

HUOM! Tämä on alkuperäisen artikkelin rinnakkaistalenne. Rinnakkaistalenne saattaa erota alkuperäisestä sivutukseltaan ja painoasultaan.

Käytä viittauksessa alkuperäistä lähdettä:

Kauttonen, J., Khan, U., Määttänen, I. & Henttonen, P. (31.08.2023) Koneoppimisella tietoa sisusta: Hyödyllinen sisu ennakoi hyvinvointia ja työssä onnistumista. eSignals PRO.
<http://urn.fi/URN:NBN:fi-fe20230831114074>

PLEASE NOTE! This is an electronic self-archived version of the original article. This reprint may differ from the original in pagination and typographic detail.

Please cite the original version:

Kauttonen, J., Khan, U., Määttänen, I. & Henttonen, P. (31.08.2023) Koneoppimisella tietoa sisusta: Hyödyllinen sisu ennakoi hyvinvointia ja työssä onnistumista. eSignals PRO.
<http://urn.fi/URN:NBN:fi-fe20230831114074>



Copyright: © 2023 by the authors and Haaga-Helia University of Applied Sciences. Licensed under the terms and conditions of the Creative Commons Attribution (CC BY NC SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Koneoppimisella tietoa sisusta: Hyödyllinen sisu ennakoi hyvinvointia ja työssä onnistumista

Janne Kauttonen, Umair Ali Khan, Ilmari Määttänen & Pentti Henttonen

Sisukysely on hyvä mittari henkilön sitkeydelle ja stressinsietokyvyille. Keräämämme aineiston perusteella näyttää siltä, että sisulla on myös laajempi yhteys henkilön hyvinvointiin, työelämään ja jopa siihen, miten ihminen sanoittaa omia ajatuksiaan ja oloaan. Olemme ”Sisu Työelämässä” -hankkeessa selvittäneet sisun yhteyttä hyvinvointiin ja jaksamiseen.

Tässä kirjoituksessa käymme läpi tuloksia joita saimme analysoimalla sisu-kyselyaineistoa koneoppimismenetelmillä. Laajennamme näin aikaisempaa, pitkälti perinteiseen tilastolliseen analyysiin perustuvaa tutkimusta muun muassa soveltamalla tekstianalyysiä avoimiin vastauksiin. Samalla demonstroimme miten koneoppiminen voi tukea perinteistä kyselyaineiston tilastollista analyysiä. Tulostemme perusteella sisukysely on hyödyllinen työkalu ennustamaan henkilön hyvinvointia ja suoriutumista työssä.

Lineaaristen mallien tuolla puolen – koneoppimismallit sisun tutkimuksessa

Aikaisemmassa tutkimuksessa (Henttonen ym. 2022) kehitettiin kuuden mittarin sisu-skaala ja osoitettiin sen vahva yhteys henkilön hyvinvointiin ja terveyteen (WHO-5 ja CES-D mittarit), työstressiin (ERI mittari) ja koettuun stressiin. Tulokset osoittivat että hyödyllinen sisu korreloi positiivisesti hyvinvoinnin kanssa. Haitallisen sisun kohdalla yhteys oli päinvastainen, minkä lisäksi löydettiin sille positiivinen korrelaatio työstressin ja yleisen stressin välillä. Linearisissa regressiomalleissa sisulla oli huomattava selittävä vaikutus sekä positiiviseen, että negatiiviseen hyvinvointiin. Sisulla oli vahva yhteys työstressiin.

Sisulla on kuitenkin vaikutusta myös laajemmin yksilön hyvinvointiin ja suoriutumiseen. Halusimme laajentaa aikaisempaa tutkimusta ja testata ilmeneekö sisu jollain tavalla myös henkilön kirjallisessa ilmaisussa. Todellisen maailman ilmiöt ovat harvoin lineaarisia ja muuttujien väliset vuorovaikutukset ovat yleensä monimutkaisia, joten otimme datan analyysissä avuksi koneoppimismallit.

