



# Anbefaling av nyhetsinnhold i praksis

*Fra algoritmer til personaliserte nyheter*

## Recommending news content

*From algorithms to personalised news*

Rolf Dyrnes Svendsen

Leder, NxtMedia  
[rdsvends@gmail.com](mailto:rdsvends@gmail.com)

Jon Atle Gulla

Professor, Institutt for datateknologi og informatikk, NTNU  
[jon.atle.gulla@ntnu.no](mailto:jon.atle.gulla@ntnu.no)

Jørgen Frøland

Prosjektleder, Polaris Media ASA  
[jorgen.froland@polarismedia.no](mailto:jorgen.froland@polarismedia.no)

### SAMMENDRAG

Automatiserte anbefalinger av nyhetsinnhold brukes i dag på nettsidene til mange medieselskaper med hensikt å presentere leserne mer relevante nyheter og tilby bedre brukeropplevelser. Slike anbefalingsløsninger drar nytte av nye teknikker fra maskinlæring og stordataarkitekturer fra informatikkfaget. I denne artikkelen diskuterer vi om personlige nyhetsanbefalinger kan øke trafikken til nettaviser i en tid hvor mediebransjen er under press, samt bidra til å gi medieaktørene større brukerforståelse. I denne artikkelen belyser vi noen av utfordringene med nyhetsanbefalinger og diskuterer spesielt teknikker som kollaborativ filtrering og innholdsbaserte anbefalinger. Vi kommer også inn på evalueringsmetriker, fordi tradisjonelle mål på *presisjon* og *recall* ikke alltid samsvarer med mediehusenes behov for å benytte løsningene strategisk for å bedre lønnsomheten. Polaris Media har over flere år eksperimentert med anbefalingsteknologi på sine nettaviser, og vi presenterer i artikkelen funn og innsikter fra denne utprøvingen. Noen hovedfunn er at anbefalingsløsninger kan øke klikkraten og lesetiden betydelig, revitalisere eldre nyhetsstoff slik at innhold får lengre levetid samt bidra til å forstå brukersegmentene bedre. Artikkelen posisjonerer seg innenfor en *design science research*-tilnærming, hvor målsetningen er å forstå industriens behov og utvikle og prøve ut artefakter som møter disse behovene.

### Nøkkelord

nyheter, nyhetsanbefaling, anbefalingssystemer, algoritmer, kunstig intelligens, maskinlæring, stordata, personalisering, personvern, språkteknologi, stordata, forretningsmodell

## ABSTRACT

Automatic recommendation of news content is now used by many online media houses. The intention is to present more relevant news stories and give the readers a more satisfying user experience, and the solutions make use of new techniques from machine learning and big data architectures. The central question is whether personal recommendations can increase traffic to online news sites and also contribute to increased user understandings. In this paper, we explain some of the challenges of news recommendation and discuss the different techniques for collaborative filtering and content-based recommendation. We briefly present evaluation metrics, since traditional measures such as precision and recall are not always consistent with the media houses' need to use the solutions to improve their profitability. Polaris Media has over some years experimented with this technology on their news sites. Their experiments confirm that it can increase click rates and reading times substantially, help us understand user segments, and revitalize older news content. The article positions itself within a design science research approach in which the aim is both to understand industry's needs and to develop and test artifacts that meet those needs.

### Key words

critical algorithm studies, machine vision, aesthetic inference, Instagram, recommendation algorithms, visual algorithms, technological determinism, non-conscious cognition

## 1. INTRODUKSJON

Personaliserte nyheter tilbys i dag på nettsidene til mange medieselskaper. Løsningene understøttes av en egen teknologi, såkalte *anbefalingssystemer*, som analyserer brukeratferd og som prøver å estimere brukernes interesse for nyhetsartikler de ikke har lest. Nyhetsanbefaling kan gjøre det lettere for leserne å finne relevante nyheter i et mylder av stadig skiftende nyhetsstoff av varierende interesse. Samtidig er det en kjensgjerning at mediehusene sliter med tapte annonseinntekter og salg, slik at det er en målsetning å benytte anbefalingssystemer også som et strategisk grep for verdiskapning. Kommersielle aktører som er mest avhengige av annonseinntekter, fokuserer typisk på mye trafikk. Aktører som har abonnementsinntekter som hovedinntektskilde, vil gjerne ha strategier som forlenger abonnementene og som skaffer nye. Offentlige aktører som NRK og BBC er derimot mindre avhengige av lønnsomhetsvurderinger, og benytter derfor anbefalingssystemer som underbygger samfunnskravet om god kunnskapsformidling. BBC, for eksempel, benytter en «Public Service Algorithm» der tanken er at brukerne over tid best tilfredsstilles ved artikler som oppfyller samfunnsoppdraget om god kunnskapsformidling, formulert som følger:

$$\text{Retention} = f(\text{Diversity, Novelty, Surprisal, Click Through Rate})$$

For kommersielle aktører vil suksess gjennom personaliseringsteknologi være avhengig av å balansere samfunnsansvar og krav til lønnsomhet. Den siste sikrer uavhengighet, den første troverdighet. Samtidig har den raske, teknologidrevne utviklingen åpnet for kritikk fra flere hold som hevder at anbefalingssystemene i liten grad tar hensyn til filterbobler, og at løsningene kan bidra til å gi et mer fragmentert og tabloidisert nyhetsbilde. Det foreligger lite innsikt i hvordan anbefalingsløsninger faktisk er teknisk skrudd sammen, og hvordan de fungerer i praksis fra medieaktørens perspektiv, og ikke kun fra forbrukerens. Denne artikkelen er et bidrag som adresserer dette kunnskapsgapet ved å illustrere nettopp

ulike teknikker og anbefalingsmetoder som brukes i praksis av mediehusene i Norge for å tilby personaliserte nyheter. Mer spesifikt adresserer vi følgende problemstilling i denne artikkelen:

- Hvordan kan anbefalingsløsninger bidra til å tilby mer relevante nyheter til leseren på den ene siden, og økt lønnsomhet for mediehus på den andre?

Vi vil i artikkelen svare på problemstillingen ved å presentere ulike teknikker som har blitt prøvd ut hos Polaris Media med hensikt om å tilby mer relevant innhold til leseren på den ene siden, og øke lønnsomheten til mediehuset på den andre. Hensikten med artikkelen er først og fremst å gi en enhetlig presentasjon og kritisk diskusjon av teknologiene bak nyhetsanbefaling. Artikkelen viser hvordan ulike komponenter fra maskinlæring, språkteknologi og datagruvedrift samspiller i et rammeverk, samt hvilke alternativer som finnes for kritiske deler av anbefalingssystemene. Utgangspunktet for artikkelen følger en *design science research* (DSR)-tilnærming, som er kjennetegnet av en metodikk som forsøker å forstå hva industriens behov er, samtidig som man henter teoretisk fundament fra akademiske kilder (Grønli, 2017). Sentralt i DSR er utviklingen av artefakter som møter et behov i industrien, som deretter testes, utprøves og korrigeres fortløpende. Rammeverket som presenteres i artikkelen, forklarer funksjonaliteten til komponentene og viser hvordan menneskeskaptede komponenter til sammen skaper en tjeneste som adresserer brukerbehov, og som kan evalueres i henhold til definerte metrikker. Vi går ikke inn på evalueringer av hver enkelt komponent i rammeverket, da disse komponentene igjen kan implementeres på en rekke forskjellige måter med både sine styrker og svakheter. Det er altså rammeverket – selve arkitekturen bak nyhetsanbefalingssystemer – som former diskusjonen i denne artikkelen, snarere enn noen underliggende teori eller metode. I tillegg viser vi gjennom konkrete prosjekter fra Polaris Media hvordan løsningene har vært tatt i bruk, evaluerer resultatene mot reelle forretningsmål og aktivt tilpasser teknologien til mediehusenes behov og strategier.

Resten av artikkelen er organisert som følger: Først presenterer vi kort Polaris Media og hvordan prosjektet og erfaringene som denne artikkelen viser til, kom til. Deretter diskuterer vi personvernsaspekter, hva som kjennetegner gode anbefalingssystemer og noen av hovedutfordringene ved nyhetsanbefalinger. Dernest gjengir vi funn og resultater fra tre tester gjennomført hos Polaris Media, før vi til slutt konkluderer.

## 2. POLARIS MEDIA OG METODE

Polaris Media ASA er et norsk børsnotert mediekonsern som består av 35 lokale og regionale mediehus hvorav Adresseavisen, Sunnmørsposten og Romsdal Budstikke er de største. Selskapet ble etablert i 2008 gjennom en fusjon av Adresseavisen Gruppen og Harstad Tidende-gruppen. Schibsted ASA er største aksjonær med 29 % av aksjene.

Polaris Media begynte tidlig med å eksperimentere på egen hånd med anbefalingsteknologi, både på sine store dagsaviser og mindre lokalaviser. Adresseavisen, med sine 60 000 abonnenter og rundt 300 000 artikkelvisninger om dagen, publiserer selv 60–70 daglige artikler og tar i tillegg inn rundt 200–300 artikler hver dag fra eksterne partnere som for

eksempel NTB. Dette er små tall sammenlignet med Google News som tilbyr nyheter på 35 språk, aggregerer nyheter fra mer enn 25 000 publisister og som har nesten 500 millioner brukere hver måned basert på bl.a. maskinlæringsanalyser (Liu et al., 2010). Dette viser at teknologien er skalerbar og takler både et mer dynamisk nyhetsbilde og mer krevende og korttidsfokuserede brukere.

