MUSTELÄISKISTÄ MUSTIIN LAATIKOIHIN

ELINA VESSONEN

Psykologisten ominaisuuksien mittareita käytetään apuna yhteiskunnallisessa päätöksenteossa. Esimerkiksi masennuslääkekokeissa käytetään psykologisia mittareita lääkkeiden tehokkuuden arvioinnissa. Monien työnhakijoiden soveltuvuutta puolestaan arvioidaan persoonallisuusmittareilla. Mittareita hyödyntävät päätöksentekijät ovat ajan hengen mukaisesti kiinnostuneita siitä, miten koneoppiminen muuttaa näitä mittareita. Mitä uutta koneoppiminen tuo psykologisten ominaisuuksien mittaamiseen? Mitkä vanhat ongelmat ovat jatkossakin ajankohtaisia?

Kone korvaa musteläiskät

Sveitsiläisen Hermann Rorschachin 1920-luvulla kehittämä musteläiskätesti on yksi maailman tunnetuimmista psykologisista testeistä. Siinä haastattelija esittää potilaalle kymmenen musteläiskältä näyttävää, mustavalkoista tai osittain värillistä kuvaa ja pyytää tätä kertomaan, mitä potilas kuvassa näkee (Searls 2018). Rorschachin ja hänen seuraajiensa mukaan osaava haastattelija voi päätellä potilaan vastauksista psykologisia faktoja: Onko potilas masentunut? Onko hän introvertti? Onko potilaalla skitsofrenia?

Rorschachin musteläiskätesti on kuulunut psykologisen mittaamisen kalustoon vuosikymmeniä, ja sillä on edelleen kannattajansa, kertoo Damion Searls tuoreessa kirjassaan. Useimmat psykologit ovat kuitenkin hylänneet testin epätieteellisenä. Musteläiskien katselemisen sijaan psykologit suosivat standardoituja kyselyjä, joissa koehenkilöitä pyydetään arvioimaan esimerkiksi avoimuuttaan tai masentuneisuuttaan numeroasteikon avulla. 1960-luvulla julkaistussa Beckin depressiokyselyssä potilas arvioi esimerkiksi surullisuuttaan asteikolla nollasta kolmeen, jossa nolla tarkoittaa ”En ole surullinen” ja kolme tarkoittaa ”Olen niin surullinen tai onneton, etten kestä enää”. Numeroarvioita käytetään potilaan masentuneisuuden päättelemiseen (Beck ym. 1961; vrt. Aalto 2016).

Musteläiskät ja numeroarviot saattavat kuitenkin pian olla historiaa. Moni psykologi suuntaa nyt katseensa ihmisten digitaaliseen jalanjälkeen. Uusimpien tutkimusten mukaan persoonallisuutemme voidaan päätellä esimerkiksi siitä, millaisia kuvia jaamme sosiaalisessa mediassa. Masentuneisuutta ja jopa itsemurha-aikeita voidaan puolestaan ennustaa Facebook-päivitystemme kielestä, siis siitä, millaisia sanoja käytämme sosiaalisessa mediassa. Uusien mittatekniikoiden merkittävin mahdollistaja on koneoppiminen.

Teoriasta datamassoihin

Vuonna 1955 julkaistussa klassikkoartikkelissaan psykologit Lee Cronbach ja Paul Meehl esittivät, että psykologien käyttämien mittareiden on perustuttava psykologiseen teoriaan. Jos tutkija haluaa mitata vaikkapa persoonallisuuden piirteitä, on ensin muodostettava teoria siitä, miten eri persoonallisuuspiirteet näkyvät käyttäytymisessä, miten piirteet liittyvät toisiinsa ja niin edelleen. Cronbachin ja Meehlin mukaan yksi mittarin validiteetin kriteeri onkin se, miten hyvin mittarin tuottamat tulokset osuvat yksiin mitattavaa ominaisuutta käsittelevän teorian kanssa. Cronbachin ja Meehlin artikkeli tiivistää teorian tärkeyden konstruktiovaliditeetin käsitteeseen, joka on yhä yksi tärkeimmistä psykologisen mittarin arvioimisen kriteereistä. Muita kriteereitä ovat esimerkiksi mittarin ennustavuus ja toistettavuus, joista jälkimmäinen tarkoittaa eri mittauskerroilla saatujen vastausten yhtenevyyttä.