Toisin kuin lineaariset mallit, käyttämämme gradient boosting -koneoppimismalli sallii muuttujien väliset monimutkaiset riippuvuussuhteet, puuttuvat havainnot, sekä yliparametrisoidut mallit, joissa on enemmän muuttujia kuin aineistossa havaintoja. Toisin kuin perinteiset tilastolliset mallit, koneoppimismallit testataan aina erillisillä testiaineistolla, jolloin voidaan aidosti varmistaa mallien yleistvyys ja erityisesti ennustekyky uudessa aineistossa. Tässä työssä käytimme 80 prosenttia aineistosta (näytteistä) mallien opetukseen ja loput 20 prosenttia mallien suorituskyvyn testaukseen.

Kyselyaineistojen yhteydessä on yleensä tapana kerätä myös avoimia vastauksia, joiden kanssa tyydytään tyypillisesti laadulliseen analyysiin. Koneoppiminen tarjoaa kuitenkin tehokkaita tapoja tutkia tätäkin osaa aineistosta määrällisesti. Erityisesti suurten kielimallien (esim. ChatGPT) avulla tekstimuotoinen syöte voidaan helposti muuntaa numeeriseen muotoon siten, että tekstin sisältämä informaatio säilyy. Tämä on huomattava ero siihen, että tutkittaisiin tekstin erillisiä osia (esim. sanoja tai sanatyyppejä) tai kuvailevia tilastollisia muuttujia (esim. merkkien tai sanojen määrä). Tässä analyysissä hyödynsimme tekstianalyysiä siinä, miten vastaajat sanoittivat oman hyvinvointinsa ja milloin he kokivat parhaaksi antaa periksi ja luovuttaa.

Laaja kyselyaineisto sisusta, hyvinvoinnista ja taustamuuttujista

Tutkimusta varten keräsimme laajan kyselyaineiston, jonka tässä työssä analysoitava osa koostui 455 näytteestä ja 41 muuttujasta (piirteestä). Näistä 41:stä muuttujasta 35 sisälsi erilaisia demografisia, työhön ja hyvinvointiin liittyviä kysymyksiä ja loput 6 mittasivat sisua. Kaikkiaan 15:ssä muuttujassa oli puuttuvia havaintoja siten, että havainnoista puuttui 20-50 prosenttia. Tämä on hyvin tavallinen tilanne, kun käsitellään todellista (ei simuloitua) aineistoa. Sen sijaan, että heittäisimme pois puuttuvat havainnot, käytimme tässä työssä menetelmää, jossa puuttuvat arvot rekonstruoidaan. Tämä on kuvattu seuraavassa osiossa.

Sisu skaala koostuu seuraavasta kuudesta mittarista (suluissa englanninkieliselle käännökselle pohjautuva lyhenne): piilevä voima (LP), toimintasuuntautuneisuus (AM), poikkeuksellinen periksiantamattomuus (EP), haitat ajattelulle (HR), haitat itselle (HS) ja haitat muille (HO). Näistä kolme ensimmäistä muodostavat ns. hyödyllisen sisun ja loput ovat ns. haitallista sisua. Tarkempi kuvaus mittaristosta löytyy alkuperäisestä tutkimusartikkelista (Henttonen ym. 2022). Muut 35 muuttujaa jakautuivat neljään kategoriaan seuraavasti:

