

Instagramin suosittelujärjestelmien toiminta

Tietotekniikka
Tietotekniikan tutkinto-ohjelma
Tietotekniikan laitos, Teknillinen tiedekunta
TkK-tutkielma

Laatija:
Emmi Ylikoski

Ohjaaja:
Erikoistutkija Petra Virjonen

Huhtikuu 2024

TkK-tutkielma
Tietotekniikan laitos, Teknillinen tiedekunta
Turun yliopisto

Oppiaine: Tietotekniikka

Tutkinto-ohjelma: Tietotekniikka

Tekijä: Emmi Ylikoski

Otsikko: Instagramin suosittelujärjestelmien toiminta

Sivumäärä: 30 sivua

Päivämäärä: Huhtikuu 2024

Suosittelujärjestelmät on alun perin kehitetty auttamaan ihmisiä tekemään päätöksiä. Nopeasti ne ovat kuitenkin levinneet erilaisille alustoille sekä nettisivuille. Sosiaalisen median suuren suosion kasvun myötä, viimeisen vuosikymmenen aikana, myös sosiaalisen median sovelluksissa on alettu hyödyntämään suosittelujärjestelmiä.

Tässä tutkielmassa perehdytään sosiaalisen median sovelluksen Instagramin suosittelujärjestelmien toimintaan. Tutkielman tarkoituksena on tutkia, miten Instagram suosittelee alustansa käyttäjilleen sisältöä ja pohtia miltä suosittelualgoritmeilta Instagramin suosittelujärjestelmät ominaisuuksiltaan vaikuttavat. Tutkielma on suoritettu kirjallisuuskatsauksena.

Tutkielmassa on havaittu, että Instagram on jakanut suosittelujärjestelmänsä erilleen jokaiseen sovelluksen eri osaan syöte, tarinat, hakusivu ja kelat, jossa jokaisessa käytetään omaa räätälöityä suosittelualgoritmia. Ominaisuuksiltaan Instagramin syöte- sekä kelaosion suosittelualgoritmit vaikuttavat tuotepohjaiselta yhteistoiminnalliselta suosittelualgoritmilta, tarinaosion suosittelualgoritmi vaikuttaa käyttäjähajaiselta yhteistoiminnalliselta suosittelualgoritmilta ja hakusivun suosittelualgoritmi vaikuttaa sisältöpohjaisen suosittelualgoritmin sekä käyttäjähajaisen yhteistoiminnallisen suosittelualgoritmin hybridisuosittelualgoritmilta.

Tehdyn tutkimuksen myötä on havaittu, että sosiaalisen median sovellusten suosittelujärjestelmät ovat jokseenkin vaikea tutkimuskohde, sillä ne ovat salattuja. Tämän takia onkin merkittävää, että aiheen tutkimusta jatketaan, jotta aihe saataisiin saataville suuremmalle yleisölle ja käyttäjien tietoisuutta saataisiin lisättyä. Täten sovellusten käyttökokemus voisi parantua entisestään ja käyttäjät tulisivat tietoiseksi myös sosiaalisen median sovelluksen käytön mahdollisista uhista sekä tietoturvariskeistä.

Asiasanat: suosittelujärjestelmät, sosiaalinen media, Instagram

Sisällysluettelo

| | | |
|----------|--|-----------|
| 1 | Johdanto | 1 |
| 1.1 | Tutkielman tarkoitus, aiheen valinta ja tutkimuskysymykset | 2 |
| 1.2 | Menetelmät ja tiedonhaku | 3 |
| 1.3 | Tutkielman rakenne | 3 |
| 2 | Suosittelujärjestelmät yleisesti | 4 |
| 2.1 | Yhteistoiminnallinen suosittelualgoritmi | 5 |
| 2.2 | Sisältöpohjainen suosittelualgoritmi | 7 |
| 2.3 | Yleiset heikkoudet ja vahvuudet | 8 |
| 2.4 | Hybridisuosittelualgoritmi | 10 |
| 3 | Instagramin suosittelujärjestelmät | 11 |
| 3.1 | Syöte ja tarinat | 12 |
| 3.2 | Hakusivu | 16 |
| 3.2.1 | Esisuodatusvaihe | 16 |
| 3.2.2 | Ranking-vaihe | 18 |
| 3.3 | Kelat | 25 |
| 4 | Pohdintaa | 27 |
| 5 | Yhteenveto | 29 |
| | Lähteet | 31 |

1 Johdanto

Ihmiset ovat jo pitkään tehneet päätöksiä muiden antamien suositusten perusteella. Usein läheisiltä tai ystäviltä kysytään mielipidettä tai suositusta jonkin päätöksen tekemiseen. Tämä yksinkertainen ajatus johti alun perin suosittelujärjestelmien syntyyn. Ihmisten päätöksentekoa haluttiin alkaa helpottamaan antamalla heille suosituksia. [1]

Suosittelujärjestelmät ovat ohjelmistotyökaluja, -tekniikoita ja -menetelmiä, jotka suosittelevat kohteita käyttäjille heidän mieltymyksien ja kiinnostuksen kohteiden pohjalta. Suosittelujärjestelmien tavoitteena on tehdä hyödyllisiä ja järkeviä suosituksia käyttäjille, jotta he löytävät tuotteita, jotka voisivat kiinnostaa heitä. Kun käytetyn sovelluksen tai nettisivun suosittelu sisältö on käyttäjän mielenkiinnonkohteiden kaltaista, voidaan käyttäjä saada ostamaan lisää tuotteita tai palaamaan palveluun aina uudestaan. [2]

Ensimmäinen suosittelujärjestelmä Tapestry kehitettiin vuonna 1992 sähköpostijärjestelmiin, jossa käyttäjät pystyivät kirjoittamaan sähköpostiin suodatussääntöjä esimerkiksi mielipiteiden tai toisten käyttäjien käytöksen pohjalta. Nopeasti Tapestryn kehittämisen jälkeen verkkokauppa Amazon.com otti suositteluteknologian laajamittaisesti käyttöön. Suosittelujärjestelmien käyttö verkkokaupassa oli suuri menestys ja Amazon.com:in jälkeen suosittelujärjestelmät levisivät nopeasti muillekin alustoille ja verkkokauppasivustoille kuten LinkedIn, Yahoo ja E-bay. [1], [3]

Sosiaalisen median käyttö on lisääntynyt huomattavasti viimeisen vuosikymmenen aikana sosiaalisten verkostojen kasvun myötä [4]. Suuren suosion kasvun myötä myös sosiaalisen median sovelluksissa on alettu hyödyntämään suosittelujärjestelmiä. Sosiaalista mediaa voidaan pitää viestintäkanavana sosiaalisen verkoston käyttäjien keskuudessa. Sosiaalisen verkoston käyttäjät pystyvät luomaan, siirtämään, vaihtamaan, jakamaan ja poistamaan sosiaalisen median sisältöä. Täten sosiaalisen median käyttäjä ei ole pelkkä kuluttaja vaan kuluttajan ja tuottajan yhdistelmä. [5] Käyttäjien tuottama sisältö voi olla esimerkiksi lyhyitä tekstipostauksia, live-videoita tai valokuvia [6].

Sosiaalisen median sovelluksen suosittelujärjestelmän kehittäminen voi olla hyvin erilaista kuin esimerkiksi verkkokaupassa käytettävän yleisen suosittelujärjestelmän kehittäminen, kun otetaan huomioon sosiaalisen median erityispiirteet ja sen monipuolisuus. Yleiset suosittelujärjestelmät ovat yleensä kehitetty auttamaan käyttäjää valitsemaan joukon hyödyllisiä tuotteita pitkästä tuoteluettelosta, esimerkiksi verkkokaupan tuotteista. Tällöin

suosittelujärjestelmän ja käyttäjän vuorovaikutus on yksinkertainen, sillä suosittelujärjestelmä antaa suosituksia käyttäjälle ja käyttäjä antaa palautetta kuten tuotearvioita takaisin suosittelujärjestelmälle. Sosiaalisen median sovelluksen suosittelujärjestelmä taas käsittelee paljon rikkaampia syötteitä sekä monimutkaisempia ihmisen ja tietokoneen välisiä vuorovaikutuksia kuten käyttäjien välisiä sosiaalisia suhteita. [5]

1.1 Tutkielman tarkoitus, aiheen valinta ja tutkimuskysymykset

Tässä tutkielmassa tutkitaan sosiaalisen median sovelluksen Instagramin suosittelujärjestelmien toimintaa sekä millä perusteella se tekee suosituksia käyttäjilleen. Aihe on rajattu koskemaan ainoastaan mitä julkaisuja Instagram suosittelee käyttäjilleen, joten tutkielma ohittaa esimerkiksi sen, mitä käyttäjiä Instagram suosittelee käyttäjänsä seuraavan tai mitä mainoksia Instagram suosittelee käyttäjilleen. Aihe on rajattu myös koskemaan vain yhtä sosiaalisen median sovellusta lähdeaineiston rajaamisen vuoksi sekä tutkielman tekijän henkilökohtaisen kiinnostuksen vuoksi. Tutkimuskysymykset ovat:

TK1: Miten Instagram suosittelee käyttäjilleen postauksia?

TK2: Mihin suosittelualgoritmeihin Instagramin suosittelujärjestelmät vaikuttavat perustuvan?

Aihetta on tärkeä tutkia, sillä useiden sosiaalisen median sovellusten suosittelujärjestelmät ovat salaisia eikä niitä ole annettu julkiseen tietoon. Tämän takia sovellusten keskiverto käyttäjä ei tiedä mitä suosittelujärjestelmät oikeasti tekevät tai miten hänestä kerättyjä tietoja voidaan mahdollisesti käyttää suositusten tekemiseen. Salauksen takia suosittelujärjestelmiä ei voida myöskään suoraan tarkastella ja analysoida vaan niistä suoritettut tutkimukset perustuvat tehtyihin kokeiluihin ja valistuneisiin arvauksiin. [7]

Tämän lisäksi tutkimusten [8], [9] mukaan suosittelujärjestelmät ovat alttiita puolueellisuudelle, sillä ne ovat niin riippuvaisia käyttäjiensä mieltymyksistä. Jos alustan käyttäjät ovat puolueellisia ja tätä dataa käytetään suosittelujärjestelmän optimointiin, voi suosittelujärjestelmä näin vahvistaa jo olemassa olevia ennakkoluuloja ja stereotypioita [9]. Myös Instagram on saanut kritiikkiä sen suosittelujärjestelmien puolueellisuudesta. Sovellusta on kritisoitu mm. siitä, että se suosii vaaleaihoisia käyttäjiä tummaihoisten sijaan. [10] Jos suosittelujärjestelmien toiminta olisi yleisemmin tiedossa, voisi tällaista puolueellisuutta olla helpompi kitkeä niistä pois. Tästäkin syystä aiheesta on syytä tehdä vielä tutkimuksia.

1.2 Menetelmät ja tiedonhaku

Tämän tutkielman tutkimusmenetelmänä on kirjallisuuskatsaus. Aineistoa tutkielmaan haettiin useasta eri tietokannasta tarkasti määritellyllä hakulauseella. Eri tietokannoissa hakua pystyi rajaamaan koskemaan vain tietyn kielisiä ja tietyn päivämäärän jälkeen julkaistuja teoksia. Aineistoa haettiin pääosin englannin kielellä, mutta myös muutama suomenkielinen lähdeaineisto löytyi. Enemmän kuin 11 vuotta vanhoja teoksia ei tässä tutkielmassa ole käytetty aineistona, jotta tieto olisi edelleen paikkansapitävää, sillä varsinkin tekninen tieto voi päivittyä nopeasti. Kielen ja julkaisupäivän rajauksen jälkeen aineistoa voitiin lähteä rajaamaan teoksen otsikon ja tiivistelmän perusteella. Tämän hakuprosessin tuloksena saatiin tässä tutkielmassa käytetyt lähdeaineistot.

Tiedonhakua suoritettiin pääosin tekniikan alan julkaisuja sisältävässä IEEE-tietokannassa. Tämän lisäksi hakua suoritettiin myös ACM Digital Library, ArXIV.org, Google Scholar ja Web of Science tietokannoissa. Näiden tietokantojen teosten lisäksi lähdeaineistoksi valittiin myös Instagramin omat nettisivut, Instagramin blogipostaus sekä Instagramin omistavan yrityksen Metan blogipostaus, sillä niiden todettiin sisältävän tutkielman kannalta merkittävää tietoa.

Tietokannoissa käytetty hakulause vaihteli hieman, riippuen siitä, minkälaista aineistoa haettiin. Haettaessa aineistoa Instagramin suosittelujärjestelmien toiminnasta hakulauseena oli alunperin ”recommend* system” AND Instagram, mutta se tuotti liian vähän hakutuloksia niin se muutettiin muotoon ”recommend* system” AND (”social media” OR Instagram). Näin hakua pystyttiin laajentamaan muihinkin sosiaalisen median sovelluksiin kuin pelkästään Instagramiin ja aineistojen tietoa pystytään soveltamaan Instagramin suosittelujärjestelmiin.

1.3 Tutkielman rakenne

Tutkielman luvussa 2 käsitellään jo tunnettuja menetelmiä, joilla voidaan luoda personoituja suosituksia käyttäjille käyttäessä jotakin nettisivua tai sovellusta. Luvussa 3 käsitellään sosiaalisen median sovelluksen Instagramin suosittelujärjestelmien toimintaa. Luvussa 4 pohditaan sosiaalisen median sovellusten suosittelujärjestelmien tulevaisuuden tutkimuksen suuntia. Luvussa 5 tiivistetään tutkielman oleelliset asiat ja vedetään ne yhteen.

