

Opinnäytetyö (AMK)

TURUN AMMATTIKORKEAKOULU

Tietojenkäsittely

2020

Marianne Alanen

TEKOÄLYN JA ALGORITMIEN HYÖDYNTÄMINEN INSTAGRAMISSA

Marianne Alanen

TEKOÄLYN JA ALGORITMIEN HYÖDYNTÄMINEN INSTAGRAMISSA

Opinnäytetyössä selvitetään millä tavoin Instagram hyötyy DeepText-tekoälystä sovelluksen omilla alustoilla ja millä tavoilla se on vaikuttanut sovelluksen kehittämiseen ja käyttäjiin. Instagram on kasvanut näiden kuluneiden vuosien aikana ja noussut yhdeksi maailmanlaajuisesti suosituksi sosiaalisen median kanavaksi. Sovelluksen jatkuva kasvu on tarjonnut myös yrityksille uusia mahdollisuuksia tavoittaa kohderyhmänsä erilaisten markkinointimenetelmien avulla.

Lähtökohtana työssä analysoidaan Instagramin työkalujen ja eri toimintojen avulla millaisia mahdollisuuksia ne tarjoavat käyttäjille ja yrityksille. Erityisesti keskitytään Instagramin algoritmiin ja siihen vaikuttaviin tekijöihin sekä DeepText-tekoälyn ja sen käyttötarkoitukseen.

Instagram-syötteen muodostumiseen vaikuttaviin tekijöihin ja millä tavoin Instagram kerää tarvittavaa dataa sovelluksensa kehittämiseen. Opinnäytetyössä hyödynnettiin aiheisiin liittyviä raportteja, artikkeleita, videoita, tutkimuksia ja tilastoja. Tarkoituksena oli analysoida ja kartoittaa Instagramin ja DeepTextin kehityksen taustat, syyt ja seuraukset, sekä sovelluksen tilanne tulevaisuudessa.

Tutkimuksista selvisi, miten DeepTextin tehokkaasta tekstinkäsittelystä ja -suodatuksesta huolimatta, sen kykenemättömyys jäljitellä ihmisen empatiakykyä aiheuttaa ristiriitaa sensuroinnin suhteen. Sen lisäksi kävi ilmi, miten tekstipohjaisten hakukoneiden sijaan päädyttiin käyttämään *Automatic Image Annotation*-tekniikkaa kuvien indeksoinnin helpottamiseksi. Instagramin suosion todetaan kasvavan parin seuraavan vuoden aikana, mutta ei kuitenkaan tule vielä hetkeen ylittämään Facebookin käyttäjämäärää.

ASIASANAT:

Instagram, sosiaalinen media, algoritmi, tekoäly, DeepText

BACHELOR'S / MASTER'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Bachelor of Information Technology

2020 | 33 pages

Marianne Alanen

UTILIZATION OF ARTIFICIAL INTELLIGENCE AND ALGORITHMS IN INSTAGRAM

In this thesis we examine how Instagram benefits from DeepText artificial intelligence on the application's own platforms and in what ways it has affected the application's development and its users. Instagram has grown over the past few years and has become one of the most popular social media channels worldwide. The continuous growth of the application has also provided companies with new opportunities to reach their target group through various marketing methods.

The starting point is to analyze the opportunities offered to users and companies with the help of Instagram tools and various functions. In particular, the focus is on the Instagram algorithm and the factors that affect it, as well as DeepText artificial intelligence and its purpose.

Factors influencing the formation of the Instagram feed and how Instagram collects the necessary data to develop its application. The thesis utilized related reports, articles, videos, research and statistics. The purpose was to analyze and map the backgrounds, causes and consequences of the development of Instagram and DeepText, as well as the future situation of the application.

Studies have shown how, despite DeepText's efficient word processing and filtering, its inability to mimic a person's ability to empathize causes a contradiction in terms of censorship. In addition, it was revealed how Automatic Image Annotation technology was used instead of text-based search engines to facilitate the indexing of images. The popularity of Instagram is set to grow over the next couple of years, but it won't surpass Facebook's user base for a while yet.

KEYWORDS:

Instagram, social media, algorithm, artificial intelligence, DeepText

SISÄLTÖ

1 JOHDANTO	1
2 SOSIAALINEN MEDIA	2
2.1 Instagram	2
2.2 Instagram algoritmi	5
3 DEEPTEXT-TEKOÄLY	8
3.1 Laajennettu älykkyys	9
3.2 Koneoppiminen	10
3.3 Syväoppiminen	11
3.4 DeepText Instagramissa	11
3.5 DeepTextin hyödyt ja haitat	12
4 INSTAGRAMIN BIG DATA JA DATA-ANALYYSI	14
4.1 Tiedon louhinta	14
4.2 Tagien kartoittaminen	15
4.3 Kuvien sisältö ja käyttäjätyypit	16
4.4 Tykkäysten poisto	19
4.5 Instagramin tulevaisuus	21
5 POHDINTA	23
LÄHTEET	25

KUVAT

Kuva 1. Suosituimmat sosiaaliset verkostot (miljoona käyttäjää) (Statista, 2020)	3
Kuva 2. Maailmanlaajuinen Instagram käyttäjien määrä vuosina 2016-2023 (miljoona käyttäjää) (Statista, 2020).	4
Kuva 3. Miten Instagram-algoritmi ajattelee, kun julkaiset jotain (Ward, 2020).	7
Kuva 4. Laajennettu älykkyys - ihmisen ja tekoälyn vahvuudet (Siukkonen & Neittaanmäki, 2019).	9
Kuva 5. Koneoppimisen prosessi (Siukkonen & Neittaanmäki, 2019).	10

Kuva 6. Massadatan viisi V:tä (Iyengar, 2018)	14
Kuva 7. Tiedonlouhinta hyödyntää koneoppimisen, tietotekniikan ja tilastotieteen menetelmiä (Siukkonen & Neittaanmäki, 2019).	15
Kuva 8. Euklidisen etäisyyden kaava (Wikipedia, 2013),	18
Kuva 9. Käyttäjien keskimääräinen viettämä aika sosiaalisissa verkostoissa. (Product Lead, 2019).	22

TAULUKOT

Taulukko 1. Teinien suosimat sosiaalisen verkostot USA:ssa (Statista, 2020).	4
--	---

1 JOHDANTO

Tämä opinnäytetyö on muodoltaan tutkimuksellinen työ, jonka tavoitteena on tarkastella Instagramin algoritmin ominaisuuksia ja toimintaa. Sen lisäksi selvitetään, miten Instagram-sovellus hyödyntää Facebookin kehittämää DeepText -tekoälyä omilla alustoillaan ja mitä hyötyjä ja haittoja se on vuosien mittaan tuonut esille. Aihe on kiintoisa siitä näkökulmasta, miten sosiaaliset mediat hyötyvät näistä ominaisuuksista.

Lähtökohtana tässä työssä tutustuttiin Instagramiin, mitä varten se kehitettiin ja mitä sillä tehdään. Tutkimuksessa keskitytään nimenomaan DeepTextiin, joka aluksi kehitettiin Facebookin käyttäjäkokemuksen parantamiseksi ja myöhemmin tuotiin myös Instagramiin. Pää tavoitteena on kartoittaa, miten DeepText on vaikuttanut Instagramiin ja sen käyttäjiin, mitä hyviä ja huonoja puolia se on nostanut esille ja miltä sovelluksen tulevaisuus näyttää.

Sen lisäksi syvennyttään Instagramin algoritmiin, jonka pohjalta muodostuu Instagram-syöte, jossa Instagramiin julkaistut kuvat näkyvät heti, kun sovellus avataan. Työssä selvitetään, millä perusteilla algoritmi valitsee kuvia ja videoita esitettäväksi sovelluksen etusivulle. Lisäkysymyksinä ovat seuraavat: Mistä eri tyypeistä ja ryhmistä nämä kuvat ja videot koostuvat ja mitä ne kertovat käyttäjistä? Miten tagit auttavat Instagramiin ladattun sisällön kartoittamisessa ja datan keruussa?

Tapaustutkimuksen kohteeksi on valittu Instagram, koska vaikka se on suosituin kuvien jakamissovellus, se ja erityisesti DeepText ovat herättäneet suhteellisen vähän huomiota tutkimusyhteisössä. Instagram on siinä nähden hieman uudempi sovellus, jonka kasvava käyttäjien määrä houkuttelee aina vain enemmän ja enemmän yrityksiä mainostamaan tuotteitaan ja palveluitaan. Työssä selvitetään, mistä johtuu sovelluksen suosion ja kehityksen eksponentiaalinen kasvu.

2 SOSIAALINEN MEDIA

Sosiaalinen media on verkon välityksellä tapahtuvaa julkista jakamista, tiedonrakentelua ja keskustelua. Se hyödyntää tietoverkkoja ja tietotekniikkaa viestinnän muotona, jossa käsitellään vuorovaikutteisesti ja käyttäjälähtöisesti tuotettua sisältöä ja luodaan ja ylläpidetään ihmisten välisiä suhteita (Opetushallitus, 2004). Sosiaalinen media on vakiintunut ilmaisu, jolla viitataan verkkopalveluihin, kuten Facebookiin, YouTubeen, Twitteriin, Instagramiin, WhatsAppiin, blogeihin ja wikeihin. (Rongas & Honkonen, 2016).

