



LUNDS UNIVERSITET

Ekonomihögskolan

Institutionen för informatik

Vägen till träffsäkra rekommendationer i e-handeln

Strategier för kvalitativa rekommendationssystem för e-handelsföretag
med begränsad datatillgång

Kandidatuppsats 15 HP, kurs SYSK16 i Informatik

Författare: Frida Hansén
Hilda Stenström

Handledare: Nicklas Holmberg

Rättande lärare: Niki Chatzipanagiotou, Benjamin Weaver

Vägen till träffsäkra rekommendationer inom e-handeln: Strategier för kvalitativa rekommendationssystem med begränsad datatillgång

ENGELSK TITEL: The Path to Accurate E-commerce Recommendations: Strategies for high-quality recommendation systems with limited data

FÖRFATTARE: Frida Hansén och Hilda Stenström

UTGIVARE: Institutionen för informatik, Ekonomihögskolan, Lunds universitet

EXAMINATOR: Osama Mansour, Docent

FRAMLAGD: Maj, 2024

DOKUMENTTYP: Kandidatuppsats

ANTAL SIDOR: 44

NYCKELORD: rekommendationssystem, e-handel, datainsamling, strategier, kallstart, gles data, begränsad data, AI, maskininlärning

SAMMANFATTNING (MAX. 200 ORD):

Rekommendationssystem (RS) inom e-handeln tenderar att stöta på signifikanta utmaningar när det kommer till begränsad datatillgång, vilket kan försämra förmågan för systemet att leverera kvalitativa och träffsäkra produktrekommendationer. Framförallt är "kallstart" och "gles data" två kritiska utmaningar kopplade till datatillgång som e-handelsföretag ständigt stöter på inom implementeringen av RS. För att utforska hur e-handelsföretag med begränsad datatillgång kan hantera dessa utmaningar presenteras forskning om olika strategier inom RS, som innefattar exempelvis hur olika datainsamlingsmetoder samt implementation av avancerade tekniker och modeller kan användas för att hantera problem kopplat till begränsad data. En djupare förståelse för hur RS kan användas i praktiken genereras från fyra intervjuer med e-handelsföretag inom varierande branscher som alla använder RS. Utifrån det empiriska och teoretiska resultatet presenteras studiens slutsats som fyra strategier vilka e-handelsföretag med begränsad datatillgång kan tillämpa för att hantera problem kopplade till begränsad datatillgång. Dessa innefattar att implementera avancerade dataanalysmetoder, fokusera på kvalitativ datainsamling, integrering av innovativa teknologier samt genom användningen av hybrida RS.

Innehåll

1 Inledning	6
1.1 Bakgrund.....	6
1.2 Problemområde och forskningsfråga.....	7
1.3 Syfte.....	8
1.4 Avgränsningar.....	8
2 Litteraturgenomgång	9
2.1 Begreppsredogörelse.....	9
2.1.1 Kallstart.....	9
2.1.2 Gles data.....	9
2.1.3 Pålitliga RS.....	10
2.2 Grundläggande funktion.....	10
2.3 Historisk utveckling.....	11
2.4 Olika typer av RS.....	11
2.4.1 Kollaborativ filtrering.....	11
Användarbaserad Kollaborativ Filtrering.....	12
Objektbaserad Kollaborativ Filtrering.....	12
2.4.2 Innehållsbaserade RS.....	13
2.4.3 Kunskapsbaserade RS.....	13
2.4.4 Hybrida RS.....	14
2.5 Strategier.....	15
2.5.1 Datainsamling.....	15
2.5.2 Integrering av algoritmer.....	16
2.5.3 Prestandamätningar.....	17
2.7 Ansvar.....	18
2.7.1 Pålitlighet.....	18
2.7.1 Integritetsskydd.....	19
2.7 Litteratursammanfattning.....	19
3 Metod	21
3.1 Metodval.....	21
3.1.1 Kvalitativ studie.....	21
3.1.2 Semistrukturerad intervju.....	21
3.1.3 Litteraturstudie.....	22
3.2 Intervjuguide.....	22
3.3 Urval av respondenter.....	24
3.4 Metod för datainsamling.....	25
3.4.1 Transkribering.....	25
3.5 Metod för dataanalys.....	25
3.6 Validitet och reliabilitet.....	26

3.7 Etik.....	27
4 Empiriskt resultat.....	28
4.1 Implementation.....	28
4.1.1 Funktion hos RS.....	28
4.1.2 Datainsamling.....	29
4.2 Datatillgång.....	30
4.2.1 Upplevda utmaningar.....	30
4.2.2 Hantering av kallstart.....	31
4.3 Strategier.....	32
4.3.1 Val av algoritmer.....	32
4.3.2 Prestandamätningar.....	33
4.4 Ansvar.....	34
4.4.2 Pålitlighet.....	34
4.4.1 Integritetsskydd.....	35
5 Diskussion.....	37
5.1 Datainsamling.....	37
5.2 Datatillgång.....	38
5.3 Strategier.....	39
5.3.1 Algoritmer och modeller.....	39
5.3.2 Prestandamätningar.....	40
5.4 Ansvar.....	41
5.4.1 Pålitlighet.....	41
5.4.2 Integritetsskydd.....	42
6 Slutsats.....	43
6.1 Vidare forskningsmöjligheter.....	44
Appendix 1.....	45
Appendix 2.....	47
Appendix 3.....	52
Appendix 4.....	57
Appendix 5.....	59
Appendix 6.....	64
Referenslista.....	65

Figurer

Figur 2.1: Kollaborativ filtrering, förenklad illustration	12
Figur 2.2: Innehållsbaserade RS, förenklad illustration	13
Figur 2.3: Kunskapsbaserade RS, förenklad illustration	14

Tabeller

Tabell 2.1: Teoretisk referensram	20
Tabell 3.1: Intervjufrågor	23
Tabell 3.2: Intervjurespondenter	24
Tabell 3.3: Kodning för dataanalys	26

1 Inledning

Inledningskapitlet syftar till att introducera det ämne som arbetet behandlar. Kapitlet inleds med en översikt över ämnets bakgrund, följt av en presentation av det problemområde som uppsatsen adresserar samt den tillhörande forskningsfrågan. Avslutningsvis avser inledningskapitlet att redogöra för uppsatsens syfte samt dess avgränsningar.

1.1 Bakgrund

Expansionen av internet och den ständiga tillgången till uppkopplade enheter har medfört stora förändringar i människors konsumtionsvanor (Holzer, 2022). Termen e-handel myntades först av den amerikanske ekonomen Theodore Levitt (1971) som förutspådde att fenomenet skulle komma att revolutionera kommersiella transaktioner genom möjliggörandet av att konsumera produkter hemifrån. Sedan Levitts tid har det skett stora förändringar inom e-handeln. Detta då e-handeln idag till stor del drivs av kundbeteende och tekniska anpassningar för att optimera användarvänligheten (Nugrahanto et al. 2023). Under det senaste decenniet har e-handelns omfattning expanderat stadigt, med en markant ökning som inträffade under åren för Covid-19-pandemin (Yltävä, 2024). Coronapandemin bidrog tillsammans med de medföljande restriktionerna till en begränsad tillgång till fysiska butiker vilket i sin tur resulterade i en kraftig tillväxt inom e-handeln (Holzer 2022). Yltävä (2024) påvisar denna ökning genom att framhålla att onlineförsäljningen höjdes från 87 miljarder kronor till 146 miljarder kronor mellan år 2019 och 2021. Holzer (2022) menar på att pandemin och restriktionerna var en bidragande faktor till att e-handeln idag används mer än någonsin tidigare.

E-handel kan beskrivas som användningen av internet för att delta i utbyte av varor, tjänster och information (Nationalencyklopedin, 2024). Utvecklingen av internet innebar en omfattande övergång från de fysiska marknaderna till att många istället började konsumera via digitala plattformar, där säljare och konsumenter kan engagera sig och genomföra transaktioner utan fysiska möten (Nugrahanto, 2023). Dessa transaktioner kan, enligt Nugrahanto (2023), förekomma mellan allt ifrån individer, företag och myndigheter för att överföra äganderätt eller rättighet att använda vissa varor eller tjänster. E-handeln kan vidare kategoriseras i följande affärsmodeller: business-to-consumer (B2C), business-to-business (B2B) och consumer-to-consumer (C2C) (Le, 2022).

Business-to-Consumer (B2C) är en affärsmodell där företag säljer produkter eller tjänster direkt till konsumenter (Le, 2022). Inom e-handeln sker denna försäljning via internet och följaktligen försvinner fysiska butiker som mellanled (Sun et al. 2022). Sun et al. (2022) förklarar att i dagens B2C-landskap inom e-handeln använder konsumenter framförallt företagets webbplatser eller digitala plattformar för att köpa produkter och tjänster. Detta inkluderar inte bara fysiska varor som kläder, elektronik och matvaror utan även digitala produkter och tjänster såsom musik, filmer och programvara. I takt med e-handelns fördelar som dess bekvämlighet, konstanta tillgänglighet och breda utbud har B2C blivit en alltmer populär och vanligt förekommande affärsmodell (Sun et al. 2022). Tillgången till globala marknader har också öppnat upp möjligheten för företag att nå konsumenter över hela världen utan geografiska begränsningar. Med tanke på att antalet konsumenter som är engagerade i omfattande onlineköp förväntas öka, understryker Sun et al. (2022) även betydelsen av den fortsatta tillväxten av B2C inom e-handel.

Utvecklingen av rekommendationssystem (RS) har spelat en avgörande roll inom e-handelns framväxt och effektivisering (Hussein et al. 2024). Närvaron av otaliga B2C-företag online har resulterat i att dagens konsumenter erbjuds ett överflöd av olika produkter. Hussein et al. (2024) beskriver att detta överflöd skapat ett dilemma för konsumenten som idag behöver navigera bland en oerhörd mängd alternativ för att hitta de produkter som bäst motsvarar deras behov och önskemål. I detta sammanhang har RS blivit angelägna och hjälpsamma för både konsumenter och företag (Hussein et al. 2024). Ett RS är ett verktyg som erbjuder användarna förslag på produkter, tjänster eller innehåll baserat på användarnas tidigare beteende, preferenser och data (Ullén, 2015). Denna teknik tillämpas över ett brett spektrum av branscher, inklusive sociala medier, streamingtjänster och e-handel för att optimera användarupplevelsen (Hussein et al. 2024). Bodduluri et al. (2024) framhåller att syftet med dessa system är att skraddarsy användarupplevelsen för att skapa anpassade förslag och således driva engagemang, generera intäkter och främja varumärkeslojalitet.

Med tanke på att RS är en funktion som tillämpas över åtskilliga branscher krävs det att en pålitlighet och tillförlitlighet för systemen upprätthålls (O'Donovan & Smyth, 2005). För att kunna försäkra detta anses det som en nödvändighet att fokusera på säkerhetsställandet av transparens, rättvisa och ansvarighet vid implementering (Beattie et al. 2022). Beattie et al. (2022) understryker vikten av att möta dessa utmaningar och efterfrågar praktiska tillvägagångssätt för att utvärdera och hantera eventuella missvisande inom RS. Således betonas värdet av utvecklarnas förståelse om att systemen upprätthåller transparens och pålitlighet samt att det respekterar användarnas integritet och dataskydd (Beattie et al. 2022).

1.2 Problemområde och forskningsfråga

Tidigare forskning visar att RS ökar konsumenternas upplevelser av en tjänst, samt deras positiva inställning till den (Li & Karahanna, 2015). Enligt Li och Karahanna (2015) är RS ett relativt nytt ämnesområde inom forskningen om informationssystem. Forskningen inom ämnet har däremot ökat under det senaste decenniet vilket beror på den snabba framväxten av digitala plattformar och tekniska framsteg inom AI och maskininlärning, som successivt bidragit till den ökande mängden tillgänglig data (Li & Karahanna, 2015). Systemen har således idag blivit allt mer komplexa, och den omfattande mängden digital data är nu större än någonsin tidigare (Bodduluri et al. 2024). Bodduluri et al. (2024) redogör för hur den tekniska utvecklingen lett fram till utvecklingen och implementeringen av många olika typer av RS över ett brett spektrum av branscher. Det snabba genomslaget av dessa system har genererat ett rikt teoretiskt landskap samt ett brett spektrum av forskningsstudier som berör RS framväxt och tillämpningsområden (Li & Karahanna, 2015).

Studier inom RS har belyst en rad utmaningar som företag stöter på när de implementerar systemen som kan potentiellt begränsa effektiviteten i algoritmernas prestanda. Hindrena varierar beroende på den specifika algoritmen men tenderar att koncentrera sig kring problem relaterade till gles data samt kallstartsproblemet, vilket är vanligt förekommande hos mindre företag eller företag med begränsad datatillgång (Bodduluri et al. 2024). Tidigare forskning belyser även utmaningen med att försäkra att RS fortsätter vara pålitliga och relevanta i verkliga situationer (O'Donovan & Smyth, 2005). För att minska risken för missledande RS bör systemets ansvarsfullhet således verka som centrala delar av implementeringen.

E-handelsföretag med begränsad tillgång till data står inför utmaningen att tillhandahålla personliga shoppingupplevelser för att öka kundnöjdhet och försäljning, samtidigt som de har

begränsade dataresurser jämfört med större konkurrenter (Bodduluri et al. 2024). Chowdhury (2022) redogör för begreppet kallstart, som är en av de största utmaningarna relaterat till implementationen av RS. Problemet uppstår när systemet är nytt eller när det finns otillräckligt med data om kunder och produkter för att systemet effektivt ska kunna generera användbara och ansvarsfulla rekommendationer (Chowdhury, 2022). Denna brist på data kan allvarligt begränsa algoritmernas förmåga att leverera personaliserade och relevanta förslag, vilket är avgörande för systemets övergripande prestanda och användarvänlighet (Payne, 2021). Det finns idag en betydande mängd befintlig forskning kring problemen kallstart och gles data. Däremot identifieras en kunskapslucka gällande rekommendationer, strategier och vägledning som e-handelsföretag med begränsad datatillgång kan använda inom implementeringen av RS för att hantera dessa problem. Dessutom finns det en begränsad tidigare forskning som erbjuder vägledning om integrering av ansvar och tillförlitlighet i implementeringen av ett pålitligt RS. Denna uppsats ämnar således att fokusera på att identifiera tekniska rekommendationer och strategier som e-handelsföretag med begränsad datatillgång kan tillämpa inom implementeringen av ett pålitligt RS.

Mot bakgrund av problemområdet som identifierats har följande forskningsfråga formulerats: *Hur kan e-handelsföretag implementera och använda rekommendationssystem för att hantera utmaningarna som följer med begränsad tillgång till data?*

1.3 Syfte

Uppsatsen syftar till att bidra med strategier och vägledning för att optimera implementering av RS hos e-handelsföretag med begränsad datatillgång. Fokus riktas mot att adressera de specifika utmaningarna som följer med gles data och kallstart. Genom att utforska hur RS används inom e-handeln ämnar uppsatsen att föreslå effektiva strategier och tekniker för e-handelsföretag att implementera inom RS, trots begränsad datatillgång.

1.4 Avgränsningar

Studien avgränsar sig till e-handelssektorn, med särskild inriktning på att identifiera strategier för företag som står inför utmaningen med begränsad datatillgång. Valet av e-handelssektorn motiveras av den växande betydelsen av RS för att öka användarupplevelse och således försäljning inom denna bransch. Vidare är valet att studera företag med begränsad datatillgång inom e-handelssektorn motiverat av att det möjliggör en analys av de unika utmaningarna och hinder som dessa företag måste övervinna på grund av begränsad data.

En ytterligare avgränsning är att endast undersöka RS för produktrekommendationer inom B2C-segmentet (business-to-consumer). Eftersom att uppsatsen avgränsas till produktrekommendationer kommer forskningen inte att beröra RS inom streamingtjänster.

2 Litteraturgenomgång

Detta kapitel inleds med redogörelse för de centrala begreppen *kallstart*, *gles data* samt *pålitliga RS*, vilka utgör kritiska komponenter i studien. Följande delar kommer att djupdyka i den existerande litteraturen och forskningen som berör RS, specifikt inom e-handel. Litteraturgenomgången syftar till att lägga en teoretisk grund för vår forskningsfråga och kommer att ge en översikt av tidigare forskning om RS funktion, systemets historiska utveckling, ansvarsfrågan kring RS samt de vanligaste typerna av RS och dess effekt på algoritmernas prestanda och träffsäkerhet i relation till mängden tillgänglig data. Kapitlets innehåll är avsett att underbygga den empiriska datainsamlingen och bistå i den efterföljande analysen och diskussionen.

2.1 Begreppsredogörelse

2.1.1 Kallstart

Kallstart är ett vanligt förekommande problem inom området för RS och avser utmaningen att på ett träffsäkert sätt kunna rekommendera objekt som nyligen lagts till i systemet och därmed saknar historisk interaktionsdata från användarna (Chowdhury, 2022). Det handlar även om svårigheten att föreslå relevant innehåll till användare som är nya för systemet och för vilka det inte finns någon tidigare aktivitetsinformation att basera rekommendationerna på (Payne, 2021). Chowdhury (2022) understryker att kallstart är ett komplext problem eftersom systemet i många fall förlitar sig på historisk data om tidigare interaktioner eller betyg för att kunna skapa träffsäkra, personliga förslag. När sådan data saknas, som är fallet med helt nya objekt, nyregistrerade användare eller helt nya e-handelsföretag, står systemet inför en betydande utmaning att leverera kvalitativa och tillförlitliga rekommendationer (Payne, 2021).

2.1.2 Gles data

Gles data är ett problem likt kallstartsproblemet inom RS, där det kan finnas ett stort antal användare, men det saknas tillräckligt med interaktioner mellan användare och objekt (Abbas, 2023). Detta fenomenet är vanligt då majoriteten av användarna interagerar med endast ett fåtal av de många tillgängliga objekten som finns i systemet (Evert & Mattisson, 2016). Abbas (2023) beskriver att denna tendens lämnar många produkter med begränsad interaktionsdata. Detta gör det svårt för algoritmerna att exakt förutspå användarnas preferenser för dessa mindre utforskade artiklar vilket potentiellt kan leda till att användarna missar värdefulla produkter. Evert & Mattisson (2016) förklarar att gles data gör det svårt för systemet att analysera användarnas preferenser och beteenden, vilket kan resultera i att användarna får missvisande och mindre relevanta rekommendationer.