2 Suositteijärjestelmät yleisesti

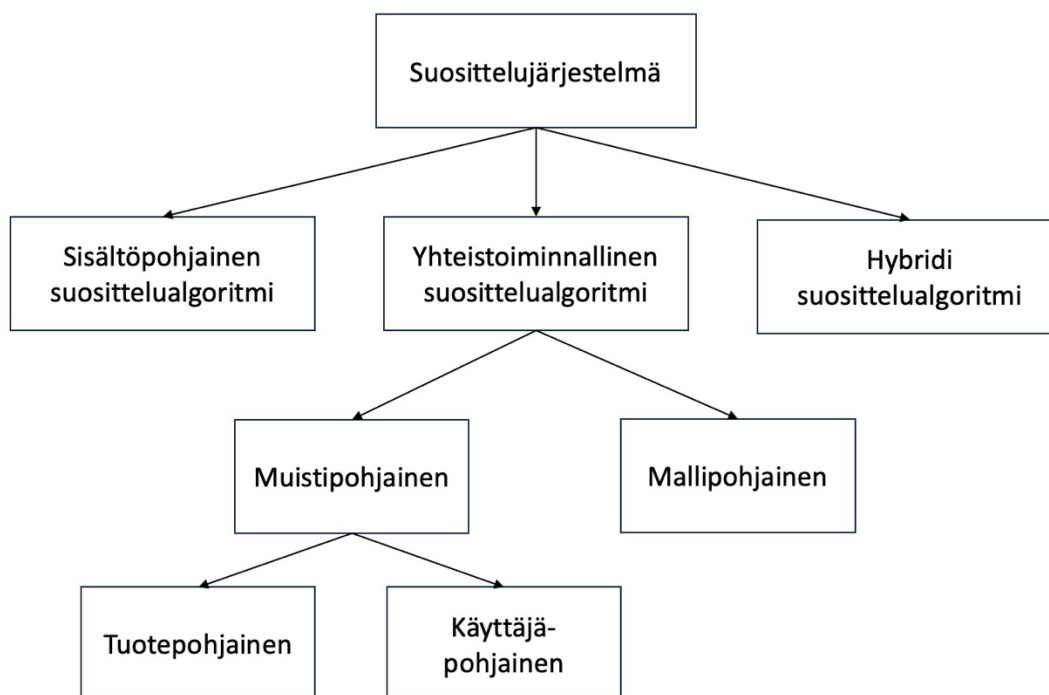
Nykyään Internetissä kohdataan valtava määrä tietoa ja digitaalista dataa joka päivä. Tämän uuden tiedon jatkuva lisääntyminen ja sen helppo saavutettavuus vaikeuttaa Internetin käyttäjien päätöksentekoa sekä varsinkin palveluntarjoajien kykyä tarjota käyttäjille vain sopivaa ja asiaankuuluvaa sisältöä. Ratkaistakseen tämän ongelman monet palveluntarjoajat ja verkkokauppasivustot käyttävät suositteijärjestelmiä (engl. recommender systems), jotta he voivat tarjota käyttäjilleen sisältöä heidän mieltymyksien ja henkilökohtaisten valintojen mukaan. [1]

Suosittelujärjestelmät perustuvat koneoppimisalgoritmeihin, jotka muun muassa oppivat käyttäjien kiinnostuksen kohteista, löytävät samankaltaisia käyttäjiä ja löytävät piilotettuja malleja käyttäjien ja kohteiden välillä [1]. Tällä tiedolla järjestelmä pystyy esittämään käyttäjälle personoitua sisältöä ja palveluita [11]. Koneoppiminen (engl. machine learning) on tekoälyn ja tietojenkäsittelytieteen ala, jossa tietokone oppii, miten suorittaa tehtäviä sille annetuista esimerkkitiedoista. Koneoppimismallit ovat tietokoneohjelmia, jotka koostuvat koneoppimisalgoritmeista. Nämä algoritmit käyvät läpi koulutusprosessin, jotta ne voivat oppia löytämään malleja ja tekemään päätöksiä aiemmin näkemättömästä tietojoukosta. Koulutusprosessissa oppiminen voi olla joko ohjattua tai ohjaamatonta. Ohjatussa oppimisessa mallia koulutetaan tietojoukolla, jonka rakenne ja tulos ovat jo ennestään tiedossa. Täten koneoppimismalli voi oppia tietojoukon sisäisiä malleja, säätämällä parametrejaan siten, että ennusteen virhe on mahdollisimman pieni. Ohjattua oppimista on esimerkiksi kuvantunnistus, jossa koneoppimismalli on opetettu datalla, joka sisältää kuvien oikeat luokat. Täten malli voi oppia tunnistamaan erilaisia objekteja kuvista. Ohjaamattomassa oppimisessa malli oppii tietojoukon rakenteen ja tuloksen itse, ilman, että se olisi ollut jo ennestään tiedossa, kuten esimerkiksi klusterointi. [12] Klusteroinnissa algoritmi voi ryhmitellä tietojoukon luokkiin näytteiden samankaltaisuuden perusteella siten, että samaan ryhmään kuuluvat ovat keskenään enemmän samankaltaisia kuin mihinkään muuhun ryhmään kuuluvien näytteiden kanssa [13]. Koulutuksen jälkeen koneoppimismallia testataan, jotta nähdään, kuinka hyvin se suoriutuu tehtävästään ennen näkemättömillä tietojoukoilla. Mallin suorituskykyä mitataan useilla parametreilla, kuten sen tarkkuudella ja muistilla. [12]

Suosittelujärjestelmistä on nykyään tullut hyvin suosittuja ja niitä käytetäänkin monilla eri sivustoilla ja sovelluksissa. Kaikista yleisimmät alueet, joissa suositteijärjestelmiä käytetään

ovat esimerkiksi mitä elokuvia katsoo, mitä musiikkia kuuntelee, mitä tuotteita ostaa ja ketä ihmisiä tapailee, sillä suosittelujärjestelmiä käytetään myös nettitaipailusivustoilla. [2]

Suosittelujärjestelmiä voidaan luokitella eri suosittelumalleihin niiden algoritmien rakenteen ja tekniikan pohjalta. Kolme yleisesti käytettyä suosittelualgoritmiä ovat sisältöpohjainen suosittelualgoritmi, yhteistoiminnallinen suosittelualgoritmi sekä hybridipohjainen suosittelualgoritmi. Yhteistoiminnallinen suosittelualgoritmi voidaan jakaa vielä muistipohjaiseen sekä mallipohjaiseen suositteluun. Muistipohjainenkin malli voidaan jakaa vielä kahteen pienempään luokitteluun, tuotepohjaiseen ja käyttäjäpohjaiseen. [11] Kuvassa 2-1 on havainnointu suosittelujärjestelmien luokittelua. Näihin suosittelualgoritmeihin tutustaan seuraavissa alaluvuissa lisää.



Kuva 2-1 Suosittelujärjestelmien luokittelu lähteen [11] perusteella.

2.1 Yhteistoiminnallinen suosittelualgoritmi

Yhteistoiminnallinen suosittelualgoritmi tekee suosituksia käyttäjälle hänen oman käytöksensä sekä muiden samankaltaisten käyttäjien käytöksen perusteella. Yleinen periaate on se, että jos kaksi tai enemmän käyttäjää pitävät samoista asioista jollakin aihealueella on

todennäköistä, että he pitävät samanlaisista asioista myös jossain muussa kategoriassa.

Algoritmi kerää käyttäjien mielipiteitä ja laskee samankaltaisuuksia käyttäjien kiinnostuksen kohteiden välillä tehdäkseen suosituksia. Käyttäjä voi saada suosituksia sellaisista kohteista, joita hän ei ole itse vielä koskaan arvostellut, mutta samankaltaiset käyttäjät ovat arvostelleet kohteen hyvin. Hyvä esimerkki yhteistoiminnallisesta suosittelualgoritmista on esimerkiksi se, kun sosiaalisen median sovelluksissa ehdotetaan käyttäjälle ihmisiä, joita hän saattaa tuntea. Tällöin suosituksia on tehty yhteisten kavereiden perusteella. On mahdollista, jos kahdella käyttäjällä on yhteisiä ystäviä, että he myös tuntevat toisensa. [1], [11], [14]

Yhteistoiminnallinen suosittelualgoritmi kerää useamman kuin yhden käyttäjän mielipiteitä ja useamman kuin yhden kohteen yksityiskohtia tehdäkseen suosituksia. Tämän takia algoritmi kerää ja analysoi erittäin suuria datamääriä, jotka ovat saatavilla sovelluksesta tai nettisivusta, jota käyttäjä käyttää. Yhteistoiminnallinen suosittelualgoritmi voidaan jakaa kahteen ryhmään, muistipohjaiseen ja mallipohjaiseen suositteluun, sen perusteella, miten ne prosessoivat saamaansa dataa suosituksien tekemiseksi. [15]

Muistipohjaisessa suosittelussa algoritmi tekee suosituksia käyttäjälle, käyttämällä koko keräämäänsä tietokantaa käyttäjäprofiileista. Tietokantaan kuuluu käytetyn sovelluksen tai nettisivun käyttäjien erilaista dataa kuten, tykkäyksiä, klikkauksia ja arvosteluja. Algoritmi sitten pyrkii etsimään samankaltaisia käyttäjiä aktiivisen käyttäjän kanssa ja muodostamaan niistä samankaltaisten käyttäjien ryhmän. Tämän jälkeen algoritmi laskee aktiiviselle käyttäjälle suosituksia muodostetun ryhmän käyttäjien antamien arvioiden perusteella. [11]

Muistipohjainen suosittelu voidaan luokitella vielä tuotepohjaiseen ja käyttäjäpohjaiseen suositteluun. Käyttäjäpohjaisessa suosittelussa algoritmi etsii samankaltaisuuksia eri käyttäjien välillä vertailemalla heidän arvostelujaan samalle kohteelle. Tämän jälkeen algoritmi tekee ennustetun arvostelun kyseiselle kohteelle aktiiviselle käyttäjälle, mikä on keskimääräinen arvostelu kuin samankaltaisilla käyttäjillä. Suositus voidaan sitten antaa aktiiviselle käyttäjälle sen mukaan, kuinka samankaltainen ennustettu arvostelu on muihin samankaltaisiin käyttäjiin verrattuna. [11], [15]

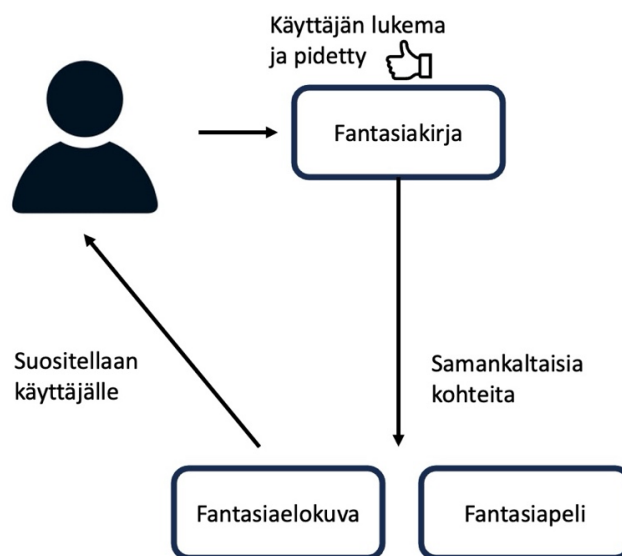
Tuotepohjaisessa suosittelussa algoritmi etsii samankaltaisuuksia kohteiden välillä, ei käyttäjien. Algoritmi luo joukon aktiivisen käyttäjän jo arvostelluista kohteista, jotka olisivat mahdollisimman samanlaisia kuin tuleva suositeltava kohde. Tämän arvosteltujen kohteiden joukon avulla algoritmi voi ennustaa, tulisiko käyttäjä pitämään suositeltavasta kohteesta. [15]

Mallipohjaisessa suosittelussa suositukset tehdään mallilla, joka on tehty käyttäjän arvioiden perusteella. Malli muodostetaan koneoppimis- tai tiedonlouhintatekniikoilla, ja se on suunniteltu poimimalla hyödyllisiä tietoja suurista datamääristä, jota yhteistoiminnallinen suosittelualgoritmi käy läpi. Poimitut tiedot liittyvät suositeltavien kohteiden tai samankaltaisten käyttäjien tiettyihin attribuutteihin. Käyttämällä tätä mallia, voidaan suosituksia tehdä ilman, että jokaista suositusta kohden täytyy aina analysoida suuria datamääriä, jota yhteistoiminnallinen suosittelualgoritmi käy läpi. [15]

2.2 Sisältöpohjainen suosittelualgoritmi

Sisältöpohjainen suosittelualgoritmi tekee suositukset ainoastaan yhden käyttäjän aikaisemman käytöksen ja datan perusteella. Käyttäjän aikaisemman käytöksen lisäksi tärkeässä roolissa on myös kohteen kuvaus, sillä suositus tehdään kohteen sisällön ja käyttäjäprofiilitietojen vertailun perusteella. Algoritmi suosittelee käyttäjälle tuotteita, joita hän ei ole vielä nähnyt tai arvostellut, mutta jotka ovat samankaltaisia kuin tuotteet, joista hän on jo pitänyt aikaisemmin. [15] Suositeltaville kohteille voidaan laskea samankaltaisuus niiden yhteisten ominaisuuksien perusteella [1]. Mitä enemmän käyttäjä on vuorovaikuttanut suositusten kanssa, kuten antanut arvosteluja tai tykkäyksiä, sitä tehokkaammaksi algoritmi tulee [15].

Sisältöpohjainen suosittelu toimii siten, että ensin jokaisen suositeltavan kohteen sisältö esitetään termeinä, yksityiskohtina, attribuutteina tai piirrevektorina. Kohteen sisällön attribuutteina voi esimerkiksi olla elokuvan tai kirjan tyyli tai jokin termi, joka toistuu sisällössä useaan kertaan. Tämän jälkeen algoritmi saa syötteenä joukon kyseisiä attribuutteja ja analysoi niitä sekä jotain kohdetta, jota käyttäjä on jo arvostellut. Analyysistä algoritmi pystyy muodostamaan käyttäjän kiinnostuksista mallin, joka luo suosituksia. Parhaimmat suositukset saadaan, kun suositeltavien kohteiden attribuutteja vertaillaan useampaan kuin yhteen jo käyttäjän arvosteltuun kohteeseen. [16] Kuvassa 2–2 on demonstroitu sisältöpohjaisen suosittelualgoritmin toimintaa.



Kuva 2-2 Sisältöpohjaisen suosittelualgoritmin toiminta, jossa algoritmi suosittelee käyttäjälle samaan tyyliin kuuluvia kohteita lähteen [17] perusteella.