Monet yksilöt käyttävät sosiaalista mediaa ystävien ja läheisten kanssa yhteydenpitoon. Jotkut ihmiset käyttävät erilaisia sosiaalisen median sovelluksia verkottamaan uramahdollisuuksiaan, etsimään ihmisiä ympäri maailmaa, joilla on samanlaisia kiinnostuksen kohteita, ja jakamaan ajatuksiaan, tunteitaan, oivalluksiaan ja tunteitaan. Näihin aktiviteetteihin osallistuvat ovat osa virtuaalista sosiaalista verkostoa (Dollarhide, 2020).

Yrityksille sosiaalinen media on välttämätön työkalu. Ne käyttävät alustaa asiakkaiden löytämiseen ja kanssakäymiseen, myynnin edistämiseen mainonnan ja myynninedistämisen avulla, kuluttajien trendien mittaamiseen ja asiakaspalvelun tai tuen tarjoamiseen (Dollarhide, 2020).

2.1 Instagram

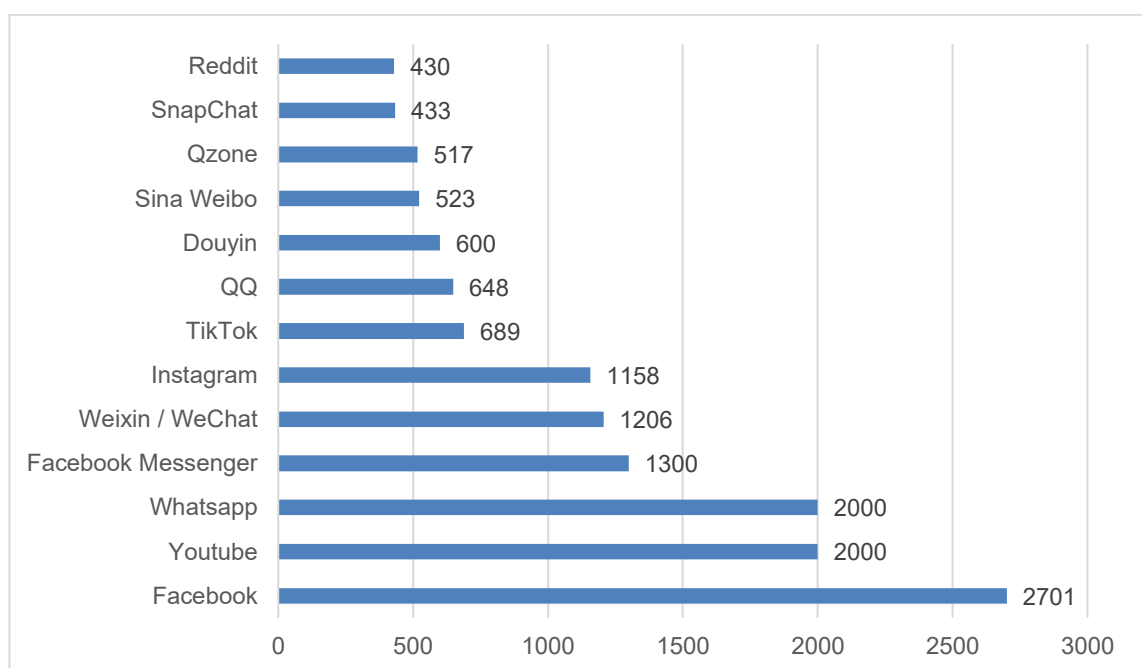
Instagram on ilmainen kuvienjako- ja yhteisöpalvelu. Sen avulla käyttäjät voivat jakaa kuvia ja videoita sekä kommentoida, tykätä ja seurata toistensa jakamaa sisältöä. Palvelu tarjoaa luovan, hauskan ja yksinkertaisen tavan kuvata, muokata ja jakaa kuvia, videoita ja viestejä joko yksityisesti tai julkisesti (Instagram, 2020).

Instagram julkaistiin lokakuun 6. päivä vuonna 2010. Heti julkaisupäivänä noin 25 000 käyttäjää oli ottanut sovelluksen käyttöönsä. Viikon kuluttua Instagramin julkaisusta sovellus oli ladattu 100 000 kertaa. Joulukuun puolivälissä käyttäjien määrä oli noussut jo miljoonaan. Elokuussa vuonna 2011 sovellukseen oli ladattu 150 miljoonaa kuvaa. 10 miljoonan käyttäjän raja ylittyi syyskuussa vuonna 2011 (Shaun and Kyra, 2020).

Instagramin nouseva suosio houkutteli sosiaalisen median jättejä, kuten Twitteriä ja Facebookia. Twitter tarjosi Instagramin luojalle Kevin Systromille 500 miljoonaa ostaakseen

Instagramin, mutta Systrom kieltäytyi. Huhtikuussa vuonna 2012 Facebook osti Instagramin noin miljardilla dollarilla (Shaun and Kyra, 2020).

Facebook oli ensimmäinen sosiaalinen verkosto, joka ylitti miljardin rekisteröidyn tilin, ja sillä on tällä hetkellä yli 2,7 miljardia aktiivista käyttäjää kuukaudessa. Yhtiöllä on tällä hetkellä myös neljä suurinta sosiaalisen median alustaa, joilla kaikilla on yli miljardi kuukausittaista aktiivista käyttäjää: Facebook (ydinalusta), WhatsApp, Facebook Messenger ja Instagram (Kuva 1). (Statista, 2020).



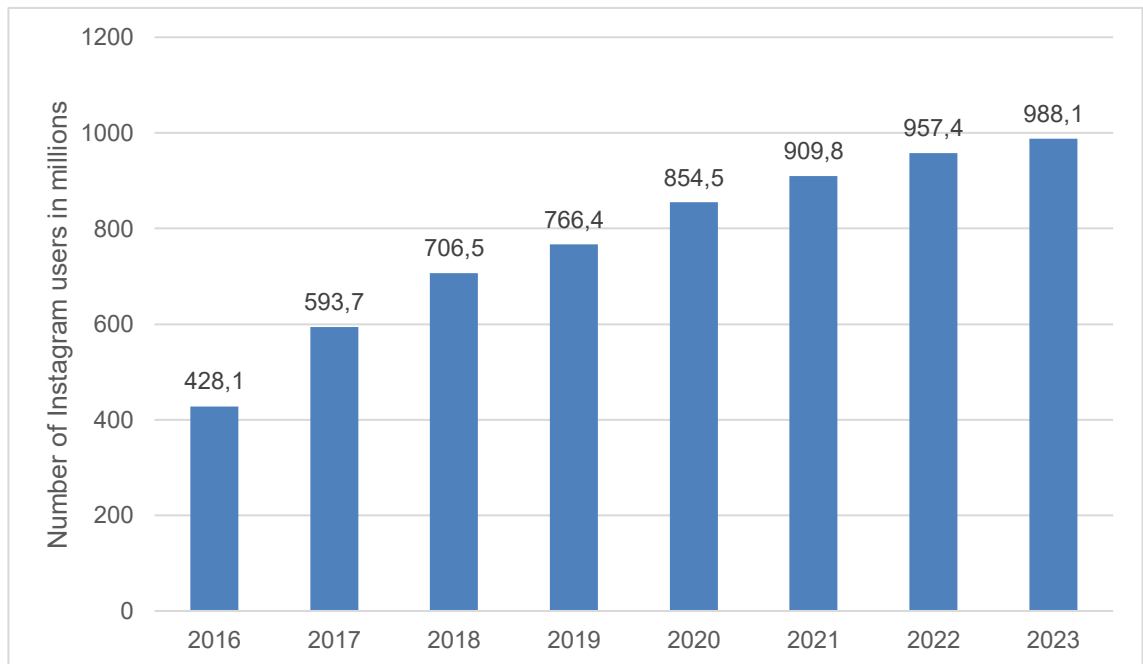
Kuva 1. Suosituimmat sosiaaliset verkostot (miljoonaa käyttäjää) (Statista, 2020)

Taulukko 1 osoittaa yhdysvaltalaisen teini-ikäisten ensisijaiset sosiaaliset verkostot. Syksyn 2019 kyselyn aikana todettiin, että Snapchat on tärkein sosiaalinen verkosto 44 %:lle Yhdysvaltain teini-ikäisistä. Nopeasti kasvava sosiaalinen videosovellus TikTok sijoittui kolmanneksi. Neljä prosenttia puolestaan ilmoitti TikTokin olevan suosikkiverkosto (Statista, 2020).

Taulukko 1. Teinien suosimat sosiaalisen verkostot USA:ssa (Statista, 2020).

	Fall 2012	Spring 2013	Fall 2013	Spring 2014	Spring 2015	Fall 2015	Spring 2016	Fall 2016	Spring 2017	Fall 2017	Spring 2018	Fall 2018	Spring 2019	Fall 2019
Snapchat	-	-	-	-	11 %	17 %	24 %	35 %	39 %	47 %	45 %	46 %	41 %	44 %
Instagram	12 %	17 %	27 %	30 %	29 %	29 %	23 %	24 %	23 %	24 %	26 %	32 %	35 %	35 %
TikTok	-	-	-	-	-	-	-	-	-	-	-	-	0 %	4 %
Twitter	27 %	30 %	31 %	27 %	21 %	18 %	16 %	13 %	11 %	7 %	9 %	6 %	6 %	3 %
Facebook	42 %	33 %	27 %	23 %	12 %	13 %	15 %	13 %	11 %	9 %	8 %	6 %	6 %	3 %
Pinterest	2 %	2 %	2 %	2 %	2 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %	1 %	3 %
Tumblr	3 %	4 %	5 %	5 %	4 %	3 %	2 %	-	-	-	-	-	-	-
Google+	6 %	5 %	4 %	4 %	1 %	1 %	1 %	1 %	-	-	-	-	-	-
Other	2 %	4 %	2 %	4 %	8 %	8 %	5 %	-	-	-	-	-	-	-

Kuva 2 näyttää Instagramin käyttäjien määrän kasvun tilastot ja ennusteet vuosina 2016-2023. Vuoden 2021 ennusteen mukaan käyttäjien määrä tulee ylittämään 900 miljoonan käyttäjän rajan. (Statista, 2020)



Kuva 2. Maailmanlaajuinen Instagram käyttäjien määrä vuosina 2016-2023 (miljoona käyttäjää) (Statista, 2020).