2.1.3 Pålitliga RS

O'Donovan och Smyth (2005) belyser rollen och betydelsen av tillit inom implementeringen av RS. Författarna påpekar att RS kan bli missledande när de enbart baseras på användbarheter utan att fokusera på tillförlitligheten hos de rekommendationer som ges. Att ta hänsyn till rekommendationens tillförlitlighet inom utvecklingen av RS är en central del av en ansvarsfull implementering av systemet (O'Donovan & Smyth, 2005). Om aspekten inte tas i beaktelse riskerar rekommendationerna att bli missvisande och påträngande för användaren, vilket i sin tur kan minska användarnas tillfredsställelse och förtroende för systemet (O'Donovan & Smyth, 2005). Gles data och kallstartsproblemet är två exempel på fenomen som kan skada systemets kvalitet, tillförlitlighet och precision, vilka också kan generera missvisande rekommendationer till användarna (Abbas, 2023).

Ytterligare en kritisk faktor inom implementering av RS är att ta hänsyn till användarens integritet. Ricci et al. (2011) belyser utmaningen med integritetsbevarande inom RS och förklarar att i och med försöken att bygga mer utvecklade system som genererar mer träffsäkra rekommendationer, samlar systemet in allt mer användardata. Detta kan ha en negativ inverkan på användarnas integritet då användarna upplever systemet som inskränkande och påträngande (Ricci et al. 2011). Det finns därför ett behov av att designa och implementera lösningar som förnuftigt använder och lagrar användardata. Ricci et al. (2011) betonar även vikten av att säkerställa att information om användarna endast används för att rekommendera objekt och inte kan nås fritt av utomstående.

2.2 Grundläggande funktion

RS grundläggande funktion handlar huvudsakligen om att samla in data om användare och objekt (Ullén, 2015). Systemen använder sedan denna information för att skapa modeller och tillämpa algoritmer som kan bedöma och rangordna objekten baserat på hur relevanta de kan tänkas vara för en viss användare (Ullén, 2015). Dessa system genererar rekommendationer genom att antingen identifiera objekt som liknar de som användaren tidigare har favoriserat, genom att identifiera likheter mellan olika objekt eller genom att hitta användare med liknande preferenser och därefter rekommendera objekt som uppskattats av dessa likasinnade användare (Bodduluri et al. 2023).

Portugal et al. (2018) undersöker den senaste utvecklingen inom RS och hur olika tekniker tillämpats för att förbättra rekommendationens precision och relevans. Maskininlärning (ML) är den teknik som utgör kärnan i moderna RS (Portugal et al. 2018). Författarna förklarar att genom att utnyttja ML-algoritmer kan RS effektivt bearbeta och analysera stora mängder användardata, inklusive tidigare interaktioner och uttryckta preferenser. Denna datainsamling och analys möjliggör för systemet att inte bara få insikter och upptäcka mönster i användarnas beteenden utan även generera en djupare förståelse om objektens attribut och likheter (Portugal et al. 2018). På detta sätt kan algoritmerna förutse framtida preferenser, vilket resulterar i personligt anpassade rekommendationer. Det finns olika typer av maskininlärningsmodeller som används inom RS, vilka redogörs för i avsnitt 2.4.

2.3 Historisk utveckling

Det ursprungliga exemplet på ett RS lanserades år 1992, med syfte att sortera ut oönskad skräppost från e-postmeddelanden (Falk & Lindeskog, 2021). Sedan dess har användningsområdet för RS expanderat och idag använder e-handelsföretag inom har implementerat dessa system för att kunna ge skräddarsydda förslag till sina kunder (Falk & Lindeskog, 2021).

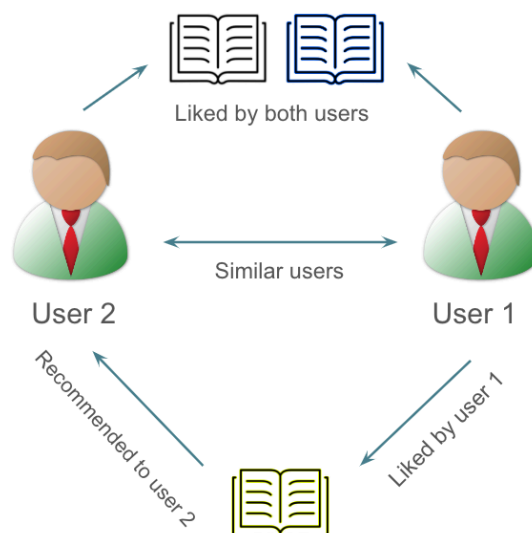
Inom e-handeln har Amazon spelat en särskilt betydande roll i utvecklingen och användningen av RS för produktrekommendationer (Smith & Linden, 2017). Systemen, som började ta form i slutet av 1990-talet, har revolutionerat sättet som företag interagerar med sina kunder på genom att erbjuda personaliserade shoppingupplevelser baserade på kundens tidigare beteenden och preferenser. Smith och Linden (2017) redogör för den produktbaserade kollaborativa filtreringen, en metod som i grunden utvecklades av Amazon och som möjliggjorde rekommendationer i en skala som tidigare inte setts. Författarna förklarar att metoden gjorde det möjligt att analysera och sortera miljontals produkter och användare för att därefter distribuera otaliga rekommendationer baserade på den insamlade datan. Lanserad år 1998, baserades denna teknik på att identifiera relaterade produkter i produktkatalogen för att skapa rekommendationer baserade på dessa likheter. Denna effektiva och skalbara metod har antagits brett, inklusive av stora webbplatser som YouTube och Netflix, tack vare teknikens kapacitet att leverera relevanta förslag baserat på användardata (Smith & Linden, 2017). Hos e-handelsföretag har RS haft en särskilt stor inverkan på försäljningen. Enligt MacKenzie et al. (2013), beräknades ca 35% av alla köp som gjordes på Amazon.com ha sitt ursprung från produktrekommendationer.

2.4 Olika typer av RS

Det finns flera olika tillvägagångssätt och metoder för att bygga upp ett RS. De tre vanligaste algoritmerna vid produktrekommendationer är kollaborativ filtrering, innehållsbaserade rekommendationer samt kunskapsbaserade rekommendationer (Bodduluri et al. 2023). Ytterligare en allt mer förekommande metod är hybrida RS, vilka kombinerar två eller flera olika typer av ML-algoritmer (Chowdhury, 2022).

2.4.1 Kollaborativ filtrering

Kollaborativ filtrering är en vanlig metod inom RS som använder historisk data från användare för att generera rekommendationer (Chowdhury, 2022). Kollaborativ filtrering utgör grunden för moderna RS och utvecklades och implementerades av Amazon år 1998 (Smith & Linden, 2017). Smith och Linden (2017) redogör för denna teknik som revolutionerade sättet kunder upptäcker produkter online på genom att tillhandahålla skräddarsydda förslag, vilka grundar sig på användarens gemensamma beteendemönster med andra användare. Det finns två huvudsakliga underkategorier av kollaborativ filtrering, användarbaserad och objektbaserad (Bentzer & Thulin, 2022). Båda varianterna bygger på preferenser mellan liknande användare eller objekt och är metoder som än idag används för rekommendationer hos företag som bland annat Netflix och Amazon (Bodduluri et al. 2023).



Figur 1: Kollaborativ filtrering, förenklad illustration

Användarbaserad Kollaborativ Filtrering

Användarbaserad kollaborativ filtrering analyserar data och beteendet hos användare med liknande intressen för att ge rekommendationer till en specifik användare (Bentzer & Thulin, 2022). Om systemet upptäcker en annan användare med liknande preferenser, en så kallad "granne", kan systemet rekommendera målkunden produkter som denna liknande användare tidigare visat intresse för (Hussein et al. 2021). Denna metod kallas grannmodellen och kretsar kring att upptäcka användare med liknande preferenser eller intressen, vilka ofta benämns som närmsta grannen (nearest neighbor). Användarbaserad kollaborativ filtrering utgår således från att användare med liknande data och historik har liknande preferenser och intressen (Bodduluri et al. 2024). Denna utgångspunkt kan enligt Hussein et al. (2021) likväl begränsa träffsäkerheten hos metoden då algoritmerna tenderar att ibland välja ut kunder som inte är representativa nog som grannar till en viss användare, vilket kan resultera i att de slutgiltiga rekommendationerna inte blir så exakta som önskat.

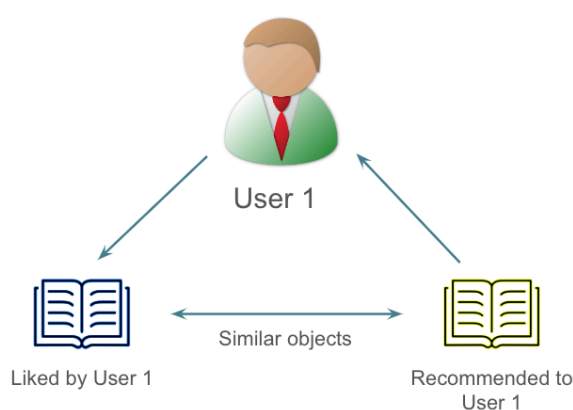
Objektbaserad Kollaborativ Filtrering

Inom objektbaserad kollaborativ uppskattar systemet hur en viss användare skulle kunna intressera sig för en produkt baserat på hur de har betygsatt eller intresserat sig för andra liknande produkter tidigare (Chowdhury, 2022). Systemet baseras på användarnas interaktiva data med olika objekt och söker efter mönster i hur olika användare har interagerat med olika objekt för att identifiera likheter mellan objekten. (Bentzer & Thulin, 2022). Om många användare som gillar objekt A även gillar objekt B, antar den objektbaserade kollaborativa filtreringen att objekt A och B är liknande (Chowdhury, 2022).

Hussein et al. (2021) redogör för bristerna inom kollaborativ filtrering och framhåller att både användarbaserad och objektbaserad kollaborativ filtrering frekvent stöter på utmaningar såsom kallstartproblem samt gles data och låg skalbarhet. Systemet förlitar sig nämligen på befintlig betygs- och användardata och om sådan saknas för vissa produkter har algoritmerna allt svårare att generera kvalitativa rekommendationer och det rekommenderade innehållet riskerar istället att bli missvisande (Hussein et al. 2021). Hussein et al. (2021) förklarar att de kollaborativa metoderna följaktligen kan vara utmanande att implementera på plattformar som saknar rik tillgång på tidigare data om användare och objekt.

2.4.2 Innehållsbaserade RS

Innehållsbaserade RS syftar till att föreslå artiklar som liknar de som en konsument tidigare uppskattat, baserat på egenskaper hos produkterna (Hussein et al. 2021). Till skillnad från kollaborativ filtrering, genomför algoritmen en mer djupgående analys av det specifika objektets innehåll för att få en förståelse för vilka egenskaper och attribut som definierar ett specifikt objekt och likheter mellan olika objekt (Bodduluri et al. 2024). Bodduluri et al. (2024) framhåller att genom att analysera och identifiera nyckelattribut hos objekten som en användare visat intresse för, kan systemet således föreslå nya, liknande objekt som matchar användarens preferenser. Hussein et al. (2021) förklarar att denna typ av RS i många fall även använder sig av information om användares beteenden, såsom klick, köp och läsaktiviteter för att beräkna en specifik användares intresse samt engagemang kring olika objekt.



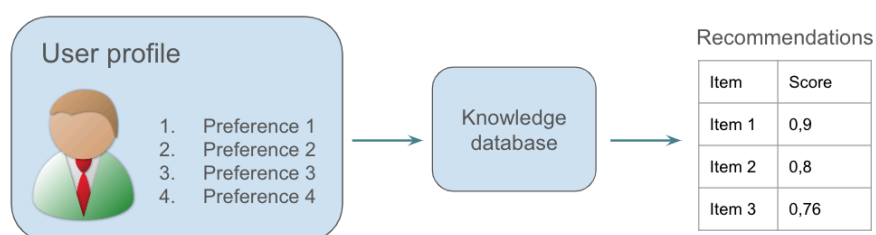
Figur 2: Innehållsbaserade RS, förenklad illustration.

Att använda denna typ av navigerings- och beteendebaserad information kan kompensera för bristen på betygsinformation som ofta förekommer inom system med kollaborativ filtreringsmetod och kan istället öka effektiviteten i systemet, vilket är en fördel med den innehållsbaserade metoden (Hussein et al. 2021). Innehållsbaserade rekommendationer är således inte drabbade av kallstartsproblemet eftersom att algoritmerna kan anpassas till introduktion av nya objekt genom att primärt analysera data om objektens attribut eller visuella egenskaper (Graham, 2024). Graham (2024) påstår att de efterföljande innehållsbaserade rekommendationerna likväl kan komma att begränsas till objekt med liknande egenskaper som det ursprungliga objektet som användaren integrerat med eller besökt, vilket inte alltid ligger i linje med det användaren söker. Hussein et al. (2021) fastställer även att den navigerings- och beteendebaserade data som systemet använder inte alltid speglar användarens verkliga preferenser, vilket är ytterligare en svaghet hos de innehållsbaserade rekommendationerna.

2.4.3 Kunskapsbaserade RS

Kunskapsbaserade RS använder en specifik kunskapsbas relaterad till en viss produktkategori för att generera förslag, enbart baserat på användarens önskemål och krav (Tran, 2007). Tran (2007) förklarar att algoritmerna inom de kunskapsbaserade systemen inte använder sig av historisk data om objekt eller användare, till skillnad från kollaborativa och innehållsbaserade RS. Systemet samlar istället in användarens specificerade preferenser eller krav för en viss produkt, via exempelvis ett formulär eller sökfält för att sedan söka genom kunskapsbasen för

att identifiera alternativ som bäst motsvarar dessa krav (Tran, 2007). Genom att inkludera användarens inmatade önskemål som den huvudsakliga datakällan, kan dessa typer av RS göra mer precisa och relevanta förslag som med stor sannolikhet matchar användarens önskemål (Hussein et al. 2021). Kunskapsbaserade RS är dock särskilt lämpade för produktkategorier som konsumeras med lägre frekvens (Tran, 2007). Tran (2007) illustrerar detta genom att ta upp fordon och fastigheter som exempel på marknader där det ofta saknas omfattande data om användares tidigare inköp och preferenser. Även om det kan finnas viss historik över tidigare preferenser gällande kategorierna, är det sannolikt att användarnas nuvarande åsikter och önskemål har förändrats över tid då dessa typer av köp sällan genomförs med hög frekvens (Tran, 2007). På dessa marknader bedöms kunskapsbaserade RS vara mer fördelaktiga jämfört med system som primärt förlitar sig på analys av historisk data om användarbeteende och produktdata (Tran, 2007). Hussein et al. (2021) beskriver att trots att kunskapsbaserade rekommendationer kan bidra till att komma över vanliga begränsningar inom traditionella RS som kallstartsproblemet och gles data, anses kunskapsbaserade RS inte vara lika träffsäkra för mindre e-handlare. Detta då systemen är mer resurskrävande då det behövs en omfattande domänkunskap och en detaljerad förståelse för produkterna som erbjuds. Författarna framhåller att mindre e-handelsföretag även kan inneha ett mer dynamiskt och snabbriktigt produktutbud, vilket kan göra det svårt att upprätthålla en aktuell kunskapsbas.



Figur 3: Kunskapsbaserade RS, förenklad illustration.

2.4.4 Hybrida RS

Hybrida RS är avancerade verktyg som skiljer sig från traditionella rekommendationstekniker genom att de kombinerar flera algoritmer, såsom kollaborativa, innehållsbaserade och kunskapsbaserade metoder, för att skapa mer precisa och anpassade förslag till användaren (Chowdhury, 2022). Bodduluri et al. (2024) förklarar att tillväxten av internet och digitala lösningar, med ökad komplexitet och volymdata, kräver mer avancerade och precisa rekommendationsalgoritmer. Genom hybridisering av RS kan man ta vara på de traditionella algoritmernas specifika fördelar och kombinera dem med andra algoritmer för att lösa utmaningar (Bodduluri et al. 2024). Vidare redogör Bodduluri et al. (2024) för hybrida systemets betydelse för e-handeln. Författarna påpekar att hybridisering av systemen kan generera mer dynamiska rekommendationer, som reflekterar de senaste användarinteraktionerna och preferenserna. Enligt Chowdhury (2022), är hybrida system särskilt användbara för att hantera kallstartsproblemet. Framgången med att använda hybrida RS för att övervinna kallstartsproblemet kan förklaras av systemets förmåga att snabbt anpassa och uppdatera rekommendationerna för att matcha användarens ständigt förändrade intressen (Bodduluri et al. 2024). Bodduluri et al. (2024) understryker att genom att inte

enbart förlita sig på användarnas tidigare beteenden kan de hybrida systemen erbjuda relevanta förslag även för helt nya användare, vilket utgör en direkt lösning av kallstartsproblemet. Alamdari et al. (2020) understryker dock att hybrida RS, särskilt de som använder big data-teknik och mer avancerade algoritmer, ofta medför stora kostnader och hög komplexitet. Dessa hybrida system kräver större bearbetningsresurser, vilket även kan leda till ökade driftskostnader (Alamdari et al. 2020).

2.5 Strategier

2.5.1 *Datainsamling*

Inom utvecklingen av RS är datakvaliteten avgörande för algoritmernas träffsäkerhet (Apsis, 2024). Det är således kritiskt att fokusera på vilken data som ska samlas in, snarare än att försöka öka kvantiteten av data. För att ge kvalitet till insamlad data är "data mining" en vanlig och effektiv datainsamlingsmetod (Han et al. 2012). Han et al. (2012) redogör för definitionen av data mining som handlar om att utvinna intressanta mönster och kunskap ur stora datamängder. Denna process innebär analys av data från olika källor såsom databaser och datalager, samt från data som kontinuerligt strömmas in i systemet (Han et al. 2012). Författarna framhåller att man genom data mining kan extrahera djupgående insikter och avslöja dold information, vilken annars inte skulle vara uppenbar vid en första anblick av rådatan.

