

Suositusjärjestelmät verkkokaupoissa:
tekniset haasteet, ratkaisumenetelmät ja
arviointi

TURUN YLIOPISTO
Tietotekniikan laitos
TkK-tutkielma
Tietotekniikka
Joulukuu 2025
Daniela Degenhardt

TURUN YLIOPISTO
Tietotekniikan laitos

DANIELA DEGENHARDT: Suositusjärjestelmät verkkokaupoissa: tekniset haasteet, ratkaisumenetelmät ja arviointi

TkK-tutkielma, 30 s.
Tietotekniikka
Joulukuu 2025

Tutkielmassa tarkastellaan verkkokauppojen suositusjärjestelmiä keskittyen niiden teknisiin haasteisiin, ratkaisumenetelmiin ja arviointiin. Suositusjärjestelmät ovat keskeinen osa sähköistä kaupankäyntiä, sillä ne personoivat käyttäjäkokemusta, helpottavat relevanttien tuotteiden löytämistä ja tukevat liiketoiminnallisia tavoitteita. Kirjallisuuskatsauksessa eritellään suositusjärjestelmien yleisimmät lähestymistavat, kuten yhteistoiminnallinen, sisältöpohjainen ja hybridisuodatus, sekä kuvataan suositusten muodostumisen perusvaiheet ja käyttäjä–kohde–vuorovaikutuksen rooli. Keskeisiksi teknisiksi haasteiksi tunnistetaan kylmäkäynnistys, datan harvuus, skaalautuvuus sekä monimuotoisuuden ja tarkkuuden välinen ristiriita. Haasteiden ratkaisemiseksi esitetään muun muassa metadatan hyödyntämistä, hybridimalleja, klusterointia sekä luottamukseen perustuvia menetelmiä. Lisäksi tarkastellaan mallipohjaisia optimointitekniikoita, kuten matriisihajotelmiin perustuvia menetelmiä, jotka voivat lieventää erityisesti harvuuteen liittyviä ongelmia.

Arviointiosiossa tarkastellaan sekä matemaattisia suorituskykymittareita, kuten tarkkuus, täsmällisyys, muistettavuus ja RMSE, että käytännön arviointimenetelmiä, kuten A/B-testausta, online-mittareita ja käyttäjätestausta. Tutkielma korostaa suositusjärjestelmien kontekstisidonnaisuutta: tehokkainta järjestelmää ei voida määritellä yleispätevästi, vaan toimivat ratkaisut riippuvat datan luonteesta, sovellusympäristöstä ja käyttäjäkäyttäytymisestä. Suositusjärjestelmien kehittäminen nähdään jatkuvana optimointiprosessina, jossa tekniset tavoitteet, käyttäjäkokemus ja liiketoiminnalliset päämäärät pyritään sovittamaan yhteen.

Asiasanat: suositusjärjestelmät, verkkokauppa, suositus, kylmäkäynnistys, datan harvuus

Sisällys

1	Johdanto	1
2	Suositusjärjestelmien perusteet	4
2.1	Suosituksen synty	6
2.2	Merkitys sähköisessä kaupankäynnissä	9
3	Järjestelmien luokittelu	11
3.1	Yhteistoiminnalliset suositusjärjestelmät	12
3.2	Sisältöpohjaiset menetelmät	14
3.3	Muita suositusjärjestelmiä	15
3.4	Hybridipohjaiset suositusjärjestelmät	16
4	Tekniset ongelmat ja ratkaisut	17
4.1	Kylmäkäynnistys	17
4.2	Datan harvuus	19
4.3	Skaalautuvuus	20
4.4	Monimuotoisuus ja tarkkuus	21
5	Järjestelmien arviointi	22
6	Pohdinta	27
7	Yhteenveto	29

Kuvat

1.1	Kaavio aineiston valinnasta ja käsittelystä	3
2.1	Suosituksen vaiheet	7
3.1	Suositusjärjestelmien luokittelu	12
3.2	Käyttäjä-kohde-matriisi	13

1 Johdanto

Suositusjärjestelmistä tuli tutkimuksen mielenkiinnonkohde 1990-luvun puolivälissä [1], kun digitaalinen informaatio aloitti eksponentiaalisen kasvun [2]. "Näistä voitit tykätä seuraavaksi", "Käyttäjät, jotka ostivat tämän ostivat myös:", "Saatat pitää myös" ja monet vastaavat ovat verkkokauppojen sivuilla toistuvia vakiolauseita. Ne ovat esimerkkejä siitä, miten suositusjärjestelmät ohjaavat käyttäjiä kohti mahdollisesti kiinnostavia tuotteita. Niiden tavoitteena on lisätä myyntiä, mutta samalla ne tekevät näkyväksi sen, kuinka laajasti suositusjärjestelmiä hyödynnetään sähköisessä kaupankäynnissä.

Suositusjärjestelmät määrittellään älykkäiksi suodattimiksi, jotka käsittelevät niille ennalta määriteltyä dataa ja tuottavat sen pohjalta personoituja suosituksia käyttäjille [3]. Verkkokaupan suositusjärjestelmät määrittellään järjestelmiksi, jotka hyödyntävät verkkokauppa-alustaa asiakkaiden tukemiseen tarjoamalla heille henkilökohtaisia suosituksia. Tavoitteena on auttaa asiakasta ostopäätöksessä samaan tapaan kuin myyjä myymäläympäristössä. Verkkokaupan suositusjärjestelmä mahdollistaa verkkosivuston mukautumisen asiakkaan yksilöllisiin tarpeisiin ja luo personoidun osto-ympäristön.[4]

Tämä on kuitenkin ihannetilanne, joka ei aina toteudu käytännössä. Suositusjärjestelmät kohtaavat useita haasteita, jotka heikentävät sekä suositusten tehokkuutta että käyttäjäkokenemusta. Tässä tutkielmassa perehdytään verkkokauppojen algoritmipohjaisten suositusjärjestelmien teknisiin haasteisiin sekä tarkastellaan, millaisia

ratkaisuja niiden minimoimiseksi on esitetty. Tavoitteena ei ole määritellä yhtä universaalia "ideaalia" suositusjärjestelmää vaan pikemminkin tarkastella, millaisia tekijöitä tulisi ottaa huomioon pyrittäessä teknisesti tehokkaaseen ja kontekstuaalisesti toimivaan järjestelmään. Tutkielmassa vastataan seuraaviin tutkimuskysymyksiin:

TK1 Millaisia teknisiä haasteita verkkokauppojen suositusjärjestelmissä on?

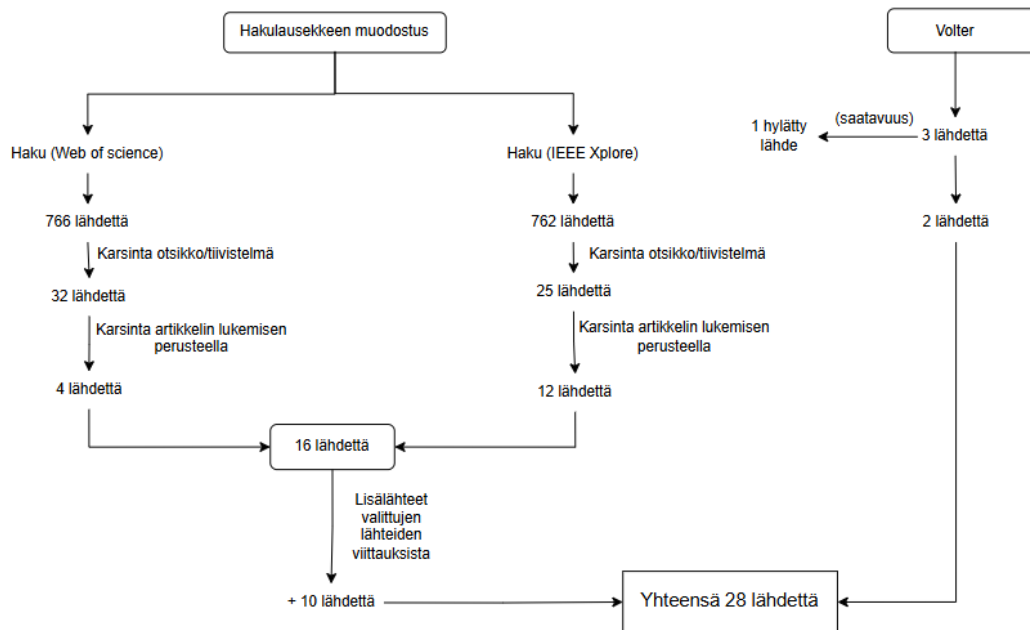
TK2 Millaisia menetelmiä ja lähestymistapoja on ehdotettu näiden ongelmien ratkaisemiseksi?

TK3 Miten suositusjärjestelmien tehokkuutta ja toimivuutta voidaan arvioida, ja millaisia mittareita niiden arvioinnissa käytetään?

Tutkielma on kirjallisuuskatsaus. Tutkielman aineistonhakua esittää Kuva 1.1. Aiheeseen liittyvää yleistietoa sisältävää kirjallisuutta etsittiin Turun yliopiston Volter-tietokannasta hakusanalla ("recommender system*"OR "recommendation system*"). Haun avulla löydettiin kolme sopivaa perusteosta, joista saatavuuden perusteella valittiin kaksi. Tarkemmin tutkimuskysymyksiin liittyvää aineistoa on haettu IEEE Xplore- ja Web of science-tietokannoista muodostetulla hakulausekkeella. Hakulausekkeella saaduista hakutuloksista on valittu artikkelit otsikoiden ja nopean lukemisen perusteella. Lisäaineistoa on kerätty valittujen artikkeleiden viitteistä.