Fra 2011 ble samarbeid med forskningsmiljøer og konkurrerende mediehus etablert. Med regionale utviklingsmidler ble et nettverk av mediehus i Norge og forskere fra Institutt for datateknologi og informatikk (IDI) ved NTNU koplet sammen i et prosjekt for å kartlegge aktuelle anbefalingssystemer og deres anvendelse i mediesammenheng. Dialogen med akademiske fagmiljøer bidro til økt kunnskap hos Polaris Media og at forsøkene og utprøvingene av anbefalingssystemene ble mer systematiske og hypotesedrevet. Arbeidet gikk fra småskalatester med redaksjonelle analyse- og intuitivt drevet observasjonsarbeid, til mer formell tilnærming. Storskalabruk av anbefalingsteknologiene ble innledet i 2014 ved opprettelse av et eget Big Data-team i Polaris Media med ansvar for stordatasatsing innen ulike fagområder i konsernet. Med bruk av iterative utviklingsprosesser dokumenterte teamet utprøvingene av personaliseringsteknologi gjennom problematisering, eksperimentering og evaluering/presentering. I en smidig utviklingsprosess utvikles en modell for datamaskinens valg og overvåking av kvaliteten på disse valgene. Modellen justeres deretter av teamdeltakerne underveis i retning ønsket utfall, ikke ulikt fra hvordan maskinlæringsløsninger optimaliseres.

De første forsøkene foregikk i tilrettelagte omgivelser for eksperimenter som nettavdeling og konsernets Medielab. Adresseavisen, som er konsernets digitale senter, begynte å legge tekstlenker inntil en artikkel til andre artikler som kunne være interessante for leseren.

Tanken var at lenkene skulle føre til økt oppholdstid på nettstedet, gi flere sidevisninger og styrke brukerlojaliteten. Anbefalingene ble etterhvert redaksjonelt utstyrt rikere med bilder og plassert i designede bokser. Eksperimentene fortsatte ved at anbefalingene blandet både kontekstuelle lenker med lenker til saker som var mye lest (trendet). Ytterligere et steg ble tatt da løsningene tok hensyn til hvilken plattform brukeren kom fra. Tester og evaluering av ulike elementer fra anbefalingsteknologiene ble gjennomført ved hjelp redaksjonelle arbeidsmetoder – det vil si gjennom utprøving direkte i produktene med etterfølgende analyser av trafikk tall og diskusjoner med fagpersoner innen selskapets organiserte forsøksvirksomhet.

Anbefalingslenkene hadde på tidspunktet 2011–2012 kontekstuell relevans og var knyttet til lesersakene uavhengig av data om brukerne som først ble muliggjort senere gjennom mobile løsninger.

Ulike hypoteser ble testet av teamet med hensikt å øke relevansen hos brukermassen ved ulike konfigurasjoner av utprøvningsmodeller. Tre slike tester var:

1. testing av anbefalinger mot brukergrupper hos Adresseavisen der effekt mot ulike brukergrupper var målet
2. testing av anbefalinger på hele mobilforsiden hos iTromsø der identitet av brukere gjennom pålogging via smarttelefoner muliggjorde individuelle nettversjoner
3. testing av anbefalinger over ulike tidsperioder hos Altaposten der man undersøkte *long tail*-effekt med anbefalinger av eldre, relevant materiale i Alta

Det er resultater fra disse tre testene vi presenterer i denne artikkelen. Eksperimentene var direkte utledet av forretningsmål om å posisjonere nettstedet som ledende i sine nedslagsfelter i konkurranse med andre aktører kombinert med mulighetene til økt kunnskap om målgruppene. Det ble utviklet en prototype i de tre testene som ble utprøvd over lengre perioder for å korrigere for blant annet variasjon i nyhetsbildet. Innsamlede data var anonyme adferds-data. De ble brukt til å analysere og identifisere effekter av prototypen. Datainnsamlingen skjedde i leverandørens systemer. Dataanalyse foregikk i deres verktøy og i lokale data-uttrekk for ytterligere analyse. Disse ble deretter sammenstilt i en presentasjonsvennlig form for å gjøres tilgjengelig for resten av Polaris Media. Underveis i utprøvingene ble et femårig forskningsprosjekt om anbefalingsteknologi startet opp i 2015 i samarbeid med NTNU fra akademia.

Redaksjonelle valg og anbefalinger overfor brukerne har tradisjonelt vært basert på journalistisk fagkunnskap, skjønn og erfaring. Datamaskiner har derimot andre egenskaper, som for eksempel lagringskapasitet, regnekraft og koordineringshastighet (såfremt den har et regelsett å forholde seg til). Etablert journalistisk praksis og datamaskinens egenskaper er på denne måten ganske forskjellig, og det var gjentatte diskusjoner om redaksjonell styring av personaliseringsteknologi i mediehuset. Diskusjonene ble ført på flere nivåer i organisasjonen, både på topp- og i medarbeiderfora, men hele tiden forankret på toppnivå. Samarbeidet med akademia gav i tillegg satsningen og prosjektet troverdighet og legitimitet internt hos Polaris Media.

Vi mener at det etterhvert ble en bevissthet internt i organisasjonen om at samspillet mellom menneske og datamaskin ville forbedre produktene. Nyhetsanbefalinger bør, optimalt sett, samle og anvende alle tilgjengelige datapunkter og tilpasse brukeropplevelsen i sanntid til den enkelte brukers situasjon. Dette er en oppgave som ikke lar seg løse uten maskinell assistanse. Individuelle tilpasninger av nyhetsprodukter må derfor benytte automatiserte løsninger, hvor løsningen hos Polaris Media ble en digital nyhetsopplevelse som kombinerer automatiserte beslutninger med redaksjonelle vurderinger. Mulighetene som gis i dag, er flere: En er å sette av spesifikke plasseringer som styres av henholdsvis automatiserte og redaksjonelle anbefalinger som kan sammenlignes. En annen er å justere modellen med redaksjonelle vurderinger, men der modellen brukes i sin helhet. En tredje er en kombinasjon av de foregående hvor modellen tvinges til å ta beslutninger på konkrete regler i spesifikke sammenhenger.

### 3. STORE DATA OG PERSONVERN

I prinsippet er all informasjon om brukerne interessant fra et mediehusperspektiv innenfor personvernets reguleringer. Medienes interesse for data om sine brukere er knyttet til deres samfunnsrolle, oppgaver og økonomisk bærekraft for å sikre uavhengighet. Samfunnsrollen er definert av retten til å drive nyhetsformidling, journalistiske undersøkelser og debatt om samfunnsforhold, ofte koblet til service- og underholdningstjenester. Tidligere hadde mediene personlig data om sine brukere tilgjengelig gjennom frivillige abonnementsavtaler. Bruk av meningsmålinger gav også mye data om brukernes interesse, deres bruk av innhold og tjenester. Digitalisering har på sin side radikalt økt datamengden om brukere og bruken av innhold og tjenester.

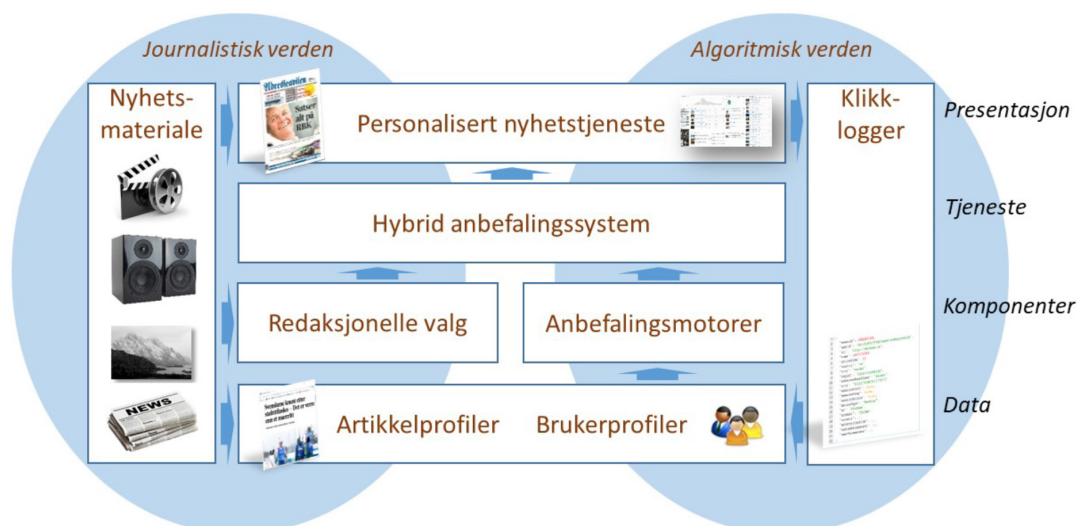
Innføringen av *GDPR (General Data Protection Regulation)* i EU har gitt personvern større betydning enn tidligere. Loven nekter innsamling av persondata uten eksplisitt samtykke. Dersom dette ikke overholdes, kan publisistens handlingsrom begrenses og føre til mer generaliserte nyhetsprodukter fremfor forslag fra anbefalingssystemer. Undersøkelser fra 2017 indikerer at brukerne er skeptiske til innholdsleverandørenes bruk av persondata (Mohallick, 2018). Til tross for at 78,7 % av brukerne anså seg som hyppige brukere av anbefalingsløsninger, trodde bare 9,0 % at innholdsleverandørene respekterte lover og regler for personvern. 33,6 % mente personvernlovene ikke blir overholdt (Craner et al., 2000; Olsen et al., 2005). Samtidig mente 63 % at de ville være mindre bekymret for personvernbrudd dersom de var i stand til å inspisere, modifisere og slette egne brukerprofiler. Åpenhet er derfor en nøkkel for å øke tilliten mellom innholdskonsumenter og innholdsleverandører.