2.3 Yleiset heikkoudet ja vahvuudet

Yllä mainituilla suosittelualgoritmeilla kaikilla on omat vahvuutensa mutta myös jotain heikkouksia ja ongelmia, joita pitäisi pystyä parhaan mukaan ratkomaan. Yleisiä ongelmia mitä algoritmit kohtaavat ovat kylmäkäynnistysongelma, skaalautuvuusongelma, harvuusongelma, yksityisyys ja eri nimisten tuotteiden samankaltaisuus. [15]

Kylmäkäynnistysongelma tarkoittaa sitä, kun sivustoa tai sovellusta käyttää uusi käyttäjä eikä hän ole vielä tykännyt tai arvostellut mitään kohteita. Uuden käyttäjän mieltymykset ovat näin suosittelualgoritmille tuntemattomat eikä se pysty tekemään hänelle osuvia suosituksia.

Skaalautuvuusongelmassa ongelmana on se, että suosittelujärjestelmän resurssit eivät riitä tiedon prosessointiin ja suositusten muodostamiseen. Käyttäjien ja kohteiden lisääntyessä sivustoilla koko ajan, sovelluksissa käytetyt suosittelujärjestelmät tarvitsevat yhä enemmän ja enemmän resursseja. [15]

Suosittelualgoritmit toimivat paremmin mitä enemmän käyttäjät ovat vuorovaikutuksessa käyttämänsä alustan kanssa, esimerkiksi tykkäävät ja antavat arvosteluja. On kuitenkin aina olemassa myös käyttäjiä, jotka ovat olleet hyvin vähän vuorovaikutuksessa käyttämänsä alustan kanssa, minkä takia hänen mieltymyksiänsä on vaikea päätellä. Kyseiselle käyttäjälle on myös vaikea löytää hänen kanssaan samankaltaisia käyttäjiä ja hänet saatetaan myös

yhdistää väärään käyttäjään, joka ei oikeasti olisi hänen kanssaan samanlainen, sillä hänestä on niin vähän tietoa suosittelualgoritmeilla. Tämä on nimeltään harvuusongelma ja se ilmenee, jos käyttäjästä on saatavilla liian vähän dataa suositusten tekemiseen. [15]

Yksityisyys on selkeästi myös suosittelujärjestelmien ongelma. Suosittelualgoritmien täytyy kerätä käyttäjistä mahdollisimman paljon tietoa, jotta se voisi tehdä mahdollisimmat hyviä ja tarkkoja suosituksia. Myös arkaluontoista tietoa kerätään, kuten tietoa käyttäjän sijainnista, iästä ja sukupuolesta. Kerätyistä tiedoista voi herätä kysymys kuinka luotettavia, eheitä tai luottamuksellisia ne oikeastaan ovat. Suosittelualgoritmit voivat törmätä myös ongelmaan, jossa samankaltaisilla tuotteilla on eri nimi (engl. synonymy). Tällöin suosittelualgoritmien on vaikea erottaa näiden kahden tuotteen välistä eroa. [15]

Erityisesti yhteistoiminnallisen suosittelualgoritmin heikkouksia ovat kylmäkäynnistys-, skaalautuvuus- ja harvuusongelma sekä samankaltaisten kohteiden eri nimisyys. Vahvuuksia taas ovat helppo toteutus ja uuden tiedon lisääminen asteittain, kun käytetään muistipohjaista suosittelua. Käyttämällä mallipohjaista suosittelua ei tarvita niin paljoa muistia tai CPU (Central Processing Unit) -aikaa. [11], [16]

Sisältöpohjaisen suosittelualgoritmin heikkouksia ovat esimerkiksi vaikeus esittää suositeltavien kohteiden sisältö joinain termeinä, yksityiskohtina, attribuutteina tai piirrevektorina sekä liiallinen erikoistumisongelma, sillä algoritmi suosittelee paljon samantyyppisiä kohteita. Vahvuutena sisältöpohjaisella suosittelualgoritmeilla on taas käyttäjän riippumattomuus, sillä toisin kuin yhteistoiminnallisessa suosittelualgoritmissa suositukset tehdään ainoastaan aktiivisen käyttäjän käytöksen pohjalta eikä se riipu muista käyttäjistä. Tämän lisäksi sisältöpohjainen suosittelualgoritmi on toiminnaltaan läpinäkyvämpi ja aktiiviselle käyttäjälle kerrotaan selitys, kuinka suosittelualgoritmi toimii. Sisältöpohjainen suosittelualgoritmi on myös parempi suosittelemaan kohteita, joita kukaan käyttäjä ei ole vielä arvioinut. [16]

Algoritmien kohtaamia ongelmia pystytään ratkaisemaan eri keinoin.

Kylmäkäynnistysongelman pystyy esimerkiksi ratkaisemaan sillä, että ennen alustan käyttöönottoa uudelle käyttäjälle esitetään kysely ja sillä tiedustellaan hänen mielenkiinnonkohteitaan. [15] Kerätyn tiedon yksityisyyttä voidaan parantaa esimerkiksi määrittämällä yksityisyyttä koskevat eettiset ohjeet ja määräykset, joita suosituksia tekevien alustojen tulisi noudattaa. Tämän lisäksi yksityisyyttä voitaisiin parantaa myös hajauttamalla kerätyn tiedon tallentamista, sillä tiedon tallentaminen ja sen käsittely keskitetysti on alttiimpi

hyökkäyksille ja uhille. Alustoilla voidaan myös ottaa käyttöön erilaisia ohjelmia ja algoritmeja, jotka parantavat yksityisyyttä, kuten obfuskointi (engl. obfuscation), jolloin tietoihin lisätään tarkoituksenmukaisesti melua, jotta siitä ei voi tunnistaa, kuka käyttäjä on kyseessä. [18] Skaalautuvuusongelmaa voidaan ratkaista parantamalla järjestelmiä fyysisesti. Yksi tehokas tapa, jolla pystyy ratkaisemaan useamman yllä mainituista ongelmista samanaikaisesti, on käyttää hybridisuosittealgoritmia. [15] Tätä algoritmia käsitellään seuraavassa alaluvussa lisää.

2.4 Hybridisuosittealgoritmi

Hybridisuosittealgoritmeilla on mahdollista saada parempia tuloksia suosituksiin, koska se yhdistää monia eri suosittelualgoritmeja keskenään. Tällä tavalla hybridisuosittealgoritmi pystyy välttämään joidenkin yksittäisten suosittelualgoritmien rajoitukset, tekniset ongelmat ja optimoimaan järjestelmää paremmin. Idea on yhdistää kaksi suosittelualgoritmia, jotka tukevat toistensa heikkouksia, jolloin algoritmi voi antaa parempia ja tehokkaampia suosituksia kuin kumpikaan suosittelualgoritmi yksinään. [11]

Suosituksen tekemiseen eri alustoilla voidaan käyttää monenlaisia tekniikoita kuten yhteistoiminnallista tai sisältöpohjaista suosittelua. Hybridisuosittealgoritmi yhdistelee kahden tai useamman tekniikan parhaimpia piirteitä keskenään. Tekniikoita voidaan yhdistellä toisiinsa monilla eri tavoin. Eräs tapa on ensin erotella eri algoritmien toteutus ja vasta lopuksi yhdistää niiden tulokset. Kun eri tekniikoita yhdistellään keskenään, pystyy hybridialgoritmi välttämään muiden tekniikoiden ongelmia kuten kylmäkäynnistys-, harvuus- ja skaalautuvuusongelman. [11], [15]

Yleinen alusta, jolla käytetään hybridisuosittealgoritmiä, on suoratoistopalvelu Netflix. Sen suosittelujärjestelmä yhdistää yhteistoiminnallista- sekä sisältöpohjaista suosittelualgoritmia, sillä se antaa suosituksia käyttäjilleen heidän katseluhistoriansa sekä kiinnostuksenkohteidensa perusteella, mutta myös hänen ja muiden Netflix-käyttäjien samankaltaisuuden perusteella. [1]

Vaikka hybridisuosittealgoritmi pystyykin yhdistämällä eri suosittelualgoritmien tekniikoita ratkaisemaan monia ongelmia, joita yksittäin nämä algoritmit kohtaisivat, ei tarkoita, että sillä ei olisi ollenkaan heikkouksia. Hybridisuosittealgoritmin heikkouksia ovat muun muassa lisääntyneet toteutuskustannukset ja lisääntynyt monimutkaisuus. [15]

3 Instagramin suosittelujärjestelmät

Instagram on yksi maailman suosituimmista sosiaalisen median sovelluksista yli 2 miljardilla aktiivisella käyttäjällään [19]. Se määrittää sovelluksen omilla nettisivuilla maksuttomaksi kuvien ja videoiden jakamissovellukseksi. Sovelluksen käyttöehdoissa on myös määritetty, että se ei myy käyttäjiensä tietoja eteenpäin, mikä voisi olla tapa rahoittaa sovelluksen toimintaa. Instagramin käyttäjät voivat jakaa kuvia ja videoita seuraajiensa tai valitsemiensa kaveriryhmien kanssa ilmaiseksi. Sovelluksessa myös pystyy katsomaan seurattujen käyttäjien julkaisuja sekä tykkäämään, että kommentoimaan niitä. Sen lisäksi että sovelluksessa näkee seurattujen käyttäjien julkaisemia postauksia, Instagram myös suosittelee käyttäjilleen uutta sisältöä kuten kuvia ja videoita. Näiden suositusten avulla aktiivinen käyttäjä pääsee tutustumaan uuteen sisältöön käyttäjiltä, joita hän ei vielä itse seuraa. [20]

Instagram on alusta, joka mahdollistaa monenlaista toimintaa. Monille ihmisille se on paikka löytää iloa, esimerkiksi pitämällä yhteyttä ystäviinsä. Instagramissa on kuitenkin mahdollista myös mainostaa omaa yritystään tai tuotettaan, ja monet valokuvaajat ja mallit käyttävät sovellusta omana gallerianaan. Tämän lisäksi Instagramia voidaan käyttää poliittisen ja yhteiskunnallisen vaikuttamisen välineenä. [19] Nykyään Instagramissa on myös käyttäjiä, jotka saavat rahaa mainostamalla muiden tuotteita omalla alustallaan.

Koska Instagramissa on niin suuri määrä käyttäjiä ja vielä suurempi määrä käyttäjiin liittyviä postauksia, suosittelujärjestelmät ovat tarpeellisia. Suosittelujärjestelmien avulla käyttäjille pystytään suosittelemaan sopivaa uutta sisältöä ja myös käyttäjiä, joiden sisältö voisi olla kiinnostavaa. Erityisesti kylmäkäynnistysongelman kannalta tämä on tärkeää, kun uudet käyttäjät eivät vielä seuraa ketään tai eivät ole olleet vuorovaikutuksessa minkään sisällön kanssa liittyessään Instagramiin. [21]

Yleisesti sosiaalisen median sovelluksen suosittelujärjestelmällä on joukko käyttäjiä $U = \{u_1, \dots, u_n\}$ ja joukko suositeltavia kohteita $O = \{o_1, \dots, o_m\}$. Jokaiselle käyttäjä-kohde-parille $\{u_i, o_j\}$ suosittelualgoritmi pystyy laskemaan arvon tai sijoituksen (engl. rank), joka mittaa käyttäjän u_i kiinnostusta kohdetta o_j kohtaan tai kohteen o_j hyödyllisyyttä käyttäjää u_i kohtaan. Tämän lisäksi suosittelualgoritmit koostuvat yleensä yhdestä tai useammasta seuraavasta komponentista. Esisuodatusvaiheessa valitaan jokaiselle käyttäjälle u_i osajoukko $O_i^c \subset O$, joka sisältää kohteita, jotka ovat hyviä ehdokkaita suositeltaviksi kohteiksi. Ranking-vaiheessa jollain luvussa 2 esitellyistä suosittelualgoritmeilla lasketaan käyttäjän u_i jokaiselle

osajoukon O_i^c ehdotetulle kohteelle o_j sijoitus. Jälkisuodatusvaiheessa muodostetaan jokaiselle käyttäjälle u_i uusi osajoukko $O_i^f \subseteq O_i^c$, sulkemalla kohteita pois alkuperäisestä osajoukosta käyttäjän ja sovelluksen välisen vuorovaikutuksesta saadun datan perusteella, esimerkiksi tykkäysten ja klikkausten. [22]

Instagram on avannut kuinka heidän suosittelujärjestelmänsä antavat sijoituksia mahdollisille suositeltaville kohteille heidän omassa blogikirjoituksessaan. Sovelluksen jokaisessa eri osassa, syöte, tarinat, hakusivu ja kelat, käytetään omaa suosittelualgoritmia, joka on räätälöity sen mukaan, kuinka ihmiset näitä sovelluksen osia käyttävät. Jotta suosittelualgoritmi voisi antaa suositeltaville kohteille sijoituksia, se kerää tietoa aktiivisesta käyttäjästä ja muista käyttäjistä ja heidän välisestään vuorovaikutuksesta. Kerätty tieto vaikuttaa siihen minkälaista sisältöä käyttäjille suositellaan. Näitä tietoja kutsutaan signaaleiksi. [23]

Seuraavissa alaluvuissa käsitellään, kuinka näissä sovelluksen osissa annetaan sijoitukset suositeltaville kohteille eli millä perusteella Instagram sisältöä käyttäjilleen suosittelee. Kuten tutkielmassa on jo todettu, suosittelujärjestelmät ovat usein salaisia, minkä takia Instagram ei kuvaile kovin tarkasti blogikirjoituksessaan, kuinka heidän suosittelujärjestelmänsä toimii. Suosittelujärjestelmiä voidaan kuitenkin pohtia luvussa 2 esiteltyjen algoritmien avulla ja tarkastella minkä algoritmin ominaisuuksia Instagramin suosittelualgoritmeissa on.

3.1 Syöte ja tarinat

Instagram-syöte (engl. Instagram Feed) on jokaiselle käyttäjälle henkilökohtainen sivu, jossa näkyy sisältöä seuratuilta käyttäjiltä, Instagramin suosittellemaa sisältöä ja mainoksia [23]. Syötteessä jokainen käyttäjä pystyy myös itse julkaisemaan sisältöä, johon hänen seuraajansa voivat reagoida. Syöteen yläosassa näytetään julkaistut postaukset, joiden Instagram uskoo kiinnostavan käyttäjää kaikista eniten. [20] Instagramin suosittelualgoritmi päättää, mitä postauksia suositella kaikista eniten kerättyjen signaalien avulla [23].