Aineistohaku tehtiin käyttäen seuraavaa hakulausetta: ("recommender system*"OR "recommendation system*") AND ("e-commerce") AND ("technical challenge*"OR "technical problem*"OR "limitation*"OR "scalability"OR "cold start"OR "performance"). Ensin keksittiin aiheeseen liittyvä alustava hakulauseke, jota testattiin yllä mainittuihin tietokantoihin. Hakutulosten pohjalta lausetta korjailtiin mahdollisimman ihanteellisen tuloksen saamiseksi. IEEE Xplore- ja Web of Science-tietokannoista löydettiin yhteensä 1528 tulosta, joista suurin osa karsittiin

otsikon tai tiivistelmän perusteella. Jäljelle jäi yhteensä 57 artikkelia, joista tarkemman lukemisen perusteella hylättiin 41. Tutkielmassa lähteitä on yhteensä 28, joista 12 käytettiin pääasiassa taustalukuihin ja 16 vastaamaan tutkimuskysymyksiin.



Kuva 1.1: Kaavio aineiston valinnasta ja käsittelystä

Tutkielman luvussa 2 määritellään suositusjärjestelmiin liittyvät keskeiset termit, suosituksen yleiset vaiheet ja merkitys sähköisessä kaupankäynnissä. Luvussa 3 perehdytään suositusjärjestelmien luokitteluun ja luvussa 4 tekniisiin ongelmiin ja niiden ratkaisuihin. Luku 5 erottelee mittarit, joita käytetään suositusjärjestelmien arvioimiseen. Tutkimuksen lopussa luvussa 6 pohditaan mahdollista ideaalimallia ja sen kontekstuaalisuutta ja luvussa 7 tiivistetään tutkielman oleelliset asiat ja vastaan tutkimuskysymyksiin.

2 Suositusjärjestelmien perusteet

Suositusjärjestelmän tehtävänä on ennustaa käyttäjälle kohteita, jotka vastaavat hänen yksilöllisiä tarpeitaan ja mieltymyksiään. Tämän saavuttamiseksi järjestelmät mallintavat sekä käyttäjiä että heidän vuorovaikutustaan tarjolla olevan sisällön kanssa. Käyttäjistä ja kohteista kerätty tieto muodostaa perustan erilaisille suositelumenetelmille.[5] Tässä luvussa esitellään suositusjärjestelmien perusmekanismi ja keskeisimmät käsitteet.

Henkilöä, jolle suositus esitetään, kutsutaan käyttäjäksi (engl. user) [6]. Suositusjärjestelmän käyttäjillä voi olla hyvin erilaisia tavoitteita ja ominaisuuksia. Suosituksen, käyttäjän ja tietokoneen välisen vuorovaikutuksen personoimiseksi järjestelmät hyödyntävät monenlaista käyttäjistä kerättyä tietoa. Tämä tieto voidaan jäsentää useilla eri tavoilla, ja se mitä tietoa mallinnetaan ja miten, riippuu käytetystä suositelumenetelmästä. Käyttäjistä kerätty tieto jaetaan yleisesti vuorovaikutusdataan ja metadataan. Vuorovaikutukseen perustuva data sisältää tiedot siitä, mitä käyttäjä tekee verkkokaupassa. Metadata taas sisältää esimerkiksi demografisia tietoja käyttäjistä eli tietoja perusominaisuuksista, kuten ikä tai tulotaso, käyttäjän antamia profilitietoja tai sosiaalisen median kautta välittyvää taustatietoa.[5]

Kohteet (engl. items) ovat objekteja, joita järjestelmä suosittelee [6]. Kohteet voidaan luokitella niiden arvon ja hyödyn perusteella. Kohteen arvo on positiivinen, jos kohde on käyttäjälle hyödyllinen, ja negatiivinen, jos kohde ei vastaa käyttäjän tarpeita ja hän on tehnyt virheellisen valinnan. Kohteisiin liittyvät tiedot ovat sekä

vuorovaikutusdataa että metadataa. Vuorovaikutusdata sisältää tietoja siitä, miten käyttäjät ovat olleet vuorovaikutuksessa kohteen kanssa. Metadata sisältää tiedot kohteen ominaisuuksista, kuten teknisiä tietoja, kategoriat tai käyttötarkoitus. Tällainen tieto auttaa järjestelmää arvioimaan tuotteiden välisiä samankaltaisuuksia. Samalla tavalla myös käyttäjämetadata auttaa järjestelmää yhdistämään samankaltaiset käyttäjät.[5]

Vuorovaikutuksella tarkoitetaan käyttäjän antamaa palautetta ja tämän ajatellaan olevan joko suoraa (explicit feedback) tai epäsuoraa (implicit feedback). Suoraa palautetta on esimerkiksi arvostelu tai arvio asteikolla 1–5.[6] Arvostelut ovat suosituin vuorovaikutuksen tyyppi, jota suositusjärjestelmät käyttävät suositusten hyväksiksi [5]. Tämän voidaan ajatella johtuvan siitä, että tällainen tieto on yleensä erittäin luotettavaa. Tieto käyttäjän mieltymyksistä on varma, koska se tulee suoraan tältä. Haasteena taas nähdään se, että arvostelun antaminen vaatii käyttäjältä ylimääräistä vaivaa.[3] Järjestelmän on hyvin vaikea muodostaa suosituksia käyttäjälle, joka ei arvostele tuotteita, ja siksi tällaisen palautteen hyvyys on kontekstisidonnaista eikä niinkään absoluuttinen totuus.

Epäsuora palaute taas välittyy siitä, miten käyttäjä reagoi kohteeseen ilmoittamatta suoranaisesti mieltymystään [6]. Epäsuora palaute on esimerkiksi tietyllä sivulla vietetty aika, avatut linkit ja klikkaukset. Tällaisen palautteen etuna voidaan pitää sitä, ettei se vaadi käyttäjältä aktiivista osallistumista.[3]

Toisin sanoen suoran palautteen tapauksessa tehtävä määritellään usein ennusteongelmaksi, jossa pyritään arvioimaan, millaisen arvosanan käyttäjä antaisi tietylle kohteelle, ja epäsuorassa palautteessa tehtävä voidaan sen sijaan määritellä kulutusasteen maksimoimiseksi. Huomioon ei niinkään oteta sitä, mistä käyttäjä pitää tai ei pidä, vaan se, minkälaisien kohteiden kanssa käyttäjä todennäköisesti on vuorovaikutuksessa.[2] Vaikka epäsuoran palautteen tarkkuus voi jäädä suoraa palautetta heikommaksi sen perustuessa päättelyyn ja todennäköisyyteen, on esitetty,

että epäsuora palaute on objektiivisempaa ja täten vähentää vinoumia, sillä sen ei ajatella vaativan käyttäjää toimimaan sosiaalisesti toivotulla tavalla.[3]

Vuorovaikutuksen muoto riippuu sovellusalueesta. Esimerkiksi videon katseluaika tai verkkosivulla vietetty aika voidaan molemmat tulkita kulutussignaaleiksi. Mainosalustoilla ollaan useammin kiinnostuneita Click-Through Rate -arvon (CTR) maksimoimisesta, eli siitä, kuinka suuri osuus näytetyistä kohteista saa käyttäjältä klikkauksen.[2] CTR:ää voidaan soveltaa myös verkkokauppaympäristöön, koska yleensä käyttäjä klikkailee juuri niitä tuotteita, joista on eniten kiinnostunut.

Suositusjärjestelmän toimintaa voidaan siis havainnollistaa prosessina, jossa järjestelmä suosittelee kohteita käyttäjille hyödyntäen dataa joko käyttäjistä, kohteista, molemmista tai käyttäjän ja kohteen välisestä vuorovaikutuksesta [2]. Yleisesti suositusjärjestelmän tekemä analyysi perustuu aikaisempaan käyttäjä-kohdevuorovaikutukseen, koska aikaisemmat mielenkiinnonkohteet ja käyttäytymistäipumukset ovat hyviä osoittimia mahdollisille tulevaisuuden valinnoille [6].

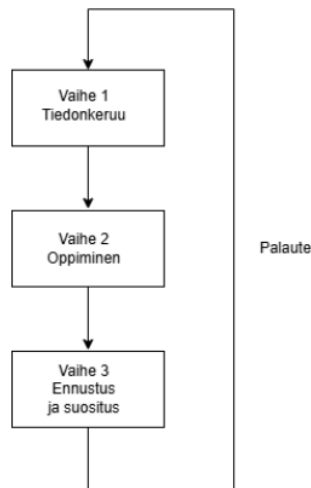
2.1 Suosituksen synty

Toteuttaakseen tärkeimmän tehtävänsä eli suosituksen käyttäjälle suosittelujärjestelmän täytyy ennustaa, onko kohde suosittelun arvoinen. Tämän saavuttamiseksi järjestelmän on kyettävä arvioimaan joidenkin kohteiden hyötyä käyttäjälle tai ainakin vertailemaan kohteiden hyötyarvoja keskenään, ja tämän vertailun perusteella päättämään, mitä kohteita suositellaan käyttäjälle.[5]

Suositusjärjestelmän kehittämisessä on tärkeää pohtia useita peruskysymyksiä. Ensinnäkin on huomioitava, (1) millaista dataa on saatavilla järjestelmälle. Usein ei ole itsestään selvää, mitä tietoja tulisi hyödyntää ja miten käsitellä puuttuvia tai vaikeasti saatavia datan osia. Toiseksi on tärkeä tarkastella (2) miten käyttäjän toiminta tulkitaan. Esimerkiksi verkkokaupassa lisää ostoskoriin-toiminto saattaa olla merkitykseltään erilainen kuin itse osta-toiminto. Samalla tavoin on tarpeen pohtia

