

Jovan Karmakka

**Suosittelujärjestelmät ja niiden vaikutukset
päätöksentekoon**

Tietotekniikan kandidaatintutkielma

14. joulukuuta 2022

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

Tekijä: Jovan Karmakka

Yhteystiedot: jovan.r.karmakka@student.jyu.fi

Ohjaaja: Tytti Saksa

Työn nimi: Suosittelujärjestelmät ja niiden vaikutukset päätöksentekoon

Title in English: Recommender systems and their influence on decision-making

Työ: Kandidaatintutkielma

Opintosuunta: Tietotekniikka

Sivumäärä: 21+0

Tiivistelmä: Suosittelujärjestelmiä käytetään laajasti suodattamaan suuri määrä informaatiota. Niiden tehtävänä on luoda käyttäjälle personoitua sisältöä ja helpottaa päätöksenteossa. Tässä tutkielmassa tarkastellaan erilaisten suosittelujärjestelmien toimintaa ja käyttökohteita. Tämän lisäksi selvitetään, mitä kirjallisuudessa kerrotaan suositusjärjestelmien vaikutuksista käyttäjän valintoihin.

Avainsanat: hakualgoritmi, suosittelualgoritmi, suosittelujärjestelmä,

Abstract: Recommender systems are widely used to filter a large amount of information. Their objective is to create personalized content for the user and ease decision-making. This bachelor's thesis examines how different recommender systems function and use-cases of various recommendation systems. We will also find out what the literature says about the influence of recommendation systems on our choices.

Keywords: search algorithm, recommendation algorithm, recommendation system,

Sisällys

1	JOHDANTO	1
2	SUOSITTELUJÄRJESTELMÄT SOVELLUKSISSA JA INTERNETISSÄ.....	2
	2.1 Järjestelmien kehitys ja tulevaisuus	2
	2.2 Erilaisia käyttökohteita.....	4
3	SUOSITTELUMENETELMÄT JA ALGORITMIT	7
	3.1 Sisältöperustaiset menetelmät	7
	3.2 Yhteistoiminnallinen suodattaminen	9
	3.3 Hybridimenetelmä.....	11
	3.4 Uudet menetelmät	11
4	SUOSITTELUJÄRJESTELMIEN VAIKUTUKSET PÄÄTÖKSENTEKOON	13
5	YHTEENVETO.....	15
	LÄHTEET	16

1 Johdanto

Informaation ja valintojen määrä internetissä on valtava. Internetissä käyttäjät kohtaavat jatkuvasti päätöksiä siitä, minkälaista sisältöä kulutetaan tai tuotteita ostetaan. Tämä valintojen määrä kasvaa jatkuvasti kiihtyvällä tahdilla, mutta käyttäjien kyky seuloa ja löytää etsimäänsä informaatiota ei kasva samalla tahdilla. Internetin käyttäjät ovatkin usein tilanteessa, jossa informaation määrän käsitteleminen on vaikeaa. On kehitetty erilaisia haku- ja suosittelujärjestelmiä auttamaan tämän informaation seulonnassa. Suosittelujärjestelmät käyttävät hyväkseen käyttäjän mieltymyksiä ja vuorovaikutushistoriaa muodostaakseen sisältöehdotuksia. Näiden järjestelmien tarkoituksena on esimerkiksi personoida sisältöä käyttäjälle. Näillä ehdotuksilla päätöksentekoa helpotetaan ja käyttäjiä suojellaan liialliselta valinnan määrältä. Näiden järjestelmien toiminnassa on kuitenkin eroja. Huomaamatta suosittelujärjestelmät voivatkin ohjata käyttäjiä valintojen tekemisessä ja tehdä ne jopa heidän puolestaan.

Kirjallisuuden perusteella suosittelujärjestelmät jaotellaan kolmeen ryhmään. Nämä ovat sisältöperustainen menetelmä, yhteistoiminnallinen suodattaminen sekä hybridimenetelmä, jossa nämä kaksi metodia yhdistetään. Sisältöperustaisessa menetelmässä ideana on muodostaa käyttäjästä profiili, jonka perusteella tuotteita suositellaan. Yhteistoiminnallinen suodattaminen kerää käyttäjäjoukosta ominaisuuksia ja muodostaa ryhmiä, joiden mukaan suosituksia luodaan. Kolmannessa menetelmässä yhdistetään molempien menetelmien hyviä puolia, jotta saataisiin luotua tarkempia suosituksia.

Tässä tekstissä suosittelujärjestelmillä tarkoitetaan sellaisia järjestelmiä, jotka käyttäjän toiminnan tai aiemman toiminnan perusteella muodostavat personoituja suosituksia. Suositukset voivat olla esimerkiksi mediasisältöä, uutisia tai tuotteita. Suosittelujärjestelmiä yleensä arvioidaan sen mukaan kuinka tarkkoja niiden tuottamat suositukset ovat. Kirjallisuuden perusteella vähemmälle huomiolle jää kuinka suosittelujärjestelmät vaikuttavat käyttäjän päätöksentekoprosessiin. Ensimmäisessä kappaleessa tarkastellaan suosittelujärjestelmiä yleisellä tasolla, jonka jälkeen katsotaan tarkemmin minkälaisia menetelmiä suosittelujärjestelmät käyttävät. Lopuksi selvitetään, onko näillä järjestelmillä vaikutusta käyttäjän mieltymysten muodostumiseen tai valintoihin.

2 Suositteijärjestelmät sovelluksissa ja internetissä

Nykyään ei voi olla törmäämättä suosituksiin. Suosituksia tulee käyttäjälle vastaan verkko-kaupassa asioidessa esimerkiksi tuotteista, joita muut saman tuotteen ostajat ovat ostaneet yhdessä käyttäjän tarkasteleman tuotteen kanssa. Suositteijärjestelmän avulla käyttäjälle voidaan luoda täsmällisempiä mainoksia ja ehdotuksia. Kirjoittajien Pazzanin ja Billsuksen (2007) mukaan tyypillinen skenaario suosittelujärjestelmälle onkin websovellus, joka esittää yhteenvedon tuotteista, joista käyttäjä sitten valitsee haluamansa. Esimerkiksi uutisivustot esittävät etusivulistauksen uutisotsikoista tai verkkokaupat esittävät yhteenvedon tuotteista tai mainoksista.