Vaikka Rorschach julkaisi mittarinsa ennen Cronbachin ja Meehlin klassikkoartikkelia, myös häntä motivoi ajatus, että musteläiskätestin tulosten perustelemiseksi tarvitaan psykologista teoriaa. Testin kehittelyn alkuvaiheessa Ror-

schach keskittyi lähinnä havainnoimiseen: Mitkä ovat yleisimpiä musteläiskissä nähtyjä hahmoja? Millaiset ihmiset keskittyvät musteläiskien yksityiskohtiin? Miten masennusdiagnoosin saaneet potilaat reagoivat läiskien väreihin? Rorschach ja hänen aikalaispsykologinsa kuitenkin himoitsivat havainnoille teoreettista selitystä. Selittäjä löytyi Rorschachin mukaan psykoanalyyttisesta teoriasta, josta Rorschach oli saanut vaikutteita testiä kehittäessään (Searls 2018).

Koneoppimiseen nojaavassa psykologisessa mittaamisessa teoria jää usein uupumaan. Sen sijaan ohjaksia pitelee alusta loppuun ”2000-luvun öljy” eli data. Menetelmiä datan hyödyntämiseen on monia, kuten paljon puhuttu neuroverkkoteknologia. Perusidea menetelmissä on se, että koneelle ei anneta hypoteeseja tai teorioita siitä, mitkä käyttäytymismallit ja vastaustavat liittyvät mihinkin psykologiseen ominaisuuteen. Sen sijaan koneoppimisalgoritmi saa suuren datamassan, josta sen annetaan itse muodostaa paras malli esimerkiksi persoonallisuuden ennustamiseen.

Käytännössä (ja yksinkertaistaen!) koneoppimiseen nojaava mittaus etenee esimerkiksi näin. Ensiksi tutkijat hankkivat datamassan, esimerkiksi 50 000 Facebook-käyttäjän ”tykkäykset” sekä numeroasteikkotestiin perustuvan arvion käyttäjien persoonallisuudesta. Tämä on niin sanotusti harjoitusdataa, josta kone opettelee yhteyksiä tykkäysten ja persoonallisuustyyppien välillä. Jos esimerkiksi suhteessa suuri määrä persoonallisuudeltaan avoimia ihmisiä tykkää Hello Kitty -tuotteista Facebookissa, kone oppii käyttämään Hello Kitty -tykkäystä sen merkkinä, että henkilö on todennäköisesti persoonallisuustyypiltään avoin (Kosinski, Stillwell ja Graepel 2013). Lopputulos yhteyksien opettelusta on monimutkainen malli, jonka perusteella ihmisen persoonallisuus voidaan ennustaa käyttäen ainoastaan Facebook-tykkäyksiä.

Halpaa ja hyvää dataa?

Perinteiseen psykologiseen mittaamiseen liittyy paljon haasteita, jotka koneoppimisen uskotaan ratkovan. Perinteiset haastattelut, liittyivätpä ne musteläiskiin tai numeroasteikkoihin, maksavat aikaa ja rahaa, varsinkin jos haastateltavia on paljon. Esimerkiksi masennuslääkekokeissa satoja koehenkilöitä haastatellaan sekä kokeen alussa että lopussa, jotta saadaan selville, miten potilaiden masentuneisuus muuttuu kokeen aikana. Tällainen tutkimus on kallista ja hidasta.

Toinen yleinen ongelma on vastausten vinoumat, eli vastaukset, jotka eivät heijasta sitä, mitä testillä halutaan mitata (esim. Saal, Downey ja Lahey 1980). Esimerkiksi persoonallisuustesteissä ihmiset saattavat kaunistella vastauksiaan sosiaalisesti hyväksyttävämmiksi. Toisin sanoen henkilö vastaa kysymykseen niin, että haastattelija, tutkija tai muu ulkopuolinen saa vastaajasta positiivisen kuvan. Kyselytuloksia voi vääristää myös vastaajan väsymys, tylsistyminen, valehtelu tai taipumukset, jotka eivät liity mitattavaan ominaisuuteen.