(3) millaista vuorovaikutusta järjestelmä pyrkii optimoimaan. Tavoitellaanko esimerkiksi katseluajan vai klikkausprosentin maksimoimista? Viimeisenä on olennaista määritellä (4) miten varsinainen tehtävä muotoillaan. Oppimistrategiat voivat erota sen mukaan, pyritäänkö algoritmilla mallintamaan käyttäjäkohtainen funktio, joka kuvaa käyttäjän kiinnostusta eri kohteisiin, vai halutaanko muodostaa lista kohteista, joista käyttäjä todennäköisesti on kiinnostunut (hakutehtävä). Lisäksi myös sovelluskohtaiset vaatimukset voivat vaihdella. Joissain tapauksissa riittää, että yksi suositus on sopiva, kun taas toisissa pyritään tarjoamaan mahdollisimman monta osuvaa ehdotusta, vaikka tarkkuus heikentyisi samalla.[2]

Valmis suositusjärjestelmä sisältää useita vaiheita, jotka voidaan karkeasti jakaa Kuvan 2.1 mukaisesti kolmeen vaiheeseen: tiedonkeruuvaihe, oppimisvaihe ja ennustusvaihe. Suositusjärjestelmä toimii ohjausjärjestelmän tavoin, sillä nykyisen suosituksen tulos vaikuttaa merkittävästi tuleviin suosituksiin palautemekanismin kautta.[3]



Kuva 2.1: Suosituksen vaiheet

Tiedonkeruuvaiheessa järjestelmä kerää olennaista ja tarkkaa tietoa käyttäjistä käyttäjäprofiilin luomiseksi. Profiili sisältää tietoa käyttäjän ominaisuuksista, käyttäytymisestä ja vuorovaikutussuhteista kohteiden kanssa. Suositus on hyvin riippuvainen tästä vaiheesta, sillä sen tuottama data-aineisto määrittää pitkälti ennusteiden tarkkuuden.[3]

Tähän vaiheeseen liittyy olennaisesti vuorovaikutus ja ylempänä mainitut palautteen tyypit, suora ja epäsuora palaute. Näistä kahdesta palautetyypistä voidaan muodostaa myös hybridipalaute, jossa palautetta kerätään sekä epäsuorana datana että suorina käyttäjien antamina palautteina.[7]

Oppimisvaiheessa suositusjärjestelmä soveltaa koneoppimisalgoritmeja käyttäjäprofiileista koostuvan aineiston suodattamiseen ja mallin kouluttamiseen. Mallille opetetaan käyttäjän mieltymykset tarkastelemalla käyttäjän ja eri kohteiden välistä vuorovaikutusta sekä ryhmittelemällä samankaltaisia käyttäjiä ja kohteita. Tämän vaiheen tuloksena muodostuu erilaisia käyttäjä- ja kohderyhmiä.[3] Koneoppimisalgoritmit ovat menetelmiä, joiden avulla voidaan tunnistaa ja mallintaa aineistosta nousevia kuvioita tai rakenteita, joita voidaan hyödyntää erilaisissa sovellustilanteissa [7].

Ennustusvaiheessa Järjestelmä tuottaa suosituksia hyödyntämällä oppimisvaiheessa koulutettua mallia. Mallin tehtävänä on ennustaa, kuinka todennäköisesti käyttäjä pitää tietystä kohteesta. Se laskee käyttäjän mieltymyksen todennäköisyyden, ja ennalta määritetyn kynnsarvon perusteella päätetään, suositellaanko kohdetta käyttäjälle vai ei.[3]

2.2 Merkitys sähköisessä kaupankäynnissä

Internet on avannut valtavan määrän uusia liiketoimintamahdollisuuksia ja mahdollistaa palveluiden jatkuvan saatavuuden. Verkkokauppojen kasvaessa ja kehittyessä yhä laajemmiksi kokonaisuuksiksi on syntynyt tarve asiakkaille tarjottavista suosituksista, jotka perustuisivat koko saatavilla olevaan valikoimaan vaihtoehtoja [5]. Digitaalisen datan valtava määrä ja internetin käyttäjämäärien lisääntyminen luovat merkittävän haasteen tietotulvan hallintaan, mikä taas vaikeuttaa käyttäjien pääsyä haluamaansa tuotteeseen tai tietoon verkossa [8].

Ongelma ratkaistiin kehittämällä suositusjärjestelmät ohjeistamaan verkkokaupan asiakkaita. Kuten aiemmin on todettu, suositusjärjestelmien keskeisimpänä tehtävänä pidetään tavoitetta tarjota käyttäjälle yksilöllisiä suosituksia [3]. Verkkokauppojen järjestelmät oppivat käyttäjän toiminnasta ja suosittelevat tälle relevantteja tuotteita, mikä personoi käyttäjäkokemuksen ja lisää käyttäjän sitoutuneisuutta verkkokauppaan.

Suositusjärjestelmät voivat parantaa verkkokaupan myyntiä kolmella keskeisellä tavalla. Usein verkkokaupan kävijät vain selailevat tuotteita ilman ostoa. Jos sivusto kuitenkin näyttää käyttäjälle tälle osuvia suosituksia, oston todennäköisyys kasvaa. Suositustekniikat voivat myös ehdottaa lisätuotteita käyttäjälle sen tuotteen lisäksi, jota hän on jo ostamassa. Tällaista menetelmää kutsutaan ristiinmyynniksi, ja sen suurin tavoite on kasvattaa keskimääräistä tilauksien arvoa ajan myötä. Lisäksi nykyisessä kilpailutilanteessa, jossa kilpailevalle sivustolle siirtyminen tapahtuu muutamalla klikkauksella, asiakasuskollisuus on erittäin tärkeää. Suositusjärjestelmien mahdollistama personointi vahvistaa käyttäjän ja sivuston välistä suhdetta. Mitä enemmän asiakas käyttää verkkokauppaa ja sitoutuu siihen, sitä paremmin järjestelmä oppii käyttäjän mieltymykset ja tuottaa tälle osuvampia suosituksia.[9]

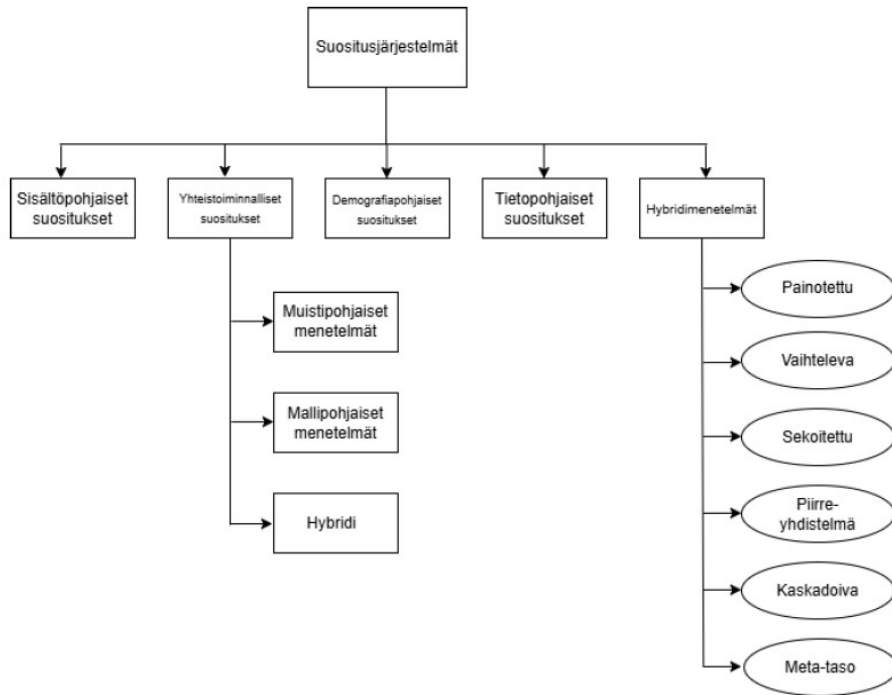
Vaikka verkkokauppojen suositusjärjestelmien voidaan todeta mullistaneen sähköisen kaupankäynnin, menestyksestään huolimatta näissä järjestelmissä on myös

useita haasteita, jotka estävät niiden optimaalisen suorituskyvyn. Yksi suurimmista haasteista on datan harvuus, jossa käyttäjä-kohde-vuorovaikutusten vähäisyys estää järjestelmää arvioimasta käyttäjän mieltymyksiä oikein, erityisesti uusien tai harvinaisten kohteiden kohdalla. Tätä ongelmaa pahentaa niin kutsuttu kylmäkäynnistysongelma, joka vaikeuttaa yksilöllisten suositusten tarjoamista täysin uusille käyttäjille tai jättää suosittelematta kohteita, joilla on vähän historiatietoja. Lisäksi verkkokaupoissa skaalautuvuutta voidaan pitää keskeisenä haasteena, sillä suositusjärjestelmien tulee käsitellä jatkuvasti kasvavaa tietomäärää.[10] Yllä nimettyihin haasteisiin perehdytään luvussa 4.

