Best Practices and Case Studies* (1st ed. 2019. utg.). Cham: Springer International Publishing : Imprint: Springer.
- Buttle, F. & Maklan, S. (2015). *Customer relationship management: concepts and technologies* (3rd. utg.)Routledge.
- Capgemini. (2020). Datadrevet innsikt har blitt avgjørende for virksomheters suksess, men nordiske aktører sover i timen. I. My news desk. Hentet fra <https://www.mynewsdesk.com/no/capgemini-norge/pressreleases/datadrevet-innsikt-har-blitt-avgjoerende-for-virksomheters-suksess-men-nordiske-aktoerer-sover-i-timen-3057513>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide* (bd. 9).

- Dahbi, A., Jabri, S., Ballouki, Y. & Gadi, T. (2017). A new method to select the interesting association rules with multiple criteria. *International Journal of Intelligent Engineering and Systems*, 10(5), 191-200.
- Dataiku. (u.å.). Dataiku. Hentet fra <https://www.dataiku.com/>
- Dubois, A. & Gadde, L.-E. (2002). Systematic combining: an abductive approach to case research. *Journal of business research*, 55(7), 553-560.
- Europris. (2019). *Europris ASA annual report 2019*. Hentet fra https://s22.q4cdn.com/579442476/files/doc_financials/annual/Europris-ASA-annual-report-2019.pdf
- Europris. (2020). Annual report Europris ASA 2020. Hentet fra https://s22.q4cdn.com/579442476/files/doc_financials/2020/ar/2020-A%CC%8Arsrapport_samlet_low.pdf
- Europris. (u.å.). Historien om Europris. Hentet 27.04.2021 fra <https://www.europris.no/ep/historie>
- Fan, H. & Poole, M. S. (2006). What is personalization? Perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3-4), 179-202.
- Faridizadeh, S., Abdolvand, N. & Rajae Harandi, S. (2018). Market Basket Analysis Using Community Detection Approach: A Real Case. I(s. 177-198). Cham: Cham: Springer International Publishing.
- Gangurde, R., Kumar, B. & Gore, D. S. (2017). Building prediction model using market basket analysis. *Int. J. Innov. Res. Comput. Commun. Eng*, 5(2), 1302-1309.
- Griva, A., Bardaki, C., Pramataris, K. & Papanikolaou, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, 100, 1-16.
- Han, J. & Kamber, M. (2006). *Data mining: Concepts and techniques* (bd. 10) Morgan Kaufmann Publishers.
- Hardy, M. (2010). Pareto's law. *The Mathematical Intelligencer*, 32(3), 38-43.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction* Springer Science & Business Media.

- Hoanca, B. & Mock, K. (2011). Using market basket analysis to estimate potential revenue increases for a small university bookstore. *Conference for Information Systems Applied Research*.
- Huang, M.-H. & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30-50.
- Jacobsen, D. I. (2005). *Hvordan gjennomføre undersøkelser? : innføring i samfunnsvitenskapelig metode* (2. utg. utg.). Kristiansand: Høyskoleforl.
- Jarek, K. & Mazurek, G. (2019). MARKETING AND ARTIFICIAL INTELLIGENCE. *Central European Business Review*, 8(2).
- Kamakura, W. A. (2008). Cross-selling: Offering the right product to the right customer at the right time. *Journal of Relationship Marketing*, 6(3-4), 41-58.
- Kelleher, J. D., Mac Namee, B. & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms* The MIT Press.
- Kumar, V. (2010). Customer relationship management. *Wiley international encyclopedia of marketing*.
- Kwon, K. & Kim, C. (2012). How to design personalization in a context of customer retention: Who personalizes what and to what extent? *Electronic commerce research and applications*, 11(2), 101-116. <https://doi.org/10.1016/j.elerap.2011.05.002>
- Liu, C. (2015). A conceptual framework of analytical CRM in Big Data age. *International Journal of Advanced Computer Science & Applications*, 1(6), 149-152.
- Lorvik, N. (2021). Europris vokste med over 30 prosent. *Nettavisen*. Hentet fra <https://www.nettavisen.no/okonomi/europris-vokste-med-over-30-prosent/s/12-95-3424083883>
- Lu, X. (2018). *An introduction to clustering techniques*. Innlegg presentert ved SAS Global Forum 2018, Denver, CO.
- Nasser, T. & Tariq, R. (2015). Big data challenges. *J Comput Eng Inf Technol* 4: 3. doi: [http://dx.doi.org/10.4172/2324.9307\(2\)](http://dx.doi.org/10.4172/2324.9307(2)).
- Nazarov, A. (2019). Big Data Driven Marketing. *International Scientific and Practical Conference on Digital Economy (ISCDE 2019)* (s. 677-680): Atlantis Press.