2.2 Instagram algoritmi

Algoritmi on rajallinen sarja hyvin määriteltyjä, tietokoneen toteuttamia ohjeita, tyypillisesti ongelmaluokan ratkaisemiseksi tai laskennan suorittamiseksi. Algoritmit ovat aina yksiselitteisiä, ja niitä käytetään eritelminä laskelmien, tietojenkäsittelyn, automaattisen päättelyn ja niiden tehtävien suorittamiseen. J.G Brookshan määrittelee algoritmin koostuvan järjestyksessä olevista, yksiselitteisistä toiminnoista, jotka voidaan suorittaa ja jotka määrittelevät lopputulokseen johtavan prosessin (Wikipedia, 2020).

Instagram algoritmin tavoitteena on maksimoida Instagramin kassavirtaa. Koska Instagram on sen käyttäjille ilmainen, maksu otetaan mainoksista. Instagramin jatkuvasti kasvava suosio houkuttelee yrityksiä mainostamaan palveluitaan ja tuotteitaan, mikä tarkoittaa suurempia mainostuloja Instagramille (Rantanen, 2019).

TechCrunchin mukaan Instagramin algoritmi hyödyntää käyttäjän aiempaan käyttäytymiseen perustuvaa koneoppimista luodakseen ainutkertaisen syötteen jokaiselle käyttäjälle. Seuraavassa on laadittu kolme tärkeintä tekijää (Leppänen, 2018), jotka määrittelevät Instagram-syötteen sisällön:

Kiinnostus

Aiempien tykkäysten ja seurattujen käyttäjien perusteella Instagram kykenee arvioimaan käyttäjän kiinnostuksen kohteet. Mitä todennäköisimmin Instagram olettaa käyttäjän tykkäävän julkaisusta, sitä korkeammalle se nousee syötteessä. Tämä perustuu aiempaa käyttäytymiseen samantyyppisen sisällön suhteen ja mahdollisesti konenäön analysointiin julkaisun sisällöstä.

Ajankohtaisuus

Algoritmi priorisoi viimeaikaiset julkaisut niiden tuoreuden ja trendikkyuden mukaan. Toiminto kannustaa käyttäjiä julkaisemaan uutta ja laadukasta sisältöä säännöllisesti pysyäkseen näkyvillä seuraajien Instagram-syötteessä.

Yhteys

Jos käyttäjä kommentoi paljon tiettyjen käyttäjien julkaisuihin tai tämä on merkitty heidän kuviinsa, viestittää tämä Instagramille, että kyseiset käyttäjät ovat todennäköisesti läheisiä keskenään. Tämä tuo käyttäjille enemmän näkyvyyttä seuraajien syötteessä.

Sisällön tuoreuden, ajankohtaisuuden ja näkyvyyden lisäksi algoritmiin vaikuttavat myös seuraavat tekijät:

Käytön tiheys

Sillä kuinka usein käyttäjä avaa Instagramin, on merkitystä, sillä heille pyritään näyttämään parhaat julkaisut sitten viime kirjautumisen. Tämä tarkoittaa käytännössä sitä, että mitä useammin sovellus avataan, sitä kronologisempi syötteestä tulee. Jos käyttäjä selaa Instagramia vain silloin tällöin, määräytyy syötteen järjestys enemmän sen mukaan, mistä Instagram olettaa käyttäjän pitävän.

Seuraaminen

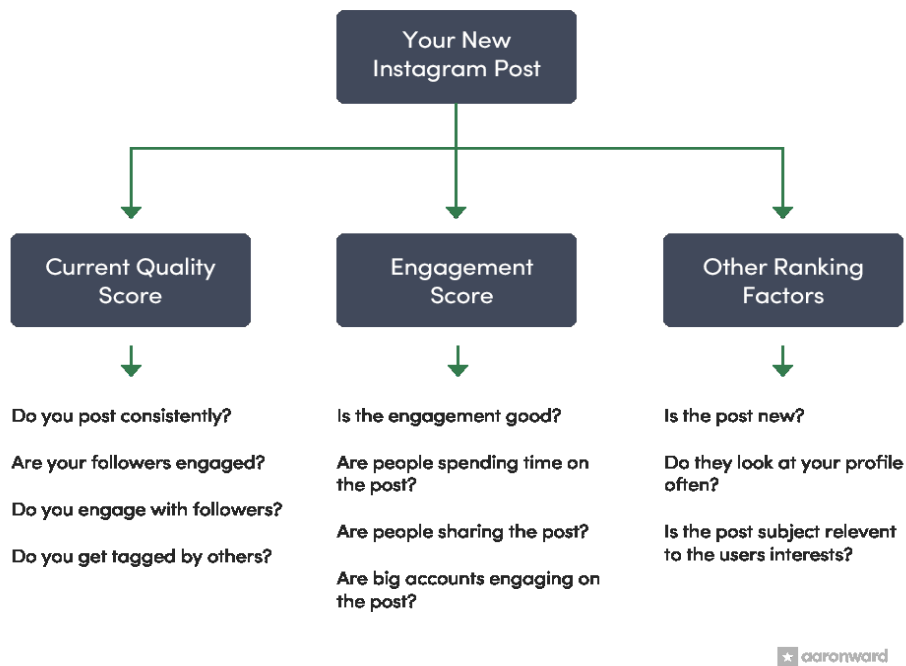
Mitä useampaa tiliä käyttäjä seuraa, sitä epätodennäköistä on, että näkee jokaisen julkaisun kaikilta tileiltä. Instagram pyrkii enemmän näyttämään julkaisuja mahdollisimman laajalta kirjoilta eri tilejä, sen sijaan, että käyttäjä näkisin vain tietyn tilin kaikki julkaisut.

Käyttö

Se, kuinka pitkään käyttäjä selaa Instagramia kerralla määrittää, näkeekö tämä algoritmin esiin nostamat kohokohdat lyhyillä selaussessioilla, vai kaivaako Instagram esiin julkaisuja syvemmältä syötteestä käyttäjän viihtyessä pidempiä aikoja sovelluksen parissa.

Kuvassa 3 esitetty Aaron Wardin esittämä kuva yksinkertaistaa Instagram algoritmin tavalla käsitellä Instagram-viestejä.

How the Instagram algorithm thinks when you post.



Kuva 3. Miten Instagram-algoritmi ajattelee, kun julkaiset jotain (Ward, 2020).

3 DEEPTEXT-TEKOÄLY

Teksti on yleinen viestinnän muoto Facebookissa, kuten monessa muussa sosiaalisen median palvelussa. Käyttäjien kokemusten parantamiseksi ja tekstin erilaisten käyttötapojen ymmärtämiseksi Facebook loi vuonna 2016 DeepTextin, joka on syväoppimiseen perustuva tekstinkäsittelyohjelma. Se pystyy ymmärtämään melkein ihmisen tarkkuudella usean tuhannen viestin sisällön sekunnissa yli 20 kielellä. (Abdulkader;Lakshmiratan;& Joy, 2016).

Tekstin ymmärtäminen vaatii useita tehtäviä, kuten yleisen luokittelun viestin sisällön ja kontekstin määrittämiseksi. Päästäkseen lähemmäs sitä, miten ihmiset ymmärtävät tekstiä, tietokoneelle on opetettava ymmärtämään slangia ja yleistämistä(engl. *word-sense disambiguation*) (Abdulkader;Lakshmiratan;& Joy, 2016).

Tekstin ymmärtäminen edellyttää hankalien skaalausten ja kielihaasteiden ratkaisemista siellä, missä perinteiset NLP-tekniikat (*Neuro-Linguistic Programming*) eivät ole tarpeeksi tehokkaita. Syväoppimisen avulla pystymme ymmärtämään tekstiä paremmin useilla kielillä ja käyttämään merkittyjä tietoja entistä tehokkaammin kuin perinteiset NLP-tekniikat (Abdulkader;Lakshmiratan;& Joy, 2016).

Koska Facebook-yhteisö on globaali, on tärkeää, että DeepText kykenee ymmärtämään mahdollisimman montaa kieltä. Perinteiset NLP-tekniikat edellyttävät laajaa esikäsitteilylogiikkaa, joka perustuu monimutkaiseen tekniikkaan ja kielitaitoon. Jokaisella kielellä on myös muunnelmia, koska ihmiset käyttävät slangia ja erilaisia kiertoilmaisuja saman idean välittämiseen. Syväoppimista käyttämällä voimme vähentää riippuvuutta kielestä riippuvaisesta tiedosta, koska järjestelmä voi oppia tekstistä ilman edeltävää tai vain vähän esikäsitteilyä. Tämä auttaa kattamaan useita kieliä nopeasti pienellä suunnittelutyöllä (Abdulkader;Lakshmiratan;& Joy, 2016).