3 Järjestelmien luokittelu

Vaikka suositusjärjestelmiä on tutkittu pitkään ja laajasti, ja niitä pidetään yleisesti merkittävänä koneoppimisen ja informaation suodatuksen alaluokkana, ei niillä todeta olevan varmaa ja vakiintunutta luokittelua. Suositusjärjestelmän perusidea, eli käyttäjälle hyödyllisten ja kiinnostavien kohteiden suosittelu, on intuitiivisesti helppo hahmottaa, mutta sitä tarkempi ja systemaattisempi määrittely kohtaa nopeasti haasteita. Suositusjärjestelmien toimintaperiaatteet, sovellusalueet ja käytettävä data vaihtelevat merkittävästi, mikä tekee yhtenäisen luokittelun muodostamisesta vaikeaa.[2]

Tässä luvussa pyritään muodostamaan suositusjärjestelmille luokittelu, joka antaisi hyvän yleiskuvan mahdollisista tekniikoista. Kirjallisuudessa perinteisenä pidetty luokittelu jakaa suositusjärjestelmät kolmeen eri kategoriaan, jotka ovat yhteistoiminnalliset suositukset (engl. collaborative recommendations), sisältöpohjaiset suositukset (engl. content-based recommendations) ja hybridimenetelmät (engl. hybrid approaches)[1]. Kuvassa 3.1 esitetty suositusjärjestelmien luokittelu tarjoaa hyödyllisen yleiskuvan keskeisimmistä menetelmistä, vaikka se ei täysin kuvaa nykyaikaisten järjestelmien monimuotoisuutta [2].



Kuva 3.1: Suositusjärjestelmien luokittelu

3.1 Yhteistoiminnalliset suositusjärjestelmät

Termi "yhteistoiminnallinen suodatus" esiteltiin ensimmäisen kerran vuonna 1992 Goldbergin ym. toimesta. He esittivät, että informaation suodatusprosessi tehostuu, kun siihen osallistuu ihmisiä [1],[11]. Yhteistoiminnallisessa suodatuksessa suositukset käyttäjälle tuotetaan vertailemalla käyttäjän mieltymyksiä muiden käyttäjien mieltymyksiin, jotka ovat arvioineet samoja kohteita samalla tavalla kuin käyttäjä itse [1]. Menetelmä perustuu ajatukseen, että samoista asioista kiinnostuneet käyttäjät ovat taipuvaisia valitsemaan samanlaisia kohteita [12]. Käytetyt samankaltaisuusmittarit vaihtelevat, mutta ne perustuvat aina menneisiin vuorovaikutuksiin ja niistä annettuun palautteeseen.[2]

Menetelmä hyödyntää käyttäjien mieltymyksistä ja arvosteluista koostuvaa tietojoukkoa muodostaakseen käyttäjä-kohde-matriisin, josta esimerkki on esitetty kuvassa 3.2. Kuvassa on matriisi $m \times n$, jossa m on käyttäjien lukumäärä $(U_1, U_2, U_3, \dots, U_m)$ ja n on kohteiden lukumäärä $(I_1, I_2, I_3, \dots, I_n)$, jotka käyttäjät ovat arvostelleet. Solun arvo $r_{i,u}$, matriisissa on kohteen u arvostelu, jonka antoi käyttäjä i . Arvostelu voi olla joko epäsuoraa tai suoraa (käyttäjän palaute asteikolla k). Yhteistoiminnallisen suositusjärjestelmän lopputulos voi olla kahdenlainen. Ensimmäinen vaihtoehto on ennuste $r_{i,u}$, joka kuvaa käyttäjän i arvioitua mieltymystä kohteeseen u . Toinen vaihtoehto on suosituslista, joka sisältää N kohdetta, joista käyttäjän arvioidaan todennäköisimmin pitävän.[12]

	I_1	I_2	I_3	I_4	\dots	I_n
U_1						
U_2						
U_3						
U_4						
\dots						
U_m						

	= Suora palaute (esim. arvostelu, joukko {1, 2, 3, 4, 5})
	= Epäsuora palaute (esim. klikkaus)
	= Tuntematon arvo

Kuva 3.2: Käyttäjä-kohde-matriisi

Yhteistoiminnalliset tekniikat voidaan jaotella tarkemmin kolmeen alaluokkaan: muistipohjaiset menetelmät, mallipohjaiset menetelmät ja hybridimallit.

Muistipohjaisista menetelmistä tunnetuin on naapuripohjainen suodatus (engl. neighborhood-based filtering). Siinä käyttäjän mieltymyksiä ennustetaan hyödyntä-

mällä samankaltaisia käyttäjiä tai kohteita, jotka ovat lähellä käyttäjää tai kohdetta vertailuavaruudessa. Menetelmän tarkkuus ja tehokkuus riippuvat siitä, miten käyttäjien tai kohteiden välinen samankaltaisuus määritellään. Muistipohjaiset menetelmät jaetaan edelleen kahteen päätyyppiin, jotka ovat käyttäjäpohjainen suodatus (engl. user-based filtering) ja kohdepohjainen suodatus (engl. item-based filtering). Käyttäjäpohjaisessa suodatuksessa arvioidaan käyttäjien välistä samankaltaisuutta vertaamalla heidän antamiaan arvosteluja samoille kohteille ja käyttämällä tätä tietoa ennusteiden laskemiseen käyttäjän mieltymyksistä. Kohdepohjaisessa suodatuksessa taas tarkastellaan kohteiden välisiä suhteita: järjestelmä hakee kaikki aktiivisen käyttäjän arvioimat kohteet, määrittää näiden kohteiden samankaltaisuuden kohteeseen, jolle ennustetta halutaan, ja valitsee tämän jälkeen N samankaltaisinta kohdetta arvioidakseen, kuinka todennäköisesti käyttäjä pitää uudesta kohteesta.[12]

Mallipohjaiset menetelmät hyödyntävät tiedonlouhinnan ja koneoppimisen tekniikoita. Näihin menetelmiin lukeutuvat esimerkiksi assosiaatiosäännöt, klusterointi, neuroverkot, päätöspuut, regressiomallit, Bayes-luokittelu ja linkkianalyysi.[12]

Hybridimallit yhdistelevät muistipohjaisten ja mallipohjaisten menetelmien vahvuudet tavoitteena yleensä parempi ennustustarkkuus ja tehokkuus.[12]

3.2 Sisältöpohjaiset menetelmät

Ennustavat käyttäjän ja kohteen välisen yhteensopivuuden ominaisuuksien perusteella [2]. Ne hyödyntävät käyttäjäprofileja, jotka muodostetaan kohteiden ominaisuuksista, joiden kanssa käyttäjä ollut vuorovaikutuksessa [3]. Sisältöpohjainen suodatus perustuu kahteen vaiheeseen, joista ensimmäisessä käyttäjäprofilissa olevia ominaisuuksia käytetään muiden samanlaisten kohteiden löytämiseen ja toisessa näitä ominaisuuksia verrataan keskenään ja suositellaan käyttäjälle niitä kohteita, joiden samankaltaisuus on suurin [12],[13]. Samankaltaisuus voidaan todeta esimerkiksi Naive Bayes -luokittelijan, päätöspuiden tai kosinietäisyyden avulla [3].

Sisältöpohjaiset menetelmät ovat yksilökeskeisiä, koska ne perustavat suosituksen yhden käyttäjän profiliin ilman muiden käyttäjien tietojen hyödyntämistä [3]. Ne ovat luonnollisin ja yksinkertaisin tapa tuottaa suosituksia, sillä käyttäjältä ei vaadita erillistä palautetta. Jo yksi käyttäjän aiemmin osoittama mieltymys voi riittää, jotta järjestelmä pystyy suosittelemaan tälle useita samankaltaisia kohteita.[12]

3.3 Muita suositusjärjestelmiä

Kolmen pääluokan vierelle on myöhemmin esitetty muutamaa lisäluokkaa. Nämä luokat ovat demografiapohjaiset ja tietopohjaiset järjestelmät [2]. Näille menetelmille on ominaista se, että ne tukeutuvat voimakkaasti metadataan silloin, kun käyttäjä–kohde–vuorovaikutusdataa ei ole saatavilla tai sitä on hyvin vähän.

Tietopohjaisen suositusjärjestelmät (engl. knowledge-based Recommendation Systems) hyödyntävät eksplisiittistä, asiantuntijatasoista metadattaa käyttäjistä sekä kohteista. Järjestelmä pyrkii kuvaamaan, miten hyvin kohteen ominaisuudet vastaavat käyttäjän vaatimuksia, ja toimii usein sääntöjen tai asiantuntijatietoon perustuvien päätösmallien avulla.[12]

Demografiset suositusjärjestelmät (engl. demographic-Based Recommendation Systems) puolestaan hyödyntävät demografisia tietoja eli käyttäjämadataa, kuten ikää, sukupuolta, asuinpaikkaa ja ammattia, suositusten muodostamisessa. Menetelmä perustuu oletukseen, että samankaltaisiin demografisiin ryhmiin kuuluvilla käyttäjillä on usein myös samankaltaisia mieltymyksiä. Tästä syystä demografiapohjaiset järjestelmät soveltuvat hyvin tilanteisiin, joissa vuorovaikutusdataa ei ole vielä kertynyt, mutta käyttäjästä on saatavilla perustason taustatietoa. Menetelmän rajoitteena on kuitenkin taipumus stereotypioihin, sillä se olettaa demografisten ryhmien olevan sisäisesti homogeenisia, mikä ei välttämättä aina pidä paikkaansa.[14]

3.4 Hybridipohjaiset suositusjärjestelmät

Järjestelmät yhdistävät kaksi tai useampia menetelmiä parantaakseen suorituskykyä ja vähentääkseen yksittäisten tekniikoiden rajoitteita [9]. Kuten Kuvassa 3.1 esitetään, hybridimenetelmät voidaan jakaa seuraaviin tyyppeihin:

Painotetuissa (engl. weighted) menetelmissä menetelmien tuottamat arviot yhdistetään yleensä lineaarisesti. Suositellijoille asetetaan painoarvot, joita säädetään sen perusteella, miten hyvin menetelmä vastaa käyttäjän todellisia arvioita.[15]

Vaihtelevissa (engl. switching) menetelmissä järjestelmä valitsee kulloinkin parhaiten soveltuvan menetelmän. Esimerkiksi sisältöpohjaisesta menetelmästä voidaan vaihtaa yhteistoiminnalliseen, jos suositus ei ole riittävän luotettava.[15]