- Neysiani, B. S., Soltani, N., Mofidi, R. & Nadimi-Shahraki, M. H. (2019). Improve performance of association rule-based collaborative filtering recommendation systems using genetic algorithm. *Int. J. Inf Technol. Comput. Sci*, 2, 48-55.
- Olson, D. L. & Lauhoff, G. (2019). Market basket analysis. I *Descriptive Data Mining* (s. 31-44). Springer.
- Peppers, D., Rogers, M. & Kotler, P. (2016). *Managing Customer Experience and Relationships: A Strategic Framework* Wiley.
- Phorasim, P. & Yu, L. (2017). Movies recommendation system using collaborative filtering and k-means. *International Journal of Advanced Computer Research*, 7(29), 52.
- Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.
- PWC. (2015). *Hva er Big Data, og hva betyr Big Data for deg?* Hentet fra <https://www.pwc.no/no/publikasjoner/information-management/big-data.pdf>
- Ragavi, R., Srinithi, B. & Anitha Sofia, V. (2018). Data Mining Issues and Challenges: A Review,". *International Journal of Advanced Research in Computer and Communication Engineering*, (7), 4-7.
- Rigby, D. (2011). The future of shopping. *Harvard business review*, 89(12), 65-76.
- Ringdal, K. (2013). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode*. (3. utg. utg.)Fagbokforl.
- SAS. (2017). Association Node. Hentet fra <https://documentation.sas.com/doc/en/emref/14.3/n16x97j506upgin1190wrfc1rg0l.htm>
- Saunders, M., Lewis, P. & Thornhill, A. (2016). *Research methods for business students* (7th ed. utg.). Harlow: Pearson.
- Seng, J.-L. & Chen, T. (2010). An analytic approach to select data mining for business decision. *Expert Systems with Applications*, 37(12), 8042-8057.
- Shafique, U. & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- Shalev-Shwartz, S. & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms* Cambridge university press.

- Tan, P.-N., Kumar, V. & Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (s. 32-41).
- Tashakkori, A. & Creswell, J. W. (2007). *The new era of mixed methods*. I: Sage Publications.
- Van Der Aalst, W. (2016). *Data science in action*. I *Process mining*. Springer.
- Verhoef, P., Kooge, E. & Walk, N. (2016). *Creating value with big data analytics: Making smarter marketing decisions* Routledge.
- Wang, J. & Oppenheim, A. (2003). The pitfalls of knowledge discovery in databases and data mining. I *Data mining: Opportunities and challenges* (s. 220-238). IGI Global.
- Wedel, M. & Kannan, P. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97-121.
- Wifstad, K., Jenssen, T. B., Eide, L. S., Grünfeld, L. A. & Skogli, E. (2018). *Konkurrans e i dagligvaremarkedet: Konkurrans e i alle ledd*. Regjeringen.no: Menon economics. Hentet fra <https://www.regjeringen.no/contentassets/4c26f095eaaa4f9c9d001762f78bcc72/virke-dagligvare---vedlegg.pdf>
- Wrzuszczak-Noga, J. (2018). Applying Basket Analysis and RFM Tool to Analyze of Customer Logs. *International Conference on Information Systems Architecture and Technology* (s. 258-266): Springer.
- Wu, S.-J., Chiang, R.-D. & Wu, T.-F. (2020). Direct mail promotion mechanisms and their application in supermarkets. *The Journal of Supercomputing*, 76(3), 1398-1415.
- Zamil, A. M., Ahmad, A. A. & Vasista, T. (2020). Enhancing Customer Loyalty with Market Basket Analysis Using Innovative Methods: A Python Implementation Approach. *International Journal of Innovation*,.
- Zhang, L., Priestley, J., DeMaio, J., Ni, S. & Tian, X. (2021). Measuring Customer Similarity and Identifying Cross-Selling Products by Community Detection. *Big Data*, 9(2), 132-143.
- Östlund, U., Kidd, L., Wengström, Y. & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: a methodological review. *International journal of nursing studies*, 48(3), 369-383.