1. **Demografiset muuttujat ja työllisyys:** Sisältää tietoja iästä, sukupuolesta, koulutustasosta ja työtilanteesta, kuten kokoaikaisesta tai osa-aikaisesta työllistymisestä, koulutuksesta, vuosilomasta, lomautuksista, työttömyydestä, eläkkeelle siirtymisestä, opiskelijastatuksesta ja muista työtyypeistä. Lisämuuttujia tässä klusterissa ovat työasema, työllisyysjakso ja viikoittaiset työtunnit.
2. **Työympäristö ja -olosuhteet:** Työpaikan olosuhteet, jotka vaikuttavat työntekijän kokemukseen työstä. Sisältää asioita kuten yhteispäätöksenteon vaikutus, ongelmanratkaisun tuki kollegoilta, etätyöjärjestelyt ja työsuorituksen taso. Lisäksi muuttujaryhmä korostaa työpaikan tasa-arvon merkitystä, sisältäen muuttujia, jotka mittaavat vahvaa sukupuoli- ja työntekijöiden oikeudenmukaisuutta sekä urakehitysmahdollisuuksia.
3. **Hyvinvointi ja terveys:** Mittaa yksilön kokonaisvaltaista hyvinvointia, mielenterveyttä ja fyysistä terveyttä sekä tyytyväisyyttä elämään. Se sisältää muuttujia, jotka osoittavat yksilön stressitasoja, taloudellista stressiä, yleistä elämäntyytyväisyyttä, terveydentilaa ja unihäiriöitä. Klusteri kattaa myös fyysisen harjoittelun tottumukset ja ulkoisiin tekijöihin liittyvät huolenaiheet, kuten Ukrainan sotaan liittyvät huolet. Lisäksi se sisältää muuttujia, jotka liittyvät ponnistelu-palkkio -epätasapainoon (ERI mittari) ja maailman terveysjärjestön hyvinvointi-indeksiin (WHO-5 mittari).

4. **Avoimet tekstivastaukset:** Kaksi avointa kysymystä: ”Mistä ihmiset tietävät, milloin kannattaa luovuttaa tai jatkaa jonkin päämäärän tavoittelua?” ja ”Miten olet voinut viimeisen kahden viikon aikana?”. Näistä ensimmäinen liittyy murtumispisteen tunnistamiseen ja toinen omaan hyvinvointiin. Molempiin kysymyksiin vastauksena saatiin tekstiä, joiden pituudet vaihtelivat muutamasta sanasta useisiin lauseisiin.

Seuraavaksi käymme läpi puuttuvien arvojen ja avointen tekstivastausten käsittelyn, gradient boosting -mallin ja saadut tulokset.

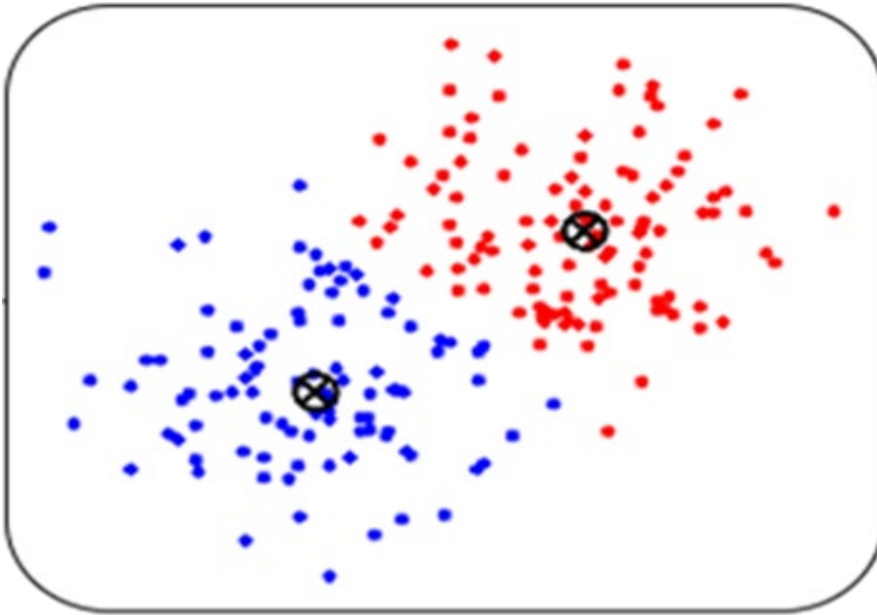
Puuttuvien arvojen rekonstruointi

Perinteisessä tilastollisessa analyysissä datan puuttuvat arvot on poistettava, jolloin jäljelle jäävä aineiston koko voi kutistua merkittävästi. Vaihtoehtoisesti puuttuvat havainnot voidaan yrittää korvata keskiarvolla (numeeriset muuttujat) tai yleisimmällä luokalla (kategoriset muuttujat). Tämä on kuitenkin ongelmallista, koska tällöin havainnot eivät ole enää toisistaan riippumattomia. Koneoppimismallien kanssa asia voidaan ratkaista aineistoa mallintamalla. Tässä analyysissä käytimme Random Forest -algoritmia ennakoimaan puuttuvaa aineistoa olemassa olevan aineiston perusteella (Stekhoven et al. 2012). Idea on se, että muuttujien välillä on korrelaatioita, jotka auttavat arvaamaan mitä puuttuvat luvut todennäköisesti ovat. Näin ollen hyödynsimme analyysissä koko aineiston (455 näytettä).