Mosseri [23] kertoo blogitekstissä, että sijoitusten antaminen suositeltaville kohteille alkaa siitä, että ensin määritellään joukko asioita, joita ylipäättään halutaan suositella. Syötteessä tähän joukkoon kuuluvat aktiivisen käyttäjän seuraamien käyttäjien tekemät postaukset sekä postaukset käyttäjiltä, joita aktiivinen käyttäjä ei vielä seuraa, mutta joista hän voisi olla kiinnostunut. Suosittelualgoritmi määrittelee kohteet, joista aktiivinen käyttäjä voisi olla kiinnostunut sen perusteella mitä käyttäjiä hän seuraa, mistä kohteista hän on tykännyt tai

ollut vuorovaikutuksessa viime aikoina. Tämä toiminta on selkeästi tunnistettavissa esisuodatusvaiheeksi ja määritelty joukko sisältää kohteita, jotka ovat ehdokkaita suositeltavaksi kohteeksi.

Tämän jälkeen joukon kohteille täytyy antaa sijoitukset, jotta tiedetään mitä suositella ensimmäisenä. Suositelualgoritmi antaa kohteille sijoitukset kerättyjen signaalien perusteella. Tärkeimmät signaalit suositellessa sisältöä syötteessä ovat aktiivisen käyttäjän aktiivisuus, julkaistujen postauksien tiedot, postauksen julkaisseen käyttäjän tiedot sekä aktiivisen käyttäjän vuorovaikutuksen historia. Suositellakseen mahdollisimman hyviä kohteita suositelualgoritmi kerää tietoa, mistä postauksista aktiivinen käyttäjä tykkää tai mitä postauksia hän kommentoi tai tallentaa. Tämän lisäksi se kerää tietoa julkaistuista postauksista kuten, kuinka moni käyttäjä on tykännyt postauksesta, kuinka nopeasti postauksesta tykätään sekä kuinka nopeasti sitä kommentoidaan ja tallennetaan. Näiden lisäksi katsotaan, milloin postaus on julkaistu ja mihin sijaintiin se on yhdistetty, jos on. Suositelualgoritmi kerää myös postauksen julkaisseesta käyttäjästä tietoa, jotta se voi päätellä kuinka mielenkiintoinen kyseinen käyttäjä voi olla toiselle ihmiselle. Tällaisia signaaleja ovat esimerkiksi, kuinka usein muut Instagramin käyttäjät ovat olleet vuorovaikutuksessa kyseisen käyttäjän kanssa viime viikkoina. Tämän lisäksi suositelualgoritmi kerää tietoa aktiivisen käyttäjän vuorovaikutuksen historiasta jonkun muun käyttäjän kanssa, esimerkiksi kommentoivatko käyttäjät toisensa postauksiin vai ei. [23]

Näiden signaalien pohjalta suositelualgoritmi antaa sijoitukset joukossa määritellyille kohteille. Annetut sijoitukset ovat ennusteita siitä, kuinka todennäköisesti aktiivinen käyttäjä on vuorovaikutuksessa postauksen kanssa eri tavoin. Mitä todennäköisemmin käyttäjä jää postauksen kohdalle muutamaksi sekunniksi, kommentoi, tykkää tai jakaa sitä tai napauttaa postauksen tehneen käyttäjän profiilikuvaa sen korkeamman sijoituksen kohde saa ja sen korkeammalla postaus näkyy aktiivisen käyttäjän syötteessä. Instagram-syötteen suositelualgoritmissa on myös havaittavissa jälkisuodatusvaihe. Suositelualgoritmi nimittäin poistaa kerättyjä signaaleja ja tehtyjä ennusteita sen perusteella miten aktiivisen käyttäjän mielenkiinnon kohteet muuttuvat ajan myötä. [23]

Näiden ominaisuuksien perusteella syöteosion suositelualgoritmi vaikuttaa olevan tuotepohjainen yhteistoiminnallinen suositelualgoritmi. Suositelualgoritmi ei voi olla sisältöpohjainen, sillä suositusten tekemiseen vaikuttaa useamman kuin ainoastaan yhden käyttäjän data ja aikaisempi käytös ja sisältöpohjainen suositelualgoritmi tekee suositukset

ainoastaan yhden käyttäjän aikaisemman käytöksen ja datan perusteella [15]. Lisäksi yhteistoiminnalliseen suositelualgoritmiin viittaa se, että suositusten tekemiseen vaikuttaa aktiivisen käyttäjän vuorovaikutuksen historia sekä postauksen julkaisseen käyttäjän tiedot, mikä tarkoittaa sitä, että suositusten tekemiseen vaikuttaa aktiivisen käyttäjän oman käytöksensä lisäksi muiden käyttäjien käytös kuten yhteistoiminnallisessa suositelualgoritmissa [1]. Suositelualgoritmi vaikuttaa olevan tuotepohjainen, sillä se kerää tietoa siitä, mistä kohteista aktiivinen käyttäjä on aikaisemmin pitänyt sekä suositeltavasta kohteesta, mikä viittaisi siihen, että se pyrkii etsimään samankaltaisia kohteita, joista käyttäjä on jo aikaisemmin pitänyt, mikä on tuotepohjaisen yhteistoiminnallisen suositelualgoritmin toimintaperiaate [15].

Instagram-tarinaosiossa (engl. Stories) Instagramin käyttäjät pystyvät jakamaan jokapäiväisiä hetkiään seuraajiensa kanssa. Tarinoissa käyttäjät näkevät seuraamiensa käyttäjien tarinat sekä mainoksia. [23] Tarinoihin voi jakaa kuvia ja videoita, jotka katoavat käyttäjän profiilista 24 tunnin kuluttua julkaisemisesta [20]. Samalla tavalla kuin Instagram syötteessä suositeluprosessi alkaa esisuodatusvaiheella, eli määrittellään joukko kohteita, jotka ovat mahdollisia suositeltavia kohteita. Tarinoissa tähän joukkoon kuuluu ainoastaan aktiivisen käyttäjän seuraamien käyttäjien tarinat. [23]

Joukon määrittelyn jälkeen, joukon kohteille täytyy antaa sijoitukset signaalien perusteella. Signaalit, jotka vaikuttavat kohteiden sijoitukseen tarinaosiossa ovat aktiivisen käyttäjän katseluhistoria, vuorovaikutushistoria tarinoissa ja käyttäjien välinen läheisyys.

Suosittelualgoritmi kerää tietoa siitä kuinka usein aktiivinen käyttäjä katsoo jonkun käyttäjän tarinoita, jotta voidaan priorisoida niiden käyttäjien tarinat, joita aktiivinen käyttäjä katsoo kaikista eniten. Tämän lisäksi algoritmi kerää tietoa siitä kuinka usein aktiivinen käyttäjä vuorovaikuttaa julkaistun tarinan kanssa, kuten tykkää siitä tai lähettää tarinan julkaisseelle käyttäjälle yksityisviestin. Näiden asioiden lisäksi suositelualgoritmi kerää tietoa siitä, kuinka läheinen aktiivinen käyttäjä on tarinan julkaisseen käyttäjän kanssa ja kuinka todennäköisesti he ovat ystäviä tai perheenjäseniä. [23]

Näiden signaalien perusteella aiemmin määritellyn joukon kohteille annetaan sijoitukset. Sijoitus on suositelualgoritmin tekemä ennustus siitä, kuinka osuva ja tarkka suositeltava kohde on aktiiviselle käyttäjälle. Ennustuksen osuvuus voidaan määrittellä sillä, kuinka todennäköisesti aktiivinen käyttäjä avaa jonkun tarinan, vastaa tarinaan tai siirtyy seuraavaan

tarinaan. Sijoituksen perusteella kohteet suositellaan käyttäjälle. Mitä parempi sijoitus, sitä korkeammalla se näkyy tarinoissa. [23]

Tarinaosion suosittelualgoritmi vaikuttaa ominaisuuksiltaan olevan käyttäjöpohjainen yhteistoiminnallinen suosittelualgoritmi. Päätelmä perustuu siihen, että suosittelualgoritmi etsii samankaltaisuuksia kahden käyttäjän välillä, esimerkiksi kuinka läheisiä käyttäjät ovat. Tämä viittaa käyttäjöpohjaiseen yhteistoiminnalliseen suosittelualgoritmiin, sillä se etsii samankaltaisuuksia eri käyttäjien välillä [11]. Sisältöpohjainen suosittelualgoritmi voidaan tämän perusteella sulkea pois, sillä suositukset tehdään enemmän kuin ainoastaan yhden käyttäjän aikaisemman käytöksen ja datan perusteella ja sisältöpohjainen suosittelualgoritmi tekee suositukset ainoastaan yhden käyttäjän aikaisemman käytöksen ja datan perusteella [15]. Tämän lisäksi suositusten tekemiseen vaikuttaa aktiivisen käyttäjän vuorovaikutushistoria, mikä myös viittaa yhteistoiminnalliseen suosittelualgoritmiin, sillä suositusten tekemiseen vaikuttaa aktiivisen käyttäjän oman käytöksensä lisäksi muiden käyttäjien käytös, mikä on tyypillistä yhteistoiminnallisissa suosittelualgoritmissa [1].

Sen lisäksi, että Instagram haluaa tehdä jokaiselle käyttäjälleen mahdollisimman hyviä suosituksia, sen täytyy myös moderoida sovellukseen julkaistua sisältöä, jotta sovellus on kaikille sen käyttäjilleen turvallinen alusta. Instagram on määritellyt nettisivuillaan yhteisösäännöt, jotka eivät koske vain syötettä ja tarinoita vaan koko Instagramia. Julkaistu sisältö, joka on yhteisösääntöjen vastainen, poistetaan. Sääntöjen rikkominen useampaan kertaan voi myös johtaa muihin rajoituksiin tai tilin poistamiseen. [23] Instagramin yhteisösääntöjen vastaista sisältöä ovat esimerkiksi postaukset, joita käyttäjä ei ole itse ottanut, postaukset, jotka sisältävät alastomuutta, väkivaltaa, itsensä vahingoittamista tai vihapuhetta. Tämän lisäksi käyttäjien tulee aina noudattaa lakia, eikä Instagramissa saa julkaista mitään lain vastaista, kuten postauksia, jotka ylistävät terrorismia tai järjestäytyntä rikollisuutta. [20]

Mahdollisesti yhteisösääntöjen vastaista sisältöä suodatetaan ja tarkastetaan teknologialla ja ihmistarkastajilla. Tekoälyteknologia on tarkistusprosessin keskiössä. Koneoppimismalleilla voidaan esimerkiksi tunnistaa, mitä kuvassa on tai analysoida julkaisun tekstiä. Täten tekoälymalleja voidaan kehittää oppimaan, esiintyykö sisällössä alastomuutta tai väkivaltaa. Tämän jälkeen mallit voivat määrittää, täytyykö sisällön suhteen ryhtyä joihinkin toimiin, kuten sen poistamiseen alustalta. Tekoäly voi myös lähettää sisällön ihmisen tekemään tarkistukseen, jos se vaatii lisätarkistusta. Tällöin Instagramin tarkistustiimi tekee lopullisen

päätöksen, miten sisällön kanssa tulisi toimia. Tekoälyteknologia oppii ja kehittyy näistä jokaisesta ihmisen tekemästä päätöksestä. [20]

3.2 Hakusivu

Instagramin hakusivu (engl. Explore page) on tarkoitettu auttamaan Instagramin käyttäjiä löytämään uusia asioita, joista he voisivat olla kiinnostuneita. Hakusivulla näytetään jokaiselle käyttäjälle Instagramin suosittelujärjestelmien tekemiä personoituja suosituksia käyttäjiltä, joita he eivät vielä seuraa. Suositukset voivat olla kuvia, videoita tai keloja. [23] Kelat ovat enintään 90 sekunnin pituisia pystysuuntaisia videoita, jotka näkyvät Instagramin kelat-välilehdellä [20].

Hakusivun suosittelualgoritmi seuraa samantyyppisiä vaiheita kuin syöte- ja tarinaosion suosittelualgoritmit. Erona on, kun syöte- ja tarinaosioissa suositellaan pääasiassa sisältöä aktiivisen käyttäjän seuraamilta käyttäjiltä, hakusivulla suositeltu sisältö on käyttäjiltä, joita aktiivinen käyttäjä ei seuraa. [23] Instagramiin julkaistaan miljardeja postauksia päivässä ja kaikki postaukset ovat kelvollisia hakusivun suosittelualgoritmin syötearvoksi [24]. Suosittelualgoritmin täytyisi pystyä tästä valtavasta syötteestä tekemään jokaiselle käyttäjille tarkkoja suosituksia.