Vedlegg

Vedlegg 1 – Intervjuguide

Vår bakgrunn

Bakgrunn informant

- Kan du fortelle litt om deg selv? (Rolle, Bakgrunn, erfaring)

Europris og datadrevet innsikt

- Fortell litt om din rolle i sammenhengen med personaliseringsprosjektet.

Personalisering i praksis

- Hvilke datakilder skal brukes i prosjektet?
- Hvordan ser dere for dere at det skal foregå?

Nåværende bruk

- Hvilke andre gevinster har det hatt?
- Hva er tankegangen til Europris bak hvorfor man skal satse på personalisering?

Utfordringer

- Hva er de største utfordringene du ser?
- Hva ville gjort at rollen din kan utføres enklere?

Muligheter

- Hvor tror du potensialet her er?

Impact for Europris

- Hvordan kan dette hjelpe Europris i konkurransen i årene fremover?
- Finnes det en klar strategi for arbeidet med CRM hos Europris?
- Hva vet du om krysningen CRM og maskinlæring/data mining?
- Hva er motivasjonen bak implementering av data analyse i din avdeling/Europris?

Vedlegg 2 – Oversikt over alle variabler i datasettet

<input type="checkbox"/>	1	net_amount_dwh	string	▼	Auto-detect +
<input type="checkbox"/>	2	_hash	string	▼	Auto-detect +
<input type="checkbox"/>	3	operator_id	string	▼	Auto-detect +
<input type="checkbox"/>	4	gross_amount_dwh	string	▼	Auto-detect +
<input type="checkbox"/>	5	payments	string	▼	Auto-detect +
<input type="checkbox"/>	6	created_at	string	▼	Auto-detect +
<input type="checkbox"/>	7	gross_amount	string	▼	Auto-detect +
<input type="checkbox"/>	8	purchase_at	string	▼	Auto-detect +
<input type="checkbox"/>	9	amount_due_dwh	string	▼	Auto-detect +
<input type="checkbox"/>	10	_customer_hash	string	▼	Auto-detect +
<input type="checkbox"/>	11	tax_lines	string	▼	Auto-detect +
<input type="checkbox"/>	12	updated_at	string	▼	Auto-detect +
<input type="checkbox"/>	13	entry_status	string	▼	Auto-detect +
<input type="checkbox"/>	14	amount_due	string	▼	Auto-detect +
<input type="checkbox"/>	15	currency	string	▼	Auto-detect +
<input type="checkbox"/>	16	id	string	▼	Auto-detect +
<input type="checkbox"/>	17	event	string	▼	Auto-detect +
<input type="checkbox"/>	18	transaction_date	string	▼	Auto-detect +
<input type="checkbox"/>	19	transaction_id	string	▼	Auto-detect +
<input type="checkbox"/>	20	receipt_id	string	▼	Auto-detect +
<input type="checkbox"/>	21	store.name	string	▼	Auto-detect +
<input type="checkbox"/>	22	store.phone_number	string	▼	Auto-detect +
<input type="checkbox"/>	23	store.id	string	▼	Auto-detect +
<input type="checkbox"/>	24	store.organization_number	string	▼	Auto-detect +
<input type="checkbox"/>	25	store.terminal_id	string	▼	Auto-detect +
<input type="checkbox"/>	26	no_of_items	string	▼	Auto-detect +
<input type="checkbox"/>	27	created_by	string	▼	Auto-detect +
<input type="checkbox"/>	28	_account_id	string	▼	Auto-detect +
<input type="checkbox"/>	29	infocode_items	string	▼	Auto-detect +
<input type="checkbox"/>	30	account_id	string	▼	Auto-detect +
<input type="checkbox"/>	31	event_delivery	string	▼	Auto-detect +
<input type="checkbox"/>	32	net_amount	string	▼	Auto-detect +
<input type="checkbox"/>	33	items	string	▼	Auto-detect +
<input type="checkbox"/>	34	customer.pays_tax	string	▼	Auto-detect +
<input type="checkbox"/>	35	customer.type	string	▼	Auto-detect +
<input type="checkbox"/>	36	customer.token.token_id	string	▼	Auto-detect +
<input type="checkbox"/>	37	customer.token.type	string	▼	Auto-detect +
<input type="checkbox"/>	38	total_discount	string	▼	Auto-detect +
<input type="checkbox"/>	39	discounts	string	▼	Auto-detect +
<input type="checkbox"/>	40	customer_id	string	▼	Auto-detect +
<input type="checkbox"/>	41	customer.status	string	▼	Auto-detect +

