

Big data: Uhka vai mahdollisuus

Big data on tullut viime vuosina tutuksi käsitteeksi, vaikka sen täsmällinen merkitys on yhä suurelle osalle hämärän peitossa. Sanaparina se ei viittaa mihinkään konkreettiseen ongelmaan tai ratkaisuun, vaan kuvaa yleisellä tasolla datamäärien kasvun ja monipuolistumisen luomaa haastetta ja näihin haasteisiin tarjottuja edistyksellisiä ratkaisuja.

Terveysteen ja hyvinvointiin liittyvää dataa kerätään yhä useammasta lähteestä ja mahdollisuudet kerätyn datan hyödyntämiseksi ovat samoin monipuolistuneet vauhdilla. Periaatteessa datan avulla saadaan täydellinen kuva ihmisen terveydentilasta menneisyydessä jo nyt. Yhdistämällä tämä tietämys malleihin, jotka on rakennettu suurten ihmismäärien datavirtojen ja -varantojen perusteella, saadaan luoduksi ennusteita. Näiden ennusteiden avulla voidaan ennustaa ja ennaltaehkäistä terveydellisiä ongelmia sekä lisätä kokonaisvaltaista hyvinvointia.

Yksinkertaisimmat esimerkit nojaavat yksittäisen datalähteen hyödyntämiseen. Sosiaalisen median datavirroista voidaan tunnistaa masentuneita tai itsetuhoisia aikeita hautovia. Veriarvojen avulla puolestaan saadaan seulotuksi ne, joilla on tai joita uhkaavat erilaiset elintasosairaudet kuten diabetes.

Haastavimmat esimerkit ovat eri datalähteitä yhdisteleviä. Monimuuttuja-analyysin avulla saadaan tarkkoja prognooseja, joiden perusteella voidaan ennustaa tarkasti sairauksien puhkeamistodennäköisyyksiä ja siten kohdentaa ennaltaehkäiseviä toimenpiteitä tehokkaasti. Mitä suurempia ja kattavampia data-aineistoja saadaan käyttöön, sitä parempi on ennustustarkkuus.

Data-aineistojen saatavuutta parantavat yhteiset ja keskitetyt tietokannat ja -varastot kuten Kanta.fi. Lisäksi mittausteknologioiden kehitys ja yleistyminen auttaa täsmällisen kokonaiskuvan luomisessa. Enää mittaaminen ei rajoitu ammattilaisten tekemiin pistokoeluontoisiin tarkastuksiin laboratorioissa tai lyhyen ajanjakson kokonaisseurantaan esimerkiksi sairaalassa. Yhä suurempi osa ihmisistä mittaa itse itseään ja tekee sitä jopa jatkuvasti.

Geenitiedossa piilee myös suuria mahdollisuuksia. Vielä vuosituhannen alussa ihmisen DNA:n kartoittamisen kustannukset olivat tähtitieteellisiä. Human Genome -projektin budjetti oli noin 3 mrd \$. Projekti kartoitti noin 1 700 ihmisen genomin. Nyt koko genomin kartoittaminen maksaa enää noin 1 000 \$ ja ennustettavissa on, että hinta tulee laskemaan huomattavasti lähivuosina. Kun yksikkökustannus on esimerkiksi 100 \$, on mahdollista esimerkiksi koko väestön kattavan geenitietopankin luominen. Suuri osa terveysongelmista on jollain tasolla perinnöllisiä. Ennaltaehkäisevälle työlle ja henkilökohtaiset erityispiirteet huomioivan hoitosuunnitelman tekemiselle tämä avaa aivan uusia mahdollisuuksia.

Kauaskantoisimmat vaikutukset tulevat olemaan tekoälyn kehittymisellä. Jo nyt ensi askeleita on ottanut mm. IBM:n Watson, jonka terveydenhuollolle kehitetyllä sovelluksessa saadaan automatisoiduksi diagnostiikkaa ja muuta perinteisesti korkeaa koulutusta ja asiantuntemusta vaativaa työtä. Datamäärien kasvaessa ja datalähteiden monipuolistuessa tulee yhä haasteellisemmaksi ihmisasiantuntijan ottaa huomioon oikeilla painoarvoilla kaikki mahdolliset muuttujat. Onkin mielenkiintoista nähdä, mikä on terveydenhuollon tilanne esimerkiksi vuonna 2050. Luultavasti emme osaa edes kuvitella, mikä on mahdollista tuolloin.

Immo Salo www.bigdata.fi