Siitä huolimatta, että verkkosivut lähettävät tiedot HTML-kuvauskielellä, käyttäjälle näytettävä verkkosivu on tyypillisesti luotu dynaamisesti (Pazzani ja Billsus 2007). Taustalla on yleisesti tietokanta, jonka avulla verkkosivuja luotaessa käyttäjälle luodaan ja näytetään lista tuotteista. Näytettävää tietoa on enemmän kuin verkkosivulle mahtuu ja tuotteista on valittava vain osa. Suositteijärjestelmän avulla oleellinen tieto saadaan esille ja järkevään järjestykseen. J. Chenin ym. (2020) mukaan suosittelujärjestelmät ovatkin tehokkain tapa lievittää tiedon ylikuormitusta.

Seuraavaksi tarkastellaan tarkemmin suosittelujärjestelmien kehitystä ja ominaispiirteitä. Tämän jälkeen katsotaan minkälaisissa tilanteissa suosittelujärjestelmiä käytetään, ja mitkä ovat niiden tyypillisiä käyttökohteita.

2.1 Järjestelmien kehitys ja tulevaisuus

Calero Valdezin, Zieflen ja Verbertin (2016) mukaan ensimmäinen suosittelujärjestelmä on Tapestry. Tämä järjestelmä käytti hyväkseen yhteistoiminnallista suodattamista suositusten antamiseen ja suodatti suuren määrän sähköposteja. Tapestry sisälsi järjestelmän, joka perustui sähköpostien merkitsemiseen ja myöhemmin sähköpostien suodattamiseen näiden mukaan. Suositusmenetelmiä tarkastellaan tarkemmin luvussa 3.

Termiä suosittelujärjestelmä (engl. recommender system) käytettiin ensimmäisen kerran 1997

kun Resnick ja Varian kirjoittivat artikkelin, jossa he kuvailivat suosittelujärjestelmien olevan enemmänkin päätöksenteon työkaluja kuin tiedonhaun työkaluja (Calero Valdez, Zieffle ja Verbert 2016). Tämän jälkeen suosittelujärjestelmien käyttö ja tutkiminen on kiihtynyt ja onkin tänä päivänä kiinnostava aihe (Afsar, Crump ja Far 2021).

Kuinka hakujärjestelmä sitten eroaa suosittelujärjestelmästä? Keyvanin ja Huangin (2022) mukaan hakujärjestelmän tehtävä on ymmärtää monimutkaisia luonnollisen kielen malleja ja palauttaa paras vastaus sekunnin murto-osassa. He kertovat tiedonhaun olevan prosessi, jossa haetaan ja palautetaan asianmukaisin tulos käyttäjän antaman tiedon perusteella. Perinteiset suosittelujärjestelmät sen sijaan pyrkivät tarjoamaan palautetta käyttäjän ja järjestelmän välisen vuorovaikutuksen perustella (Keyvan ja Huang 2022). Hakujärjestelmän antamat tulokset ovat aina tuotettu käyttäjän syötteen perusteella, mutta suosittelujärjestelmän tulokset muodostetaan erilaisten menetelmien kautta ja voivat olla alttiita virheille.

Verkkosivujen aikakaudella sisältöpohjaista suosittelujärjestelmää käytettiin menestyksellä suositteluun verkkosivuja (Lops ym. 2019). Pian huomattiin, että pelkästään sisältöön perustuvalla suodattamisella on heikkouksia, kun tämä järjestelmä ei ota huomioon tuotteiden laatua suositusprosessin aikana. Lops ym. (2019) esittävät, että yhteistoiminnallinen suodattaminen yleisesti ottaen ottaa tämän asian huomioon eksplisiittisen tuotearvion avulla. Heidän mukaansa sisältöön perustuva järjestelmä sen sijaan ottaa epäsuorasti huomioon tuotteiden suosion implisiittisellä palauteskenaariolla.

Vuosikymmenten ajan on yritetty löytää ennustustarkkuudessaan parasta suosittelumallia. Ennusteen tarkkuuden vuoksi puhtaasti sisältöön perustuvat menetelmät vaikuttivat jo vanhentuneilta (Lops ym. 2019). Niitä käytettiin toisten järjestelmien tukemiseen niin sanotuissa kylmäkäynnistämistilanteissa. Viimevuosina onkin yleistyneet niin sanotut hybridijärjestelmät, jotka kykenevät keräämään tietoa erilaisista uusista tietolähteistä, kuten käyttäjän tuottamasta sisällöstä. Nämä järjestelmät pystyvät keräämään tietoa käyttäjän tuottamasta jatkuvasta virrasta ja tekemään parempia suosituksia. Lopsin ym. (2019) mukaan tulevaisuudessa tullaan näkemään tutkimussuuntia metatietojen ja sivutietojen roolin selvittämiseksi suositusjärjestelmissä.

Lopsin ym. (2019) mukaan sisältöön perustuvia suosittelujärjestelmiä on viime aikoina käy-

tetty myös luonnollisten kielten käsittelyssä (engl. natural language processing, NLP). Esimerkiksi sanaupotukset on koettu hyödyllisiksi ja ne ovat saaneet viime aikoina laajempaa käyttöä sen jälkeen, kun niiden huomattiin toimivan hyvin monessa luonnollisen kielen käsittelyä vaativassa tehtävässä. Zhangin, Lun ja Jinin (2021) mukaan suosittelujärjestelmät ovat suuressa roolissa myös digitaalisessa äänijärjestelmissä kuten Siri ja Google home.

Viime vuosina on nähty kiinnostuksen kohdistuvan selitettävään tekoälyyn (engl. explainable artificial intelligence) monimutkaisen koneoppimisen kehityksen myötä (Lops ym. 2019). Tämä kiinnostus näkyy erityisesti terveydenhuollon osa-alueella, jossa vastuullisuus ja selitettävyys ovat tärkeitä ominaisuuksia.