3.2.1 Esisuodatusvaihe

Kuten syöte- ja tarinaosioissa hakusivullakin, suositteluprosessi alkaa siitä, että määritellään joukko kohteita, jotka olisivat mahdollisesti hyviä kohteita suositeltavaksi [23], mikä on tunnistettavissa esisuodatusvaiheeksi. Hakusivulla tämän joukon määrittäminen on hieman hankalampaa, koska syöte on niin iso. Pystyäkseen määrittämään tämän joukon suuresta syötteestä suosittelualgoritille on kehitetty hakuputkisto (engl. retrieval pipeline), joka keskittyy Instagramin käyttäjätilien tietoihin yksittäisen median tietojen sijaan. Tämä on tehokas lähestymistapa joukon määrittämiseen, sillä Instagramissa on suuri määrä tilejä, jotka perustuvat tiettyihin teemoihin, ja ne pyrkivät tavoittamaan käyttäjiä, jotka ovat kiinnostuneita kyseisestä teemasta. Määrittääkseen joukon mahdollisesti suositeltavia kohteita, suosittelualgoritmi ensin etsii käyttäjätilejä, jotka ovat aiheellisesti samantyyppisiä. Ensimmäinen suosittelualgoritmi hyödyntää käyttäjätilejä, joiden kanssa aktiivinen käyttäjä on ollut vuorovaikutuksessa, esimerkiksi tykännyt jostain käyttäjän postauksesta. Näitä käyttäjätilejä sanotaan siementileiksi. Seuraavaksi suosittelualgoritmi määrittelee siementilien kanssa samankaltaiset tilit käyttämällä tilien upotuksia (engl. account embeddings). [24]

Tarkastellaan tilien upotuksia samankaltaisen tekniikan sanojen upotuksen (engl. word embedding) avulla. Word embedding on tekniikka, jonka avulla sanajonoja voidaan esittää toisessa muodossa. Esittääkseen sanoja toisessa muodossa, word embedding hyödyntää tekstin uudessa esitysmuodossa vektoreita. Sanat, joilla on sama merkitys ja syntaksi saavat saman vektorimerkityksen. [25] Tilien upotuksessa sen sijaan, että tekstin esitysmuotoa muutetaan, muutetaan Instagramin käyttäjätilien esitysmuotoa ja esitetään ne vektoreina. Hakusivun suosittelualgoritmi pääättelee tilien upotukset käyttämällä ig2vec-mallia, joka on samankaltainen kuin word2vec-malli [24]. Word2vec on matala neuroverkkomalli, jota käytetään sanojen vektoriesitysten saamiseksi. Malli muuntaa sanaesityksen, joka on aakkosnumeeristen merkkien yhdistelmä, vektoreiksi, käyttäen neuroverkkoa (engl. neural network), jossa on täysin yhdistetty piilotettu kerros. [25]

Neuroverkko on koneoppismalli, joka on inspiroitunut ihmisen aivojen biologisesta hermoston toiminnasta. Neuroverkot koostuvat kerroksista, jotka sisältävät nk. neuroneita. Yleensä kerrokseen kuuluu syötekerros, yksi tai useampi piilotettu kerros ja ulostulokerros. [26] Matalissa ja kevyissä neuroverkoissa on tyypillisesti vain yksi tai kaksi piilotettua kerrosta, kun taas syvissä neuroverkoissa niitä voi olla paljon enemmän [27]. Kerroksissa olevat neuronit ovat yhteydessä toisiinsa ja jokaiseen liittyy paino ja vakio-termi. Neuroverkot voivat oppia esimerkkien perusteella, täten koulutusdatan avulla niitä voidaan kouluttaa erilaisiin tehtäviin, kuten tietojen luokitteluun, kuvioiden tunnistukseen ja ennustusten tekemiseen. [26] Neuroverkkoja voidaan kouluttaa päivittämällä ja säätämällä neuroneille asetettuja painoja [27]. Word2vec:ssa mallin piilokerroksen painomatriisia käytetään sanojen muuntamiseen vektoreiksi [25].

Käyttämällä samoja tekniikoita kuin word2vec, hakusivun suosittelualgoritmi pystyy ennustamaan tilejä, joiden kanssa aktiivinen käyttäjä todennäköisesti on vuorovaikutuksessa tietyn istunnon aikana Instagramissa. Instagramin ig2vec-malli käsittelee siementilien käyttäjätunnuksia sanajonona lauseessa ja esittää ne vektoreina. Kuten word2vec-mallikin, ig2vec-malli oppii sanajonon vektoriesityksen sen kontekstin perusteella. Kun käyttäjätunnukset esitetään vektoreina, kahden käyttäjän välinen samankaltaisuus voidaan laskea käyttämällä joko kosinietäisyyttä tai pistetuloa. Tämän perusteella suosittelualgoritmi tekee k-lähimmän naapurin haun (engl. K-nearest neighbor, KNN) löytääkseen aiheellisesti samankaltaisia tilejä siementileille. [24] KNN on luokittelu- ja regressioalgoritmi, jonka ideana on luokitella tuntematon kohde osaksi samaa ryhmää, kuin kohteen lähimmät naapurit.

Lähin naapuri voidaan määritellä samankaltaisuusmittarin avulla [28], mikä tässä tapauksessa lasketaan kosinietäisyydellä tai pistetulolla.

Tämän lisäksi suosittelualgoritmi käyttää Facebookin uusinta lähimmän naapurin hakukonetta FAISS:ia¹ tukevana hakuinfrastruktuurina. Lopuksi näiden siementilien kanssa samankaltaisten käyttäjätilien perusteella, suosittelualgoritmi pystyy löytämään mediaa, jota nämä tilit ovat julkaisseet tai jonka kanssa he ovat olleet tekemisissä, ja siitä mediasta muodostuu joukko kohteita, jotka ovat ehdokkaita suositeltaviksi kohteiksi. [24]

Hakusivulla jopa kymmenet tuhannet postaukset voivat olla mahdollisia ehdokkaita suositeltavaksi kohteeksi keskivertokäyttäjälle. Sen lisäksi, että Instagram poistaa alustalta sen yhteisösääntöjen vastaisia julkaisuja, Instagram haluaa pitää huolta siitä, että sen suosittelema sisältö on turvallista ja asianmukaista kaikenikäisille käyttäjilleen. Tämän takia ennen kuin suosittelualgoritmi määrittelee joukon mahdollisesti suositeltavia kohteita, Instagram suodattaa erilaisten signaalien avulla pois sisällön, jonka voi todeta olevan epäsovivaa sisältöä suositeltavaksi. [24]

Instagram on määritellyt suositusohjeet (engl. Recommendation Guidelines), joiden perusteella voidaan todeta sisällön olevan epäsovivaa suositeltavaksi. Sisältö, joka on suositusohjeiden vastainen, on sallittu Instagramissa, mutta sitä ei suositella muille käyttäjille, sillä sitä voidaan pitää arkaluontoisena tai loukkaavana. Instagram esimerkiksi pyrkii olemaan suosittelematta sisältöä, joka kuvaa väkivaltaa kuten tappeluita, tai sisältöä, joka kuvaa tiettyjen säänneltyjen tuotteiden, kuten tupakan, sähkö tupakan tai huumausaineiden käyttöä. Tämän lisäksi väärää tai harhaanjohtavaa sisältöä ja arkaluonteista tai heikkolaatuista terveyttä tai rahoitusta koskevaa sisältöä ei suositella. [23] Kuten tarkastaessa mahdollisesti yhteisösääntöjen vastaista sisältöä, Instagram hyödyntää koneoppimisjärjestelmiä havaitakseen ja suodattaakseen sisältöä, joka ei sovi suositeltavaksi [24].

3.2.2 Ranking-vaihe

Tämän jälkeen, kun joukko mahdollisia suositeltavia kohteita on määritelty, tulee joukon kohteille antaa sijoitukset. Joukko voi sisältää tuhansia kelvollisia postauksia keskivertokäyttäjää kohden ja koko joukon kohteiden arvioiminen vaatisi paljon resursseja. Joukosta otetaan 500 ehdokkaan näyte, joita käytetään hakusivun suosittelualgoritmin

¹ <https://faiss.ai>

kolmivaiheisessa sijoitusinfrastruktuurissa. Tämä infrastruktuuri pyrkii tasapainottamaan sijoituksen osuvuuden ja laskennan tehokkuuden välisiä kompromisseja. Kuitenkin mitä enemmän postauksia arvioidaan jokaista käyttäjää kohti, sitä suurempi mahdollisuus suosittelualgoritmillä on löytää paras personoiduin media käyttäjän määritellystä joukosta. Jotta median määrä voitaisiin maksimoida jokaista sijoituspyyntöä kohden, hakusivun suosittelualgoritmi käyttää ensimmäisessä sijoitusvaiheessa ranking-tislausmallia, joka valitsee joukosta ehdokkaat ennen monimutkaisempien sijoitusmallien käyttöä. [24]

Ranking-tislausmalli on koneoppimismalli, joka jäljittelee sijoitusinfrastruktuurin kahden muun sijoitusvaiheen yhdistelmää mahdollisimman minimaalisilla ominaisuuksilla. Mallia koulutetaan oppimaan sijoitusvaiheen myöhemmistä monimutkaisemmista malleista ja se yrittää approksimoida niitä mahdollisimman paljon. Monimutkaisempien sijoitusmallien syöte-ehdokkaat ja niiden ominaisuudet sekä niiden tulokset tallennetaan ja tislausmallia koulutetaan tämän tallennetun datan perusteella yksinkertaisemmalla neuroverkkomallirakenteella tulosten replikoimiseksi. Tislausmallin tavoitteena on optimoida pääsijoitusmallin tuloksien normalisoidun diskontatun kumulatiivisen hyödyn (NDCG) sijoitushäviö. [24] NDCG on suositusjärjestelmien tai hakutulosten laadun arvioimiseen käytetty mittari [29]. Ranking-tislausmalli valitsee 150 osuvinta ehdokasta 500 ehdokkaan joukosta, jotka lähetetään myöhemmän vaiheen monimutkaisempiin sijoitusmalleihin. Tämä tekniikka mahdollistaa sen, että suosittelualgoritmi voi arvioida suuremman joukon mediaa löytääkseen osuvimman median jokaista sijoituspyyntöä kohti, pitäen samalla laskennalliset resurssit hallinnassa. [24]

Sijoitusinfrastruktuurin toisessa vaiheessa kevyt neuroverkkomalli valitsee 50 osuvinta ehdokasta, tiheämmillä ominaisuuksilla kuin ranking-tislausmalli. Kolmannessa vaiheessa syvä neuroverkkomalli valitsee 25 osuvinta ehdokasta jäljelle jääneiden ehdokkaiden joukosta, joista muodostuu käyttäjän hakusivun ensimmäinen sivu. Kahdessa viimeisessä vaiheessa suosittelualgoritmi ennustaa kuinka todennäköisesti käyttäjä vuorovaikuttaa postauksen kanssa, kuten tykkää siitä, jotta se voi päätellä kuinka kiinnostunut käyttäjä voisi kyseisestä postauksesta olla. Jaetun monikerroksisen preseptoniverkon (MLP), avulla, suosittelualgoritmi kerää erilaisia signaaleja, joiden perusteella ennustukset tehdään. [24] MLP on eteenpäinsyöttävä neuroverkko, mikä tarkoittaa sitä, että neuroverkon informaatio virtaa ainoastaan eteenpäin kerroksesta toiseen [30].

Kerättyjä signaaleja ovat julkaistun postauksen tiedot, aktiivisen käyttäjän aktiivisuus hakusivulla, aktiivisen käyttäjän ja postauksen julkaisseen käyttäjän vuorovaikutuksen historia ja postauksen julkaisseen käyttäjän tiedot. Suositteuvalgoritmi kerää tietoja siitä, kuinka nopeasti julkaistusta postauksesta tykätään, kommentoidaan tai tallennetaan. Nämä signaalit merkitsevät paljon tehdessä suosituksia hakusivulla, sillä ne kertovat kuinka suosittu jokin postaus on. Tämän lisäksi suosittelualgoritmi kerää tietoa siitä, miten aktiivinen käyttäjä on vuorovaikuttanut hakusivun postausten kanssa aikaisemmin. Lisäksi suosittelualgoritmi kerää tietoa postauksen julkaisseesta käyttäjästä, kuten kuinka monta kertaa eri ihmiset ovat vuorovaikuttaneet kyseisen käyttäjän kanssa. [23]

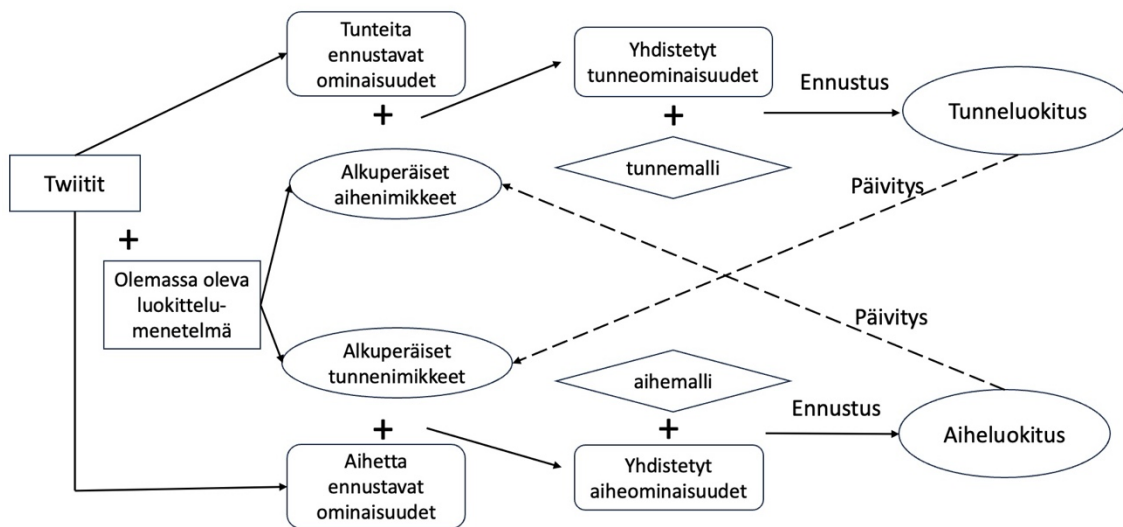
Hakusivun suosittelualgoritmi käyttää multi-task multi-label (MTML) -neuroverkkoa ennustusteiden tekemiseen [24]. MTML-neuroverkko hyödyntää monitehtäväoppimista (engl. multi-task learning) sekä moninimikeoppimista (engl. multi-label learning) [31].