#### 4. HVA KJENNETEGNER ET GODT ANBEFALINGSSYSTEM?

Tradisjonelt vurderes anbefalingssystemer etter hvor nøyaktig anbefalingene treffer brukernes dokumenterte interesser. Dette forutsetter at det foreligger et testdatasett som kan brukes til å sammenlikne faktiske brukerevalueringer (eller artikkelvisninger) med anbefalingene gjort av anbefalingsmotoren. En *evalueringsmetrikk* spesifiserer en framgangsmåte og en rekke kriterier for å evaluere et datasystem. For nyhetsanbefaling er det vanlig å bruke *presisjon (precision)* for å vurdere kvaliteten til systemet. Presisjon viser til andelen av de anbefalte artiklene som viste seg faktisk å være relevante for brukeren. En annen metrikk er *recall*, som angir andelen av de relevante artiklene i settet som ble anbefalt av systemet. Dersom det foreligger mer eksplisitte og finmaskete scorer fra brukeren, kan metrikkene *Root-Mean-Square Error (RMSE)* eller *Mean-Square Error (MSE)* benyttes.

For et mediehus er det derimot flere metrikker enn presisjon og RMSE som er nyttige. I praksis brukes ofte såkalte AB-tester med *klikkrater* for å overvåke hvor mange artikler en gjennomsnittlig leser kikker på under en leseøkt. Høye klikkrater betyr at leseren presenteres for mange andre relevante artikler som han også klikker på før vedkommende avslutter leseøkten. Et annet alternativ er *oppholdstid* eller *lesetid*, dvs. tiden leseren bruker på å lese en artikkel eller tilbringe på avisens forside. Antakelsen er at brukeren benytter lengre tid på å lese en artikkel som fanger interesse. Sentralt i denne sammenheng er også at metrikkene fanger opp aspekter ved brukeratferden som kan ha innflytelse på avisens lønnsomhet. For eksempel er det ikke urimelig å anta at høye klikkrater og lang lesetid resulterer i flere annonseklikk og større sannsynlighet for konvertering til abonnenter. Verdt å merke er at mange mediehus også har andre metrikker som er mer indirekte relatert til lønnsomhet og abonnementsalg. Det kan være et ønske om en viss variasjon i de anbefalte nyhetene (*diversity*), et visst innslag av helt nye temaer (*novelty*) eller at så mye som mulig av nyhetskatalogen tas i bruk (*catalog coverage*) (Shani & Gunawardana, 2011, Vargas & Castells, 2011).

Kvaliteten av den personaliserte nyhetsløsningen er likevel bare til en viss grad gitt av kvaliteten på selve anbefalingsmotorene. Et aspekt er at den integrerte, hybride nyhetstjeneren ofte kombinerer flere anbefalingsmotorer i tillegg til innslag av redaksjonelle anbefalinger, som vist i arkitekturen i Figur 1.



**Figur 1:** Lagdelt arkitektur for nyhetspersonalisering.

Et annet aspekt er presentasjonslag hvor layout, farger, illustrasjoner og liknende benyttes for å gjøre anbefalingene så attraktive som mulig for leserne. De senere årene er det publisert flere arbeider som tar for seg hele brukeropplevelsen av anbefalingssystemer (f.eks. Knijnenburg et al., 2012; Zins et al., 2011).

## 5. ANBEFALINGSTEKNOLOGI OG NYHETSINNHOOLD

Anbefalingssystemer bruker maskinlæring for å finne mønstre i store datasett. Anbefalingssystemer har blitt viktige i takt med eksplosjonen av informasjon på nettet, der behovet for å prosessere store mengder data i sanntid har blitt sentralt. Anbefalingssystemer har siden 90-tallet framstått som en egen disiplin, nært beslektet med informasjonsfiltrering og søkesystemer (Belkin & Croft, 1992). Grunntanken er å definere ulike teknikker for å spå – eller predikere – brukerreaksjonen på et gitt sett av alternativer, basert på informasjon om alternativene, brukerne og deres avhengigheter. Disiplinen bygger på forskning fra kognitiv vitenskap, informasjonsgjenfinning, beslutnings- og ledelsesteorier samt fra semantisk web og datautvinning i senere år (Borges & Lorena, 2010). Nyhetsdomenet anses som ett av de vanskeligste domenene for anbefalingssystemer fordi nyhetsartikler i hovedsak er fritekst, og det er stor variasjon i hvordan verden beskrives i naturlig språk. Det finnes ingen åpenbar, maskinell metode for å trekke ut meningsinnholdet av en tekst, slik at det er nødvendig å finne andre tilnærminger som fanger noen viktige aspekter av innholdet. Sentralt her er konseptet «entitet», som er et *objekt av en bestemt type (f.eks. person) som refereres i teksten ved bruk av egennavn, pronomener og andre elementer*.<sup>1</sup> Før nyhetsartikler legges inn i anbefalingssystemet, tar man ut stoppord og spesialtegn, norma-

1. Ifølge *Store Norske Leksikon* brukes entiteter om noe som er eller har vært i vid forstand, og kan brukes om noe i naturen eller noe som er menneskeskapt. Det kan for eksempel være en ting, en hendelse, et konsept eller et faktum. En entitet er noe selvstendig og entydig, og det kan gis en entydig identifikator. I anbefalingssystemer forenkles dette noe ved at en antar at egennavn og grammatiske strukturer som refererer til egennavn, utgjør de relevante referansene til entitetene i teksten.

liserer ord og bruker ulike statistiske teknikker for å avdekke de mer prominente ordene i teksten. Det er vanlig å trekke ut entiteter, for eksempel personnavn som «*Lionel Messi*», lokasjoner som «*Barcelona*» eller organisasjoner som «*Samsung*». Tanken er at slike entiteter kan utgjøre et diskriminerende sammendrag av artiklenes semantiske innhold. Samtidig er noen artikler vanskelige å forholde seg til fordi de er korte, tvetydige eller inkonsistente, eller det kan komme senere artikler som gjør de publiserte nyhetene utdaterte eller direkte feilaktige. Mye av innholdet i nyhetsartikler må også forstås ut fra en underforstått kontekst som datamaskinen ikke har tilgang til. Entiteten «*Rosenborg*» kan for eksempel i en artikkel referere til et fotballag, mens det i andre artikler er snakk om en bydel eller en bedrift (Ingvaldsen & Gulla, 2015).

En analyse av trafikken hos Adresseavisen over 7 dager i februar 2017 viste at en artikkel i gjennomsnitt hadde en lengde på bare 519 ord. Av til sammen 923 artikler var 43 % rene nyhetsartikler, som i snitt inneholdt 5,9 entiteter. Sport- og kulturartikler hadde henholdsvis 17 % og 10 % av trafikken. Noen kategorier var preget av lengre tekster som kan være lettere å analysere innholdsmessig. Ferie- og kulturartikler, for eksempel, inneholdt rundt 11 entiteter hver som gir en litt bedre indikasjon på artiklenes innhold, enn de 5–6 entitetene en finner i nyhetssaker (Gulla et al., 2017).

Tolkningen av nyhetstekstene er likevel bare en av mange utfordringer som må håndteres av anbefalingssystemer for medieinnhold (Karimi et al., 2018; Özgöbek, 2014). Andre utfordringer er kaldstart, kort levetid, implisitte evalueringer, varierende og flyktige interesser, serendipitet og hullede matriser. Vi vil nå presentere disse utfordringene kort, før vi diskuterer i neste avsnitt den underliggende tekniske praksisen for to varianter av anbefalingssystemer, mer spesifikt kollaborativ filtrering og innholdsbaserte anbefalinger, samt hybrider av disse.