Avointen tekstivastausten ryhmittelyanalyysi

Kahdesta kysymyksestä ensimmäinen liittyi luovuttamiseen (kysymys A) ja toinen omaan hyvinvointiin (kysymys B). Avoimia tekstivastauksia analysoidaan perinteisesti joko laadullisin keinoin, missä tutkija tulkitsee tekstit valitusta näkökulmasta, tai laskemalla sanojen esiintymistiheyksiä (esim. sanapilvi). Suuret kielimallit mahdollistavat kuitenkin tekstin muuntamisen numeeriseksi siten, että tekstiä voidaan analysoida puhtaasti laskennallisesti ilman merkittävää informaation menetystä. Muunnos voidaan tehdä esimerkiksi ChatGPT kielimallilla. Tässä työssä käytimme muunnokseen BERT kielimallia (Reimers et al. 2019), joka ei vaadi syötesuunnittelua.

Muunsimme tekstit ensin numeerisiksi 768-ulotteiseksi vektoreiksi, jotka edelleen muunnettiin kahteen ulottuvuuteen UMAP-mallilla (McInnes et al. 2018). Tämä on hyödyllistä, koska liian korkea datan ulottuvuus haittaa ryhmittelyä. Lopuksi nämä kaksiulotteiseksi pisteiksi muunnetut tekstit ryhmiteltiin K-means-klusterointialgoritmillä. Lopputulos on visualisoitu kuvaan 1. Tekstiryhmien tarkastelumme paljasti, että tekstien ryhmittely toimi parhaiten kahdelle ryhmälle. Ryhmät jakautuivat kysymyksen A osalta siten, kuinka hyvin henkilö osasi sanoittaa ja tunnistaa murtumispisteen ja omat rajansa. Kysymyksen B osalta vastaukset osuivat selkeästi positiivinen-negatiivinen -akselille koetun hyvinvoinnin mukaan.



Kuva 1. Avointen tekstivastausten ryhmittely kahteen osaan. Jokainen piste adustaa yhtä avointa tekstivastausta.

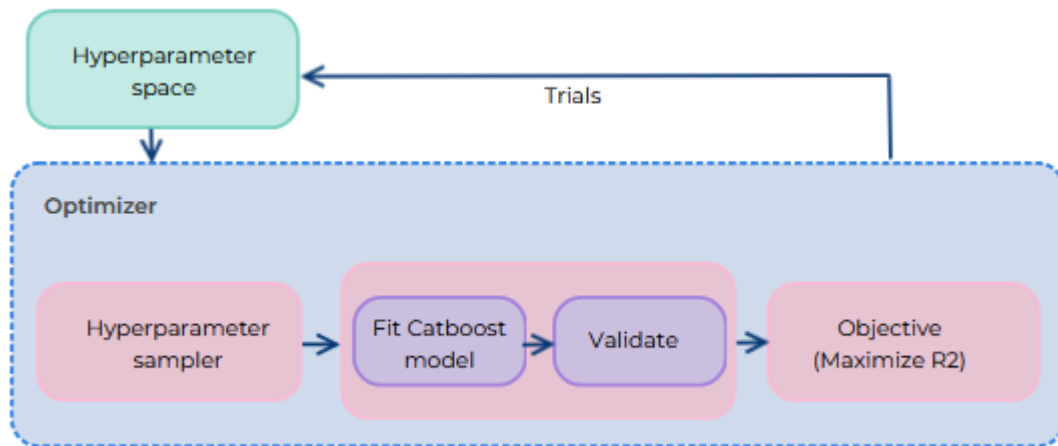
Tämän tuloksen perusteella päätelimme, että kahteen ryhmään jako on toimiva ja tekstinäytteet annotoitiin kolmen riippumattoman henkilön toimesta edellisten havaintojen mukaisesti kahteen kategoriaan. Annotoinnilla varmistimme, että klusterointi tapahtuu valittujen kriteerien perusteella. Ryhmittelyanalyysiä käytettiin näin ollen eksploratiivisen analyysin tukena, mikä on tyypillistä datan mallintamisessa. Annotoinnissa kolme asiantuntijaa luki kaikki avoimet vastaukset ja merkitsivät kumpaan kategoriaan tekstit heidän mielestään kuuluivat. Annotoijat olivat pitkälti yksimielisiä (riippumattomuutta kuvaavat kappi-arvot 0.71 ja 0.81 kysymyksille A ja B, eli yli suositellun 0.5). Tämän jälkeen data-analyysiä jatkettiin ohjatun oppimisen menetelmin luokitteluanalyysillä.