Useita muitakin tekoälyn sovelluksia on viime vuosina sovitettu suosittelujärjestelmiin. Tämän vuoksi suosittelujärjestelmät ovat saaneet uuden aikakauden, kun voidaan luoda parempia suhteita käyttäjien ja tuotteiden välille. Tekoälyyn perustuvat suositukset ovatkin osoittautuneet paremmiksi kuin tavalliset suositusjärjestelmät (Zhang, Lu ja Jin 2021). Tämän lisäksi se mahdollistaa myös monimutkaisempien tietojen esittämisen väestöstä, rakenteesta sekä virtuaalisesta ja kontekstuaalisesta tiedosta.

Seuraavassa luvussa tarkastellaan, minkälaisia käyttökohteita suosittelujärjestelmillä on.

2.2 Erilaisia käyttökohteita

Suosittelujärjestelmiä käytetään laajasti verkkoalustoilla, jotta saataisiin kerättyä käyttäjän mieltymykset ja kiinnostuksen kohteet. Näitä puolestaan käytetään luomaan tuote- ja palvelusuosituksia. Nämä suositukset ovat tärkeitä tuomaan hyötyä ja tuottoa pienellä riskillä (Calero Valdez, Ziefle ja Verbert 2016). Sähköisen kaupankäynnin alalla suosittelujärjestelmät ovatkin osoittautuneet erittäin suosituiksi juuri hakemaan ja suosittelemaan kaikista olennaisinta informaatiota tavaroista, tuotteista ja palveluista (Sharma, Shaikh ja Li 2021).

Yksi suuri käyttökohde suosittelujärjestelmille on verkossa sijaitseva sähköinen kaupankäynti ja vähittäismyynti (Sharma, Shaikh ja Li 2021). Tällaisessa yhteydessä nämä järjestelmät ovat tuomassa lisäarvoa yritykselle ja lisätäkseen myyntiä. Adomaviciuksen ym. 2013 mukaan 10-30 prosenttia verkkokaupan myynnistä voi olla suosittelujärjestelmän ansiota.

Hyvä suosittelujärjestelmä voi lisätä myös asiakkaiden luottamusta ja näin olla suuri etu kilpailijoihin verrattuna (Cosley ym. 2003). Lisäarvoa suosittelujärjestelmistä tulee, kun alustat kuten Netflix tai YouTube voivat kerätä suuren määrän tietoa käyttäjien kaupankäynnistä ja käyttäytymisestä ja luoda näiden pohjalta assosiaatioita. Tätä tietoa hyödyntämällä ne voivat tarjota räätälöityjä ohjelmia tai ostosuosituksia katsojille sekä asiakkaille. Sharman, Shaikhin ja Lin (2021) mukaan nämä ominaisuudet tuovat hyvän tilaisuuden kaupallistamiselle sekä auttavat sähköistä kaupankäyntiä lisäämään ristimyyntiä. Zhangin ym. (2021) mukaan motivaatio suosittelujärjestelmien tutkimustyölle tuleekin liiketoiminnan puolelta, jotta potentiaalista voittoa voidaan kasvattaa.

Suosittelujärjestelmiä käytetään laajasti myös suoratoistopalveluissa, kun käyttäjälle laaditaan sisältölistauksia. Weidhaasin ym. (2021) mukaan laadukkaalla suosittelujärjestelmällä on tärkeä rooli sovelluksen käyttökokemuksessa. Hänen havaintojensa mukaan suoratoistopalveluissa on liian paljon valinnanvaraa ja valinnan vaikeus johtaa tyytymättömyyteen. Tämän lisäksi hän huomasi, että käyttäjät luottivat sovelluksen luomiin suosituksiin, jos ne olivat aikaisemmin olleet onnistuneita. Sharman ym. (2021) mukaan suosittelujärjestelmillä voidaan myös lyhentää aikaa, jonka käyttäjät käyttävät etsiessään haluamaansa sisältöä.

Tämän hetken suurin suoratoistopalvelu Netflix on kertonut, että se voisi menettää jopa miljardin Yhdysvaltain dollaria vuosittain menetettyinä asiakkaina, jos se ei olisi ottanut suositusjärjestelmää käyttöön ja parantanut sitä (Sharma, Shaikh ja Li 2021). Netflix onkin järjestänyt avoimia kilpailuja, joissa tarkoituksena on kehittää paras suosittelujärjestelmä. Uuden järjestelmän on täytynyt olla 10 prosenttia parempi kuin sen tämänhetkinen järjestelmä ja palkintona on ollut miljoona Yhdysvaltain dollaria. Aggarwalin ym. (2016) mukaan Netflixin palkintokilpailu on vaikuttanut merkittävästi tutkimusyhteisöön ja suosittelujärjestelmien kehitykseen.

Tilaukseen perustuvat musiikinkuuntelualustat kuten Spotify ovat yleistyneet viime aikoina. Schedlin, Kneesin ja Gouyon (2017) mukaan suosittelujärjestelmät ovat nykyään keskeinen osa näitä palveluita. Ennen suosituksia on tehty kertaluontoisesti ja suositeltu esimerkiksi albumia tai yksittäistä kappaletta. Näiden uusien alustojen myötä suositukset täytyy tehdä kokonaisista soittolistoista ja ehdotuksia täytyy luoda myös uudesta sisällöstä, josta kuuntelija voisi pitää. Nämä muutokset tuovat uusia haasteita suosittelujärjestelmille. Schedlin ym.

(2017) mukaan suosittelujärjestelmien kehitystä pitäisi lisätä musiikkialalla juuri sen erityispiirteiden vuoksi.

Näiden lisäksi suosittelujärjestelmiä käytetään sosiaalisen median palveluissa ja uutisten suositteluun. Karimin, Jannachin ja Jugovacin (2018) mukaan varsinkin uutissuosituksset ovat aktiivinen tutkimusaihe, ja viime vuosina on tapahtunut merkittävää edistymistä eri suuntiin. Yksi käyttökohte on myös sosiaalisen median palvelut, jotka eroaakin tavanomaisesta suositusten tekemisestä, kun huomioon täytyy ottaa myös sosiaaliset suhteet (Sperli ym. 2018).

3 Suositelumenetelmät ja algoritmit

Tässä kappaleessa tarkastellaan erilaisia suositusmenetelmiä ja -algoritmeja, joita suositelujärjestelmissä käytetään. Suositelujärjestelmiä on pääasiassa kolmenlaisia, sisältöperustaisia, yhteistoiminnallisia ja hybridimenetelmiä (Zhang ym. 2019). Ensiksi tarkastellaan näitä kahta menetelmää, jonka jälkeen katsotaan hybridimenetelmää, joka tyypillisesti yhdistää näitä kahta menetelmää. Lopuksi tarkastellaan uusia suositelumenetelmiä ja kuinka tekoälyä käytetään kyseisissä järjestelmissä.

3.1 Sisältöperustaiset menetelmät

Sisältöperustaiset menetelmät ovat suosittu suositusmenetelmä. Niiden menetelmissä on hie-man eroja, mutta pääidealtaan ne ovat kaikki samanlaisia. Pazzanin ym. (2007) mukaan näillä järjestelmillä on keino kuvailla suositeltavan tuotteen tietoja sekä keino rakentaa käyttäjästä profiili, joka sisältää kuvaukset käyttäjän mieltymyksistä. Profiili luodaan ja päivitetään usein automaattisesti käyttäjän tuottaman toiminnan perustella (Pazzani ja Billsus 2007). Sisältöperustainen menetelmä vertailee näitä kahta tietoa ja luo suositukset niiden pohjalta. Erilaisia oppimisalgoritmeja on mukautettu oppimaan käyttäjäprofileja ja oppimisalgoritmin valinta riippuu yleensä sisällön esityksestä. Sisältöperustaisen menetelmän käyttökohteita ovat esimerkiksi nettisivujen, uutisten, ravintoloiden, ohjelmien sekä erilaisten myytävien tuotteiden suositukset.

Sisältöperustaisen menetelmän pääidea on luokitella tuotteet avainsanoilla, selvittää mistä käyttäjä pitää ja suositella samankaltaisia tuotteita. Ricci, Rokach ja Shapira (2011) esittävät, että suositusten muodostaminen sisältöperustaisella mallilla on arkkitehtuuriltaan kolmiosainen, joista jokainen on erillinen komponentti.

Riccin ym. (2011) mukaan ensimmäinen on sisällönerittelykomponentti. Siinä tietokohteet analysoidaan ja niistä erotellaan niiden ominaispiirteet, esimerkiksi nettisivut voidaan muuttaa avainsanavektoreiksi. Tämän vaiheen tarkoitus on muuttaa tuotteiden sisältö sellaiseen muotoon, että sitä voidaan hyödyntää seuraavissa vaiheissa. Seuraavaksi on profiilin oppimisvaihe. Tässä vaiheessa kerätään ja yleistetään käyttäjä mieltymykset. Tässä käytetään

yleensä apuna koneoppimisen malleja, joiden avulla voidaan päätellä profiili käyttäjän kiinnostuksen kohteista, joista hän pitää tai ei pidä. Lopuksi on suodatusvaihe. Tämä komponentti käyttää hyväkseen käyttäjän profiilia suositellakseen tuotteita. Tuloksena on joko binäärinen tai jatkuva relevanssiarvo (engl. relevance judgment), joka lasketaan käyttämällä vektoreiden laskutoimituksia kuten pistetuloa.

Riccin ym. (2011) mukaan sisältöperustaisissa menetelmissä on yhteistä piirteidenirrotusmenetelmä TF-IDF (term frequency–inverse document frequency), joka laskee olennaiset piirteet sanojen painokertoimien avulla. Yleinen tapa on avainsanapohjainen vektoriavaruusmalli. Kuvataan seuraavaksi Riccin ym. (2011) esittämän avainsanapohjaisen vektoriavaruusmalli.

Vektoriavaruusmallissa (engl. vector space model) jokainen asiakirja esitetään vektorilla n -ulotteisessa avaruudessa. Jokainen ulottuvuus vastaa sanaa asiakirjan kokonaissanastosta, toisin sanoen jokainen asiakirja esitetään sanojen painoarvojen vektorina, jossa jokainen paino ilmaisee assosiaatioasteen asiakirjan ja sanan välillä (Ricci, Rokach ja Shapira 2011).

Olkoon $D = \{d_1, d_2, \dots, d_N\}$ joukko asiakirjoja ja $T = \{t_1, t_2, \dots, t_N\}$ sanakirja asiakirjan sanoista. T saadaan luonnollisen kielen prosessoinnin työkaluilla. Jokainen asiakirja d_j on kuvattuna vektorina n -ulotteisessa vektoriavaruudessa joten $d_j = \{W_{1j}, W_{2j}, \dots, d_{nj}\}$ missä W_{kj} on painoarvo sanalle t_k asiakirjassa d_j .

Sisältöperustaisella menetelmällä on useita vahvuuksia, etenkin yhteistoiminnalliseen menetelmään verrattuna. Riccin ym. (2011) mukaan se pystyy tekemään suositukset pelkästään yhden käyttäjän toiminnan perusteella eikä tarvitse muita käyttäjiä. Se on myös tarvittaessa läpinäkyvä, sillä suosituksen antamisen perusteet on helposti saatavilla. Se pystyy myös suosittelemaan tuotteita, joita kukaan ei ole vielä arvostellut (Sharma, Shaikh ja Li 2021).

Sisältöperustainen suosittelemenetelmä ei ole vailla heikkouksia. Mun (2018) mukaan yksityisyydensuojan vahvistumisen myötä käyttäjän profiilin luominen on vaikeutunut eikä se pysty pelkästään sisällön perustella ottamaan huomioon seikkoja, joita tarkkaan suositukseen tarvitaan. Se ei pysty myöskään olemaan kovin intuitiivinen ja suositukset ovat usein hyvin samankaltaisia (Ricci, Rokach ja Shapira 2011). Suositukset myös uusille käyttäjille ovat huonot sillä käyttäjäprofiilissa ei ole riittävästi tietoa hyvän suosituksen tekemiseen.

3.2 Yhteistoiminnallinen suodattaminen

Kollaboratiivinen eli yhteistoiminnallinen suodattaminen (engl. collaborative filtering) on yksi käytetyimmistä ja menestyneimmistä suosittelujärjestelmistä (R. Chen ym. 2018). Se kerää ensiksi käyttäjän mieltymyksistä tietoa esimerkiksi tuotteiden arvioinnin tai ostohistorian perusteella ja sitten vertailee käyttäjiä keskenään muodostaakseen suositukset (Melville ja Sindhvani 2010). Tämä menetelmä kykenee tekemään laadukkaita suosituksia mutta sen tehokkuus heikkenee datan ja käyttäjien määrän kasvaessa. Melvillen ja Sindhwanin (2010) mukaan yhteistoiminnallinen suodattaminen voidaan jakaa kahteen menetelmään: tilastollisia menetelmiä käyttävä muistiin perustuva menetelmä (engl. memory-based method) ja malliperustainen määrittely (engl. model-based).

Muistiin perustuvasta mallista käytetään myös nimitystä naapuruspohjainen menetelmä (engl. neighborhood-based). Tässä mallissa aktiivisen käyttäjän mieltymysten samankaltaisuutta verrataan muihin käyttäjiin ja näin saadaan valittua muiden käyttäjien osajoukko. Tästä joukosta tuotetaan suositukset käyttäen painotettua yhdistelmää muiden käyttäjien tykkäyksistä.

Monet naapuruspohjaiset menetelmät toimivat samalla idealla. Tarkastellaan seuraavaksi Melvillen ym. (2010) esittelemää muistiin perustuvaa mallia.

Ensiksi määritetään painoarvo kaikille käyttäjille suhteessa aktiiviseen käyttäjään samankaltaisuuden perusteella. Tässä painoarvo $W_{a,u}$ on samankaltaisuuden mitta käyttäjän u ja aktiivisen käyttäjän a kanssa. Kirjallisuudessa on käytetty useita samankaltaisuuden mittareita kuten Spearmanin tai Kendallin järjestyskorrelaatiokerrointa, mutta Pearsonin korrelaatiokerroimen on havaittu pärjäävän parhaiten (Melville ja Sindhvani 2010). Tämä on kuvattu alla kaavassa 3.1.

$$W_{a,u} = \frac{\sum_{i \in I} (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_{a,i} - \bar{r}_a)^2 \sum_{i \in I} (r_{u,i} - \bar{r}_u)^2}} \quad (3.1)$$

I on joukko tuotteita joita molemmat käyttäjät ovat suosittelleet, $r_{u,i}$ on käyttäjän u arvostelu tuotteelle i ja \bar{r}_u on käyttäjän u antama keskiarvoluokitus. Vastaavasti $r_{a,i}$ on käyttäjän a arvostelu tuotteelle i ja \bar{r}_a on käyttäjän a antama keskiarvoluokitus. Sitten valitaan K mää-

rä käyttäjiä, joilla on suurin samankaltaisuus aktiivisen käyttäjän kanssa. Lopuksi ennusteet lasketaan yleisesti poikkeamien painotettuna keskiarvona naapurin keskiarvosta. Kuvattu alla kaavassa 3.2.

$$P_{a,i} = \bar{r}_a = \frac{\sum_{u \in K} (r_{u,i} - \bar{r}_u) \times W_{a,u}}{\sum_{u \in K} W_{a,u}} \quad (3.2)$$

Tässä $P_{a,i}$ on ennustus käyttäjälle a tuotteesta i . $W_{a,u}$ on käyttäjien a ja u samankaltaisuus ja K kaikista samankaltaisten naapureiden joukko.

Malliperustaisessa menetelmässä käytetään koneoppimista sekä tiedonlouhintaa ennakoivien ennustusten tuottamiseen (Aggarwal ym. 2016). Niissä tapauksissa, joissa malli on parametrisoitu, parametrit opitaan optimointimenetelmiä käyttäen. Tällaisia mallipohjaisia menetelmiä ovat esimerkiksi päätöspuut ja Bayesin teoreema. Näillä menetelmillä voidaan luoda kattavia ennustuksia, vaikka käyttäjistä saadut tiedot ovat vähäiset.

Vaikka malliperustaista määrittelyä pidetään arvossaan sen yksinkertaisuuden vuoksi, on se kuitenkin luonteeltaan heuristinen ja siitä syystä se ei sovellu kaikkiin tilanteisiin. Aggarwalin ym. (2016) mukaan ero naapuruuspohjaisen ja malliperustaisen menetelmän ero on muutenkin keinotekoinen, sillä muistiin perustuvia menetelmiä voidaan pitää samankaltaisuuteen perustuvina malleina vaikkakin heuristisina.

Yhteistoiminnallisen suodattamisen vahvuus on, että se ei tarvitse tuotteiden tietoja tehdäksseen suosituksia ja se on helppo toteuttaa (Aggarwal ym. 2016). Yhteistoiminnallisen menetelmän huono puoli on sen riippuvuus muista käyttäjistä. Jos toisten käyttäjien mieltymyksistä ei ole tietoa, suosituksen muodostaminen on vaikeaa (Ricci, Rokach ja Shapira 2011; Mu 2018). Siinä missä sisältöperustaisesta suosituksesta saadaan helposti tieto tuotteen ominaisuuksista, joiden perustella suositus on tehty, sisältöperustainen on kuin musta laatikko. Sen suositus perustuu vain siihen, että tuntematon käyttäjä pitää samanlaisista asioista.

3.3 Hybridimenetelmä

Hybridimenetelmä pyrkii saamaan aikaan parempia suosituksia yhdistämällä sisältöperustaisen menetelmän ja yhteistoiminnallisen suodattamisen (Melville ja Sindhvani 2010). Hybridijärjestelmiä voidaan käyttää tehostamaan suosituksia, sillä se voi hyödyntää useaa tietolähdettä samanaikaisesti. Sitä voidaan myös käyttää parantamaan jo olemassa olevaa suositusjärjestelmää. Edellä käytiin läpi kahden erilaisen menetelmän hyviä ja huonoja puolia, hybridimenetelmällä näitä huonoja puolia voidaan vähentää (Batmaz ym. 2019). Toiset suositusmenetelmät pärjäävät huonosti kylmäkäynnistystilanteessa, kun tietoa muista käyttäjistä on niukasti. Toisaalta kun datan määrä kasvaa, ongelmana voi olla myös suositusten yksipuolisuus. Hyödyntämällä useaa menetelmää yritetään ratkaista tällaiset ongelmat ja saada yleisesti ottaen paremmat suositukset. On havaittakin, että hybridimallilla saadaan paremmat suositukset kuin pelkästään yksittäisellä suosittelujärjestelmällä (R. Chen ym. 2018). Siitä syystä se onkin käyttökelpoinen suosittelujärjestelmä.