Monitehtäväoppiminen on koulutusparadigma, jossa koneoppimismalleja koulutetaan oppimaan yhteisiä ideoita toisiinsa liittyvien tehtävien välillä, kouluttamalla niitä useiden tehtävien datalla samanaikaisesti jaettujen esitysten avulla [32]. Moninimikeoppiminen määrittää tietojoukon alkioille useita nimikkeitä samanaikaisesti. Jokainen nimikemäärittäminen voidaan nähdä ennustavan omistaako alkio nimikkeeseen vai ei. [33]

Huang ja muut [31] esittävät tutkimuksessaan MTML-luokitusmallin, joka suorittaa sekä syötteen tunneanalyysin että aiheuokituksen samanaikaisesti. Kirjoittajat ovat käyttäneet tutkimuksessaan mallin syötteenä sosiaalisen median alustan X:n (entinen Twitter) postauksia nk. twiitteja. Tutkimus keskittyy twiittien tunneanalyysiin, joka vangitsee käyttäjän subjektiivisen tunnelman, kuten ”positiivinen” ja ”negatiivinen” sekä twiitin aiheuokitukseen, joka osoittaa sisällön aiheen ennalta määritetyistä näkökulmista, kuten ”uutiset” ja ”kohteliaisuus”. Kun otetaan huomioon sattumanvarainen kokoelma twiitteja, aiheuokitus voi paljastaa tietyt aiheet ja näkökulmat, joista käyttäjät puhuvat tai jotka ovat hallitsevia, kun taas tunneanalyysi kertoo osuuden käyttäjiä, jotka suhtautuvat aiheeseen myönteisesti tai negatiivisesti. Usein tunneanalyysia ja aiheuokitusta on käsitelty erillään toisistaan, vaikka ne voivat usein liittyä toisiinsa. Twiittien aiheet voivat auttaa edistämään tunneanalyysia ja päinvastoin. Koska nämä kaksi tehtävää liittyvät toisiinsa, monitehtäväoppiminen voi vahvistaa tehtävien tuloksia kuin niiden suorittaminen itsenäisesti. Samaan aikaan jokainen tehtävä voi sisältää useampia nimikkeitä, eli yksi twiitti voi viitata useampaan kuin yhteen aiheeseen. Moninimikeoppimisella syötteen alkioille voidaan määrittää useampi nimike samanaikaisesti. Täten nämä kaksi luokittelutyölä on yhdistetty

MTML-luokitusmalliksi. Tutkimuksen esittämä malli yhdistää molempien analyysien tulokset aikaisemmista vaiheista edistääkseen ja vahvistaakseen nykyisiä tuloksia iteratiivisesti. Yhden tehtävän opitut luokkanimikkeet sisällytetään osaksi toisen tehtävän ominaisuuksien ennakkointia.

Kuvassa 3-1 on havainnollistettu Huang ja muut [31] tutkimuksessa käytettyä luokittelumallia. Ensin syötteenä olevista twiiteista poimitaan tunteita ja aiheita ennustavia ominaisuuksia kuten avainsanoja. Samaan aikaan jo jollain olemassa olevalla luokittelumenetelmällä twiittien alkuperäiset luokkatunnisteet voidaan tunnistaa. Tämän jälkeen sisällyttämällä alkuperäiset luokkatunnisteet ennakoiviin ominaisuuksiin, saadaan yhdistetyt tunne- ja aiheominaisuudet. Mallia voidaan kouluttaa estimoimalla koulutustietojoukon kertoimia. Kun malli on koulutettu, se voi yhdistettyjen ominaisuuksien avulla luoda uusia alkuperäisiä aihe- että tunnenimikkeitä. Kahden luokittelun toistaminen iteratiivisesti pitää luokkatunnisteet ajan tasalla, kunnes ne eivät enää muutu merkittävästi ja haluttu tulos on saavutettu.



Kuva 3-1 Huang ja muut [31] tutkimuksessa esitetty luokittelumalli (Kuvaa muokattu lähteestä [31])

Tarkastellaan Instagramin hakusivun suosittelujärjestelmän MTML-mallia Huang ja muut [31] tutkimuksen avulla. Hakusivun suosittelualgoritmi käyttää MTML-mallia ennustaakseen, miten aktiivinen käyttäjä tulee vuorovaikuttamaan postauksen kanssa. Olipa vuorovaikutus sitten positiivinen, kuten tykkääminen tai negatiivinen, kuten ”näytä vähemmän tällaisia postauksia”-napin painaminen. [24] Voidaan päätellä, että samaan tapaan kuin tutkimuksenkin MTML-malli, Instagramin hakusivun suosittelujärjestelmän MTML-malli

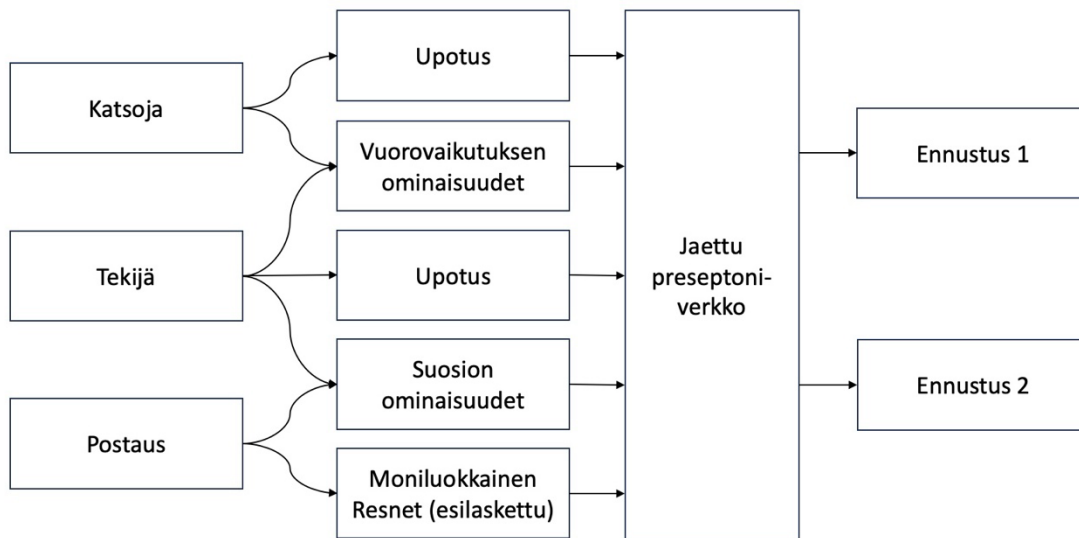
pyrkii määrittämään, mihin aiheiluokkaan suositeltava postaus kuuluu ja suhtautuuko käyttäjä kyseiseen aiheeseen myönteisesti vai negatiivisesti. Tutkimuksessa käytetyt twiitit sisältävät tekstiä, mutta Instagramin sisältö on pääosin kuvia tai videoita. Moninimike-kuvaluokittelun avulla voidaan pyrkiä tunnistamaan kuvissa olevat eri objektit tai attribuutit [34]. Postauksista saadut luokkanimikkeet edistävät tunneanalyysin tekoa ja täten suosittelualgoritmi pystyy ennustamaan kuinka todennäköisesti käyttäjä vuorovaikuttaa postauksen kanssa, joko negatiivisesti tai positiivisesti.

Kuvassa 3–2 on esitetty Instagramin hakusivun suosittelualgoritmin käyttämä MTML-malli. Hakusivun suosittelualgoritmin MTML-mallin ja Huang ja muut [31] tutkimuksen MTML-mallin välillä voidaan nähdä yhtäläisyyksiä. Hakusivun suosittelualgoritmin käyttämä MTML-malli kerää tutkimuksessa esitetyn mallin mukaisesti erilaisia ominaisuuksia syötejoukostaan, jotta se voi ennustaa kuinka todennäköisesti käyttäjä vuorovaikuttaa postauksen kanssa. Syötejoukko on hakusivun suosittelualgoritmin MTML-mallille tutkimukseen verrattuna paljon suurempi, sillä tutkimuksessa keskityttiin ainoastaan luokiteltavien twiittien sisältöön. Hakusivun suosittelualgoritmin tehdessä suosituksia, ennustuksen tekemiseen vaikuttaa myös itse aktiivinen käyttäjä sekä postauksen julkaissut käyttäjä. Postauksen tekijästä ja aktiivisesta käyttäjästä kerätään vuorovaikutuksen ominaisuuksia, mikä voi tarkoittaa esimerkiksi mistä postauksista he ovat tykänneet aikaisemmin tai keiden käyttäjien kanssa he ovat olleet vuorovaikutuksessa. Tämän lisäksi postauksen tekijästä kerätään suosion ominaisuuksia. Tämän voidaan päätellä auttavan ennustamaan sitä, kuinka kiinnostava kyseinen käyttäjä todennäköisesti on. Näiden asioiden lisäksi aktiivisesta käyttäjästä sekä postauksen tehneestä käyttäjästä tehdään aiemmin esitelty tilien upotus. Tilien upotuksella molemmille käyttäjille löydetään samankaltaisia tilejä, joiden avulla voidaan määritellä, mihin aiheiluokkaan postauksen tehnyt käyttäjä kuuluu ja minkälaisista aiheista aktiivinen käyttäjä on aiemmin tykännyt. Tilien upotusta voidaan pitää tutkimuksessa esitetyn mallin mukaisesti alkuperäisten luokkatunnisteiden määrittelynä. Täten saadaan tietää, mihin aiheiluokkaan postauksen tehnyt käyttäjä kuuluu ja onko aktiivinen käyttäjä pitänyt aiemmin tämän kaltaisesta sisällöstä.

Kuvassa 3–2 on myös huomattavissa, että itse postauksesta kerätään suosion ominaisuuksia, minkä voidaan päätellä auttavan ennustamaan, kuinka kiinnostava postaus todennäköisesti on. Tämän lisäksi postauksesta määritellään ominaisuuksia esilasketun ResNet:in avulla (engl. Residual Network). ResNet on syväoppimismalli, joka on alun perin suunniteltu vastaamaan erittäin syvien neuroverkkojen kouluttamisen haasteisiin, käyttämällä ohitusyhteyksiä (engl.

skip connection) tai jäännösyhteyksiä (engl. residual connection). ResNet koostuu jäännöslohkoista (engl residual block), jotka taas koostuvat yhdestä tai useammasta neuroverkkokerroksesta, joissa on oikotie tai ohitusyhteys. Oikotie tai ohitusyhteys ohittaa yhden tai useamman neuroverkkokerroksen. Ohitusyhteys yhdistää lohkon syötteen sen tulostukseen, minkä ansiosta se voi oppia säättämään tai muokkaamaan kyseisen neuroverkkokerroksen tulosta menettämättä syötteen lähtötietoa. [35]

ResNet on erittäin tehokas tekniikka kuvien luokittelussa [35] sekä ominaisuuksien poimimisesta kuvista [36]. Täten voidaan päätellä, että hakusivun MTML-mallissa ResNet:iä käytetään poimimaan postauksen ominaisuuksia. Tämän voidaan päätellä olevan postauksen luokkatunnisteiden määrittelyä, kuten tutkimuksessa Huang ja muut [31] on määritelty twiittien luokkatunnisteet.



Kuva 3-2 Instagramin hakusivun suosittelualgoritmin käyttämä MTML-malli [24]

Huang ja muut tutkimuksessa [31] yhdistämällä poimitut ominaisuudet ja alkuperäiset luokkatunnisteet pystyttiin määrittämään uusia luokkatunnisteita, jota toistamalla iteratiivisesti päästiin haluttuun tulokseen. Instagramin hakusivun suosittelualgoritmissa oleva jaettu preseptoniverkko (MLP), määrittelee minkä signaalien perusteella ennustukset tehdään [24]. Täten voidaan päätellä, että hakusivun suosittelualgoritmin MTML-mallissa oleva MLP, tutkimuksen MTML-mallin mukaan, yhdistää kerätyt ominaisuudet ja määritellyt alkuperäiset luokkanimikkeet, jotta se pystyy tekemään ennustuksia, kuinka todennäköisesti käyttäjä vuorovaikuttaa kyseisen postauksen kanssa. Päätelmää tukee se, että hakusivun suosittelualgoritmi yhdistää tehtyjä ennustuksia käyttämällä ennusteille painotettua summaa,

päätäessä onko postaus relevantti vai ei [24]. Tämä tarkoittaisi sitä, että kuten tutkimuksessakin ennusteita tehtäisiin iteratiivisesti, hyödyntäen aiemman ennustuksen tulosta seuraavan tekemiseen. Täten suosittelualgoritmi on ennustanut, kuinka todennäköisesti käyttäjä vuorovaikuttaa postauksen kanssa ja suosittelee postauksia ennusteiden mukaan käyttäjälle.

Tämän lisäksi Instagram on lisännyt painotettuun summaan yksinkertaisen heuristisen säännön, jotta hakusivulla näkyy tarpeeksi monimuotoista sisältöä. Hakusivun suosittelualgoritmi lisää rangaistustekijän saman tekijän tai siementilin postauksiin, jotta postaus ei saisi niin korkeaa sijoitusta ja täten ei näy niin korkealla aktiiviselle käyttäjälle. Tämä takaa sen, että aktiivinen käyttäjä ei näe hakusivulla useita postauksia samalta tekijältä tai siementililtä. [24]

Näiden ominaisuuksien pohjalta Instagramin hakusivun suosittelualgoritmi vaikuttaa olevan yhteistoiminnallisen käyttäjäpohjaisen sekä sisältöpohjaisen suosittelualgoritmin hybridisuosittelualgoritmi. Algoritmi vaikuttaa olevan sisältöpohjainen, sillä tehdessään suosituksia se ottaa huomioon sisällön, jota aktiivinen käyttäjä ei ole todennäköisesti vielä nähnyt, sillä suositukset ovat pääosin käyttäjiltä, joita aktiivinen käyttäjä ei seuraa. Tämä viittaa sisältöpohjaiseen suosittelualgoritmiin, sillä se suosittelee käyttäjälle kohteita, joita hän ei ole vielä nähnyt tai arvostellut. Tämän lisäksi sisältöpohjainen suosittelualgoritmi suosittelee kohteita, jotka ovat samankaltaisia kuin kohteet, joista käyttäjä on jo pitänyt aikaisemmin. [15] Hakusivun suosittelualgoritmi tekee suosituksia samalla tekniikalla. Se pyrkii suosituksia tehdessä päättelemään, kuinka todennäköisesti aktiivinen käyttäjä voisi pitää kohdetta kiinnostavana sen perusteella, mistä kohteista ja teemoista aktiivinen käyttäjä on pitänyt aikaisemmin. Sisältöpohjaisessa suosittelussa käyttäjän aikaisemman käytöksen lisäksi tärkeässä roolissa on myös kohteen kuvaus [15]. Tämän mukaisesti hakusivun suosittelualgoritmi kerää tietoa myös siitä, kuinka suosittu mahdollisesti suositeltava postaus on, jotta se voi päätellä kuinka kiinnostava se on.

