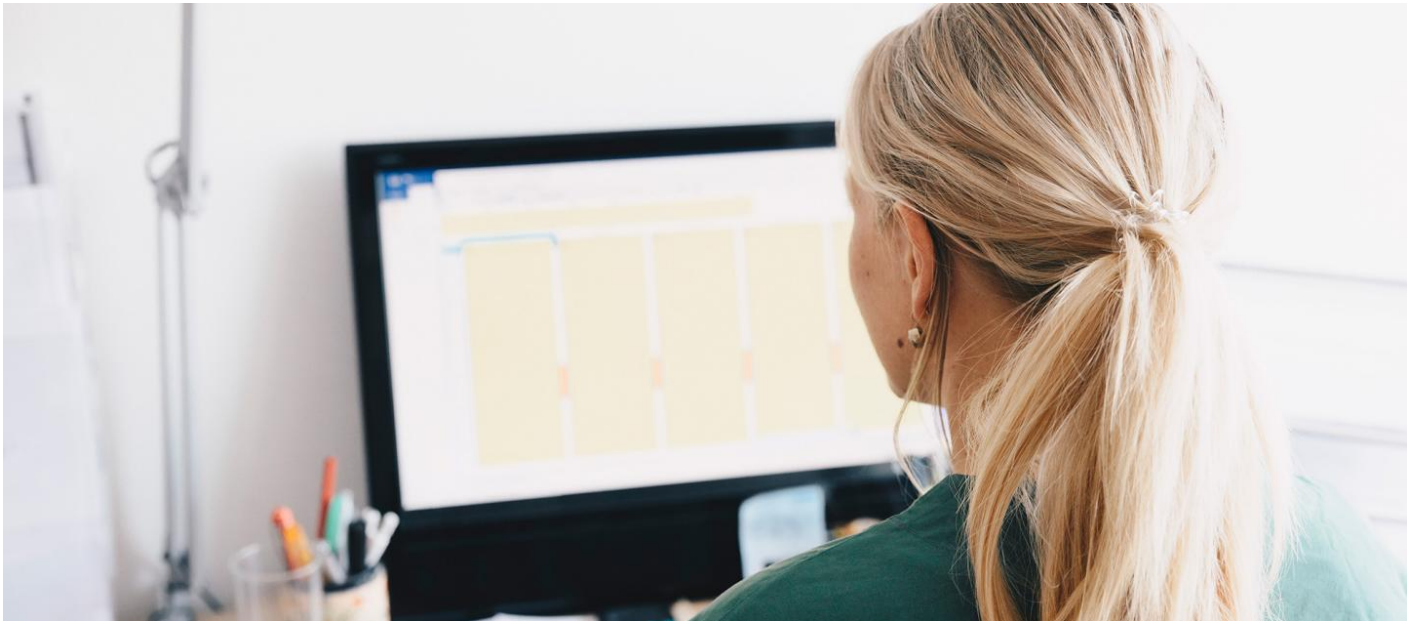


**BLOGI** 08.10.2021

# Koneoppimisen vahvuudet ja heikkoudet yhteiskuntatieteellisessä tutkimuksessa

Mullistaako koneoppiminen yhteiskuntatieteellistä tutkimusta? Mikä on sen rooli tulevaisuudessa? Vastauksia kannattaa etsiä pohtimalla, mihin tekoälyyn perustuvat menetelmät parhaiten soveltuvat ja missä muut menetelmät toimivat paremmin.

**Pekka Varje**

tutkimuspäällikkö

Koneoppiminen on tehnyt tuloaan yhteiskuntatieteelliseen tutkimukseen jo pitkään. Tekoälyyn liittyy paljon odotuksia siitä, kuinka kone analysoi valtavia aineistoja valmiiksi tutkijalle ja samalla muuttaa koko tieteen tekemisen tapaa.

Toistaiseksi koneoppimisen rooli on silti jäänyt yhteiskuntatieteellisen tutkimuksen kentällä suhteellisen pieneksi ja tulokset monelta osin alustaviksi.

Työsuojelurahaston rahoittaman ja Työterveyslaitoksen toteuttaman tutkimushankkeen ”Paremmalla ennakkoinnilla kestävämpään mielen hyvinvointiin työssä” lähtökohtana on ollut koneoppimisen

menetelmien soveltaminen sosiaalilääketieteelliseen mielenterveyden tutkimukseen. Keskenäinen hanke on antanut jo useita opetuksia näiden menetelmien vahvuuksista ja heikkouksista.