Sekoitetuissa (engl. mixed) menetelmissä useiden menetelmien tulokset esitetään rinnakkain, esimerkiksi samassa tuotelistassa, ilman että niitä yhdistetään.[15]

Piirre-yhdistelmä (engl. feature combination) menetelmissä eri menetelmien tuottamia ominaisuuksia yhdistetään samaan malliin, kuten yhteistoiminnallisen suodatuksen piirteitä sisältöpohjaiseen malliin.[15]

Kaskadoivissa (engl. cascade) menetelmissä menetelmät suoritetaan hierarkkisesti: ensisijainen menetelmä tuottaa suositukset, ja toissijainen tarkentaa niitä muuttamatta alkuperäistä tulosta.[15]

Piirre-laajennus (engl. feature augmentation) menetelmissä ensimmäisen menetelmän tuottamat piirteet lisätään suoraan toisen menetelmän syötteenä.[15]

Meta-tason (engl. meta-level) menetelmissä toisen menetelmän syötteenä toimii kokonainen ensimmäisen menetelmän oppima malli, ei vain yksittäisiä piirteitä.[15]

Hybridimenetelmien keskeinen etu on se, että ne yhdistävät eri suositelutekniikoiden vahvuudet ja kompensoivat heikkouksia, kuten kylmäkäynnistys- ja datan harvuusongelmia.[6]

4 Tekniset ongelmat ja ratkaisut

Tässä luvussa käsitellään verkkokauppojen suositusjärjestelmien yleisimpiä haasteita, joihin lukeutuvat kylmäkäynnistys, datan harvuus, skaalautuvuus ja monimuotoisuus. Lisäksi tarkastellaan mahdollisia ratkaisumenetelmiä.

4.1 Kylmäkäynnistys

Kylmäkäynnistys terminä on peräisin automaailmasta: moottorin ollessa kylmä sen käynnistäminen on vaikeaa, mutta saavuttaessa optimaalisen lämpötilan, se toimii ongelmitta. Sama ongelma esiintyy myös suositusjärjestelmissä. Kun käytettävissä ei ole riittävästi tietoa, järjestelmä ei toimi optimaalisesti. Ongelmaa esiintyy sekä kohteilla että käyttäjillä [14]. Kun uusi kohde lisätään verkkokauppaan, sillä ei ole vielä arvosteluja, koska käyttäjäinteraktiota ei ole. Tätä havainnollistaa Kuvan 3.2 käyttäjä-kohde-matriisin kohde U_3 . Jos käyttäjien vuorovaikutusta ei ole riittävästi, järjestelmä ei tiedä, milloin suositella kohdetta oikein. Kun käyttäjä luo tilin ensimmäistä kertaa verkkokauppaan, ei hänellä ole aiempia mieltymyksiä tai historiaa, johon tulevat suositukset voitaisiin perustaa. Kuvan 3.2 käyttäjä-kohde-matriisissa käyttäjä I_1 kärsii kylmäkäynnistysongelmasta. Ongelmaa esiintyy myös jo olemassa oleville käyttäjille. Kun käyttäjä on tehnyt ostopäätöksen tietystä tuoteryhmästä, hänen kiinnostuksensa kyseiseen kategoriaan tyypillisesti vähenee. Suositusjärjestelmä ei kuitenkaan välttämättä tunnista tätä muutosta, mikä johtaa tilanteeseen, jossa järjestelmä jatkaa vanhaan tietoon perustuvia suosituksia, vaikka käyttäjän

kiinnostus on jo siirtynyt muualle. Kuten Kuvan 3.2 matriisissa käy ilmi, järjestelmä tallentaa vain aiemmat interaktiot eikä käyttäjän muuttuneita tarpeita. Vanha käyttäjä on vastaavassa tilanteessa kuin uusi käyttäjä: järjestelmällä ei ole tarpeeksi tietoa käyttäjän mieltymyksistä.[9]

Kylmäkäynnistysongelman suurin haaste on tiedonpuute. Ehdotetut ratkaisumenetelmät pyrkivät täyttämään tämän vajeen valmiiksi saatavilla olevan metadatan avulla. Yksinkertaisimmin ongelma voidaan ratkaista laittamalla käyttäjä arvostelemaan joitakin tuotteita tai kuvailemaan omia mieltymyksiään heti tämän liittyessä verkkokaupan asiakkaaksi. Tällöin järjestelmä saa käyttöönsä heti luotettavaa dataa, jonka pohjalta tehdä suosituksia [16]. Koska jotkin suositusjärjestelmät, kuten demografia- ja tietopohjaiset menetelmät, hyödyntävät vahvasti metadataa ja toimivat paremmin vähäisen datan tilanteissa, voidaan kylmäkäynnistysongelmaa lievittää hyödyntämällä hybridimalleja. Yhteistoiminnallinen suositusjärjestelmä kärsii väistämättä kylmäkäynnistyksestä käyttäjien, kohteiden tai molempien osalta. Ongelmaa voidaan lieventää yhdistämällä järjestelmään demografia- tai sisältöpohjaisia menetelmiä. Näin yhteistoiminnallinen suositusjärjestelmä saa käyttöönsä lisäinformaatiota eli metadataa käyttäjistä tai kohteista, mikä mahdollistaa suositusten tuottamisen jo ennen käyttäjäinteraktioiden kertymistä.[6]

Lisäksi yksi varhaisista metadatapohjaisista menetelmistä on Naive Bayes -luokittelijan käyttö. Menetelmä soveltuu hyvin niukasti havaintoja sisältäviin tilanteisiin, koska se pystyy tekemään ennusteita pelkkien kohteen kuvailevien ominaisuuksien perusteella. Näin järjestelmä voi arvioida esimerkiksi uuden tuotteen todennäköisen kategorian tai käyttäjäryhmän, jolle se todennäköisimmin sopii, vaikka tuotteella ei vielä olisi aiempia arvioita tai klikkauksia.[17] Kylmäkäynnistystä voidaan käsitellä myös mallipohjaisilla menetelmillä, kuten WALSOptimoinnilla (Weighted Alternating Least Squares). Sen etuna on, että uusi kohde voidaan liittää olemassa olevaan malliin ilman koko matriisihajoituksen uudelleenlaskutusta.

Käyttäjistä aiemmin opitut latentit edustukset pidetään muuttumattomina, ja järjestelmä oppii uuden kohteen esityksen nopeasti muutaman ensimmäisen käyttäjävuorovaikutuksen perusteella. Tämä mahdollistaa sen, että uudet tuotteet pääsevät mukaan suositteluihin jo varhaisessa vaiheessa. Metadataan perustuvia heuristisia menetelmiä voidaan hyödyntää erityisesti silloin, kun käyttäjä- tai kohdehistoriaa ei ole lainkaan. Tällöin suosituksia voidaan tuottaa arvioimalla uuden kohteen kiinnostavuutta sen kategorian, kuvauksen, hintaluokan tai muiden metadataan perustuvien ominaisuuksien perusteella. Esimerkiksi uudelle käyttäjälle voidaan suositella hänen ilmoittamiensa kiinnostusalueiden mukaisia kategorioita tai yleisesti suosittuja tuotteita kyseisiltä alueilta, vaikka aiempaa ostohistoriaa ei olisi.[9]

4.2 Datan harvuus

Datan harvuus johtuu tilanteesta, jossa käyttäjät arvioivat vain pienen osan kaikista kohteista [9]. Useimmat suositusjärjestelmät ryhmittelevät samankaltaisten käyttäjien suositukset: käyttäjä-kohde-matriisissa on kuitenkin tyhjiä tai tuntemattomia arvoja (jopa 99%)[18] kuten Kuvan 3.2 käyttäjä-kohde-matriisissa havainnollistetaan. Näin ollen järjestelmä voi tuottaa perustelemattomia suosituksia käyttäjille, jotka eivät anna arvioita tai palautetta. Esimerkiksi jos verkkokirjakaupassa on kaksi miljoonaa eri kirjaa ja suuri määrä käyttäjiä, jokainen käyttäjä voidaan kuvata 2 miljoonan elementin matriisilla, jossa jokainen arvo vastaa käyttäjän antamaa arvosanaa tietylle kirjalle [9]. Kun käyttäjien ja kohteiden määrä kasvaa suureksi, käyttäjä-kohde-matriisi muuttuu väistämättä hyvin harvaksi, eli suurin osa sen alkiosta on tyhjiä tai tuntemattomia.

Monet menetelmät pyrkivät lieventämään ongelmaa mallintamalla käyttäjien mieltymykset tämän käyttäytymisen ja luotettujen sosiaalisten yhteyksien perusteella [9]. Tässä yhteydessä luotetuilla sosiaalisilla yhteyksillä tarkoitetaan muiden käyttäjien arvioita, joihin käyttäjä voi perustaa oman päätöksensä. Luottamusta on

käytetty laajasti suositusjärjestelmien vahvistamiseen [19]. Luottamus määritellään uskona siihen, että muut käyttäjät antavat tarkkoja arvioita (suoria tai epäsuoria). Yksi ehdotettu luottamukseen perustuva lähestymistapa on merge-lähestymistapa [20]. Siinä otetaan huomioon käyttäjän läheiset naapurit eli mahdollisimman vastaavat käyttäjät ja pyritään parantamaan suositusjärjestelmien yleistä ennustustarkkuutta. Tarkemmin käyttäjän ja tämän luotettavan naapurin antamat arviot yhdistetään laskemalla niiden keskiarvo usein arvioituista kohteista käyttäjän ja naapurin välisen samankaltaisuuden perusteella.[9]

Ongelma voidaan ratkaista myös hybridinäkökulmasta. Esimerkiksi pelkän sisältöpohjaisen suositusjärjestelmän sijaan voimme yhdistää sisältöpohjaisen ja yhteistoiminnallisen järjestelmän hybridimalliksi, joka ratkaisee matriisin harvuutta.[14]

4.3 Skaalautuvuus

Skaalautuvuus on noussut merkittäväksi ongelmaksi verkkokauppojen nopean kasvun vuoksi. Suositusjärjestelmien on nykyaikana kyettävä tuottamaan tuloksia nopeasti laajamittaisissa sovelluksissa. Useiden mahdollisten naapureiden löytäminen ei enää riitä vaan verkkokaupat vaativat tarkempia tuloksia vielä suuremmilta hakualueilta [9]. Algoritmit kohtaavat suorituskykyyn liittyviä ongelmia erityisesti silloin, kun käyttäjistä on valtavasti tietoa [20]. Jos esimerkiksi verkkosivulla on kymmenen tuhatta havaintoa tietystä käyttäjistä, algoritmilla voi olla vaikeuksia löytää käyttäjää vastaavia käyttäjiä eli läheisiä naapureita tälle.