Koneoppimiseen nojaava mittaaminen välttää osan näistä perinteisen mittaamisen ongelmista. Suuri osa mittaamiseen soveltuvasta datasta on helposti ja ilmaiseksi saatavilla – ainakin toistaiseksi. Tutkijoiden ei tarvitse tuhlata aikaa ja rahaa laboratoriossa koehenkilöitä haastatellen, sillä netti on pullollaan Instagram-tykkäyksiä, Twitter-viestejä ja muuta nykyaikaista tutkimusdataa.

Digiajan menetelmät välttävät myös joitakin vastausten vinoumia. Yleensä netissä tehdyt persoonallisuustestit tuntuvat yksityisiltä, sillä haastattelija tai tutkija ei ole testitilanteessa fyysisesti läsnä. Tämä saattaa motivoida testin suorittajia rehellisyyteen. Digiajan psykologit myös pyrkivät tekemään testitilanteesta hauskan, jopa pelimäisen, tai muutoin motivoimaan vastaajia. Esimerkiksi Cambridgen yliopiston Discover My Profile -testisivulta testin suorittaja saa lopuksi palautteen vastauksistaan, siis esimerkiksi raportin omasta persoonallisuudestaan. Tämäkin voi motivoida laadukkaampaan ja keskittyneempään vastaamiseen.

Jos koneoppimisalgoritmille annettu harjoitusdata kuitenkin sisältää vinoumia, nämä vinoumat vaikuttavat ohjelmiston tekemien päätelmien laatuun. Kuvitellaan esimerkiksi, että monella masennuskyselyyn vastaavalla henkilöllä on taipumus joko liioitella tai vähätellä kokemaansa masentuneisuutta. Toisin sanoen henkilön vastaukset eivät kerro ainoastaan vastaajan masentuneisuudesta vaan myöskin hänen taipumuksestaan liioitteluun tai vähättelyyn. Jos tällaisia vinoutuneita vastauksia käytetään harjoitusdatana ohjelmiston opettamisessa, kone oppii tekemään vääristynei-

tä päätelmiä vastaajien masentuneisuudesta. Koneen tekemien ennusteiden ja päätelmien laatu on siis riippuvainen harjoitusdatan laadusta sekä sen kyselypatteriston laadusta, jolla harjoitusdata on kerätty.

Massavaikuttamisen väline

Mitä kaikkea koneoppiminen sitten mahdollistaa? Digitaalisen jalanjäljen mahdollistamat sovellutukset tulivat suurelle yleisölle tutuiksi viimeistään Cambridge Analytica -yrityksen tiedonkäyttöskandaalin myötä. Yritys hyödynsi Facebook-käyttäjien tietoja ja näiden perusteella tehtyjä psykologisia profiileja poliittisen mainonnan ja tiedotuksen kohdistamiseen Yhdysvaltojen presidentinvaalien yhteydessä. Toistaiseksi ei tiedetä, vaikuttiko Cambridge Analytican toiminta vaalien lopputulokseen.

Psykologiseen profilointiin perustuvaa massavaikuttamista on kuitenkin tutkittu muissa olosuhteissa. Vuonna 2017 julkaistussa tutkimuksessa Facebookissa esitettyjä kauneustuotemainoksia kohdennettiin sen mukaan, onko käyttäjä introvertti vai ekstrovertti (Matz ym. 2017). Kohdennus tehtiin siis sen mukaan, miten käyttäjä suhtautuu sosiaalisiin tilanteisiin. Introverteille näytetyissä mainoksissa vedottiin hiljaisuuden ja vetäytyneisyyden kaltaisiin ominaisuuksiin esimerkiksi mainoslauseella ”Beauty doesn’t have to shout” eli ”Kauneuden ei tarvitse olla äänekästä”. Ekstroverteille suunnatuissa mainoksissa puolestaan vedottiin esimerkiksi energisyyteen ja puheliaisuuteen, siis ekstroverteille ominaisiin ominaisuuksiin.

Tutkijoiden mukaan psykologisen profiilin mukaisesti kohdennetut mainokset olivat tehokkaampia kuin ei-kohdennetut mainokset. Toisin sanoen kauneustuotemainoksen nähnyt käyttäjä osti tuotteen todennäköisemmin, jos mainos oli kohdennettu käyttäjän psykologiseen profiiliin sopivaksi. Joissain tilanteissa koneoppimiseen perustuva psykologinen mittaaminen ja profilointi ovat siis tehokkaita massavaikuttamisen keinoja.

Vaalitulosten manipulointi ja meikkien myyminen eivät ehkä ole sellaisia yhteiskuntaa hyödyttäviä sovellutuksia, joita tieteeltä toivotaan. Onneksi psykologiaan sovellettu koneoppiminen sopii myös yhteiskunnallisen hyvän luomiseen. Harvardin yliopistosta hiljattain tohtoriksi väitellyt Andrew Reece kollegoineen tutkii, voiko Instagram-kuvia ja Twitter-viestejä käyttää psyykkisten sairauksien diagnosoimiseen ja ennustamiseen (Reece ja Danforth 2017; Reece ym. 2017). Kuvien kirkkautta ja väriä analysoimalla sekä viestien sävyä ja sanamäärää seuraamalla koneoppimisalgoritmi pyrkii päättelemään, onko kuvan tai viestin lähettäjällä masennus tai traumaperäinen stressihäiriö.

Reecen ja kollegoiden tutkimusten mukaan Instagram- ja Twitter-käyttäytymiseen perustuva diagnostiikka on toistaiseksi epätäydellistä. Alustavasti vaikuttaa kuitenkin siltä, että koneoppimismalleilla on vähintään yhtä hyvä menestys masennuksen diagnosoimisessa kuin yleislääkäreillä keskimäärin. Lisäksi sosiaaliseen mediaan perustuva diagnostiikka on verrattain edullista ja reaaliaikaista, ja se tavoittaa parhaimmillaan miljoonia netinkäyttäjiä. Onkin odotettavissa, että mallien tarkkuuden parantuessa nämä uudet menetelmät auttavat halvempien ja helposti saavutettavien mielenterveyspalveluiden rakentamisessa.

Mustan laatikon läpivalaisu

Vaikka koneoppimiseen nojaava psykologinen mittaaminen on kiehtovaa ja mahdollisesti tehokasta, liittyy siihen myös isoja eettisiä ongelmia. Niin yritykset, poliitikot, tutkijat kuin tavalliset netinkäyttäjätkin joutuvat miettimään datan keräämiseen, säilyttämiseen ja käyttöön liittyviä moraalisia kysymyksiä: Kuka saa kerätä tietoa kansalaisten nettikäyttäytymisestä? Miten tietoa saa hyödyntää? Pitääkö käyttäjälle kertoa, mihin tietoja hyödynnetään?

Toukokuussa 2018 nämä kysymykset nousivat jälleen julkisuuteen, kun uusi EU:n yleinen tietosuoja-asetus pantiin täytäntöön. Asetus velvoittaa yrityksiä ja muita toimijoita kertomaan asiakkailleen entistä tarkemmin siitä, mitä tietoa käyttäjistä kerätään, mihin tietoa käytetään ja kuinka kauan tietoa säilytetään (European Commission, 11.10.2018). Asetuksessa määrätään myös, että mikäli käyttäjästä kerättyä tietoa käytetään automatisoituun päätöksentekoon – esimerkiksi automaattiseen lainapäätökseen – on käyttäjällä oikeus tietää, mihin päätös perustuu.

Tietosuoja-asetuksen on tarkoitus suojella ihmisten yksityisyyttä ja muita oikeuksia. Toisaalta tiedonkäytön rajoitteet myös nostavat esiin uu-

sia ongelmia. Monet koneoppimismallit ovat niin monimutkaisia, etteivät tutkijat kykene tulkitsemaan, mihin mallin johtopäätökset tai ennustukset perustuvat. Tutkijat tietävät toki alkutilanteen eli käytetyn datan ja menetelmät sekä lopputuloksen eli mallin tuottamat ennustukset, mutta väliin jäävä prosessi on usein tuntematon. Koneoppimismalleista näkeekin usein käytettävän termiä *black box* eli musta laatikko, millä viitataan siihen, ettei niiden toiminnan logiikkaa tunneta – ainakaan toistaiseksi. Tietosuoja-asetukset pakottavat tutkijat uuden oppimisongelman eteen: miten kone oppii perustelemaan johtopäätöksensä?

Viimeaikaisessa tutkimuksessa päätelmien perusteluja on usein haettu samasta paikasta ja samoin keinoin kuin itse päätelmiäkin, eli datamassasta koneoppimisen avulla. Koneoppimisalgoritmi saattaa esimerkiksi poimia pitkästä tekstistä sanoja tai yksittäisiä lauseita, jotka vaikuttavat merkittävästi mallin antamiin ennustuksiin. Jos koneoppimismallin on tarkoitus päätellä, onko Twitter-käyttäjä masentunut, saattaa se perustella päätelmänsä korostamalla niitä sanoja, joilla oli suurin painoarvo masentuneisuuden ennustamisessa. Esimerkiksi runsas kieltosanojen (”ei”, ”ei koskaan”, ”älä”) ja joidenkin kirosanojen käyttö näyttää olevan yhteydessä masentuneisuuteen (Reece ym. 2017).

Rorschachin ja Cronbachin kaltaisille teorianhimoisille tutkijoille tällaiset selitykset tuskin riittävät, sillä ne eivät vastaa isoihin miksi-kysymyksiin. Miksi kieltosanojen käyttö ennustaa masentuneisuutta? Muokkaako kielenkäytön sävy ajatusten sävyä, vai kenties päinvastoin? Mittariskeptikko saattaa myös epäillä, että kieltosanojen käyttö ei liitykään masentuneisuuteen vaan vastaustyyliin, joka saa henkilön vaikuttamaan testin perusteella masentuneelta. Toisin sanoen: jos koneoppimismallin saama masennustestidata sisältää vinoumia, kuten liioittelua tai vähättelyä, on mahdollista että kielto- ja kirosanojen käyttö on yhteydessä näihin vinoumiin eikä niinkään masentuneisuuteen.

Kun koneoppimiseen nojaavien mittarien perusteella tehdään päätelmiä ja päätöksiä, on muistettava, että yleensä nämä uudet mittarit on ”harjoitettu” perinteisempien numeroasteikkoarvioiden ja testien avulla. Näiden perinteisten mittarien laadusta, tarkoituksenmukaisuudesta ja oikeasta validointitavasta on paljonkin tieteellistä väittelyä. Tästä voidaan pitää esimerkkinä mittaria nimeltä HRSD (Hamilton Rating Scale of Depression). HRSD on laajalti masennuslääkekokeissakin käytetty depressiokysely, jonka validiudesta ja hyödyllisyydestä kiistellään sen laajasta käytöstä huolimatta – tai ehkä juurikin sen takia (Bagby ym. 2004). Koneoppiminen ratkaisee joitakin psykologisen mittaamisen ongelmia, mutta teoriaa, kritiikkiä ja jatkuvaa arviointia tarvitaan jatkossakin.

Kirjallisuutta

Aalto, A. (2016). Beckin depressiokysely 21-osioinen (käyttö väestötutkimuksiin). Toimia-tietokanta, THL. http://www.thl.fi/toimia/tietokanta/mittariversio/83/. Alkuperäinen julkaisupäivä 26.1.2011, viitattu versioon, joka on päivätty 13.4.2016.

Bagby, R. M., Ryder, A. G., Schuller, D. R. ja Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? *American Journal of Psychiatry* 161 (12): 2163–77.

Beck, A.T., Ward, C. H., Mendelson, M., Mock, J. ja Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571.

Cronbach, L. J. ja Meehl, P. E. (1955). Construct Validity in Psychological Tests. *Psychological Bulletin* 52 (4): 281–302.

European Comission. ”What information must be given to individuals whose data is collected?” https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/what-information-must-be-given-individuals-whose-data-collected_en. Avattu 11.10.2018.

Kosinski, M., Stillwell, D. ja Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110 (15), 5802–5805.

Matz, S. C., Kosinski, M., Nave, G. ja Stillwell, D. J. (2017). Psychological Targeting as an Effective Approach to Digital Mass Persuasion. *Proceedings of the National Academy of Sciences of the United States of America* 114 (48): 12714–19.

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

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M. ja Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific reports*, 7, 13006.

Saal, F. E., Downey, R. G. ja Lahey, M. A. (1980). Rating the Ratings: Assessing the Psychometric Quality of Rating Data. *Psychological Bulletin* 88 (2): 413–28.

Searls, D. 2017. *The Inkblots: Hermann Rorschach, His Iconic Test, and the Power of Seeing*. Crown.

Kirjoittaja on tohtorikoulutettava Cambridgen yliopistossa.