### 5.1. Kaldstart

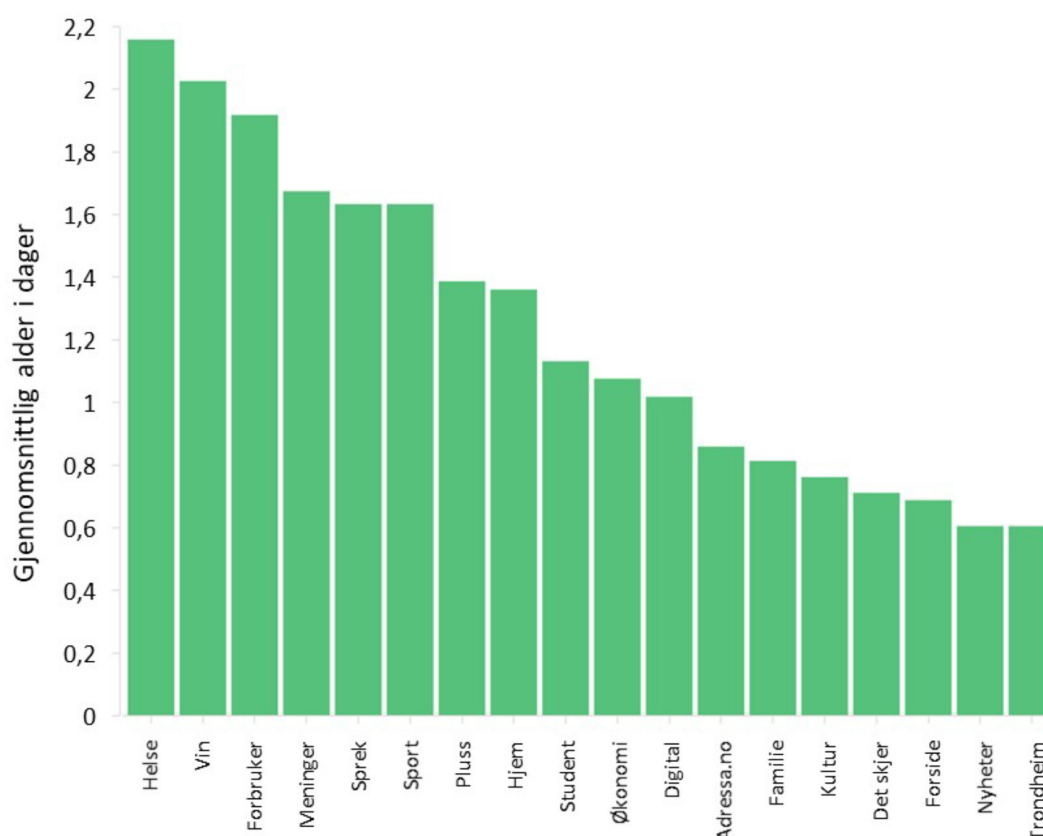
I nyhetsdomenet oppstår ofte situasjoner der systemet skal anbefale publiserte nyheter til nye lesere, eller prøve å anbefale nylige publiserte nyheter til eksisterende lesere. Dette er krevende fordi de fleste anbefalingsalgoritmene bruker historikk fra leserne og artiklene til å generere nye anbefalinger. Uten en historikk vites det ikke hva brukeren er interessert i, og det er derfor også vanskelig å gruppere brukeren sammen med andre liknende brukere eller artikkelen sammen med liknende artikler.

### 5.2. Kort levetid

En annen utfordring er at nyhetsartikler har langt kortere levetid enn for eksempel bøker og filmer. Ordinære nyhetsartikler om ulike hendelser som har skjedd, leses vanligvis i løpet av de to første dagene etter publisering, til tross for at levetiden varierer en del mellom aviser og nyhetskategoriene. En sammenligning av nettrafikken hos fire norske aviser fra 3 måneder i 2014, viste at gjennomsnittlig levetid for en avisartikkel kan variere fra rundt én dag til 2–3 dager (Gulla et al., 2016). Levetid betegner tiden det tok fra artikkelen ble publisert til en leser har lest den. En artikkels levetid (alder) er gjennomsnittet for alle leserne som har kikket på artikkelen. Figur 2. som viser levetiden per innholdskategori for



Adresseavisen hentet fra samme undersøkelse, viser at det er viktig også å ta hensyn til artikkelens tematiske kategori. Mens nyhetsartikler om trafikk, små dagsaktuelle begivenheter og saker med trivielt innhold hadde en gjennomsnittlig levetid på 0,6 dager, hadde saker med et dypere temamessig innhold, som for eksempel helserelaterte artikler, en levetid på 2,2 dager (Figur 2.):



**Figur 2:** Levetiden til ulike nyhetskategorier i Adresseavisen.

### 5.3. Implisitte evalueringer

Anbefalingssystemer vil gjerne ha eksplisitte brukerevalueringer av tidligere leste artikler for å generere nye anbefalinger, men lesere har vist lite interesse for å angi en score på hvor interessant en artikkel er (Lerche, 2016). Løsningen på denne utfordringen har typisk vært å utlede enkle scorere på bakgrunn av brukerhistorikk. Tidlige systemer brukte artikkelklikk som indikasjon på interesse, men noen klikker ved uhell og andre forlater artikkelen når de ser at innholdet ikke var som forventet (Epure et al., 2017). Mange nyere anbefalingsløsninger bruker en variant av lesetid for å utlede leserens interesse for artikkelen (De Pessemier et al., 2015, Gulla et al., 2014). Ideen er at en leser trolig har interesse for artikkeltemaet dersom vedkommende tok seg tid til å lese artikkelen grundig og til siste slutt.

#### 5.4. Varierende og flyktige interesser

Samtidig har nyhetslesere ofte mindre stabile interesseprofiler enn for eksempel boklesere eller film- og musikkinteresserte (Billsus & Pazzani, 1999; Liu et al., 2011). Leserens interesser kan variere mellom sesonger og kan knyttes til kontekstuelle faktorer som tid på døgnet, været eller om leseren befinner seg hjemme eller på kontoret. Nyhetsanbefalings-systemer finner det derfor ofte nødvendig å operere med to profiler for brukeren; en lang-tidsprofil som lagrer generelle interesser innenfor sport, kultur, og liknende, og en kort-tidsprofil som ser på nyhetsbildet akkurat nå (se for eksempel Li et al., 2011).

#### 5.5. Serendipitet

En annen utfordring for anbefalingssystemer er *serendipitet*, som betyr «heldig oppdagelse». Begrepet viser til graden av overraskelse i vellykkede anbefalinger. Utgangspunktet er at mediehus ønsker å bedre brukeropplevelsen ved å inkludere nyheter som ikke er forventet eller som er åpenbart interessante, men som likevel fanger leserens interesse. Ettersom anbefalingsstrategiene baserer seg på brukerhistorikk, er det ikke trivielt å trekke inn relevante nyheter som ikke likner på noe av det vedkommende har lest før (Kotkov et al., 2016).

#### 5.6. Hullete matriser

Den siste utfordringen vi skal nevne, er hullete matriser. Som vi skal se nedenfor, er det brukernes tidligere evalueringer av artikler som danner grunnlaget for å anbefale nye artikler. I nyhetsdomenet er det mange flyktige lesere, og det kommer inn flere tusen nye artikler hver eneste dag. Det betyr at én bestemt bruker har lest veldig få av de tilgjengelige artiklene, og at en gjennomsnittsartikkel er lest av veldig få brukere. Det er med andre ord mange hull i matrisen som skal beregne anbefalinger, ofte bare 0,1–2 % av feltene som har en evaluering, noe som er utfordrende for maskinlæringsteknikkene (Papagelis et al., 2005).

Med disse utfordringene i bakhodet vil vi nå presentere to hovedkategorier av anbefalingsstrategier som er hyppig brukt, og som vi har prøvd ut hos Polaris Media: kollaborativ filtrering og innholdsbaserte anbefalinger.

## 6. TO HOVEDKATEGORIER AV ANBEFALINGSSTRATEGIER

Nyhetsanbefaling handler om å estimere og rangere evaluering av artikler som er ukjente for brukeren. For å beregne slike estimater trekkes evalueringer av andre artikler fra samme bruker eller evalueringer fra andre brukere med liknende interesser inn. Utgangspunktet er altså en matrise med brukere i en dimensjon og nyhetsartikler i en annen. Elementene i matrisen er brukernes evalueringer av artikler og kan for eksempel representeres ved en score mellom verdiene 1 og 5. Et tomt element i matrisen betyr at denne brukeren ikke har angitt noen score for denne artikkelen. Vi kan se for oss at et anbefalingssystem er en preferansefunksjon som prøver å fylle ut alle tomme elementer i matrisen med estimerte score. For en bestemt bruker rangerer en alle estimater, slik at de høyeste estimatene

kommer først og utgjør anbefalingene av artikler fra systemet. For å få dette til må systemet dra nytte av alle andre scorere i systemet, i tillegg til annen innsikt som foreligger om artiklene i form av titler, kategorier, publiseringsdatoer, publisister, entiteter og liknende. Ett perspektiv på anbefalingssystemene er altså utviklingen av teknikker for å *komplettere en delvis utfylt matrise* med estimerte evalueringer. I de fleste tilfellene er vi imidlertid ikke avhengige av eksakte estimater på evalueringene. Det vi trenger, er relative estimater som gjør det mulig å rangere de mest relevante artiklene for en bruker. Dette er *rangeringsperspektivet* på anbefalingssystemer.

Det er som nevnt to hovedkategorier av anbefalingsstrategier; kollaborativ filtrering og innholdsbaserte anbefalinger, samt hybride tilnærminger som kombinerer de to. I tillegg brukes ofte noen enklere strategier som ikke er avhengige av noen lagret brukerprofil, men som drar nytte av kortvarig sesjonsinformasjon eller generelle popularitetsmål eller trender. Selve maskineriet bygger på *vektorrommodellen* og ulike likhetsfunksjoner samt hele batteriet av maskinlæringsteknikker. En vektorrommodell er en modell som brukes til å representere tekstdokumenter og andre objekter som baserer seg på tekstdokumenter. Vektorene inneholder et bestemt antall elementer, og for hvert element lagres en vekt som angir hvor viktig dette elementet er i vektoren. I et søkesystem er antall elementer typisk gitt av hele vokabularet som brukes i dokumentene. Anbefalingssystemer bruker ofte kortere vektorer ved bare å inkludere gjenkjente entiteter og spesielle fraser.

### 6.1. Kollaborativ filtrering

Tanken bak kollaborativ filtrering av nyhetsartikler er at lesere med de samme interessene eller preferansene ikke nødvendigvis har lest de samme artiklene. Det prøves først derfor å kartlegge hvilke lesere som har sammenfallende interesser. Deretter anbefales artikler som leseren ikke har lest tidligere, men som har blitt godt evaluert (eller bare lest) av andre lesere med samme interesser. Selve innholdet av artiklene ignoreres av denne strategien. Metoden kan beskrives som følgende:

Kollaborativ filtrering er en type anbefalingsteknikk som predikerer en brukers interesse for en artikkel utelukkende basert på denne brukerens likhet med andre brukere og disse liknende brukernes score for denne artikkelen.

