Laadukas data vaatii harkintaa

4.10.2021

Usein kuulee sanottavan, että ”data on uusi öljy”. Tällä viitataan data tuomaan lisäarvoon esimerkiksi tuotteiden ja palveluiden yhteydessä. Data ei ole uusi öljy, mutta se mahdollista parhaimmillaan uusien tuotteiden ja palveluiden synnyn. Tämä tarkoittaa, että myös datan laatuun tulee kiinnittää huomiota samoin tavoin, kuten tuotteiden ja palveluiden tuotannossa. Data tulee nähdä organisaation yhtenä resurssina, josta hyötyy koko organisaatio. Datan laatuun vaikuttaa useampi tekijä, joihin yritysten tulisi kiinnittää lähtökohtaisesti huomiota ennen kuin lähtee kehittää dataperustaisia tuotteita tai palveluita. Kuva 1 esittää datan laatuun vaikuttavia tekijöitä.



Kuva 1. Datan laatuun vaikuttavia tekijöitä (Sánchez et al., 2019)

Kuvassa 1 on esimerkkejä datan laatuun vaikuttavista tekijöistä. Datan täytyy olla saatavissa olevaa, joka ei ole itsestään selvyys, sillä osa tietojärjestelmistä saattaa olla suljettuja. Dataa tulisi pystyä käsittelemään ja analysoimaan, jotta yritys voisi tuottaa siitä esimerkiksi koneoppimismalleja osaksi omia tuotteitaan ja palveluita. Datan tulee olla riittävän tarkkaa ja relevanttia, jotta siitä on hyötyä – kaikki data ei ole hyödyllistä.

Käytettävyys on tärkeää. Datan tulee olla paikkansapitävää ja virheetöntä, jotta siihen voitaisiin luottaa esimerkiksi liiketoimintaa kuvaavien mittareiden yhteydessä ja koneoppimismalleissa.

Data itsessään ei tuo lisäarvoa – onhan se vain dataa. Datasta pitää luoda merkityksiä ja sitä kautta taas lisäarvoa asiakkaalle tai loppukäyttäjälle. Dataan liittyy lähes poikkeuksetta mallinnettavuus, josta on hyötyä muun muassa koneoppimissovellutuksia kehitettäessä. Jos data on huonolaatuista, niin lopputulos on myös huonolaatuinen. ”Garbage in, garbage out” pitää näin ollen paikkansa. Datan keräämistä harkittaessa tulee huomioida koko data-analytiikan prosessi, eli datan keräämisestä ei yksi ole hyötyä.

Ari Sivula
Tutkimus- ja kehittämispäällikkö, yliopettaja
SeAMK Tekniikka

*KTT Ari Sivula toimii Seinäjoen ammattikorkeakoulussa tutkimus- ja kehittämispäällikkönä sekä digitaalisten innovaatioiden yliopettajana. Lisäksi Sivula on tuotantotalouden dosentti (erityisalana datainnovaatiot) Vaasan yliopistossa. Sivula on julkaissut kymmeniä tieteellisiä artikkeleita erilaisissa tieteellisissä aikakausilehdissä sekä konferensseissa.***Lähteet**

Sánchez, R. Á., Iraola, A. B., Unanue, G. E., & Carlin, P. (2019). TAQIH, a tool for tabular data quality assessment and improvement in the context of health data. Computer methods and programs in biomedicine, 181, 104824.