Yan (2017) presenterar en rad strategier för datainsamling inom data mining som är centrala för utvecklingen av ett användaranpassat RS inom e-handeln. Författaren betonar vikten av att analysera och bearbeta data om användarbeteende, såsom webbläsarhistorik, transaktionshistorik och produktbetyg för att kunna erbjuda användaren personifierade rekommendationer. Strategin är avgörande eftersom att den möjliggör för RS att inte bara förstå utan även förutse användarpreferenser och beteenden (Yan, 2017).

Datainsamlingsmetoden som Yan (2017) presenterar kvalificerar sig som data mining på grund av dess fokus på att extrahera värdefull information och insikter från rådata. Metoden använder data mining för att identifiera mönster, trender och relationer i användarnas beteendedata (Yan, 2017). Dessa insikter är inte direkt uppenbara utan kräver en noggrann analysprocess där data omvandlas till användbar kunskap (Han et al. 2012). Han et al. (2012) förklarar att genom tillämpning av data mining kan systemet upptäcka underliggande tendenser bland användarna, vilket möjliggör skapandet av skraddarsydda rekommendationer som speglar individuella intressen och preferenser. Yans (2017) forskning pekar på vikten av att noggrant välja vilken data som ska samlas in och analyseras inom RS. All data är nämligen inte lika värdefull för att bygga ett kvalitativt system (Apsis, 2017). Till exempel, medan beteendecentrerade dataobjekt som transaktionshistorik och sökhistorik kan ge djupgående insikter om användarnas intressen, kan data från enkäter bidra med mindre värde då de ofta inte speglar användarnas faktiska preferenser ur ett brett, föränderligt perspektiv (Apsis, 2017). Dessa mindre relevanta interaktioner kan istället förvirra algoritmen till att generera missvisande data.

2.5.2 Integrering av algoritmer

Valet av modeller inom RS bör baseras på flera faktorer, inklusive företagets mål, den tillgängliga datans natur samt de specifika utmaningarna som e-handelsföretaget står inför (Salunke & Nichite, 2022). Salunke och Nichite (2022) förespråkar en balanserad tillämpning av dessa metoder för att öka precisionen i rekommendationerna, förbättra kundupplevelsen och i sin tur även driva försäljningen. Inom e-handeln har hybrida RS ansetts särskilt framgångsrika för att tackla problem som kallstart (Bodduluri et al. 2024). Detta eftersom att de kombinerade systemen inte förlitar sig på historisk data och kan således rekommendera både helt nya produkter och rikta rekommendationer till helt nya användare (Bodduluri et al. 2024). För nya användare eller objekt (vilka är en central del av kallstartsproblemet) erbjuder hybrida RS en lösning genom att utnyttja tillgänglig information utöver användarbetyg (Basiri et al. 2010). Vidare klargör Basiri et al. (2010) även att till exempel demografisk information om användare eller innehållsinformation om objekt kan användas för att göra initiala rekommendationer innan tillräckligt med betygdata samlats in.

Karn et al. (2022) redogör för en hybrid metod inom RS som syftar till att förbättra precision och relevans i rekommendationer inom e-handeln. Författarna beskriver ett hybridiserat RS som kombinerar kollaborativa och innehållsbaserade algoritmer tillsammans med neurala nätverk för att kunna ta hänsyn till användarnas beteendemönster, åsikter och reaktioner utöver de traditionella rekommendationsmetoderna (Karn et al. 2022). Inom den kombinerade metoden används kollaborativ filtrering för att identifiera och rekommendera produkter baserat på likheter mellan olika användares historik. Hybriden inkluderar även innehållsbaserade metoder som används för att generera rekommendationer baserade på likheter i produktattribut och användarpreferenser. Slutligen kombineras algoritmerna med neurala nätverk, vilka tillämpas för att förbättra systemets förmåga att hantera stora och komplexa datamängder (Karn et al. 2022). Det flerskiktade systemet kombineras sedan med sentimentanalys, vilket utgör processen att genom analys av digital text, avgöra den känslomässiga tonen i det som skrivs (Karn et al. 2022). Sentimentanalysen används för att identifiera och analysera känslomässiga uttryck i produktrecensioner, vilket kan generera en mer djupgående förståelse för användarnas preferenser (Karn et al. 2022). Vidare förklarar författarna att genom kombinationen av två klassiska RS algoritmer integrerade med neurala nätverk och sentimentanalys, förbättras inte bara träffsäkerheten och relevansen i produktrekommendationerna utan användarna erbjuds även en personlig och användarcentrerad shoppingupplevelse utöver de traditionella rekommendationsmetoderna.

Bodduluri et al. (2024) framhåller vikten av att e-handelsföretag kontinuerligt uppdaterar sina kunskaper om de senaste teknologiska utvecklingen för att konstant förbättra prestandan i sina RS. Exempelvis kan integrationen av djupinlärningsmodeller och naturlig språkbehandling (Natural Language Processing) förbättra systemets förmåga att leverera relevanta rekommendationer då algoritmerna kan förstå användarrecensioner och feedback ytterligare (Bodduluri et al. 2024). Författarna förklarar vidare att integrering av djupinlärning och neurala nätverk kan, liksom Karns (2022) exempel, hjälpa företag att undgå kallstartsproblemet och utmaningarna som följer med gles data. Detta eftersom att djupinlärningsalgoritmer och andra avancerade tekniker kan hantera komplexa mönster och relationer i data på ett sätt som traditionella rekommendationstekniker inte kan (Bodduluri et al. 2024). Författarna beskriver att de avancerade algoritmerna är särskilt effektiva för att identifiera dolda preferenser och därmed förutsäga användarbeteenden. En god förutsägbarhet hos algoritmerna är avgörande för att minska effekterna av gles data och kallstartsproblem hos företag med mindre datatillgång. Följaktligen framhåller författarna att e-handelsföretag bör använda

hybridiserade RS som integrerar avancerade algoritmer för att hjälpa till med att angripa dessa utmaningar (Bodduluri et al. 2024).

2.5.3 Prestandamätningar

Att mäta prestandan och träffsäkerheten av RS är av avgörande betydelse för att säkerställa att systemet ger relevanta och användbara förslag till sina användare. Shani och Gunawardana (2011) understryker vikten av att utvärdera RS på genom mätningar av ett flertal egenskaper som kan påverka användarupplevelsen, såsom systemets kvalitet och täckning. Författarna redogör för kvalitet inom RS, som är ett mått på hur väl systemets förutsägelser eller rekommendationer speglar användarnas faktiska preferenser eller intressen. Systemets täckning utgör andelen av objekt- eller användarutrymmet som systemet effektivt kan ge rekommendationer för, vilket innefattar ett mått på hur väl systemet hanterar gles data (Shani & Gunawardana, 2011). Genom att noggrant mäta och utvärdera både kvaliteten och omfattningen av rekommendationerna kan ett företag garantera ett mer tillförlitligt RS. Ett sådant system kommer bättre att stämma överens med användarens preferenser och inkludera rekommendationer för både nya och befintliga användare och produkter (Shani & Gunawardana, 2011).

Det finns flera olika mått och nyckeltal som används för att mäta RS träffsäkerhet, täckning och noggrannhet (Deutschman, 2023). Vid val av prestandamått för RS betonar Deutschman (2023) vikten av att anpassa måtten efter systemets specifika egenskaper och användning. Om systemet använder innehållsbaserad filtrering eller kollaborativ filtrering, bör mått som speglar likhet respektive noggrannhet och kvalitet av algoritmerna prioriteras. Tillgängligheten av annoterade data och om rekommenderade objekt har en specifik rangordning är ytterligare faktorer som påverkar val av lämpliga mått, där A/B-testning framhålls som ett oundgängligt verktyg för att förstå systemets faktiska affärsvärde (Deutschman, 2023). Slutligen understryker författaren att ingen enskild uppsättning mått kan fånga alla aspekter av ett systems prestanda och att användarfeedback är avgörande för att säkerställa att systemet lever upp till både användarnas och företagets behov.

Shani och Gunawardana (2011) redogör för två olika typ av prestandamätningar som visat sig vara effektiva inom implementeringen av RS. Det första metoden som författarna beskriver är ett offline-experiment. Experimentet utförs genom att använda och analysera förinsamlade dataset. Genom dessa kan utvecklare estimeras RS prestanda baserat på kvaliteten i algoritmernas förutsägelser utan att involvera riktiga användare (Shani & Gunawardana, 2011). Författarna förklarar att offline-experiment är kostnadseffektiva och tillåter jämförelse av ett brett spektrum av algoritmer. Experimentet begränsas däremot till att endast generera svar på frågor om hur träffsäkra algoritmernas förutsägelser är och kan således inte mäta rekommendationernas inverkan på användarbeteenden (Shani & Gunawardana, 2011). Den mest tillförlitliga utvärderingen av RS prestanda är att i realtid mäta hur användarbeteenden förändras som ett resultat av systemets rekommendationer. Genom ett sådant onlineexperiment kan e-handelsföretag få insikter om RS verkliga värde och effekt (Shani & Gunawardana, 2011). Utvärderingarna som författarna presenterar kan hjälpa utvecklare och företag att välja mellan olika algoritmiska tillvägagångssätt, finjustera systemparametrar och förbättra systemets övergripande prestanda och kvalitet.

2.6 Ansvar

2.6.1 Pålitlighet

Saxborn et al. (2024) introducerar att strategier för att bygga förtroende i e-handelns RS är viktiga. Författarna understryker att genom att undersöka hur kundförtroende påverkas av RS kvalitet och autenticitet, visar studien på att rekommendationers relevans och anpassning efter användarens smak och intressen är avgörande för att bygga och bibehålla förtroendet för hemsidan. Att tidigt i kundinteraktioner erbjuda träffsäkra förslag har visat sig även vara kritiskt för att öka förtroendet och användarupplevelsen hos e-handlare, då de första rekommendationerna spelar en viktig roll i att forma kundens förtroende för webbplatsen (Saxborn et al. 2024). Bodduluri et al. (2024) förklarar vikten för användarförtroende inom RS. Författarna beskriver att för att bygga förtroende hos användaren krävs det att systemen inte bara är träffsäkra utan också transparenta och förklarliga. E-handelsföretag bör sträva efter att utveckla system där användarna enkelt kan förstå varför vissa rekommendationer görs, vilket kan bidra till en ökad känsla av kontroll och acceptans av rekommenderade produkter (Bodduluri et al. 2024).

Saxborn et al. (2024) lyfter fram ett antal nyckelkomponenter som tillsammans bidrar till att skapa ett mer pålitligt och ansvarsfullt system. Att ta hänsyn till komponenterna kan bidra till ett system som respekterar både användarnas preferenser och deras integritet (Saxborn et al. 2024). Den första faktorn handlar om den upplevda rekommendationskvaliteten. Författarna förklarar att det är viktigt hur användarna uppfattar kvaliteten på det service som erbjuds genom RS. En hög upplevd kvalitet innebär nämligen att rekommendationerna känns relevanta, personliga och tillförlitliga, vilket kan stärka användarnas förtroende för e-handelsplattformen (Saxborn et al. 2024). För att öka kvaliteten hos rekommendationerna och för att ett RS ska kunna ge förslag som verkligen passar användarnas önskemål och intressen lyfter Zhang (2023) fram ett område som utvecklare inom RS bör fokusera på. Zhang (2023) påpekar att det är betydelsefullt att adressera och lösa situationer där det finns en diskrepans mellan systemets träningsdata och de faktiska målen med rekommendationerna. Med andra ord, ibland kan den information som samlats in och som systemet lär sig ifrån, inte helt och hållet stämma överens med vad användarna faktiskt är intresserade av eller vad man som utvecklare vill att systemet ska uppnå. Det kan exempelvis handla om att datan inte längre är aktuell eller att den inte täcker alla aspekter av användarnas preferenser (Zhang, 2023). Författaren föreslår att använda kausalitetsanalys för att hantera denna diskrepans. Analysen innebär att man undersöker orsakssamband - alltså att förstå vilka faktorer som leder till vilka resultat. Författaren förklarar vidare att genom integreringen av kausalitetsprinciper kan utvecklarna av RS bättre förstå hur och varför vissa data inte fullständigt återspeglar användarens verkliga preferenser och hur de kan justera systemet för att rätta till dessa skillnader. Författarna understryker att användarens tidigare erfarenheter och interaktioner med e-handelsplattformen och dess RS även påverkar deras förväntningar och förtroende. Positiva erfarenheter bygger upp förtroende över tid, medan negativa erfarenheter kan minska tilliten till framtida rekommendationer (Saxborn et al. 2024). En annan viktig faktor som författarna lyfter fram är äktheten i rekommendationerna. Äktheten hänvisar till hur genuina och opartiska rekommendationerna uppfattas av användarna. Förtroendet stärks när användarna känner att rekommendationerna verkligen syftar till att matcha deras individuella preferenser utan dolda agendor (Saxborn et al. 2024).

2.6.1 Integritetsskydd

För att ta hänsyn till integritetsskydd och respekt av personuppgifter framhåller Saxborn et al. (2024) en faktor som handlar om användarens intention att dela interaktionsdata. Denna faktor handlar om användarnas vilja att dela personlig och beteendedata med plattformen (Saxborn et al. 2024). Författarna tydliggör att genom att förstå vad som motiverar eller avskräcker användare från att dela sin data kan e-handelsföretag anpassa sina strategier för datainsamling på ett sätt som respekterar användarnas integritet och bygger förtroende.

Hur RS tillåts samla in personuppgifter har förändrats markant i takt med de nya kraven och regler som framkommit i samband med GDPR (Mohallick et al. 2018). Mohallick et al. understryker att GDPR har förstärkt användarnas rättigheter online genom att kräva tydligt och explicit samtycke från användarens sida för insamling av personuppgifter eller användardata. Författarna påpekar att detta förbättrar möjligheterna för användarna att kontrollera sin egna information samt vilken data de väljer att dela. Vidare betonas vikten av att inkludera detta samtycke inom RS som samlar in användardata. Detta eftersom att användarna högt värderar att själva kunna bestämma hur deras personliga data används (Mohallick et al. 2018). Författarna framhåller även att användarnas möjlighet att när som helst modifiera eller radera sina delade uppgifter bidrar till att öka deras upplevelser och förtroende för systemet. Detta stärker i sin tur systemet och det aktuella företagens pålitlighet (Mohallick et al. 2018).

2.7 Litteratursammanfattning

I litteraturgenomgången utforskas de centrala begreppen och fenomenen inom området för RS, särskilt inom e-handel, vilka har varit avgörande för att definiera de teoretiska grunderna för studien. Gällande utmaningar kring datainsamling belyser litteraturen särskilt de centrala problemen inom RS, kallstartsproblemet och gles data, vilka båda påverkar systemens förmåga att leverera träffsäkra rekommendationer. Vidare presenterar det teoretiska resultatet olika strategier och algoritmer att integrera inom utvecklingen av RS. Tidigare forskning belyser vikten av välja rätt algoritmer och modeller baserat på företagets bransch, datatillgång och specifika utmaningar. Tilläggsvis visar det teoretiska resultatet på betydelsen av att integrera tillförlitlighetsaspekter och integritetsskydd i systemdesignen för att förbättra användarupplevelsen och bevara användarnas förtroende.

Genomgången pekar avslutningsvis på teknologiska framsteg som maskininlärning och olika filtreringstekniker som centrala verktyg och strategier att integrera för att hantera utmaningar kopplade till datatillgång samt för att förbättra systemens prestanda och träffsäkerhet. Utifrån de teoretiska resultaten som presenteras i kapitel 2, har fyra kritiska områden som är avgörande för träffsäkerheten hos RS identifierats. De fyra kritiska områdena presenteras i Tabell 1.0: implementation, datatillgång, strategier och ansvar.

Tabell 1.0: Teoretisk referensram

Huvudområde	Sub-område	Litteratur
RS: Implementation	Historisk utveckling av RS	Falk & Lindeskog 2021; Smith & Linden 2017
	Funktion hos RS	Ullén 2015; Bodduluri et al. 2023; Portugal et al. 2018
RS: Datatillgång	Kallstart	Chowdhury 2022; Payne 2021
	Gles data	Abbas 2023; Evert & Mattisson 2016
RS: Strategier	Datainsamling	Apsis 2024; Han et al. 2012; Yan 2017
	Algoritmer	Hussein et al. 2024; Salunke & Nichite 2022; Bodduluri et al. 2024; Basiri et al. 2010; Karn et al. 2022; Chowdhury 2022; Tran 2007
	Prestandamätningar	Shani & Gunawardana 2011; Deutschman 2023
RS: Ansvar	Pålitlighet	Saxborn et al. 2024; O'Donovan & Smyth 2005; Bodduluri et al. 2023; Abbas 2023; Zhang 2023
	Integritetsskydd	Ricci et al. 2011; Saxborn et al. 2024;

3 Metod

I metodkapitlet presenteras en noggrann genomgång av den metodik som tillämpats i studien. Kapitlet inleds med en detaljerad beskrivning av de forskningsmetoder som valts, inklusive en motivering för detta val. Därefter följer en detaljerad presentation av intervjuguiden, med de noggrant utformade frågor som kommer att användas under studiens semi-strukturerade intervjuer, samt en beskrivning av urvalet av respondenter för dessa. Avsnittet avslutas med en genomgång av metodens validitet och reliabilitet, samt en granskning av de etiska övervägandena som ligger till grund för forskningsansatsen. En redogörelse för hur AI används inom framställningen av uppsatsen återfinns i Appendix (1).

3.1 Metodval

3.1.1 Kvalitativ studie

En kvalitativ metod har valts för studien då metoden anses lämplig i hänsyn till studiens syfte att bidra med strategier för implementering av RS hos e-handelsföretag med begränsad datatillgång. Genom kvalitativa metoder strävar uppsatsen efter att åstadkomma en helhetsbeskrivning av det utforska ämnet (RS) (Nationalencyklopedin, 2024). En kvalitativ metod är enligt Creswell (2014), ett lämpligt val vid en studie av mer komplexa avgränsningar eftersom att metoden möjliggör en med djupgående förståelse för olika aktörers erfarenheter och perspektiv. Alvehus (2019) framhåller att det främsta syftet med kvalitativa metoder är att visa på komplexitet och nyansrikedom. Studien ämnar att uppnå detta syfte genom en omfattande litteraturstudie, tillsammans med intervjuer med respondenter från företag vilket genererar nyanserade åsikter och tolkningar från ett flertal branscher inom e-handeln. Jacobsen (2019) beskriver att forskning med kvalitativa metoder klarlägger begrepp i specifika situationer, vilket i studiens fall kan beröra företags hantering av *gles data* och *kallstart*. Dessutom redogör Starrin och Svensson (1996) för att kvalitativ forskning förbättrar förståelsen av människor och fenomen i deras unika kontexter. Redogörelsen ligger i linje med studiens val av semistrukturerade intervjuer med respondenter från olika branscher vilka alla innehar unika perspektiv på RS.

3.1.2 Semistrukturerad intervju

Med studiens syfte i beaktelse, kan en djupare diskussion av relevanta frågeställningar vara relevant. Därför har semistrukturerade intervjuer valts som metod för genomförandet av intervjustudien. Oates et al. (2022) påpekar att semistrukturerade intervjuer tillåter en omfattande granskning som överträffar de mer begränsade möjligheterna som medföljer med en strukturerad intervju. Metoden tillåter nämligen flexibiliteten att ställa kompletterande frågor som är anpassade till varje specifik intervju eller respondent, vilket tillåter respondenten att bidra med ytterligare insikter, utan att vara bunden till förutbestämda frågor (Oates et al. 2022). Enligt Bryman (2008) innebär användningen av semistrukturerade

intervjuer att man arbetar inom definierade teman, vilket i studiens fall omfattar de fyra huvudområden som specificeras i intervjuguiden. Alvehus (2019) anser att en effektiv intervjumetod ger utrymme för respondentens insikter att utvecklas till nya frågeställningar, vilket berikar diskussionen om hur det specifika företaget tar sig an de berörda utmaningarna.

Valet av semistrukturerade intervjuer grundar sig likaså i att respondenternas individuella svar och perspektiv varierar. Skillnaderna i respondenternas branschtillhörighet och yrkesroller kan leda till diversitet i exempelvis vilka områden de kan bidra med mest insikter kring. Denna variation motiverar användningen av en intervjumetod som tillåter ett balanserat samtal och djupare diskussioner inom specifika kunskapsområden där respondenten besitter större expertis. Respondentens möjlighet att mer detaljerat belysa ämnet utifrån sina unika erfarenheter och tankar kan resultera i ett mer givande empiriskt resultat. De följdfrågor som ställs under intervjun kommer att tillsammans med respondenternas svar återfinns i Appendix (2-5).

3.1.3 Litteraturstudie

Enligt Oates et al. (2022) är en litteraturgenomgång ett relevant verktyg för att öka förståelsen kring ett visst ämnesområde. Genom att granska och analysera befintlig forskning kan forskare upptäcka kunskapsluckor och urskilja inom vilka områden som behov av ytterligare forskning finns (Oates et al. 2022). En litteraturstudie genomfördes följaktligen med syftet att skapa en djup förståelse och inblick av RS och för att identifiera ett problemområde. Problemområdet som presenteras i avsnitt 1.2 definierades genom att, utifrån Oates et al. (2022) principer, studera den tidigare forskningen som gjorts om RS och e-handel och utifrån den forskningen identifiera en kunskapslucka inom ämnet.

Insamlingen av litteratur har utförts via de vetenskapliga sökmotorerna Google Scholar samt LUBSearch. Sökresultat via LUBSearch har sedan lett till databaser som exempelvis Association for Computer Machinery (ACM). Inom framställningen av litteraturgenomgången har följande sökord använts för att hitta lämplig tidigare forskning som berör ämnet:

- Recommender systems in E-commerce
- Recommender systems Cold start
- Recommender systems Sparse data
- Collaborative filtering
- Content-based recommendation systems
- Knowledge based recommendation systems
- Hybrid recommendation systems
- Privacy in recommendation systems

3.2 Intervjuguide

Enligt Oates et al. (2022) syftar utformningen av en intervjuguide till att verka som en lista av frågor som inkluderas i intervjun. Författaren redogör för att en intervjuguide däremot inte ämnar att endast användas som en uppsättning frågor som måste ställas i en förutbestämd ordning utan att guiden istället bör beaktas som en lista på de ämnen, med tillhörande frågor, som man önskar att intervjun ska avhandla. Intervjufrågorna har således utformats för att spegla strukturen och de centrala huvudområdena som identifierats i litteraturstudien, en

koppling som tydligt framgår i Tabell 3.1. Genom denna struktur ämnar intervjuguiden att etablera förankring till litteraturgenomgången som presenteras i kapitel 2. Frågorna som intervjuguiden innehåller ställs till anställda inom företag som är lämpliga för forskningsområdet *RS inom e-handel*. Vid insamlingen av empirin används Tabell 3.1 som innefattar de intervjufrågor som ställs till samtliga respondenter. Tabellen fungerar som en detaljerad intervjuguide, där varje intervjufråga är direkt relaterad till något av huvudområdena som studien avser att utforska.

Tabell 3.1: Intervjufrågor

Frågor	Huvudområde
<ul style="list-style-type: none"> - Hur samlar ni in och analyserar användardata för att skapa relevanta och pålitliga rekommendationer till era kunder? Vilken typ av användar- och objektdata samlar systemet in? 	Implementation
<ul style="list-style-type: none"> - Vilka utmaningar har ni stött på vid implementeringen och användningen av rekommendationssystem? Har ni upplevt några utmaningar kopplade till datatillgång? - Hur hanterar ni kallstartsproblemet för nya användare eller produkter i ert RS? 	Datatillgång
<ul style="list-style-type: none"> - Vilka huvudsakliga algoritmer eller metoder använder ni för ert RS, hur har dessa påverkat systemets effektivitet och användarupplevelsen? Vilka fördelar i prestanda tror ni att valet av algoritmer har genererat? - Hur mäter ni prestandan och effektiviteten av era rekommendationer? 	Strategier
<ul style="list-style-type: none"> - Hur säkerställer ni att ert RS respekterar användarnas integritet? - Hur säkerställer ni pålitlighet hos rekommendationerna? 	Ansvar

3.3 Urval av respondenter

Inom intervjustudien har ett ändamålsenligt urval gjorts, vilket enligt Denscombe (2014) innebär att valet av respondenter baseras på specifika attribut för att få ut relevant information. I fallet av denna studie är det gemensamma attributet för de valda respondenterna att de ska ingå i ett företag som använder eller utvecklar RS samt ingå i den tekniska sektionen av företaget. Urvalet av respondenter innebär även att välja respondenter som anses kunna bidra med ett värde till studien, vilket ligger i linje med det ändamålsenliga urval som Denscombe (2014) beskriver.

Kontakt med respondenterna för planering av intervjun har skett via mail. Majoriteten av personerna kontaktades via hemsidan hos de företag som ansetts lämpliga för studiens forskningsområde. Utöver detta har även LinkedIn använts för att initialt kontakta företag där kontaktuppgifter till medarbetarna ej fanns tillgängliga på hemsidan. Intervjuerna hölls på svenska eller engelska beroende på respondentens modersmål. Samtliga svar presenteras dock på svenska i kapitel 4. Den nedanstående tabellen presenterar en sammanställning av respondenter som deltagit i intervjustudien, samt de företag de är associerade med. För att upprätthålla och skydda individens och organisationens integritet har identifierande uppgifter såsom namn på personer och företag anonymiserats. Respondenterna kommer således att refereras till som R1-R5, och företagen till Företag A-D i presentationen av empirin (Kapitel 4).

Tabell 3.2: Intervjurespondenter

Respondent	Organisation	Roll	Marknad	Plats för intervju	Appendix
R1	Företag A	Data Intelligence Director	E-handel för kläder och accessoarer	Microsoft Teams	2
R2	Företag B	Software Developer	E-handel för läkemedel	Microsoft Teams	3
R3	Företag B	Software Developer	E-handel för läkemedel	Microsoft Teams	3
R4	Företag C	Director of Operations	E-handel för elektroniska produkter	Microsoft Teams	4
R5	Företag D	Co-Founder and CEO	Distributionstjänst för artister	Microsoft Teams	5

3.4 Metod för datainsamling

Med hänsyn till respondenternas tillgänglighet och geografiska förhållanden genomförs alla intervjuer inom studien digitalt, med hjälp av kommunikationsplattformen Microsoft Teams. Detta beslut möjliggör en flexibel och tillgänglig metod för att samla in data från respondenter

oavsett deras fysiska placering, vilket underlättar en bredare deltagarbaserad forskning. För att noggrant dokumentera och bevara de talade orden i sin mest autentiska form, spelas varje intervju in med tillstånd från de intervjuade. Dessa inspelningar transkriberas till skriftlig text.

3.4.1 Transkribering

Transkribering av intervjuerna genomförs för att säkerställa en korrekt representation av empirin. Transkribering utgör överföringen av verbal kommunikation till textform (Widodo, 2014). Widodo (2014) beskriver att den snabba tillväxten av digital teknologi resulterat i att inspelningar av innehåll idag används regelbundet inom den kvalitativa forskningen, vilket förenklar processen för forskare att omvandla längre tal och intervjuer till text för att i efterhand kunna genomföra en djupgående analys och bearbetning av det inspelade innehållet. Författaren understryker att transkriberingsprocessen således är avgörande för att säkerställa en korrekt och detaljerad analys av det insamlade materialet inom studien, vilket i sin tur även bidrar till arbetets övergripande kvalitet och integritet.

Alla intervjuer genomförs och spelas in via Microsoft Teams. För att transkribera dessa inspelningar används Teams inbyggda AI-baserade transkriberingsverktyg. Utförlig beskrivning om hur AI använts för transkribering återges i Appendix (1). Efter att en intervju spelats in och transkriberats går transkripten noggrant igenom för att kunna redigera och korrigera eventuella felaktigheter eller oklarheter. Denna process innebär en renskrivning för att säkerställa att transkriptet är korrekt och tydligt. Följaktligen görs en grundlig granskning för att säkerställa att transkriberingsverktyget genererat korrekta svar. Inspelningarna av intervjuerna utförs enbart i syfte att underlätta transkriberingen och raderas efter färdigställd transkribering.

3.5 Metod för dataanalys

I syfte att analysera det material som inhämtats genom intervjuerna och litteraturgenomgången, används metoden tematisk analys. Denna metodologiska ansats möjliggör identifiering av genomgående teman eller mönster inom det insamlade datamaterialet, vilket i sin tur tillåter utvinning av djupare mening ur ett brett omfång kvalitativ information (Braun & Clarke, 2006). Den tematiska analysen verkar enligt Braun och Clarke (2006) för att förenkla sökande och analys av litteratur och intervjuerna genom att organisera datauppsättningen efter ett antal huvudområden .

Nowell et al. (2017) framhåller att en tematisk analysmetod är särskilt lämplig för intervjustudier eftersom den tillåter forskaren att systematiskt organisera och tolka deltagarnas erfarenheter och perspektiv på ett meningsfullt sätt. Genom att noggrant koda och kategorisera data baserat på återkommande teman i studien kan forskare generera insikter som är både rika och komplexa, vilket bidrar till en djupare förståelse av forskningsfrågan (Nowell et al. 2017). Den empiriska redovisning är således strukturerad efter de huvudområden som påträffats i litteraturstudien, vilka ligger till grund för intervjufrågornas struktur och presenterats i Tabell 2.0. Dessa huvudområden har identifierats genom att urskilja återkommande teman eller mönster i datamaterialet tillsammans med en analys om den forskningsfråga som studien är ämnad att besvara. I den tematiska analysen har vi identifierat och redovisat fyra centrala huvudteman som presentationen av empirin kommer att vara strukturerad utifrån, med utgångspunkt i forskningsfrågan och syftet och i enlighet med

metoden som Braun och Clarke (2006) beskriver, har vi identifierat och redovisat fyra centrala huvudteman som presentationen av empirin kommer att vara strukturerad utifrån.

För att korrekt koppla samman intervju svaren med relevanta teman utförs initial kodning baserad på huvudområdena som specificeras i Tabell 2.0. Denna kodningsprocess dokumenteras i Appendix (2-5), där varje huvudområde representeras genom en förkortning. Dessa kodningar utgör sedan grunden för strukturen i den empiriska presentationen av respondenternas svar. Förkortningarna och deras betydelser redovisas tydligt i Tabell 3.3.

Tabell 3.3: Kodning för dataanalys

Huvudområde	Kodning
Implementation	I
Datatillgång	D
Strategier	S
Ansvar	A

3.6 Validitet och reliabilitet

För att säkerställa studiens kvalitet och validitet har uppsatsen granskats löpande av handledare för kontinuerlig feedback och förslag på förbättringar. Validitet beskrivs enligt Alvehus (2019) som en bedömning om huruvida rapporten undersöker det som faktiskt är tänkt att undersöka. Studien har förhållit sig till att validitet följs genom noggrant utformade frågeställningar samt användandet av insamlingsmetoder i form av intervjuer och insiktsgivande källor som är relevanta för studiens syfte. Validiteten stärks genom användandet av tematisk analys, då det säkerställer en organiserad uppsättning av empiri och teoretiska resultat (Braun & Clarke, 2006). Vidare redogör Bryman (2008) för extern validitet som innefattar om studien kan generaliseras och appliceras på andra miljöer baserat på den information som samlats in i form av intervjuer och litteratur. Trots att studien utgår från intervjuer av 6 respondenter, presenterar dessa respondenter insikter från ett diversifierat omfång av branscher där RS används. Därav kan en viss mån av extern validitet uppnås, som hade stärkts ytterligare av att intervjuer med ett större omfång respondenter inom varje bransch för att kunna identifiera branskmässiga skillnader och likheter. För att säkerställa intervjuernas validitet är det kritiskt att garantera att respondentens svar ligger i linje med det problemområde som studien berör (Alvehus, 2019). Därför tillhandahålls samtliga respondenter en beskrivning av studiens forskningssyfte, både i god tid innan- samt i början av intervjun. Detta steg utförs även i syfte att se till att respondenten är lämplig inom ämnet och besitter kunskap om RS i någon mån, vilket stärker intervjuens validitet.

Reliabiliteten upprätthålls genom att följa en strikt forskningsdesign och metodologi. Validitet och reliabilitet av intervjun stärks genom att en pilottestning av intervjun genomförts med en utvald, oberoende respondent innan de primära intervjuerna. Testet genomförs för att kunna identifiera eventuella problem samt otydligheter och i sin tur genomföra nödvändiga justeringar av frågor innan den huvudsakliga datainsamlingen påbörjas. Gani et al. (2020)

förklarar att genom att tillämpa ett pilottest ökar sannolikheten att frågorna är tydliga och att intervjuernas resultat blir mer konsekventa och tillförlitliga.

3.7 Etik

Oates et al. (2022) presenterar en detaljerad uppsättning principer som bör ligga till grund för forskningsetiska åtgärder som ska skydda deltagarnas rättigheter. Dessa etiska principer, som har varit vägledande för studiens etiska överväganden, innefattar följande: deltagarnas frihet att välja att inte delta i studien, deltagarnas möjlighet att när som helst återkalla sitt bidrag, kravet på att erhålla informerat samtycke från alla deltagare, samt garantierna om deltagarnas anonymitet och sekretess. Detta ramverk syftar till att säkerställa att deltagarna känner sig respekterade och skyddade genom hela forskningsprocessen (Oates et al. 2022).

För att säkerställa att behandlingen av respondenternas personuppgifter ligger i linje med GDPR samt för att kunna garantera deltagarna att deras uppgifter behandlas konfidentiellt, har ett konfidentialitetsavtal upprättats vilket presenteras i Appendix (6). Avtalet skickas till deltagarna via mail i god tid inför intervjun för att säkerställa att respondenterna har haft tillräckligt med tid att formulera och framföra eventuella frågor om avtalets innehåll. Detta steg är avgörande för att garantera att deltagarna fullt förstår vad deras godkännande innebär, vilket är en central aspekt i att värna om deltagarnas rättigheter enligt Oates et al. (2022). Avtalet läses igenom i början av intervjun och respondenten ombeds sedan godkänna avtalet muntligen.

Respondenterna ombeds i samband med muntligt godkännande av avtalet, att lämna samtycke till inspelning och transkribering av intervjun. Respondenten har rätt att delta anonymt, enligt Oates et al.s (2022) principer. Vidare har anonymitet upprätthållits för samtliga respondenter och företag som deltagit i studien, vilket innebär att ingen personlig eller företagsrelaterad information offentliggörs i studien. Respondenterna refereras till som "R1-R5" i uppsatsen, och företaget till "Företag A-D" för att kunna skydda känslig företagsrelaterad information. Intervjuerna används endast i forskningssyfte och inspelningen kommer därför att raderas efter transkriberingens färdigställande för att säkerställa deltagarnas sekretess. I avtalet informeras även deltagarna om deras rätt att när som helst avbryta intervjun. Respondenten erbjuds även möjligheten att granska forskningsrapporten innan inlämning och har därmed möjlighet att dra tillbaka sina uttalanden, vilket ligger i linje med redogörelsen som Oates et al. (2022) lägger fram.

4 Empiriskt resultat

I detta avsnitt genomförs en presentation av den empiriska datan som inhämtats genom intervjuer. De huvudområden som introducerades i litteratursammanfattningen har legat till grund för struktureringen av intervjufrågorna. För en detaljerad översikt över vilka specifika frågor som hör till respektive huvudområde, hänvisas läsaren till Intervjuguiden (3.2). Dessa områden återges i detta kapitel som underrubriker, organiserade i en sekvens som speglar den ordningsföljd som etablerades i Tabell 1.0.

4.1 Implementation

4.1.1 Funktion hos RS

R1 förklarar att Företag A använder sig av RS både på sin webbsida och inom marknadsföring via e-post. Dessa typer av rekommendationer skiljer sig däremot åt. R1 förklarar att företagets webbplats inte erbjuder personligt anpassade rekommendationer. Istället fokuserar deras system på att samla in data om produkterna för att automatiskt föreslå liknande produkter till användaren. R1 klargör att om en användaren besöker en specifik produktsida, kommer systemet att erbjuda samma rekommendationer till alla kunder som tittar på den produkten. Rekommendationerna är således endast beroende av objekt och inte av användardata. Ett mer personifierat RS används inom företagets marknadsföring via e-post där R1 förklarar att RS används för att generera riktad marknadsföring baserat på användarens köphistorik och betyg.

R2 beskriver att Företag B har implementerat ett eget RS som har utvecklats över tid och initialt baserades på användardata insamlade genom Google Analytics. Systemet baseras primärt på orderdata som genereras när kunderna gör ett köp. R2 framhäver att beslutet att utveckla sitt eget RS internt, istället för att köpa ett färdigt från en extern leverantör baseras på företagets starka IT-kompetenser och tillgängliga utvecklarresurser. Genom att bygga systemet själva siktar de på att bättre kontrollera kostnader och öka kunskapen om sina data och kunder. R2 noterar att externa system ofta är dyra och att det kan vara svårt att verifiera deras påstådda ökning i försäljning. De understryker vikten av att noggrant utvärdera alla aspekter av RS prestanda och kostnadseffektivitet.

“Det man kan säga är att vi gör ganska mycket själva, vissa moduler som vi har valt är redan färdiga men vi har valt att inte gå ut på stan och köpa ett färdigt system.” - R2 (Appendix 3 #35)

R4 förklarar att Företag C erbjuder medlemskap som bland annat hjälper att samla in användardata som grund för att skapa relevanta rekommendationer till sina kunder. När kunderna går med i företagets kundklubb får företaget tillgång till viss personlig information, såsom ålder, kön, geografisk plats, samt e-post och telefonnummer. Denna information kopplas sedan till användarens köphistorik för att bygga en bättre bild av vem kunden är och vad de tenderar att köpa. Enligt R4 analyserar Företag C dagligen dessa data för att förstå kundernas köpvanor och utveckla sitt produktutbud.

“Vi analyserar alltså och har stor nytta av köphistorik kopplat till dig som medlem.” - R4 (Appendix 4 #2)

R5 beskriver användningen av ett system som, även om det inte är ett traditionellt RS, liknar sådana system genom att det använder användardata för att optimera erbjudanden och förbättringar som ökar användarnyttan och försäljning. Användardata, såsom interaktions- och köpbeteenden, är en central del av företagets arbete för att skapa relevanta erbjudanden och förbättringar.

4.1.2 Datainsamling

R1 förklarar att Företag A har behövt anpassat sin strategi för datainsamling och personanpassade rekommendationer efter att ha stött på rättsliga hinder. Enligt R1, har företaget övergett användningen av Google Analytics för att identifiera specifika kunder och har istället övergått till en mer anonymiserad form av datainsamling. Denna ändring innebär att de inte längre kan erbjuda högst personliga rekommendationer baserade på individuella användardata via företagets webbplats. Istället används denna typ av användardata huvudsakligen för målinriktad, personlig marknadsföring via e-post.

R1 redogör för ett experiment som företaget utförde där de jämfört träffsäkerheten i sitt eget RS med Googles, vilket resulterade i en strategisk omorientering mot mindre personliga, men mer kostnadseffektiva system. R1 påpekar att detta tillvägagångssätt inte bara minskar behovet av datorkraft, vilket sänker kostnaderna, utan också att det generellt inte kompromissar avkastningen på investeringarna. R1 förklarar att företaget därför idag fokuserar på att samla in data om attribut för objekten samt likheter mellan objekt inom deras RS. Detta görs bland annat genom användningen av AI-verktyg som analyserar bilder på objekten för att ta fram olika attribut som färg, mönster, produktkategori och passform. R1 förklarar vidare att det även finns data som samlas in kopplat till relationen mellan de produkter som användaren bara tittar på, och den produkt som personen i slutändan köper. Den datan kan i sin tur användas för att identifiera liknande produkter och generera rekommendationer.

“We then saw that specific product recommendations highly outperforming the personalized recommendations we had from Google.” - R1 (Appendix 2 #2)

R2 redogör att Företag B samlar in en rad olika typer av användardata för sitt RS. Denna data inkluderar visningar av produkter, klickningar, tillägg till kundvagnar och framför allt köp. Vidare samlas data även in när produkter visas i sökresultat och på rekommendationspaneler, vilka båda lagras för analys. R3 tillägger att Google Analytics används för att samla in ytterligare information såsom enhetstyper och användargrupper.

“Vi samlar in visningar av objekt, alltså när produkter visas upp, och klick, add-to-basket och framförallt köp. Sedan har vi data kopplade till sökresultat, när produkter visas i sök och rekommendationspaneler, båda dessa sparas.” - R2 (Appendix 3 #7)

R4 klargör att Företag C fokuserar på att analysera köpbeteende snarare än personlig identifiering. Företaget undersöker bland annat vilka typer av produkter kunderna köper, hur mycket de spenderar per transaktion, genomsnittligt kvitto och om de köper från flera avdelningar eller bara en. För att skapa relevanta rekommendationer använder de denna data för att identifiera köpmönster och rikta in sig på produkter som sannolikt är av intresse för kunden. R4 nämner att denna data också hjälper Företag C att ge mer informerade rekommendationer. Dessutom används den insamlade datan från alla 5,5 miljoner medlemmar

för att se större köptrender och mönster, vilket hjälper företaget att utveckla mer effektiv marknadsföring och produktsortiment.

Företag D samlar in en mängd olika användar- och objektdata, inklusive besöksdata och användarbeteende på webbplatsen, transaktionsdata (köpfrekvens, köpvoly) och konsumtionsdata, vilket i detta fall inkluderar streams av musik och geografiska lyssnartrender. R5 tydliggör att denna data används för att förstå och förbättra användarinteraktioner och effektivisera tjänsten.

“Användardata är centralt i allt vi gör och tittar mycket på användarbeteende. Hur de klickar och köper.” - R5 (Appendix 6 #2)

4.2 Datatillgång

4.2.1 Upplevda utmaningar

I takt med att Företag A utvecklats från en e-handel inom mode till ett online-varuhus, framhåller R1 att kundbasen blivit allt mer diversifierad, vilket ytterligare komplicerar möjligheterna att dra tillförlitliga slutsatser om individuella kunders preferenser. R1 framhåller att det har varit en stor utmaning att få systemet att generera inspirerande rekommendationer som leder till köp och ökar användarupplevelsen, och inte bara föreslå de mest liknande produkterna.

“It's also a lot about how you can make it inspirational instead of just showing similar items. Because It's very easy to make a recommendation that will show you very identical items.” - R1 (Appendix 2 #4)

R1 förklarar att deras RS har stött på utmaningar relaterade till gles data, vilket uppstod då de hade en stor mängd produkter i lager samtidigt som antalet aktiva användare var relativt lågt. R1 understryker att obalansen gjorde det svårt att samla tillräckligt med användardata för att skapa kvalitativa rekommendationer.

Även R2 lyfter fram flera utmaningar med deras RS. En större utmaning är att även om datainsamling kan vara enkelt, så är inte datans kvalitet alltid tillfredsställande, vilket kräver manuell eller semi-manuell justering. Dessutom har exporten av data från Google Analytics ändamål för maskininlärning varit begränsad, förklarar R3, vilket hindrat tillgången till historisk data. Företag B har nyligen slagit på en tjänst i Google Analytics som bidragit till en mer omfattande tillgång till exporterad data. R3 betonar att om de hade haft möjlighet att börja om, skulle de ha slått på denna tillgänglighet till data redan från början.

R4 berättar att en av de centrala utmaningarna som företag 3 stött på är att övertyga kunder att bli medlemmar och därmed kunna samla in användardata. Tilläggsvis är en annan utmaning att kunder inte behöver klicka på "godkänn" när de läggs till som nya medlemmar på företaget. Detta kan leda till att kunder får tillgång till vissa förmåner utan att bli medlemmar, vilket gör att företaget inte kan använda deras data för rekommendationer. Detta kan påverka både kundupplevelsen och företagets affärsstrategi, enligt R4.

En av de centrala utmaningarna som R5 lyfter fram kopplade till RS är relevansen i de automatiserade rekommendationerna. R5 illustrerar detta med ett exempel där användare kan känna sig förföljda av produkter de klickat på men inte köpt, vilket kan vara resultatet av en dålig implementering av RS. Denna typ av aggressiv marknadsföring kan skapa negativa användarupplevelser, där produkter kontinuerligt "följer efter" användaren på internet, vilket kan leda till irritation snarare än till köp.

“Har du någon gång varit inne på en webbsida och klickat på en vara men inte köpt den, men varan förföljer än i 30 dagar eller längre? För det här kan man se som en väldigt dålig implementering av rekommendationssystem” - R5 (Appendix 5 #6)

4.2.2 Hantering av kallstart

R1 framhåller att kallstartsproblemet och gles data är stora utmaningar för dem, särskilt eftersom de arbetar med säsongsbetonade produkter som finns i många varianter och ofta har korta livscykler. För att hantera utmaningarna, speciellt för nya produkter, använder Företag A olika metoder för att snabbt samla in data om nya artiklar så snart de introduceras. R1 förklarar att dessa metoder inkluderar att främja nya kollektioner och använda insikter från företagets kategoriansvariga för att justera rekommendationerna efter säsong och aktuella trender. R1 betonar att kallstart är ett komplext problem och det är svårt att hitta en heltäckande lösning för problemet.

“When we do personalization, we now have the cold start issue for new items. So whenever we add a new season for example, we need to figure out how it compares to existing products. And have we solved it? I don't know.” - R1 (Appendix 2 #9)

R3 berättar att deras Företag B nyligen har börjat hantera ett problem relaterat till kallstart inom RS, vilket beror på att beteendedata endast varit tillgänglig under en kort period, cirka en månad. Denna begränsade tillgång till data skapar utmaningar i att effektivt förutsäga konsumenternas säsongsspecifika köpbeteenden, som till exempel varierad efterfrågan på sommar- och vinterprodukter.

För att anpassa sitt RS efter säsongsspecifika köpmönster, använder Företag B en metod där två dataset integreras olika i modellträningen. Det första datasetet, som består av aktuell och direkt relevant beteendedata, viktas högre för att stärka dess inflytande i modellen. Det andra datasetet, som innehåller information från icke-klickade visningar, får en lägre viktning för att minska dess effekt på resultatet. R3 beskriver att denna strategi syftar till att finjustera systemet så att det kan ge mer träffsäkra rekommendationer och kringgå problemet med kallstart. R3 framhåller att målet är att snart implementera denna modell i den aktiva produktionen.

R4 framhåller att Företag C har haft problem med gles data och kallstart när nya produkter lanseras, eftersom det finns en tidsfördröjning innan kundernas feedback kan samlas in. Detta gör det svårt att ge rekommendationer för nya artiklar utan historik. För att lösa detta använder företaget data från liknande befintliga produkter. Om de lanserar en ny produkt, kopieras attribut från en annan produkt med redan insamlad data, vilket ger en grund för rekommendationer. På så sätt kan Företag C skapa rekommendationer för nya produkter även med begränsad data, genom att återanvända insikter från liknande befintliga produkter.

R5 beskriver specifikt utmaningarna kopplade till kallstart och gles data som innebär svårigheter i att tillhandahålla relevant innehåll eller tjänster när det inte finns tillräcklig eller ingen tidigare data tillgänglig om användarna eller objekten. Denna brist på data gör det utmanande att driva förutsägelser och skapa effektiva strategier för att engagera nya användare eller lansera nya produkter. För att hantera dessa problem betonar R5 vikten av att jobba hypotesdrivet i de tidiga skedena av en produktlansering. Det innebär att de skapar teorier om vad användarna kan tänkas vilja eller behöva, baserat på begränsad information, och sedan snabbt testat dessa hypoteser genom att samla in och analysera data från tidiga interaktioner. R5 påpekar att detta gjort det möjligt för Företag D att snabbt justera och förbättra sina erbjudanden baserat på faktiska användarreaktioner.

4.3 Strategier

4.3.1 Val av algoritmer

R1 redogör för den viktigaste delen av deras datainsamlingsstrategi som är att få systemet att få en djup förståelse för produkterna. AI-verktyget för bildanalys är en av de primära modellerna som används. Verktyget extraherar visuella egenskaper från bilderna och RS jämför sedan dessa för att hitta likheter med andra objekt. Baserat på dessa likheter presenteras sedan rekommendationer till användarna. Metoden som R1 presenterar har möjliggjort relevanta förslag utan att behöva förlita sig på personlig användardata.

“We are able to see a lot just by using AI to analyze images of a product. Then, we can add the next layer on top of that where we create product groups based on that data and that's what the recommendation models are doing, trying to match products that are similar.” - R1 (Appendix 1 #11)

R3 redogör för de modeller som Företag B använder inom sitt RS för att förbättra precisionen i produktförslagen. De primära teknikerna inkluderar machine learning-modeller för att identifiera produkter som ofta köps tillsammans, samt för att föreslå relevanta produkter baserat på vad kunden redan har i kundkorgen. Denna process kallas "retrieval", där systemet väljer ut relevanta produkter från företagets stora produktkatalog. Vidare används en rankingsmodell som anpassar produkterna som visas baserat på användarens tidigare köphistorik och andra relevanta faktorer, som till exempel månad eller märkespreferenser.

“Framförallt så använder vi machine learning för att hitta produkter som ofta köps tillsammans. Även för att hitta produkter som är relevanta för det du har i kassan.” - R3 (Appendix 3 #18)

R3 nämner även att deras system involverar tekniker som kollaborativ filtrering. Utöver detta använder Företag B neurala nätverk och embeddings för att skapa vektorrepresentationer av produkter. Detta görs för att kunna dela in produkter i grupper beroende på egenskaper och attribut, vilket underlättar att hitta likheter mellan produkterna. Denna hybridiserade teknik hjälper till att mer precist föreslå produkter som är relaterade till användarens intressen.

R4 från Företag C förklarar att företaget använder olika algoritmer för produktutveckling och personaliserade rekommendationer. R4 är uttrycker en osäkerhet om vilka specifika modeller och algoritmer som används inom företagets RS men understryker att målet med

integreringen av algoritmerna är att kunna skapa mer personliga rekommendationer till medlemmar som Företag C har en relation till, men också att kunna rekommendera innehåll till helt nya kunder, vilket kräver andra typer av algoritmer.

I fråga om algoritmer och metoder beskriver R5 inte användningen av specifika rekommendationsalgoritmer, eftersom deras system inte är ett traditionellt RS. Istället fokuserar Företag D på att analysera data genom diverse verktyg som exempelvis BigQuery och Google Analytics för att kunna samla insikter om användarbeteende som kan öka försäljning och förbättra användarupplevelsen.

4.3.2 Prestandamätningar

R1 framhåller att Företag A utvärderar prestandan på sina algoritmer och RS genom att använda en blandning av A/B-tester och split-tester. Dessa tester tillämpar olika algoritmer på olika användargrupper för att etablera en jämförelsegrund och identifiera vilken metod som är mest effektiv. R1 understryker att de viktigaste mätparametrarna som de beaktar inkluderar antalet klick, konverteringsgraden och antalet objekt som användare bläddrar igenom. Indikatorerna har hjälpt Företag A att förstå om rekommendationerna faktiskt inspirerar användarna att utforska fler produkter.

Företag B genomför även de A/B-tester för att jämföra förändringar i systemet mot en baslinje, vilket hjälper dem att identifiera om uppdateringar förbättrar eller försämrar prestandan. R2 nämner även att de använder externa jämförelser med en "black box"-modell för att bedöma sitt systems relativa träffsäkerhet.

R2 detaljerar hur Företag B vidare mäter prestandan i sitt RS. Företaget utvärderar huvudsakligen prestanda baserat på hur produkter läggs i varukorgen och köps, vilket ger dem konkret data om systemets framgång. För att ytterligare förstå engagemanget hos användaren, beskriver R2 att klickfrekvenser analyseras via Google Analytics. R2 uttrycker även en önskan om att i framtiden förbättra hur de mäter prestanda i sitt RS. De vill kunna se både om användare klickar på rekommendationer samt följa upp om dessa klick leder till faktiska köp. Förbättringen är ämnad att ge en mer exakt bild av rekommendationernas verkliga affärsinverkan.

“Sen har vi även lite framtida önskemål om att kunna mäta det ännu lite bättre. Vi kan nämligen se om någon har klickat på våra rekommendationer, men vi kan inte se om någon har klickat och sedan köpt den” - R2 (Appendix 3 #23)

Företag C mäter prestandan och träffsäkerheten av sina rekommendationer genom att spåra flera indikatorer. De följer antalet medlemmar i deras kundklubb, hur många de kan rekrytera och undersöker hur många som senare avregistrerar sig. De mäter också hur många som blir medlemmar men inte godkänner villkoren, vilket ger en indikation på användarnas engagemang och acceptans av systemet. R4 berättar att företaget även övervakar den genomsnittliga varukorgens storlek, det totala beloppet kunderna köper för, storleken på ordern samt hur ofta medlemmen handlar under ett år för att kunna generera relevanta förslag till medlemmen. Mätvärden som exempelvis att en klubbmedlem gör ett visst antal köp per år används för att bedöma hur kvalitativa rekommendationerna är och hur de påverkar kundernas beteende och företagets försäljning.

“Vi följer även hur stor genomsnittliga varukorgen är, vilket värde du köper för, antalet artiklar och hur ofta du köper under ett år.” - R4 (Appendix 4 #9)

R5 beskriver att för att mäta prestanda och kvaliteten av sina system, följer Företag D nyckeltal som visar på förändringar över tid och strävar efter att kontinuerligt förbättra dessa metriker för att maximera kundvärdet.

4.4 Ansvar

4.4.2 Pålitlighet

R1 diskuterar hur Företag A arbetar för att säkerställa pålitliga och träffsäkra rekommendationer, vilket är en kontinuerlig och komplex utmaning på grund av olika synsätt inom företaget. R1 betonar att diverse grupper har olika prioriteringar; vissa betonar modeaspekten och fokuserar på att rekommendationerna visuellt matchar användarens estetik, medan andra grupper mer inriktar sig på försäljningsmål, som att rensa lagret och sälja produkter till rätt pris vid rätt tillfälle.

Dataavdelningen hos Företag A, där R1 ingår, fokuserar framför allt på datadrivna insikter för att bedöma och säkerställa kvaliteten i rekommendationerna, undersöka hur de fungerar i praktiken och vilken inverkan de har på verksamheten i stort. R1 påpekar att det inte finns en universell definition av vad en pålitlig rekommendation innebär, vilket gör det svårt att mäta och säkerställa kvaliteten konsekvent.

“I don't even think we have a definition for what is a reliable recommendation. In that case, it's a tough game because there's no right and wrong.” - R1 (Appendix 2 #17)

R1 lyfter vidare fram användarresans komplexitet i deras RS, där tidigare köp och aktuella behov kan variera stort. R1 exemplifierar att en kund som nyligen köpt en vinterjacka är inte nödvändigtvis intresserad av att köpa en till, men kan istället behöva köpa en present, vilket kräver en helt annan typ av rekommendationer.

När det gäller Företag B och deras pålitlighet och träffsäkerhet av rekommendationer, beskriver R3 utmaningarna med att säkerställa att systemet levererar korrekta resultat. Företag B använder en kombination av A/B-tester och andra jämförande metoder för att utvärdera rekommendationens kvalitet. Dessa tester jämförs mot en baseline, och företaget använder även en "black box" för att mäta systemets prestanda mot externa standarder.

R3 förklarar att företaget ibland använder manuella kontroller för att utvärdera pålitligheten. De har även själva i teamet testat tjänsterna när de varit inloggade på sina personliga konton på hemsidan. Ett exempel är när R3 nyligen fick rekommendationer för linser trots att han varken har glasögon, synfel eller någonsin köpt linser. Denna avvikelse ledde till en utredning där företaget upptäckte en bugg i träningsystemet, vilket åtgärdades.

“Det var ett tydligt fall där man kunde se att något blivit fel, då fick man gräva lite i datan och det visade sig att träningsdatorn hade en bugg i sig och så man fick reda ut det.” -R3 (Appendix 3 #28)

R2 nämner att A/B-testerna kan ge en övergripande bild av systemets prestanda, men att de inte alltid kan avslöja detaljerade problem. Även om en rekommendation kan verka träffsäker på ytan, kan den underliggande logiken ibland behöva justeras för att säkerställa kontinuerlig förbättring.

Företag B:s strategi för att säkerställa pålitliga rekommendationer är således en blandning av automatiserade tester och manuell granskning, med ett fokus på att arbeta för att upptäcka eventuella fel som kan uppstå för att åtgärda dem. Kombinationen av dessa metoder har hjälpt företaget att leverera mer träffsäkra och pålitliga rekommendationer, samtidigt som användarnas integritet respekteras. En svårighet kring att veta hur bra systemen fungerar för kunden tas även upp, där R3 beskriver att trots att man själv tycker det ser okej ut betyder det inte nödvändigtvis att det fungerar för kunderna. En given vilja grundas i att rekommendationerna ska bidra till högre försäljning men att det är svårt att mäta detta hos kunden.

“Det är svårt med rekommendationer, det är svårt att veta om de fungerar bra och visar bra saker.”- R3 (Appendix 3 #32)

Företag C säkerställer att deras RS respekterar användarnas integritet genom att strikt följa GDPR-reglerna och ha stor respekt för kundens privatliv. Enligt företagets policy raderas kunddata om en användare inte har handlat hos dem på 12 månader. Kundens data behålls således inte längre än nödvändigt, vilket minskar risken för integritetsbrott och bidrar till att skydda kundernas personliga information. Det kan med detta inte finnas sparad data som inte är relevant för kundens nuvarande affärsbehov.

För att säkerställa pålitliga och relevanta rekommendationer fokuserar Företag D på att skapa en balanserad och väl integrerad upplevelse för användarna. R5 betonar att de strävar efter att deras rekommendationer inte ska uppfattas som påträngande utan snarare som värdefulla förslag baserade på användarens tidigare beteenden och preferenser. Genom att noggrant analysera användardata och beteenden kan R5 anpassa sina tjänster för att bättre möta kundernas behov och önskemål.

Företag D använder även feedback från användarna för att kontinuerligt justera och förbättra relevansen i sina rekommendationer. Om företagets RS visar sig vara mindre träffsäker, berättar R5 att de snabbt anpassar sina algoritmer för att förbättra både precisionen och användarnas upplevelse. Detta dynamiska tillvägagångssätt tillåter Företag D att hålla sina rekommendationer både relevanta och uppskattade, vilket minskar risken för negativa användarupplevelser och ökar den allmänna tillfredsställelsen.

4.4.1 Integritetsskydd

En av de största utmaningarna som R1 belyser är integreringen av GDPR-regleringar som påverkat hur företaget hanterar personuppgifter, vilket kräver en ny strategi för att ge rekommendationer till användare vilka tas fram utan att skrida mot GDPR. Med hänsyn till detta förklarar R1 att företaget har valt att fokusera på objektbaserade rekommendationer utan insamling av personlig användardata. R1 framhäver att Företag A även prioriterar användarnas integritet inom sina RS genom att erbjuda möjligheten att välja bort spårning vid webbplatsens första besök. Detta val stoppar all spårning och användning av data för riktad marknadsföring, vilket garanterar respekt för användarnas privatliv. För varumärken hanterar

Företag A önskemål om integritet i reklamkampanjer och ökar transparensen så att varumärken kan övervaka hur deras produkter presterar på plattformen.

Enligt R2 samlar Företag B:s RS endast in data från kunder som uttryckligen gett tillstånd för det. Detta inkluderar information från användare som accepterat cookies, men det handlar om en allmän modellering av användargruppen snarare än en individuell spårning. R2 påpekar att detta tillvägagångssätt säkerställer att personlig information inte används för rekommendationer utan användarens godkännande.

R3 tillägger att företaget har inbyggda säkerhetsåtgärder för att skydda användarnas integritet. Alla användardata är anonymiserade, vilket innebär att företaget inte kan identifiera enskilda kunder. Även om en anställd känner till ett namn eller en e-postadress, kan denna information inte användas för att spåra eller hämta specifik information om den personen. Detta resulterar i att systemet inte kan dra personliga slutsatser om enskilda kunder, vilket säkerställer ett högt integritetsskydd. Vidare understryker R3 att företagets rekommendationer inte delas externt, och att användarnas rekommendationer inte spåras över olika plattformar. Detta innebär att ingen utomstående kommer att kunna se vilka rekommendationer en specifik användare har fått, vilket ytterligare bidrar till att bevara integriteten.

“Givet att jag exempelvis vet ditt namn eller email så kan jag inte skicka in det i systemet och få ut vad du är intresserad av eller vad du har köpt förut. Isåfall hade man ju kunnat dra slutsatser, så på så vis är det anonymiserat” - R3 (Appendix 3 #26)

Företag C säkerställer sina rekommendationer exempelvis genom att följa mönster som visar att produkter med många rekommendationer tenderar att vara mer attraktiva för kunderna. Om en kund till exempel har tio glödlampor i sin historik, varav tre av dem har många rekommendationer, kan det vara lättare för nästa kund att välja en av de tre rekommenderade lamporna. Dessutom, om en kund köper en lampa, kan företaget utifrån den informationen förutse att kunden också är benägen att lägga till en glödlampa i varukorgen. Genom att analysera kunders köpmönster och beteenden kan företaget öka pålitligheten hos sina rekommendationer och förbättra träffsäkerheten, vilket i sin tur ökar chansen för att kunderna ska hitta relevanta produkter och genomföra köp.

R5 betonar att säkerställandet av användarnas integritet är av yttersta vikt för Företag D. För att upprätthålla denna integritet arbetar företaget med anonymiserad data, vilket innebär att datainsamlingen och analysen sker utan att individuella användaridentiteter exponeras eller sparas på ett sätt som kan kopplas direkt till en specifik person. R5 framhåller att detta tillvägagångssätt säkerställer att även om omfattande data samlas in för att förbättra och anpassa tjänsterna, kan inte individuella användare identifieras personligen baserat på de insamlade datamängderna.

“Vi tror att om man ska bygga en sådan typ av tjänst som vi gör handlar det mycket om förtroende som byggs i att man sätter användaren i första rummet och deras integritet.” - R5 (Appendix 5 #13)

5 Diskussion

I det här kapitlet ämnar vi att sammanställa och analysera empirin i relation till de teoretiska resultaten som framkommit genom litteraturstudien. Genom att integrera empiriska observationer med teoretisk förståelse, strävar studien efter att skapa en omfattande diskussion kring hur ett pålitligt RS kan hanteras, utvecklas och implementeras ansvarsfullt och kvalitativt inom e-handeln. Vidare fokuserar diskussionen särskilt på att undgå de utmaningar som uppstår i samband med kallstart och gles data.

5.1 Datainsamling

Litteraturen föreslår en datainsamlingsstrategi för RS som fokuserar på datakvalitet snarare än kvantitet. Genom att använda data mining-metoder, som beskrivs av Han et al. (2012), kan djupgående insikter och dold information extraheras från omfattande datamängder. Yan (2017) betonar särskilt vikten av att välja rätt data för insamling, där beteendedata ses som mest värdefulla för att skapa personaliserade rekommendationer. Forskningen föreslår att genom att fokusera på data som ger insikt om användarnas faktiska preferenser och beteenden, kan RS mer effektivt förutse och möta individuella kundbehov.

Jämförelsen av datainsamlingsstrategier mellan de intervjuade företagen visar på en diversifiering i tillvägagångssätt beroende på företagets resurser, kompetenser och regelverk. Företag A och B har båda anpassat sina strategier för att möta både interna och externa krav och möjligheter där Företag B har utvecklat ett eget RS som utnyttjar användardata och framförallt beteendedata insamlad genom Google Analytics och andra interna verktyg. Denna strategi speglar den föreslagna strategin som Yan (2017) presenterar eftersom att företagen använder data mining eller analys av komplex användardata för att förbättra rekommendationernas precision och relevans. R3 och R5 beskriver att beteendedata som systemet samlar in exempelvis innefattar klick, sökhistorik, visningar av objekt samt köp. Även Företag C beskriver en insamling av användardata där analys av kunders köphistorik, varor i varukorgen och antal köp per månad görs för att kunna ta fram relevanta rekommendationer. Företag B har valt att utveckla detta system internt för att kunna anpassa det efter deras specifika behov och för att ha full kontroll över teknologin samt kostnaderna. Detta tillvägagångssätt ger dem förmågan att djupare analysera kundbeteenden och skapa mer personliga kundupplevelser.

Även R5 beskriver att Företag D använder en mer komplex dataanalys för att extrahera värde från rådata. Forskningen från Han et al. (2012) betonar betydelsen av att extrahera nyttig kunskap från stora mängder data, en process som Företag 4 tillägnar sig åt genom att använda sofistikerade analysmetoder för att identifiera mönster som kan inte är uppenbara vid första anblicken. R5 redogör att Företag 4 fokuserar på att samla in ett brett spektrum av data som sträcker sig från användarbeteende och transaktionsdata till detaljerad konsumtionsdata. Detta stämmer överens med Yans (2017) beskrivning av hur viktigt det är att använda användardata, såsom webbläsarhistorik och transaktionshistorik, för att skraddarsy och förbättra användarupplevelsen.

Enligt Apsis (2024), är kvaliteten på data avgörande och det är viktigt att inte bara samla in data utan att också säkerställa att den är relevant och användbar. R5 och R2s metod kan i vissa

avseenden diskuteras mer kvantitativt fokuserade, vilket kan medföra risker för att samla in data som är mindre relevanta eller mindre värdefulla för att bygga träffsäkra RS. Båda företagen beskriver däremot noggranna och etablerade metoder för prestandamätningar, vilket i sin tur kan stärka kvaliteten hos den insamlade datan. Företag A, har däremot, enligt R1, riktat in sig på mindre personanpassade rekommendationer på sin webbplats och fokuserat mer på produktbaserade data snarare än användardata. Detta beror delvis på de juridiska begränsningarna som påverkat deras möjlighet att använda persondata. Detta skiljer sig från den presenterade forskningen som betonar vikten av att använda djupgående beteendedata för att skapa mer skräddarsydda användarupplevelser. Företag A tycks istället prioritera en mer generaliserad tillämpning som minskar komplexiteten och kostnaderna för datan, vilket visserligen stämmer överens med litteraturens diskussion om att fokusera på datakvalitet framför kvantitet. R1 förklarar att detta val har visat sig ge bättre kostnadseffektivitet samt genererat mer kvalitativa rekommendationer.

Forskningen pekar på att ett noggrant urval av vilken data som ska samlas in är avgörande för att utveckla ett pålitligt RS som inte bara reagerar på, utan även förutser, användarnas önskemål och intressen. Intervjustudien visar på att detta val av datainsamlingsstrategier kan variera, beroende på företagets kompetens, resurser och prioriteringar. En betydande upptäckt i det empiriska resultatet är att två av de intervjuade företagen, A och B, tidigare anlitat externa leverantörer, som exempelvis Google som säljer färdiga RS. Respondenterna uppger att företagen emellertid har övergått till att utveckla egna interna system för produktrekommendationer. Övergången motiveras dels av kostnadseffektivitet, men även för att de såg en förbättring i kvaliteten på rekommendationerna. R1 framhäver att denna strategi även möjliggör för Företag A att framställa högkvalitativa rekommendationer utan att kompromissa användarnas integritet eller bryta mot regelverk som GDPR.

5.2 Datatillgång

Litteraturen belyser kallstart och gles data som centrala utmaningar inom området för RS. Evert och Mattison (2016) förklarar att gles data kan hindra systemet från att göra en tillförlitlig analys användarpreferenser och beteenden. Även kallstart är ett känt problem inom RS, som enligt Payne (2021) kan leda till missvisande och mindre relevanta rekommendationer. Tre av fyra respondenter som deltagit i intervjun har varit bekanta med både kallstart och gles data vilket styrker att begreppen är två kritiska och förekommande problem inom RS. De intervjuade företagen redogör tillika att de arbetar aktivt för att motverka dessa problematiska fenomen, genom att tillämpa olika strategier och tekniska hjälpmedel.

Företag A:s tillvägagångssätt, som handlar om att använda AI för att analysera bilder, är en innovativ metod som hanterar kallstartsutmaningar genom att omedelbart tillhandahålla insikter baserade på visuella egenskaper snarare än användarinteraktioner. Även R4 redogör för användandet av denna teknik som visat sig vara framgångsrik för att gruppera produkter och leverera rekommendationer utan historisk användardata vilket är essentiellt för hantering av kallstartsproblemet.

R1 tillägger även att Företag A implementerat ett antal metoder för att snabbt samla in data för nya produkter, inklusive marknadsföring av nya kollektioner genom en prioritering av nya produkter i rekommendationerna, vilket är en strategi som adresserar kallstartsproblemet direkt. Företag B tar till ett liknande, innovativt tillvägagångssätt för att finjustera sitt RS och

på så sätt hantera kallstart och likaså bättre anpassa systemet till säsongsspecifika köpmönster. Genom att använda sig av en teknik där två olika dataset integreras och viktas olika under modellträningen, siktar de på att förbättra rekommendationernas relevans och precision. Modellträningen gör likt Företag A:s metod, en prioritering av nya kollektioner. Denna differentierade viktning av datatyper adresserar utmaningen med kallstart, där nya eller mindre integrerade objekt kan vara svåra att korrekt klassificera och rekommendera. R2 förklarar även att genom att prioritera beteendedata som är mest relevanta och aktuella, minskar risken att nya eller temporärt populära objekt felaktigt föreslås till fel användare.

R5 belyser specifika utmaningar och strategier i deras användning av RS som kan kopplas till forskningslitteraturen om kallstart och gles data. Liksom Chowdhury (2022) och Payne (2021) beskriver, identifierar R5 problematiken kring kallstart där nya användare eller produkter saknar historik, vilket komplicerar skapandet av relevanta rekommendationer. För att hantera problemen betonar R5 vikten av att arbeta hypotesdrivet i avsaknad av data, vilket är en strategi för att testa och iterera lösningar trots begränsad initial information. Denna ansats speglar forskningens diskussion om kallstartsproblematiken där bristen på initial data skapar en barriär för relevanta och personliga rekommendationer.

5.3 Strategier

5.3.1 Algoritmer och modeller

För att hantera problem kopplade till datatillgång, föreslår Payne (2021) att tekniker som kan generera användbara rekommendationer även utan historisk interaktionsdata bör användas, särskilt för att kringgå kallstart. Historisk användardata är fundamentalt för klassiska rekommendationsmodeller som exempelvis kollaborativ filtrering. Mot denna bakgrund redogör Bodduluri et al. (2023) för att hybridiserade RS, som kombinerar flera algoritmtyper, kan vara särskilt effektiva för att övervinna kallstartsproblemet och hantera gles data. Dessa system utnyttjar både befintlig användardata och tillgänglig objektinformation för att skapa mer omfattande och dynamiska rekommendationer.

Företag B använder, i linje med vad Bodduluri et al. (2023) förespråkar, ett hybridiserat system som använder sig av klassiska modeller som, likt kollaborativ filtrering, rekommenderar liknande produkter som de som användaren intresserat sig för, kombinerat med mer avancerade tekniker som neurala nätverk och embeddings. Bodduluri et al. (2023) framhåller betydelsen av att e-handelsföretag, likt Företag B kombinerar med avancerad och modern teknik inom RS, inte endast för att förbättra träffsäkerheten av rekommendationerna, utan även för att hantera problem kopplat till datatillgång såsom kallstart och gles data. Abbas (2023) belyser även vikten av att använda avancerade analyser som kan extrahera insikter från begränsade datamängder. Detta kan innebära att man använder sig av mer sofistikerade maskininlärning- eller AI-tekniker eller algoritmer som är speciellt utformade för att hantera gleshet. Ett AI-verktyg för bildanalys som R1 redogör för ligger således i linje med ståndpunkter från Abbas (2023) och Bodduluri et al. (2023) om att en inkorporering av avancerade algoritmer kan bidra med att motverka kallstart och gles data inom RS. Eftersom att den AI-genererade bildanalysen som R1 beskriver endast analyserar objektens visuella attribut är tekniken oberoende av historisk data och pålitliga rekommendationer av nya objekt kan således genereras utan att företagen drabbas av begränsad datatillgång om objekten.

R1 förklarar att Företag A inte använder personliga rekommendationer. Istället fokuserar företagets RS på att identifiera likheter mellan objekt. Denna strategi kan jämföras med innehållsbaserade RS eftersom den inte drabbas av kallstartsproblemet då de inte är beroende av historisk användardata, enligt Graham (2024). R1 betonar dessutom att denna metod är mer kostnadseffektiv än personliga RS och att den kräver mindre datahantering och historiskt sett genererat en högre avkastning på investeringen för företaget. Detta påstående stöds av Alamdari et al. (2020) som understryker att mer komplexa RS som kräver stora mängder användardata för analys innebär högre bearbetningsresurser och därmed ökade kostnader för företaget. Denna diskussion väcker frågor om lämpligheten av objektbaserade rekommendationer för företag med begränsad tillgång till data och tekniska resurser. Om sådana företag kan dra nytta av en mindre datakrävande metod som fortfarande uppfyller kundbehov och ökar avkastningen, kan objektbaserade system vara ett fördelaktigt alternativ. Vidare integrerar Företag A sitt objektbaserade RS med ett AI-verktyg, vilket skulle kunna diskuteras som ett hybridiserat system. Strategin leder till ytterligare överväganden om potentialen i hybridiserade system eftersom att systemet som Företag A använder sig av kombinerar fördelarna från olika tekniska modeller och kan således erbjuda en balans mellan funktionalitet och kostnadseffektivitet.

5.3.2 Prestandamätningar

I det empiriska resultatet framgår det att de intervjuade företagen använder flera olika metoder för att mäta träffsäkerheten i sina RS. Företag A och B använder A/B-tester och split-tester för att identifiera vilka algoritmer som fungerar bäst. Dessa metoder som R1 och R2 beskriver, ligger i linje med Deutschman (2023) och framhålls som ett outhärligt verktyg för att testa samt förstå algoritmers prestationer utan att påverka den verkliga användarupplevelsen negativt. R2 betonar även vikten av att följa konverteringsgraden, klickfrekvens och antalet objekt användare interagerar med, vilket direkt mäter användarens engagemang och systemets förmåga att generera intresse.

Företag C, som R4 representerar, mäter systemets prestanda genom direkt affärsimpakt såsom försäljningsdata och kundengagemang i form av klubbmedlemskap och köpfrekvens. Dessa metoder ger praktiska mått på systemets träffsäkerhet som direkt korrelerar med företagets ekonomiska resultat, vilket i sin tur även rekommenderas av Shani och Gunawardana (2011) för att förstå det verkliga värdet av RS. R3 redogör dessutom för en användning av externa jämförelser och "black box"-modeller för att bedöma relativa effektiviteten, vilket är ett tillvägagångssätt som inte direkt nämns av Shani och Gunawardana (2011) men som speglar Deutschmans (2023) betoning på att anpassa prestandamått efter specifika systemegenskaper. Metoden illustrerar betydelsen av kontextuella och målanpassade utvärderingsstrategier för RS.

Forskningen om prestandamätningar inom RS som presenteras i litteraturgenomgången ger värdefulla ramar för hur företag kan utvärdera sina system. Dessa ramar bör enligt Shani och Gunawardana (2011) fokusera på noggrannhet, täckning, och användarrespons. Detta fokus är i linje med R5s beskrivning av hur Företag D utvärderar rekommendationernas prestanda. Företaget använder feedback från användarna för att kunna justera och utveckla rekommendationernas träffsäkerhet. R5 berättar att företaget förändrar sina system dynamiskt utefter användarnas respons, vilket är något som litteraturen betonar är av största vikt för att förstå systemets affärsvärde. Även R3 förklarar att de emellanåt genomför manuella tester av företagets RS för att utvärdera vilka rekommendationer som ges till de anställdas personliga konton. Detta manuella test tyder även på ett fokus på en noggrann och praktisk

prestandamätning vilket överensstämmer med testen som Shani och Gunawardana (2011) föreslår.

5.4 Ansvar

5.4.1 Pålitlighet

R1 från Företag B påpekar att det är svårt att definiera vad som utgör en pålitlig rekommendation. Följaktligen kompliceras mätningen av rekommendationens prestanda, eftersom det som anses pålitligt för en användare kan vara mindre relevant för en annan användare. Denna utmaning berörs av O'Donovan och Smyth (2005), som belyser vikten av att ta hänsyn till rekommendationens tillförlitlighet inom utvecklingen av RS. De framhåller att utan detta fokus kan rekommendationerna bli påträngande eller missvisande för användarna. R1 hos Företag A nämner också att användares intressen och behov kan förändras snabbt, vilket gör det svårt att förutsäga vad som kommer att vara relevant för dem. Denna komplexitet betonas ytterligare genom O'Donovan och Smyths (2005) som redogör för att om RS inte är tillräckligt flexibla och pålitliga, riskerar de att bli påträngande och missvisande. R5 berättar att denna påträngande effekt som RS riskerar att inneha, är något de jobbar mycket aktivt för att motverka. R5 förklarar vidare att om rekommendationerna inte är trovärdiga eller pålitliga, riskerar systemet att få en omvänd effekt där användaren tänker att företaget försöker sälja dem något de inte vill ha, vilket upplevs irriterande och påträngande. Detta perspektiv framhäver således behovet av att anpassa rekommendationer baserat på användarens preferenser och föränderliga intressen.

Företag B:s strategi för att förbättra pålitligheten innefattar även att själv logga in som användare och kontrollera vad som initialt rekommenderas, vilket kan ge insikter om potentiella fel. Denna metod är i linje med vad Saxborn et al. (2024) rekommenderar, där de betonar att tidigt erbjuda träffsäkra rekommendationer för att bygga förtroende hos användare. De första rekommendationerna spelar en nyckelroll i att forma användarnas upplevelse av RS, så det är kritiskt att dessa är relevanta och rättvisande. Genom att R3 hos Företag B loggade in och utforskade de första rekommendationerna från ett kundperspektiv kunde han få insikten om huruvida rekommendationerna faktiskt matchade hans egna preferenser. Bevisligen fungerade testet då R3 fick en missvisande rekommendation vilket resulterade i en korrigerig av en bugg i testdatan som lösning.

Zhang (2023) tar upp avvikelser mellan träningsdata och de faktiska målen med rekommendationerna, vilket kan orsaka felaktiga eller missvisande förslag. R1:s uttalande reflekterar denna problematik. R5 förklarar att om det inte finns en klar definition av pålitlighet, kan systemet basera sig på data som inte nödvändigtvis korrelerar med användarnas faktiska preferenser, vilket skapar en avvikelse som minskar systemets träffsäkerhet. Även R3 på Företag B belyser detta genom att uttrycka svårigheten kring att veta om systemet fungerar bra då man själv kan anse att systemet ser okej ut men att det inte nödvändigtvis behöver fungera för kunderna. Att säkerställa pålitlighet och trovärdighet i rekommendationerna är en komplex utmaning och både R1 och R3 framhåller det som nästintill omöjligt att säkerställa att man lyckas med ett fullkomligt pålitligt RS. Gemensamt för samtliga intervjuade företag är att de alla anser sig arbeta aktivt för att mäta prestanda,

integrera ny teknik och konstant evaluera rekommendationerna för att arbeta mot mer pålitliga RS, vilket ligger i linje med vad det teoretiska resultatet föreslår.

5.4.2 Integritetsskydd

Bodduluri et al. (2023) föreslår att för att öka förtroendet hos användaren bör RS vara transparenta och förklarliga. Företag B strävar efter detta genom att se till att rekommendationerna baseras på data som är anonymiserade och att systemet inte spårar individuella användares personliga information, vilket skyddar deras integritet. Perspektivet stöds även av Ricci et al. (2011), som betonar att RS måste bevara integriteten genom att begränsa insamlingen av personlig information och förhindra att den delas med utomstående. Alla företag som intervjuats arbetar hårt för att säkerställa användarens integritet inom deras RS vilket tyder på en integritetsbevarande betydelse för implementering av dessa typer av system.

Saxborn et al. (2024) betonar vikten av att förstå användarnas intention att dela interaktionsdata för att anpassa datainsamlingstrategier på ett sätt som respekterar användarnas integritet och bygger förtroende. Företag A demonstrerar sitt engagemang för användarnas integritet genom att erbjuda möjligheten att avstå från spårning, vilket möjliggör för användare att kontrollera delningen av beteendedata med plattformen. Enligt R1 har de nya reglerna introducerade genom GDPR lett till omfattande förändringar inom företagets RS. Numera använder företaget uteslutande objektbaserade RS på sin hemsida, vilket garanterar att användarnas personliga integritet bevaras. R1 redogör att strategin även förenklar arbetet för dem, eftersom att det minskar det omfattande ansvar som tidigare var associerat med integritetsskydd inom deras RS. Företag B säkerställer integritetsskydd genom att på ett liknande sätt endast samla in data från användare som uttryckligen gett tillstånd för det. Deras anonymiserade datahantering och inbyggda säkerhetsåtgärder förhindrar att personlig information används för rekommendationer utan användarens godkännande, vilket även stämmer överens med den ansvarsfulla implementeringen av RS som understryks av Ricci et al. (2011).

R5 tar även upp vikten av integritet, vilket harmonierar med Ricci et al. (2011) som betonar den stora betydelsen av att fokusera på integritetsbevarande inom RS. Genom att använda anonymiserade data och säkerställa att all datahantering följer stränga säkerhetsprotokoll, strävar Företag D efter att skydda användarnas personuppgifter samtidigt som de genererar relevanta rekommendationer som ej anses påträngande. Strategin svarar mot det teoretiska resultatets uppmaning till ansvarsfull hantering av användardata för att undvika att systemen blir uppfattade som inskränkande eller påträngande.

Det teoretiska resultatet visar på stora förändringar inom hur RS tillåts samla in användardata i och med GDPR (Mohallick et al. 2018). I frågan om företagets syn på dataskydd och integritet inom RS framgår det att samtliga respondenter konstaterar att regulationerna krävt att de inom implementeringen av RS prioriterar användarnas integritet högt. Samtliga respondenter nämner även att användaren måste ge samtycke till insamling av användardata via företagets hemsida för att systemet ska tillåtas analysera beteendedata och historik, vilket även understryks av Mohallick et al. (2018) som av stor betydelse för användarens förtroende för företagets RS.

6 Slutsats

Detta kapitel kommer att presentera studiens resultat från ett brett perspektiv för att besvara uppsatsens centrala forskningsfråga. Dessutom kommer slutsatsavsnittet att utforska ytterligare forskningsmöjligheter och föreslå riktningar för framtida studier inom ämnet.

Hur kan e-handelsföretag implementera och använda rekommendationssystem för att hantera utmaningarna som följer med begränsad tillgång till data?

Det empiriska resultatet, tillsammans med tidigare forskning, föreslår att e-handelsföretag med begränsad datatillgång effektivt kan tillämpa RS genom framförallt fyra tillvägagångssätt: genom att implementera avancerade dataanalysmetoder, fokusera på kvalitativ datainsamling, integrering av innovativa teknologier samt genom användningen av hybrida RS.

Implementera avancerad dataanalys: Studiens resultat visar att användningen av avancerade dataanalysmetoder kan hjälpa till att extrahera djupgående insikter från befintliga datamängder. Genom att analysera och tolka användardata kan företag identifiera dolda mönster och preferenser som inte är uppenbara vid första anblicken av datan. Dessa insikter gör det möjligt för RS att på ett kvalitativt sätt förutse och tillgodose individuella kundbehov även när datamängden är begränsad. Detta stöds av forskningen, som betonar användningen av avancerad dataanalys och data mining för att förstå djupgående användarbeteenden och preferenser. De empiriska fallen från företagen B, C och D speglar även denna strategi genom att de använder avancerade dataanalysmetoder för att skapa personliga och relevanta rekommendationer. Företagen samlar in beteende- och transaktionsdata om användare för att extrahera värdefull information från rådata, vilket har visat sig vara betydande för att generera träffsäkra RS. Företag A, B och D använder sig även av teknologier som Google Analytics och interna analysverktyg för att djupdyka i användar- och objektdata som även speglar den teoretiska rekommendationen om att utvinna värdefulla insikter ur insamlad data.

Fokus på kvalitativ datainsamling: För att maximera nyttan av tillgänglig data tyder studiens resultat på att e-handelsföretag bör välja rätt typ av data att samla in och att det är av stor vikt att prioritera datans kvalitet framför kvantitet. Beteendedata såsom användarnas interaktioner på webbplatsen, deras transaktionshistorik och sökhistorik är särskilt värdefulla för att kunna leverera personliga rekommendationer av kvalitet. R5 beskriver att det är grundläggande inom RS att fokusera på kvalitet och träffsäkerhet för att användaren inte ska känna sig påträngd av rekommendationerna. Det är därför av vikt att utvärdera vilken data som är lämplig att samla in för det specifika företaget. Resultatet tyder däremot på att insamling av användardata inte alltid är nödvändig eller lönsam. Kvalitativ data som genererar träffsäkra rekommendationer kan nämligen extraheras från endast objektdata och inte användardata, vilket är en metod som implementerats av Företag A. Objektbaserade RS är ett mer kostnadseffektivt sätt att hantera kallstart och erbjuda kvalitativa rekommendationer vilket kan vara användbart för e-handelsföretag med mer begränsade resurser.

Integrering av innovativ teknologi för att hantera kallstart och gles data: För att övervinna specifika utmaningar som kallstart och gles data, tyder det teoretiska och empiriska resultatet på att RS är i behov av integrerade modeller och tekniker som ej baseras på historisk data. Det empiriska resultatet tyder på att problemen kring begränsad datatillgång kan hanteras genom implementeringen av innovativa lösningar som exempelvis AI-driven

bildanalys, vilket R1 beskriver, för att omedelbart generera insikter baserade på ett objekts visuella egenskaper. R2 och R3 beskriver även användningen av neurala nätverk och embeddings som skapar grupperingar av objekt baserade på dess likheter, vilket är en strategi som inte heller berörs av kallstart. De avancerade teknikerna kan nämligen tillämpas för att effektivt kategorisera och rekommendera nya produkter baserat på deras likhet med befintliga objekt, utan att förlita sig på historiska användardata. Strategin lämpar sig följaktligen väl för företag med begränsad tillgång till historisk data.

Hybrida RS: Enligt det teoretiska resultatet kan hybridmodeller som kombinerar flera algoritmtyper, såsom innehållsbaserade RS och kollaborativ filtrering, vara särskilt effektiva för att hantera kallstart och gles data. Utöver det kan dessa modeller även verka för att begränsa liknande problem som följer med begränsad datatillgång. Hybrida system kan dra nytta av både befintlig användardata och objektinformation för att skapa mer omfattande och dynamiska rekommendationer. Genom att integrera olika datakällor och algoritmer med tekniker som neurala nätverk likt strategin som Företag B använder, kan e-handelsföretag utveckla RS som är mer flexibla och kapabla att hantera de variationer samt brister som finns i datamängderna. Det empiriska resultatet tyder även på att implementering av en viss typ av hybridmodell inom RS kan genomföras utan att medföra stora kostnader. R1 redogör för ett system som använder AI-verktyg för bildanalys tillsammans med ett innehållsbaserat RS som endast samlar in objektdata. Denna strategi innefattar således ett kostnadseffektivt kombinerat system som inte bara hjälpt Företag A hantera utmaningar som kallstart och gles data, utan även underlätta hanteringen av personuppgifter som kompliceras markant i och med GDPR.

Det empiriska resultatet visar på att genom att implementera avancerade dataanalysmetoder, fokusera på kvalitativ datainsamling, integrera innovativa teknologier eller använda hybrida RS, är möjligt att effektivt leverera kvalitativa rekommendationer. Tilläggsvis har de intervjuade företagen även visat på tillgodoseende av individuella kundbehov trots begränsade datamängder eller resurser. De fyra presenterade tillvägagångssätten erbjuder inte bara lösningar på vanliga problem inom RS som kallstart och gles data, de bidrar även till att skapa personaliserade och relevanta kundupplevelser. Dessutom pekar studiens resultat på framtida forskningsmöjligheter inom utvecklingen och integreringen av nya modeller och moderna tekniker för att ytterligare förbättra systemens kvalitet, precision och kostnadseffektivitet i e-handelssammanhang.

6.1 Vidare forskningsmöjligheter

Under studiens genomförande identifierades, som tidigare diskuterats i problemområdet, en brist i forskningen angående hur e-handelsföretag med begränsad datatillgång hanterar kallstart och gles data inom implementering och användning av RS. De två problemen är avgörande eftersom de påverkar systemets förmåga att ge korrekta och relevanta produktförslag. För att djupare förstå dessa utmaningar valdes en kvalitativ metod, vilket har tillåtit en ingående förståelse för hur olika e-handelsföretag strategiskt anpassar sina teknologier för att adressera dessa problem. Valet av metod har således erbjudit värdefulla insikter om olika företags tillvägagångssätt. För framtida forskning hade det däremot varit givande att utforska hur dessa anpassningar istället påverkar slutanvändarnas upplevelse och nöjdhet, vilket skulle kunna bidra med en ytterligare dimension inom ämnet och en djupare förståelse av systemets användarvänlighet i praktiken.