Ta for eksempel matrisen i Figur 3 nedenfor, som viser hva fem lesere synes om seks nyhetsartikler. Mens slike matriser generelt inneholder en score som angir hvor godt en bruker liker et produkt, er eksplisitte scorere sjelden tilgjengelig for nyhetsdomenet. En må derfor gjøre antakelser ut fra leserens oppførsel, for eksempel ved å bruke normalisert lesetid som en indikasjon på hvor godt en leser likte en artikkel. For enkelhets skyld har vi i Figur 2 bare brukt binærverdier. Hvis den normaliserte lesetiden er over en viss terskel, for eksempel 20 sekunder pr. 100 ord, antar vi at leseren likte artikkelen og setter verdien 1 inn i matrisen. En verdi på 0 betyr at leseren ikke brukte tilstrekkelig tid på artikkelen, mens ? indikerer at leseren ikke har hatt anledning til å lese den. Det blir nå mulig å beskrive hver leser gjennom en vektor som angir leserens scorere på artiklene som utgjør brukerprofilen for leseren. For eksempel har leser 5 følgende brukerprofil:

$\langle 0,?,1,0,1,1 \rangle$

Vi kan nå bruke kollaborativ filtrering til å anbefale en ulest artikkel, dvs. artikkel 1 eller 6, til leser 3. Det er en hel rekke maskinlæringsteknikker å velge mellom, men vi bruker her en enkel teknikk kalt *k-nærmeste nabo* (*k-nearest neighbor*) som benytter cosinus-likhet mellom profilene. Når cosinus-likheten beregnes, kan man se for seg at brukerprofilene utgjør vektorer i et N-dimensjonalt rom (6 dimensjoner i eksempelet fra Figur 2). Cosinus-likheten mellom to vektorer er definert som cosinus av vinkelen mellom vektorene og har en maksimal verdi på 1. To brukerprofiler er altså veldig like dersom cosinus-likheten mellom de to er nær 1.

Kolonnen til høyre i Figur 3 viser cosinus-likheten mellom leser 3 og alle andre lesere i tabellen. Vi konkluderer med at leser 3s to nærmeste naboer er leser 5 (likhet 1) og leser 1 (likhet 0,87). Alternativer til cosinus-likhet er ofte *Persons korrelasjonskoeffisient* mellom lesere og *den euclidske distansen* mellom leservektorene.

Artikkel i → Leser ↓	1	2	3	4	5	6	Cosinus(i,3)
1	1	1	1	1	1	1	0,87
2	1	1	?	1	0	1	0,5
3	?	1	1	0	1	?	1,0
4	0	0	0	1	0	1	0
5	0	?	1	0	1	1	1

**Figur 3:** Kollaborativ filtrering med 5 lesere og 6 artikler

Når vi skal predikere leser 3s interesse for artikkel 1 og 6, kan vi bruke *Pearson-vektet gjennomsnitt* av scorene for de to leserne 1 og 5 til å fylle matrisen for leser 3. Beregningen bruker alle scorene til leser 1 og 5, vektet scorene på basis av likheten mellom leserne, og antar at de manglende scorene for leser 3 står i samme forhold til scorene for leser 1 og 5 som de scorene vi allerede har for leser 3. I dette tilfellet gir metoden et estimat på 1,0 for leserens interesse for artikkel 6 og et estimat på 0,47 for artikkel 1. Systemet anbefaler derfor artikkel 6 til leser 3.

I praksis har man gjerne et mer finmasket nett enn binærverdier i slike matriser, og man normaliserer verdiene for å kompensere for at noen brukere legger seg på generelt høyere verdier enn andre. En brukerprofil kan omfatte flere hundre tusener med verdier. Vi bruker også mye større grupper av brukere når vi skal finne brukere med flere sammenfallende interesser enn bare to.

Eksempelet over viser det vi kaller *bruker-sentrert kollaborativ filtrering* der grupperte brukere er utgangspunktet for å generere anbefalinger. I *artikkel-sentrert kollaborativ filtrering* grupperes derimot artiklene sammen dersom de er lest av de samme brukerne. En ny artikkel anbefales på denne måten dersom den er gruppert sammen med mange andre artikler som leserne tidligere har lest og gitt positive tilbakemeldinger på (Aggarwal, 2016).

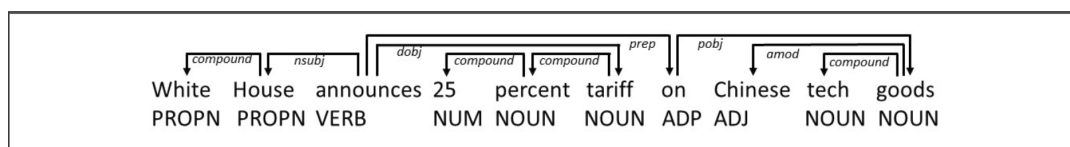
Selv om *k-nearest neighbor* er et eksempel på en av de enkleste maskinlæringsteknikkene, er den mye brukt og gir ofte gode resultater i form av hvor nøyaktig anbefalingene treffer brukernes dokumenterte interesser. Et alternativ er å bruke mer komplekse maskinlæringsteknikker for å generere en modell av datasettet på forhånd og bruke denne på alle senere prediksjoner. Det gir mindre beregninger under selve prediksjonsfasen, men samtidig må det med jevne mellomrom brukes beregningsmessig krevende rutiner til å oppdatere modellen med hensyn til nye artikler som publiseres. Andre mye brukte maskinlæringsteknikker i modell-basert kollaborativ filtrering er *Bayesian classifiers*, *support vector machines*, *singular value decomposition* og nevralt nett (Adomavicius & Tuzhilin, 2005; Rajaraman & Ullman, 2011).

## 6.2. Innholdsbaserte anbefalinger

En annen anbefalingsstrategi er innholdsbaserte anbefalinger. Tanken bak denne strategien er at leserens historikk avslører mye om hva vedkommende er interessert i og kan tenke seg å lese mer av. Hvis loggene for eksempel viser at en leser har brukt mye tid på nyheter om amerikansk politikk, vil det være hensiktsmessig å anbefale flere artikler om samme eller beslektede temaer. Sentralt i slike anbefalingssystemer er derfor logging av brukerens lesevaner og lingvistiske analyser av nyhetsinnhold. Vi kan beskrive denne metoden som følgende:

Innholdsbasert anbefaling er en type anbefalingsteknikk som predikerer en brukers interesse for en artikkel utelukkende basert på denne artikkelens likhet med enten (1) tidligere artikler likt av brukeren eller (2) en generert brukerprofil.

Innholdet av artikler må brytes ned i representasjoner som kan håndteres av anbefalingssystemet. Vi representerer innholdet av en artikkel som en N-dimensjonal vektor, der noen elementer i vektoren kan være metadata om publisist og kategori, mens de aller fleste elementene er ord, fraser eller begreper som sier noe om hva artikkelen handler om. Hvert element har så en vekt som angir hvor prominent elementet er for artikkelens innhold. For å komme fram til disse elementene, må det foretas en lingvistisk analyse som begynner med å normalisere skrivemåter, segmentere separate ord og fjerne stoppord og lingvistisk usignifikant innhold. Ulike parsingsteknikker brukes til å annotere (brekke opp) teksten med ordklasser (*Part-of-speech tagging*), redusere bøyde former til stammeformer (*stemming*) eller oppslagsformer (*lemmatization*), gruppere etterfølgende ord til nominalfraser (*NP chunking*) og trekke ut entiteter som organisasjoner, personer og lokasjoner (*entity extraction*) fra teksten. La oss for eksempel anta en engelsk nyhet inneholder setningen «*White House announces 25 percent tariff on Chinese tech goods.*» Dette kan settes opp som vist i figur 4a og 4b. Figur 4a viser annotering av ordklasser og syntaktiske roller i setningen. «*White House*» er for eksempel gjenkjent som egennavn og subjektet i setningen. Figur 4b viser hvilke entiteter som kan trekkes ut av en slik setning. «*White House*» gjenkjennes som en organisasjon (*ORG*), mens «*Chinese*» analyseres som en *NORP* (nasjonalitet, religiøs eller politisk organisasjon):



**Figur 4a:** Annotering av tekst med ordklasser og syntaktiske roller

[White House]<sub>ORG</sub> announces [25 percent]<sub>PERCENT</sub> tariff on [Chinese]<sub>NORP</sub> tech goods

**Figur 4b:** Identifisering av entiteter og deres typer

For å avgjøre hvor prominent (sentralt) et ord eller en frase er i en nyhetstekst, beregner vi en vekt som settes inn i vektoren. Den mest brukte vektingsmetoden er *tf.idf* (*term frequency – inverse document frequency*), som gir en høy vekt for ord som forekommer hyppig i den aktuelle artikkelen, men sjelden i andre artikler. For anbefalingssystemer er det vanlig å la artikkelvektoren fokusere på entiteter og ord/fraser med høye *tf.idf*-vekter, såkalte nøkkelord. Det er mulig å la artikkelvektoren inneholde alle ord i alle nyheter, noe som delvis gjøres i storskala søkemaskiner, men antakelsen i anbefalingssystemene er at interesseområder er tilstrekkelig representert ved entiteter og betydningsfulle fraser eller begreper. I setningen i Figur 4a kan for eksempel både «Chinese tech goods» og «tech goods» utgjøre nominalfraser, men sannsynligvis opptrer «tech goods» mye oftere og blir da den nominalfrasen som tas med i vektoren som skal representere artikkelens innhold.

