

Aaro Melchy

SUOSITTELUJÄRJESTELMÄT JA VÄÄRISTYMÄT

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Toukokuu 2022

TIIVISTELMÄ

Aaro Melchy: Opinnäytetyön otsikko
Kandidaattitutkielma
Tampereen yliopisto
Tietojenkäsittelytieteiden tutkinto-ohjelma
Elokuu 2022

Suosittelujärjestelmiä käytetään nykyään useissa erilaisissa ympäristöissä käyttäjien tukena. Niiden avulla suositellaan sisältöä käyttäjälle, minkä seurauksena käyttäjät voivat nopeasti ja mutkattomasti löytää itselleen mieleisiä kohteita monien erilaisten kohteiden seasta. Suosittelujärjestelmien kohtaamat vääristymät voivat tehdä suositteluprosessista puolueellista. Tässä työssä tutkitaan suosittelujärjestelmien toimintaa ja siihen vaikuttavia vääristymiä kartoittaen ongelmaa. Työn tavoite on selvittää, minkälaisia vääristymiä suosittelujärjestelmät kohtaavat, ja minkälaisin keinoin näitä vääristymiä voidaan ehkäistä ja lieventää.

Työn ensimmäisessä osassa esitellään suosittelujärjestelmien keskeisimmät suodatusmenetelmät: Yhteistoiminnallinen suodatus, sisältöpohjainen suodatus sekä hybridimenetelmät. Lisäksi ensimmäisessä osassa esitellään suosittelujärjestelmien käyttöympäristöjä. Toisessa osassa käsitellään vääristymiä avaten ongelman taustaa ja eritellen vääristymien eri tyyppisiä. Vääristymien eri tyypeille esitetään erilaisia menetelmiä, joilla niitä voidaan ehkäistä ja lieventää.

Työn tutkimuksen tuloksena vääristymien eri ilmentymien tyyppisiä saatiin eriteltyä. Vääristymät pohjautuvat monenlaisiin eri syihin, kuten kohteiden suosioon, näkyvyyteen ja käyttäjien toimintaan. Vääristymien eri ilmentymiä yhdistää se, että niitä ilmenee järjestelmässä usein luonnollisesti. Näitä on myös usein hankalaa havaita ilman tarkempaa selvitystä. Tämän vuoksi vääristymien ennaltaehkäisy tulisi ottaa huomioon jo järjestelmän suunnitteluvaiheessa. Tutkimuksen myötä ilmeni myös, että vääristymiä voidaan lieventää erilaisilla menetelmillä. Jokaista vääristymätyyppiä saadaan lievennettyä sille soveltuvilla menetelmillä. Menetelmät painottuvat muun muassa datan käsittelyyn, erilaisiin koneoppimismenetelmiin ja mallintamiseen. Menetelmiä tulisi ottaa järjestelmään käyttöön jo järjestelmän käyttöönoton aikana, jotta vääristymiä voitaisiin ehkäistä mahdollisimman paljon turvaten järjestelmän tarkoituksenmukainen toiminta.

Avainsanat: Suosittelujärjestelmät, vääristymät, puolueellisuus, palautesilmukka

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

Sisällysluettelo

1	Johdanto	1
2	Suosittelujärjestelmät.....	1
2.1	Suosittelujärjestelmien menetelmät	2
2.1.1	Yhteistoiminnallinen suodatus	2
2.1.2	Sisältöpohjainen suodatus	2
2.1.3	Hybridimenetelmä	2
2.2	Suosittelujärjestelmien käyttöympäristöt	3
2.2.1	Verkkokaupat	3
2.2.2	Suoratoistopalvelut	3
2.2.3	Sosiaalinen media	4
2.2.4	Uutispalvelut	4
3	Vääristymät.....	4
3.1	Palautesilmukka	5
3.2	Vääristymät datassa	6
3.2.1	Valintavääristymät	6
3.2.2	Altistusvääristymät	6
3.2.3	Konformisuusvääristymät	7
3.2.4	Positiivivääristymät	7
3.3	Vääristymät ja epärealisuus tuloksissa	8
3.3.1	Suosiovääristymät	8
3.3.2	Epärealisuus	9
3.4	Vääristymät mallissa	10
3.5	Vääristymien vahvistuminen palautesilmukassa	10
4	Keskustelu	12
5	Yhteenveto.....	13
	Lähdeluettelo.....	13

1 Johdanto

Suosittelujärjestelmät (engl. *recommender systems*) ovat nykyisin käytössä laajasti erilaisissa palveluissa, esimerkiksi verkkokaupoissa, sosiaalisessa mediassa ja uutissivustoilla. Niiden avulla palveluista saadaan muovattua jokaiselle käyttäjälle henkilökohtaisempi kokonaisuus suosittelemalla käyttäjille kohteita pohjautuen heidän omiin mieltymyksiinsä. Täten niiden toiminta vaikuttaa palveluiden käyttäjiin vahvasti, joten järjestelmien tehokas ja tarkoituksenmukainen toiminta on tärkeää.

Tästä huolimatta tarkoituksenmukainen toiminta voi olla paikoin haastavaa saavuttaa suosittelujärjestelmien kohtaamien ongelmien ja haasteiden takia. Vääristymät (engl. *bias*) ovat yksi ongelmien ilmentymä ja niiden seuraukset voivat olla arvaamattomia ja laajoja. Vääristymät ovat oleellinen ongelma suosittelujärjestelmiin liittyen, minkä takia niiden ehkäiseminen on tärkeää. Ongelmana se on kuitenkin moniulotteinen, minkä vuoksi ongelman kartoitus ja sopivien ratkaisujen löytäminen on oleellista.

Tutkielma on toteutettu kirjallisuuskatsauksena ja siinä perehdytään suosittelujärjestelmiin ja tarkemmin niiden kohtaamiin vääristymiin. Tarkoituksena on tutkia, millaisia vääristymiä esiintyy ja millaisia keinoja voidaan käyttää niiden ehkäisemistä ja lieventämistä varten. Täten tutkimuskysymyksenä on: ”Millaisia vääristymiä suosittelujärjestelmät kohtaavat ja miten näitä voidaan lieventää?”. Ilmiön laajuudesta ja monimuotoisuudesta johtuen ehkäisymenetelmiä ei kuvata hyvin yksityiskohtaisesti, vaan tarkoituksena on pohjustaa näitä ja antaa yleiskuvaa ratkaisuista.