Med tanke på att teknologier och datahanteringsmetoder kontinuerligt utvecklas vore det även relevant att genomföra longitudinella studier för att följa hur e-handelsföretag anpassar sina RS över tid. Tilläggsvis kan detta vara relevant i och med att det empiriska och teoretiska resultatet visat att många företag framgångsrikt integrerat den absolut senaste tekniken inom RS. En sådan forskning skulle kunna generera en djupare förståelse för hur företagen hanterar förändrade förutsättningar inom data. Exempel på detta kan vara införandet av nya dataskyddsregler som GDPR, eller tekniska innovationer inom AI och maskininlärning, som enligt studiens resultat har haft stor påverkan på implementeringen av RS.

Slutligen identifierar studiens resultat en variation inom hur olika e-handelsföretag inom flera olika branscher hanterar och upplever problem kopplat till datatillgång. Ytterligare forskning skulle följaktligen kunna undersöka specifika branschskillnader eller jämföra skillnader och likheter mellan olika geografiska marknader. En jämförande forskning skulle kunna avslöja vilka strategier som är mest effektiva för olika kontexter och därmed bidra till en mer anpassad tillämpning av RS i branshmässig eller global skala.

Appendix 1

AI-redogörelse för examensarbete

I framställandet av denna studie har vi använt oss av AI-baserade verktyg för att effektivisera arbetets gång och förbättra kvaliteten på uppsatsen. Dessa verktyg har spelat en roll i förberedelsefasen, research, idégenerering och redigering av vårt arbete. Här följer en detaljerad redogörelse för användningen av dessa verktyg.

Verktyg:

Microsoft Teams AI-integrerad transkribering: Detta verktyg har använts för att automatiskt transkribera intervjuer som vi genomfört under forskningsprocessen.

ChatGPT: Verktuget har använts för att föreslå ändringar eller rättningar i vår egna text, samt i forskningsfasen för att sammanfatta forskningstexter och artiklar.

Grad av användning:

Microsoft Teams AI-integrerade transkribering: Vi har använt Microsoft Teams inbyggda AI-verktyg för att transkribera våra intervjuer. Transkriberingen skedde automatiskt under intervjuerna, vilket sparade oss betydande tid och ansträngning som annars skulle ha lagts på manuell transkribering.

Efter transkriberingen har vi manuellt granskat och justerat texten för att säkerställa noggrannhet. Detta inkluderade att korrigera eventuella fel som AI-verktyget genererat på grund av exempelvis bakgrundsljud eller oklara uttal. Transkriberingarna har använts som grund för dataanalysen i kapitel 4 (Empiriskt resultat) där citat och information från intervjuerna presenteras.

ChatGPT: ChatGPT har använts för att förbättra språket i vårt examensarbete. Detta inkluderar omformulering av specifika meningar för att öka textens tydlighet och flyt eller för att få förslag på grammatiska förbättringar och stilistiska justeringar.

Vi har även använt oss av ChatGPT för att snabbt sammanfatta stora mängder forskningstexter, vilket hjälpt oss att effektivisera och öka kvaliteten i vår litteraturgenomgång. Detta har varit speciellt värdefullt i kapitel 2 (Litteraturgenomgång), där vi har kunnat integrera insikter från en bred källbas på ett effektivt sätt. Detta har även varit kritiskt för identifieringen av kunskapsluckor inom problemformuleringen då många större forskningsarbeten bearbetades.

Appendix 6

Konfidentialitetsavtal för Kandidatuppsats inom Informatik

Detta konfidentialitetsavtal ("Avtalet") ingås mellan intervjuare ("Intervjuare") och respondenten ("Respondent").

Insamling och Användning av Information:

Respondenten samtycker till inspelning och transkribering av intervjun. Informationen kommer endast att användas för akademiska ändamål relaterade till kandidatuppsatsen.

Konfidentialitet:

Intervjuarna åtar sig att inte agera på eller sprida insamlad information. Alla uppgifter kommer att behandlas konfidentiellt i enlighet med GDPR.

Anonymitet:

Om respondenten väljer anonymitet, kommer ingen personlig eller företagsrelaterad information att offentliggöras i uppsatsen. Respondenten kommer att refereras till som "Respondent 1-5" och företaget till "Företag A-X"

Avbrytande och Granskning:

Respondenten har rätt att när som helst avbryta intervjun. Respondenten erbjuds även möjligheten att granska forskningsrapporten innan inlämning.

Radering av Inspelning:

Inspelningen kommer att raderas efter transkriberingens färdigställande.

Avtalet upprättas muntligt under intervjun

Vänliga hälsningar,
Frida Hansén & Hilda Stenström

Referenslista

- Abbas, A. (2023). First time? Recommender Systems and the ‘Cold Start’ Problem. techopedia.com. <https://www.techopedia.com/first-time-recommender-systems-and-the-cold-start-problem> [Accessed 5 April. 2024].
- Alamdari, P.M., Navimipour, N.J., Hosseinzadeh, M., Safaei, A.A. and Darwesh, A. (2020). A Systematic Study on the Recommender Systems in the E-Commerce. IEEE Access, 8, pp.115694–115716. <https://doi.org/10.1109/access.2020.3002803>
- Alvehus, J. (2019). Skriva uppsats med kvalitativ metod: en handbok, *2a upplagan*, Liber AB: Stockholm
- Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B. and Malthouse, E. (2021). User-centered Evaluation of Popularity Bias in Recommender Systems, *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (s. 119-129). <https://doi.org/10.1145/3450613.3456821>
- Apsis.se (2024). Stärk ditt produkt rekommendationssystem. APSIS. <https://apsis.se/blog/hur-du-forbattar-ditt-produkt-rekommendationssystem> [Accessed 10 April 2024]
- Basiri, J., Shakery, A., Moshiri B. and Zi Hayat, M. (2010). Alleviating the cold-start problem of recommender systems using a new hybrid approach, *5th International Symposium on Telecommunications*, pp. 962-967, http://individual.utoronto.ca/zihayatm/Papers/IST_2.pdf
- Beattie, L., Taber, D. and Cramer, H., (2022). Challenges in Translating Research to Practice for Evaluating Fairness and Bias in Recommendation Systems, *Proceedings of the 16th ACM Conference on Recommender Systems*, pp. 528-530. <https://dl-acm-org.ludwig.lub.lu.se/doi/10.1145/3523227.3547403>
- Bentzer, C. and Thulin, H. (2023). Recommender Systems Using Limited Dataset Sizes. <https://kth.diva-portal.org/smash/get/diva2:1779222/FULLTEXT01.pdf> [Accessed 1 April 2024]
- Bodduluri, K.C., Palma, F., Jusufi, I., Kurti, A. and Löwenadler, H., (2024). Exploring the Landscape of Hybrid Recommendation Systems in E-commerce: A Systematic Literature Review. IEEE Access, pp.1–1. <https://bth.diva-portal.org/smash/get/diva2:1843233/FULLTEXT01.pdf>

- Bryman (2008). Samhällsvetenskapliga metoder, *2a upplagan*, Liber AB: Malmö
- Bryman, A. and Bell, E. (2017). Företagsekonomiska forskningsmetoder, 3:e upplagan, Stockholm: Liber AB
- Chowdhury, S. (n.d.). Evaluating Cold-Start in Recommendation Systems Using a Hybrid Model Based on Factorization Machines and SBERT Embeddings. <https://kth.diva-portal.org/smash/get/diva2:1710395/FULLTEXT01.pdf> [Accessed 1 April. 2024].
- Creswell, J. W. (2014). Forskningsdesign: Kvalitativa, kvantitativa och blandade metoder. Thousand Oaks: Sage Publications, Inc.
- Denscombe, M. (2014). The Good Research Guide: For Small-Scale Social Research Projects. Fifth edition. New York: Open University Press
- Deutschman, Z. (2022). Recommender Systems: Machine Learning Metrics and Business Metrics. neptune.ai. <https://neptune.ai/blog/recommender-systems-metrics>. [Accessed 1 April. 2024]
- Evert, A.-K. and Mattisson, A. (2016). <http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A927885>. KTH Publications. <https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A927885&dswid=-1258>
- Falk, A. and Lindeskog, J.S. (2021). Considerations when Constructing a Food Recommender System with Sparse Data. LU-CS-EX. <http://lup.lub.lu.se/student-papers/record/9067009> [Accessed 7 April 2024]
- Gani, N., Rathakrishnan, M., Krishnasamy, H. A Pilot Test For Establishing Validity and Reliability of Qualitative Interview in The Blended Learning English Proficiency Course. https://www.researchgate.net/profile/Mohan-Rathakrishnan/publication/342864562_Journal_of_critical_reviews_Journal_of_Critical_Reviews_A_PILOT_TEST_FOR_ESTABLISHING_VALIDITY_AND_RELIABILITY_OF_QUALITATIVE_INTERVIEW_IN_THE_BLENDED_LEARNING_ENGLISH_PROFICIENCY_COURSE/links/5f0946cc45851550509c732a/Journal-of-critical-reviews-Journal-of-Critical-Reviews-A-PILOT-TEST-FOR-ESTABLISHING-VALIDITY-AND-RELIABILITY-OF-QUALITATIVE-INTERVIEW-IN-THE-BLENDED-LEARNING-ENGLISH-PROFICIENCY-COURSE.pdf [Accessed 15 April 2024]
- Graham, S (2024). Skapa ett innehållsbaserat rekommendationssystem - Azure Architecture Center. learn.microsoft.com. <https://learn.microsoft.com/sv-se/azure/architecture/solution-ideas/articles/build-content-based-recommendation-system-using-recommender> [Accessed 7 April 2024]
- Han, J.H., Kamber, M. and Pe, J. (2012). Data Mining: Concepts and Techniques | ScienceDirect. <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques> [Accessed 15 April 2024]

- Holzer, T., (2022). E-handeln ökade under pandemin. scb.se.
https://www.scb.se/hitta-statistik/temaomraden/sveriges-ekonomi/fordjupningsartiklar/Sveriges_ekonomi/e-handeln-okade-under-pandemin/ [Accessed 25 March 2024].
- Hussien, F.T.A., Rahma, A. and Wahab, H. (2021). Recommendation Systems For E-commerce Systems An Overview. *Journal of Physics: Conference Series*.
<https://iopscience.iop.org/article/10.1088/1742-6596/1897/1/012024/pdf> [Accessed 1 April. 2024]
- Jacobsen, D.I. (2002). Vad, hur och varför? Om metodval i företagsekonomi och andra samhällsvetenskapliga ämnen, Lund: Studentlitteratur
- Karn, A.L., Karna, R.K., Kondamudi, B.R., Bagale, G., Pustokhin, D.A., Pustokhina, I.V. and Sengan, S. (2022). Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. *Electronic Commerce Research*, 23. <https://doi.org/10.1007/s10660-022-09630-z> [Accessed 25 March 2024]
- Le, T. (2022). An Investigation into Customer Satisfaction of Customer Service Quality of an E-commerce Platform, *Proceedings of the 2022 3rd International Conference on Internet and E-Business* pp. 7–13.
<https://dl-acm-org.ludwig.lub.lu.se/doi/pdf/10.1145/3545897.3545899> [Accessed 1 May 2024]
- Li, S. and Karahanna, E. (2015). Online Recommendation Systems in a B2C E-Commerce Context: A Review and Future Directions. *Journal of the Association for Information Systems*, 16(2), pp. 72–107. <https://doi.org/10.17705/1jais.00389>
- MacKenzie, I., Meyer, C. and Noble, S. (2013). How retailers can keep up with consumers. McKinsey & Company.
<https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers> [Accessed 1 April. 2024]
- Merriam, S. B., and Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass Inc.
- Mohallick, I., De Moor, K., Özgöbek, Ö. and Gulla, J.A. (2018). Towards New Privacy Regulations in Europe: Users' Privacy Perception in Recommender Systems. *Security, Privacy, and Anonymity in Computation, Communication, and Storage*, 11342, pp. 319–330. https://doi.org/10.1007/978-3-030-05345-1_27
- Nationalencyklopedin. (2022). kvalitativ metod.
<https://www.ne.se/uppslagsverk/encyklopedi/l%C3%A5ng/kvalitativ-metod> [Accessed 20 April. 2024]
- Nationalencyklopedin. (2024). e-handel.
<http://www.ne.se/uppslagsverk/encyklopedi/lång/e-handel> [Accessed 1 April. 2024]
- Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative*

- methods. vol. 21, no. 3.
<https://journals.sagepub.com/doi/abs/10.1177/1609406917733847>
- Nugrahanto, H.R., Fadillah, T.H., Rizkyna, P.C. and Zhafira, W.A. (2023). E-Commerce: A New Media that Creates New Disasters,
<https://dl-acm-org.ludwig.lub.lu.se/doi/pdf/10.1145/3512576.3512599> [Accessed 7 April. 2024]
- Oates, B.J. Griffiths, M. and McLean, R. (2022) Researching information systems and computing. Second edition. London: SAGE.
- O'Donovan, J. and Smyth, B. (2005). Trust in recommender systems. Proceedings of the 10th international conference on Intelligent user interfaces - IUI '05.
<https://doi.org/10.1145/1040830.1040870> [Accessed 3 April. 2024]
- Payne, M. (2021). Recommender Systems For Business - A Gentle Introduction | Width.ai.
<https://www.width.ai/post/recommender-systems-recommendation-systems>. [Accessed 20 March. 2024].
- Portugal, I. Alencar, P. Covan, D. (2018) The use of machine learning algorithms in recommender systems: A systematic review. Expert Systems with Applications.
<https://doi.org/10.1016/j.eswa.2017.12.020>
- Ricci, F., Rokach, L. and Shapira, B. eds., (2022). Recommender Systems Handbook. New York, NY: Springer US. <https://doi.org/10.1007/978-1-0716-2197-4>
- Salunke, T. and Nichite, U. (2022). Recommender Systems in E-commerce.
<https://doi.org/10.13140/RG.2.2.10194.43202> [Accessed 10 April. 2024]
- Saxborn, M., Pan, Y. and Said, A. (2024). Trust Through Recommendation in E-commerce.
<https://doi.org/10.1145/3627508.3638294> [Accessed 7 April. 2024]
- Shani, G. and Gunawardana, A. (2010). Evaluating Recommendation Systems. Recommender Systems Handbook, pp.257–297. https://doi.org/10.1007/978-0-387-85820-3_8
- Smith, B and Linden, G. (2017). Two Decades of Recommender Systems at Amazon.com. *IEEE Internet Computing*, vol. 21, no. 3, pp. 12-18.
https://ieeexplore.ieee.org/abstract/document/7927889?casa_token=5uUKbchOVxkA AAAA:xbhMmuCR4G9tfJ1BcstbR-0usqGTHFakYRNjNw4xbVPRo7Y1FwhUGXPY XKodFRzKWCYugOrN3Kc
- Starrin, Bengt and Svensson, Per-Gunnar (red.), 1996, Kvalitativ metod och vetenskapsteori. Lund: Studentlitteratur.
- Sun, X., Ma, H., Song, L. and Yang, B. (2021). International reference and Enlightenment of value-added tax system for B2C cross-border service transactions under Digital Economy, *Proceedings of the 2021 9th International Conference on Information Technology: IoT and Smart City* pp .123–128.
<https://dl-acm-org.ludwig.lub.lu.se/doi/pdf/10.1145/3512576.3512599>

- Tran, T. (2007). Combining Collaborative Filtering and Knowledge-Based Approaches for Better Recommendation Systems. <https://www.site.uottawa.ca/~nelkadri/CSI5389/Papers/34-Combining%20Collaborative%20Filtering%20and%20Knowledge-Based%20Approaches.pdf> [Accessed 10 April 2024]
- Ullén, J. (2015). Recommender System Validation Platform. LU-CS-EX 2015-25. <https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=7445203&fileOId=7445323> [Accessed 25 March 2024]
- Widodo, H.P. (2014). Methodological Considerations in Interview Data Transcriptions. *International Journal of Innovation in English Language*, 3(1), pp.101–107, 111. https://d1wqtxts1xzle7.cloudfront.net/34269037/Interview_Data_Transcription_2014-1ibre.pdf?1406074245=&response-content-disposition=inline%3B+filename%3DInterview_Data_Transcription.pdf&Expires=1715173493&Signature=CoXECuU4t6Ni~s8qG~xq05x~vZSYMhdSEnE8Ty3DlZqOTkq0DG5-k0zZfDjJOWJ~I9sUF-1J2IySz5uL-oFeUioB5O00UfrIUdZdy5uAZ1bHM-YrFyYIicKHfzKhnCTKaK82pExcBfxSNyVqsvuHyz2a2qn4u46RZ7PK56lPgRoEJXlbcHfoxw2Kwn1VQrcCl3W52T4CVTJXdArkV0GuXDcBKvrD5~B6MRucqO6GBlg8H-agWsRt5vE1GjYbFsl~w4-a0OQ4eaGMp5cvKhti3APOrCkf3miVo4irLK3qLjVvy7gatRFA13tVFpSsKpWe2iOZlufAzzibh-VvJQ6Q_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- Yan, L. (2017). Personalized Recommendation Method for E-Commerce Platform Based on Data Mining Technology. 2017 International Conference on Smart Grid and Electrical Automation (ICSGEA). <https://doi.org/10.1109/icsgea.2017.62> [Accessed 15 April 2024]
- Yltävä, L. (2024). E-commerce revenue in Sweden in 2023, by product category, *Statista*, <https://www.statista.com/statistics/684157/e-commerce-turnover-in-sweden-by-product-category/> [Accessed 30 March 2024]
- Zhang, Y. (2023). Towards Trustworthy Recommender System: A Faithful and Responsible Recommendation Perspective. *International ACM SIGIR Conference on Research and Development in Information Retrieval*. <https://doi.org/10.1145/3539618.3591798> [Accessed 7 April 2024]