Vedlegg 4 – Informasjonsskriv NSD

FRA KUNDEAVIS TIL PERSONLIG KOMMUNIKASJON: EN CASESTUDIE OM DATADREVET KUNDEINNSIKT

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke dagens situasjon i Europris sitt CRM-arbeid, og se på ulike løsninger for bruk av dataen som er tilgjengelig i dag ved å se på blant annet data mining. I dette skrevet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Vi ønsker å undersøke dagens situasjon, utfordringer og potensialet innenfor datadrevet innsikt hos Europris. Til slutt ønsker vi å se på hvilken verdi implementering av data mining kan ha for Europris sitt CRM-arbeid. Vårt ønske er å gjennomføre en kvalitativ undersøkelse, hvor datainnsamling i hovedsak skal gjøres via intervjuer med de ansatte i Europris som jobber med CRM og kundedata. Det vil også gjøres analyser av transaksjonsdata. Problemstillingen vil være av interesse både for Europris og videre vekst, men også resten av bransjen som har likhetstrekk med Europris.

Forskningsspørsmålene vi skal besvare er:

- *Hvordan jobber Europris med personalisering og datadrevet innsikt i dag, og hvilke utfordringer har de knyttet til dette?*
- *Hvordan kan analyse av transaksjonsdata hjelpe Europris forbedre personalisering av kommunikasjonen til medlemmer av kundeklubben?*
- *Hva er den forventede forretningsverdien for Europris ved bruk av personalisering og datadrevet innsikt?*

Informasjonen skal brukes som en del av vår masteroppgave.

Hvem er ansvarlig for forskningsprosjektet?

Norges miljø- og biovitenskapelige universitet er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Vi har bedt deg om å delta da du er ansatt i Europris og arbeider enten med CRM og kundedata eller har en annen relevant arbeidsoppgave for å kunne besvare våre spørsmål.

Hva innebærer det for deg å delta?

Vi ønsker å benytte oss av kvalitative undersøkelser i form av semistrukturerte intervju. Intervjuene vil i hovedsak bli gjennomført digitalt på tjenesten teams. Hvis du deltar i dette prosjektet innebærer det at du gjennomfører et intervju på ca. 45 minutter. I intervjuet vil vi tar for oss blant annet dine arbeidsoppgaver, hvordan dere behandler dagens data for å nå deres kunder i CRM-arbeidet og kunnskap rundt maskinlæring/AI. Intervjuet vil bli tatt opp på video for at vi skal kunne transkribere i ettertid.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- *Det vil kun være vi som skriver denne masteroppgaven og veiledere ved NMBU som vil ha tilgang til opplysningene som blir gitt eller gjengitt i oppgaven.*
- *Vi vil anonymisere navn og annen kontaktinformasjon når data blir lagret. Vi vil også lagre datamaterialet på forskningsservere ved NMBU.*

Opgaven vil være konfidensiell og vil ikke publiseres før avtalt tid. Det kan forekomme steder i oppgaven der du vil bli gjenkjent i form av stilling eller kunnskapen som er gitt. Men oppgaven skal bare være tilgjengelig for Europris, sensor og veiledere ved NMBU.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Opplysningene slettes innen prosjektslutt, noe som etter planen er juni 2021.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

1. innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
2. å få rettet personopplysninger om deg,
3. å få slettet personopplysninger om deg, og
4. å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra NMBU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- NMBU ved:
Tuva Brynildsen, tuva.marie.josefine.brynildsen@nmbu.no, telefon: 926 68 738.
Elise Løining, elise.loining@nmbu.no, telefon: 478 84 566
(Veileder) Mike Riess, mike.riess@nmbu.no.
- Vårt personvernombud:
Hanne Pernille Gulbrandsen, personvernombud@nmbu.no, telefon: 402 81 558

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Mike Riess

Tuva Brynildsen og Elise Løining

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i intervju.
- at opplysninger om meg publiseres slik at jeg kan gjenkjennes, det vil si at din stilling kan være gjengitt i oppgaven – med publiseres menes at studentene, veileder, sensor og Europris får tilgang til oppgaven.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Signert av prosjektdeltaker, dato)

Vedlegg 5 – Utrengninger

Resultater:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
1	Hvor mye vil 15% økt salg av antecedent(A), øke salget av consequent (B)?																		
2	Hva er verdi av salget av consequent (B) uten bruk av regel? (expected confidence)																		
3																			
4																			
5	Antecedent	Consequent	Expected C	Confidence	Support	Lift	Transaction Cou	Pris Anteced	Pris Consequ	Antall Antece	Antall Consequ	Total omsetning							
6	BALSAM	SHAMPOO	3,33	52,94	0,88	15,90	9,00	34,00	50,00	17,00	34,00	2278,00							
7	SHAMPOO	BALSAM	1,67	26,47	0,88	15,90	9,00	50,00	34,00	34,00	17,00	2278,00							
8	SHAMPOO	DUSISÅPE	3,43	23,53	0,78	6,86	8,00	50,00	31,00	34,00	35,00	2785,00							
9	DUSISÅPE	SHAMPOO	3,33	22,86	0,78	6,86	8,00	31,00	50,00	35,00	34,00	2785,00							
10	SØPPELOSER	OPPVASKBØRSTE	4,90	22,22	0,78	4,54	8,00	10,00	8,00	50,00	36,00	788,00							
11	OPPVASKBØRSTE	SØPPELOSER	3,53	16,00	0,78	4,54	8,00	8,00	10,00	36,00	50,00	788,00							
12																			
13																			
14																			
15																			
16																			
17																			
18	Antecedent	Consequent	15% økning Antecedent				Økt salg av Consequent gitt confidence				15% økning Antecedent, Expected confidence Consequent				Økt omsetning regel				
19			Økning Antall	Økning i omsetning			Antall	%Økning i antall	Økning i omsetning			Antall	%Økning i antall	Økning i omsetning			Økning i omsetning A+B	Økning i omsetning A + B	
20	BALSAM	SHAMPOO	2,55	86,70			1,35	4,0 %	67,50			0,08	0,25 %	4,25			154,20	90,95	
21	SHAMPOO	BALSAM	5,10	255,00			1,35	7,9 %	45,90			0,09	0,50 %	2,90			300,90	257,90	
22	SHAMPOO	DUSISÅPE	5,10	255,00			1,20	3,4 %	37,20			0,17	0,50 %	5,42			292,20	260,42	
23	DUSISÅPE	SHAMPOO	5,25	162,75			1,20	3,5 %	60,01			0,17	0,51 %	8,74			222,76	171,49	
24	SØPPELOSER	OPPVASKBØRSTE	7,50	75,00			1,67	4,6 %	13,33			0,37	1,02 %	2,94			88,33	77,94	
25	OPPVASKBØRSTE	SØPPELOSER	5,40	43,20			0,86	1,7 %	8,64			0,19	0,38 %	1,91			51,84	45,11	
26																			
27	Dersom antall solgte	BALSAM	øker med 15%, så vil antall solgte			SHAMPOO	øke med	4,0 %											
28	Dersom antall solgte	SHAMPOO	øker med 15%, så vil antall solgte			BALSAM	øke med	7,9 %											
29	Dersom antall solgte	SHAMPOO	øker med 15%, så vil antall solgte			DUSISÅPE	øke med	3,4 %											
30	Dersom antall solgte	DUSISÅPE	øker med 15%, så vil antall solgte			SHAMPOO	øke med	3,5 %											
31	Dersom antall solgte	SØPPELOSER	øker med 15%, så vil antall solgte			OPPVASKBØRSTE	øke med	4,6 %											
32	Dersom antall solgte	OPPVASKBØRSTE	øker med 15%, så vil antall solgte			SØPPELOSER	øke med	1,7 %											