Yhteistoiminnalliseen käyttäjäpohjaiseen suosittelualgoritmiin taas viittaa se, että suositusten tekemiseen ei vaikuta ainoastaan aktiivisen käyttäjän data ja käytös vaan myös esimerkiksi postauksen julkaisseen käyttäjän data ja käytös. Tämän lisäksi hakusivun suosittelualgoritmin käyttämä MTML-malli määrittää tilien upotuksella luokkatunnisteet aktiiviselle sekä postauksen julkaisseelle käyttäjälle. Tilien upotuksessa pyritään löytämään käyttäjän kanssa samankaltaisia tilejä. Nämä viittaavat yhteistoiminnalliseen käyttäjäpohjaiseen

suosittelualgoritmiin, sillä käyttäjäpohjaisessa suosittelussa algoritmi etsii samankaltaisuuksia eri käyttäjien välillä vertailemalla heidän arvosteluitansa samalle kohteelle [15]. Hakusivun suosittelualgoritmi kerää tietoa siitä, mistä aktiivinen käyttäjä sekä postauksen julkaissut käyttäjä ovat tykänneet aikaisemmin, jotta se pystyy löytämään heidän kanssaan samankaltaisia tilejä.

3.3 Kelat

Instagram kelat (engl. Instagram Reels) on sovelluksen osa, jossa Instagramin käyttäjät pystyvät kuvamaan ja muokkaamaan enintään 90 sekunnin pituisia pystysuuntaisia videoita. Kuvattuihin videoihin pystyy lisäämään musiikkia, tehosteita, tarroja sekä ääninauhoja. Julkaistut kelat näkyvät kelat-välilehdellä. [20] Kelat-välilehdellä käyttäjät pystyvät löytämään uutta kiinnostavaa sisältöä, sillä aivan kuten hakusivullakin suurin osa välilehdellä näytettävästä sisällöstä on sellaisilta käyttäjiltä, joita aktiivinen käyttäjä ei vielä seuraa [24].

Koska kelat-välilehti on hyvin samanlainen kuin hakusivu, on sen sisällön suositteluprosessikin hyvin samankaltainen. Ensin suosittelualgoritmi kerää joukon keloja, joista aktiivinen käyttäjä voisi pitää ja tämän jälkeen joukon kohteille annetaan sijoitukset sen perusteella, kuinka mielenkiintoisia ne ovat aktiiviselle käyttäjälle suosittelualgoritmin mukaan. [24] Nämä kaksi vaihetta voidaan tunnistaa esisuodatusvaiheeksi ja ranking-vaiheeksi.

Pystyäkseen määrittämään joukon kohteita, joista aktiivinen käyttäjä voisi pitää, käyttäjille esitetään kyselyitä, joissa heiltä kysytään, pitivätkö he jotain tiettyä kelaa hauskana tai ajankäyttönsä arvoisena. Näiden kyselyiden vastausten pohjalta suosittelualgoritmi oppii käyttäjien mielenkiinnon kohteista ja pystyy määrittämään joukon kohteita, joista se uskoo aktiivisen käyttäjän pitävän. [24]

Ranking-vaiheessa pystyäkseen antamaan sijoituksia määritellyn joukon kohteille suosittelualgoritmi tekee ennustuksia siitä, kuinka todennäköisesti käyttäjä jakaa kelan uudelleen, katsoo koko kelan loppuun asti, tykkää kelasta ja siirtyy kelan äänisivulle. Ennustus siirtyykö käyttäjä kelan äänisivulle voi kertoa siitä, saako käyttäjä inspiraatiota oman kelan tekemiseen. Ennustusten tekemiseen suosittelualgoritmi kerää signaaleja aktiivisen käyttäjän aktiivisuudesta, aktiivisen käyttäjän vuorovaikutuksen historiasta kelan julkaisseeseen käyttäjän kanssa, kelan tiedoista sekä kelan julkaisseeseen käyttäjän tiedoista. [24]

Suosittelualgoritmi kerää tietoa esimerkiksi siitä, mistä keloista aktiivinen käyttäjä on tykännyt tai mitä hän on jakanut ja kommentoinut viime aikoina. Nämä signaalit helpottavat suosittelualgoritmia ymmärtämään, mikä sisältö on aktiiviselle käyttäjälle olennaista. Tämän lisäksi suosittelualgoritmi kerää tietoa julkaistusta kelasta kuten sen sisällöstä, ääniraidasta ja suosiosta. Kelan julkaisseesta käyttäjästä kerätään tietoa, kuinka suosittu kyseinen käyttäjä on, esimerkiksi seuraajien lukumäärä ja kuinka paljon käyttäjän ja hänen sisältönsä kanssa vuorovaikutetaan. Tämä auttaa suosittelualgoritmia määrittämään, kuinka kiinnostava käyttäjä on kyseessä. Kuten hakusivullakin, selatessa kelat-välilehteä julkaistut videot ovat usein käyttäjiltä, joita aktiivinen käyttäjä ei seuraa. Suosittelualgoritmia kuitenkin kiinnostaa, jos aktiivinen käyttäjä on ollut kelan julkaisseen käyttäjän kanssa vuorovaikutuksessa, sillä se antaa käsityksen siitä kuinka kiinnostunut aktiivinen käyttäjä saattaa olla hänen sisällöstään. [24]

Näiden kerättyjen signaalien perusteella suosittelualgoritmi pystyy ennustamaan, kuinka todennäköisesti aktiivinen käyttäjä on jonkinlaisessa vuorovaikutuksessa tietyn kelan kanssa ja pystyy antamaan sille sijoituksen. Mitä paremman sijoituksen kela saa sitä korkeammalla se näkyy käyttäjälle. Sijoitusten lisäksi keloja koskee myös samat suositusohjeet kuin hakusivuakin. Jos julkaistu kela on näiden ohjeiden vastainen, sen näkyvyyttä pyritään vähentämään. [24]

Instagramin kelaosion suosittelualgoritmi ominaisuuksiltaan vaikuttaa eniten olevan tuotepohjainen yhteistoiminnallinen suosittelualgoritmi. Päätelmä perustuu siihen, että kelaosiossa suositukset ovat pääosin käyttäjiltä, joita aktiivinen käyttäjä ei vielä seuraa, minkä vuoksi suositteluprosessissa on mukana useamman kuin vain aktiivisen käyttäjän dataa. Tämän takia kyseessä ei voi olla sisältöpohjainen suosittelualgoritmi, sillä se tekee suositukset ainoastaan yhden käyttäjän aikaisemman käytöksen ja datan perusteella [15].

Suosittelualgoritmin voidaan päätellä olevan tuotepohjainen, koska se etsii samankaltaisuuksia suositeltavien kohteiden välillä eikä käyttäjien, mikä on tuotepohjaisen yhteistoiminnallisen suosittelualgoritmin toimintamalli [15]. Esimerkiksi määrittäessään joukkoa kohteista, joista aktiivinen käyttäjä voisi pitää, algoritmi pyrkii kyselyiden avulla löytämään samankaltaisia keloja, joista aktiivinen käyttäjä on jo aikaisemmin tykännyt. Tämän lisäksi ranking-vaiheessa suosittelualgoritmi pyrkii määrittämään, minkälaisista kohteista aktiivinen käyttäjä on aikaisemmin tykännyt, pystyäkseen antamaan osuvan suosituksen suositeltavalle kohteelle.

4 Pohdintaa

Usein eri sosiaalisen median alustoilla on käytössä ainoastaan yksi suosittelualgoritmi kaikille alustan osa-alueille. Instagram on kuitenkin päättänyt jakaa suosittelujärjestelmänsä tarkempiin käyttötarkoituksiin alustansa sisällä [23]. Jokainen alustan sisäinen suosittelualgoritmi on optimoitu kyseisen osa-alueen ominaisuuksiin, minkä takia algoritmin ei tarvitse ottaa huomioon muiden alustan osa-alueiden eri käyttötarkoituksia. Taktiikka näyttää toimineen hyvin, sillä se varmistaa, että suosittelualgoritmin tekemät suositukset ovat varmasti osuvia ja kiinnostavia. Tämä varmasti parantaa Instagramin käyttökokemusta ja saa käyttäjät palaamaan sovellukseen aina uudestaan. Täysin ongelmattomia eivät Instagramin suosittelujärjestelmät kuitenkaan ole.

Kuten tutkielmassa on jo todettu, sosiaalisen median sovellusten suosittelujärjestelmien tutkiminen on osittain hankalaa, sillä usein ne ovat salaisia julkiselle yleisölle [7]. Suosittelujärjestelmien ollessa kuitenkin niin yleisiä nykypäivän digitaalisilla alustoilla ja yritysten tavoite parantaa niitä koko ajan, motivoi tutkimuksen tekoa suosittelujärjestelmiä kohtaan. Samaan aikaan suosittelujärjestelmien käyttö herättää huolta sovellusten käyttäjissä niiden ei-toivotuista vaikutuksista sovelluksen käyttöön.

Yksi tulevaisuuden tärkeä tutkimuskohde varmasti on suosittelujärjestelmien puolueellisuus. Aivan kuten mikä tahansa digitaalinen alusta myös suosittelujärjestelmät voivat olla puolueellisia. Tällä tarkoitetaan sitä, että järjestelmä suosittelee selkeästi johonkin ryhmään kuuluvia kohteita enemmän kuin toisia [37]. Usein esimerkiksi vähemmistöjen edustajien julkaisemia postauksia suositellaan sosiaalisen median alustoilla vähemmän kuin valtaväestön.

Suosittelujärjestelmien puolueellisuutta voi aiheuttaa esimerkiksi data, jolla järjestelmien neuroverkkoja koulutetaan. Jos koulutusdatassa jokin ihmisryhmä on enemmän esitettyä, suosittelujärjestelmä oppii puolueelliseksi [37]. Instagramin osalla tämä voi johtaa ihmisten eriarvoistumiseen, jo olemassa olevien ennakkoluulojen ja stereotyyppien vahvistamiseen sekä siihen, että kaikki erilaiset mielipiteet eivät enää saisi näkyvyyttä Instagramissa.

On tärkeää, että nykypäivänä sosiaalisen median alustoilla myös vähemmistöt ja eriävät mielipiteet saisivat näkyvyyttä, ja sen sijaan, että jo olemassa olevia ennakkoluuloja vahvistettaisiin, saataisiin niitä purettua sosiaalisen median sovelluksien avulla. Sosiaalisen median sovellukset saavuttavat miljardeja ihmisiä päivässä ja oikein käytettynä ne voisivat

tehokkaasti auttaa purkamaan vallitsevia ennakkoluuloja ja stereotyyppioita. Täten jokaiselle tulisi taata samanlaiset mahdollisuudet verkostoitua ja saada näkyvyyttä sosiaalisen median sovelluksissa.

Möller ja muut [38] ovat kuitenkin tutkimuksessaan havainneet sen, että huolellisella algoritmin suunnitellulla puolueellisuus voitaisiin minimoida. Jos sosiaalisen median sovelluksien suosittelujärjestelmät olisivat julkisia, puolueellisuutta ja algoritmien toimintaa saataisiin tutkittua paremmin ja täten puolueellisuus voitaisiin saada kitkettyä suosittelujärjestelmistä kokonaan pois. Tämänhetkinen tilanne, jossa suosittelujärjestelmät ovat yrityssalaisuuksia ei kuitenkaan tarjoa tarvittavia resursseja siihen.

Tulevaisuuden tutkimuskohteena tulisi myös olla suosittelujärjestelmien yksityisyys. Suosittelujärjestelmien salauksen vuoksi, keskiverto sovelluksen käyttäjä ei tiedä kuinka suuria tietoturvaohjeita sosiaalisen median sovelluksien käyttöön liittyy. Tulevaisuudessa tulee panostaa tutkimuksessa siihen, että sovelluksen käytön tietoturvariskit tuotaisiin käyttäjille selkeästi ilmi sekä, että tietoturvariskejä saataisiin minimoitua ja käyttäjien yksityisyyttä parannettua.

Yllä mainituilla huomioilla tuntuu olevan yksi yhteinen tekijä. Suosittelujärjestelmien salaisuus vaikeuttaa niiden korjaamista. Tutkimusta näiden aiheiden tiimoilta on jatkettava, jotta suosittelujärjestelmien toiminta ja toimintaperiaatteet saataisiin enemmän julkiseen tietoon ja täten niiden epäkohtia saataisiin helpommin parannettua.

5 Yhteenveto

Tämän tutkielman tutkimuskysymykset olivat:

TK1: Miten Instagram suosittelee käyttäjilleen postauksia?

TK2: Mihin suosittelualgoritmeihin Instagramin suosittelujärjestelmät vaikuttavat perustuvan?

Instagram on jakanut suosittelujärjestelmänsä erilleen jokaiseen sovelluksen eri osaan (syöte, tarinat, hakusivu ja kelat). Näissä sovelluksen osissa käytetään omaa suosittelualgoritmia, joka on räätälöity sen mukaan, kuinka ihmiset näitä sovelluksen osia käyttävät. Jokaisen sovelluksen osan suosittelualgoritmi käy läpi suosittelualgoritmeille tyypillisen esisuodatus- sekä ranking-vaiheen. Esisuodatusvaiheessa suosittelualgoritmi määrittelee osajoukon kohteita, jotka ovat hyviä ehdokkaita mahdollisesti suositeltaviksi kohteiksi. Ranking-vaiheessa suosittelualgoritmi antaa jokaiselle osajoukon kohteelle sijoitukset sen tekemän ennustuksen mukaan kuinka osuva kohde mahdollisesti voisi käyttäjälle olla. Sijoitukset lasketaan jollain jo olemassa olevalla suosittelualgoritmillä.

Syöteosiossa suosittelualgoritmi tekee ennustuksen osajoukon kohteiden kiinnostavuudesta aktiivisen käyttäjän aktiivisuuden, julkaistujen postauksien tietojen, postauksen julkaiseman käyttäjän tietojen sekä aktiivisen käyttäjän vuorovaikutuksen historian perusteella.