Perinteinen NLP-tekniikka muuntaa sanat muotoon, jonka tietokonealgoritmi voi oppia, antamalla yksittäisille sanoille kokonaislukutunnuksen (engl. *integer ID*). Esimerkiksi sanat "*brother*" ja "*bro*", jotka molemmat tarkoittavat sanaa "*veli*", saavat aivan erilaiset kokonaislukutunnukset (Abdulkader;Lakshmiratan;& Joy, 2016). Syväoppimistekniikalla voimme puolestaan käyttää sanan upotuksia, matemaattista käsitettä, joka säilyttää semanttisen suhteen sanojen välillä. Joten oikein laskettuna voidaan nähdä, että sanojen

"brother" ja "bro" upotukset vastaavat toisiaan. Tämä mahdollistaa sanojen syvemmän semanttisen merkityksen ymmärtämisen (Abdulkader;Lakshmiratan;& Joy, 2016).

Sanojen upotuksia käyttämällä voidaan myös ymmärtää saman semantiikan useilla kielillä huolimatta pintamuodon eroista. Esimerkiksi englanniksi ja espanjaksi "happy birthday" ja "feliz cumpleaños" tulisi olla hyvin lähellä toisiaan yhteisessä upotustilassa. Karitoittamalla sanat ja lauseet yhteiseksi upotustilaksi DeepText pystyy rakentamaan kielidiagnostiikkamalleja (Abdulkader;Lakshmiratan;& Joy, 2016).

3.1 Laajennettu älykkyys

Tekoälyllä ei ole tarkkaa yleistä määritelmää, van määritelmä riippuu kontekstista. Tekoälyn käsite viittaa ohjelmiin ja laitteisiin, jotka kykenevät oppimaan ja jotka auttavat ihmistä niissä toiminnoissa, joita varten ne on suunniteltu. Kun taas tekoäly on luotu toimimaan ja reagoimaan ihmisten tavoin, laajennettu älykkyys (engl. *augmented intelligence*) käyttää samoja koneita parantaakseen ihmisen työtä. (Kuva 4.)

Ihmisen vahvuuksia	Tekoälyn vahvuuksia
<ul style="list-style-type: none">• Terve järki• Mielikuvitus• Moraali• Abstrahointi• Myötätunto• Yleistäminen	<ul style="list-style-type: none">• Luonnollinen kieli• Koneoppiminen• Hahmontunnistus• Harhojen vähentäminen• Tiedon löytäminen• Toistaminen• Johdonmukaisuus

Kuva 4. Laajennettu älykkyys - ihmisen ja tekoälyn vahvuudet (Siukkonen & Neittaanmäki, 2019).

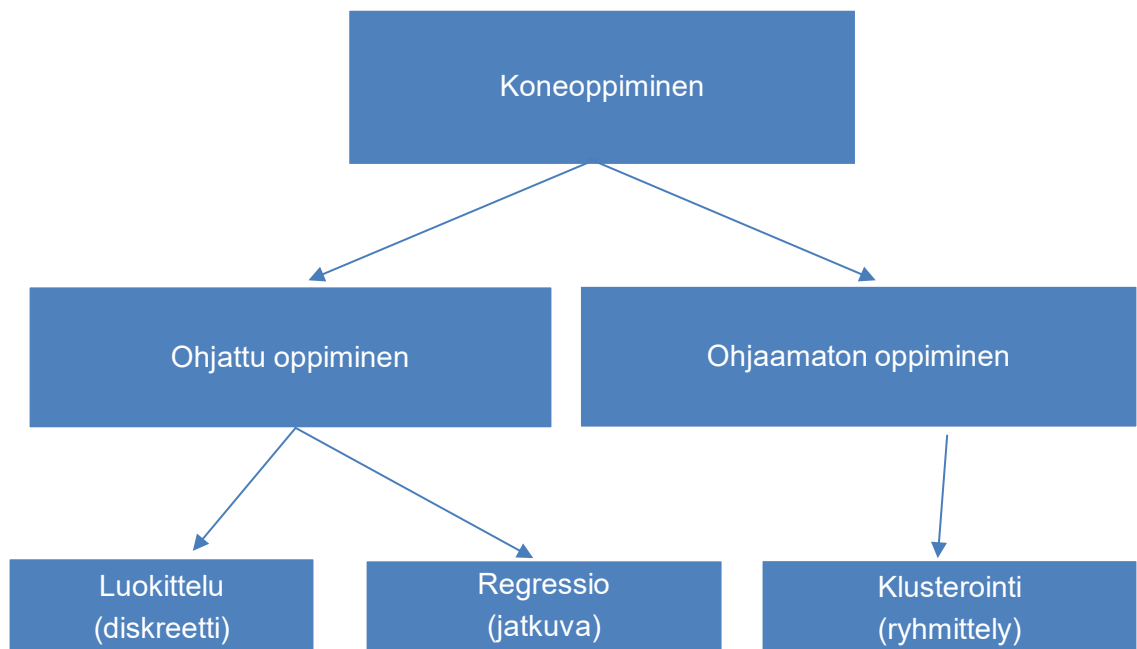
Laajennettu älykkyys vaatii sekä ihmisen, että koneen yhteistyötä, joissa molempien osapuolten vahvuudet auttavat saavuttamaan suuremman liiketoiminnallisen arvon. Toisin

sanoen laajennetun älykkyyden ensisijainen tehtävä on auttaa ihmistä toimimaan paremmin ja älykkäämmin (Roberts, 2020).

3.2 Koneoppiminen

Koneoppiminen (engl. *machine learning*) on tekoälyn osa-alue, jonka tarkoituksena on saada ohjelmisto toimimaan entistä paremmin pohjatiedon ja mahdollisen käyttäjän toiminnan perusteella. Koneoppimistilanteessa kone oppii toistuvista tapahtumista ilman, että ihminen erikseen opettaa sitä. Koneoppimisella pyritään automatisoimaan tiedon tulkintaa ja laajentamaan koneen havainnointikykyä monimutkaisten algoritmien avulla perinteisen raja-arvoihin tukeutuvan mallin sijasta.

Koneoppiminen voidaan jakaa kahteen eri kategoriaan: ohjattuun oppimiseen ja ohjaamattomaan oppimiseen (Kuva 5). Ohjatulla oppimisella tarkoitetaan tilannetta, jossa koneelle opetetaan erilaisia tilanteita, joita koneen tulee tarkkailla ja huomioida. Ohjaamaton oppiminen on tilanne, jossa kone itse löytää tilanteita ja ehdottaa niiden lisäämistä tarkkailuun (Siukkonen & Neittaanmäki, 2019).



Kuva 5. Koneoppimisen prosessi (Siukkonen & Neittaanmäki, 2019).

3.3 Syväoppiminen

Facebookin mukaan, DeepText kykenee lukemaan ja ymmärtämään tuhansia viestejä sekunnissa syväoppimiskyvyn ansiosta. Syväoppiminen on koneoppimisen haara, joka perustuu neuroverkkojen kerroksiin (Gaudin, 2016). Syväoppiminen (engl. *deep learning*) on tekniikka, jossa syväoppiminen ja neuroverkot matkivat ohjelmallisesti ihmisen aivojen toimintaa. Ilmiöiden opettaminen tietokoneelle neuroverkkojen avulla vaatii suuria määriä opetusdataa, sillä syvät neuroverkot voivat sisältää miljoonia säädettäviä parametrejä (Siukkonen & Neittaanmäki, 2019).

3.4 DeepText Instagramissa

Seuraavassa on koottu lista (The Science Times, 2019) Instagramin toiminnoista, jotka hyödyntävät DeepText -tekoälyä.

Roskapostisuodatin

Roskapostin välttämiseksi, roskapostisuodatin pystyy havaitsemaan vääränlaiset ja epäilyttävät viestit ja poistamaan ne yhdeksällä eri kielellä kirjoitetulta Instagram-tililtä, kuten englantia, venäjää, arabia ja kiina. Tekoälyn havaitessa väärennetyt viestit ne poistetaan pysyvästi. Viestien lisäksi Instagram kykenee erottamaan myös tilit, joilla on väärennetyjä seuraajia.

Ennakkokäyttäjäkokemus (UX)

Instagramin sisäisen sisällön määrän kasvaessa, tekoäly hyödyntää koneoppialgoritmia tietojen lajitteluun muokatakseen syötteitä enemmän käyttäjäkohtaisiksi. Instagram-syöte pyrkii näyttämään postaukset siinä järjestyksessä, joka on käyttäjien mieleen. Koneoppialgoritmi oppii, mikä on relevanttia ja arvostettua käyttäjien keskuudessa, mikä helpottaa tietojen lajittelussa sopivan syötteen luomiseksi.

Kohdemarkkinointi

Arvioimalla käyttäjien hakumäärytykset ja sitoutumistiedot, Instagram voi myydä mainoksia yrityksille, jotka haluavat tavoittaa kyseisen asiakasprofiilin ja jotka saattavat olla eniten kiinnostuneita tietyn markkinointiviestin vastaanottamisesta.

Facebookilla on 1,8 miljardia käyttäjää, ja se omistaa Instagramin. Yrityksellä on tehokas analyysitietoverkko, joka auttaa kohdentamaan mainontaa sen perusteella, mistä ihmiset pitävät ja ketä he seuraavat. (Bernard Marr & Co., 2020)

Sivutoiminnon tutkiminen

Tagien ja trendien hyödyllisen käytön avulla Instagramin käyttäjät voivat helposti etsiä kuvia tietyistä tapahtumista, paikoista ja (verkko)kaupoista ympäri maailma, jotka ovat valtakunnallisesti trendikkäitä. Hakuominaisuutta tukevat tagit ovat erittäin hyödyllinen ominaisuus Instagramin käyttäjille ajankohtaisesti kiinnostavien asioiden etsimisessä miljoonien postattujen kuvien joukosta.