Mun (2018) mukaan hybridimenetelmät voidaan jakaa kolmeen erilliseen menetelmään. Monoliittiseen menetelmään (engl. monolithic method), rinnakkaiseen menetelmään (engl. parallel method) ja liukuhihnamenetelmään (pipeline method). Monoliittisessa metodissa yhdistetään monta erilaista suosittelustrategiaa yhdeksi algoritmiksi. Kahdessa muussa metodissa tarvitaan vähintään kaksi erillistä suosittelumetodia. Rinnakkaisessa järjestelmässä useampi erillinen suosittelujärjestelmä tuottaa omat suosituslistaukset, jotka sitten lopussa yhdistetään yhdeksi suositukseksi. Liukuhihnamallissa yhdistetään kaksi tai useampi metodi ketjussa, Tällöin edellisen menetelmän tulosta käytetään lähtöarvona seuraavalle metodille.

3.4 Uudet menetelmät

Tekoälyä (engl. artificial intelligence) ja erityisesti koneoppimista käytetään Zhangin ym. (2021) mukaan suositusjärjestelmien ennusteiden parantamiseen ja kylmäkäynnistysongelmien ratkaisemiseen. Se auttaa myös tilanteissa, joissa tietoa on niukasti käytettävissä suositusten muodostamiseen. Tekoälyn lisääminen suositusjärjestelmiin on parantanut käyttäjäkokemusta ja lisännyt käyttäjän tyytyväisyyttä (Zhang, Lu ja Jin 2021). sillä saadaan myös aikaan paremmat suositukset kuin tavallisilla metodeilla. Tekoälyn käyttäminen onkin aloit-

tanut uuden ajan suositusten muodostamisessa, kun käyttäjän ja tuotteen välisistä suhteista saadaan edistyneempää tietoa. Zhang, Lu ja Jin (2021) esittävät, että kaikista tekoälyn metodeista suositusten muodostamisessa hyödyllisimpiä ovat puheentunnistus (engl. speech recognition) sekä tietokonenäkö (engl. computer vision).

Kun muodostetaan suosituksia esimerkiksi elokuvista, on niitä voitu muodostaa tähtiarvostelun pohjalta aikaisemmin mainituilla menetelmillä. Puheentunnistuksen kehityksen myötä voidaan myös hyödyntää myös metadata, elokuvan kuvaus ja käyttäjien kirjoittamat arvostelut (Zhang, Lu ja Jin 2021). Useita mielipiteen louhimista (engl. opinion mining) ja aihe-mallinnusta (engl. topic modeling) käyttävää metodia on jo kehitetty luonnollisen kielen käsittelyyn. Suositusten muodostamisessa näitä metodeja on alettu käyttää enenevässä määrin ja niiden on havaittu hyödyttävän monia suosittelujärjestelmiä etenkin täyttämään arvostelumatriiseja silloin kun arvosteluja on niukasti (Zhang, Lu ja Jin 2021).

Vapaasti kirjoitetun tekstin informaatiosta on etua silloinkin, kun tietoa on riittävästi. Zhangin ym. (2021) mukaan käyttäjien kirjoitettua tekstiä hyödynnetään löytämään piileviä ominaisuuksia käyttäjästä ja näin saadaan tarkempia sekä läpinäkyvämpiä suosituksia. Heidän mukaansa arvostelutekstejä voidaan sisällyttää verkkotunnusten välisiin suositusmenetelmiin, joissa käyttäjävektorit kartoitetaan epälineaaristen funktioiden avulla.

Suositusjärjestelmät ovat hyötyneet myös tietokonenäön kehitymisestä etenkin muodin ja korujen myynnissä. Zhangin ym. (2021) mukaan tietokonenäön ja syväoppimisen yhdistämisellä on saatu erinomaisia parannuksia suositustuloksiin. Tietokonenäön avulla tuotteiden kuvista saadaan eroteltua piirteitä, joita voidaan yhdistää muihin suositusjärjestelmiin.

Syväoppimisen (engl. deep learning) kehittymisen myötä sitä on alettu käyttää myös suosittelujärjestelmissä. Sen on havaittu parantavan huomattavasti suositusjärjestelmien suorituskykyä (Covington, Adams ja Sargin 2016). Chenin ym. (2018) mukaan syväoppimisen hyödyntäminen suositusten muodostamisessa on hyödyllistä. Se pystyy löytämään piileviä assosiaatioita ja sillä on kyky erottaa piirteitä käyttäjän tuotearviosta ilman samankaltaisuusmittareita. Zhangin ym. (2021) mukaan syväoppimisen saralla varsinkin usean kerroksen neuroniverkkoa on hyödynnetty puheentunnistuksessa ja tietokonenäössä.

4 Suosittelujärjestelmien vaikutukset päätöksentekoon

Edellä käsiteltiin suosittelujärjestelmien käyttökohteita ja tarkasteltu laajalti erilaisia menetelmiä, joilla käyttäjälle voidaan luoda suosituksia ja sisältöä. Huomattiin, että kyseisillä järjestelmillä on paljon tietoa käyttäjän mieltymyksistä sekä keinot vertailla käyttäjiä toisiinsa. Tässä luvussa tarkastellaan mitä kirjallisuudessa on kerrottu suosittelujärjestelmien vaikutuksesta mieltymysten muodostumiseen ja päätöksentekoon sekä voiko suosittelujärjestelmät vaikuttaa käyttäjän valintoihin.

Suosittelujärjestelmillä on suuri vaikutus siihen, minkälaista sisältöä käyttäjälle näytetään. Jos tällaisiin järjestelmiin vaikutetaan ja suosituksia muutetaan pahantahtoisesti voi se vaikuttaa siihen, kuinka monelle käyttäjälle suosituksia esitetään (Fang ym. 2018). Esimerkki tällaisesta on niin sanottu opetusdatan myrkyttäminen (engl. poisoning attack). Myös Adomavicius ym. (2013) osoittavat, että puolueellisella suosittelujärjestelmällä voi olla merkittävä vaikutus kuluttajien mieltymyksiin. Tämän lisäksi Cosley ym. (2003) esittävät hieman vanhemmassa julkaisussa seikkoja, joilla suosittelujärjestelmät voivat vaikuttaa käyttäjän mieltymyksiin. Esimerkiksi tilanteessa, jossa tuotetta arvioidaan ja samaan aikaan näytetään tuotteen jo annettu arvio, on mahdollista, että käyttäjän antama arvio on erilainen kuin jos tuotteen arvostelua ei näytettäisi.

Suosittelujärjestelmät usein yhdistetään myös informaatiokuplien (engl. filter bubble) muodostumiseen (Nguyen ym. 2014). Möller ym. (2018) selvittivät, kuinka paljon erilaisilla suosittelujärjestelmillä on vaikutusta uutisten monimuotoisuuteen. Hänen saamiensa tulosten mukaan henkilökohtaiset suositukset eivät vähentäneet uutisten monimuotoisuutta. Sen sijaan yhteistoiminnallisella menetelmällä saatiin monipuolisempia suosituksia erilaisista uutisaiheista. Sen sijaan Helbergertin, Karppisen ja D'acunton (2018) mukaan suosittelujärjestelmät vaikuttavat informaatiokuplien muodostumiseen. Heidän mukaansa suositukset muodostuvat käyttäjän aikaisempien valintojen sekä mieltymysten mukaan ja siitä syystä suosittelujärjestelmät vaikuttavat käyttäjien valintoihin enenevässä määrin. Jiang ym. (2019) ovat havainneet myös suositusten monimuotoisuuden kaventuvan suosittelujärjestelmien vuoksi mutta pitää ilmiötä nykytietämyksen mukaan kuitenkin rajallisena ja kokeellisena.

Suosittelujärjestelmien on havaittu vaikuttavan myös tuotteiden katselukertoihin sekä myyntiin. Lee ja Hosanagar (2021) osoittavat, että tällainen vaikutus on enemmän käyttötavaroissa kuin mielihyvää tuottavissa tuotteissa. Heidän mukaansa suositus toimiikin eräänlaisena käyttäjäarvioiden korvaajana ja täten lisää tuotteen kiinnostavuutta. Fleder ja Hosanagar (2009) ovat selvittäneet vaikuttavatko suosittelujärjestelmät myynnin keskittymiseen. Hänen mukaansa suositut ja hyvin näkyvät tuotesuositukset voivat korvata suositukset, jotka olisivat käyttäjälle paremmat ja näin vaikuttaa myynnin keskittymiseen ja tuotteen valitsemiseen.

Weidhaasin ym. (2021) mukaan internetissä ja sosiaalisessa mediassa tapahtuvalla kuulopuheella (engl. word of mouth) on suuri vaikutus ihmisten tuote tai palveluvalintoihin. Hänen mukaansa se voi käynnistää niin kutsuttua jäljittelykäyttäytymistä, joka voi olla positiivista kuin negatiivista. Hän on havainnut, että esimerkiksi elokuvan ennakkolevityksen Facebook-tykkäyksillä ja myyntituloilla on selvä yhteys. Milano, Taddeo ja Floridi (2020) ovat havainneetkin, että suosittelujärjestelmät muokkaavat käyttäjien mieltymyksiä ja ohjaa valintoja niin yksilöllisesti kuin sosiaalisesti. Weidhaas ym. (2021) mukaan suosittelujärjestelmillä käyttäjät saadaan kuluttamaan sisältö he eivät välttämättä muuten kuluttaisi.

5 Yhteenveto

Tässä työssä tarkasteltiin yleisimpiä suosittelujärjestelmiä ja katsottiin näiden järjestelmien käyttökohteita. Lopussa selvitettiin suosittelujärjestelmien vaikutusta käyttäjien mieltymykseen ja valintoihin.

Suosittelujärjestelmien kehitys on kiihtynyt viimeisten vuosikymmenten aikana, ja ne ovat tällä hetkellä suuren kiinnostuksen kohteena. Nämä järjestelmät ovatkin intuitiivinen työkalu selättämään suuri määrä informaatiota ja auttamaan käyttäjää päätöksenteossa. Suosittelujärjestelmät ovat levinneet moneen paikkaan, ja niitä löytää tänä päivänä niin verkkokaupoista, suoratoistopalveluista kuin uutispalveluistakin.

Erilaisia suosittelujärjestelmiä on paljon. Näistä suosituimpia ovat sisältöperustainen menetelmä, yhteistoiminnallinen suodattaminen sekä hybridimenetelmä. Viime aikoina tekoälyä ja koneoppimista on käytetty parantamaan suositusten muodostamista. Näistä erityisesti tietokonenäön ja puheentunnistuksen on havaittu hyödyttävän suosittelujärjestelmiä.

Kirjallisuuden perusteella sähköisessä kaupankäynnissä personoidut tuotesuosituksot vaikuttavat ostopäätökseen enemmän kuin personoimattomat suosituksot. Erityisesti silloin kun kyseessä on markkinat, joissa on suuri valinnanvara ja makuasioilla suuri merkitys. Tämän lisäksi suosittelujärjestelmien on havaittu vaikuttavan informaatiokuplien muodostamiseen, mutta nämä tulokset ovat vielä rajalliset.

Suosittelujärjestelmien on havaittu vaikuttavan suuresti siihen mitä sisältöä käyttäjä kuluttaa. Tämä avaa myös mahdolliset riskit suositusten pahantahtoiseen muuttamiseen. Nämä järjestelmät voivat lisätä myös jäljittelykäyttäytymistä, kaventaa käyttäjien kuluttaman sisällön monimuotoisuutta sekä saada käyttäjät kuluttamaan sisältöä, jota he eivät muuten kuluttaisi.

Lähteet

- Adomavicius, Gediminas, Jesse C Bockstedt, Shawn P Curley ja Jingjing Zhang. 2013. “Do recommender systems manipulate consumer preferences? A study of anchoring effects”. *Information Systems Research* 24 (4): 956–975.
- Afsar, M Mehdi, Trafford Crump ja Behrouz Far. 2021. “Reinforcement learning based recommender systems: A survey”. *ACM Computing Surveys (CSUR)*.
- Aggarwal, Charu C, ym. 2016. *Recommender systems*. Nide 1. Springer.
- Batmaz, Zeynep, Ali Yurekli, Alper Bilge ja Cihan Kaleli. 2019. “A review on deep learning for recommender systems: challenges and remedies”. *Artificial Intelligence Review* 52 (1): 1–37.
- Calero Valdez, André, Martina Ziefle ja Katrien Verbert. 2016. “HCI for recommender systems: the past, the present and the future”. Teoksessa *Proceedings of the 10th ACM Conference on Recommender Systems*, 123–126.
- Chen, Jiawei, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang ja Xiangnan He. 2020. “Bias and debias in recommender system: A survey and future directions”. *arXiv preprint arXiv:2010.03240*.
- Chen, Rui, Qingyi Hua, Yan-Shuo Chang, Bo Wang, Lei Zhang ja Xiangjie Kong. 2018. “A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks”. *IEEE Access* 6:64301–64320.
- Cosley, Dan, Shyong K Lam, Istvan Albert, Joseph A Konstan ja John Riedl. 2003. “Is seeing believing? How recommender system interfaces affect users’ opinions”. Teoksessa *Proceedings of the SIGCHI conference on Human factors in computing systems*, 585–592.
- Covington, Paul, Jay Adams ja Emre Sargin. 2016. “Deep neural networks for youtube recommendations”. Teoksessa *Proceedings of the 10th ACM conference on recommender systems*, 191–198.

- Fang, Minghong, Guolei Yang, Neil Zhenqiang Gong ja Jia Liu. 2018. "Poisoning attacks to graph-based recommender systems". Teoksessa *Proceedings of the 34th annual computer security applications conference*, 381–392.
- Fleder, Daniel, ja Kartik Hosanagar. 2009. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity". *Management science* 55 (5): 697–712.
- Helberger, Natali, Kari Karppinen ja Lucia D'acunto. 2018. "Exposure diversity as a design principle for recommender systems". *Information, Communication & Society* 21 (2): 191–207.
- Jiang, Ray, Silvia Chiappa, Tor Lattimore, András György ja Pushmeet Kohli. 2019. "Degenerate feedback loops in recommender systems". Teoksessa *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 383–390.
- Karimi, Mozghan, Dietmar Jannach ja Michael Jugovac. 2018. "News recommender systems—Survey and roads ahead". *Information Processing & Management* 54 (6): 1203–1227.
- Keyvan, Kimiya, ja Jimmy Xiangji Huang. 2022. "How to approach ambiguous queries in conversational search? A survey of techniques, approaches, tools and challenges". *ACM Computing Surveys (CSUR)*.
- Lee, Dokyun, ja Kartik Hosanagar. 2021. "How do product attributes and reviews moderate the impact of recommender systems through purchase stages?" *Management Science* 67 (1): 524–546.
- Lops, Pasquale, Dietmar Jannach, Cataldo Musto, Toine Bogers ja Marijn Koolen. 2019. "Trends in content-based recommendation". *User Modeling and User-Adapted Interaction* 29 (2): 239–249.
- Melville, Prem, ja Vikas Sindhwani. 2010. "Recommender systems." *Encyclopedia of machine learning* 1:829–838.
- Milano, Silvia, Mariarosaria Taddeo ja Luciano Floridi. 2020. "Recommender systems and their ethical challenges". *Ai & Society* 35 (4): 957–967.
- Mu, Ruihui. 2018. "A survey of recommender systems based on deep learning". *Ieee Access* 6:69009–69022.

- Möller, Judith, Damian Trilling, Natali Helberger ja Bram van Es. 2018. “Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity”. *Information, Communication & Society* 21 (7): 959–977.
- Nguyen, Tien T, Pik-Mai Hui, F Maxwell Harper, Loren Terveen ja Joseph A Konstan. 2014. “Exploring the filter bubble: the effect of using recommender systems on content diversity”. Teoksessa *Proceedings of the 23rd international conference on World wide web*, 677–686.
- Pazzani, Michael J, ja Daniel Billsus. 2007. “Content-based recommendation systems”. Teoksessa *The adaptive web*, 325–341. Springer.
- Ricci, Francesco, Lior Rokach ja Bracha Shapira. 2011. “Introduction to recommender systems handbook”. Teoksessa *Recommender systems handbook*, 1–35. Springer.
- Schedl, Markus, Peter Knees ja Fabien Gouyon. 2017. “New paths in music recommender systems research”. Teoksessa *Proceedings of the Eleventh ACM Conference on Recommender Systems*, 392–393.
- Sharma, Ravi S, Aijaz A Shaikh ja Eldon Li. 2021. “Designing Recommendation or Suggestion Systems: looking to the future”. *Electronic Markets* 31 (2): 243–252.
- Sperli, Giancarlo, Flora Amato, Fabio Mercorio, Mario Mezzanzanica, Vincenzo Moscato ja Antonio Picariello. 2018. “A social media recommender system”. *International Journal of Multimedia Data Engineering and Management (IJMDEM)* 9 (1): 36–50.
- Weidhaas, Raphael, Stephan Schlögl, Veikko Halttunen ja Teresa Spieß. 2021. “Watch This! The Influence of Recommender Systems and Social Factors on the Content Choices of Streaming Video on Demand Consumers”. Teoksessa *International Conference on Wirtschaftsinformatik*, 738–753. Springer.
- Zhang, Qian, Jie Lu ja Yaochu Jin. 2021. “Artificial intelligence in recommender systems”. *Complex & Intelligent Systems* 7 (1): 439–457.
- Zhang, Shuai, Lina Yao, Aixin Sun ja Yi Tay. 2019. “Deep learning based recommender system: A survey and new perspectives”. *ACM Computing Surveys (CSUR)* 52 (1): 1–38.