Tarinaosiossa ennustukset tehdään aktiivisen käyttäjän katseluhistorian, vuorovaikutushistorian tarinoissa sekä käyttäjien välisen läheisyyden perusteella. Kelaosiossa suosittelualgoritmi tekee ennustukset aktiivisen käyttäjän aktiivisuuden, aktiivisen käyttäjän vuorovaikutuksen historian kelan julkaisseen käyttäjän kanssa, kelan tietojen sekä kelan julkaisseen käyttäjän tietojen perusteella.

Hakusivulla esisuodatusvaiheessa käytetään tilien upotuksia, jonka avulla pystytään määrittämään samankaltaisia tilejä, joista aktiivinen käyttäjä on jo aikaisemmin pitänyt. Näiden tilien postauksista pystytään määrittämään joukko mahdollisesti suositeltavia kohteita. Ranking-vaiheessa hakusivun suosittelualgoritmi käyttää kolmivaiheista sijoitusinfrastruktuuria. Ensimmäisessä vaiheessa koneoppimismalli valitsee 150 osuvinta kohdetta määrittelystä joukosta, toisessa vaiheessa 50 ja kolmannessa vaiheessa 25. Kahdessa viimeisessä vaiheessa suosittelualgoritmi ennustaa kuinka todennäköisesti käyttäjä vuorovaikuttaa postauksen kanssa, jotta se voi päätellä kuinka kiinnostunut käyttäjä voisi

kyseisestä postauksesta olla. Hakusivun suosittelualgoritmi käyttää multi-task multi-label (MTML) -neuroverkkoa ennustusteiden tekemiseen.

Tutkielmassa huomataan, että Instagramin syöte- sekä kelaosion suosittelualgoritmit vaikuttavat ominaisuuksiltaan tuotepohjaiselta yhteistoiminnalliselta suosittelualgoritmilta. Tarinaosion suosittelualgoritmi vaikuttaa käyttäjäpohjaiselta yhteistoiminnalliselta suosittelualgoritmilta ja hakusivun suosittelualgoritmi vaikuttaa sisältöpohjaisen suosittelualgoritmin sekä käyttäjäpohjaisen yhteistoiminnallisen suosittelualgoritmin hybridisuosittelualgoritmilta.

Sosiaalisen median sovellukset kehittyvät jatkuvasti niiden käyttäjäkunnan kasvaessa ja teknologian kehittyessä. Yritykset pyrkivät kehittämään sovelluksistaan parhaimman mahdollisen version, jotta käyttäjät palaisivat sinne aina uudestaan. Tästä syystä tässä tutkielmassa tehdyt päätelmät koskevat ainoastaan tämänhetkisiä Instagramin suosittelujärjestelmiä. Instagram voi kehittää suosittelujärjestelmiään eteenpäin lähitulevaisuudessa, jolloin niiden toiminta voi muuttua ja tämän tutkielman päätelmät eivät ole enää ajankohtaisia.

Tutkielman aikana on huomattu, että Instagramin suosittelujärjestelmistä kertovan akateemisen kirjallisuuden määrä on hyvin vähäinen. Tämän voidaan päätellä johtuvan siitä, että sosiaalisen median sovellusten suosittelujärjestelmät ovat hyvin vaikea tutkimuskohde, sillä ne eivät ole julkisesti, eivätkä tutkimuksen tekoon, nähtävissä. Tästä syystä sosiaalisen median sovellusten suosittelujärjestelmien tutkimusta on tärkeä jatkaa tulevaisuudessa, jotta aihe saataisiin saataville suuremmalle yleisölle. Tämän lisäksi tutkimusta aiheen parissa olisi syytä jatkaa suosittelujärjestelmissä nousevien ongelmien vuoksi. Esimerkiksi suosittelujärjestelmien puolueellisuus ja yksityisyys ovat ongelmia, joita kannattaa tulevaisuudessa tutkia.

Tutkimuksen jatkaminen aiheen parissa hyödyttäisi sosiaalisen median sovellusten käyttäjiä, jakamalla heille tietoa, miten sovellus oikeasti toimii. Tällä tavoin käyttäjille voitaisiin tuoda ilmi sovellusten mahdolliset tietoturvaongelmat ja yksityisyysriskit. Tämän lisäksi tutkimuksen jatkamisella suosittelujärjestelmissä nousevia ongelmia pystyttäisiin paremmin minimoimaan.

Lähteet

- [1] R. Sharma, S. Rani, ja S. Tanwar, ”Machine Learning Algorithms for building Recommender Systems”, teoksessa *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 2019, ss. 785–790. doi: 10.1109/ICCS45141.2019.9065538.
- [2] S. Jain, A. Grover, P. S. Thakur, ja S. K. Choudhary, ”Trends, problems and solutions of recommender system”, teoksessa *International Conference on Computing, Communication & Automation*, 2015, ss. 955–958. doi: 10.1109/CCAA.2015.7148534.
- [3] D. Jannach, P. Pu, F. Ricci, ja M. Zanker, ”Recommender Systems: Past, Present, Future”, *AI Mag.*, vsk. 42, ss. 3–6, syys 2021, doi: 10.1609/aimag.v42i3.18139.
- [4] M. Ge, G. Pilato, F. Persia, ja D. D’Auria, ”Recommender System for Social Media: Research Challenges and Future Applications”, teoksessa *2023 Fifth International Conference on Transdisciplinary AI (TransAI)*, 2023, ss. 253–257. doi: 10.1109/TransAI60598.2023.00033.
- [5] M. Ge, F. Persia, ja D. D’Auria, ”Advanced Recommender Systems by Exploiting Social Networks”, teoksessa *2019 IEEE International Conference on Humanized Computing and Communication (HCC)*, 2019, ss. 118–125. doi: 10.1109/HCC46620.2019.00025.
- [6] M. Ge ja F. Persia, ”Factoring Personalization in Social Media Recommendations”, teoksessa *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, 2019, ss. 344–347. doi: 10.1109/ICOSC.2019.8665624.
- [7] K. Bishqemi ja M. Crowley, ”TikTok Vs. Instagram: Algorithm Comparison”, *J. Stud. Res.*, vsk. 11, nro 1, helmi 2022, doi: 10.47611/jsrhs.v11i1.2428.
- [8] M. Mansoury, B. Mobasher, R. Burke, ja M. Pechenizkiy, ”Bias Disparity in Collaborative Recommendation: Algorithmic Evaluation and Comparison”, heinäkuu 2019.
- [9] V. Tsintzou, E. Pitoura, ja P. Tsaparas, ”Bias Disparity in Recommendation Systems”, *CoRR*, vsk. abs/1811.01461, 2018, [Verkossa]. Saatavissa: <http://arxiv.org/abs/1811.01461>
- [10] A. Mosseri, ”Ensuring Black Voices are Heard”, Instagram. Viitattu: 19. maaliskuuta 2024. [Verkossa]. Saatavissa: <https://about.instagram.com/blog/announcements/ensuring-black-voices-are-heard>

- [11] B. Patel, P. Desai, ja U. Panchal, "Methods of recommender system: A review", teoksessa *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, ss. 1–4. doi: 10.1109/ICIIECS.2017.8275856.
- [12] R. Farhat, Y. Murali, M. Jemni, ja H. Ezzedine, "An overview of Machine Learning Technologies and their use in E-learning", teoksessa *2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA)*, 2020, ss. 1–4. doi: 10.1109/OCTA49274.2020.9151758.
- [13] T. Rincy N ja R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques"; teoksessa *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2020, ss. 1–6. doi: 10.1109/SCEECS48394.2020.190.
- [14] R. Prabakaran, J. Pradeepkandhasamy, ja M. Arun, "A Survey on Recommendation Systems using Collaborative Filtering Techniques", teoksessa *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2023, ss. 1445–1450. doi: 10.1109/ICSSIT55814.2023.10060889.
- [15] F. Mansur, V. Patel, ja M. Patel, "A review on recommender systems", teoksessa *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017, ss. 1–6. doi: 10.1109/ICIIECS.2017.8276182.
- [16] 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, ss. 1–4. doi: 10.1109/ICIIECS.2017.8276172.
- [17] M. H. Mohamed, M. H. Khafagy, ja M. H. Ibrahim, "Recommender Systems Challenges and Solutions Survey", teoksessa *2019 International Conference on Innovative Trends in Computer Engineering (ITCE)*, 2019, ss. 149–155. doi: 10.1109/ITCE.2019.8646645.
- [18] Y. Himeur, S. S. Sohail, F. Bensaali, A. Amira, ja M. Alazab, "Latest trends of security and privacy in recommender systems: A comprehensive review and future perspectives", *Comput. Secur.*, vsk. 118, s. 102746, 2022, doi: <https://doi.org/10.1016/j.cose.2022.102746>.
- [19] Y. Register, L. Qin, A. Baughan, ja E. S. Spiro, "Attached to "The Algorithm": Making Sense of Algorithmic Precarity on Instagram", teoksessa *Proceedings of the*

- 2023 CHI Conference on Human Factors in Computing Systems*, teoksessa CHI '23. New York, NY, USA: Association for Computing Machinery, 2023. doi: 10.1145/3544548.3581257.
- [20] Instagram, Instagram Ohje- ja tukikeskus. Viitattu: 5. lokakuuta 2023. [Verkossa]. Saatavissa: <https://help.instagram.com>
- [21] F. M. Mehlhose, M. Petrifke, ja C. Lindemann, ”Evaluation of Graph-based Algorithms for Guessing User Recommendations of the Social Network Instagram”, teoksessa *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, 2021, ss. 409–414. doi: 10.1109/ICSC50631.2021.00075.
- [22] F. Amato, V. Moscato, A. Picariello, ja G. Sperlí, ”Recommendation in Social Media Networks”, teoksessa *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, 2017, ss. 213–216. doi: 10.1109/BigMM.2017.55.
- [23] A. Mosseri, ”Instagram Ranking Explained”, Instagram. Viitattu: 5. lokakuuta 2023. [Verkossa]. Saatavissa: <https://about.instagram.com/blog/announcements/instagram-ranking-explained/>
- [24] I. Medvedev, ”Powered by AI: Instagram’s Explore recommender system”, Meta AI blog. Viitattu: 26. tammikuuta 2024. [Verkossa]. Saatavissa: <https://ai.meta.com/blog/powered-by-ai-instagrams-explore-recommender-system/>
- [25] I. R. Hendrawan, E. Utami, ja A. D. Hartanto, ”Comparison of Word2vec and Doc2vec Methods for Text Classification of Product Reviews”, teoksessa *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2022, ss. 530–534. doi: 10.1109/ICITISEE57756.2022.10057702.
- [26] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, ja H. Arshad, ”State-of-the-art in artificial neural network applications: A survey”, *Heliyon*, vsk. 4, nro 11, s. e00938, 2018, doi: <https://doi.org/10.1016/j.heliyon.2018.e00938>.
- [27] A. Koutsoukas, K. J. Monaghan, X. Li, ja J. Huan, ”Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data”, *J. CHEMINFORMATICS*, vsk. 9, kesäkuu 2017, doi: 10.1186/s13321-017-0226-y.
- [28] Y. Sari, M. Maulida, E. Gunawan, ja J. Wahyudi, ”Artificial Intelligence Approach For BAZNAS Website Using K-Nearest Neighbor (KNN)”, teoksessa *2021 Sixth International Conference on Informatics and Computing (ICIC)*, 2021, ss. 1–4. doi: 10.1109/ICIC54025.2021.9632954.

- [29] O. Baker ja Q. Yuan, "Machine learning: Factorization Machines and Normalized Discounted Cumulative Gain for Tourism Recommender System Optimisation", teoksessa *2021 IEEE International Conference on Computing (ICOCO)*, 2021, ss. 31–36. doi: 10.1109/ICOCO53166.2021.9673502.
- [30] K. M. Mishra ja K. J. Huhtala, "Fault Detection of Elevator Systems Using Multilayer Perceptron Neural Network", teoksessa *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFa)*, 2019, ss. 904–909. doi: 10.1109/ETFa.2019.8869230.
- [31] S. Huang, W. Peng, J. Li, ja D. Lee, "Sentiment and topic analysis on social media: a multi-task multi-label classification approach", teoksessa *Proceedings of the 5th Annual ACM Web Science Conference*, teoksessa WebSci '13. New York, NY, USA: Association for Computing Machinery, 2013, ss. 172–181. doi: 10.1145/2464464.2464512.
- [32] M. Zhao, Y. Zhang, L. Qiao, ja D. Sun, "A Deep Neural Networks Based on Multi-task Learning and Its Application", teoksessa *2021 40th Chinese Control Conference (CCC)*, 2021, ss. 6201–6206. doi: 10.23919/CCC52363.2021.9549842.
- [33] Y. Huang, W. Wang, L. Wang, ja T. Tan, "Multi-task deep neural network for multi-label learning", teoksessa *2013 IEEE International Conference on Image Processing*, 2013, ss. 2897–2900. doi: 10.1109/ICIP.2013.6738596.
- [34] Z. Yan, W. Liu, S. Wen, ja Y. Yang, "Multi-Label Image Classification by Feature Attention Network", *IEEE Access*, vsk. 7, ss. 98005–98013, 2019, doi: 10.1109/ACCESS.2019.2929512.
- [35] E. D ja N. P. G. Bhavani, "An Effective DNN Based ResNet Approach for Satellite Image Classification", teoksessa *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, 2023, ss. 1055–1062. doi: 10.1109/ICOSEC58147.2023.10276330.
- [36] D. McNeely-White, J. R. Beveridge, ja B. A. Draper, "Inception and ResNet features are (almost) equivalent", *Cogn. Syst. Res.*, vsk. 59, ss. 312–318, 2020, doi: <https://doi.org/10.1016/j.cogsys.2019.10.004>.
- [37] M. Marcuzzo, A. Zangari, A. Albarelli, ja A. Gasparetto, "Recommendation Systems: An Insight Into Current Development and Future Research Challenges", *IEEE Access*, vsk. 10, ss. 86578–86623, 2022, doi: 10.1109/ACCESS.2022.3194536.
- [38] J. Möller, D. Trilling, N. Helberger, ja B. van Es, "Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content

diversity”, *Inf. Commun. Soc.*, vsk. 21, nro 7, ss. 959–977, 2018, doi:
10.1080/1369118X.2018.1444076.