Klusterointitekniikat tarjoavat ratkaisun skaalautuvuuteen jakamalla käyttäjät tai kohteet ryhmiin eli klustereihin siten, että saman klusterin jäsenet ovat keskenään samankaltaisia. Näin jokainen klusteri toimii omana rajattuna “naapurustonaan”, jonka sisällä suosituksia tuotetaan. Kun haku rajataan koko tietomassan sijaan vain klusterin jäseniin, sekä vertailtavan datan määrä että laskennallinen työ pienenevät merkittävästi. [21] Tämä tekee ennusteiden tuottamisesta nopeampaa ja

mahdollistaa suositusjärjestelmän toiminnan suurissa ja jatkuvasti kasvavissa tietokokonaisuuksissa.[9]

4.4 Monimuotoisuus ja tarkkuus

Suositusjärjestelmät voivat tehdä joko samankaltaisia tai monipuolisia suosituksia. Kun suositusjärjestelmä optimoi liiaksi tarkkuutta, seurauksena on liiallinen erikoistuminen (engl. over-specialization), joka ilmenee erityisesti sisältöpohjaisissa järjestelmissä. Tällöin suositukset jäävät kapeiksi ja ennalta arvattaviksi: järjestelmä ehdottaa vain käyttäjän jo tuntemien kohteiden lähellä olevia vaihtoehtoja. Tämä vähentää mahdollisuuksia löytää uusia tai monipuolisia kohteita, mikä voi olla haitallista myös kaupallisille tavoitteille ja heikentää käyttäjäkokemusta.[22] Suositusjärjestelmän monimuotoisuus mahdollistaa myös sellaisten kohteiden löytymisen, joita käyttäjä ei välttämättä löytäisi yksin.[9]

Monimuotoisuutta suosituksissa voidaan arvioida kahdella mittarilla: yllättävyydellä ja personalisoinnilla. [23] Yllättävyys kuvaa järjestelmän kykyä tuottaa odottamattomia, ennalta arvaamattomia suosituksia, kun taas personalisointi mittaa eri käyttäjien saamien suosituslistojen keskinäistä erilaisuutta eli käyttäjien välistä monimuotoisuutta.[9] Liiallista erikoistumista voidaan vähentää hyödyntämällä naapureihin perustuvaa yhteistoiminnallista suodatusta. [14] Todennäköisyyspohjaisessa naapurivalinnassa käytetään menetelmää, joka valitsee k naapuria painotetun otannan kautta ottaen huomioon käyttäjien tai kohteiden samankaltaisuudet mahdollisten naapureiden kanssa.[22]

5 Järjestelmien arviointi

Tässä luvussa tarkastellaan suositusjärjestelmien arviointimittareita, joiden avulla voidaan tunnistaa järjestelmän toiminnan kannalta keskeiset ominaisuudet ja mitata niitä. Koska suositusalgoritmit eroavat toisistaan sekä toimintalogiikaltaan että käyttökonektiltaan, myös niiden arviointiin soveltuvien mittareiden valinta on kontekstisidonnainen eikä yksiselitteinen [24].

Perinteisesti suositusjärjestelmiä arvioidaan kahdella tavalla riippuen datan luonteesta. Luokittelumittarit perustuvat binääriseen dataan eli tietoihin siitä, tapahtuiko jokin käyttäjän toiminto (esimerkiksi ostaminen tai klikkaaminen). Tällöin suositusjärjestelmä nähdään luokittelutehtävänä, jossa arvioidaan mallin kykyä erottaa käyttäjälle relevantti ja ei-relevantti sisältö [25]. Regressiomittarit puolestaan perustuvat ordinaaliseen dataan, esimerkiksi arvosteluihin, joissa käyttäjä ilmaisee mieltymyksensä asteikolla.

Binäärinen arviointi (luokittelu) olkoon O havaintojen joukko. Binäärinen kohdemuuttuja on funktio $y : O \rightarrow \{0, 1\}$, jossa arvo $y = 1$ kuvaa kiinnostavaa luokkaa (esim. käyttäjä klikkasi/osti tuotteen). Binäärinen ennustemalli on funktio $\hat{y} : O \rightarrow \{0, 1\}$, joka on oppimisalgoritmin tuottama arvio kohdemuuttujasta. Seka-

vuusmatriisi laskee todellisten arvojen y ja mallin ennusteiden \hat{y} ristiintaulukoinnin:

	$\hat{y} = 0$	$\hat{y} = 1$
$y = 0$	a	b
$y = 1$	c	d

jossa sarakkeet kuvaavat mallin ennusteita ja rivit todellisia arvoja.[26] Matriisissa määritellään neljä mahdollista lopputulosta eli (a) totet negatiiviset eli ei suositeltu- ja eikä hyviä kohteita, (b) väärät positiiviset eli ei-relevantit suositukset, (c) väärät negatiiviset eli relevantteja kohteita, joita ei suositeltu ja (d) totet positiiviset eli onnistuneet suositukset [9]. Kohteiden kokonaismäärä: $n = a + b + c + d$.

Suositusjärjestelmää voidaan arvioida sekavuusmatriisista johdetuilla mittareilla. Näitä keskeisiä mittareita ovat tarkkuus (eng. accuracy), täsmällisyys (eng. precision) ja muistettavuus (eng. recall). Tarkkuus mittaa mallin oikein tekemien ennusteiden suhteellista osuutta ja se voidaan laskea kaavalla:

$$Tarkkuus = \frac{a + d}{a + b + c + d}$$

Täsmällisyys taas mittaa, kuinka suuri osa järjestelmän suosittlemista kohteista on käyttäjän mielestä myös relevantteja ja voidaan laskea seuraavasti:

$$precision = \frac{d}{(b + d)}$$

Muistettavuus mittaa, kuinka suuren osan kaikista käyttäjälle sopivista kohteista järjestelmä onnistuu suosittelemaan tälle ja lasketaan

$$recall = \frac{d}{(c + d)}$$

Hyvä suositusjärjestelmä pyrkii optimoimaan kaikki mittarit samanaikaisesti. Esimerkiksi suuri suositeltujen tuotteiden määrä voi kasvattaa muistettavuutta, mutta samalla heikentää tarkkuutta, jos mukana on paljon ei-relevantteja kohteita.[9]

F-mitta määritellään:

$$F\text{-measure} = \frac{2PR}{P + R}$$

Se yhdistää täsmällisyyden (P) ja muistettavuuden (R) yhdeksi luvuksi ottamalla niiden harmonisen keskiarvon, mikä painottaa erityisesti tilanteita, joissa jompikumpi arvo on matala. Näin F-mitta kuvaa tasapainoa kahden eri näkökulman välillä: kuinka monta löydetystä osumista on oikein (täsmällisyys) ja kuinka suuri osa kaikista relevanttista osumista löydetään (muistettavuus). Saatu yksittäinen arvo helpottaa järjestelmien vertailua erityisesti silloin, kun täsmällisyys ja muistettavuus ovat keskenään ristiriidassa.[27]

Regressiopohjainen arviointi (arvosanaennustaminen) tilastolliset menetelmät arvioivat suositusjärjestelmän tarkkuutta vertaamalla järjestelmän ennustamia arvosteluja käyttäjän oikeisiin arvosteluihin. Mean Absolute Error (MAE), Root Mean Square Error (RMSE) ja korrelaatio ovat yleisimmin käytettyjä mittareita:[27]

MAE lasketaan seuraavanlaisesti:

$$\text{MAE} = \frac{1}{N} \sum_{ui} |p_{ui} - r_{ui}|,$$

jossa p_{ui} on ennustettu arvostelu käyttäjälle u kohteesta i . r_{ui} on käyttäjän oikea arvostelu ja N kohdejoukon arviointien kokonaismäärä. Matalampi MAE-arvo kertoo suositusten hyvästä tarkkuudesta. RMSE lasketaan seuraavanlaisesti:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{ui} (p_{ui} - r_{ui})^2},$$

ja se huomio MAE-arvoa paremmin absoluuttisen virheen mahdollisuuden. Mitä matalampi RMSE-arvo on, sitä parempi suosituksen tarkkuus.[27]

Käytännön arviointimenetelmät edellä esitetyt mittarit kuvaavat järjestelmän teknistä suorituskykyä. Käytännössä suositusjärjestelmien arviointi täydentyy käyttäjäperustaisilla ja reaaliaikaisilla menetelmillä.

Vaikka tarkkuuden ja peittävyuden, eli järjestelmän kyvyn hyödyntää mahdollisimman suuren osan koko järjestelmän kohde- ja käyttäjäavaruudesta [5], ajatellaan olevan suositusjärjestelmien arvioinnin keskeisimpiä tekijöitä, monet toissijaiset tekijät, kuten uutuusarvo, luottamus ja sattumanvarainen löytämisen ilo (serendipiteetti), ovat hyvin merkittäviä käyttäjäkokemuksen kannalta. Niillä on lisäksi todettu olevan huomattavia sekä lyhyen että pitkän aikavälin vaikutuksia konversioprosentteihin. Konversioprosentti on verkkokaupan keskeinen mittari, joka kuvaa, kuinka suuri osa käyttäjistä suorittaa halutun toiminnon eli konversion. Ongelmaksi kuitenkin muodostuu näiden tekijöiden määrällinen arviointi, joka on usein varsin subjektiivista eikä mittaamisen tuoksi ole olemassa selkeitä numeraalisia menetelmiä kuten tarkkuuden mittaamiselle on.[6]

Käytännön arviointitekniikat voidaan esimerkiksi jakaa karkeasti kolmeen ryhmään: offline-mittareihin, A/B-testaukseen ja online-mittareihin.[25] Online-mittareiden toiminnalle on elintärkeää käyttäjä ja tältä saatavat tiedot, kun taas offline-mittarit eivät tarvitse aktiivista käyttäjää [6]. Online-menetelmiä ovat muun muassa CRT (Click-Through rate), konversioaste ja keskimääräinen verkkosivulla vietetty aika. Niiden pohjalta arvioidaan järjestelmän toimivuutta reaaliaikaisien käyttäjävuorovaikutusten avulla. Offline-mittarit puolestaan perustuvat historialliseen dataan, ja niihin kuuluvat muun muassa tarkkuutta mittaavat ominaisuudet, kuten täsmällisyys, muistettavuus ja F-mitta. Lisäksi monimuotoisuus- ja uutuusviihätys-mittarit arvioivat suositusten omaperäisyyttä. A/B-testauksella tar-

koitetaan kokeita, joilla mittareiden toimivuutta tarkastellaan todellisessa käyttöympäristössä.[25]

Lisäksi suositusjärjestelmiä voidaan arvioida käyttäjä tutkimusten avulla, joissa testikäyttäjät ovat vuorovaikutuksessa järjestelmän kanssa ja antavat suoraa palautetta sen toiminnasta. Käyttäjätestauksen etuna on, että siitä saatava data on usein tarkempaa ja selkeämmin tulkittavaa kuin passiivisesti kerätty vuorovaikutusdata. Menetelmässä on kuitenkin heikkouksia: testikäyttäjien tietoisuus osallistumisestaan voi vaikuttaa heidän valintoihinsa, jolloin tulokset eivät täysin vastaa todellista käyttötilannetta. Lisäksi rekrytointi on haastavaa, eikä otos yleensä edusta koko käyttäjäkantaa, sillä rekrytointiprosessi itsessään toimii vinoutuneena suodattimena, jota ei voida täysin kontrolloida [6].

Yhteenvetona suositusjärjestelmien arviointi perustuu sekä matemaattisesti määriteltyihin mittareihin että käytännön menetelmiin, jotka kuvaavat järjestelmän suorituskykyä todellisessa käyttöympäristössä. Teknisiä mittareita, kuten tarkkuutta, täsmällisyyttä, muistettavuutta ja regressiopohjaisia virhemittareita, hyödynnetään arvioitaessa mallin ennustekykyä. Näitä täydentävät käyttäjakeskeiset ja reaaliaikaiset menetelmät, kuten online-mittarit, A/B-testit ja käyttäjätestaukset, jotka tuottavat tietoa siitä, miten suositukset vaikuttavat käyttäjien toimintaan ja kokemukseen. Eri arviointitapojen yhdistäminen on välttämätöntä, sillä mikään yksittäinen mittari ei riitä kattavasti kuvaamaan suositusjärjestelmän tehokkuutta monimuuttavassa ja muuttuvassa käyttöympäristössä.

6 Pohdinta

Suositusjärjestelmiä käsittelevissä artikkeleissa korostuu toistuvasti kaksi käsitettä: kontekstuaalisuus ja monikäsitteisyys. Näiden taustalla on havainto, että suositusjärjestelmiä ei voida tarkastella yhden universaalien määritelmän tai yhden parhaan ratkaisun kautta. Järjestelmien tavoitteet, datan saatavuus, verkkokaupan tuotekategoriat sekä organisaation tekniset ja taloudelliset resurssit ovat esimerkkejä tekijöistä, jotka muodostavat yhdessä sen kontekstin, jossa suositusjärjestelmää arvioidaan ja kehitetään. Näin ollen ratkaisu, joka on ideaalinen yhdelle alustalle, voi olla täysin hyödytön toisessa.

Tämä ilmenee selvästi myös tässä tutkielmassa käsitellyissä teknisissä haasteissa. Esimerkiksi kylmäkäynnistys, datan harvuus ja skaalautuvuus eivät ole yleispäteviä ongelmia vaan sidoksissa siihen, millaista vuorovaikutusdataa käyttäjistä ja kohteista voidaan kerätä, kuinka laaja tuotevalikoima on ja miten dynaamista käyttäjien toiminta on. Erityisesti verkkokaupoissa ostokäyttäytymisen ennustaminen on haastavaa, koska yksittäisen käyttäjän mielenkiinnon kohteet voivat muuttua hyvinkin nopeasti: tänään käyttäjä etsii kameran muistikorttia, ja huomenna hänen huomionsa voi kohdistua täysin eri tuotekategoriaan. Suositusjärjestelmän näkökulmasta tämä tarkoittaa, että historian painottaminen ei aina tuota luotettavinta tulosta, ja järjestelmän on pystyttävä tasapainottamaan lyhyen ja pitkän aikavälin käyttäytymisen signaaleja.

Tulosten pohjalta voidaan hahmotella ajatus ”konseptuaalisesta ideaalimallista”, vaikka varsinaista universaalia ideaalijärjestelmää ei voida määritellä. Tällainen järjestelmä yhdistäisi useita tavoitteita, jotka todellisissa ympäristöissä ovat usein keskenään ristiriidassa: korkean tarkkuuden, riittävän monimuotoisuuden, käyttäjäkokemuksen optimoinnin ja liiketoiminnallisten tulosten tukemisen. Lisäksi ideaalimalli olisi kykenevä oppimaan jatkuvasti muuttuvista signaaleista ja reagoimaan niihin reaaliajassa ilman, että suositukset lukkiutuvat menneisyyden toimintamalleihin.

Viime vuosien kehitys erityisesti koneoppimisen ja generatiivisen tekoälyn alueella on tuonut suositusjärjestelmille uusia mahdollisuuksia. Syväoppimiseen perustuvat mallit, kontekstuaaliset suositusalgoritmit ja laajemmat käyttäjäprofiilit voivat lieventää perinteisiä ongelmia kuten datan harvuutta ja monimuotoisuus–tarkkuusristiriitaa. Samalla ne kuitenkin tuovat mukanaan uusia haasteita, kuten läpinäkyvyyden, reiluuden ja mallien tulkittavuuden, joista suositusjärjestelmäkirjallisuudessa keskustellaan yhä aktiivisemmin. Tekoälypohjainen suosittelu voi siis ratkaista olemassa olevia teknisiä rajoitteita, mutta se ei poista tarvetta ymmärtää järjestelmän toimintaa laajemmassa kontekstissa.[28]

Kokonaisuutena voidaan todeta, että suositusjärjestelmät ovat ennen kaikkea adaptiivisia ja kontekstuaalisia teknologioita, joiden kehittäminen ei ole kertaluontoinen suunnitteluprojekti vaan jatkuva optimoinnin ja arvioinnin sykli. Verkkokauppojen kaltaisessa ympäristössä, jossa käyttäjien mieltymykset, tuotteet ja vuorovaikutukset muuttuvat jatkuvasti, suositusjärjestelmän on oltava joustava, oppiva ja monitasoinen. Tulevaisuuden järjestelmät eivät todennäköisesti pyri täydellisyyteen vaan dynaamiseen tasapainoon: niiden on kyettävä mukautumaan tilanteisiin, joissa täydellinen informaatio puuttuu, ja tuotettava silti käyttäjälle tarkoituksenmukaisia, perusteltuja ja hyödyllisiä suosituksia.

7 Yhteenveto

Tässä tutkielmassa tarkasteltiin kirjallisuuskatsauksen avulla verkkokauppojen suositusjärjestelmien teknisiä haasteita, niihin ehdotettuja ratkaisumenetelmiä sekä järjestelmien arvioinnissa käytettäviä mittareita. Suositusjärjestelmät ovat keskeinen osa nykyaikaista sähköistä kaupankäyntiä, sillä ne henkilökohtaistavat käyttökoke-
musta, tehostavat tuotteiden löydettävyyttä ja pyrkivät tukemaan verkkokaupan liiketoiminnallisia tavoitteita. Järjestelmien toimivuus ei kuitenkaan ole itsestäänselvyys, vaan ne kohtaavat teknisiä rajoitteita, jotka vaikuttavat suositusten tarkkuuteen ja luotettavuuteen. Aineiston avulla vastattiin tutkimuksen alussa esiteltyihin kysymyksiin seuraavasti:

TK1 Keskeisiksi suositusjärjestelmien teknisiksi haasteiksi tunnistettiin kylmäkäynnistys, datan harvuus, skaalautuvuus sekä monimuotoisuuteen ja tarkkuuteen liittyvät ristiriidat. Kylmäkäynnistys syntyy, kun järjestelmällä ei ole riittävästi tietoa uusista käyttäjistä tai kohteista. Datan harvuus puolestaan johtuu siitä, että käyttäjät arvioivat vain murto-osan tarjolla olevista tuotteista, mikä vaikeuttaa luotettavien samankaltaisuuksien laskemista. Suurissa verkkokaupoissa myös skaalautuvuus nousee merkittäväksi haasteeksi, sillä järjestelmien on kyettävä käsittelemään kasvavia käyttäjä- ja kohdejoukkoja ilman suorituskyvyn heikkenemistä. Lisäksi tarkkuuden optimointi voi kaventaa suositusten monimuotoisuutta ja johtaa liialliseen erikoistumiseen.

TK2 Kylmäkäynnistystä voidaan lieventää esimerkiksi demografisilla tiedoilla, hybridimalleilla tai koneoppimistekniikoilla, kuten Naive Bayes -luokittelijalla ja WALS-optimoinnilla. Datan harvuutta pyritään vähentämään hyödyntämällä sosiaaliseen luottamukseen perustuvia menetelmiä tai yhdistämällä eri suositelutekniikoita hybridijärjestelmiksi. Skaalautuvuusongelmissa tehokkaita ratkaisuja tuovat klusterointimenetelmät, jotka ryhmittelevät käyttäjiä ja kohteita pienempiin osajoukkoihin. Monimuotoisuus-tarkkuus -ristiriitaa voidaan tasapainottaa esimerkiksi painotettujen naapuruusmenetelmien avulla tai lisäämällä suositusten satunnaisuutta ja uutuusarvoa.

TK3 Arviointimenetelmiä voidaan jakaa teknisiin ja käytännön mittareihin. Teknisiä mittareita ovat muun muassa tarkkuus, täsmällisyys, muistettavuus sekä virhepohjaiset regressiomittarit, kuten MAE ja RMSE. Käytännön arviointia puolestaan toteutetaan online-mittareilla, kuten klikkausprosentteilla ja konversioasteella, sekä A/B-testauksella ja käyttäjätutkimuksilla. Koska yksittäinen mittari ei kuvaa järjestelmän suorituskykyä kokonaisvaltaisesti, arviointi edellyttää useiden menetelmien yhdistämistä.

Tutkielman perusteella voidaan verkkokauppojen suositusjärjestelmien todeta olevan monimutkaisia ja kontekstisidonnaisia kokonaisuuksia, joiden kehittäminen edellyttää tasapainottelua teknisten rajoitteiden, käyttäjäkokemuksen ja liiketoiminnallisten tavoitteiden välillä. Vaikka yksittäisiä ongelmia voidaan lieventää eri menetelmillä, ei universaalista ratkaisua ole olemassa. Parhaat tulokset saavutetaan yhdistämällä useita lähestymistapoja ja arvioimalla järjestelmän toimintaa jatkuvasti muuttuvassa käyttöympäristössä. Näin ollen ideaalia, kaikkialle soveltuvaa mallia ei ole, vaan paras suositusjärjestelmä on dynaaminen, adaptiivinen ja kontekstinsa mukaan kehittyvä kokonaisuus.

Lähdeluettelo

- [1] K. Shah, A. Salunke, S. Dongare ja K. Antala, ”Recommender systems: An overview of different approaches to recommendations”, teoksessa *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, s. 1–4. DOI: 10.1109/ICIIECS.2017.8276172.
- [2] M. Marcuzzo, A. Zangari, A. Albarelli ja A. Gasparetto, ”Recommendation Systems: An Insight Into Current Development and Future Research Challenges”, *IEEE Access*, vol. 10, s. 86 578–86 623, 2022. DOI: 10.1109/ACCESS.2022.3194536.
- [3] A. Saxena, ”A Systematic Survey of Recommendation Systems”, *2023 5th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, 2023. DOI: 10.1109/ICAC3N60023.2023.10541702.
- [4] M. Chen, ”Research on recommender technology in E-commerce recommendation system”, teoksessa *2010 2nd International Conference on Education Technology and Computer*, vol. 4, 2010, s. V4-409-V4-412. DOI: 10.1109/ICETC.2010.5529654.
- [5] F. Ricci, L. Rokach ja B. Shapira, *Recommender Systems Handbook*. Springer Nature, 2010. DOI: 10.1007/978-0-387-85820-3_1.
- [6] C. C. Aggarwal, *Recommender Systems: The Textbook*, 1. painos. Springer Nature, 2016, s. 41–46. DOI: 10.1007/978-3-319-29659-3.

-
- [7] M. V. Kumar ja P. N. V. S. P. Kumar, "A Study on Different Phases and Various Recommendation System Techniques", *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, 5C 2019.
- [8] S. Malik, A. Rana ja M. Bansal, "A Survey of Recommendation Systems", *Information Resources Management Journal*, vol. 33, 4 2020. DOI: 10.4018/IRMJ.2020100104.
- [9] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim ja R. Kashef, "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities", *Advances in Artificial Intelligence: Machine Learning, Data Mining and Data Sciences*, 2020. DOI: 10.3390/app10217748.
- [10] M. Rishabh Sisodiya ja M. Kanika, "E-Commerce Recommendation System", *International Journal of Scientific Research in Engineering and Management (IJSREM)*, vol. 8, 4 2024. DOI: 10.55041/IJSREM30996.
- [11] D. Goldberg, D. Nichols, M. Oki Brian ja D. Terry, "Using collaborative filtering to weave an information tapestry", *Communications of the ACM*, vol. 35, s. 61–70, 12 1992. DOI: 10.1145/138859.138867.
- [12] K. Raghuwanshi Sandeep ja K. Pateriya R, "Recommendation Systems: Techniques, Challenges, Application, and Evaluation", teoksessa *Soft Computing for Problem Solving*, vol. 2, 2018, s. 151–164. DOI: 10.1007/978-981-13-1595-4_12.
- [13] J. Pazzani Michael ja D. Billsus, "Content-based recommendation systems", teoksessa Springer, 2007, luku 10, s. 325–341. DOI: 10.1007/978-3-540-72079-9_10.
- [14] S. Jain, A. Grover, P. Singh Thakur ja S. Kumar Choudhary, "Trends, problems and solutions of recommender system", teoksessa *International Con-*

- ference on Computing, Communication and Automation*, 2015. DOI: 10.1109/CCAA.2015.7148534.
- [15] R. Burke, "Hybrid Recommender Systems: Survey and Experiments", teoksessa *User Modeling and User-Adapted Interaction*. Springer Nature, 2002, luku 12, s. 331–370. DOI: 10.1023/A:1021240730564.
- [16] S. Khusro, Z. Ali ja I. Ullah, "Recommender Systems: Issues, Challenges, and Research Opportunities", teoksessa *Information Science and Applications (ICISA)*, 2016, s. 1179–1189. DOI: 10.1007/978-981-10-0557-2_112.
- [17] A. I. Schein, A. Popescul, L. H. Ungar ja D. M. Pennock, "Methods and metrics for cold-start recommendations", teoksessa *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2002. DOI: 10.1145/564376.564421.
- [18] F. Al-Bakri Nadia ja S. Hassan Hashim, "Reducing Data Sparsity in Recommender Systems", *Journal of Al-Nahrain University*, vol. 21, s. 138–147, 2018. DOI: 10.22401/JNUS.21.2.20.
- [19] G. Guo, "Integrating trust and similarity to ameliorate the data sparsity and cold start for recommender systems", teoksessa *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, s. 451–454. DOI: 10.1145/2507157.2508071.
- [20] G. Guo, Z. jie ja D. Thalmann, "Merging trust in collaborative filtering to alleviate data sparsity and cold start", *Knowledge-Based Systems*, vol. 57, s. 57–68, 2014. DOI: 10.1016/j.knosys.2013.12.007.
- [21] M. Sarwar Badrul, G. Karypis, J. Konstan ja J. Riedl, "Recommender Systems for Large-scale E-Commerce: Scalable Neighborhood Formation Using Clustering", teoksessa *In Proceedings of the Fifth International Conference on Computer and Information Technology*, vol. 5, 2002, s. 291–324.

- [22] P. Adamopoulos ja A. Tuzhilin, ”On over-specialization and concentration bias of recommendations: probabilistic neighborhood selection in collaborative filtering systems”, *Proceedings of the 8th ACM Conference on Recommender Systems*, s. 153–160, 2014. DOI: 10.1145/2645710.2645752.
- [23] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. Rushton Wakeling ja Y.-C. Zhang, ”Solving the apparent diversity-accuracy dilemma of recommender systems”, *Proc. Natl. Acad. Sci. USA*, 2010. DOI: 10.1073/pnas.1000488107.
- [24] K. S. Elattar ja K. M. Fouad, ”A Survey on Recommender Systems Challenges and Solutions”, teoksessa *2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2022, s. 296–301. DOI: 10.1109/MIUCC55081.2022.9781739.
- [25] L. El Harrouchi, H. Moussaoui, M. Karmoudi ja N. El Akkad, ”A review of recommendation systems”, teoksessa *2025 5th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, 2025, s. 1–9. DOI: 10.1109/IRASET64571.2025.11008191.
- [26] S. Foulle, ”Mathematical Characterization of Better-than-Random Multiclass Models”, *Transactions on Machine Learning Research*, 2025.
- [27] F. O. Isinkaye, Y. O. Folajimi ja B. A. Ojokoh, ”Recommendation systems: Principles, methods and evaluation”, *Egyptian Informatics Journal*, vol. 16, nro 3, s. 261–273, 2015. DOI: 10.1016/j.eij.2015.06.005.
- [28] L. V. Nguyen, ”Next-Generation Recommendation Systems in E-Commerce”, teoksessa *2025 International Conference on Computing, Intelligence, and Application (CIACON)*, 2025, s. 1–6. DOI: 10.1109/CIACON65473.2025.11189672.