Tutkielman toisessa luvussa käydään läpi suosittelujärjestelmiä yleisellä tasolla, avataan niiden tyypillisiä menetelmiä ja erilaisia käyttöympäristöjä. Tämän jälkeen kolmannessa luvussa tarkastellaan vääristymien eri tyyppisiä ja näiden taustoja. Tämän lisäksi kyseisessä kappaleessa pohjustetaan keinoja lieventää näitä vääristymiä. Neljännessä luvussa tuodaan omaa pohdintaa esille enemmän ja tutkielman viimeisessä luvussa käydään vielä oleellisimmat asiat läpi.

2 Suosittelujärjestelmät

Suosittelujärjestelmät toimivat käyttäjän tukena sisällön seulomisessa ja valintojen tekemisessä erilaisissa sähköisissä ympäristöissä. Niiden tärkeys on korostunut viime aikoina, sillä internetin sisältämä data ja informaatio on kasvanut erittäin paljon vuosien varrella. Ilman suosittelujärjestelmiä kaiken tämän informaation ja datan seasta voisi olla hyvin vaivalloista löytää käyttäjälle itselle sopivaa sisältöä. Esimerkiksi verkkokaupoissa käyttäjille usein suositellaan tuotteita, joita he todennäköisesti voisivat päätyä ostamaan perustuen esimerkiksi käyttäjän ostohistoriaan. Nimensä mukaisesti suosittelujärjestelmien toiminta perustuu siihen, että käyttäjälle suositellaan kohteita

järjestelmän seulomien kohteiden seasta ja tätä seulontaa toteutetaan eri menetelmin. (Shah, Salunke, Dongare & Antala, 2017)

2.1 Suositelujärjestelmien menetelmät

Suosittelujärjestelmien toiminta voidaan rakentaa erilaisten menetelmien mukaan. Tyypillisesti menetelmät jaotellaan kolmeen eri luokkaan: yhteistoiminnalliseen suodatukseseen, sisältöpohjaiseen suodatukseseen ja hybridimenetelmiin. Jokainen menetelmä sisältää erilaisia vahvuuksia ja heikkouksia. (Shah et al., 2017) Tämän takia on tärkeää selvittää sopiva menetelmä kehittäessä järjestelmää.

2.1.1 Yhteistoiminnallinen suodatus

Kollaboratiivinen eli yhteistoiminnallinen suodatus pohjautuu suositusten tekemiseen muiden käyttäjien kanssa, jotka ovat alkuperäisen käyttäjän kanssa samankaltaisia aiempien valintojensa perusteella. Pääperiaate on, että käyttäjät, jotka ovat aiemmin päätyneet samanlaisiin arvioihin erilaisten kohteiden suhteen, ovat jatkossakin luultavasti samaa mieltä. Kollaboratiivisen suodatuksen menetelmät voidaan jaotella läheisyyspainotteisiin (engl. *neighborhood based*) ja mallipainotteisiin (engl. *model-based*) menetelmiin. (Shah et al., 2017) Yhteistoiminnallista suodatusta käytetään usein esimerkiksi verkkokaupoissa. Käyttäjälle voidaan suositella tuotteita, joita samantyyppisen ostohistorian omaavat käyttäjät ovat ostaneet.

2.1.2 Sisältöpohjainen suodatus

Sisältöpohjainen suodatus perustuu suositusten tekemiseen perustuen kohteiden sisällön kuvaukseen ja ominaisuuksiin. Käyttäjän mieltymyksiä pohjalta luodaan malli pohjautuen aiempien kohteiden analysointiin. (Shah et al., 2017) Tämä malli on havainnollistus käyttäjän mieltymyksistä ja kiinnostuksista toimien pohjana uusille suosituksille (Lops, Gemmis & Semeraro, 2011).

Lops ja muut (2011) kiteyttävät prosessin siten, että uusia kohteita seulotaan suositeltavaksi vertaamalla niiden kuvausta ja ominaisuuksia käyttäjäprofiiliin, joka pohjautuu edellä kuvattuun malliin. Tämän vertaamisprosessin myötä tulokseksi saadaan arvio kohteen osuvuudesta ja olennaisuudesta käyttäjälle (Lops et al., 2011). Esimerkkinä sisältöpohjaisesta suodatuksesta Spotify-musiikkipalvelun suosittelu: Jos käyttäjä on kuunnellut ja lisännyt suosikkeihin jazz-kappaleita, hänelle voidaan jatkossa suositella jazz-kappaleita.

2.1.3 Hybridimenetelmä

Hybridimenetelmät perustuvat usean eri suodatusmenetelmän hyödyntämiseen. Hybridimenetelmien tarkoituksena on parantaa järjestelmien tehokkuutta (Shah et al. 2017). Tämän myötä on mahdollista karsia eri menetelmien heikkouksia ja myös

hyödyntää niiden vahvuuksia. Hybridimenetelmät tuovat paljon joustavuutta, sillä menetelmät voidaan toteuttaa monin eri tavoin (Shah et al., 2017).

Esimerkkinä tästä on painotettu (engl. *weighted*) toteutus, jonka avulla kohteen painoarvo suosittelun kannalta lasketaan kaikkien järjestelmässä toteutettujen menetelmien pohjalta (Shah et al., 2017). Täten järjestelmään saadaan luotua enemmän moniulotteisuutta.

2.2 Suosittelevjärjestelmien käyttöympäristöt

Suosittelujärjestelmiä voidaan käyttää monenlaisissa eri ympäristöissä. Kehityksen myötä suosittelujärjestelmiä opitaan hyödyntämään yhä enemmän ja niiden käyttö onkin vakiintunut jo tietyissä ympäristöissä kuten verkkokaupoissa. Erilaisten palvelujen toimintaa saadaan suosittelujärjestelmillä tehostettua sekä palveluntarjoajan että käyttäjän näkökulmasta. Suosittelevjärjestelmät tarjoavat monenlaisia mahdollisuuksia palveluntarjoajille, mutta ne voivat tuoda myös erilaisia ongelmia ja haasteita.

2.2.1 Verkkokaupat

Suosittelujärjestelmät ovat erittäin tärkeä osa verkkokaupankäyntiä yritysten sekä käyttäjien kokemuksen kannalta nykypäivänä (Alamdari, Navimipour, Hosseinzadeh, Safaei & Darwesh, 2020). Verkkokaupat sisältävät usein laajasti informaatiota erilaisten vaihtoehtojen muodossa, minkä takia asiakkaalle on tärkeää löytää itselle parhaat vaihtoehdot näiden joukosta. Suosittelevjärjestelmät toimivat asiakkaan tukena antamalla erilaisia suosituksia vaihtoehtojen suhteen, minkä seurauksena asiakkaan ei tarvitse käyttää yhtä paljoa aikaa tuotteiden hakemiseen. Tällä tavalla saadaan tehtyä palvelusta helppokäyttöisempi ja asiakkaan käyttäjäkokemus paranee, mikä kasvattaa tyytyväisyyttä palvelua kohtaan (Alamdari et al., 2020).