Figur 5 illustrerer et enkelt eksempel på fire nyhetsartikler som nylig er publisert. Vektene avslører at artikkel 1 primært tar for seg tollbarrierer på teknologiprodukter, artikkel 2 og 3 dreier seg mest om amerikansk politikk, mens artikkel 4 fokuserer på Donald Trump og Xi Jinping. Brukerloggen viser hvilke artikler brukeren har lest på hvilke tidspunkt. Hvis brukeren gir en score på alle artiklene han har kikket på, kan en definere en brukerprofil som inneholder alle artiklene han har lest, med tilhørende score. Når det senere skal predikeres om en ny artikkel bør anbefales til denne leseren, kan naboskapsmetoden beskrevet over (*k-nearest neighbor*) brukes til å undersøke scorene til de *k* leste artiklene som ligger nærmest den nye artikkelen. Det er som tidligere nevnt dog sjelden at slike eksplisitte scorere er tilgjengelige. Det er derfor mer vanlig å lage en egen brukerprofilvektor, som gir et gjennomsnitt av alle artiklene vedkommende har lest, og som vi antar var av interesse for nettopp denne leseren. Ettersom det antas at gammel historikk er mindre relevant enn den nyeste, brukes ofte en *decay*-funksjon som gradvis reduserer innflytelsen av eldre artikler på brukerprofilen (Darvishy et al., 2015) eller rett og slett ignorerer artikler over en viss alder (Fortuna et al., 2010). Resultatet er at hver bruker representeres med en brukerprofil som består av en vektor med de samme dimensjonene som artikkelvektorene.

I Figur 5 vises profilen til bruker *u* på høyre side. Denne brukeren har åpenbart lest mange artikler om Donald Trump ettersom vekten for Donald Trump er på hele 0,9, mens få artikler har handlet om tollproblemer (vekten på tariff er bare på 0,3).

Artikkel $i \rightarrow$	1	2	3	4		Bruker $u$
Nøkkelord $\downarrow$						
White House	0	0,5	0,7	0,3		0,3
Donald Trump	0,2	0,8	0,5	0,8		0,9
Xi Jinping	0	0,1	0	0,8		0,1
Tariff	0,9	0,2	0,2	0,2		0,3
Tech goods	0,9	0	0	0,1		0,1
Cosinus( $u,i$ )	0,42	0,97	0,82	0,81		

**Figur 5:** Innholdsbaserte anbefalinger for bruker  $u$ .

For å avgjøre hvilke artikler som skal anbefales til bruker  $u$  i Figur 5, beregner vi cosinuslikheten mellom brukerprofilvektoren og hver av artikkelvektorene. Resultatet, som er vist nederst i Figur 5, forteller at artikkel 2 matcher brukerprofilen veldig godt med en likhet på 0,97. Det er dermed en fornuftig artikkel å anbefale. Dette er ikke så overraskende når en ser at begge vektorene domineres av Donald Trump og har få andre prominente nøkkelord.

I praktiske løsninger vil det være tusenvis av nøkkelord i slike artikkel- og brukervektorer, og det foreligger ofte metainformasjon som artikkellengde og publisist integrert i maskineriet. I stedet for term-metoden  $tf.idf$  kan statistiske tester som Gini-indeksen eller entropi for å avgjøre hvilke nøkkelord som bør inkluderes i vektorene, brukes. Et populært alternativ til naboskapsmodellen i innholdsbaserte anbefalinger er Bayesianske klassifikatorer. Med *Bayes* estimeres sannsynligheten for at en nyhet  $a$  tilhører klassen av interessante nyheter  $c$  for en leser, basert på annen sannsynlighetsinformasjon om  $c$  og  $a$ . Dette er komplekse beregninger, og i praksis benyttes ofte en enklere variant, som *naiv Bayes*, som drar nytte av samme type vektorer som naboskapsmodellen (Bellugín & Said, 2018).

### 6.3. Hybridløsninger og alternative tilnærminger

Både kollaborativ filtrering og innholdsbaserte anbefalinger som vi har presentert ovenfor, brukes på nettsidene for nyhetsmedier. Innholdsbaserte anbefalinger har dessuten vist seg å håndtere godt *kaldstart*-problemer. Innholdsbaserte anbefalinger har derimot blitt kritisert for å generere *filterbobler* eller *ekkokamre*. Fordi anbefalte artikler baserer seg på leserens lesehistorikk, er det en risiko for at de innholdsbaserte anbefalingene forsterker det leseren allerede har vist interesse for og at vedkommende skjermes fra nye temaer eller perspektiver. Et annet potensielt problem med innholdsbaserte systemer er at det ikke foreligger en forståelse av kvaliteten på artiklene. Artikler anbefales utelukkende basert på likhet med andre artikler. Ettersom systemet anbefaler artikler som noen har lest og vurdert som interessante gjennom kollaborativ filtrering, må en viss grad av menneskelig kvalitets sjekk til for å hindre at veldig dårlige artikler blir anbefalt.

Senere forskning viser til gode resultater når kollaborativ filtrering og innholdsbaserte anbefalinger sammenstilles i hybride løsninger (Liu et al., 2010). Resultatene fra de separate anbefalingsmotorene kan kombineres på ulikt vis (Burke, 2002), men mange løsninger

bruker ulik vektning på motorene, situasjonsavhengig svitsjing mellom dem eller mer eller mindre tilfeldig blanding av resultater fra flere motorer. Et spesialtilfelle er når resultater fra en anbefalingsmotor blandes med redaksjonelt utvalgte artikler.

I nyhetsdomenet er det også andre faktorer som kan være hensiktsmessige for å finne gode anbefalinger. Ferske nyheter er for eksempel gjerne mer interessante enn gamle, nyheter i lokalområdet kan ofte være interessante selv om de ikke matcher brukerprofilen, og det kan noen ganger være bedre å fokusere på store viktige hendelser fremfor leserens generelle interesser innenfor en eller annen kategori. *Kontekstuelle anbefalingssystemer* bruker derfor tid, lokasjon og andre kontekstuelle faktorer til å anbefale kontekst-relevante nyheter, ofte uten å involvere brukerprofilen i det hele tatt. *Sesjonsbaserte anbefalingssystemer* genererer derimot anbefalinger utelukkende basert på hva brukeren har lest i den sesjonen han er i, og trenger derfor ingen lagrede brukerprofiler over leserens generelle interesser. Hver slik sesjon inkluderer typisk 8–10 artikkellesninger (Gulla et al., 2017), noe som er nok til å bygge en begrenset korttidsprofil over brukeren. Hvis systemet ikke har noen brukerprofil over leseren, kan også *popularitets-drevne motorer*, som rett og slett anbefaler de mest populære nyhetene akkurat nå, være aktuelle.

Til tross for flere muligheter er det brukerprofilene som det jobbes mest med innen forskning på nyhetsanbefalinger (Karimi et al., 2018).

Vi vil nå presentere funn fra tre forskjellige tester gjennomført hos Polaris Media basert på ulike metoder beskrevet ovenfor som belyser problemstillingen vi skisserte innledningsvis.

## 7. FUNN OG TESTRESULTATER

### 7.1. Test 1: Adresseavisen og forsøk med sosiale medier

Adresseavisen utformet januar 2015 hypotesen: «Er brukere fra sosiale medier annerledes enn brukere ellers?». Utgangspunktet var at alle artikler hadde tre lenker under selve artikkelen. Lenkene var presentert med bilde og tekst på lik linje med hvordan artikler presenteres på forsiden av ulike nettaviser. I løpet av det første kvartalet 2015 gjennomførte Adresseavisen tre AB-tester på artikkelsidene. Det var en kjensgjerning at brukere fra Facebook var vesentlig mindre lojale enn andre brukere ettersom slike lesere hadde et lavere antall saker per bruker enn lesere som landet fra andre nettsider.

Først ble det etablert en referanse for bruk i videre testing. Testingen hadde flere steg: Ved AB-testing sammenlignes prototyper med resultater fra tidligere prototyper for å få fram best utfall. Utviklingen er smidig ved at nye versjoner kan introduseres underveis. Slik optimaliseres utfallet i en iterativ prosess.

Adresseavisen sammenlignet til å begynne med populært (*trending*) innhold med artikkelen som nettopp var lest (kontekstuell). De kontekstuelle anbefalinger hadde 5–15% høyere klikkrate enn de sakene som ble plukket ut basert på popularitet. Dette viste at hele brukermassen samlet sett ble mer engasjert av å få anbefalt innhold som lignet på det de akkurat hadde lest, fremfor å få presentert generelt populære artikler her og nå.

Neste steg var å utføre tilsvarende AB-test utelukkende på brukere som kom fra Facebook. På disse brukerne var det mulig å sammenligne de ulike personaliserings-alternativene på tvers av hele gruppen og under-grupperingen. Testen viste at Facebook-brukere



hadde større sannsynlighet for å klikke på en *trending*-sak, fremfor en sak som likner på den aktuelle saken. *Trending*-nyhetene hadde hele 20–50 % høyere klikkrate i testperioden. Fordi Adresseavisen er en regional avis og Facebook-brukere kommer fra hele landet, kan det tyde på at Facebook-brukere ikke har samme tilhørighet til det lokale innholdet som brukergruppen generelt.