## Aineistolähtöinen tutkimus

Koneoppimisen menetelmät soveltuvat erityisen hyvin aineistolähtöiseen tutkimukseen. Ne eivät edellytä huolellisesti muotoiltuja hypoteeseja eli etukäteen tehtyjä oletuksia muuttujien välisistä suhteista.

Malliin voidaan syöttää lukemattomia ulottuvuuksia sisältävä valtava aineisto, josta tekoäly hetkessä poimii tutkittavan ilmiön kannalta olennaiset tekijät ja sulkee pois merkityksettömät tekijät. Siten sen etu perinteisiin tilastomenetelmiin verrattuna on tehokkuus piilevien ja yllättävien yhteyksien etsimisessä.

Koneoppimisen tuottamat tulokset ovat tilastomenetelmiin verrattuna kuitenkin vaikeampia tulkita. Koneen tekemät valinnat aineiston luokittelussa jäävät monilta osin ihmissilmältä piiloon.

Hypoteesien testaamisessa tilastomenetelmät puoltavat hyvin paikkaansa koneoppimisen rinnalla.

## Luonnollisen kielen prosessointi

Yhteiskuntatieteellisessä ja humanistisessa tutkimuksessa erityisiä odotuksia on kohdistettu tekoälyn kykyyn analysoida luonnollista ihmisten tuottamaa kieltä. Ajatus on, että kone saataisiin tuottamaan tietoa tekstimassoista, jotka ovat niin suuria, että kukaan ihminen ei voisi niitä käydä järjestelmällisesti läpi.

Myös omassa hankkeessamme on hyödynnetty koneoppimista luonnollisen kielen analysointiin. Käyttämämme aihemallinnus tunnistaa dokumenteista sanojen välisiä suhteita yli koko aineiston ja ryhmittelee niitä aiheiksi väsymättömällä tarkkuudella.

Ryhmittelyn jälkeen on mahdollista analysoida aiheiden esiintymistä aineistossa ajallisen, henkilöryhmittäisen tai muun ulottuvuuden perusteella. Aiheita voidaan hyödyntää jopa matemaattisten ennustemallien muuttujina.

Koneen tuottama analyysi luonnollisesta kielestä ei kuitenkaan ole samanlainen kuin se, joka syntyy ihmisen suorittamasta tekstin lähiluvusta.

Koneen tunnistama aihe koostuu sanoista, joiden yhdistävä tekijä on jakaumiin perustuva monimutkainen matemaattinen suhde eikä inhimillisesti tulkittava kieliopillinen suhde. Aiheet saattavat olla hankalia tulkita selkeän kertomuksen puuttuessa.

Aihemallinnus ei voi korvata perinteistä laadullista analyysiä, vaan tuo sen rinnalle tavoitteiltaan erilaisen analyysin.

## Koneoppiminen – tutkimuksen tulevaisuus?

Koneoppiminen tuo tieteelliseen tutkimukseen paljon uutta. Sen voi odottaa vakiinnuttavan paikkansa tutkimusmenetelmien joukossa.

Sen kyky käsitellä ja luokitella suuria ja monia ulottuvuuksia sisältäviä aineistoja on ylivoimainen. Koneoppiminen mahdollistaa myös monimutkaisten ja hienostuneiden ennustemallien tuottamisen.

Koneoppimisen ei kuitenkaan kannata odottaa korvaavan vanhoja menetelmiä, vaan tuovan niiden rinnalle yhden uuden tavan tehdä tutkimusta.

### Lisätietoja:

**[Paremmalla ennakkoinnilla kestävämpään mielen hyvinvointiin työssä, ENNAKKO](#)**

**[\(/tutkimus/hankkeet/paremmalla-ennakkoinnilla-kestavampaan-mielen-hyvinvointiin-tyossa-ennakko\)](#)** (Työterveyslaitoksen tutkimushankkeen esittely)

# KOMMENTOINTI

**Kommenttisi**

Mitä ajatuksia juttu herätti?

//

**Nimimerkki**

Vapaavalintainen nimimerkki

**Tallenna**

**KOMMENTIT (0)**