Asiakkaalle saadaan suositeltua myös ylimääräisiä tuotteita, joita asiakas ei olisi muuten välttämättä päätenyt ostamaan, mikä lisää myyntiä entisestään ja parantaa asiakasuskollisuutta (Alamdari et al., 2020). Esimerkiksi Amazon ja Ebay hyödyntävät suosittelujärjestelmiä palveluissaan.

2.2.2 Suoratoistopalvelut

Monet suoratoistopalvelut kuten Netflix ja Spotify hyödyntävät suosittelujärjestelmiä. Netflix sisältää valtavan määrän erilaisia elokuvia ja sarjoja, joiden joukosta käyttäjä voi valita itselleen mieleistä katsottavaa palvelun kotisivulta, joka on täysin personalisoitu käyttäjälle (Gomez-Urbe & Hunt, 2015).

Tutkimuksen mukaan tyypillinen Netflixin käyttäjä menettää kiinnostuksensa 60–90 sekunnin välillä tehdessään valintoja katseltavan sisällön suhteen. Joko käyttäjä löytää jotain mielenkiintoista tai muuten riski, että käyttäjä hylkää palvelun, kasvaa huomattavasti. Tämän takia suosittelujärjestelmiä tarvitaan Netflixissä: niiden avulla

käyttäjälle saadaan suositeltua mielenkiintoista sisältöä ja saadaan ymmärrystä sen suhteen, minkälaisen sisällön käyttäjä kokee mielenkiintoiseksi. (Gomez-Uribe et al., 2015)

Voidaan siis olettaa, että suosittelujärjestelmät ovat oleellinen osa suoratoistopalveluja, koska niiden avulla voidaan taata palvelun menestyminen varmemmin käyttäjän kokeman tyytyväisyyden kautta.

2.2.3 Sosiaalinen media

Suosittelujärjestelmistä on tullut vakiintunut osa sosiaalista mediaa viime vuosien aikana. Tyypilliseen tapaan suosittelujärjestelmien tarve sosiaalisessa mediassa on suuri suurien informaatiomäärien takia. Muun muassa Youtube, Facebook ja Twitter hyödyntävät suosittelujärjestelmiä palveluissaan. Sosiaalisessa mediassa käyttäjästä kerätään paljon tietoa ja suosittelujärjestelmät voivat hyödyntää tätä informaatiota analysoiden esimerkiksi käyttäjän kiinnostuksenkohteita ja sosiaalisia verkostoja. Näiden pohjalta saadaan suositeltua käyttäjälle erilaisia asioita, kuten ihmisiä, tuotteita ja paikkoja. (Campana & Delmastro, 2017)

2.2.4 Uutispalvelut

Uutisten käyttö sähköisessä muodossa on kasvanut vuosien varrella. Laajasti käytössä olevat uutispalvelut tarjoavat käyttäjilleen ajankohtaisia ja mielenkiintoisia uutisia käyttäjilleen suosittelujärjestelmien avulla. Uutisten tiiviin julkaisutahdin ja laajan saatavilla olevan informaation takia käyttäjien on yhä vaivalloisempaa löytää mielenkiintoisia ja itselle sopivia uutisia (Karimi, Jannach & Jugovac, 2018). Suosittelujärjestelmien avulla saadaan paikattua tätä ongelmaa uutispalveluissa. Karimin ja muiden (2018) tutkimuksesta selviää, että sisältöpohjainen suodatus on tehokkain menetelmä uutispalveluiden kannalta tutkijoiden näkökulmasta. Tämä ei kuitenkaan tarkoita, että muita menetelmiä ei tulisi hyödyntää.

3 Vääristymät

Yksi oleellinen suosittelujärjestelmiin liittyvä ongelma on vääristymien ilmeneminen. Aiheeseen liittyvä tutkimus on kasvanut viime vuosien aikana huomattavasti (Chen, Dong, Wang, Feng, Wang & He, 2020), mikä osoittaa aiheen tärkeyden ja ajankohtaisuuden. Vääristymät johtavat siihen, että järjestelmän käyttäjälle antamat suositukset eivät ole optimaalisia käyttäjän näkökulmasta ja itse järjestelmän tehokkuus kärsii, mikä huonontaa käyttäjäkokemusta sekä mahdollisesti myös palvelun mainetta (Chen et al., 2020).

Vääristymiä esiintyy monissa eri muodoissa ja niihin vaikuttavat tekijät vaihtelevat paljon, mikä voi tehdä ongelman ratkaisemista monimutkaista ja haastavaa. Vääristymien

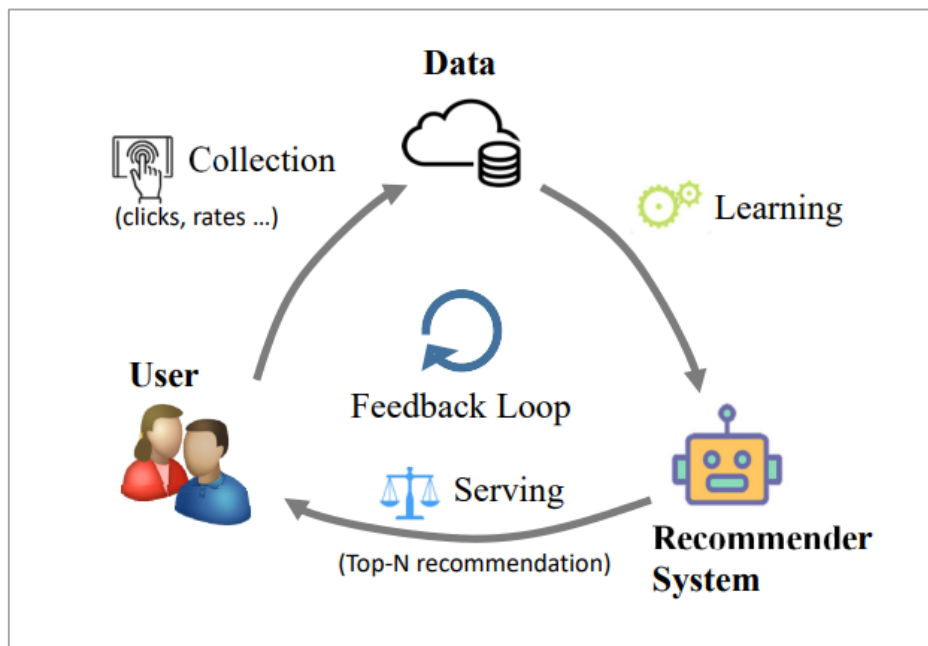
ilmetessä järjestelmän suositteluprosessi muuttuu käytännössä puolueelliseksi. On siis tärkeää kyetä havaitsemaan näitä vääristymiä niiden ilmetessä, jotta niitä voidaan kitkeä järjestelmästä pois ja näin ollen taata optimaalinen ja puolueeton toiminnallisuus. Mahdolliset vääristymät tulisi ottaa jo järjestelmän suunnittelun yhteydessä huomioon, jotta voidaan parhaassa tilanteessa jopa ehkäistä niiden ilmeneminen kokonaan.

