Arttu Heikkilä

KOGNITIIVISTEN TEKNOLOGIOIDEN HYÖDYNTÄMINEN MIELENTERVEYDEN HÄIRIÖIDEN HAVAITSEMISESSA SOSIAALISESSA MEDIASSA

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2017

TIIVISTELMÄ

Heikkilä, Arttu
Kognitiivisten teknologioiden hyödyntäminen mielenterveyden häiriöiden havaitsemisessa sosiaalisessa mediassa
Jyväskylä: Jyväskylän yliopisto, 2017, 30 s.
Tietojärjestelmätiede, kandidaatintutkielma
Ohjaaja(t): Kollanus, Sami

Mielenterveyden häiriöiden määrä ja taloudellinen taakka on kasvanut viimeisten vuosikymmenien aikana merkittävästi. Maailmanlaajuisesti masennuksesta kärsii yli 300 miljoonaa ihmistä, ja mielenterveyden häiriöiden kustannukset ovat maailmanlaajuisesti noin 2,5 biljoonaa USA:n dollaria. Sekä määrää että kustannuksia voidaan vähentää ennakoivilla toimenpiteillä. Ihmisille ei ole tyypillistä kiinnittää merkittävää huomiota mielenterveyteensä normaalin terveyden tavalla. Siksi on tärkeä havaita ja rohkaista ihmisiä kliinisiin tutkimuksiin hyvissä ajoin, ennen kuin ongelma etenee pidemmälle. Tässä kirjallisuuskatsauksessa kartoitin sitä, miten sosiaalista mediaa voidaan käyttää ennakoivana alustana. Tämä suoritetaan usein tekoälyteknologioita hyödyntäen analysoimalla ihmisten tekemiä julkaisuja, ja niistä mahdollisia merkkejä tai oireita löytämällä. Aikaisemmissa tutkimuksissa on päästy merkittäviin tuloksiin analysoimalla sosiaalisesta mediasta haettua teksti- tai kuvadataa, tai analysoimalla muuta toimintaa sosiaalisissa verkostoissa. Tämä tapahtuu usein yhdistämällä useita eri muuttujia, liittyen tehtyihin julkaisuihin. Näitä muuttujia ovat esimerkiksi kielen käyttö, julkaisujen tiheys ja ajankohta, saadut kommentit tai julkaistujen kuvien eri ominaisuudet. Tekoälyteknologioita hyödyntäen ihmisten tekemistä julkaisuista voidaan päätellä merkittävän tarkasti, kärsiikö henkilö potentiaalisesti piilevästä mielenterveyden häiriöstä, tai onko hän tulevaisuudessa potentiaalinen uhri mielenterveysongelmille. Tämän kaltaisia analyysin tuloksia voi hyödyntää monella tavalla ohjaamaan ja rohkaisemaan käyttäjää kliinisiin tutkimuksiin.

Asiasanat: Tekoäly, luonnollisen kielen prosessointi, mielenterveyden häiriöt, koneoppiminen, masennus, tekstianalytiikka

ABSTRACT

Heikkilä, Arttu
Using cognitive technologies to detect mental disorders in social media
Jyväskylä: University of Jyväskylä, 2017, 30 p.
Information Systems, Bachelor's Thesis
Supervisor(s): Kollanus, Sami

Mental disorders in general are a great burden to society when considering the amount and costs of different mental disorders. Globally, over 300 million people are affected by depression, and the costs of different mental disorders are estimated to have been around 2.5 trillion US dollars. Both the costs, and the amount of people affected can be lowered with pre-emptive measures. It is not typical for generally healthy people to give great notice to their mental health. That is why it is important to notice these disorders, and encourage people to seek medical assistance, before the problem escalates any further. In this literature review I took a look on how social media can be used as a platform to early screening. This is done by using different artificial intelligence technologies to analyze posts made by people, and search for signs and symptoms. Previous researchers have found great success in this by analyzing text or image data. Researches have also analyzed other activities in social networking sites, such as networking or other interactions between users. All this is usually done by combining different kind of variables from posts. These variables include for example linguistic factors, frequency and timestamp of posts, received comments and different characteristics of a published image. Using machine learning, it is possible to detect depressive behavior or other potential mental disorders, to a significant accuracy. The results of these kind of analysis's can be used in various ways to encourage the user to seek medical assistance.

Keywords: Artificial intelligence, natural language processing, major depressive disorder, mental disorders, text-analysis, machine learning

KUVIOT

TAULUKOT

SISÄLLYS

1 JOHDANTO

Maailmanlaajuisesti arviolta 300 miljoonaa ihmistä kärsii masennuksesta (World Health Organization, 2017). Masennuksen ja muiden vähemmän yleisten mielenterveyden häiriöiden kustannukset ovat merkittävä kulu julkiselle terveydenhuollolle. Esimerkiksi vuonna 2007 Euroopan talousalueen mieliala- sekä ahdistushäiriöiden kustannusten arvioitiin olevan yli 136 miljardia euroa (McDaid & Park, 2011). Maailmanlaajuisesti mielenterveyden häiriöiden aiheuttamien kokonaiskustannuksien arvioitiin olevan vuonna 2010 noin 2,5 biljoonaa USA:n dollaria, ja kasvavan $6,0 biljoonaan vuoteen 2030 mennessä (Bloom ym., 2012). Ennaltaehkäisevien ohjelmien on havaittu vähentävän masennusta (WHO, 2017), joten ennakoinnille on selkeä, ja etenkin kasvava tarve.

Kognitiivisilla teknologioilla, eli tekoälyä hyödyntävillä teknologioilla, on valtava merkitys nykypäivän tietojenkäsittelyssä. Tämän lisäksi ihmisten aktiivisuus sosiaalisen median palveluissa on jatkuvassa kasvussa, sekä uusia palveluita syntyy ja kuolee paljon (Obar & Wildman, 2015). Suurimmat alan toimijat, kuten Facebook, hyödyntävät erilaisia koneoppimisteknologioita suoramainontatarkoituksessa. Ihmisten julkaisuja analysoidaan ja niistä poimitaan mainostamistarkoituksessa kiinnostavaa dataa. Vastaavalla tavalla teksti- tai kuvajulkaisuja sekä muuta sosiaalisissa verkostoissa ja internetissä tapahtuvaa käyttäytymistä voidaan analysoida tekoälyteknologioiden avulla, ja löytää merkkejä käyttäjien mahdollisista mielenterveyden häiriöistä, tai niihin viittaavista oireista (Wongkoblap ym., 2017).

Tämän tutkimuksen tarkoituksena on kartoittaa sitä, miten erilaisia kognitiivisia teknologioita voidaan hyödyntää mielenterveyden häiriöiden havaitsemisessa eri sosiaalisissa medioissa. Tämä tapahtuu pääasiassa teksti- ja kuvaanalytiikan avulla, jossa koneoppivia teknologioita opetetaan tunnistamaan pieniä, mutta tärkeitä merkkejä ihmisten julkaisuista. Näiden merkkien, ja suuren otannan kautta voidaan löytää viitteitä mielenterveysongelmista.

Tutkimus keskittyy pääosin masennuksen havaitsemiseen, johtuen sen yleisyydestä sekä valtaväestössä että aiheeseen liittyvässä kirjallisuudessa. Tutkimuksen tarkoitus ei ole myöskään diagnosoida ketään, vaan löytää merkkejä ja

teknologioita, joilla voidaan saada vihjeitä tai viitteitä käyttäjän tai käyttäjien mahdollisista mielenterveysongelmista.

Sosiaalisia medioita ei ole rajattu, mutta tutkimus keskittyy yleisimpiin medioihin, jotka tarjoavat käytettävää ja julkista dataa. Sosiaalinen media käsitteenä on kuitenkin hankala, ja tässä tutkimuksessa sillä tarkoitetaan medioita, joissa käyttäjien tuottama sisältö ja vuorovaikutus ovat palvelun pääpainona. Käsittelen tätä asiaa tarkemmin luvussa 3.

Ensimmäisenä tutkielmassa, luvussa 2, keskityn tekoälyyn ja koneoppimiseen. Avaan kognitiivisten teknologioiden käsitteitä ja pyrin avaamaan tarkemmin niitä teknologioita, jotka ovat tutkimuksen kannalta oleellisimpia. Käsittelen myös tekoälyn historiaa ja koneoppimisen syntyä, sekä teknologioina että tieteenaloina. Seuraavaksi luvussa kolme kartoitetaan mielenterveyden häiriöitä, sekä tarkastellaan sosiaalisia medioita niiden käytettävyyden puolesta. Neljännessä luvussa olen nostanut esille muutaman esimerkkitutkimuksen, jotka avaavat hyvin tutkimuksen ongelmaa. Tässä luvussa myös määritellään tärkeimmät havainnot useammasta tutkimuksesta, sekä tehdään niille koonti ja tarkastelu suhteuttaen siihen, kuinka niitä voi hyödyntää. Viimeisenä luvussa 5 kootaan löydökset ja havainnot. Pyrin myös löytämään muita mahdollisia lähestymistapoja kuin mitä on jo tehty, sekä tarkastelemaan aiheen potentiaalia kokonaisuudessaan.

2 TEKOÄLY JA KONEOPPIMINEN

Kognitiivisella teknologialla, tai yleisimmin tekoälyllä, tarkoitetaan tietokonetta, tietokoneohjelmaa, järjestelmää, tai algoritmia joka kykenee toimintoihin, jotka voidaan lukea älykkäiksi tai ihmisaivojen toimintaa imitoivaksi. Tekoälyn määritelmä on erittäin avoin. Useimpien määritelmien mukaan nykypäivänä 'älykkäiksi' sovelluksiksi voidaan lukea todella merkittävä osa palveluista ja ohjelmistoista. Tästä johtuen, pyrin keskittymään hieman suppeampaan tekoälyn alaluokkaan, koneoppimiseen.

Koneoppimisella tarkoitetaan tietokoneohjelmaa tai järjestelmää, joka on kykenevä oppimaan sille syötetyn datan perusteella. Tämä tarkoittaa sitä, että järjestelmälle ei tarvitse ohjelmoida sääntöjä jokaista tilannetta varteen, vaan se on rakennettu itse oppimaan ja luomaan ohjelmasääntöjä erilaisille tilanteille. (Domingos, 2012)

2.1 Koneoppimisen ja tekoälyn lyhyt historia

Koneoppiminen, kuten tekoäly kokonaisuudessaankin, on yleisestä käsityksestä poiketen erittäin käytetty elementti nykyteknologiassa. Suuri osa tekoälystä, ja erityisesti koneoppimisesta, on usein järjestelmän, ohjelmiston tai verkkopalvelun pinnalle näkymätöntä teknologiaa, joka saattaa oppia suorittamaan käyttäjän näkökulmasta yksinkertaisiakin asioita. Tällaiset asiat eivät välttämättä ole ohjelmistolle yksinkertaisia, jolloin itsestään oppiva komponentti voi olla tarpeen.

2.1.1 Tekoälyn synty

Tieteenalana koneoppiva tekoäly on lähes yhtä vanha kuin tekoäly itsessään. Määritelmien vaihtelevuudesta ja aihealueiden jatkuvasta kehityksestä johtuen on mahdoton määrittää, kuinka pitkä historia koneoppimisella on. Stuart Russelin (1995) mukaan kuitenkin aikaisin työ, joka hyväksytään yleisesti tekoälyksi, on Warren McCullohin ja Walter Pittsin suunnittelema alkukantaisen neuroverkon simulaatio vuonna 1943. Tämä verkko simuloisi kaukaisesti ihmisaivojen rakennetta keinotekoisilla, binäärisillä neuroneilla. Nämä neuronit siirtyisivät ”päällä”-tilaan tarpeeksi merkittävän osan läheisten neuroneiden stimulaation seurauksena (Russell, 1995).

McCullohin ja Pittsin työ voidaan nähdä tulevien tekoälysovellusten innoittajana (Russell, 1995). 1950-luvun aikana kehitettiin jo ensimmäisiä ohjelmia pelaamaan tammea ja shakkia. Myös ensimmäinen neuroverkkotietokone rakennettiin vuonna 1951 Princetonin yliopistossa (Russell, 1995).

2.1.2 Hyödyntäminen teollisuudessa ja kehitys

Optimistisista tekoälyn alkutaipaleen ennustuksista huolimatta tekoälyn kehitys kohtasi ongelmia seuraavina vuosikymmeninä. Ensimmäinen merkittävä vaikeus syntyikin tekoälyn skaalautuvuudessa. Aikaisemmat tekoälysovellukset eivät niinkään ”tunteneet” käsittelemäänsä aihetta, vaan pärjäsivät syntaktisella manipulaatiolla., eli olivat erittäin sääntöpohjaisesti ohjelmoituja (Russell, 1995).

Ensimmäinen onnistunut kaupallinen tekoälyjärjestelmä *R1* tuli käyttöön 80-luvun alkupuolella *Digital Equipment Corporation*ille. *R1* oli tietokoneohjelma, joka konfiguroi tietokonejärjestelmiä asiakkaan tarpeiden mukaan. (McDermott, 1982). Vuoteen 1986 mennessä, järjestelmä oli tuottanut yritykselle arviolta $40 miljoonan vuosittaiset säästöt. Kokonaisuudessaan tekoälyteknologia teollisuudenalana kasvoi rahassa mitattuna yli tuhat kertaiseksi vuosien 1980 - 1988 aikana. (Russell, 1995)

Nykypäivään mennessä tekoäly, kognitiiviset teknologiat ja koneoppimisen sovellutukset ovat kehittyneet merkittävästi. Näiden kehitykselle merkittävää on myös tietokoneiden kehittyminen, sillä laskentateholla on suuri merkitys tekoälysovellusten käyttömahdollisuuksille. Kognitiivisten teknologioiden levinneisyys on erittäin yleistä, ja sitä voidaan havaita lähes kaikilla teollisuudenaloilla.

2.2 Koneoppiminen

Koneoppiminen käsitteenä on hieman hankala, eikä universaalia määritelmää ole. Koneoppiminen tieteenalana tarkastelee järjestelmiä, jotka oppivat uutta niille annetusta datasta (Domingos, 2012), ja niiden kehittämistä ja käyttöä. Riippuen määritelmästä, koneoppiminen jaetaan joko kahteen tai kolmeen alaluokkaan. Tässä tutkielmassa koneoppiminen jaetaan kahteen pääluokkaan, ohjattuun ja ohjaamattomaan oppimiseen, joiden sisällä luokittelua voidaan jatkaa pidemmälle (Chapelle, Schölkopf & Zien, 2006).

2.2.1 Ohjaamaton oppiminen

Chapellen ja kumppanien (2006) mukaan ohjaamattoman oppimisen (unsupervised learning) tarkoituksena on löytää kiinnostavia rakenteita rakenteettomasta datasta. Tämä tarkoittaa käytännössä sitä, että koneelle annettu data on merkitsemätöntä, eikä siihen sisälly erillisiä luokitteluja. Kone ei opi mitään ohjatusti, eli ei saa palautetta. Kaikki mitä järjestelmä oppii, pohjautuu annettuun dataan. Ohjaamaton oppiminen on kykenevä tunnistamaan datasta muutoksia, ryhmittymiä ja anomalioita. Käytännön sovelluksia ohjaamattomasta oppimisesta voi löytää esimerkiksi kuvien ja videoiden pakkausalgoritmeista, tai esimerkiksi

markkinointitarkoituksessa tehdystä asiakassegmentoinnista. Esimerkkityyppejä ohjaamattomasta oppimisesta ovat esimerkiksi klusterointi (clustering) ja anomaliatunnistus (anomaly detection tai outlier detection).

Kluseteroinnilla tai klusterianalyysilla (cluster analysis) tarkoitetaan aikaisemmin mainittua ryhmittelyä. Aggarwal ja Chandan (2013) määrittää klusteroinnin seuraavalla tavalla: ”Given a set of data points, partition them into a set of groups which are as similar as possible” (Aggarwal & Chandan, 2013, s. 2).

Anomaliatunnistuksella tarkoitetaan ongelmaa, jossa on tarkoitus löytää sarjasta dataa muodostumia, jotka eivät ‘käyttäydy’ oletetulla tavalla. Nämä havainnot siis poikkeavat jostain tietystä odotetusta ’kaavasta’, eli ovat siis anomalioita. Tämänkaltaista koneoppimista voi löytää esimerkiksi erilaisissa kyberturvallisuusratkaisuissa, terveydenhuollossa ja vianhavaitsemisjärjestelmissä. (Chandola, Banerjee & Kumar, 2009) Tämänkaltainen koneoppiminen tulee olemaan myös myöhemmässä vaiheessa merkittävässä roolissa tätä tutkimusta.

2.2.2 Ohjattu oppiminen

Yksinkertaistettuna ohjatun oppimisen (supervised learning) tarkoituksena on ennustaa tulosteiden (output) arvoja tai tuloksia syötteiden (input) perusteella (Hastie, Tibshirani & Friedman, 2009). Ohjatussa oppimisessa annettu esimerkkidata on merkittyä, ja kone opetetaan niiden perusteella tunnistamaan, erottamaan tai luokittelemaan merkitsemätöntä dataa opitun perusteella. Ohjattua oppimistakin voidaan jakaa pidemmälle, tai sen rinnalle voidaan liittää vielä muun tyyppistä oppimista. Tyypillisimpiä esimerkkejä näistä ovat esimerkiksi aktiivinen oppiminen, vahvistettu oppiminen ja puoli-ohjattu oppiminen.

Aktiivisella oppimisella tarkoitetaan oppimista, joissa usein budjetista johtuvan merkityn opetusdatan vähäistä määrää korvataan tarkasti ja spesifisti merkityllä ja ohjeistetulla datalla. ”when the examples to be labeled are selected properly, the data requirements for some problems decrease drastically”. (Schohn & Cohn, 2000)

Vahvistettua oppimista on koneen opetus tavalla, jossa sille annetaan sen toiminnan tai tuloksen pohjalta ’palaute’ (reinforcement signal). Tämä palaute voi olla yksinkertaisimmillaan mikä vaan merkki siitä, oliko valittu toiminta tai tuloste hyvä vai huono. Tätä voidaan laajentaa pidemmälle, kuten esimerkiksi pisteytysjärjestelmiin. (KaelBling, Littman & Moore, 1996).

Puoliohjattu oppiminen (semi-supervised learning) asettuu ohjatun ja ohjaamattoman oppimisen väliin. Se yhdistää molempien tapausten pääpiirteet käyttämällä sekä merkittyä, että merkitsemätöntä harjoitusdataa oppimiseen (Zhu, 2011).

2.3 Luonnollisen kielen prosessointi

”Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things” (Chowdhury, 2003, s. 1). Luonnollisen kielen prosessoinnilla tarkoitetaan siis tapaa tai teknologiaa, jolla tietokone pystyy ymmärtämään luonnollisella kielellä tuotettua tekstiä tai puhetta. Luonnollisella kielellä tarkoitetaan tämänkaltaisissa sovellutuksissa tekstiä tai puhetta, joka on ihmiselle luonnollista, eli tietojenkäsittelytieteellisestä näkökulmasta strukturoimatonta dataa. Luonnollisen kielen prosessointi ei pohjimmiltaan ole koneoppimisteknologia, ja sitä on aikaisemmin tehty käsin sääntöpohjaisesti. Kuitenkin koneoppimisen kehittymisen myötä myös luonnollisen kielen prosessointi kehittyi laajalti soveltamaan koneoppimisen algoritmeja.

Luonnollisen kielen prosessoinnin tehtävät voidaan jakaa myös muutamaan alaluokkaan. Näitä ovat esimerkiksi puheentunnistus (speech recognition), luonnollisen kielen ymmärrys (Natural Language Understanding, NLU), luonnollisen kielen tuottaminen (Natural Language Generation, NLG), tai jokin yhdistelmä näitä. Esimerkiksi ääniohjattu keskustelujärjestelmä tarvitsisi kaikkia edellä mainituista toimiakseen oikein. Tutkimuksen tarkoituksen myötä aion keskittyä pääosin vain luonnollisen kielen ymmärrykseen, NLU:hun.

2.4 Luonnollisen kielen ymmärrys

Tavoitetilassa luonnollista kieltä ymmärtävä tietokoneohjelma ymmärtäisi ihmisen puhetta yhtä hyvin kuin ihminen. NLU:n haaste kuitenkin on kielenymmärryksen laajassa tietotaidossa. Chowdhuryn (2003) mukaan NLU sisältää kolme isompaa ongelmaa: Ajatusprosessin, kielellisen ilmaisun esityksen ja sen tarkoituksen, sekä ongelman maailman tai ympäristön ymmärryksessä (world knowledge). Luonnollisen kielen ymmärrys on kehittynyt paljon vuosikymmenien aikana, ja tämänkaltaisia teknologioita hyödyntäviä sovellutuksia on ollut markkinoilla jo pidemmän aikaa. NLU:ta hyödyntävät esimerkiksi sovellukset tai teknologiat jotka parsivat tai tiivistävät informaatiota pidemmistä teksteistä, tai pyrkivät tunnistamaan semanttisia ja morfologisia tarkoituksia käännöstarkoituksessa. Myös luonnollisen kielen käyttöliittymät ovat esimerkkejä NLU:n hyödyntämisestä.

Useiden lupaavien tutkimustulosten myötä voidaan todeta, että luonnollisen kielen ymmärrys, sekä NLP kokonaisuudessaan on erittäin vahvalla pohjalla nykypäivänä. Viimeisen vuosikymmenen aikana eri tutkijaryhmät ovat panostaneet testidatan saatavuuteen (Chowdhury, 2003), mikä auttaa koneoppimisen tyypillisimmän ongelman ratkaisemisessa; opettamisdatan suuren määrän hankinnassa.

2.5 Kielen rakenne NLP:ssä ja NLU:ssa

Luonnollisen kielen prosessoinnissa tai ymmärryksessä lause, virke, sana tai mikä vain dokumentti puretaan pienempiin osiin analysointia varten. Chowdhury (2003) jakaa luonnollisen kielen analyysin seitsemään (TAULUKKO 1) alueeseen tai tasoon. Luonnollisen kielen prosessoinnin tai ymmärryksen järjestelmä voi hyödyntää kaikkia tai osaa tämän analyysin tasoista.

Foneettinen tai fonologinen taso (Phonetic level).	Tämä taso ottaa kantaa lausumiseen, äänenpainoon ja muuhun vokaaliseen toimintaan. Tällä ei siis ole vaikutusta tekstin prosessoinnissa.
Morfologinen taso (morphological level).	Morfologialla tarkoitetaan yksittäisen sanan 'rakennuspalikoita'. Näitä ovat esimerkiksi sanan 'lemma' eli perusmuoto sekä sijapäätteet ym.
Sanastotaso tai Sanastollinen taso (lexical level).	Tällä tarkoitetaan sanojen sanastollista merkitystä.
Syntaktinen taso (syntactic level).	Lauseiden rakenne, syntaksi ja kielioppi.
Semanttinen taso (Semantic level).	Tämä taso tarkastelee sanojen ja lauseiden tarkoitusta ja merkitystä. Eli mitä esimerkiksi jokin lause tarkoittaa?
Diskurssitaso (Discourse level).	Tämä määrittää tekstin tai dokumentin tyypin pohjautuen sen rakenteeseen ja tyyliin.
Pragmaattinen taso (Pragmatic level).	Tällä tasolla tarkoitetaan dokumentin tai tekstin ulkopuolelta tulevaa tietoa, esimerkiksi metatietoa dokumentista.

TAULUKKO 1 Luonnollisen kielen analyysin tasot (Chowdhury, s.5, 2003.)

3 MIELENTERVEYDEN HÄIRIÖT JA SOSIAALINEN MEDIA

Tutkimuksen tarkoituksena on kartoittaa koneoppimisteknologioiden hyödyntämistä mielenterveysongelmien havaitsemisessa tai ennakoinnissa. Tämä tarkoittaa sitä, että alkuun on pystyttävä paikantamaan merkkejä nimenomaisista mielenterveysongelmista, ennen kuin voidaan miettiä sitä, voiko näitä merkkejä tunnistaa koneen avulla. Tästä asiasta on tehty paljon tutkimusta, ja siitä löytyy useita case-tapauksia. Suurin osa tapauksista ja tutkimuksista on tehty kognitiivisia teknologioita hyödyntäen, mutta osa on voitu tehdä esimerkiksi ihmisten tutkimana tai yksinkertaisia, sääntöpohjaisia ohjelmia käyttäen. Kaikki nämä tutkimukset ovat kuitenkin hyödyllisiä, sillä niistä saadaan hyviä esimerkkejä mahdollisista mielenterveysongelmista.

Tässä luvussa määrittelen sosiaalisen median käsitteenä, ja tarkastelen sitä, mitkä palvelut voidaan lukea sosiaaliseksi mediaksi tai käytettäväksi tässä tutkimuksessa. Toinen tärkeä osa tässä luvussa on mielenterveyden häiriöiden määrittäminen, ja tarkastelu siitä mitkä häiriöistä ovat tutkimuksen kannalta merkityksellisiä tai tärkeitä.

Lisäksi tämä luku sisältää tarkastelun verkkopalveluiden tarjoamista mahdollisuuksista. Jos jonkin sosiaalisen median tarjoamaa dataa halutaan tutkia, etenkin koneoppimisen näkökulmasta, on datan oltava saatavilla. Tämä tarkoittaa esimerkiksi palveluiden tarjoamia 'API:ja' (Application Programming Interface) eli ohjelmointirajapintoja.

3.1 Sosiaalinen media ja mielenterveyden häiriöt käsitteinä

3.1.1 Sosiaalinen media

Sosiaalisen median määritelmät ovat todella laajoja. Obarin ja Wildmanin mukaan (2015) käsitteen määrittelemisen vaikeuden syiksi kaksi tekijää. 'Sosiaaliset mediat' kehittyvät jatkuvasti, niitä kehitetään lisää ja niitä kuolee koko ajan. Erot näiden palveluiden välillä saattavat olla todella suuria, joten määritelmän lukkoon lyöminen on mahdotonta. Toisekseen, 'sosiaaliset mediat' käyttävät paljolti samoja teknologioita ja elementtejä kuin aikaisemmat, ei internetissä toimivat palvelut. (Obar & Wildman, 2015)

Tästä syystä tutkimuksessa sosiaalisesta mediasta puhuttaessa tarkoitetaan esimerkiksi Twitterin, Facebookin, Instagramin ja Snapchatin kaltaisia palveluita, joiden pääasiallinen tarkoitus on välittää käyttäjien tuottamaa sisältöä toisille käyttäjille. Palvelun sisällöstä vain todella pieni osa tulisi olla palvelun tarjoamaa. Tämä määritelmä ei tuota ongelmia tutkimuksen kannalta, koska kaikki mielenkiinnon kohteena olevat palvelut ovat suuria toimijoita alalla, eikä sekaannuksia

määritelmän myötä pitäisi tapahtua. Käytän myös termiä 'käyttäjälähtöinen sosiaalinen media', jolla tarkoitetaan sitä, että pääpaino on käyttäjissä ja palveluntarjoaja tarjoaa vain alustan.

3.1.2 Mielenterveysongelmat tai mielenterveyden häiriöt

Psykologisesta ja lääketieteellisestä näkökulmasta mielenterveyden häiriöt tai mielenterveysongelmat ovat todella laaja käsite. Mielenterveyden häiriöiden määritelmä tai sisältö vaihtelee hieman, esimerkiksi maasta riippuen. Tässä tutkimuksessa pohjaan määritelmäni American Psychiatric Associaton:in (APA) kirjaan "Diagnostics and Statistical Manual of Mental Disorders, Fifth Edition" (DSM-5). Kirja on julkaistu APA:n toimesta vuonna 2013. Myös suuri osa tähän tutkimukseen liittyvistä julkaisuista käyttää samaa kirjaa alkuperänä. Kyseinen kirja määrittelee mielenterveyden häiriöt (a mental disorder) seuraavalla tavalla.

> A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behaviour that reflects a dysfunction in the psychological, biological, or developmental process underlying mental functioning. (American Psychiatric Association, 2013[1])

Tässä tutkimuksessa lähteiden tyypistä johtuen keskityn pääasiassa mielialahäiriöiden, ja eritoten masennuksen havaitsemiseen. Suuri määrä tutkimusta aiheesta johtunee masennuksen määrästä valtaväestössä sekä selkeiden, tunnistettavien oireiden näkyvyydestä. En kuitenkaan karsi muita mielenterveyden ongelmia pois, sillä aihe itsessään ideaalitilanteessa on skaalautuva mahdollisimman moneen mielenterveyden häiriöön.

Diagnosointivaiheessa masennus jaetaan taudin vakavuuden perusteella eri luokkiin. Oireet ovat kuitenkin samankaltaisia kaikilla vakavuusasteilla, eikä luokituksella ole merkitystä tutkimuksen kannalta. Oireet ovat monenlaisia, ja myös erittäin yksilöllisiä. Yleisimpiä oireita ovat kuitenkin huomattava mielialan lasku, yleisen mielenkiinnon tai mielihyvän puute sekä energiatason ja keskittymiskyvyn lasku (American Psychiatric Association, 2013). Muita oireita tarkastellaan luvussa 4.3, jossa tutkitaan sitä, kuinka tekstianalyysin avulla voidaan havaita viitteitä kliinisiin oireisiin sosiaalisen median julkaisuissa.

3.2 Suosituimmat sosiaaliset verkostot

Facebook on maailman suosituin sosiaalinen verkosto käyttäjämäärällä mitattuna (Statista, 2017). Facebook ei kuitenkaan tarjoa kehittäjille mahdollisuutta päästä käsiksi yksityisten käyttäjien julkaisuihin ilman heidän suostumusta,

[1] Sivunumero ei saatavilla

mitkä ovat tämän tutkimuksen kannalta mielenkiintoisimmat. Tämän lisäksi Facebook ei anna mahdollisuutta tutkia ihmisten käyttäytymistä, verkostoa tai vuorovaikutusta palvelussa. Tästä johtuen kaikki aiheeseen keskittyvät tutkimukset keskittyvät pääasiassa mikroblogeihin, kuten Twitteriin ja Sina Weiboon (Wongkoblap, Vadillo & Curcin, 2017).

Twitter on mikroblogi, jota pidetään Facebookin rinnalla tärkeimpänä sosiaalisena mediana. Twitterissä oli vuoden 2017 3. kvartaalilla 330 miljoonaa aktiivista käyttäjää (Statista, 2017). Palvelun pääasiallinen sisältörakenne koostuu käyttäjien luomista yksittäisistä 'twiiteistä', joidenka pituus on rajoitettu. Yhden twiitin pituus oli yli kymmenen vuotta 140 merkkiä, mutta marraskuussa 2017 Twitter nosti rajan 280 merkkiin. Palveluun voi lisätä myös kuvia ja videoita. Twiittiin voi sisällyttää myös mainintoja sekä tunnisteita. Mainintа liittää jonkin toisen käyttäjän twiittiin ja ilmoittaa siitä hänelle. Tunniste on tapa kertoa, mihin asiaan tai avainsanaan twiitti liittyy. Sina Weibo on vastaavanlainen palvelu kuin Twitter, mutta on erittäin suosittu Kiinassa. Sina Weibon omistaa Weibo Corporation.

Instagram on Facebook Incorporated:in omistama sosiaalinen verkosto. Palvelun päärakenne on kuvissa, eikä palveluun voi julkisesti jakaa mitään muuta kuin kuvia tai videoita. Kuvan mukana käyttäjä voi kirjoittaa pitkänkin kuvatekstin. Kuten Twitterissäkin, tunnisteet pelaavat suurta roolia kuvateksteissä, mutta Instagramin käyttö kuitenkin perustuu lähes täysin kuvien ja lyhyiden videoiden jakamiseen. Instagram tarjoaa myös erittäin käyttökelpoisen rajapinnan datan hankintaan. Tämän lisäksi suuri osa käyttäjistä pitää julkista tiliä, mikä helpottaa datan hankintaa.

3.3 Yleisimmät sosiaaliset mediat teknologisesta näkökulmasta

Koneoppivien, tai minkä tahansa suurta määrää tietoa hyödyntävien sovellusten kehityksen merkittävä kompastuskivi on datan hankinta. Osa suosituimmista sosiaalista medioista tarjoavat kuitenkin hyvinkin käytettävät ohjelmointirajapinnat. API on sovelluksen käyttöliittymä, jolla kaksi ohjelmaa pystyvät 'keskustelemaan' keskenään (Lomborg & Bechmann, 2014). Usein sosiaaliset mediat tekevät heidän sovelluksensa ohjelmointirajapinnan julkiseksi. Tämä tarjoaa kolmansille osapuolille mahdollisuuden esimerkiksi integroida palvelu omaansa (Lomborg & Bechmann, 2017).

Rajapinnan avulla pystyy ohittamaan tarpeen käyttää visuaalista käyttöliittymää verkkoselaimessa tai sovelluksessa, ja kerätä dataa suoraan oman ohjelman kautta. Tämä tapahtuu usein suoraan http-pyynnöillä palvelun palvelimille. Esimerkiksi tutkijat hyödyntävät tätä datan hankinnassa. Jos jonkin sovelluksen

dataa halutaan kerätä tutkimusta varten suuria määriä, on käyttökelpoinen rajapinta lähes välttämätön. Tämä karsii esimerkiksi Snapchatin[2] suhteellisen käyttökelvottomaksi. Huono rajapinta on aiheen tutkimuksen kannalta menetys, sillä Snapchat on todella suosittu palvelu länsimaisten nuorten keskuudessa, ja se keskittyy todella paljon arkipäiväisten kuvien ja videoiden lähettämiseen ja vastaanottamiseen.

Kuten aiemmin mainittu, suosituimmista sosiaalisista medioista etenkin Twitter ja Instagram tarjoavat erittäin laajat ja käyttökelpoiset rajapinnat (Twitter Inc., 2017), (Facebook, Inc., 2017). Näiden rajapintojen avulla on mahdollisuus hankkia suuri määrä julkisia kuvia tai tekstijulkaisuja esimerkiksi analysointia varten. Kuten aikaisemmin todettu, koneoppivien järjestelmien avulla saavutetaan parempia tuloksia suuremmalla opetusdatamäärällä (Batista, Prati & Monard, 2004).

[2] Snapchat on Snap Incorporated:in omistama sosiaalinen media, jossa voi joko lähettää tai julkaista 'snappejä' (snap). Yksi 'snap' on joko kuva tai video, joka näkyy lähetettynä vain yleensä vain 1-10 sekuntia, tai julkaistuna on katseltavissa 24 tuntia.

4 MIELENTERVEYSONGELMIEN HAVAITSEMINEN

Tässä luvussa esittelen muutaman erittäin tuoreen tutkimuksen, jotka antavat todella hyvää suuntaa siihen, millaisiin muuttujiin on kiinnitettävä huomiota. Tutkimukset ovat myös erilaisia, jotta mahdollisimman erilaisia merkkejä löytyisi. Tämän jälkeen luvun lopuksi kokoan ja tarkastelen näitä merkkejä. Tarkoitus on tutkimusten pohjalta löytää yleisimmät ja käyttökelpoisimmat lähestymistavat. Näiden yhdistäminen toisiinsa tuo lisää potentiaalia, ja osa tutkimuksista onkin tehnyt tutkimusta hyödyntäen useampaa lähestymistapaa.

Koneoppimisessa ja etenkin sen hyödyntämisessä enemmän on parempi. Koneella on helppo tunnistaa mitkä käyttäjät sanovat mitäkin sanoja tai lisäävät millaisia kuvia. Tämä ei kuitenkaan kerro muuta kuin sen, millaista sisältöä he julkaisevat. Sovelluskohteissa, ja etenkin jossain niinkin tärkeässä asiassa kuin mielenterveysongelmien ennakoinnissa on panostettava oikeellisuuteen. Oikeellisuutta, tarkkuutta ja virheiden vähyyttä voidaan parantaa huomioimalla useita muuttujia. Koneoppimisen ja neuroverkkojen hyöty verrattuna sääntöpohjaiseen ohjelmointiin tuleekin esille vasta kun käsitellään suurta määrää itsessään merkityksettömiä muuttujia.

4.1 Masennuksen havaitseminen Twitteristä

Shenin ja kumppanien (2017) tekemässä tutkimuksessa pystyttiin luomaan datasarjat terveille, sekä masennuksesta kärsiville käyttäjille. Käyttäjät jaoteltiin tähän sarjaan ottaen huomioon masennuksen kliiniset, oikeassa maailmassa näkyvät oireet sekä verkkopalvelussa ilmaantuvat oireet. Tämän pohjalta rakennettiin koneoppimismalli, joka tunnistaa masennukseen liittyvää sanastoa. Tutkijat suorittivat rakentamansa mallin avulla case-tapauksia, joilla pyrittiin tunnistamaan selkeästi erot masentuneiden ja ei-masentuneiden käyttäjien välillä. Tätä kautta saatiin pieni määrä esimerkkitapauksia, mutta nämä esimerkkitapaukset ovat hyvin merkittyä dataa. (Shen ym., 2017.) Tutkimuksen loppuosa on siis luvussa 2.2.2 mainittua aktiivista oppimista. Tämän koneoppimismallin avulla pystyttiin tutkimaan suurta määrää dataa Twitterin rajapinnan avulla, ja tutkimus esittää muutamia tilastollisia merkkejä oletetusta masennuksesta. Nämä löydöt ovat:

1. Masentuneet käyttäjät tekevät 44 % enemmän päivityksiä normaalikäyttäjään verrattuna kello 22:00 – 06:00 välillä. Tämä voi viitata unettomuuteen.
2. Masentuneet käyttäjät puhuvat tunteistaan enemmän päivityksissään. ”Depressed users have 0.37 positive words and 0.52 negative words per tweet, which surpasses those of non-depressed users by 0.17 and 0.23” (Shen ja kumppanit, s.3843, 2017).

3. Masentuneet käyttäjän käyttävät päivityksissään ensimmäisen persoonan pronomineja kolme kertaa enemmän kuin muut käyttäjät. Tutkijoiden mukaan tämä voi viitata tukahdutettuihin monologeihin ja itsetietoisuuteen.
4. Masentuneen käyttäjän päivitykset sisältävät masennuksen oireisiin viittaavaa sanastoa 165 % enemmän kuin terveen käyttäjän. Tämä viittaa siihen, että he päivittävät siitä, mitä ovat kohdanneet oikeassa elämässä.

Osa merkeistä on sellaisia, jotka voidaan soveltaa mihinkä tahansa mediaan, joka sisältää tyypillisiä, tekstipohjaisia päivityksiä (esimerkiksi Facebook). Löytö numero 1 on myös sovellettavissa mihin tahansa sosiaaliseen mediaan, riippumatta sisällön tyypistä.

4.2 Erilaisia lähestymistapoja masennuksen havaitsemisessa

Toinen vastaava tutkimus on tehty jo aiemmin, vuonna 2013. Choudhuryn, Gamonin, Countsin ja Horvitzin (2013) tutkimuksessa löydettiin samanlaisia merkkejä kuin Shenin kumppanien (2017) tutkimuksessa. Tutkimus suoritettiin alkuun kyselyillä sekä terveille, itseilmoitetusti masentuneille sekä diagnosoidusti masentuneille henkilöille. Tutkimukseen osallistujilta kerättiin Twitter-julkaisut ja aktiviteetit joko vuoden ajalta ennen kyselyä, tai vuoden ajalta ennen ilmoitettua masennuksen alkua. Tämä tehtiin, jotta saatiin kerättyä erittelevien merkkien lisäksi ennakoivia merkkejä masennuksesta. (Choudhury ym., 2013.)

De Choudhuryn ja kumppanien (2013) tutkimus ”Predicting Depression via Social Media” keskittyi käyttäjien osalta lyhyesti sanottuna viiteen eri asiaan heidän Twitter-syötteessään; osallistuminen, egosentrinen sosiaalinen verkosto, tunnetilat, kielellinen tyyli sekä masennukseen liittyvä sanasto. Osallistumisella tarkoitetaan yksinkertaistettuna sitä, kuinka paljon, milloin, miten ja minkä tyylisiä (vertaa julkaisut, ja vastaukset) twiittejä käyttäjä julkaisee. Sosiaalisella verkostolla viitataan siihen, miten käyttäjä vuorovaikuttaa muiden käyttäjien kanssa. (Choudhury ym., 2013.)

Tunnetilat, kielellinen tyyli sekä masennukseen liittyvä sanasto keskittyvät twiittien sisältöön ja niiden analysointiin. Tunnetila ja kielellinen tyyli tehtiin analysoimalla käytettyä kieltä hyödyntäen tekstin analysointityökalua ’LIWC’. (Choudhury ym., 2013.) LIWC on luonnollisen kielen prosessoinnin työkalu, joka vertaa tekstistä kerättyjä avainsanoja valmiiksi kategorioituihin sanastoihin. Nämä sanastot ovat kategorioitu esimerkiksi sanojen positiivisuuden tai negatiivisuuden mukaan, tai sitten kieliopillisesti esimerkiksi verbeihin, substantiiveihin ja muihin sanaluokkiin. (Tausczik & Pennebaker, 2010).

Käyttäjien pohjalta rakennettiin kaksi luokkaa, terveet sekä masentuneet. Tässäkin tutkimuksessa löydettiin samankaltaisia tuloksia kuin aikaisemmin luvussa 4.1 esitellyssä. Mielenkiintoisempaa tässä tutkimuksessa merkkien sijaan ovat lupaavat tulokset heidän rakentamassaan ohjatussa koneoppimismallissa. Tämän järjestelmät tarkoituksena on tunnistaa, onko yksittäinen käyttäjä altis

masennukselle, jo ennen sen laukeamista. (Choudhury ym., 2013.) Taulukossa (TAULUKKO 2) on listattu sekä eroteltu rakennetun mallin tarkkuudet, eli kuinka suuri osa sen ennustuksista meni oikein. Tuloksista näkee myös hyvin sen, kuinka useammalla tarkastelun kohteella pääsee parempiin tuloksiin. Tämä ei kuitenkaan tässä tutkimuksessa ole kovin merkittävä asia, sillä lähes yhtä hyviin tuloksiin päästiin pelkällä twiittien sisällön analyysilla.

Taulukon kolmas kolumni on mielenkiintoisin, sillä se kuvaa koneoppimismallin osuvuutta juurikin masentuneiden ihmisten löytämisessä annetusta datasta. Rivit kuvaavat aiemmin luvussa kerrottuja tutkimuksen tarkastelun kohteita, ja viimeinen kohta kuvastaa pelkästään käyttäjien demografisten ominaisuuksien pohjalta tehtyä vertailua. Alin, lihavoitu rivi kuvaa parasta rakennettua mallia. Kuten tuloksista näkyy, rakennettu koneoppimismalli saavutti noin 70% tarkkuuden.

	precision	recall	acc. (+ve)	acc. (mean)
engagement	0.542	0.439	53.212%	55.328%
ego-network	0.627	0.495	58.375%	61.246%
emotion	0.642	0.523	61.249%	64.325%
linguist. style	0.683	0.576	65.124%	68.415%
dep. language	0.655	0.592	66.256%	69.244%
demographics	0.452	0.406	47.914%	51.323%
all features	0.705	0.614	68.247%	71.209%
dim. reduced	**0.742**	**0.629**	**70.351%**	**72.384%**

TAULUKKO 2 tutkimuksen ”Predicting Depression via Social Media” (Choudhury ym., 2013, s.136) koneoppimismallin tekemien ennustuksien tarkkuudet.

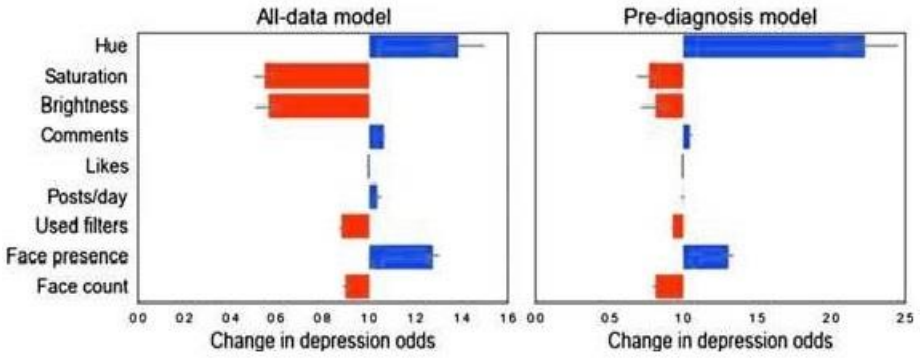
4.3 Ennakoivien merkkien havaitseminen Instagramista

Reecen ja Danforthin (2017) tekemässä tutkimuksessa “Instagram photos reveal predictive markers of depression” toteutettiin tutkimus, jonka tarkoituksena oli erotella masennuksesta kärsivien ja terveiden ihmisten Instagramiin lisäämiä kuvia. Tutkimuksen testiryhmät (masentuneet ja terveet) suorittivat eri kyselyt. Masentuneet ihmiset suorittivat ensin CES-D-kyselyn (Center for Epidemiologic Studies Depression Scale), jolla kartoitetaan ’masennustasoa’ (depression-scale) (Radloff, 1977). Tutkimukseen osallistui näiden lisäksi joukko ihmisiä, joita pyydettiin arvioimaan testihenkilöiden jakamia kuvia. (Reece & Danforth, 2017)

Tutkimuksessa löytyi useita tekijöitä, jotka voidaan yhdistää masentuneen ihmisen käyttäytymiseen kyseisessä palvelussa. Osa löydöksistä liittyy kuviin ja

osa kuviin liittyvään metadataan, esimerkiksi valittuun filtteriin[3] tai julkaisuihin liittyvään interaktioon[4]. (Reece & Danforth, 2017) Tutkijat opettivat kaksi koneoppimismallia, joita he kutsuivat annetusta datasta riippuen ”All-data”-malliksi ja ”Pre-diagnosis-malliksi. Aikaisempi käytti nimensä mukaan kaikkea kerättyä dataa, ja ”Pre-diagnosis”-malli käytti kaikkea terveiltä osallistujilta kerättyä dataa, sekä masentuneilta ihmisiltä dataa, joka oli julkaistu ennen heidän saamaansa diagnoosia. (Reece & Danforth, 2017) Eli ”Pre-diagnosis”-mallin tuloksia voidaan soveltaa nimenomaan esimerkiksi masennuksen ennakoinnissa.

Reece ja Danforth päätyivät seuraavanlaisiin tuloksiin, jotka näkyvät myös kuviossa (KUVIO 1). Masentuneiden käyttäjien julkaisemat kuvat olivat yleensä vähemmän värikylläisiä (hue), eli enemmän sini-, harmaa- ja tummasävyisiä. Mitä enemmän kommentteja julkaisu sai, sitä todennäköisemmin julkaisun oli tehnyt masennuksesta kärsivä käyttäjä. Tykkäysten kohdalla efekti on kuitenkin päinvastainen. Myös korkeampi julkaisutiheys pystyttiin liittämään masennukseen, kun masentuneilta käyttäjiltä kerättiin myös diagnoosin jälkeiset julkaisut (”all-data”-malli). Masentuneet käyttäjät julkaisevat todennäköisemmin kasvot sisältäviä kuvia, eli kuvia ihmisestä. Terveiden käyttäjien kuvat kuitenkin sisälsivät keskiarvolta enemmän ihmisiä kuin masentuneiden käyttäjien julkaisemat kuvat. Masentuneet käyttäjät käyttivät kuvissaan filttereitä vähemmän todennäköisesti. (Reece & Danforth, 2017)



KUVIO 1 Reecen ja Danforthin tutkimuksen löydökset (Reece & Danforth, 2017, s.7).

4.4 Kielen analyysi kliinisiä oireita varten

Aldarwish ja Ahmad (2017) käyttivät tutkimuksessa hieman erilaista lähtökohtaa. He käyttivät luonnollisen kielen analysointia parsimaan eri sosiaalisten medioiden (Facebook, LiveJournal ja Twitter) julkaisuja. Sen sijaan että etsittiin sosiaalisille medioille uniikkeja merkkejä, he tekivät koneoppimismallin, joka hakee

[3] Filtterillä tarkoitetaan palvelun tarjoamaan vaihtoehtoon lisätä kuvaan muokkaus, joka muuttaa esimerkiksi kuvien värikylläisyyttä, valaistusta tai saturaatiota.

[4] Muut käyttäjät voivat Instagramissa tykätä julkaisuista, tai kommentoida niitä.

julkaisun sisällöstä kliinisiin oireisiin viittaavaa puhetta. Esimerkit lauseista kuvataan kuviossa 2 (KUVIO 2). Tutkimus ei tuottanut luotettavaa mallia, johtuen kielellisistä (arabi) haastavuuksista, sekä vaikeudesta löytää aktiivisesti sosiaalisia medioita käyttäviä, masentuneita yksilöitä. (Aldarwish & Ahmad, 2017).

Tarkoitukseni ei olekaan löytää toimivaa ja tarkkaa mallia, vaan hankkia esimerkkejä lähestymistavoista ja merkeistä, joiden pohjalta voitaisiin tehdä erilaisia johtopäätöksiä. Tämän tutkimuksen lähestymistapa on erilainen, ja sitä voidaan yhdistää muihin merkkeihin, joka taas edistää mahdollisen koneoppimismallin vahvuutta ja tarkkuutta.

Symptom	Post
Sadness	I just found out my mom never wanted me in the first place.; That just ruined my day.
Loss of Interest	What, if anything, is there to live for?
Appetite	I was a little depressed that I ate so much last night there were no leftovers today
Sleep	It was a sleepless night
Thinking	I can't concentrate.
Guilt	I feel bad for doing it
Tired	too worried and tired to post tonight
Movement	I think I just gave myself permission to be lazy.
Suicidal ideation	I did it again. I don't know what I was thinking. I cut a star-like design into my upper left arm, and then took a whole bunch of pills and strong scotch. life is not going well.

KUVIO 2 Esimerkit kutakin masennuksen oiretta kuvaavasta virkkeestä, lauseesta tai lauseenpätkästä (Aldarwish & Ahmad, 2017, s.279).

4.5 Merkit tai oireet ja niiden potentiaalit koneoppimiselle

Esitellyn kirjallisuuden perusteella voidaan todeta, että jos lähdetään havaitsemaan tai ennakoimaan mielenterveysongelmia käyttäjän julkaisemasta datasta, on mahdollista keskittyä joko julkaisuihin liittyviin tekijöihin, tai lähteä suoraan etsimään kliinisiin oireisiin viittaavaa puhetta rakenteettomasta datasta. Tarkastelen ensiksi merkkeihin liittyvää havainnointia, ja niihin käytettyjä teknologioita. Tämän jälkeen avaan hieman myös kliinisiin oireisiin perustuvaa lähtöasetelmaa.

4.5.1 Julkaisujen kielen käyttöön ja sisältöön sidotut merkit

Julkaisuiden sisältöön sidotuilla merkeillä tarkoitan tässä tutkimuksessa kahta asiaa. Joko julkaisun tekstiin liittyviä kielellisiä indikaattoreita henkilön mielenterveyden tilasta, tai muun tyyppisen sisällön tarjoamia mahdollisuuksia havaita vastaavia merkkejä. Tämä tarkoittaa käytännössä pelkästään kuvia.

Näitä merkkejä voi olla monenlaisia. Esimerkiksi on havaittu, että masennuksesta kärsivät ihmiset käyttävät kielessään ensimmäisen – persoonan pronomineja merkittävästi enemmän kuin ihmiset, jotka eivät ole koskaan kärsineet masennuksesta (Rude, Gortner & Pennebaker, 2004). Tätä havaintoa tukee myös aiemmin luvussa 4 esitellyt tutkimukset, ja niiden kautta voidaan myös olettaa, että sosiaalinen media ei ole poikkeus tälle kielelliselle ilmiölle. Masentuneet ihmiset käyttävät puheessaan ja kirjoituksissaan myös negatiiviin tunteisiin sidottuja sanoja enemmän kuin ei masentuneet (Rude ym., 2004). Sisältöön liittyviä merkkejä ovat myös esimerkiksi lisääntynyt huoli lääketieteellisistä ongelmista sekä kasvava ilmaisu uskonnollisista ajatuksista (Choudhury ym., 2013). Myös esimerkiksi traumaperäisestä stressihäiriöstä kärsiville potilaille on tyypillistä puhua traumastaan (Harman & Dredze, 2014), jonka havaitseminen käyttäen kielen prosessoinnin tekniikoita ei ole mahdotonta. Myös erilaisia tunteita ja psykologisia ahdinkoja on pystytty havaitsemaan todella tehokkaasti koneoppimismallilla (Saleem ym., 2012). Näitä tunnetiloja ja ahdinkoja voidaan käyttää löytämään merkkejä, jotka viittaavat esimerkiksi lääketieteellisiin diagnooseihin. Lisäksi masentuneiden käyttäjien twiiteistä analysoidu mielentila vaihtelee enemmän ajan kuluessa, ja kokonaisuudessaan lähettävät enemmän negatiivis-sävytteisiä twiittejä (Kang, Yoon, Kim, 2016). Voidaan siis todeta, että kielestä analysoituja tunnetiloja on mahdollista hyödyntää myös masennuksen havaitsemisessa.

4.5.2 Kuvien sisältöön liittyvät merkit

Myös kuvista voidaan saada vastaavanlaista informaatiota. Luvussa 4.2 esitellyssä tutkimuksessa tutkijat pystyivät erottamaan selkeitäkin tilastollisia eroja masentuneiden ja terveiden ihmisten kuvissa. Tutkijat pääsivät n. 70% tarkkuuteen luokittelijallaan, ja saavuttivat pienemmän väärin-diagnosoitujen terveiden käyttäjien määrän verrattuna aikaisempiin malleihin (Reece & Danforth, 2017). Aiheeseen liittyvää edistystä siis tapahtuu koko ajan, mikä saattaa tarkoittaa huomattavasti tarkempia koneoppimismalleja myöhemmin.

4.5.3 Muut, ei-sisältöön sidotut merkit

Ei-sisältöön sidotut tai liittyvät merkit tarkoittavat tässä kontekstissa kaikkia muita merkkejä, jotka voidaan saada tutkimalla ihmisen käyttäytymistä sosiaalisissa medioissa. Tämä voi tarkoittaa esimerkiksi metadataa julkaisuista, sosiaalisia verkostoja palvelussa, erityyppisiä vuorovaikutuksia palveluiden sisällä, tai yksinkertaisesta aktiivisuutta tai muutosta aktiivisuudessa.

Ei-sisältöön sidotut merkkejä on havaittu, mutta niihin on keskitytty selkeästi vähemmän tutkimuksissa. Tämä johtunee siitä, että kieleen ja julkaisuiden sisältöön liittyvät analyysit antavat selkeästi parempia tuloksia. Myös tämänkaltaisen datan analysointi voi olla todella paljon vaikeampaa ja monimutkaisempaa kuin julkaisuista saadun raa'an tekstidatan analyysi. Rosenquist, Fowler ja Christakis (2011) havaitsivat tutkimuksessaan kuitenkin esimerkiksi sen, että yksittäinen henkilö osoittaa masennukseen viittaavia oireita lähes kaksi kertaa todennäköisemmin, mikäli hänen lähipiirissään on masentuneita henkilöitä. Tämänkaltaisten löydösten hyödyntäminen sosiaalisten verkostojen analysoinnissa on vähäistä, mutta niillä voi todennäköisesti saada huomattavaa lisäarvoa.

Kuitenkin merkittäviä havaintoja sosiaalisista medioista on saatu. Masentuneet käyttäjät tekevät päivityksiä terveitä käyttäjiä todennäköisemmin myöhään illalla tai yöllä (Choudhury ym., 2013) (Shen ym., 2017). Myös Reecen ja Danforthin tutkimuksessa havaittiin Instagram-kuvien metadataan sidottuja merkkejä. Merkit liittyivät paljolti kommenttien ja tykkäysten määrään ja tyypin tilastollisiin ominaisuuksiin (Reece & Danforth, 2017). Tämä tutkimus ei ota kantaa siihen, kuinka paljon lisäarvoa tämänkaltaisen datan käyttö koneoppimismallissa tuo, verrattuna kuvien sisältöön liittyvän datan käyttö.

Nämä ei-sisältöön liittyvät datalähteet voivat siis antaa lisäarvoa, kun pyritään havaitsemaan tai ennakoimaan mielenterveysongelmia sosiaalisessa mediassa. Koneoppivat luokittelijat voivat myös hyödyntää tämänkaltaista dataa. On kuitenkin havaittavissa, että julkaisuiden sisällön analyysi antaa kokonaisuudessaan paremman lähtökohdan. Se ei kuitenkaan poissulje näiden käyttöä, vaan päinvastoin voi parantaa jo rakennettujen koneoppimismallien tarkkuutta.

5 YHTEENVETO

5.1 Tulokset ja havaintoja

Tutkimuksen tarkoitus oli kartoittaa sitä, miten kognitiivisia teknologioita voidaan hyödyntää mielenterveysongelmien havaitsemisessa sosiaalisessa mediassa. Katsauksen myötä löytyi useita erilaisia merkkejä ja tekijöitä, joita voidaan havaita koneoppivilla ohjelmilla. Koneoppivat ohjelmat ovat päässeet näitä merkkejä hyödyntäen merkittävän hyviin tarkkuuksiin. Tämä prosessi tapahtuu analysoimalla eri osia sosiaalisesta mediasta, ja opettamalla kone tunnistamaan kyseisiä merkkejä. Käyttäjien julkaisut on helpoin hankkia hyödyntämällä palveluntarjoajan rajapintoja. Mielenterveyden häiriöihin viittaavia merkkejä on muun muassa:

- **Tekstissä ja metadatassa**. Esimerkiksi masentuneet käyttäjät puhuvat enemmän ensimmäisessä persoonassa, julkaisevat päivityksiä myöhään yöllä ja puhuvat tunteistaan sekä oireistaan enemmän kuin terveet käyttäjät. Myös tekstistä analysoituja tunnetiloja voidaan hyödyntää (Kang, Yoon, Kim, 2016).
- **Kuvissa**. Masentuneilla käyttäjillä on kuvissa useammin ihmisiä, mutta terveillä ihmisten keskiarvomäärä on suurempi. Masennuksesta kärsivät julkaisevat myös värisävyltään siniharmaampia kuvia kuin terveet. (Reece & Danforth, 2017)
- **Vuorovaikutuksessa ja verkostossa**. Mikäli ihmisellä on lähipiirissä masennusta, osoittaa hän itse masennukseen viittaavia oireita tervettä käyttäjää todennäköisemmin (Fowler & Christakis, 2011).
- Parhaan mahdollisen tuloksen saavuttamiseksi analyysin tulisi yhdistää kaikkia näitä alueita.

Näitä merkkejä voidaan siis opettaa koneoppiville ohjelmille, ja niiden avulla luokitella käyttäjiä. Tutkimuksissa on päästy metodista riippuen n. 70% tarkkuuteen malleissa. Aikaisemmat tutkimukset ovat keskittyneet pääsääntöisesti masennuksen havaitsemiseen, mutta vastaavia tuloksia on saatu esimerkiksi traumaperäisen stressihäiriön havainnoinnista samankaltaisin keinoin Twitter-julkaisuista (Harman & Dredze, 2014). Myös Saleem ja kumppanit pystyivät havaitsemaan erilaisia mielentiloja foorumi-viestien tekstianalytiikalla (Saleem ym. 2012), joka todettiin käyttökelposeksi lähestymistavaksi luvussa 4.5.1. Tämä ei sikäli suoraan ole mielenterveyden häiriöihin liittyvää, mutta tietyt tunnetilojen oireilut liittyvät eri mielenterveyden häiriöihin (American Psychiatric Association, 2013).

Suurin osa tutkimuksista aiheeseen liittyen perustuu tekstianalytiikkaan, ja itsekseen käytettynä se antaa parhaat tulokset. Yleisesti ottaen tutkimuksissa

käytetään hyödyksi jotakin luonnollisen kielen ymmärrys – palvelua, jolla muodostetaan kaksi tai useampi luokitusta, liittyen käyttäjien mielenterveyteen. Kuva-analyysiin pohjautuu myös tutkimuksia, mutta ei niin paljoa. Kuvien perusteella onnistuttiin löytämään masennukseen viittavia merkkejä (Reece & Danforth, 2017), sekä käyttäjän stressiin liittyvää analyysia on tehty pohjautuen Twitterissä twiittien lisäksi kuviin sekä vuorovaikutukseen (Lin ym. 2014). Myös sosiaalisiin verkostoihin ja vuorovaikutuksiin liittyvää tutkimusta on tehty, mutta niitä on merkittävästi vähemmän. Asiaa ei kannata kuitenkaan sivuttaa. Esimerkiksi Wang, Zhang ja Sun (2013) pystyivät parantamaan luokittelijansa tarkkuutta merkittävästi, kun huomioon otettiin tyypillisen tekstidatan lisäksi vuorovaikutukseen liittyvää analyysia.

Tutkimuksissa on käytetty pääsääntöisesti palveluntarjoajien rajapintoja datan hankintaa, mikäli sellainen on. Osa palveluista ei tarjoa käyttäjien dataa kehittäjille, mikä käytännössä estää koneoppivien mallien opettamisen, ja vähentää tutkimuksien määrää. Tämä on merkittävä tekijä sovellettavuudelle, mitä käsittelen tarkemmin seuraavassa luvussa. Tekstianalytiikan työkaluna psykologisissa tutkimuksissa käytetään pääasiassa LIWC – ohjelmistoa (Wongkoblap ym., 2017).

5.2 Pohdintaa, huomiot ja muut lähestymistavat

Kielen käyttöön liittyvät havainnot ovat Wongkoblapin ja kumppanien (2017) mukaan suhteellisen universaaleja, eli löydöksiä voidaan soveltaa muuhunkin kuin löytöjen alkuperäiseen kieleen. Ongelmia tulee kuitenkin vastaan etenkin pienemmillä kielillä, joissa käytettäviä työkaluja koneoppimiseen ja luonnollisen kielen ymmärtämiseen ja prosessointiin ei välttämättä ole vielä saatavilla.

Löytämistäni tutkimuksista yksikään ei pyrkinyt esimerkiksi julkaistuihin teksteihin tai kuviin annettujen kommenttien analyysiin. Tämä voisi teoriassa antaa lisää viitteitä etenkin mahdollisiin poikkeamiin julkaisun tekijän käytöksessä. Tutkimuksia liittyen keskustelufoorumeihin tai blogeihin, jotka itsessään ovat teknisesti sosiaalisia medioita, on vähän. Näissä kuitenkin on usein paljon pidempiä tekstijulkaisuja verrattuna mikroblogeihin, ja data todella usein julkisesti saatavilla. Näissä kuitenkin voi tulla vastaan käyttäjien anonymiteetti ja vaikea tavoitettavuus.

Wongkoblapin ja kumppanien (2017) suorittamassa katsauksessakaan ei löydetty ainuttakaan tutkimusta, jossa tutkittavat olisivat lääketieteellisesti diagnosoituja. Vaikkakin aikaisemmin esitelty CES-D on hyvä lähtökohta, voisi lääketieteellisesti diagnosoitujen potilaiden sosiaalisen median data antaa paremman tarkkuuden löydöissä (Wongkoblap ym., 2017). Tähän tietysti liittyvät lainsäädännölliset ja eettiset ongelmat potilasdatan jaosta ja käytöstä.

Miten näitä löydöksiä tulisi siis hyödyntää? Aiheeseen liittyy monta fundamentaalista ongelmaa, jotka estävät tai vähintään hidastavat tätä prosessia. Ensinnäkin, vaikka käytetty data on usein julkista, on ihmisten datan käyttö varsinkin tämänkaltaisessa käyttötapauksessa eettisesti kyseenalaista. Käyttäjiltä luvan kysyminen on vaihtoehto, mutta se ei itsessään sovi ajatukseen, jossa tämänkaltaisen teknologian etuus on nimenomaan suuren datamäärän hallinnassa ja analysoinnissa. Tähän tietynlainen ratkaisu voisi olla esimerkiksi lääketieteellisten instituutioiden ja organisaatioiden yhteistyö palveluntarjoajien kanssa. Toiseksi, kun puhutaan ihmisten terveydestä, voivat väärin luokitellut ihmiset (ns. false positive) tuottaa ongelmia. Lisäksi aiheeseen liittyy myös ongelmia liittyen ihmisten tavoittamiseen, vastuullisuuskysymykseen[5], sekä muita, kaupallisia ongelmia.

Mikäli näihin ongelmiin löydetään kuitenkin ratkaisuja, voisi sosiaalisen median analyysia hyödyntää esimerkiksi julkisesti hallinnoidussa terveydenhuollossa. Tämän lisäksi myös palveluntarjoaja voisi saada lisäarvoa, mikäli tällaista dataa myytäisiin kaupallisessa mielessä terveydenhuollon organisaatioille.

[5] Filosofinen pohdinta siitä, kuka, miten ja milloin on vastuussa ilmoittaa henkilön potentiaalisesta terveysongelmasta kolmannelle taholle.

LÄHTEET

Aldarwish, M. M. & Ahmad, H. F. (2017). Predicting depression levels using social media posts. *Proceedings - 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems, ISADS 2017,* 277-280.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®).* Arlington, VA: American Psychiatric Publishing.

Batista, G. E., Prati, R. C. & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM Sigkdd Explorations Newsletter*, 6(1), 20-29.

Bloom, D., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L., Fathima, S., Mowafi, M. (2012). *The global economic burden of noncommunicable diseases*. The World Economic Forum.

Chapelle, O., Schölkopf, B. & Zien, A. (2006). *Semi-supervised learning*. Cambridge, Mass.: MIT Press. Haettu osoitteesta http://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/detail.action?docID=3338523

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37(1), 51-89.

De Choudhury, M., Gamon, M., Counts, S. & Horvitz, E. (2013). Predicting depression via social media. *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, ICWSM 2013* 128-137. Palo Alto, CA: AAAI Press.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM55*, 78-87.

Harman, C. T. & Dredze, M. H. (2014). Measuring post-traumatic stress disorder in twitter. *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media.* 579-582. Palo Alto, CA: AAAI Press.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). Overview of supervised learning. Teoksessa T. Hastie, R. Tibshirani & J. Friedman (toim.), *The elements of statistical learning: Data mining, inference, and prediction* 9-41. New York, NY: Springer New York.

Kaelbling, L. P., Littman, M. L. & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237-285.

Kang, K., Yoon, C. & Kim, E. Y. (2016). Identifying depressive users in twitter using multimodal analysis *IEEE Big Data and Smart Computing (BigComp), 2016 International Conference*. 231-238.

Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Feng, L. (2014). User-level psychological stress detection from social media using deep neural network. *Proceedings of the 22nd ACM international conference on multimedia* 507-516. New York, NY: ACM.

Mcdaid, D. & Park, A. (2011). Investing in mental health and well-being: Findings from the DataPrev project. *Health Promotion International,* 26(1), i108-i139.

McDermott, J. (1982). R1: A rule-based configurer of computer systems. *Artificial Intelligence, 19(1),* 39-88.

Obar, J. A. & Wildman, S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy, 39(9),* 745-750.

R. Katikalapudi, S. Chellappan, F. Montgomery, D. Wunsch & K. Lutzen. (2012). Associating internet usage with depressive behavior among college students. *IEEE Technology and Society Magazine, 31(4)*, 73-80.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1(3),* 385-401.

Reece, A. G. & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science, 6(1),* 15.

Rosenquist, J. N., Fowler, J. H. & Christakis, N. A. (2011). Social network determinants of depression. *Molecular Psychiatry, 16(3),* 273-281.

Rude, S., Gortner, E. & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion, 18(8),* 1121-1133.

Russell, S. & Norvig, P. (1995). *Artificial intelligence: A modern approach* Pearson Education. Alan Apt.

S. Saleem, M. Pacula, R. Chasin, R. Kumar, R. Prasad, M. Crystal, T. Speroff. (2012). Automatic detection of psychological distress indicators in online forum posts. *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference,* 1-4. IEEE.

Schohn, G. & Cohn, D. (2000). Less is more: Active learning with support vector machines. *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000)*, 839–846. Morgan Kaufmann.

Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Zhu, W. (2017) Depression detection via harvesting social media: A multimodal dictionary learning solution. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17),* 3838-3844.

Tausczik, Y. R. & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29(1),* 24-54.

Chandola V., Banerjee A. & Kumar V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41(3)*, 32.

Wang, X., Zhang, C. & Sun, L. (2013). An improved model for depression detection in micro-blog social network. *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference*. IEEE. Dallas, USA.

Wongkoblap, A., Vadillo, M. A. & Curcin, V. (2017). Researching mental health disorders in the era of social media: Systematic review. *Journal of Medical Internet Research, 19(6)*, e228.

Zhu, X. (2011). Semi-supervised learning. *Encyclopedia of machine learning*, 892-897, Springer.

Statista (2017, 9) Most famous social network sites 2017, by active users. Haettu 16.12.2017 osoitteesta https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Twitter Inc. (2017). Twitter API Documentation. Haettu 16.12.2017 osoitteesta https://developer.twitter.com/en/docs

Instagram, Facebook Incorporated (2017). Instagram API Documentation. Haettu 16.12.2017 osoitteesta https://www.instagram.com/developer/