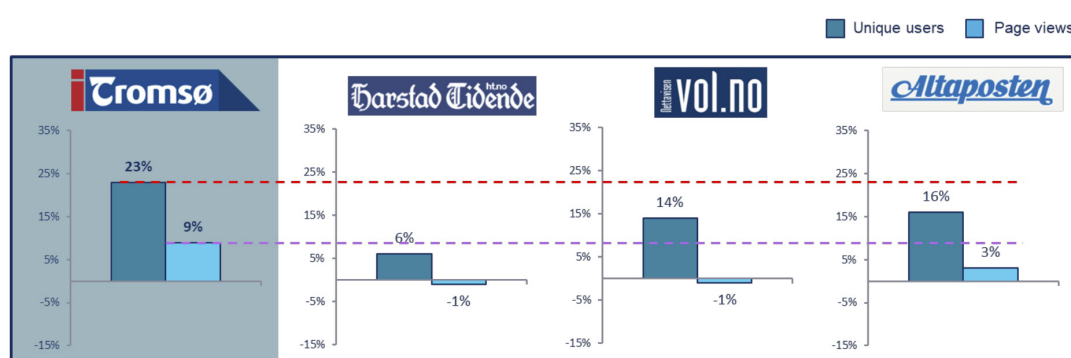
Den siste testen ble utført for å undersøke om *trending*-innhold generelt var mer populært enn *trending*-innhold som hadde trafikk fra Facebook. Klikkraten her viste seg å være 10–40 % høyere for de generelle *trending*-nyhetene. Den bekreftet ytterligere at brukergruppen fra Facebook hadde stort engasjement rundt generelt populære nyheter.

Dette eksperimentet viser hvordan man gjennom iterative tester identifiserer et avgrenset område (her en spesifikk brukergruppe) hvor relevans er høyere enn øvrige områder. Testene forteller to ting: Brukere som kommer fra Facebook, må håndteres ulikt, og brukergrupper er noe som bør eksperimenteres mer med for å identifisere andre sammenhenger hvor vi har økt relevans.

## 7.2. Test 2: Eksperiment på mobilfronten til iTromsø

Høsten 2015 startet initiativet til eksperimentet med personaliseringsteknologi i stor skala hos iTromsø. Det ledet til hypotesen «Kan algoritmen styre en hel mobil forside med individuell tilpasning?». Frem til eksperimentet ble lansert utviklet Big Data-teamet en prototype, tett oppfulgt med innspill fra iTromsø. Man tok også høyde for potensielle filterbøller. Løsningen ble satt til å vise en viss spredning av innhold. Dessuten kunne redaksjonen overstyre den øverste artikkelen ved behov. Målet var å øke trafikk, engasjement og dermed relevans. Forsiden skulle dynamisk endre seg i takt med brukerens adferd. Eksperimentet samlet data i Cxense sine systemer i 3 måneder før analysen ble startet. Som det fremgår av figurene, er noen analyser gjort etter 5 måneder.

Analysen viste en økning på 23 % av antall unike brukere og 9 % økning av klikkratene. Nettstedet hadde signifikant høyere vekst i brukere og trafikk enn aviser i samme geografiske område og av samme størrelsesorden (se Figur 6).



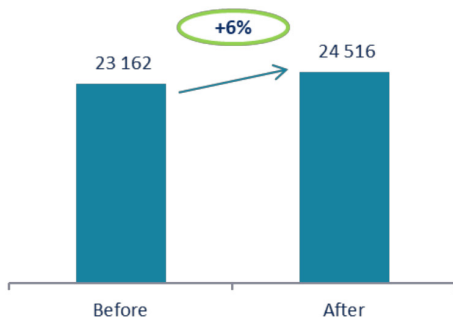
**Figur 6:** Endring av unike brukere og klikkrater for nordnorske aviser

I tillegg viste resultatene at trafikken på forsiden økte, både sammenlignet med tilsvarende periode før og samme periode ett år tidligere. Som Figur 7 viser, hadde iTromsø 23 162

klikk i 3-månedersperioden før den helautomatiske mobilfronten ble innført i 2016, mens en i 3-månedersperioden etter hadde økt med 6 % til 24 516 klikk.

#### Traffic on front page – before and after

Source: Cxense, 3 months before and after launch



#### Traffic on front page VS last year

Source: Cxense, 3 months before and after launch

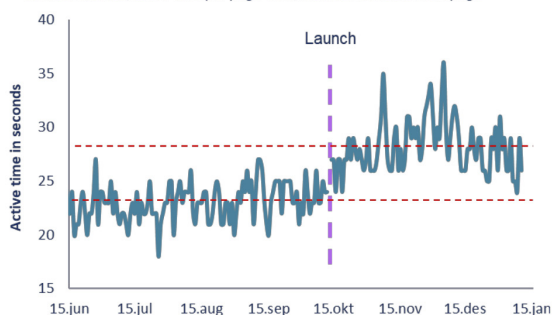


**Figur 7:** Økning av trafikken på frontsiden

Resultatene viste videre at tidsbruken på forsiden økte over natten (Figur 8).

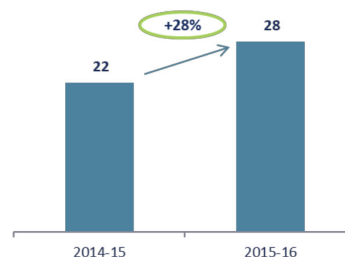
#### Significant increase in time spent on site from day one

Source: Cxense, active time per page view, itromso.no mobile front page



#### Average time spent on front page - seconds

Source: Cxense, 3 months after launch VS one year earlier



**Figur 8:** Økt tidsbruk på forsiden på kort og lang sikt

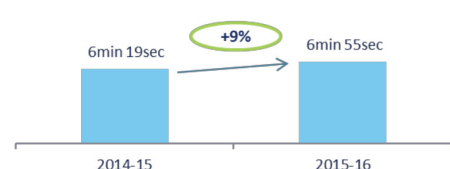
Det mest overraskende var at tidsbruken på artiklene som kom fra den nye forsiden, økte fra et gjennomsnitt på 54 sekunder før testen startet, til 62 sekunder etter (15 % økning), selv om det ikke var gjort noen endringer på artiklene i eksperiment-perioden. Sammenlagt økte mediekonsumet på iTromsø sin mobile nettside med 31 % i løpet av de fem månedene etter at vi startet eksperimentet. Det betyr at konsumet over fem måneder økte fra rundt 92 500 timers nyhetskonsum til 120 833 timer i løpet av ett enkelt år. Figur 9 viser hvordan denne økningen brytes ned i en økning av antall mobile brukere på 20 % og en økning av mobile brukeres sesjonstid på 9 % til 6 minutt og 55 sekunder.

**Total number of users - mobile**

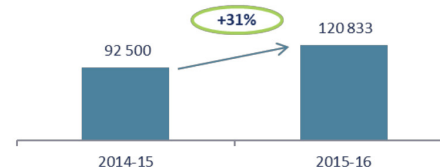
Source: Cxense, first 5 months after launch VS one year earlier

**Average active time per user - mobile**

Source: Cxense, first 5 months after launch VS one year earlier

**Total media consumption on iTromsø in hours**

Source: Cxense, first 5 months after launch VS one year earlier, mobile only

**Figur 9:** Økt mediekonsum med automatiserte anbefalinger i iTromsø.

## 7.3. Test 3: Revitalisering av eldre innhold i Altaposten

Under et prosjekt i Altaposten tidlig 2017 ble det gjort observasjoner som ønsket «å måle effekten av å trekke eldre avisartikler inn i anbefalingsmotoren». Dette ble testet mars/april samme år. Hver test ble kjørt i en 3-ukersperiode. Økt tilgjengelig innholdsvolum i modellen skulle generere mer relevante anbefalinger til tross for at innholdet kunne være betydelig eldre. Økningen av klikkratene, som er vist i Figur 10, er overraskende sterke, selv når artikler som er opptil ett år gamle trekkes inn. Når klikkraten for mobile brukere øker med 17 % for anbefalte nyheter som er opptil 4 uker gamle i stedet for to dager, kan det tyde på at mindre aviser publiserer en del saker som ikke er så tidskritiske.

Brukere	Opprinnelig alder på anbefalte artikler	Artikkelalder ved utvidelse av artikkelsett	Økning i klikkrater
Mobile lesere	0–2 dager	0–4 uker	17 %
	0–12 uker	0–52 uker	23 %
Desktop-lesere	0–2 dager	0–4 uker	42 %
	0–12 uker	0–52 uker	37 %

**Figur 10:** Anbefalinger gjør eldre innhold mer relevant i Altaposten

For å returnere til problemstillingen som var utgangspunktet for denne artikkelen – Hvordan kan anbefalingsløsninger bidra til å tilby mer relevante nyheter til leseren på den ene siden, og økt lønnsomhet for mediehus på den andre? – viser innsikt fra de tre testene at anbefalingsløsninger kan gi økt bruk av nettutgavene, økt lesetid, lengre levetid for innhold, større engasjement blant tilbakevendende brukere og økt kundelojalitet. For Polaris Media har anbefalingsløsningene resultert i både høyere klikkrater og mer lesetid. Teknologien økte innholdets levetid, gav mer gjenbruk av eldre nyhetsstoff og bidro til å forstå egne leserssegmenter bedre. Men utstrakt eksperimentering måtte til for å forstå teknologien tilstrek-

kelig og evaluere effekten av den. Samspeilet mellom automatiserte anbefalinger og redaksjonelle vurderinger var og er viktig, støttet av gode brukergrensesnitt og visuelle effekter.