Tässä luvussa keskitytään vääristymien tyypeihin, joita on koottu ja avattu Chenin ja muiden (2020) tutkimuksessa. Kyseiseen tutkimukseen on valittu vääristymien tyyppejä, joihin pohjautuen on aiemmin tehty tieteellistä tutkimusta.

3.1 Palautesilmukka

Suosittelujärjestelmän toiminta koostuu kolmesta eri osapuolesta: käyttäjästä, datasta ja itse suosittelujärjestelmästä. Käyttäjän toiminnasta kerätään dataa, jonka avulla koulutetaan suosittelujärjestelmää. Suosittelujärjestelmä antaa suosituksia käyttäjälle dataan pohjautuen ja käyttäjä tekee näiden pohjalta valintoja, jota käytetään taas datana käyttäjän muun toiminnan ohessa. Näin syntyy palautesilmukka (engl. *feedback loop*), johon suosittelujärjestelmien toiminta tyypillisesti perustuu. Tätä kokonaisuutta havainnollistaa kuva 1, josta havaitaan silmukan eri osat ja niiden välillä kulkevat vaiheet.

Vääristymien eri tyypit ilmenevät silmukan eri vaiheissa; esimerkiksi suosiovääristymiä (engl. *popularity bias*) voi ilmetä järjestelmän tarjotessa suosituksia käyttäjälle. Itse silmukka voi vahvistaa vääristymiä entisestään pahentaa ongelmaa, mikä on myös yksi vääristymän muoto. (Chen et al., 2020) Tähän palataan tarkemmin myöhemmin työssä.



Kuva 1. Palautesilmukan havainnollistus (Chen et al., 2020).

3.2 Vääristymät datassa

Käyttäjien toimintaan pohjautuva data on havainnollista kokeellisen sijaan, minkä seurauksena data voi olla vääristynyttä (Chen et al., 2020). Käytännössä tämä johtaa siihen, että mallit, joiden avulla suosittelujärjestelmiä koulutetaan, alkavat ilmentämään näitä vääristymiä ja jopa kasvattamaan niitä. Chen ja muut (2020) ovat nimenneet ilmiön datavääristymiksi (engl. *data bias*) ja he luokittelevat sen neljään eri alaluokkaan: Valintavääristymiin, konformisuusvääristymiin, altistusvääristymiin ja positiovääristymiin.

3.2.1 Valintavääristymät

Valintavääristymät (engl. *selection bias*) pohjautuvat käyttäjien vapautteen arvostella kohteita. Tutkimuksien mukaan käyttäjillä on tapana valita arvosteltavaksi kohteita, joista he pitävät. Tämän lisäksi käyttäjät todennäköisimmin arvostelevat tuotteita, joista he joko vahvasti pitävät tai eivät pidä. Wangin ja muiden tutkimuksen (2021) mukaan suurin osa saaduista arvioista on nimenomaan positiivisia arvioita ja negatiiviset arviot ovat saatujen arvioiden joukossa aliedustettuina. Ongelma kiteytyy siihen, että osa arvioista jää puuttumaan kokonaan käyttäjän jättäessä tiettyjä kohteita arvostelematta, minkä seurauksena data vääristyy, eikä ilmennä käyttäjien mielipiteitä todenmukaisesti.

Chenin ja muiden (2020) tutkimuksessa on esitetty lukuisia eri menetelmiä valintavääristymien vähentämiseen. Yksi keinoista on liittää tehtyihin havaittuihin arvosteluihin painotettu arvo, joka pohjautuu marginaaliseen todennäköisyyteen havaita arvostelu tietyllä käyttäjän ja kohteen muodostamalle parille. Tämän lisäksi voidaan hyödyntää ATOP-mittaria [Steck, 2010]. Näiden menetelmien heikkoudet piilevät siinä, että tarkkoja ja totuudenmukaisia vaadittuja määreitä ei aina voida määrittellä täysin oikein. Muut valintavääristymiä vähentävät menetelmät keskittyvät suosittelujärjestelmän kouluttamiseen mallin kautta. (Chen et al., 2020)

3.2.2 Altistusvääristymät

Epäsuoraa palautetta (engl. *implicit feedback*), joka koostuu muun muassa käyttäjän klikkauksista ja aiemmista ostoista, hyödynnetään laajasti suosittelujärjestelmien datankeruussa. Tämän tyyppinen palaute ei luonnollisesti tarjoa kovin monipuolisia keinoja muodostaa dataa. Epäsuoran palautteen kautta voidaan saada ainoastaan osittaisia merkkejä siitä, mistä käyttäjä pitää. Sen sijaan merkkejä siitä, mistä käyttäjä ei pidä ei voida muodostaa tämän tyyppisen palautteen kautta suoraan. Tämän myötä nämä merkit pohjataan havaitsematta jääneeseen vuorovaikutukseen. (Chen et al., 2019) Käytännössä siis saatetaan päätyä vääränlaisiin oletuksiin ja päätelmiin havaitsematta jääneen vuorovaikutuksen kannalta.

Tähän ilmiöön pohjautuvat altistusvääristymät (engl. *exposure bias*). Chenin ja muiden (2020) määritelmän mukaan altistusvääristymät johtuvat siitä, että käyttäjät ovat

altistettuina vain osalle tietyistä kohteista, minkä seurauksena huomioimatta jäänyt vuorovaikutus ei aina ole osoitus käyttäjän negatiivisista mielipiteistä. Käytännössä osa kohteista ei välttämättä näy käyttäjälle, jolloin on väärin olettaa, ettei käyttäjä olisi vain ollut kiinnostunut tietystä kohteesta ja sen vuoksi jättänyt sen huomioimatta. Suosittelujärjestelmät, jotka eivät kykene tunnistamaan huomioimattomia vuorovaikutuksien todellisia syitä, kohtaavat tämän myötä laajavaltaisia vääristymiä (Chen et al., 2020).