3.5 DeepTextin hyödyt ja haitat

DeepText -järjestelmä kannustaa käyttäjiä vapaaseen itseilmaisuun ja torjuu loukkaavat kommentit. Tekstilukitteluohjelman tarkoitus on auttaa koneita tulkitsemaan sanoja ja niiden kontekstia roskapostin estämiseksi.

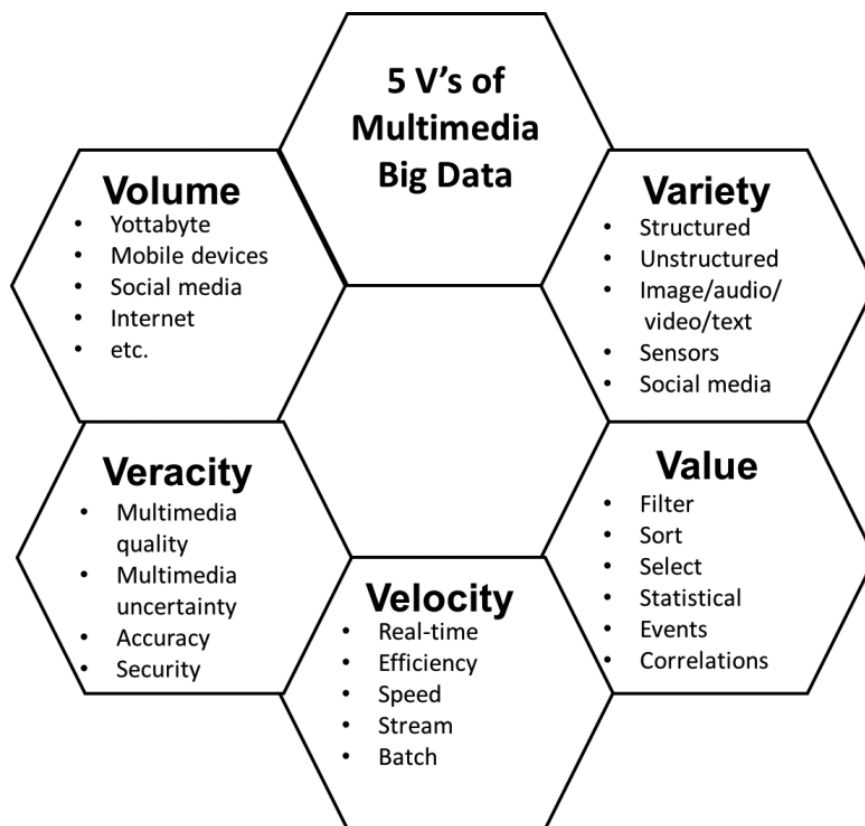
Siitä huolimatta, tällä algoritmisella sensuurilla on kuitenkin sama ongelma, kuin kaikilla stereotyyppioilla ja profiloinnilla. Ilman asiayhteyttä monia harmittomia sisältöjä tuomitaan epäoikeudenmukaisesti. Monet sanat ja lauseet välittävät eri merkityksen eri puolilla maailmaa. Erilaisia slangeja saatetaan pitää jossakin maassa vihapuheena. Jotkin kappaleet käyttävät loukkaavaa kieltä, ja joidenkin sanojen vakavuus riippuu myös käyttäjien välisistä suhteista. Algoritmia on melkein mahdoton ohjelmoida ymmärtämään vastaavanlaisia vaihtelevia konteksteja (Singh, 2017).

Instagramin mukaan uusi järjestelmä on valmis sietämään tietyn virhemarginaalin. Tällä asenteella on kaikenlaisia ongelmia, kuten sananvapaus. Facebook on joutunut aiemmin kärsimään muutamasta kiistasta tällä alalla (Singh, 2017).

Instagram pyrkii luomaan turvallisen yhteisön uuden sensuurijärjestelmän avulla. Siihen liittyy kuitenkin omat ongelmansa. Instagramin on todettu olevan jo paljon ystävällisempi kuin useimmat muut Internetin sosiaaliset verkostot. Tämä tekniikka saattaa pian vuotaa Facebookiin ja muihin sosiaalisiin medioihin (Singh, 2017).

4 INSTAGRAMIN BIG DATA JA DATA-ANALYYSI

Big Datalla eli massadatalla tarkoitetaan massiivisten, jatkuvasti kasvavien, strukturoitua ja ei-strukturoitua tietoa, kuvia, äänitteitä ja videoita sisältävien tietojoukkojen keräämistä, säilyttämistä ja tiedon käyttämistä. Näin valtavien datamäärien hallitseminen ja analysoiminen perinteisellä tietokantatyökalulla olisi käytännössä mahdotonta. Big Datalle tyypillisiä tunnusmerkkejä ovat viisi V:tä: määrä (volume), valikoima (variety), nopeus (velocity), arvo (value) ja todenmukaisuus (veracity) (Siukkonen & Neittaanmäki, 2019). (Kuva 6.)

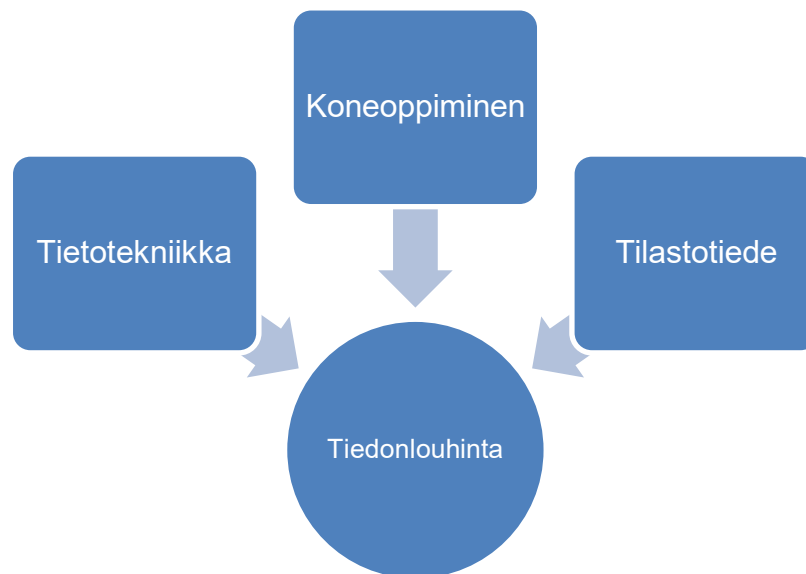


Kuva 6. Massadatan viisi V:tä (Iyengar, 2018)

4.1 Tiedon louhinta

Tiedon louhinta (engl. data mining) tarkoittaa joukkoa menetelmiä, joilla pyritään oleellisen tiedon löytämiseen suurista tietomassoista. Toimintaan liittyy tietomassojen keräämistä, säilyttämistä, jakamista, etsimistä, analysointia sekä esittämistä tilastotiedettä,

tietotekniikkaa ja koneoppimisen menetelmiä hyödyntäen (Siukkonen & Neittaanmäki, 2019).



Kuva 7 Tiedonlouhinta hyödyntää koneoppimisen, tietotekniikan ja tilastotieteen menetelmiä (Siukkonen & Neittaanmäki, 2019).

4.2 Tagien kartoittaminen

Twitterin helpompi saavutettavuus tekee siitä yhden tutkituimmista sosiaalisen median alustoista, sillä sen käyttöön liittyviä prosesseja (twiittaus, hashtagit, uudelleentwiittaus) voidaan käyttää kerätyn datan seuraamiseen ja analysointiin Twitter-sovellusliittymän kautta (Twitter API), eriävistä aiheista, kielestä tai kontekstista huolimatta (Highfield & Leaver, 2015).

Vaikka Instagram ja Twitter ovat sisällöiltään hyvin erilaiset, niiden välillä on joitain rakenteellisia yhtäläisyyksiä, kuten hashtagit. Instagramin ”tägäämisen” käsitteellinen ja käytännön merkitys ovat rinnastettavissa Twitterin hashtageihin. Molempia käytetään tunnisteina esitettyjen pääaiheiden, ideoiden, tapahtumien, paikkojen ja tunteiden merkitsemiseksi sekä kuvissa, että twiiteissa (Highfield & Leaver, 2015).

Verkkoan ladattujen kuvien hakeminen on erittäin haastavaa paitsi tehokkuuden (oikean kuvan hakeminen käyttäjän tarpeiden / kyselyjen mukaan) ja suorituskyvyn lisäksi myös nähtävyyden (paikannettavuuden) kannalta. Nykyaikaiset hakukoneet hakevat kuvia tekstipohjaisesti, koska suurin osa loppukäyttäjistä tuntee tekstipohjaiset kyselyt verkkosivujen ja digitaalisten asiakirjojen hakemiseksi. Tekstipohjaisten kuvien hakemiseksi

kuvien on oltava jotenkin sidoksissa tiettyihin avainsanoihin tai tekstikuvaan. Tällainen tekstikuvaus saadaan yleensä verkkosivuilta tai asiakirjasta, joka sisältää HTML-vaihtoehtoisen tekstin, kuvien tiedostonimet, metadatatunnisteet ja ympäröivän tekstin (Giannoulakis & Tsapatsoulis, 2016).

Siitä huolimatta sosiaalisen median kuvia ei kuitenkaan voida indeksoida (poimia asi-aankuuluvaa tekstikuva) puhtailla verkkopohjaisilla tekniikoilla, lähinnä siksi, että sosiaalisen median käyttäjän sivut eivät noudata klassista verkkosivurakennetta. Tämän seurauksena tunnettu sisältöpohjainen kuvanhakukenttä elpyi ja syntyi AIA (engl. *Automatic Image Annotation*). AIA viittaa matalan tason ominaisuuksien purkamiseen kuvasta ja yhden tai useamman semanttisen käsitteen määrittämiseen siihen (Giannoulakis & Tsapatsoulis, 2016).

Suuri osa AIA:ta hyödyntää koneoppimistekniikoita, jonka juuret ovat oppimisen paradigmassa. Monia erilaisia malleja ja koneoppimistekniikoita kehitettiin ns. 'visuaalisten mallien' rakentamiseen, jotka sieppaavat kuvaominaisuuksien ja tekstisanojen välisen korrelaation. Visuaalisiin malleihin syötetään sitten näkymättömistä kuvista poimittuja kuvaominaisuuksia niiden tunnistamisen ennustamiseksi. Olettaen, että hyviä visuaalisia malleja voidaan saavuttaa, kuvien noutaminen tarjoaa lupaavan vaihtoehdon tekstipohjaisille menetelmille (Giannoulakis & Tsapatsoulis, 2016).

4.3 Kuvien sisältö ja käyttäjätyypit

Tässä otsikossa esittelen löytämäni tutkimuksen (Hu;Manikonda;& Kambhampati, What We Instagram: A First Analysis of Instagram Photo Content and User Types, 2014), jossa on tutkittu Instagramiin ladattujen kuvien sisältöä ja kartoitetaan sen käyttäjien arkkityypit. Tutkimus hyödyntää kvantitatiivista ja kvalitatiivista tutkimusmenetelmiä, jossa on käytetty konenäkötekniikoita kuvien sisällön tutkimiseen. Erityyppiset aktiiviset käyttäjät tunnistetaan Instagramissa klusteroinnin avulla. Tutkimuksen tarkoituksena on saavuttaa parempi ymmärrys sosiaalisesta, kulttuurisesta ja ympäristöllisestä ihmisten toimintaa koskeviin kysymyksiin. Tutkimuksen tulokset paljastavat seuraavat asiat:

- kahdeksan suosituinta kuvakategoriaa
- viisi erilaista Instagram-käyttäjätyyppiä
- käyttäjien yleisö on riippumaton jaetusta mediasta riippumatta.

Tutkimuksessa käytetty data on kerätty Instagram API:n avulla, joka sisältää käyttäjien profiilitietoja, kuvia, kuvatekstejä, tageja ja käyttäjien sosiaaliset verkostot sisältäen ystävät ja seuraajat.

Tutkimuksen metodiikka

Tutkimuksen metodiikassa kerrotaan, miten satunnaisesti valituilta Instagram-käyttäjiltä poimitut otokset saadaan heidän lähettämien kuvien ja videoiden ilmestyessä julkiseen aikajanaan. Tutkimuksen kohteiksi valittiin 37 käyttäjää, joista suurin osa olivat julkkiksia. Sen jälkeen näiden valittujen käyttäjien ystävien ja seuraajien käyttäjät indeksoitiin ja yhdistettiin, jonka tuloksena syntyi 95 343 uniikkia seed-käyttäjää. Tämän pohjalta tutkijat rakensivat satunnaisen otoksen säännöllisesti aktiivisista Instagram-käyttäjistä.

Säännölliset käyttäjät määriteltiin seuraavasti:

- eivät ole organisaatioita, tuotemerkkejä tai roskapostittajia
- käyttäjällä on vähintään 30 ystävää, 30 seuraajaa ja on lähettänyt vähintään 60 kuvaa

Kaiken kaikkiaan 13 951 käyttäjää täyttivät nämä kriteerit, joiden joukosta satunnaisesti valittiin 50 käyttäjää.

Tutkimusprosessi

Tuhannen kuvan joukosta valittiin 200 kuvaotosta tematisointiin ja luokitteluun. Sisältökategorioiden luominen osoittautui hankalaksi, johtuen kuvien olevan sisällöltään ja ominaisuuksiltaan rikkaampia kuin teksti. Tämän vuoksi konenäkötekniikoita käytettiin helpottamiseksi. Tutkimuksessa käytettiin *Scale Invariant Feature Transform* (lyhyesti *SIFT*) algoritmia paikallisten erottuvien ominaisuuksien havaitsemiseen ja näytteiden poimimiseen valokuvista. Kuvien ominaisvektorit koostuvat 128 ulottuvuudesta. Standardia kuvavektorikvantisointimenetelmää noudattamalla saatiin ulos koodikirjavektorit jokaiselle kuvalle. Lopuksi käytettiin *k-means* klusterointia 15 kuvaryhmän luomiseksi, jossa kuvien välinen samankaltaisuus lasketaan Euklidinen etäisyys niiden koodikirjavektorien välillä.

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

Kuva 8. Euklidisen etäisyyden kaava (Wikipedia, 2013),

Automatisoidun luokittelun parantamiseksi käytettiin kahta ihmiskoodaria analysoimaan 15 kuvaryhmän kuvia. He analysoivat luokkien sisäisiä ja välisiä teemoja ja tarvittaessa mukauttivat manuaalisesti kuvien luokituksia sopivimmiksi.

Lopputuloksena oli luotu lista seuraavanlaisista kuvakategorioista:

- ystävät
- ruoka
- laitteet
- kuvatekstit
- lemmikit
- aktiviteetit
- omakuvat / selfiet
- muoti.

Tutkimuksen tulokset

Tutkimuksen päätavoitteena oli kartoittaa ja ymmärtää paremmin valokuvien ja aktiivisten käyttäjien eri tyypit Instagramissa. Pyrkimyksenä oli käsitellä seuraavat tutkimuskysymykset:

- Millaisia valokuvia ihmiset yleensä julkaisevat Instagramiin?
- Kuinka käyttäjät eroavat lähettämistään kuvatyypeistä?
- Kuinka nämä erot käyttäjien valokuvien välillä vaikuttaa seuraajamäärään?

Tutkimus osoittaa, että lähes puolet tietojoukkoihin tallennetuista kuvista ovat omakuvia (selfieitä). *Lemmikit* ja *Muoti* olivat vähiten suosittuja luokkia, joiden osuus valituista kuvista oli vain viisi prosenttia. Muiden luokkien, kuten *Ruoka*, *Laitteet* ja *Kuvatekstit* osuudet olivat yli 10 % yksittäin. Tämä osoittaa Instagramin olevan enimmäkseen käyttäjien itsensä mainostamiseen ja sosiaaliseen verkostoitumiseen ystävien kanssa.

Vastaus kysymykseen eroavatko käyttäjien lähettämät kuvatyypit keskenään saatiin luomalla 8-ulotteinen vektori jokaiselle käyttäjälle (vastaten jokaista kahdeksaa kuvakategoriaa), jossa jokainen ulottuvuus edustaa käyttäjien valokuvien osuutta vastaavassa luokassa. Tuloksena saatiin erotettua viisi erilaista käyttäjätyyppiä, jotka selvästi erottuvat ominaisuuksiltaan toisistaan jakamiensa kuvien suhteen. On C4 eli ”selfien ystäviä”, jotka julkaisevat yksinomaan omakuvia, toinen käyttäjätyyppi C2, lähettää enimmäkseen kuvatekstejä, joiden upotetuissa teksteissä mainitaan mottoja, runoja, lainauksia ja jopa suosittuja hashtageja. Toisaalta on olemassa käyttäjiä, kuten C1, jotka keskittyvät lähinnä ruokakuvien lähettämiseen, mutta lähettävät myös muunlaisia kuvia. Tämän vuoksi C1 -ryhmän entropia osoittautui korkeimmaksi. C3 ryhmä julkaisee enimmäkseen kuvia liittyen erilaisiin aktiviteetteihin ja laitteisiin. Lopuksi C5-luokan käyttäjät lähettävät yhtä paljon kuvia ystävistään kuin kuvia itsestään.

Vastaus kolmanteen kysymykseen liittyen kuvatyypin vaikutuksesta seuraajamääriin, omakuvia enimmäkseen julkaisevilla käyttäjillä on huomattavasti enemmän käyttäjiä. Siitä huolimatta tutkimuksessa todetaan, että seuraajien määrä on riippumaton käyttäjäryhmistä. Analyysi ei osoita mitään tilastollista merkitsevyyttä ”seuraajien lukumäärä - käyttäjätyypit” korrelaatioita, voidaan päätellä, että käyttäjän yleisön koko on riippumaton käyttäjän tyypistä.

4.4 Tykkäysten poisto

Tällä hetkellä Instagramilla on yli miljardi aktiivista käyttäjää, jotka kirjautuvat tililleen päivittäin. Alustan ollessa erittäin kilpailukykyinen, Instagramin luojat ovat suunnitelleet poistavan joidenkin käyttäjien tykkäykset vähentääkseen vaikuttajien välistä kilpailua. Käyttäjät itse voivat silti nähdä viestiensä tykkäysten määrän, mutta tämä tieto ei ole kenenkään muun nähtävissä. Kokeilun yhteydessä käyttäjille haluttiin viesti, että halutaan seuraajien keskittyvän jakamiseen eikä siihen, miten monta tykkäystä postaus saa.

”Haluamme, että seuraajat keskittyvät siihen, mitä jaat, eikä siihen, kuinka monta tykkäystä postauksistasi saa” (Berman, 2020).

Syy

Instagram perustelee syyn tykkäysten poistamiselle olevan se, että alusta on tulossa liian kilpailukykyiseksi ja ihmiset keskittyvät enemmän tykkäyksiin kuin laadukkaan ja kiinnostavan sisällön julkaisemiseen muiden käyttäjien välisen kommunikaation kannustamiseksi. On havaittu, että monissa tapauksissa käyttäjillä on tapana poistaa viestejä, jotka eivät saa tarpeeksi tykkäyksiä. Monien sosiaalisen median vaikutusta ihmisiin koskevan tutkimusten avulla, tutkijat ovat löytäneet monien käyttäjien epäterveellisen pakkomielteen tykkäyksille, joka on melko haitallista henkiselle hyvinvoinnille. Instagram pyrkii vähentämään sosiaalista ahdistusta ja kilpailua tuomalla uuden ominaisuuden käyttöön ja poistamaan tämän osan vähentääkseen kilpailun kielteiset vaikutukset (Berman, 2020).

Seuraus

Ajatus tykkäysten poistamisesta sai aikaan ristiriitaisen vastaanoton. Tähän asti tykkäykset ja seuraajat ovat olleet tärkein prioriteetti monille Instagram vaikuttajille, koska nämä olivat tärkeimpiä mittareita, jotka auttoivat heitä saamaan näkyvyyttä alustalla ja olivat hyviä vertaamaan suosiota (Berman, 2020).

Vaikka monet olisivatkin mielissään saadessaan rauhallisemman kokemuksen Instagramissa, monet vaikuttajat ovat myös herättäneet huolta, koska Instagram on yksi heidän ensisijaisista tulovirroistaan. Heidän koko liiketoimintamallinsa on kerätä lisää tykkäyksiä ja seuraajia varmistaakseen, että eri tuotemerkit sponsorivat heitä markkinoitintarpeisiinsa (Berman, 2020).

Tämä ominaisuus on tarkoitus ottaa käyttöön valinnaisilla alueilla, kuten Kanadassa, ja on melko epävarmaa, että he ottavat sen käyttöön kaikkialla maailmassa. Jos käyttäjät pysyvät myönteisinä muutoksen, Instagram voisi pian päästä eroon tykkäyksistä toivossa luoda terveellinen ympäristö kaikille käyttäjille (Berman, 2020).

4.5 Instagramin tulevaisuus

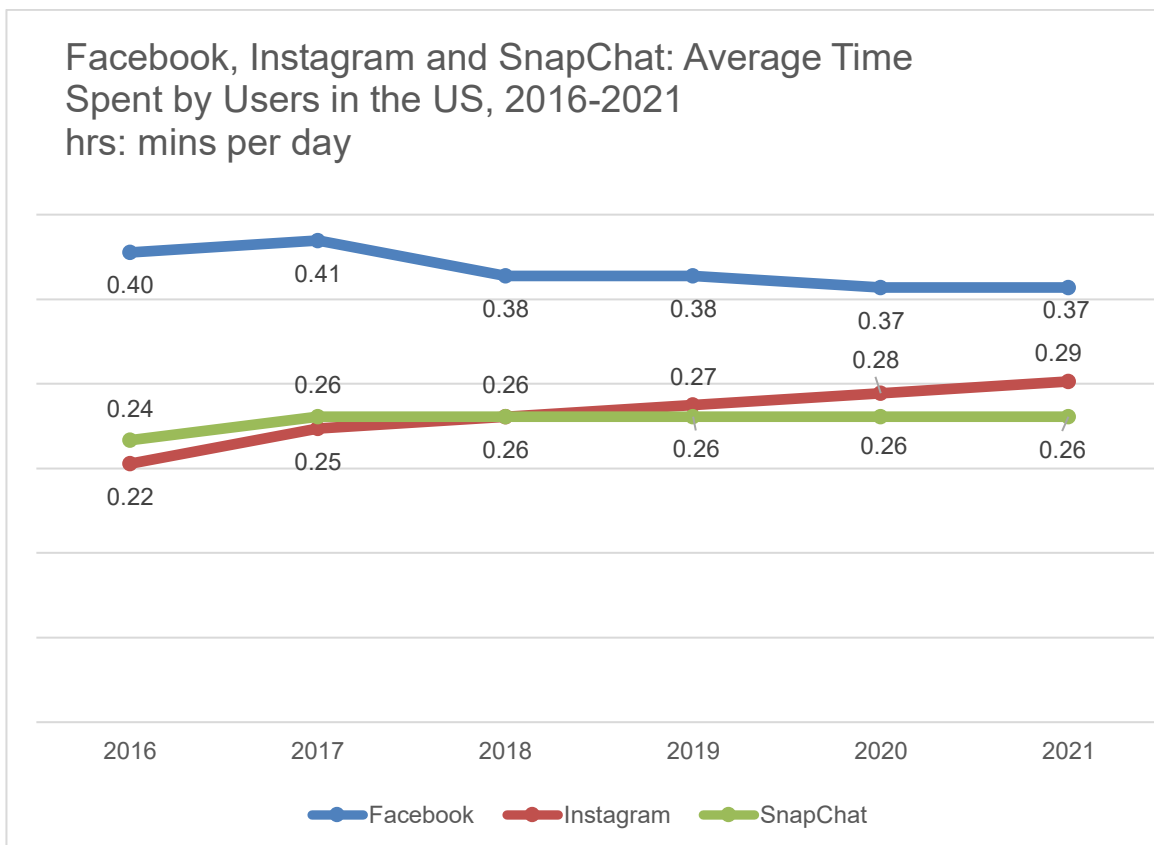
Liiketoiminnan näkökulmasta Instagram-yhteisöt ovat kultakaivos. Koskaan aikaisemmin tuotemerkit eivät olleet voineet olla yhteydessä kohdeyleisöönsä visionäärisellä tasolla tähän asti. (Product Lead, 2019).

Instagram on tuonut yrityksille valikoiman mainoskanavia: Short-term Stories, Swipe Up -ominaisuus, ei-häiritsevät mainospaikat ja kokonaan uudentyyppinen mobiili video – IGTV (Product Lead, 2019).

Instagram on julkaissut dataraportteja, jotka sijaitsevat uuden Insight-ominaisuudessa. Tämä ominaisuus on ollut suureksi avuksi sisällöntuottajille, jotka harjoittavat yhteisön kasvua ja brändeille, jotka työskentelevät mainosten kanssa (Product Lead, 2019).

Instagramin visuaalisesti houkuttelevasta ulkomuodosta huolimatta, tosiasiallisesti tuloja tuottavat tilisegmentit ja yritykset ovat edelleen kiinnostuneita toiminnan seurannasta. Tästä syystä Instagramin on työskenneltävä yrityksen sisäisen infrastruktuurin parissa ennen vahvemman laajennussuunnitelman käynnistämistä (Product Lead, 2019).

Vuoden 2020 ennusteet keskittyvät sovelluksen työkalujen kehittämiseen. Pidempiä ajanjaksoja datan keräämiseksi, datahistoriaa, tulkintoja ym. Vaikka Instagramilla olisi parantamisen varaa analysoinnin suhteen, muut alustat onnistuvat täyttämään tämän aukon. Esimerkiksi sivustot, kuten ProductLead Analytics, joka tarjoaa kuukausittain raportteja mm. kilpailijoiden seurannasta, käyttäjien luomasta sisällöstä ja brändien lähettiläiden löytämisestä (Product Lead, 2019). Kuvassa 9 voidaan nähdä miten Instagramin keskimääräinen käyttöaika on ylittänyt SnapChatin ja jatkaa kasvuaan vuoden 2021 puolella.



Kuva 9. Käyttäjien keskimääräinen viettämä aika sosiaalisissa verkostoissa. (Product Lead, 2019).

5 POHDINTA

Tavoitteena oli tarkastella DeepText-tekoälyn ja Instagram-algoritmin toimintaa ja miten Instagram on pyrkinyt niiden avulla parantamaan sovelluksen käyttäjäkokemusta ja vähentämään haitallisia tekijöitä, kuten roskapostia ja ikäviä kommentteja. Tässä opinnäytetyössä syvennyttiin Instagramista aiemmin tehtyihin tutkimuksiin, joiden keskiössä toimi Instagram API. Valitettavasti Instagramin *Legacy*-sovellusliittymän käyttöoikeus poistettiin 29.6.2020, mikä estää kolmannen osapuolen sovellusten pääsyn Instagram-syötteisiin. Työn tekijällä ei myöskään ollut käyttöoikeuksia *Instagram Basic Display API*:in, minkä vuoksi dataa ei kyetty tarkastelemaan.

Löydetyt tutkimukset perustuivat lähes yksinomaan Instagram-algoritmiin ja siihen pohjautuvaan dataan. Lopuksi päädyttiin kartoittamaan tutkimuksista Instagramin kuvakategorioiden ja käyttäjätyyppien vaikutusta seuraajien määrään. Instagram-sovellus turvautuu muihin palveluihin tarvittavan tärkeän datan keräämiseksi. Myös käyttäjät voivat hyötyä näistä tutkimuksista, sillä ne kuvaavat perusteellisesti, miten Instagram algoritmi oikein toimii ja minkä tyyppiset kuvat ovat eniten suosiossa.

Tutkimusten perusteella Instagram osoittaa kasvavan tasaisesti ainakin parin seuraavan vuoden aikana sen käyttäjämäärään nähden. Kuvista on huomattavasti hankalampi kerätä dataa tekstipohjaisilla hakukoneilla. Sen sijaan suurin osa datasta saadaan kartoittamalla kuviin liitetyistä tageista, jotka tuovat ilmi kuviin liittyvän kontekstin, paikan, tunteet ym. tekstin muodossa. Tekoälyn ansiosta Instagram-algoritmi kykenee luomaan käyttäjilleen visuaalisesti houkuttelevan ja mielekkään ulkoasun ja syötteen sovelluksen käyttäjien käyttäytymismallien perusteella.

Selvisi myös, mitkä ovat Instagramin yleisimmät kuvakategoriat ja käyttäjätyypit ja miten ne vaikuttavat seuraajien määrään. Vaikka omakuvia julkaisevat käyttäjät saavat tutkittu eniten seuraajia ja tykkäyksiä, seuraajien määrä ei ole riippuvainen julkaistujen kuvien sisältöön.

Tämä työ osoittaa koneoppimisen mahdollisuuden auttaa meitä saamaan uusia oivalluksia tutkittaessa ihmisiä sekä sosiaalisia, taloudellisia ja kulttuurillisia tekijöitä ympäri maailmaa. Tutkimukset osoittavat massadatan ja tekniikoiden, kuten konenäön, automatisoidun analyysialgoritmien ja koneoppimisen todellisen potentiaalin käsitellä valtavia määriä tietojoukkoja, jotka sosiaalisen median sivustot kuovat ymmärtääkseen ympäröivää

maailmaa. Alustojensa parantamiseksi ja väärennetyn tai loukkaavan sisällön löytämiseksi ja poistamiseksi Instagram hyödyntää kerättyä dataa kehittyäkseen. Toiset voivat puolestaan löytää lukuisia mahdollisuuksia näin valtavasta määrästä tietoa ihmisten käyttämisen ja kulttuurin tutkimiseen.

LÄHTEET

- Abdulkader, A.; Lakshmiratan, A.; & Joy, Z. (1. kesäkuuta 2016). *Introducing DeepText: Facebook's text understanding engine*. (Facebook) Haettu 9. marraskuuta 2020 osoitteesta Facebook Engineering: <https://engineering.fb.com/2016/06/01/core-data/introducing-deeptext-facebook-s-text-understanding-engine/>
- Berman, M. (27. huhtikuuta 2020). *Will Instagram Remove the Like Option in 2020?* Haettu osoitteesta Programming Insider: <https://programminginsider.com/will-instagram-remove-the-like-option-in-2020/>
- Bernard Marr & Co. (2020). *The Amazing Ways Instagram Uses Big Data And Artificial Intelligence*. (Forbes) Haettu 8. Marraskuu 2020 osoitteesta Bernard Marr & Co.: <https://bernardmarr.com/default.asp?contentID=1396>
- Dollarhide, M. E. (syyskuuta 2020). *Social media Definition*. Haettu 8. marraskuu 2020 osoitteesta Investopedia: <https://www.investopedia.com/terms/s/social-media.asp>
- Gaudin, S. (2. kesäkuuta 2016). *Facebook's DeepText A.I. can read, understand your posts*. (Computer World) Haettu 9. marraskuuta 2020 osoitteesta Computer World -sivusto: <https://www.computerworld.com/article/3078519/facebooks-deeptext-ai-can-read-understand-your-posts.html>
- Giannoulakis, S.; & Tsapatsoulis, N. (joulukuu 2016). *Evaluating the descriptive power of Instagram hashtags*. Haettu osoitteesta Science Direct: <https://www.sciencedirect.com/science/article/pii/S2352664516300141>
- Highfield, T.; & Leaver, T. (2015). *A Methodology for Mapping Instagram*. Haettu osoitteesta ePrints: https://eprints.qut.edu.au/79684/3/__staffhome.qut.edu.au_staffgroup%24_hollambc_Desktop_A%20Methodology%20for%20Mapping%20Instagram%20Hashtags_submission.pdf
- Hu, Y.; Manikonda, L.; & Kambhampati, S. (2014). *What We Instagram: A First Analysis of Instagram Photo Content and User Types*. Haettu osoitteesta <http://149.169.27.83/instagram-icwsm.pdf>

- Hu, Y.; Manikonda, L.; & Kambhampati, S. (2014). *What We Instagram: A First Analysis of Instagram Photo Content and User Types*. Haettu osoitteesta Semantic Scholar: <https://www.semanticscholar.org/paper/What-We-Instagram%3A-A-First-Analysis-of-Instagram-Hu-Manikonda/8a92831841307b513ca9c1b2c65499b08eb4c156>
- Instagram. (2020). *Instagram*. Haettu 8. marraskuuta 2020 osoitteesta Instagram: <https://www.instagram.com>
- Iyengar, S. (2018). *Multimedia Big Data Analytics: A Survey*. Haettu osoitteesta https://www.researchgate.net/publication/322373798_Multimedia_Big_Data_Analytics_A_Survey
- Leppänen, J. (8. kesäkuuta 2018). *Instagramin uusi algoritmi – miten se toimii?* (Meltwater) Haettu 11. Marraskuu 2020 osoitteesta Meltwater blogi: <https://www.meltwater.com/fi/blog/instagramin-uusi-algoritmi-miten-se-toimii>
- Oberlo. (2020). *Most Popular Social Media Platforms In 2020*. (Oberlo) Haettu 14. Marraskuu 2020 osoitteesta Statistics: <https://www.oberlo.com/statistics/most-popular-social-media-platforms>
- Opetushallitus. (2004). *Laatua E-oppimateriaaleihin*. Haettu 8. Marraskuu 2020 osoitteesta Opetushallitus: https://www.oph.fi/sites/default/files/documents/144415_laatua_e-oppimateriaaleihin_2.pdf#page=25
- Product Lead. (2019). *Instagram Predictions for 2020 – How the IG Future Looks Like*. Haettu 16. Marraskuu 2020 osoitteesta <https://productlead.me/blog/instagram-predictions-ig-future/>
- Rantanen, S. (30. heinäkuuta 2019). *Miten Instagram-algoritmi toimii? (Instagram algoritmi 2019)*. (Emine) Haettu 11. Marraskuu 2020 osoitteesta Eminen palvelusivusto: <https://emine.fi/miten-instagram-algoritmi-toimii-2019/>
- Roberts, K. (2020). *What is Augmented Intelligence and why should you know about it?* Haettu osoitteesta AuraQuantic: <https://www.auraquantic.com/what-is-augmented-intelligence/>

- Rongas, A.;& Honkonen, K. (2016). *Mikä ihmeen sosiaalinen media?* (Opetushallitus) Haettu 8. marraskuuta 2020 osoitteesta Opetushallituksen sivusto: <https://www.oph.fi/fi/koulutus-ja-tutkinnot/mika-ihmeen-sosiaalinen-media>
- Rouse, M. (huhtikuuta 2020). *Definition: social media.* (TechTarget) Haettu 8. Marraskuu 2020 osoitteesta WhatIs sivusto: <https://whatIs.techtarget.com/definition/social-media>
- Shaun and Kyra. (31. toukokuuta 2020). *What is Instagram? History of Instagram social media platform.* Haettu 8. marraskuuta 2020 osoitteesta Youtube: https://www.youtube.com/watch?v=SvFFyyVjdRw&ab_channel=ShaunandKyra
- Singh, A. (30. kesäkuuta 2017). *Instagram Now Using AI to Moderate Comments. Going Too Far?* Haettu 16. marraskuuta 2020 osoitteesta Intellect Digest: <https://www.intellectdigest.in/instagram-now-using-ai-moderate-comments-going-far-32547/>
- Siukkonen, T.;& Neittaanmäki, P. (2019). *Mitä tulisi tietää tekoälystä.* Jyväskylä: Docendo Oy.
- Statista. (29. lokakuuta 2020). *Global social networks ranked by number of users 2020.* Haettu 15. marraskuuta 2020 osoitteesta Statista: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Statista. (23. heinäkuuta 2020). *Most popular social networks of teenagers in the United States from fall 2012 to fall 2019.* Haettu 16. marraskuuta 2020 osoitteesta Statista: <https://www.statista.com/statistics/250172/social-network-usage-of-us-teens-and-young-adults/>
- Statista. (22. huhtikuuta 2020). *Number of Instagram users worldwide from 2016 to 2023.* Haettu 16. marraskuuta 2020 osoitteesta Statista: <https://www.statista.com/statistics/183585/instagram-number-of-global-users/>
- The Science Times. (29. marraskuuta 2019). *How Instagram Uses Artificial Intelligence in 2019.* Haettu 8. marraskuuta 2020 osoitteesta The Science Times: <https://www.sciencetimes.com/articles/24361/20191129/instagram-uses-artificial-intelligence-2019-works.htm>

Ward, A. (12. lokakuuta 2020). *Exactly How The Instagram Algorithm Works in 2020*.
Haettu 14. marraskuuta 2020 osoitteesta aaronward:
<https://www.aaronward.com/blog/how-the-instagram-algorithm-works>

Wikipedia. (26. maaliskuuta 2013). *Euklidinen metriikka*. Haettu osoitteesta Wikipedia:
https://fi.wikipedia.org/wiki/Euklidinen_metriikka

Wikipedia. (2020). *Algorithm*. Haettu osoitteesta Wikipedia:
<https://en.wikipedia.org/wiki/Algorithm>