## 8. AVSLUTNING

Mediehus investerer i nye plattformer for personaliserte nyheter, der tunge maskinlærings-teknikker kjøres i sanntid på store datasett for å predikere lesernes interesse for innkommende nyheter. Teknologien har muliggjort en finmasket nyhetstilpasning som overgår tidligere manuelle prosesser og redaksjonelle vurderinger i presisjon og nytte for brukeren. Mens noen mediehus bruker teknologien som et supplement til manuelle rutiner, har andre gått over til helautomatiserte personaliserte nyhetssider.

Resultatene fra Polaris Media, som er presentert i denne artikkelen, viser ikke bare mulighetene som ligger i anbefalingsløsninger, men det er også fortsatt fundamentale utfordringer knyttet til håndteringen av nyhetsstoff og forståelsen av hva som utgjør de beste, personaliserte nyhetene. Dessuten kan det også være en viss uenighet mellom leserne og innholdsleverandørene om hva som er «beste nyheter». Fordi anbefalingssystemene avhenger av mye og riktig data om brukerne, er det også en diskusjon om hvordan mediehusene skal forholde seg til personvern og lovverk om persondata. Anbefalingsløsninger er på mange måter et spill med store data. Maskinlærings-teknikker har en tendens til å gi bedre resultater med flere eksempler å lære fra og mer nøyaktig informasjon om eksemplene. Da kan det være fristende å samle inn store mengder data om brukeratferd og forsøke å kople disse til andre datakilder som gir enda mer utfyllende informasjon om brukerne. Utfordringen er å finne en god balanse mellom det som gir gode tjenester til innholdskonsumertene, og samtidig respekterer de samme konsumentenes krav til privatliv og personvern.

## REFERANSER

- Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.
- Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer.
- Belkin, N. J. & Croft, W. B. (1992). Information filtering and information retrieval: two sides of the same coin. *Communications of the ACM*, 25(12), s. 29-38.
- Bellogín, A. & Said, A. (2018). Information Retrieval and Recommender Systems, pp. 79–99. I S. Said & V. Torra (Red.), *Data Science in Practice*. Springer. <https://doi.org/10.1007/978-3-319-97556-6>
- Billsus, D. & Pazzani, M. (1999). A hybrid user model for news story classification. *Proceedings of the 7<sup>th</sup> International Conference on User Modeling*. Springer.
- Borges, H. L. & Lorena, A. C. (2010). A Survey on Recommender Systems for News Data. I E. Szczerbicki & N. T. Nguyen (Red.), *Smart Information & Knowledge Management, SCI 260*, s. 129–151. Springer.
- Cranor, L.F., Reagle, J. & Ackerman, M.S. (2000). Beyond concern: Understanding net users attitudes about online privacy. *The Internet upheaval: raising questions, seeking answers in communications policy*, 47–70. MIT Press.
- Darvishy, A., Ibrahim, H., Mustapha, A. & Sidi, F. (2015). New Attributes for Neighborhood-based Collaborative Filtering in News Recommendation. *Journal of Emerging Technologies in Web Intelligence*, 7(1), 13–19.

- De Pessemier, T., Vanhecke, K. & Martens, L. (2015). A Personalized and Context-Aware News Offer for Mobile Devices. *Proceedings of the 11<sup>th</sup> International Conference on Web Information Systems and Technologies (WEBIST 2015)*, s. 147–168.
- Epure, E., Deneckere, R., Salinesi, C., Kille, B. & Ingvaldsen, J. E. (2017). *Atelier interdisciplinaire sur les systèmes de recommandation / Interdisciplinary Workshop on Recommender Systems*.
- Fortuna, B., Fortuna, C. & Mladenic, D. (2010). Real-Time News Recommender System. I *Machine Learning and Knowledge Discovery in Databases*, volum 6323 i Lecture Notes in Computer Science, s. 583–586. Springer.
- Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., & Bruttin, C. (2014). Offline and online evaluation of news recommender systems at swissinfo.ch. *Proceedings of the 8<sup>th</sup> Conference on Recommender Systems (Recsys 2014)*, s. 169–176. ACM.
- Gulla, J. A. & Fidjestøl, A. D., Su, X. & Castejon, H. (2014). Implicit User Profiling in News Recommender Systems. *Proceedings of the 10th International Conference on Web Information Systems and Technologies (WEBIST 2014)*, s. 185–192.
- Gulla, J. A., Marco, C., Fidjestøl, A. D., Ingvaldsen, J. E. & Özgöbek, Ö. (2016). The Intricacies of Time in News Recommendation. *User Modeling, Adaptation and Personalisation (UMAP 2016), Extended Proceedings*. ACM.
- Gulla, J. A., Zhang, L., Liu, P., Özgöbek, Ö. & Su, X. (2017). The Adressa dataset for news recommendation. I *Proceedings of the International Conference on Web Intelligence*, s. 1042–1048. ACM.
- Grønli, Tor Morten (2017). Artefaktutvikling og problemløsning i IT. I Næss og Pettersen (red.), *Metodebok for kreative fag* (s. 230–239). Oslo: Universitetsforlaget.
- Ingvaldsen, J. E. & Gulla, J. A. (2015). Taming news streams with linked data. *9<sup>th</sup> IEEE International Conference on Research Challenges in Information Science*, s 536–537, Athens. IEEE.
- Karimi, M., Jannach, D. & Jugovac, M. (2018). News Recommender Systems – Survey and Roads Ahead. *Information Processing & Management*, 54(6), 1203–1227. DOI: <https://doi.org/10.1016/j.ipm.2018.04.008>.
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H. & Newell, C. (2012). Explaining the user experience of recommender systems. *User Model & User-Adapted Interaction*. DOI: [10.1007/s11257-011-9118-4](https://doi.org/10.1007/s11257-011-9118-4).
- Kotkov, D., Veijalainen, J. & Wang, S. (2016). Challenges of Serendipity in Recommender Systems. *Proceedings of the 12<sup>th</sup> International Conference on Web Information Systems and Technologies (WEBIST'16)*, s. 251–256.
- Lerche, L. (2016). *Using Implicit Feedback for Recommender Systems: Characteristics, Applications and Challenges*. PhD thesis. Technische Universität Dortmund.
- Li, L., Zheng, L. & Li, T. (2011). LOGO: A Long-short User Interest Integration in Personalized News Recommendation. *Proceedings of the 5<sup>th</sup> Conference on Recommender Systems (Recsys'11)*, s. 317–320. ACM.
- Liu, J., Dolan, P. & Pedersen, E. R. (2010). Personalized News Recommendation Based on Click Behavior. *Proceedings of the 15<sup>th</sup> International Conference on Intelligent User Interfaces (IUI'10)*, s. 31–40. DOI: [10.1145/1719970.1719976](https://doi.org/10.1145/1719970.1719976).
- Mohallick, I., De Morr, K., Özgöbek, Ö. & Gulla, J. A. (2018). Towards New Privacy Regulations in Europe: Users' Privacy Perception in Recommender Systems. *Proceedings of the 10<sup>th</sup> International Symposium on UbiSafe Computing (UbiSafe 2018)*, Melbourne, desember 2018.
- Nicastro, D. (2016). How Personalized Web Experience Saved a Newspaper. *CMS Wire*, 30 March 2016. Hentet fra <https://www.cmswire.com/digital-experience/how-personalized-web-experience-saved-a-newspaper/>.
- Olson, J.S., Grudin, J. & Horvitz, E. (2005). A study of preferences for sharing and privacy. *CHI'05 extended abstracts on Human factors in computing systems*, s. 1985–1988. ACM.

- Özgöbek, Ö., J. A. Gulla & R. C. Erdur. A Survey on Challenges and Methods in News Recommendation. *Proceedings of the 10<sup>th</sup> International Conference on Web Information System and Technologies (WEBIST 2014)*, Barcelona, April 2014.
- Papagelis, M., Plexousakis, D. & Kutsuras, T. (2005). Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences. *Proceedings of the 3<sup>rd</sup> International Conference on Trust Management (iTrust'05)*, s. 224–239.
- Park, K., Lee, J. & Choi, J. (2017). Deep Neural Networks for News Recommendations. *Proceedings of the 2017 ACM Conference on Information and Knowledge Management (CIKM'17)*, side 2255–2258. DOI:[10.1145/3132847.3133154](https://doi.org/10.1145/3132847.3133154).
- Rajaraman, A. & Ullman, J. D. (2011). *Mining of Massive Datasets*. Cambridge University Press.
- Shani, G. & Gunawardana, A. (2011). Evaluating Recommendation Systems. I F. Ricci, L. Rokach, B. Shapira & P. B. Kantor (red.), *Recommender Systems Handbook* (s. 257–297). Springer.
- Vargas, S. & Castells, P. (2011). Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems. *Proceedings of the 5<sup>th</sup> Conference on Recommender Systems (Recsys'11)*, s. 109–116. ACM.
- Zins, A. H., Bauernfeind, U., Del Missier, A., Venturini, A. & Rumetshofer, H. (2004). An Experimental Usability Test for different Destination Recommender Systems. *Information and communication technologies in tourism 2004, the 11<sup>th</sup> ENTER International Conference*. DOI: 10.1007/978-3-7091-0594-8 22