Altistusvääristymiä voidaan vähentää samalla tavalla kuin valintavääristymiä liittämällä jokaiseen havaintoon painotettu arvo. Pääperiaatteena on antaa usein havaituille vuorovaikutuksille matalampi painoarvo ja harvinaisemmille taas korkeampi. Toinen vaihtoehto on keskittyä altistusvääristymien vähentämiseen suosittelujärjestelmän koulutuksen yhteydessä mallin kautta, mikä sisältää monia eri menetelmiä, etenkin painotukseen perustuvia. (Chen et al., 2020)

3.2.3 Konformisuusvääristymät

Konformisuusvääristymät (*engl. conformity bias*) perustuvat käyttäjien taipumukseen toimia samalla tavalla muiden käyttäjien kanssa ryhmän sisällä, silloinkin kun tämä toiminta olisi käyttäjän omasta toiminnasta poikkeavaa, minkä myötä käyttäjältä saatu palaute ei aina vastaa täysin omaa todellista mielipidettä (Chen et al., 2019).

Käytännössä käyttäjät saattavat mukauttaa omat mielipiteensä muiden mielipiteiden pohjalta, mikä ilmenee vääristyneenä datana suosittelujärjestelmien kannalta, koska se rakennetaan käyttäjien palautteeseen pohjautuen. Krishnanin ja muiden (2014) tutkimuksesta selviää, että sosiaalisella vaikutuksella on vaikutusta käyttäjien antamiin arvioihin kohteisiin liittyen käyttäjien altistuessa sekä etu- että jälkikäteen julkisille mielipiteille. Kyseisessä tutkimuksessa käsitellään ilmiötä termillä ”sosiaalisen vaikutuksen vääristymä” (*engl. social influence bias*). Kyseinen ilmiö on läsnä melkein kaikissa suosittelujärjestelmissä, sillä tietyn kohteen tilastot pohjautuen arvioiden keskiarvoihin ovat näkyvissä usein julkisesti käyttäjille (Krishnan et al., 2014).

Konformisuusvääristymien vähentämiseksi on esitetty kaksi eri keinoa. Ensimmäinen perustuu suosion vaikutuksen mallintamiseen, jotta konformisuuden vaikutukset saadaan irrotettua käyttäjien todellisista mieltymyksistä. Toinen taas perustuu sosiaalisen vaikutuksen mallintamiseen, jolloin se saadaan irrotettua käyttäjien todellisista mieltymyksistä. Heikkoutena tämän tyypisessä menettelyssä on se, että datan muodostaminen vaatii olettamuksia, mikä voi johtaa epätarkkuuteen. (Chen et al., 2019)

3.2.4 Positiivävääristymät

Chenin ja muiden (2020) mukaan positiivävääristymät (*engl. position bias*) ovat vääristymiä, joiden syynä on käyttäjien taipumus valita kohteita, jotka ovat suosituslistan kärjessä, huolimatta kohteiden todellisesta relevanttiudesta, minkä seurauksena valitut

kohteet eivät välttämättä ole kovinkaan relevantteja. Heidän mukaansa tämän tyyppiset vääristymät ovat erittäin yleisiä suosittelujärjestelmien kohdalla, etenkin mainosjärjestelmissä sekä hakukoneissa.

Käyttäjät saattavat siis jättää kohteet, jotka eivät ole suosituslistan kärjessä, kokonaan huomioimatta olettaen, että kärjessä olevat kohteet ovat eniten relevantteja. Collinsin ja muiden (2018) mukaan tämän tyyppisiä vääristymiä ilmenee, koska useissa tapauksissa kohteiden relevanssin määrittelee käyttäjien klikkauksiin pohjautuva data, mutta tämän datan suhteen ei välttämättä huomioida positiovääristymiä. Tämän seurauksena saadut arviot voivat olla virheellisiä. Heidän tutkimuksestaan selviää, että vaikka merkittävä osa käyttäjistä selaakin suosituslistan läpi antamatta liikaa painoarvoa kohteiden positiolle, ovat positiovääristymät silti vahvasti läsnä.

Positiovääristymiä ja niiden ehkäisymenetelmiä on viime vuosien aikana tutkittu lukuisissa tutkimuksissa. Erinäiset mallit sopivat positiovääristymien ehkäisemiseen. Esimerkiksi klikkausmallien kautta saadaan hypoteeseja käyttäjien selauskäyttäytymisestä ja voidaan arvioida kohteiden todellinen relevanssi optimoimalla havaittujen klikkausten todennäköisyydet. Näin ollen saadaan mallinnettua klikkausten prosessia. Toinen vaihtoehto on Agarwalin ja muiden [2019] esittämä malli, joka tasapainottaa kohteiden position vaikutuksia käyttäjien päätöksiensä osalta. Näiden mallien lisäksi voidaan hyödyntää positioita huomioivia painoarvoja datan käsittelyssä. (Chen et al., 2020)

3.3 Vääristymät ja epärealisuus tuloksissa

3.3.1 Suosiovääristymät

Suosiovääristymiä (engl. *popularity bias*) esiintyy suosittelujärjestelmän antamissa tuloksissa (Chen et al., 2020). Ne ovat yleinen ja tunnettu ilmiö suosittelujärjestelmien kohdalla, pohjautuen pitkä häntä -ilmiöön (engl. *long-tail effect*) (Abdollahpouri & Mansoury, 2020). Abdollahpourin ja Mansouryn (2020) tutkimuksessa on kuvan avulla havainnollistettu pitkä häntä -ilmiötä: suositut kohteet vastaavat vain pientä osaa kohteista, mutta ne saavat kuitenkin eniten arvosteluja. Vähemmän suositut kohteet taas vastaavat suurta osaa kohteista, mutta eivät saa läheskään yhtä paljon arvosteluja. Suosiovääristymät voimistavat pitkä häntä -ilmiötä entisestään ja vaikuttavat kohteiden näkyvyyteen, mikä tarkoittaa, että suosiovääristymät johtavat myös altistusvääristymiin (Abdollahpouri & Mansoury, 2020).

Käytännössä siis suosittuja kohteita arvostellaan enemmän kuin vähemmän suosittuja, mikä johtaa siihen, että suosittuja kohteita suositellaan käyttäjille myös entistä enemmän (Abdollahpouri & Mansoury., 2020). Suosittujen kohteiden suosittelu ei välttämättä ole optimaalista, koska kohteet ovat jo ennestään hyvin tunnettuja, minkä seurauksena vähemmän tunnettujen kohteiden löytämisen todennäköisyys heikkenee

(Abdollahpouri, Burke & Mobasher. 2019). Chenin ja muiden (2020) mukaan suosiovääristymät johtavat kohdetarjonnan monipuolisuuden heikkenemiseen, mikä taas johtaa käyttäjäkokemuksen heikkenemiseen johtuen käyttäjien mieltymysten laajasta vaihtelusta.

Suosiovääristymiä voidaan vähentää esimerkiksi regularisoinnilla (engl. *regularization*), jonka avulla voidaan parantaa mallia antamaan tasapainoisempia suosituslistoja. Tutkijat ovat esittäneet monia erilaisia menetelmiä regularisoinnin suhteen. Tämän lisäksi voidaan hyödyntää adversiaalista oppimista (engl. *adversarial learning*), jonka pääideana on parantaa vähemmän suosittujen kohteiden suosittelumahdollisuuksia. Käytännössä kuuluu suosittujen ja vähemmän suosittujen kohteiden välillä pienenee ja pitkän häntä -ilmiön negatiivinen vaikutus heikkenee. Kausaalisten graafien (engl. *causal graphs*) avulla voidaan tulkita ja tehdä suosiovääristymät näkyviksi, jolloin näitä voidaan lievittää. (Chen et al., 2020)

3.3.2 Epäreiluus

Toinen olennainen tuloksiin liittyvä olennainen vääristymä on epäreiluus (engl. *unfairness*), joka on ilmiönä saanut yhä enemmän huomiota viime vuosien aikana. Friedmanin ja Nissenbaumin [1996] määritelmän mukaan tietoteknisissä ympäristöissä käsite epäreiluus tarkoittaa järjestelmän systemaattista ja epäreilua diskriminointia tiettyjä yksilöitä tai ryhmiä kohtaan muiden eduksi. (Chen et al., 2020)

Suosittelujärjestelmien kohdalla tasavertaisuuden säilyttäminen voi siis olla haastavaa toimintamallinsa vuoksi. Chenin ja muiden (2020) mukaan epäreilouden ilmiö on ollut esteenä suosittelujärjestelmien vakiinnuttamisessa yhteiskuntamme normien mukaisesti. Mikäli suosittelujärjestelmien toiminnassa ei oteta tärkeitä yhteiskunnallisia arvoja suunnittelussa huomioon, todennäköisyys epäreilulle diskriminaatiolle kasvaa. Erilaiset ryhmät ovat usein epäoikeudenmukaisesti edustettuna datassa pohjautuen erityisesti ominaisuuksiin, kuten etninen tausta, sukupuoli, ikä, koulutuksen taso ja varallisuus (Chen et al., 2020).

Suosittelujärjestelmien koulutus tapahtuu malleilla, jotka käyttävät tätä mahdollisesti puolueellista dataa, mikä taas voi mahdollisesti johtaa systemaattiseen diskriminointiin ja vähentyneeseen näkyvyyteen tiettyjen ryhmien osalta. Järjestelmä voi esimerkiksi alkaa ilmentämään etniseen taustaan ja sukupuoleen liittyviä stereotyyppisiä tai vähentämään vähemmistöjen näkyvyyttä palvelussa. (Chen et al., 2020 [Lin, Sonboli, Mobasher & Burke, 2019])

Epäreilouden ilmenemistä voidaan vähentää tasapainottamalla data tai suositustulokset (engl. *rebalancing*) jonkin reiluuteen perustuvan tavoitteen suhteen. Tätä voidaan toteuttaa esimerkiksi asettamalla koulutusdatan positiiviset tunnisteet tasaisesti eri asemassa olevien ryhmien välillä. Myös aiemmin käsitellyt menetelmät, regularisoinnilla ja adversiaalinen oppiminen, sopivat epäreilouden vähentämiseen.

Regularisoinnin ideana tässä yhteydessä on muodostaa kriteerit reiluudelle, minkä pohjalta regularisointi tapahtuu. Näin mallia saadaan optimoitua. Adversiaalisen oppimisen avulla voidaan eristää tiettyjen ryhmien ominaisuuksien vaikutukset. Näiden menetelmien heikkoutena on se, että ne saattavat heikentää suositusten osuvuutta. (Chen et al., 2020)

3.4 Vääristymät mallissa

Suosittelujärjestelmät tyypillisesti perustuvat koneoppimiseen, jossa hyödynnetään datan pohjalta rakennettuja malleja, joiden avulla tehdään suosituksia. Näihin malleihin kohdistuvat induktiiviset vääristymät (engl. *inductive bias*) eivät ole haitallisia, vaan niitä hyödynnetään mallien suunnittelussa niistä saatavien hyötyjen ansiosta (Chen et al., 2020).

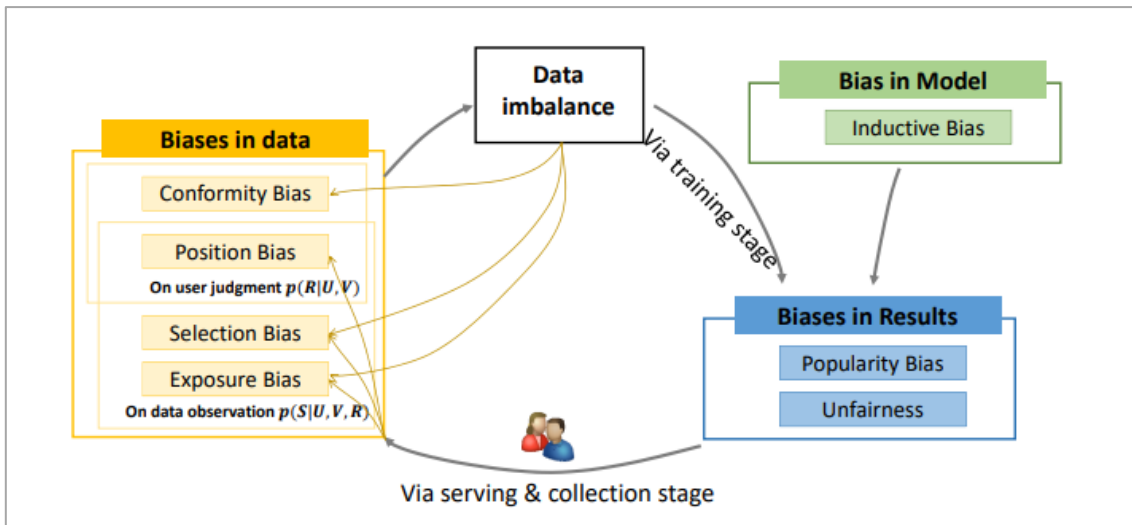
Chenin ja muiden (2020) määritelmän mukaan induktiiviset vääristymät ilmentävät mallin suorittamia olettamuksia oppiakseen tavoitefunktion paremmin ja tehdäkseen yleistyksiä muunkin kuin opetusdatan pohjalta. Tämän tyyppinen toiminta on koneoppimisen ydin: Dataan ja malliin liittyvät muodostetut oletukset ovat olennaisia järjestelmän toiminnan kannalta, koska ennustamattomat tapaukset voivat johtaa sattumanvaraisuuteen (Chen et al., 2020).

Suosittelujärjestelmän rakentaminen edellyttää samalla tavalla oletuksia tavoitefunktion luonteesta, jotta tavoitteellinen toiminta voidaan saavuttaa tehokkaammin (Chen et al., 2020). Induktiivisten vääristymien ansiosta suosittelujärjestelmien mallit voivat potentiaalisesti suorittaa toimintaansa tehokkaammin tarjoten paremmin seulottuja suosituksia. Täten induktiivisiä vääristymiä ei ole optimaalista pyrkiä vähentämään.

3.5 Vääristymien vahvistuminen palautesilmukassa

Suosittelujärjestelmien toimintaa pohjautuu palautesilmukkaan, kuten työssä on aiemmin käsitelty. Vääristymien eri muodot ilmenevät palautesilmukan eri vaiheissa ja ne voivat ajan myötä vahvistua entisestään palautesilmukan myötä (Chen et al., 2020).

Kuva 2 ilmentää kyseistä ilmiötä. Kuvasta havaitaan, että vääristymät datassa johtavat epätasapainoiseen dataan. Epätasapainossa oleva data taas voi vahvistaa tuloksissa ilmeneviä. Vääristymät tuloksissa puolestaan voivat vahvistaa datassa ilmeneviä vääristymiä vaikuttaessa käyttäjien toimintaan ja valintoihin (Chen et al., 2020). Näin ollen tämän tyyppisen silmukan myötä vääristymät voivat aiheuttaa entistä enemmän ongelmia eri osa-alueilla vääristymien vahvistuessa palautesilmukan myötä. Kyseisestä kuvasta ilmenee myös, että epätasapainossa oleva data voi vahvistaa datassa ilmeneviä vääristymiä. Käytännössä siis osa datavääristymistä voi vahvistaa itseään.



Kuva 2. Vääristymien vahvistuminen palautesilmukan myötä (Chen et al., 2020).

Chenin ja muiden (2020) tutkimuksesta ilmenee, että tutkimukset ovat osoittaneet palautesilmukan vahvistavan suosiovääristymiä suosittujen kohteiden suosion kasvaessa entisestään ja vähemmän suosittujen kohteiden suosien laskiessa entisestään. Jiangin ja muiden (2019) tutkimus osoittaa, että tämän tyyppiset vahvistuvat vääristymät johtavat informaatiokupliin (engl. *filter bubble*) ja kaikukammioihin (engl. *echo chamber*) laskiessa sisällön monipuolisuutta ja kasvattaessa käyttäjien yhdenmukaisuutta.

Yhtenäisen datan käyttäminen on suoraviivaisin keino lievittää palautesilmukan aiheuttamia vääristymäongelmia. Yhtenäistä dataa kerätään satunnaisuuteen perustuvilla tiedonkeruumenetelmillä normaalien suosittelumenetelmien sijasta. Tämän yhtenäisen datan avulla saadaan rikottua palautesilmukka, eivätkä erilaiset vääristymät vaikuta siihen. Tämän tyyppinen toiminta kuitenkin huonontaa käyttäjäkokemusta ja palvelun tuottoja, minkä vuoksi oleellinen tutkimuskysymys onkin, kuinka vääristymiä saadaan heikennettyä kyseisellä menetelmällä parhaalla mahdollisella tavalla. (Chen et al., 2020)

Toinen keino vähentää palautesilmukan aiheuttamia vääristymisongelmia on vahvistusoppiminen (engl. *reinforcement learning*), joka ei edellä kuvatun menetelmän tavoin heikennä suositusten suorituskykyä. Suosittelujärjestelmiin liittyy dilemma, joka koostuu siitä, että järjestelmän tulisi suositella kohteita, joiden ennustetaan vastaavan parhaiten käyttäjän mieltymyksiä, mutta myös satunnaisia kohteita, jotta saadaan kerättyä vääristymätöntä palautetta käyttäjiltä. Tämän palautteen avulla voidaan paremmin muodostaa kuva käyttäjän mieltymyksistä. Edellä kuvattua ongelmaa voidaan lieventää rakentamalla vahvistusoppimisagentti, koska se mahdollistaa interaktiivisen suosittelun. (Chen et al., 2020)

Vahvistusoppiminen eroaa perinteisistä menetelmistä siten, että informaation keräämiseen perustuvat tehtävät pohjautuvat vahvistusoppimisagentin ja käyttäjien välisiin sekvenssimäisiin vuorovaikutuksiin. Tämän vuorovaikutuksen vallitessa agentti voi jatkuvasti kehittää strategioitaan käyttäjän aiempaan informaatioon ja palautteeseen pohjautuen. Agentti voi strategioihinsa pohjautuen muodostaa listan kohteista, jotka parhaiten vastaavat käyttäjän mieltymyksiä tai tutkia käyttäjän mieltymyksiä, millä on pitkän aikavälin hyötyjä. Tämän jälkeen käyttäjä antaa palautetta agentille pohjautuen esimerkiksi arvosteluihin ja klikkauksiin. Tämän palautteen avulla päivitetään agentti. Täten aiemmin kuvailtua dilemmaa saadaan tasapainotettua. (Chen et al., 2020)

4 Keskustelu

Viimeaikaiset tutkimukset ovat osoittaneet vääristymien olevan hyvin monimuotoinen ilmiö suosittelujärjestelmiin liittyen. Aihe on vuosien varrella saanut huomiota yhä enemmän, mikä on mielestäni oikea suunta ilmiön kannalta. Ilman aiheellista tutkimusta, vääristymät saattavat jäädä huomioimatta, mikä aiheuttaa ongelmia suosittelujärjestelmien toimivuuden suhteen. Tämä pätee myös muihin suosittelujärjestelmien kohtaamiin ongelmiin. Suosittelujärjestelmien kasvava hyödyntäminen ja yleisyys erilaisissa ympäristöissä edellyttää sitä, että niitä ja niiden erilaisia ongelmia kyetään ymmärtämään hyvin, jotta negatiiviset seuraukset saadaan minimoitua.

Suosittelujärjestelmiin ja erityisesti vääristymiin liittyvä tutkimus on vielä suhteellisen uutta, mikä herättää kysymyksen sen suhteen, mitä kaikkea on vielä löytämättä ja tutkimatta ongelmien suhteen. Esimerkiksi mielipidevääristymät (engl. *sentiment bias*) ovat vasta löydetty uusi vääristymien muoto (Lin, Liu, Xv & Li, 2021), mikä osoittaa, että vääristymien ilmenemistä ei vielä täysin ymmärretä kokonaan ja osa ongelmista on pimennossa. Täten, onko ylipäätään realistista olettaa, että suosittelujärjestelmä ilman ongelmia on toteutettavissa tällä hetkellä? Kenties tämä voi olla tavoitteena, minkä saavuttaminen helpottuu tutkimuksen määrän kasvaessa ja tekniikoiden kehittyessä.