Med formler:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
1	Hvor mye vil 15% økt salg av antecedent(A), øke salget av consequent (B)?																	
2	Hva er verdi av salget av consequent (B) uten bruk av regel? (expected confidence)																	
3																		
4																		
5	Antecedent	Consequent	Expected Confide	Confidence	Support	Lift	Transactio	Pris Antecedent	Pris Consequent	Antall Antecedent	Antall Consequent	Total omsetning						
6	BALSAM	SHAMPOO	3,33	52,94	0,88	15,9	9	34	50	17	34	=((H6*G)/H6*G)						
7	SHAMPOO	BALSAM	1,67	26,47	0,88	15,9	9	50	34	34	17	=H7*I7/H7*K7						
8	SHAMPOO	DUSISÅPE	3,43	23,53	0,78	6,86	8	50	31	34	35	=H8*J8/H8*K8						
9	DUSISÅPE	SHAMPOO	3,33	22,86	0,78	6,86	8	31	50	35	34	=H9*J9/H9*K9						
10	SØPPELOSER	OPPVASKBØRSTE	4,9	22,22	0,78	4,54	8	10	8	50	36	=H10*I10/H10*K10						
11	OPPVASKBØRSTE	SØPPELOSER	3,53	16	0,78	4,54	8	8	10	36	50	=H11*I11/H11*K11						
12																		
13																		
14																		
15																		
16																		
17																		
18	Antecedent	Consequent	15% økning Antecedent				Økt salg av Consequent gitt confidence				15% økning Antecedent, Expected confidence Consequent				Økt omsetning regel			
19			Økning Antall	Økning i omsetning			Antall	%Økning i antall	Økning i omsetning			Antall	%Økning i antall	Økning i omsetning			Økning i omsetning A+B	Økning i omsetning A + B
20	BALSAM	SHAMPOO	=H19*1,151-J6	=H19*H6			=D19*(D6/100)	=H19/K6	=H19*G6			=D19*(C6/100)	=L19/K6	=L19*G6			=E19+I9	=E19+I9
21	SHAMPOO	BALSAM	=J7*1,151-J7	=D20*H7			=D20*(D7/100)	=H20/K7	=H20*G7			=D20*(C7/100)	=L20/K7	=L20*G7			=E20+I20	=E20+I20
22	SHAMPOO	DUSISÅPE	=J8*1,151-J8	=D21*H8			=D21*(D8/100)	=H21/K8	=H21*G8			=D21*(C8/100)	=L21/K8	=L21*G8			=E21+I21	=E21+I21
23	DUSISÅPE	SHAMPOO	=J9*1,151-J9	=D22*H9			=D22*(D9/100)	=H22/K9	=H22*G9			=D22*(C9/100)	=L22/K9	=L22*G9			=E22+I22	=E22+I22
24	SØPPELOSER	OPPVASKBØRSTE	=J10*1,151-J10	=D23*H10			=D23*(D10/100)	=H23/K10	=H23*G10			=D23*(C10/100)	=L23/K10	=L23*G10			=E23+I23	=E23+I23
25	OPPVASKBØRSTE	SØPPELOSER	=J11*1,151-J11	=D24*H11			=D24*(D11/100)	=H24/K11	=H24*G11			=D24*(C11/100)	=L24/K11	=L24*G11			=E24+I24	=E24+I24
26																		
27	Dersom antall solgte	=A19	øker med 15%, så vil antall solgte			=B19	øke med	=I19										
28	Dersom antall solgte	=A20	øker med 15%, så vil antall solgte			=B20	øke med	=I20										
29	Dersom antall solgte	=A21	øker med 15%, så vil antall solgte			=B21	øke med	=I21										
30	Dersom antall solgte	=A22	øker med 15%, så vil antall solgte			=B22	øke med	=I22										
31	Dersom antall solgte	=A23	øker med 15%, så vil antall solgte			=B23	øke med	=I23										
32	Dersom antall solgte	=A24	øker med 15%, så vil antall solgte			=B24	øke med	=I24										



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway