


AI:s roll i att främja hållbara konsumtionsval

– En studie av digital detaljhandel



Författare:
Tomas Müllern
Ulf Johansson

Forskningsrapport 2025:2



Forskningsrapport 2025:2
*"AI:s roll i att främja hållbara konsumtionsval
– En studie av digital detaljhandel"*
ingår i Handelsrådets rapportserie.
Rapporten är finansierad av Handelsrådet,
men forskarna är själva ansvariga
för rapportens innehåll.
Publiceringsår 2025.
Grafisk produktion: Fotoskrift AB
Tryck: Typografiska Ateljén AB
www.handelsradet.se
ISBN: 978-91-89922-05-1

Förord

Vi har i detta projekt haft förmånen att kunna studera viktiga fenomen kring digital handel – hållbarhet (produktreturer och hållbar konsumtion) och digitalisering (artificiell intelligens). Genom ett fruktbart samarbete mellan två akademiska discipliner – Datavetenskap och Företagsekonomi (marknadsföring) har vi kunnat skapa ny kunskap som dess discipliner, var för sig, hade haft svårt att åstadkomma.

Projektet har haft en hög ambition, men det har också varit av största betydelse att vi haft ett nära samarbete med vårt partnerföretag, Gina Tricot, men också andra företag i kläd- och modebranschen. Vår ambition är att fortsätta utveckla forskningen kring digital handel och förädla det multivetenskapliga arbetssättet, samtidigt som vi fortsätter utveckla relationerna till företag och andra intressenter.



Vi vill med detta förord tacka vår finansiär Handelsrådet för deras stöd under projektets gång. Här kan framför allt nämnas rådets stora intresse för spridning av forskningsresultat som varit en inspiration för oss under arbetet. Flera av de aktiviteter vi deltagit i har gett oss unika möjligheter att både sprida våra tankar, men också att få interagera med kloka människor. Här kan bland annat nämnas Ylva Åkesson som föredömligt arrangerat både trendspaningar och podcasts.

Vi vill också tacka berörda på Gina Tricot för deras hjälpsamhet i att göra kunddata tillgängliga för våra returprediktioner. En djupt känd tanke går också till medlemmarna i vår referensgrupp som bidragit med ovärderliga synpunkter och tankar, som ökat relevansen i det forskningsarbete vi genomfört.

Jönköping, februari 2025

Tomas Müllern (professor i Företagsekonomi vid Internationella Handelshögskolan i Jönköping) och Ulf Johansson (professor i Datavetenskap vid Tekniska Högskolan i Jönköping)

Sammanfattning och slutsatser

Detta forskningsprojekt har som övergripande syfte att undersöka om artificiell intelligens kan användas för att dels styra konsumenter mot mer hållbara konsumtionsbeteenden, och dels att utveckla bättre algoritmer för att förutsäga om kunder i digital handel kommer att returnera en vara.

Artificiell intelligens i dess olika former har kommit att spela en stor och växande roll inom digital handel. Företagen har idag stora möjligheter att med grund i stora mängder kunddata, både styra och förutsäga kundbeteenden. Användning av AI inom detaljhandeln (och specifikt digital handel) är ett forskningsfält som utvecklats starkt de senaste tio åren. Forskningen har bland annat kunnat visa på en viss skepticism från konsumenter när det gäller användningen av AI. Frågor kring integritet och data-säkerhet har verkat avskräckande på konsumenter, trots AI:s dokumenterade förmåga att både ge träffsäkra och skraddarsydda lösningar (till exempel i form av riktade produktrekommendationer).

I projektet har det övergripande syftet undersökts på två olika sätt. För det första har projektet undersökt hur konsumenter engagerar sig i hållbar konsumtion, och i vilken utsträckning dom är villiga att dela personlig information med företag. För det andra har projektet utvecklat algoritmer för att förutsäga kunders retur-beteenden vid digital handel.

Tre mål har ställts upp för projektet:

Mål 1

Undersöka hur konsumenters attityder och köpintentioner ser ut när dom får AI-genererade hållbara produktrekommendationer, jämfört med när dessa rekommendationer gäller "vanliga" produkter (utan en tydlig hållbarhetsprofil)?

Mål 2

Demonstrera hur konsumenters vilja att dela personliga data och att engagera sig i positivt word-of-mouth, påverkas av om dom mottar produktrekommendationer genererade av AI jämfört med rekommendationer genererade av en människa?

Mål 3

Designa och driftsätta, samt sammanställa ett rikt dataset för produktreturer som innehåller mängder av data kring kunder, köporder och produkter som gör det möjligt att använda maskinlärning för att generera modeller som kan förutsäga produktreturer.

Dessa tre mål har undersökts med en kombination av kontrollerade experiment med paneler av konsumenter (mål 1 och 2), samt tillämpande av maskinlärande för att utveckla och förfina algoritmer för att förutsäga retur beteenden (mål 3). Projektet har genererat ett antal viktiga resultat och dessa kan sammanfattas per mål:

Mål 1

Baserat på de genomförda konsumentexperimenten kan vi konstatera att det *inte* finns någon signifikant skillnad vad gäller konsumenters intention att välja hållbara kontra konventionella produkter. Detta gäller både när de mottar en AI-genererad produktrekommendation och en rekommendation från en människa. Det finns därmed inte stöd för tanken att AI-genererade produktrekommendationer skulle kunna användas än mer för att bidra till mer hållbara konsumtionsval. Däremot bekräftar experimenten att det finns en aversion mot AI där rekommendationer genererade via AI (för båda typerna av produkter) skapar lägre köpintentioner jämfört med rekommendationer från en människa.

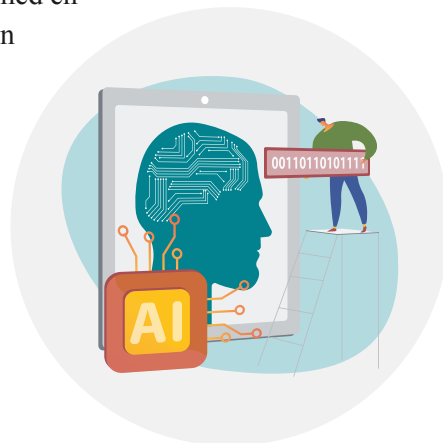
Mål 2

I en serie av experiment prövades olika aspekter av konsumenters vilja att dela personlig information med företag, samt att engagera sig i positivt word-of-mouth. Sammanfattningsvis kan två resultat lyftas fram. För det första demonstrerar projektet att konsumenter är mer positiva till att dela data och att tala gott om företaget/produkten när dom får en specifik produktrekommendation jämfört med en mer allmänt hållen sådan. För det andra visar experimenten att man är mer positiv till att dela data och tala gott om företaget, om det gäller ett mindre företag jämfört med ett stort.

Mål 3

Under detta mål har projektet använt sig av data från Gina Tricot för att med hjälp av algoritmer kunna förutsäga om lagda order kommer resultera i en retur. Den resulterande datamängden är relativt stor men samtidigt typisk för e-handlare inom kläder och mode.

Vi tror därför att många andra aktörer än Gina Tricot kan hämta inspiration från det här projektet och dess resultat. Det framtagna systemet har prövats med omfattande experiment och vi anser att träffsäkerheten är tillräckligt hög för att det bör kunna användas som beslutsstöd avseende om man bör agera på en förväntad retur eller inte. I experimenten har förstagångskunder och återkommande kunder studerats separat, och resultaten visar att systemet fungerar väl för båda dessa kategorier. Tyvärr har systemet inte hunnit driftsättas inom projektiden. Därmed har det förstås heller inte kunnat prövats skarpt, alltså i realtid med lagda order. Dock har utvärderingen följt praxis inom maskininlärning och dataanalys, varför vi är trygga med studiens validitet och specifikt våra resulterande rekommendationer. Detaljer kring såväl metod som resultat och analys finns noggrant beskrivna senare i rapporten.



Innehåll

1. Inledning – projektets syfte och mål	7
2. Teori och metod	10
2.1 Teoretisk referensram för projektet	10
2.1.1 Teoriområde 1: Integritetsfrågor kopplade till AI	10
2.1.2 Teoriområde 2: AI-motstånd/motvilja	11
2.1.3 Teoriområde 3: Förmänskligande av AI	11
2.1.4 Teoriområde 4: Returprediktion	12
2.2 Metodupplägg för projektets två delar	13
2.2.1 Konsumentfokus – experiment som forskningsmetod	13
2.2.2 Datavetenskapligt fokus – prediktiv modellering som forskningsmetod	14
3. Resultat och analys	17
3.1 Mål 1: Hur bidrar AI till hållbar konsumtion?	17
3.2 Mål 2: Hur villiga är konsumenter att dela personliga data?	19
3.3 Mål 3: Hur kan vi förutsäga kunders retur beteenden?	23
4. Diskussion och slutsatser	28
Referenser	31

Inledning – projektets syfte och mål

Detta projekt har tagit sig an den stora och viktiga frågan hur e-handelsföretag med hjälp av artificiell intelligens kan förutsäga kundbeteenden, men också påverka konsumenter att göra mer hållbara val. Vi har undersökt detta i en bransch som är alldeles speciellt intressant ur både ett AI- och ett hållbarhetsperspektiv, nämligen kläd- och modebranschen.

Studien har tittat på frågan ur två olika men sammanlänkade perspektiv. För det första undersöks frågan ur ett konsumentperspektiv, där frågor kring hur konsumenter påverkas av AI i sina hållbara konsumtionsval har stått i fokus. Denna del av projektet faller inom det akademiska fältet konsumentmarknadsföring. För det andra använder vi AI-tekniken maskinlärning för att titta närmare på ett centralt område kring konsumentbeteende nämligen produktreturer, vilket är ett stort problem just inom kläder och mode. Ambitionen har där varit att förbättra precisionen i att kunna förutsäga om en viss kund kommer att returnera en vara denne köpt. I denna inledning utvecklar vi projektets övergripande syfte samt de frågeställningar som projektet avsett att besvara.

Den snabba teknologiska utvecklingen inom artificiell intelligens har skapat nya och unika möjligheter för detaljhandelsföretag att både förutsäga och påverka konsumenters beteenden (Jussopow med flera, 2020). AI är ett stort och komplext fält och det är inledningsvis viktigt att ge en kort introduktion till vad det är och hur det kommit att användas inom kläd- och modebranschen. Enkelt uttryckt kan man säga att AI handlar om att skapa algoritmer som kan utföra uppgifter som traditionellt ansetts vara unika för människor (Song med flera, 2022). Man brukar också lägga in i definitionen av AI att det handlar om att skapa ”system” med en förmåga att lära och att inte bara utföra uppgifter som är programmerade för. Vi tillskriver därmed AI en viss autonomi som tidigare varit förbehållen människor (Formosa, 2021). Lite slarvigt säger vi att AI-system är intelligenta, och med det menar vi att dom på ett autonomt sätt uppvisar en förmåga att lära och att med en viss grad av självständighet kunna lösa svåra problem. Inom detaljhandeln har AI kommit till en hel del användning, och då framförallt inom digital handel. Bra exempel på hur AI används är till exempel att generera produktrekommendationer, att analysera stora mängder kunddata, samt att kommunicera med kunder via chatbotar. Listan kan dock göras längre och den är snabbt växande. Ett viktigt nyckelord när det gäller artificiell intelligens är data (för att vara tydlig – stora mängder



data) och att AI har gjort det möjligt att bygga system som både kan analysera stora mängder data, samt att skapa innehåll baserat på data, allt detta på ett autonomt sätt.

Möjligheten att kunna bidra till mer hållbara kundbeteenden är speciellt intressant mot bakgrund av att inte minst modebranschen är utsatt för ett hårt tryck från allmänheten, men också politiskt, att minska sitt klimatavtryck samt att agera på ett socialt hållbart sätt (till exempel genom att erbjuda bra arbetsvillkor vid produktion i lågkostnadsländer). Modebranschen är påverkad av åtminstone två stora problem med tydliga miljökonsekvenser, som båda har kopplingar till konsumenters köpbeteenden (men givetvis också till företagens agerande – det är ett samspel mellan företag och konsument). Det första problemet som präglar modebranschen, och framför allt dom delar som kan karaktäriseras som ”fast-fashion”, är en överkonsumtion som formas i en ”slit-och-slägg”-kultur där kläder har en kort livscykel hos den enskilde konsumenten. Nära kopplat till detta överkonsumtionsproblem har vi problemet med produktreturer, vilket sedan den digitala handelns intåg varit ett stort och växande problem. Det ska återigen poängteras att det leder helt fel att säga att detta endast är konsumentrelaterade problem. Företagen är i högsta grad medaktörer i att forma köpbeteenden. Det är till exempel svårt att tänka sig att överkonsumtion skulle vara ett så stort problem om inte företagen agerade med lättillgängliga varor, till låga priser, och marknadsförda med bitvis aggressiv reklam. På samma sätt samspelar konsumenters retur beteenden med vilka returvillkor företagen har. Mot denna bakgrund utvecklar vi nedan de mer precisa frågeställningar projektet arbetat med under drygt två år.

Den snabba teknologiska utvecklingen inom artificiell intelligens har skapat nya och unika möjligheter för detaljhandelsföretag att både förutsäga och påverka konsumenters beteenden.

De första två frågeställningarna i projektet rör sig kring konsumenters beteenden. En viktig aspekt på konsumentbeteende är hur konsumenter kan stimuleras att handla på ett mer hållbart sätt, genom att välja mer hållbara produkter (med lägre klimatavtryck, högre kvalitet och livslängd, men också producerade på ett socialt hållbart sätt) (Reczek med flera, 2018; Rios med flera, 2015). I en digital värld är det en intressant fråga om användning av artificiell intelligens kan spela en roll i att forma mer hållbara köpbeteenden. Mer specifikt har vi undersökt *om konsumenters attityder och köpintentioner ser olika ut när dom får AI-genererade hållbara produktrekommendationer, jämfört med när dessa rekommendationer gäller ”vanliga” produkter (utan en tydlig hållbarhetsprofil)*. Vi har också kontrollerat om skillnaden mellan hållbara och vanliga produkter påverkas av om rekommendationen är genererad av AI eller om det framgår att den genererats av en människa. Tidigare forskning har visat att så är fallet, men här får vi en möjlighet att undersöka om detta också gäller när vi jämför hållbara och konventionella produkter. I resultatdelen nedan berättar vi mer om hur experimenten kopplat till denna frågeställning lagts upp och vilka resultat vi fått.

Den andra frågeställningen kopplad till konsumentbeteenden fokuserar ännu mer på användningen av artificiell intelligens och hur konsumenter reagerar på detta. När företag använder AI så har dom till sitt förfogande kraftfulla redskap för att bearbeta data från kunder, i syfte att generera till exempel produktrekommendationer och skicka riktad reklam. Detta är dock inte oproblemiskt – företags användning av AI-genererade produktrekommendationer väcker etiska frågor kring integritet, anonymitet och datasäkerhet. I detta projekt är vi mer specifikt intresserade av att *undersöka om konsumenters vilja att dela personliga data och att engagera sig i positivt word-of-mouth, påverkas av om dom mottar produktrekommendationer genererade av AI jämfört med rekommendationer genererade av en människa*. På samma som ovan så är det också en intressant fråga om det finns en signifikant skillnad i konsumenters vilja att dela data om det gäller hållbara produkter jämfört med konventionella produkter. *Frågan är om konsumenter som uppvisar en benägenhet att välja hållbara produkter är mer eller mindre känsliga för företags användning av AI*. Är dom etiskt mer medvetna om AI:s risker vad gäller integritet, eller tvärtom mer öppna för att dela data när det gäller hållbara produkter.

När det gäller projektets tredje frågeställning så flyttar vi över till ett datavetenskapligt perspektiv där fokus riktas mot ett centralt hållbarhetsproblem, nämligen produktreturer (El Kihal och Shehu, 2022; Janakiraman med flera, 2016). Returer har negativa konsekvenser vad gäller åtminstone två hållbarhetsområden. Det har dels negativa finansiella konsekvenser för företag, vilket riskerar påverka deras finansiella hållbarhet. Den finansiella frågan är dock komplex, där det finns dynamiska samband till exempel mellan företags returpolicier och försäljningssiffror, där högre kostnader för returer kan vara ett ”pris” företag är beredda att betala för en ökad försäljning. Returer har också miljömässiga konsekvenser, och dessa är mer tydliga och det är utan tvekan så att det finns ett rakt samband mellan antalet returer och den ”miljömässiga kostnaden”.

Produktreturer är ett område som är omdebatterat och det sker en hel del förändringar just nu kring hur företag förhåller sig till returer, till exempel frågan om man ska introducera tuffare returregler. Returer är en oundgänglig del av handeln och det är givetvis så att en kund ska kunna returnera en vara som har fel storlek eller som ur kundens perspektiv är fel. I många fall är det dock möjligt att förhindra returer genom att ge bättre möjligheter för kunden att göra ett informerat val, till exempel genom tydligare produkt- och storleksinformation. Skälen till returer kan dock vara många och komplexa, och det ligger i företagets intresse att bli bättre på att förstå varför kunder returnerar och att kunna prognosticera om en kund kommer att returnera en viss vara.

I detta projekt använder vi AI i form av maskinlärning för att utveckla algoritmer för att prognosticera returer. När en kund lägger vissa typer av varor i varukorgen är ambitionen att med hjälp av välkalibrerade algoritmer kunna förutsäga om innehållet i varukorgen kommer att generera en retur. Mer precist uttryckt har projektet *designat och utvärderat ett system som med hjälp av maskininlärning och en rik datamängd innehållande information om kunder, order och produkter, med tillräcklig träffsäkerhet kan förutsäga produktreturer*.

2 Teori och metod

Detta forskningsprojekt bygger på flera olika teoretiska och metodmässiga utgångspunkter, vilket framförallt relaterar till de två ämnesinriktningar forskargruppen tillhör – marknadsföring och datavetenskap. Teori och metodval är därmed tydligt kopplade till de frågeställningar och mål med projektet som beskrevs ovan. Här följer först en kort redogörelse för teorier som haft stor betydelse för projektet, följt av en kort genomgång av metodupplägg i projektet.

2.1 Teoretisk referensram för projektet

Under de senaste åren har det vuxit fram en rik forskning kring artificiell intelligens och hur användningen av AI påverkar konsumentbeteenden. Detta har givit vårt projekt goda möjligheter att bygga vidare på tidigare forskning vad gäller AI:s roll. Forskningen är bred och snabbt växande, men baserat på en omfattande litteraturgenomgång har vi identifierat följande forskningsområden, som alla har en hög relevans för området digital handel:

Teoriområde 1: Integritetsfrågor kopplade till AI

Detta första forskningsområde är det största området vad gäller AI och konsumentbeteende, och har genererat flera olika forskningsspår. Här kommer vi att kort beskriva de viktigaste som också har format detta projekt. Viktiga delar av forskningen kretsar kring frågan hur konsumenter reagerar på företags insamling och bearbetning av data från individen (köpdata, surfvanor, kommunikation online, bilder, med mera).

En central fråga som genererat en del forskning (men inte tillräckligt) är under vilka villkor konsumenter är villiga att dela data om sig själva (Ackermann, K.A. med flera, 2022; Hu och Min, 2023). Det är en teoretiskt intressant fråga ”hur hårt” individen bevakar och håller i data om sig själv. I praktiken är detta inte ett tydligt val individen gör, utan delande av data sker mer eller mindre automatiskt när personen rör sig online. Vi är med andra ord sällan medvetna om att vi delar data på en daglig basis. I takt med ett ökat intresse från myndigheter att reglera användningen av AI så är det dock inte orimligt att tänka sig en nära framtid där val att dela data är ännu mer tydliggjorda till exempel i webshoppar. Tidigare forskning kring detta område har bland annat fokuserat på att mäta hur viktigt dataintegritet är för den enskilde, och hur detta påverkar deras vilja att dela data. Här har flera studier uppmärksammat att detta påverkas



av vilken kontroll (upplevd kontroll) den enskilde har över sina data (Brandimarte med flera, 2012).

Frågan om konsumenters vilja att dela data kopplar tydligt till en central och framväxande teori kring AI och konsumentbeteende, den så kallade PP-paradoxen (PI om vi översätter den till svenska). PP-paradoxen bygger på den grundläggande insikten att AI ger helt nya möjligheter för företag att skraddarsy kommunikation för den enskilde – lite av reklamfältets heliga graal (Cloarec, 2020). *Personalisering* beskrivs ofta som ett positivt värde, det ger företaget möjlighet att skaffa information för att öka kundupplevelsen, genom mer skraddarsydd kommunikation till konsumenten. Kommunikation, potentiellt, blir mer relevant genom att bygga på data kring den enskildes vanor och beteenden. Denna personalisering kommer dock med ett pris – det kan upplevas som stående i konflikt med den enskildes *integritet*. Paradoxen ligger i att ju mer skraddarsydd kommunikationen blir desto större riskerar känslan av integritetskränkning bli. Denna paradox har demonstrerats i ett antal studier och den utgör en viktig del av vår referensram i studien (Castelo med flera, 2019; Puntoni med flera, 2021).

Under de senaste åren har det vuxit fram en rik forskning kring artificiell intelligens och hur användningen av AI påverkar konsumentbeteenden.

Teoriområde 2: AI-motstånd/motvilja

Ett andra teoriområde som har stor relevans för detta projekt är tendens hos människor att hysa en misstro mot AI, det som kommit att kallas AI/algorithm-aversion (Dietvorst med flera, 2018; Longoni och Cian, 2020). I den ursprungliga formuleringen handlade AI-aversion om att människor är mer benägna att se fel hos AI snarare än att se att AI ofta gör mer korrekta bedömningar än människor. Detta har dock kommit att handla om en bredare syn där människor är skeptiska (averse) till att lita på bedömningar genererade av AI, trots att dom vet att maskinen ofta genererar mer korrekta bedömningar. Denna tendens till aversion är dock inte allmän utan den verkar uppstå i vissa situationer, under vissa villkor. Den växande forskningen har på olika sätt försökt att bekräfta under vilka villkor denna aversion uppstår.

Här förtjänar det att nämnas att vissa studier har demonstrerat att kunder som handlar mer frekvent online uppvisar en lägre aversion än kunder som handlar mer i fysiska affärer (Nichifor med flera, 2021).

Teoriområde 3: Förmänskligande av AI

Det tredje forskningsområdet ligger lite utanför detta projekt, men är av stort allmänt intresse och förtjänar att omnämnas här. Det ligger i definitionen av AI ("intelligens") att den i någon mening efterliknar (och kanske till och med överträffar) mänsklig intelligens. Samtidigt är och förblir AI en maskin. Det faktum att det är en maskin

påverkar, som vi sett ovan, människors tilltro till AI. En alltmer vanlig strategi för att hantera detta förtroendegap är att ”ge” AI olika mänskliga egenskaper, till exempel vad gäller tal och utseende (Hermann, 2022). Det är kanske enklast att se detta framför sig i utvecklingen av ”mänskliga” robotar, men strategin är frekvent använd till exempel i chatbotar och andra former av helpdesk/servicefunktioner.

Dessa tre teman är generella men det kan noteras hur framför allt tema 1 och 2 kopplar tydligt samman med frågor kring social hållbarhet, och framför allt frågan om integritet. Dessa tre teoriområden har vi använt när de olika experimenten har konstruerats. När det gäller frågan om vi kan se andra konsumentbeteenden när det gäller hållbara produkter så visar våra litteraturgenomgångar att det i stor utsträckning saknas empirisk forskning kring dessa. Det också detta som är det stora nyhetsvärdet i denna del av projektet – att studera de olika aspekterna av AI när det gäller hållbara produkter.

Teoriområde 4: Returprediktion

Den övergripande målsättningen är att genom exempel visa att det är möjligt för ett företag inom kläder och mode att framgångsrikt använda dataanalys som ett verktyg för att minska mängden returer (Daniel och Ülkü, 2018; Sweidan med flera, 2020). Mer konkret ska en AI (en algoritm) då en kund precis genomfört ett köp online förutsäga om den lagda ordern kommer att resultera i en retur eller inte. Exakt vilken åtgärd företaget i så fall genomför ligger utanför studien, alltså algoritmen ska i det här stadiet ses som beslutsstöd, snarare än automatiskt beslutsfattande.

Det är inte helt uppenbart vad ”framgångsrikt” i stycket ovan innebär. Dock bör man vara medveten om att de flesta företag i dag helt enkelt tillåter fria returer. Varje lösning som kan förutsäga i alla fall några returer med tillräcklig säkerhet måste därför anses ha ett värde.

Mer tekniskt klassificerar algoritmen varje order i det ögonblick de läggs som antingen retur eller inte. Med en retur avses här om kunden kommer returnera minst en av de köpta produkterna. Algoritmens prestanda kan därför mätas som dess träffsäkerhet (engelska accuracy) vilket helt enkelt är andelen korrekta förutsägelser. Dock måste man här notera att de två fel som algoritmen kan göra är olika allvarliga; om algoritmen missar en retur är det inte värre än att som i dag inte genomföra några prediktioner alls och tillåta fria returer. Om algoritmen däremot felaktigt klassificerar en order som en retur (det vill säga kunden hade inte returnerat något från ordern) och företaget agerar på denna förutsägelse finns risken att kunden inte bara stryker ordern utan eventuellt även helt slutar handla hos företaget. En felaktig prediktion om att en order kommer returneras, en så kallad *falsk positiv*, kan därför vara mycket kostsam för företaget. Detta innebär att ett system för returprediktioner måste undvika falska positiva (Urbanke med flera, 2015).

Baserat på ovanstående resonemang har studiens syfte varit att skapa ett datadrivet system för returprediktion där algoritmen kompletterar varje prediktion med en *konfidens*, alltså hur säker algoritmen är. Den tänkta användningen är då förstas att

företaget bara agerar på de returprediktioner där algoritmen är tillräckligt säker. Specifikt, om konfidensen uttrycks som en välkalibrerad sannolikhet, så ger det ett väldigt gott beslutsstöd. Om företaget exempelvis väljer att agera på alla order som prediceras som returer med en konfidens högre än 95 procent, så kommer högst fem procent av dessa prediktioner att vara felaktiga. Alltså, om företaget agerar på 20 dylika prediktioner, så kommer minst 19 av dessa att vara korrekta, i snitt.

2.2 Metodupplägg för projektets två delar

Konsumentfokus – experiment som forskningsmetod

Vilka konsekvenser på konsumenternas köpbeteenden får hållbara produktrekommendationer, och vilken roll spelar AI i att förmedla dessa rekommendationer? Förändras konsumentens vilja att köpa hållbart baserat på om rekommendationerna är genererade av AI eller av en människa? Vi har genomfört kontrollerade och upprepade experiment för att förstå hur konsumenters attityder till ett varumärke förändras när dom får AI-genererade hållbara produktrekommendationer jämfört med när dom får AI-genererade konventionella produktrekommendationer.

I en serie av experiment har vi studerat olika aspekter av dessa övergripande frågeställningar som alla har ett stort intresse för digital handel. Experimenten är i korthet upplagda så att deltagarna i varje experiment ställs inför ett scenario som inrymmer den faktor eller aspekt som vi vill undersöka i det aktuella experimentet.

Det kan till exempel handla om att konsumenten i scenariot erhåller en specifik produktrekommendation – vi är här mest intresserade av att studera rekommendationer för hållbara produkter. Vi undersöker också vilken roll det spelar om rekommendationen är genererad via AI eller ej. Den specifika rekommendationen kallas i experimenttermer för den oberoende variabeln. Denna variabel tänker vi oss sedan påverkar en eller flera beroende variabler. Beroende variabler i experimenten är till exempel deltagarens intention att köpa produkten, attityd till varumärket, men också sådant som förtroende för rekommendationen och intention att tala gott om produkten/varumärket (det som kallas word-of-mouth, WOM).

Experimenten är designade i en digital plattform som heter Qualtrics, där man också kan analysera resultaten från experimenten. Deltagarna i studien får en länk där man klickar in sig till experimentet och datan lagras på ett säkert och anonymt sätt hos Qualtrics.

En viktig fråga i all experimentell samhällsforskning är hur man rekryterar deltagare till experimenten på sådant sätt att man undviker obalanserade urval och snedvridning. För att få robusta resultat så har vi konsekvent valt att använda 400 deltagare per experiment, 200 som får svara på frågor kring experimentfaktorn (AI genererar



rekommendationen), och 200 som agerar som en kontrollgrupp och har ett scenario utan den undersökta faktorn (rekommendation genererad utan AI). Genom att jämföra resultaten i de två grupperna går det, statistiskt, att säga om den undersökta faktorn har en signifikant effekt.

Vi använder oss av så kallade crowdsourcingsajter där man rekryterar deltagare som får en mindre ersättning för sitt deltagande. Dessa sajter har goda möjligheter att rekrytera tillräckligt stora antal deltagare från breda konsumentgrupper. I vårt fall har vi valt breda konsumentpaneler med amerikanska och europeiska, vuxna, deltagare och med en jämn könsfördelning.

Vilka konsekvenser på konsumenternas köpbeteenden får hållbara produktrekommendationer, och vilken roll spelar AI i att förmedla dessa rekommendationer?

Datavetenskapligt fokus – prediktiv modellering som forskningsmetod

Studien utgår från en datamängd insamlad mellan 210101–230131. Datamängden innehåller totalt 1 268 000 order lagda online. Den övergripande metoden är så kallad *prediktiv modellering*, det vill säga en algoritm används för att till exempel skapa en modell vilken sedan används för prediktionerna. Då modellens syfte är att kunna förutsäga om en lagd order kommer resultera i en retur eller inte utgörs exemplen av just order, samt huruvida dessa returnerades eller inte. Varje order beskrivs här med ett antal variabler (features eller attribut) samt ett binärt målattribut, vilket helt enkelt säger om ordern returnerades eller inte. Algoritmen lär sig därmed vilka attributvärden som kännetecknar en order som resulterar i en retur. Den resulterande modellen kan sedan användas för att predicera om en ny order, utifrån dess attributvärden, kommer returneras eller inte.

På en övergripande nivå innehåller attributen data om själva ordern, kunden, samt i ordern ingående produkter. Nedan presenteras de olika attributen i detalj:

Order:

- Order value: summan av varornas pris
- Order discount: total rabatt i ordern
- Total items: antalet produkter i ordern
- Most expensive: priset på dyraste varan i ordern
- Diff sizes: binär variabel, 1 om ordern innehåller samma plagg i olika storlekar
- Diff colours: binär variabel, 1 om ordern innehåller samma plagg i olika färger
- Order weekday: vardag (måndag–torsdag), fredag eller helg (lördag–söndag)
- Order hour: dag (6–17), kväll (17–00), natt (00–06)
- Freight cost: fraktkostnad

Kund:

- C orders: kundens antal tidigare order
- C returns: kundens antal tidigare returer
- C return ratio: C returns / C orders

Produkt:

- P max ratio: tidigare returandel för den produkt i ordern med högst andel returer
- P maxLP ratio: laplace-korrigerad tidigare returandel för den produkt i ordern med högst andel returer
- P avg ratio: medelvärdet av i ordern ingående produkters tidigare returandel

Viktigt att notera är här att två speciella attribut (Diff sizes och Diff colours) togs fram för varje order, nämligen innehåller ordern samma plagg i olika storlekar respektive samma plagg i olika färger. Att köpa ett plagg i flera storlekar för att kunna prova dessa hemma och sedan returnera de storlekar som inte passar är ett vanligt beteende starkt förknippat med returer. Detsamma gäller, fast i mycket mindre mån, för att prova olika färger.

I datamängdens nuvarande form finns inte beskrivet vilken sorts produkt som ingår i en order, utan bara produktnummer. Varken vi eller algoritmen vet alltså om ett visst produktnummer motsvarar exempelvis en skjorta eller ett par byxor. Däremot är produktnummer konsistenta över tid, så vi kan ta fram historik för hur frekvent en viss produkt returnerats tidigare. Ett rimligt antagande är att vissa produkter kommer returneras oftare än andra, exempelvis för att deras storlekar är små eller stora, alternativt att plagget inte ser lika bra ut i verkligheten som på webshoppens bild. Genom att inkludera returhistorik för produkter som attribut kan algoritmen använda sig av den informationen.

Självklart finns kundhistorik bara tillgängligt för kunder som handlat förut. I studien valdes därför att behandla (och modellera) förstagångskunder och återkommande kunder separat. En rimlig förväntan var här att det skulle visa sig enklare att modellera återkommande kunder.

Totalt användes för förstagångskunder alltså tolv attribut, medan ytterligare tre rörande kundhistorik lades till för återkommande kunder.

När man utvärderar en prediktiv modell måste detta göras på data som inte använts för skapandet av modellen. Detta är naturligt då syftet ju är att uppskatta förväntad prestanda för ny data, alltså i det här fallet för order där vi inte vet om de kommer returneras eller inte. Standardsättet är här att emulera prestanda på ny data genom att skapa modellen med hjälp av en viss del av den tillgängliga datamängden, och utvärdera den på resten. De resultat som då fås bör vara representativa för vad man kan förvänta sig vid skarp användning. I studien användes de första 70 procenten av datan (i tidsordning) för att skapa modellen, och de sista 30 procenten för utvärderingen.

De utvärderingsmått som används är följande:

- Träffsäkerhet (accuracy): andelen av alla prediktioner som är korrekta
- Precision: andelen av alla positiva prediktioner (prediktioner att en order kommer resultera i en retur) som är korrekta
- Recall: andelen av alla faktiska returer som prediceras som returer

Enligt målsättningen ovan, samt existerande litteratur om datadrivna prediktionssystem för returer, bör ett framgångsrikt system ha (mycket) hög precision med acceptabel recall.

För själva modelleringen har olika algoritmer prövats. De resultat som redovisas nedan är för algoritmen *XGBoost*, vilken tidigare har visat sig vara bland de starkaste i många olika tillämpningar.

De flesta klassificerare, inklusive *XGBoost*, predicerar ett värde för respektive klass, och dessa värden summerar till ett. Det är därför frestande att tolka värdena som sannolikheter, men det är strikt talat inte korrekt. Hur väl dessa värden motsvarar faktiska sannolikheter beror bland annat på vilken algoritm som används vid modelleringen. Specifikt behöver vissa modeller kalibreras medan andra kan användas utan kalibrering. *XGBoost* är generellt någorlunda välkalibrerad, åtminstone om den tränats på tillräckligt mycket data. Oavsett kalibrering eller inte kan man alltid utgå från att ett högre värde för exempelvis klass 1 (retur) faktiskt också innebär att den ordern har en större sannolikhet att returneras. Om vi då väljer att bara agera på de order som har ett högre värde (konfidens) än en viss nivå, så kommer detta att ge oss en direkt trade-off mellan precision och recall. Ju högre gränsvärde vi väljer, desto färre order kommer vi att predicera som returer (där vi egentligen menar ”returer som vi ska agera på”), det vill säga recall går ner, men vi kommer samtidigt att vara säkrare på de vi faktiskt predicerar, så precision går upp.

Tekniskt kallas det vi mäter då för ”top x procent accuracy”, alltså vilken träffsäkerhet får vi på de x procent där vi är mest säkra. I den givna situationen vill vi, enligt ovan, maximera precision, så vi kommer förstås välja de x procent där vi är mest säkra på att de utgör returer.

Om vi antar att konfidensvärdena är någorlunda välkalibrerade kan vi dessutom i förväg uppskatta precision för olika gränsvärden, vilket så klart är oerhört viktigt. I den genomförda studien har vi valt att inte kalibrera *XGBoost*, utan vi använder dess konfidenser.

I resultaten nedan redovisar vi då precision och recall för ett antal ”top x procent” beroende på hur selektiva vi vill vara. Vi redovisar också en förväntad precision, alltså modellens uppskattade precision, för dessa värden.

Resultat och analys

I detta avsnitt redogör vi för resultaten från de olika studier som genomförts i projektet. Vi följer här upplägget från tidigare i rapporten där vi beskriver resultaten utifrån forskningsfrågorna och de tre målen i projektet.

3.1 Mål 1: Hur bidrar AI till hållbar konsumtion?

Konsumenter som handlar digitalt utsätts för olika former av påverkan från det aktuella företaget, och vi har här intresserat oss mer specifikt för produktrekommendationer. Det är ett stort och viktigt område för företagen (men också för konsumenter) och det ses allmänt som ett viktigt redskap för marknadskommunikation. Vi kan se det som ett av vår tids viktigare redskap för reklam, där internet har skapat nya medier som snabbt och kostnadseffektivt kan nå många. Här är det av avgörande betydelse att poängtera AI:s roll i att göra reklam individanpassad. Produktrekommendationer genererade med AI är inte bara en form av reklam som når många, det är dessutom en skräddarsydd reklam för den enskilde.



Vi beskrev ovan hur forskningen kring produktrekommendationer har vuxit snabbt och vi vet redan mycket om deras roll, betydelse och avigsidor. Under mål 1 har vi mer specifikt undersökt om AI:s roll förändras om produktrekommendationer gäller hållbara produkter jämfört med om det gäller konventionella produkter. För företag är det intressant att veta om man kan påverka konsumenter att göra mer hållbara val med hjälp av AI-genererade produktrekommendationer.

För att pröva detta genomfördes två experiment där vi prövade med olika produkter i de två experimenten (hörlurar och ryggsäckar). I båda experimenten delades de 400 deltagarna slumpmässigt in i en av fyra grupper. Deltagarna i varje grupp fick inledningsvis ta del av ett scenario som kunde se ut så här:

Tänk dig att du funderar på att köpa ett par nya hörlurar. För att göra ett så bra köp som möjligt så söker du på internet efter information om hörlurar. Medan du söker så dyker en produktrekommendation från en virtuell aktör baserad på artificiell intelligens upp som rapporterar följande: Hej, har du sett Urban Brands nya over-earhörlurar. Dom är byggda i ett material från återvunnen plast.

En annan grupp kunde få ett scenario där det framgick att rekommendationen genererats från en annan konsument, men fortfarande med en hållbar produkt. De övriga två grupperna fick produktrekommendationer för hörlurar som beskrevs utan en hållbarhetsprofil där man i stället framhåller tekniska egenskaper (bra ljud).

Sammantaget undersökte vi följande fyra fall (100 deltagare per sådan grupp):

- Hörlurar med hållbar profil, rekommendation genererad av AI
- Hörlurar med hållbar profil, rekommendation genererad av en människa
- Konventionella hörlurar, rekommendation genererad av AI
- Konventionella hörlurar, rekommendation genererad av en människa

Alla 400 deltagarna fick sedan svara på ett antal frågor som relaterade till det scenario dom just läst. De första frågorna mätte i vilken utsträckning deltagaren skulle kunna tänka sig (var intresserad av) att köpa produkten i scenariot. Vi hade också ett antal följdfrågor för att kontrollera för ålder, kön och politisk profil (liberal/konservativ), samt inkomst.

Kunde vi då se en statistiskt säkerställd skillnad mellan de fyra grupperna? Mer precist uttryckt så var vi intresserade av att se om det fanns en skillnad mellan hållbara produkter och konventionella och om det gick att påvisa någon direkt effekt av AI för hållbara produkter.

Vi var intresserade av att se om det fanns en skillnad mellan hållbara produkter och konventionella och om det gick att påvisa någon direkt effekt av AI för hållbara produkter.

Resultaten visar att det *inte* fanns någon signifikant skillnad mellan de två grupper som fick hållbara produkter och de som fick konventionella. Det gick därmed inte att påvisa någon skillnad beroende på typ av produkt – man hade ungefär samma inställning till AI-genererade rekommendationer i båda grupperna. Det vi kunde se var att det fanns en skillnad mellan de som fick en rekommendation via AI och de som fick en rekommendation via en människa (en annan konsument). När rekommendationen var genererad via AI så var konsumenten mer skeptisk till produkten även om den skillnaden inte var fullt ut signifikant. Däremot finns ingen skillnad i resultat beroende på om det var en hållbar eller konventionell produkt. För båda typerna av produkt var man något mer skeptisk till AI-rekommendationer än mänskligt genererade rekommendationer.

Vi kontrollerade också om resultaten påverkades av om konsumenten hade ett högt förtroende för AI. Resultaten visade att om konsumenter hade ett högt förtroende för AI så var dom mer benägna att påverkas av AI-genererade produktrekommendationer, även om detta resultat inte var statistiskt signifikant.

För att undanröja risken att dessa resultat berodde på den aktuella produktkategorin, så valde vi att upprepa samma experiment men använda oss av en annan produkt. Valet föll på ryggsäckar – i övrigt var experimentet upplagt på exakt samma sätt. Inte heller denna gång fick vi någon signifikant skillnad mellan grupperna hållbar och konventionell produkt. Detta resultat är intressant och pekar mot en utmaning för arbetet med att få konsumenter att göra mer hållbara val. Det tycks som att användning av AI inte har någon påverkan på konsumenters vilja att göra hållbara val, och det väcker intressanta frågor om det finns andra sätt (än AI-genererade hållbara produktrekommendationer) att väcka intresset för hållbara produkter?

Vi kunde också demonstrera samma nivå av AI-aversion i detta experiment. Tidigare forskning har liksom här påvisat en tendens till aversion mot AI som är konsistent över olika typer av produkter, geografiska områden, kännedom om AI, samt i viss mån ålder. Vårt bidrag till den forskningen är att vi valt att studera detta i ett område som är intressant ur ett hållbarhetsperspektiv nämligen kläder och mode. Vår studie befäster därmed en viktig grundbult i forskningen kring AI och konsumentbeteende – att det finns en tydlig aversion mot AI.

3.2 Mål 2: Hur villiga är konsumenter att dela personliga data?

Den andra stora frågeställningen i projektet handlade om den viktiga frågan om konsumenters vilja att dela personliga data om sig själva till företag. Frågan är viktig då det handlar om konsumenters gränser för hur långt företag kan gå i att samla in data om den enskilde. Som vi visade ovan i teoriavsnittet finns en växande forskning kring integritetsfrågor i digital handel, men det saknas empirisk forskning kring den mer specifika frågan om konsumenters vilja att dela data med företag. Företag i kläder- och modebranschen efterfrågar/samlar regelmässigt data från den enskilde. I takt med att användningen av AI börjar regleras så blir detta av ökande intresse, och inte minst i de situationer när kunden ges en tydlig möjlighet att kunna välja om man delar data eller ej. Precis som i den första delen av projektet var vi här också intresserade av att se om benägenheten att dela data skilde sig åt vad gäller hållbara produkter jämfört med konventionella produkter. Här genomfördes ett första experiment för att demonstrera om det fanns en skillnad mellan hållbara och konventionella produkter vad gäller viljan att dela personlig information till företag.

Vi har också undersökt den relaterade frågan under vilka förutsättningar konsumenter är benägna att prata positivt om en vara/varumärken (det som kallas positivt word-of-mouth PWOM). I tekniska termer har vi i denna del använt oss av två beroende variabler: vilja att dela data, samt benägenhet att engagera sig i positivt word-of-mouth. Dessa två faktorer har vi undersökt under flera olika förutsättningar. Inledningsvis så prövade vi om det fanns en skillnad mellan hållbara och konventionella produkter vad gäller konsumentens vilja att dela data och engagera sig i PWOM. I en serie av

experiment gick vi sedan vidare och undersökte konsumenters vilja att dela data och engagera sig i PWOM när de mottar specifika eller allmänna produktrekommendationer. Frågan om specifika kontra allmänna produktrekommendationer har föreslagits som en viktig dimension i tidigare forskning, men det saknas empirisk forskning kring det.

Vi genomförde här tre experiment där vi prövade denna dimension under ett antal olika förutsättningar:

- Inledningsvis för ett okänt varumärke för att hålla experimentet neutralt vad gäller varumärket
- Experiment nummer två prövade med ett välkänt varumärke – Nike
- Experiment nummer tre prövade om det fanns en skillnad mellan ett stort och ett litet företag

Nedan följer en kort beskrivning av upplägg samt resultat från de fyra experimenten under mål 2.

Experiment ett: Hållbar produkt jämfört med konventionell produkt

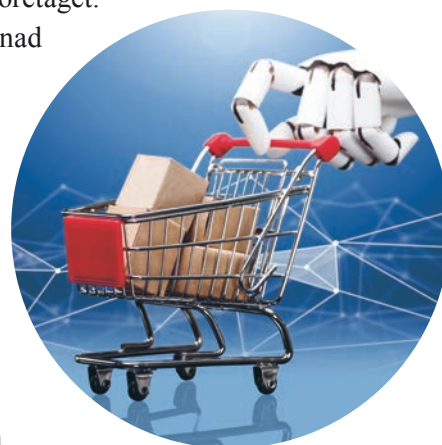
I ett första experiment fick deltagarna (400 personer) ta del av ett scenario som var formulerat enligt följande:

Antag att du får en hållbar produktrekommendation. Om företaget ber dig att dela personlig information, är du mer benägen att dela denna information jämfört med om du får en rekommendation för en konventionell produkt?

Detta experiment var upplagt på ett liknande sätt som experimenten under mål 1, med den viktiga skillnaden att denna gång så var det inte deltagarens köpintention som stod i fokus utan deras vilja att dela personlig information med företaget.

Vi kan konstatera att det inte fanns någon signifikant skillnad mellan hållbara produkter och konventionella. Vi får med andra ord samma resultat som vid experimenten som genomfördes under mål 1. Även om vi inte sett någon skillnad mellan de båda produktkategorierna så är det ändå ett robust resultat som kvarstår när den beroende variabeln ändras.

Under mål 2 i projektet låg det också i frågeställningarna att släppa lite på fokuset på hållbara produkter, och i stället gräva djupare i användningen av AI för att generera produktrekommendationer. Den djupare frågan som vi ställde var: Under vilka förutsättningar är konsumenter benägna att dela personlig information och under vilka förutsättningar är dom benägna att engagera sig i positivt WOM? Denna fråga har vi undersökt i tre skilda experiment.



Experiment två: Specifika versus allmänna produktrekommendationer för ett okänt varumärke

Inledningsvis designade vi ett experiment för att pröva konsumenters vilja att dels dela personliga data med företaget, och dels att engagera sig att tala väl om företaget (positivt word-of-mouth, PWOM). I experimentet prövade vi med andra ord två olika beroende variabler. För att undanröja risken att deltagaren påverkades av att produkten kom från ett känt varumärke så byggde det första experimentet på att varumärket var okänt (vi kallade det för UrbanWave). Produkten vi använde oss av var t-shirts från detta varumärke. Varje deltagare i experimentet fick liksom tidigare ta del av ett scenario och det var uppbyggt enligt följande:

Du är hemma, och i samma rum som du befinner dig i så finns en röststyrd, smart högtalare. Du sitter och samtalar med en god vän och konversationen ser ut på följande sätt:

Du: ”Jag gillar verkligen dom här nya t-shirtarna från UrbanWave. Det är väl ett populärt märke nu?”

Vännen: ”Absolut! UrbanWave är jättetrendigt och dom är kända för stilfulla t-shirts. På senaste tiden har dom fått mycket uppmärksamhet i sociala medier och det finns ett antal artiklar om dom i modetidskrifter.”

Följande dag ser du en AI-rekommendation som visar några produkter från UrbanWaves. När du tittar närmare på rekommendationen så ser du den t-shirt ni hade pratat om dagen innan.

Under vilka förutsättningar är konsumenter benägna att dela personlig information och under vilka förutsättningar är dom benägna att engagera sig i positivt WOM?

Det vi sedan ville pröva var om de två beroende variablerna påverkades av om man fick en specifik eller en allmän rekommendation. I scenariot ovan så visar scenariot på en specifik rekommendation – den visar exakt samma t-shirt som personerna pratat om dagen innan. I kontrollgruppen var detta ändrat till en mer allmänt hållen rekommendation. Det framgår i det fallet att det inte är den t-shirt dom pratat om, utan bara produkter i allmänhet från UrbanWave.

Resultaten är intressanta och visar att deltagarna har en högre vilja att dela data när dom får en specifik rekommendation jämfört med den allmänna. Det kan i förstone framstå som märkligt, man är mer positiv till att dela data när den smarta högtalaren har fångat upp exakt vilken t-shirt de två vännerna pratat om. En tolkning av detta resultat är att deltagarna uppskattar att rekommendationen ”träffar” helt rätt vad gäller produkten

och att det uppfattas som positivt att rekommendation är i linje med deras behov och önskemål.

Experiment tre: Specifika versus allmänna produktrekommendationer för ett välkänt varumärke

För att pröva om resultaten ändras när deltagaren ställs inför ett välkänt varumärke så upprepade vi det föregående experimentet. Den enda skillnaden var att vi i stället för det fiktiva UrbanWaves använde oss av det kända varumärket Nike. Resultaten från detta uppföljande experiment var i linje med det föregående. Deltagarna var mer benägna att dela personliga data och engagera sig i PWOM när de fick en specifik rekommendation från Nike.

Experiment fyra: Specifika versus allmänna produktrekommendationer för ett stort kontra litet företag

De resultat vi fick från de två föregående experimenten är intressanta och inte helt intuitiva. Vi hade ett antal kontrollfrågor för att se om resultaten påverkas av olika personliga faktorer, som till exempel deras allmänna inställning till integritet och deras lojalitet med varumärket (det senare kunde vi av förklarliga skäl endast undersöka i fallet Nike). Dessa personliga faktorer gjorde ingen signifikant skillnad vad gäller effekten av specifika kontra allmänna produktrekommendationer. Deltagarna var med andra ord mer benägna att dela personliga data samt engagera sig i PWOM, oavsett om det gällde ett okänt eller välkänt varumärke, vid specifika produktrekommendationer jämfört med allmänna.

I ett uppföljande experiment prövade vi om resultaten förändras när vi jämförde om rekommendationen gällde ett litet kontra ett stort företag. Här ville vi försöka isolera effekten av om det handlar om ett litet eller stort företag och hur detta påverkar deltagarens vilja att dela data. Detta experiment var lite annorlunda uppbyggt och fokuserade på storleken på företaget, snarare än typ av produktrekommendation (specifik versus allmän).

Deltagarna (400 personer) ställdes inför följande scenario:

Antag att du ska köpa kläder. Efter att ha sökt online så hittar du ett företag som är en stor organisation med tiotusentals anställda.

Kontrollgruppen (hälften av de 400) fick ett scenario där det i stället gällde ett litet företag med ett dussintal anställda). I båda scenarierna undersöktes deltagarens vilja att dela information. Resultaten visade att deltagarna var mer benägna att dela data med ett litet företag jämfört med ett stort. En möjlig tolkning av detta resultat är att man har ett större förtroende för att ett mindre företag inte missbrukar deras data och att dom upplevs som mer pålitliga och transparenta.

3.3 Mål 3: Hur kan vi förutsäga kunders retur beteenden?

Tabell 1 nedan visar resultaten för de båda datamängderna om alla order prediceras. Träffsäkerheten är strax under 75 procent, och precision strax över 75 procent, på båda datamängderna. Däremot är recall betydligt högre för återkommande kunder. En uppenbar delförklaring är här att återkommande kunder faktiskt returnerar oftare än förstagångskunden. Den viktigaste iakttagelsen är emellertid att precision är alldeles för låg för att modellen ska kunna användas på det här viset. Ungefär en returprediktion av fyra skulle då vara felaktig.

Tabell 1. Övergripande prediktiv prestanda.

Datamängd	Accuracy	Precision	Recall
Förstagångskunder	0,737	0,750	0,458
Återkommande kunder	0,740	0,765	0,670

Vi analyserar därför vilken precision vi kan uppnå om vi bara agerar på en mindre del av alla returprediktioner. Tabell 2 nedan sammanfattar resultaten för förstagångskunder. Om vi, som exempel, tittar på raden för 10 procent (där vi alltså bara agerar på de 10 procent av returprediktionerna där vi är mest säkra) så motsvarar det totalt 16 474 order, och där har vi en precision på 90,8 procent, det vill säga vår prediktion är rätt nio gånger av tio. Motsvarande recall är 24,0 procent, alltså vi upptäcker en retur av fyra. Så, av de totalt 164 790 order som vi prövar vår modell på, identifierar vi 16 474 som returer, och av dessa är 14 975 verkligen returer. Vill vi vara mer försiktiga, kan vi istället välja att bara agera på topp 5 procent, och då har vi i stället en precision på 96,5 procent (vi gör fel mer sällan än en gång på tjugo), men vi upptäcker då bara ungefär 13 procent av alla returer. Ett intressant resultat är här att modellen relativt väl uppskattar precisionen, även utan speciell kalibrering. Vi kan slutligen notera att även dessa relativt konservativa nivåer (topp 5 procent och topp 10 procent) innebär att systemet kommer rekommendera att man agerar ett antal gånger varje dag.

Tabell 2. Prediktiv prestanda för olika konfidensnivåer. Förstagångskunder.

Topp x%	#order	Uppskattad precision	Precision	Recall
1%	1 648	0,993	0,987	0,026
2%	3 295	0,986	0,984	0,052
5%	8 237	0,965	0,963	0,127
10%	16 474	0,908	0,909	0,240
15%	24 710	0,836	0,841	0,334
20%	32 947	0,773	0,782	0,414

Tabell 3 nedan sammanfattar, på motsvarande sätt, resultaten för återkommande kunder. Vi ser här att resultaten generellt är betydligt bättre, det är alltså – som förväntat – enklare att förutsäga returerna med tillgång till kundhistorik. Om vi exempelvis tillåter 5 procent felaktiga returprediktioner kan vi, genom att agera på topp 15 procent, identifiera nästan 30 procent av alla returerna.

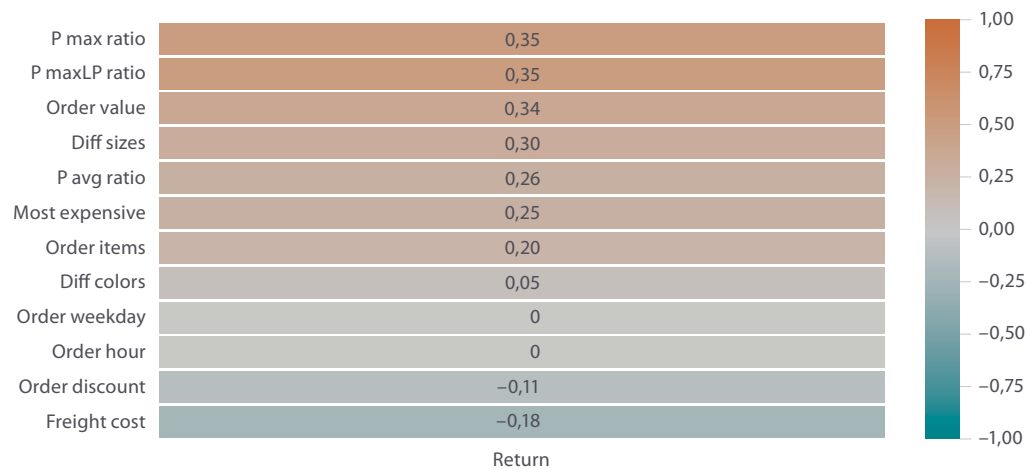
Tabell 3. Prediktiv prestanda för olika konfidensnivåer. Återkommande kunder.

Topp x%	#order	Uppskattad precision	Precision	Recall
1%	2 151	0,995	0,989	0,020
2%	4 301	0,993	0,988	0,041
5%	10 752	0,985	0,977	0,101
10%	21 504	0,972	0,966	0,199
15%	32 256	0,953	0,947	0,293
20%	43 008	0,925	0,919	0,379

För att bättre förstå hur enskilda attributvärden påverkar sannolikheten för en retur genomfördes en tvådelad analys. Dels beräknades korrelationerna mellan enskilda attributvärden och målvariabeln (retur eller inte) och dels extraherades modellens så kallade attributviktighet (feature importance). Här är det viktigt att förstå att även om båda dessa analyser är tänkta att besvara samma fråga, så görs det på helt olika sätt, och resultaten har därmed olika betydelse. När vi beräknar korrelationer så görs det direkt på datan, det vill säga utan att på något sätt beakta modellen. Som vanligt måste vi här komma ihåg att korrelation bara visar hur två variabler samvarierar, vi kan inte se det som att den ena variabelns värde orsakar den andras. Feature importance, å andra sidan, presenterar en bild av hur relativt viktiga de olika attributen var för modellen i att avgöra klassen, alltså här retur eller inte.

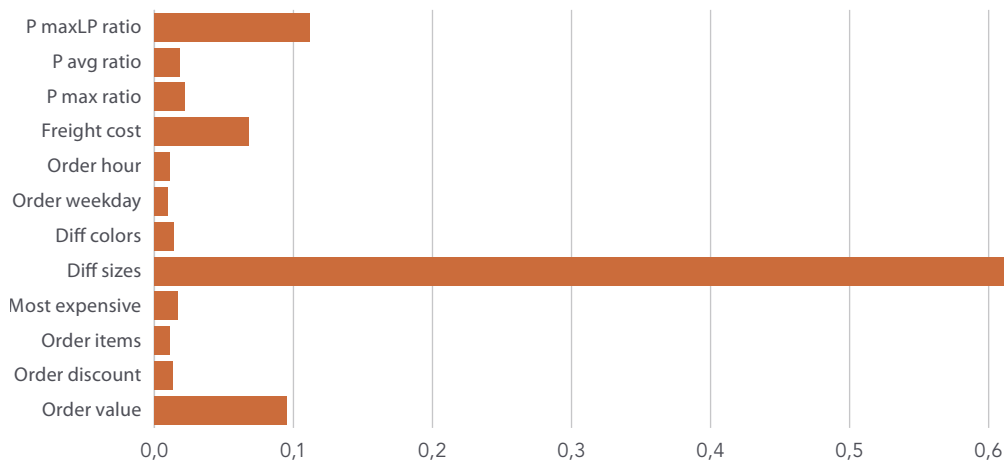
Figur 1 nedan visar de olika attributens korrelation med retur för förstagångskunder. De två attribut med starkast korrelation är båda kopplade till returhistoriken för de produkter som ingår i ordern. I övrigt ser vi väldigt väntade samband; ju fler eller dyrare produkter som ingår i ordern, desto större sannolikhet för att minst ett plagg returneras. Förekomsten av samma plagg i olika storlekar är en relativt stark indikation på en retur, vilket dock inte gäller för samma plagg i olika färger. Vi kan även notera att när ordern lades (veckodag eller tid) inte verkar påverka risken för en retur. Slutligen ser vi att rabatt och fraktkostnad båda har en svagt negativ korrelation, vilket då alltså betyder att ju högre rabatt och fraktkostnad, desto lägre risk för retur.

Figur 1. Korrelationer mellan attribut och målvariabeln det vill säga retur. Förstagångskunder.



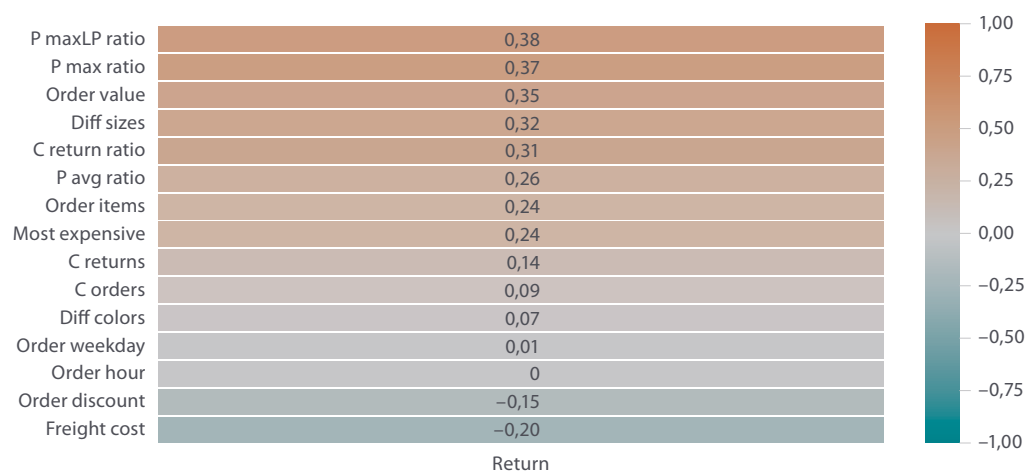
Figur 2 nedan visar modellens attributviktighet. Här framträder en delvis annan bild, för modellen är det klart viktigaste attributet huruvida ordern innehåller ett plagg i flera storlekar. Det näst viktigaste attributet handlar om ingående produkters returhistorik, men även orderns totala värde är viktigt för modellen.

Figur 2. Attributviktighet för algoritmen XGBoost vid prediktion av retur. Förstagångskunder.



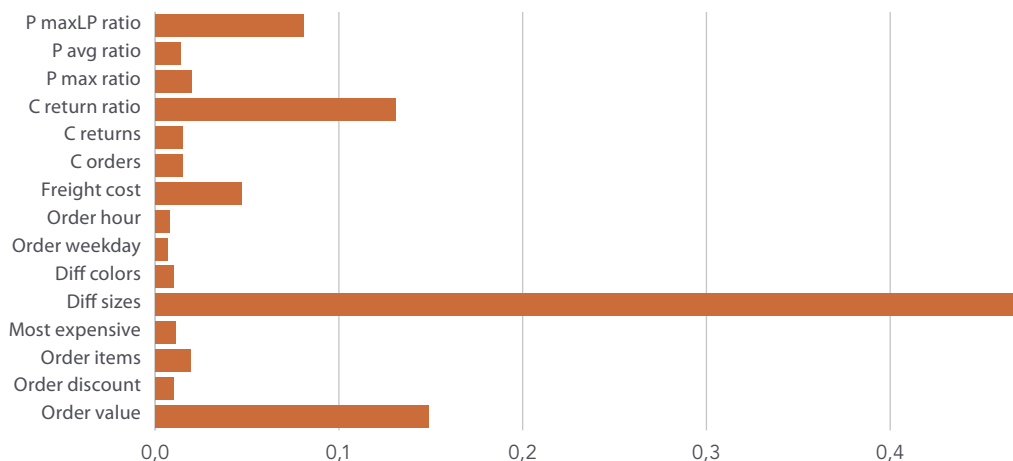
Om vi övergår till återkommande kunder ser vi i figur 3 nedan att kundernas historik har en tydlig betydelse. Korrelationen mellan en kunds historiska returandel och retur eller inte är 0,31. Även antalet tidigare retur, och kanske inte lika självklart, antalet tidigare order, är positivt korrelerade med risken för retur. I övrigt ser vi en ganska stor överensstämmelse med figur 1 ovan, vilket alltså indikerar att de valda attributen samvarierar med risken för retur på ungefär samma sätt för förstagångskunder och återkommande kunder.

Figur 3. Korrelationer mellan attribut och målvariabeln det vill säga retur. Återkommande kunder.



Figur 4 nedan visar slutligen modellens attributviktighet för återkommande kunder. Även här är mönstret med samma plagg olika storlekar det viktigaste, men inte lika dominerande som för förstagångskunder. I själva verket utnyttjar modellen alla typer av attribut det vill säga om ordern (Diff sizes och Order value), ingående produkters returhistorik (P maxLP ratio) och kundens returhistorik (C return ratio).

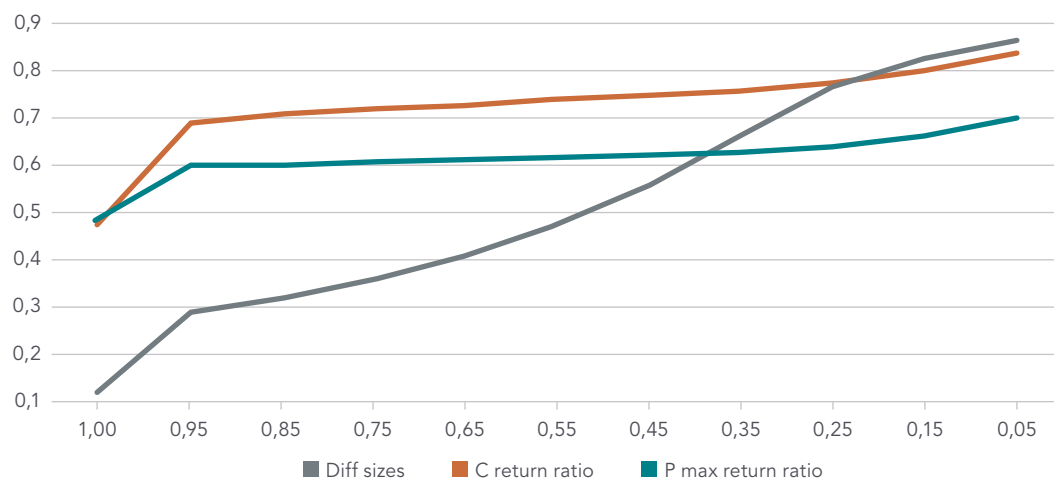
Figur 4. Attributviktighet för algoritmen XGBoost vid prediktion av retur. Återkommande kunder.



Vi avslutar resultatgenomgången med att betrakta figur 5 nedan. Här ser vi hur medelvärdet för tre viktiga attribut varierar med modellens predicerade konfidens för att en order är en retur. Mer konkret motsvarar x-axelns värden topp x procent. Alltså, för värdet 1,0, längst till vänster, ser vi vad som gäller för hela datamängden. Lite drygt 10 procent av alla order innehåller samma plagg i flera storlekar, en genomsnittlig kund har returnerat lite knappt hälften av alla order, samt i en order har den produkt som innan

returernas oftast gjort det i nästan hälften av fallen. Flyttar vi oss längre till höger, alltså vi tittar bara på de order som modellen blir allt säkrare på kommer att returneras, så ser vi att samtliga dessa värden går upp. Specifikt, för Topp 5 procent, innehåller fler än 85 procent av dessa order samma plagg i olika storlekar, medan kundens returgrad är ungefär 80 procent och ordern innehåller en produkt som innan returernas vid 65 procent av alla köp. En alltför förenklad slutsats skulle här kunna vara att det enda modellen gör är att upptäcka order som innehåller ett plagg i olika storlekar, vilket enligt tidigare visar på ett välkänt kundbeteende som driver returerna. I själva verket visar dock analysen snarare att det mönstret, som förväntat, visserligen är mycket viktigt, men att returhistorik hos såväl kunder som ingående produkter också har stor påverkan. En tydlig indikation på detta är att modellen, enligt genomgången ovan, uppenbart är kapabel att med mycket hög konfidens (högre än 95 procent) identifiera order som inte innehåller samma plagg i olika storlekar som returerna.

Figur 5. Viktiga attributs medelvärden vid olika konfidensnivåer.



4

Diskussion och slutsatser

Mål 1 och 2:

Den konsumentfokuserade delen av projektet har delvis haft som mål att undersöka i vilken mån konsumenter är benägna att välja hållbara produkter framför konventionella produkter. Viss forskning kring hållbarhet och marknadsföring har antytt att konsumenter skulle kunna vara benägna att välja hållbara produkter. Vi ville pröva detta genom att testa om AI-genererade (kontra genererade av en människa) produktrekommendationer skulle ge en starkare intention att köpa hållbara produkter framför konventionella. Hade så varit fallet hade det gett stöd för tanken att konsumenter, via AI, skulle kunna ”pushas” att göra mer hållbara val. Nu visade sig detta inte vara fallet och man kan därmed fundera över varför.



En möjlig förklaring skulle kunna vara att konsumenter inte har så tydliga preferenser för hållbara produkter som man trott, och att det därmed är andra faktorer som spelar större roll i deras konsumtionsval. Denna förklaring kunde vi dock utesluta då vi i experimenten frågade deltagarna om deras värderingar kring hållbarhet, och som visade på relativt starka värderingar. Det kan naturligtvis vara så att man ”i allmänhet” har starka värderingar kring hållbarhet, men att detta inte påverkar det specifika köpet man funderar över. Detta stöds av en omfattande forskning kring människors tendens att hysa vissa intentioner som man sedan inte följer upp i verklig handling – vi säger en sak och gör något annat. Det är också tänkbart att de produkter vi använde i experimenten inte var lämpliga för att visa på intentioner att köpa hållbart. Valet av hörlurar och ryggsäckar var baserat på produktkategorier som prövats i annan forskning och som kan anses ha hög validitet. I båda produktkategorierna var det relativt enkelt att skapa en tydlig skillnad mellan en hållbar och en konventionell i scenariot för experimentet. Trots det så såg vi ingen preferens för hållbara produkter.

En stor och viktig fråga när det gäller användningen av AI (både för hållbara och konventionella produkter) är i vilken utsträckning konsumenter är villiga att dela data om sig själva, och i vilken mån dom är benägna att tala gott om företag dom handlar hos. Mot bakgrund av tidigare forskning kring AI och marknadsföring vet vi att det finns en tydlig aversion mot AI som på olika sätt påverkar konsumenter i hur man förhåller sig till olika användningar av AI. Vi har framför allt tittat på AI-genererade produktrekommendationer där AI kommit till stor användning.

De experiment vi genomfört kopplade till mål 2 i projektet har genererat flera intressanta resultat som har stor relevans för företag inom digital handel. Inledningsvis kan vi

konstatera att benägenheten att vilja dela data inte skiljer sig åt beroende på om det är en hållbar eller konventionell produkt. Detta resultat var förväntat baserat på resultaten från experimenten under mål 1, men det var ändå viktigt att konstatera att resultatet var det väntade eftersom frågeställningen var annorlunda i detta experiment (det gällde konsumentens vilja att dela data).

Om vi i stället riktar fokus mot hur produktrekommendationer utformas så ser vi flera intressanta saker. Vi redogjorde ovan för ett experiment där AI via en röststyrd, smart högtalare gav en produktrekommendation som antingen var specifik (exakt samma produkt som två vänner talat om) eller allmän (samma varumärke men inte en specifik produkt). Intressant nog så fann vi att man var mer positiv till att dela data när det gällde en specifik produktrekommendation. Vi prövade också om specifika kontra allmänna rekommendationer påverkat konsumentens benägenhet att tala positivt om företaget. Även här gav specifika rekommendationer ett högre värde. Man kan tänka sig att det skulle kännas obehagligt att en smart högtalare (med hjälp av AI) fångar upp en konversation och ger en väldigt specifik rekommendation. Samtidigt kan det vara så att man upplever det som positivt att man får en rekommendation som så tydligt ansluter till något man önskar, att man är beredd att ge avkall på eventuella värderingar kring integritet.

För att nyansera denna bild så prövade vi också om det spelade någon roll om företaget man handlade från var ett stort eller ett litet företag för viljan att dela data. Vi kunde där konstatera att man i högre utsträckning kunde tänka sig att dela data med ett litet företag jämfört med ett stort.

Mål 3:

Avseende produktreturer är det viktigaste resultatet tveklöst att projektet har visat på en data-driven möjlighet att med tillräcklig träffsäkerhet förutsäga om en order kommer resultera i en retur eller inte. Givet kravet på att undvika falska positiva så långt det är möjligt, innebär den föreslagna lösningen att man bara agerar på en delmängd av de order som prediceras som returer.

Med hjälp av välkalibrerade maskininlärningsmodeller blir det möjligt för en e-handlare att i förväg bestämma sig för en kombination av acceptabel risk för en falsk positiv och motsvarande andel av returer som kommer fångas. För förstagångskunder kan en rimlig nivå vara att man accepterar att en returprediktion av tio är felaktig, och då fångar man (= agerar på) ungefär 25 procent av alla returer. På motsvarande sätt kan man för återkommande kunder kanske välja endera 95 procent precision med 30 procent recall eller 92 procent precision och 38 procent recall. Samtliga dessa alternativ kommer innebära att ett system i drift skulle rekommendera åtgärder ett antal gånger varje dag.

En intressant fråga, som projektet inte studerat, är vad en åtgärd skulle kunna vara. Självklart finns här ett antal alternativ; alltifrån någon form av mild nudging till att

ordern helt enkelt inte tillåts. En uppenbar möjlighet är så klart att inte låta returen vara gratis.

Beteendet att köpa samma plagg i olika storlekar är relativt vanligt, lite drygt 10 procent av alla order. Av dessa returneras de flesta. En bra början vore därför antagligen att ge sig på dessa order först. På motsvarande sätt finns ett fåtal kunder som nästan alltid returnerar. På produktnivå är det tydligt att vissa produkter returneras extremt ofta, så en djupare analys av varför bör vara intressant.

Vi ser det föreslagna systemet som ett mycket viktigt första steg. Innan ett praktiskt införande behöver ett stort antal beslut fattas. Även om projektet inte studerat detta kan det vara av värde att presentera några av författarnas åsikter kring detta. Först och främst kommer ett framgångsrikt system behöva vara helt automatiserat, det vill säga AI:n bör inte bara genomföra själva prediktionen utan också bestämma vilken eventuell åtgärd som ska tas. Här kommer beslutet att bero inte bara på prediktionen och dess konfidens utan även på faktorer som orderns storlek och kundens uppskattade värde. Systemet behöver därför ha möjlighet att i realtid inhämta den typen av information.

AI:n bör inte bara genomföra själva prediktionen utan också bestämma vilken eventuell åtgärd som ska tas.

Omvänt – om man för en viss order kan uppskatta kostnaden för en falsk positiv, och man har tillgång till korrekta sannolikheter från själva prediktionen, blir beslutet om åtgärd eller matematiskt enkelt. Om vi, som ett exempel, antar att vi för en viss order beräknat kostnaden för en falsk positiv till 200 SEK, och på motsvarande sätt vinsten för en korrekt tagen åtgärd till 50 SEK, så ska åtgärden förstås genomföras om algoritmens konfidens är högre än 0,8.

Rent tekniskt kan systemet designas så att det i stället för att genomföra prediktionen när kunden verkligen lagt ordern helt enkelt löpande analyserar kundkorgen och eventuellt agerar direkt. Rimliga åtgärder är då förstås av det mildare slaget, exempelvis lätt nudging.

Ett mycket intressant alternativ, som vi rekommenderar för fortsatt forskning, är att även i nästa steg utnyttja en datadriven ansats. Mer konkret skulle då en AI, över tid, själv kunna lära sig en optimal policy, omfattande när den ska agera och på vilket sätt. Med tillgång till prediktiva modeller av motsvarande kvalitet som projektets för själva returprediktionerna, ser vi framtagandet av ett sådant system som inte bara fullt möjligt, utan relativt okomplicerat. Detta skulle kunna kombineras med en konsumentorienterad approach där man också undersöker hur kunder skulle reagera på ett sådant automatiserat system. Integritetsfrågan torde vara central i ett sådant system.

Referenser

- Ackermann, K.A. med flera (2022). Willingness to share data: Contextual determinants of consumers' decisions to share private data with companies. *Journal of Consumer Behavior*, 21, 375–386.
- Brandimarte, L. med flera (2012). Mismatched Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science*, 4(3), 340–347.
- Castelo, N. med flera (2019). Task-Dependent algorithm aversion. *Journal of Marketing Research*, 56(6), 809–825.
- Cloarec, J. (2020). The personalization-privacy paradox in the attention economy. *Technological Forecasting & Social Change*, 161, 1–7.
- Dailey, L. C. och Ülkü, M. A. (2018). Retailers beware: On denied product returns and consumer behavior. *Journal of Business Research*, 86, 202–209.
- Dietvorst, B. J. med flera (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170.
- El Kihal, S. och Shehu, E. (2022). It's not only what they buy, it's also what they keep: Linking marketing instruments to product returns. *Journal of Retailing*, 2(29), 1–14.
- Formosa, P. (2021). Robot autonomy vs. Human autonomy: Social robots, artificial intelligence (AI), and the nature of autonomy. *Minds and Machines*, 31, 595–616.
- Hermann, E. (2022). Anthropomorphized artificial intelligence, attachment and consumer behavior. *Marketing Letters*, 33, 157–162.
- Hu, Y. och Min, H. (2023). The dark side of artificial intelligence in service: The “watching-eye” effect and privacy concerns. *International Journal of Hospitality Management*, 11, 1–9.
- Janakiraman, N. med flera (2016). The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic review. *Journal of Retailing*, 2, 226–235.
- Jussopow, E. med flera (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion, Research papers, 168.
https://aisel.aisnet.org/ecis2020_rp/168

Longoni, C. och Cian, L. (2020). Artificial intelligence in utilitarian vs. hedonic contexts: The “word-of-machine” effect. *Journal of Marketing*, 1–18.

Nichifor, E. med flera (2021). Artificial intelligence in electronic commerce: Basic chatbots and the consumer journey. *Amfiteatru Economic*, 23(56), 87–101.

Puntoni, S. med flera (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, 85(1), 131–51.

Reczek, R. W., Trudel, R. och White, K. (2018). Focusing on the forest or the trees: How abstract versus concrete construal level predicts responses to eco-friendly products. *Journal of Environmental Psychology*, 57, 87–98.

Rios, K., Finkelstein, S. R. och Landa, J. (2015). Is there a “fair” in fair-trade? Social dominance orientation influences perceptions of and preferences for fair-trade products. *Journal of Business Ethics*, 130(1), 171–180.

Song, M. med flera (2022). Will artificial intelligence replace human customer service? The impact of communication quality and privacy risks on adoption intention. *Journal of Retailing and Consumer Services*, 66, 1–17.

Sweidan, D. med flera (2020). Predicting returns in men’s fashion. I *Proceedings of the 14th international FLINS conference*, 1506–1513.

Urbanke, P., Kranz, J., Kolbe, L. (2015). Predicting product returns in ecommerce: the contribution of mahalanobis feature extraction. *Proceedings of the 36th International Conference on Information Systems*.

” Forskning för att stärka handelns konkurrenskraft och skapa goda villkor för branschens medarbetare.



Handelsrådet | 103 29 Stockholm
Besöksadress: Regeringsgatan 60
Telefon växel 010-471 85 80
www.handelsradet.se