Vääristymistä ja muista ongelmista huolimatta uskon, että suosittelujärjestelmien tulevaisuus on vahvalla pohjalla. Niiden tarjoamat mahdollisuudet käyttäjille ja palveluntarjoajille sisältävät paljon potentiaalia tuoden hyötyä molemmille osapuolille. Nykyisen kehityssuunnan vallitessa voidaan saavuttaa hyödyt entistä paremmin minimoidessa vääristymät ja muut ongelmat. Tämä toki edellyttää sitä, että myös palveluntarjoajat panostavat omien suosittelujärjestelmien toteutukseen ja kehitykseen. Vääristymien luonteesta johtuen eheän järjestelmän toteutuksen tulisi olla jatkuva prosessi, sillä suosittelujärjestelmissä ilmenevät vääristymät ilmenevät ajan myötä. Ei

täten voida olettaa, että suosittelujärjestelmä tulee toimimaan moitteettomasti, vaikka sen käyttöönoton kohdalla saattaisi siltä vaikuttaakin.

Oleellinen ongelma aiheeseen liittyen on mielestäni se, että suosittelujärjestelmien toteutus ja vääristymien ehkäiseminen voi mahdollisesti jäädä täysin mielivaltaiseksi prosessiksi palveluntarjoajien osalta. Suosittelujärjestelmien toiminta ei usein ole käyttäjille läpinäkyvää, minkä seurauksena käyttäjät eivät helposti tiedosta järjestelmän toiminnan puutteellisuutta, vaikka se heihin vaikuttaisikin negatiivisesti. Vääristymien ja muiden ongelmien lieventäminen voi olla tilanteesta riippuen kallista sekä työlästä, minkä seurauksena kaikki saatetaan jättää huomioimatta palveluntarjoajien toimesta. Mikäli palveluntarjoajan vahvana motiivina ovat taloudelliset tuotot, voi suosittelujärjestelmän parantaminen vaikuttaa epäedulliselta vaihtoehdolta, mikäli sen hetkisellä puutteellisella versiolla palvelu menestyy taloudellisesti kaikesta huolimatta läpinäkyvyyden puuteen ansiosta. Tämän tyyppinen menettely mahdollisesti toki kustautuu pidemmällä aikavälillä, mutta tätä ei välttämättä tiedosteta. Kaikesta huolimatta tämä asettaa käyttäjät huonoon asemaan ja antaa palveluntarjoajille liikkumavaraa käyttäjien kustannuksella.

5 Yhteenveto

Suosittelujärjestelmien kohtaamat vääristymät ovat hyvin laaja ja monimuotoinen ilmiö, joka ilmenee vääristymien tyyppien moninaisuutena. Tutkielmassa eriteltiin vääristymien ilmentymiä ja näiden syitä antaen kokonaiskuvaa ilmiöstä sekä sen laajuudesta. Vääristymät johtavat negatiivisiin seurauksiin sekä käyttäjän että palveluntarjoajan osalta, mikäli ne jäävät huomioimatta. Täten suosittelujärjestelmän suunnittelun yhteydessä on tärkeää tiedostaa tekijät, jotka johtavat vääristymiin. Tällä tavalla järjestelmästä voidaan luoda mahdollisimman eheä. Monia työssä läpi käytyjä vääristymien tyyppisiä yhdistää se, että ne ilmenevät järjestelmissä luonnostaan, minkä vuoksi vääristymien torjuminen muodostuu erittäin tärkeäksi ja oleelliseksi.

Vaikka vääristymiä ilmenee, niitä voidaan kuitenkin lieventää ja ehkäistä monella erilaisella menetelmällä. Vääristymiin ja ehkäisymenetelmiin liittyvä tutkimus on kasvanut viime vuosina ja uusien tutkimuksien myötä mahdollisuudet vääristymien lieventämiseksi kasvavat ymmärryksen sekä tiedon kasvaessa.

Lähdeluettelo

- Abdollahpouri, H., Burke, R., & Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. In The thirty-second international flairs conference.
- Abdollahpouri, H., & Mansoury, M. (2020). Multi-sided exposure bias in recommendation. arXiv preprint arXiv:2006.15772.

- Agarwal, A., Wang, X., Li, C., Bendersky, M., & Najork, M. (2019). Addressing trust bias for unbiased learning-to-rank. In *The World Wide Web Conference* (ss. 4-14).
- Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A., & Darwesh, A. (2020). A systematic study on the recommender systems in the E-commerce. *IEEE Access*, 8, 115694-115716.
- Campana, M. G., & Delmastro, F. (2017). Recommender systems for online and mobile social networks: A survey. *Online Social Networks and Media*, 3, 75-97.
- Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2020). Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240*.
- Collins, A., Tkaczyk, D., Aizawa, A., & Beel, J. (2018). A study of position bias in digital library recommender systems. *arXiv preprint arXiv:1802.06565*.
- Gomez-Uribe, C. A., & Hunt, N. (2015). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4), 1-19.
- Jiang, R., Chiappa, S., Lattimore, T., György, A., & Kohli, P. (2019). Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (ss. 383-390).
- Karimi, M., Jannach, D., & Jugovac, M. (2018). News recommender systems—Survey and roads ahead. *Information Processing & Management*, 54(6), 1203-1227.
- Krishnan, S., Patel, J., Franklin, M. J., & Goldberg, K. (2014). A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. *Proceedings of the 8th ACM Conference on Recommender systems* (ss. 137-144).
- Lin, C., Liu, X., Xu, G., & Li, H. (2021). Mitigating sentiment bias for recommender systems. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (ss. 31-40).
- Lin, K., Sonboli, N., Mobasher, B., & Burke, R. (2019). Crank up the volume: preference bias amplification in collaborative recommendation. *arXiv preprint arXiv:1909.06362*.
- Lops, P., Gemmis, M. D., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. Teoksessa Ricci, F., Rokach, L., & Shapira, B. (2011). *Recommender systems handbook*, (ss.73-105). Springer, Boston, MA.
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *Ai & Society*, 35(4), 957-967.
- Shah, K., Salunke, A., Dongare, S., & Antala, K. (2017). Recommender systems: An overview of different approaches to recommendations. *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)* (ss. 1-4). IEEE.
- Steck, H. (2010, July). Training and testing of recommender systems on data missing not at random. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ss. 713-722).
- Wang, X., Zhang, R., Sun, Y., & Qi, J. (2021, March). Combating selection biases in recommender systems with a few unbiased ratings. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (ss. 427-435)