Gradient boosting menetelmä ja selitettävät mallit

Regressio- ja luokitteluanalyyseihin käytimme Catboost-algoritmia, joka on gradient boosting-tyyppinen menetelmä ja tukee useita erilaisia ohjatun koneoppimisen tehtäviä, joista tärkeimpänä regressio ja luokittelu (Dorogush et al. 2018). Gradient boosting-koneoppimisen menetelmänä on sovittaa peräkkäin uusia malleja (tässä työssä päätöspuita), jotka tuottavat entistä parempia ennusteita vastemuuttujasta (Friedman et al. 2001). Gradient boosting menetelmät ovat osoittautuneet käytännössä erittäin tehokkaiksi, koska ne toimivat hyvin jo pienellä näyttemäärällä ja sietävät hyvin sotkuistakin aineistoa, toisin kuin useimmat lineaariset mallit.

Koneoppivien mallien sovitusta vaatii yleensä myös ns. hyperparametrien asettamista, jotka ohjaavat oppimisprosessia. Hyperparametrit liittyvät mallin muun muassa oppimisnopeuteen (eng. learning rate), regularisaation voimakkuuteen ja mallien monimutkaisuuteen. Tässä työssä hyperparametrit säädettiin opetusdatalla käyttämällä

kuvan 2 mukaista silmukkaa, joka maksimoi mallin selityksasteen eli korrelaatiokertoimen neliön (R^2 luku).



Kuva 2. Catboost-mallin hyperparametrien säätäminen osana opetusta.

Toisin kuin lineaariset mallit, Catboost-menetelmän tuottama malli on epälineaarinen, jonka vuoksi mallin tulosten tulkinta on hankalampaa, eikä onnistu suoraan mallin parametreja katsomalla. Tähän on kuitenkin kehitetty menetelmiä, joista eräs tunnetuin ja käytetyin on SHAP-menetelmä (Lundberg, 2020). SHAP, eli SHapley Additive exPlanations, perustuu peliteoriaan, jossa datan näytteet ovat eräänlaisia ”pelejä” ja selittävät muuttujat ”pelaajia”. Pelaajien vaikutus voidaan kvantifioida SHAP-lukuina, joiden yhteenlaskettu tulos yli kaikkien piirteiden muodostaa lopullisen mallin ennusteen. Selitettävien muuttujien vaikutus selitettäviin muuttujiin on mitattavissa siitäkin huolimatta, että mallissa on vahvasti epälineaarisia tekijöitä ja keskinäisiä vuorovaikutuksia. Seuraavaksi käymme läpi mallinnuksen tuloksia ja hyödynnämme tässä SHAP lukuja.

Sisu ennakoi sanoitettua murtumispistettä ja hyvinvointia

Aluksi kysymme miten henkilön sisu ilmenee avoimissa kirjoitetuissa vastauksissa. Eli toisin sanoen ovatko tietyt sisun osa-alueet oleellisia siinä, miten henkilö määrittelee periksiantamisen (kysymys A) ja miten hän kokee oman hyvinvointinsa (kysymys B). Opetimme aineistolla luokittelumallin, jonka tarkkuudeksi saatiin lopulta 84.6 prosenttia (kysymys A) ja 73.8 prosenttia (kysymys B). Molemmat ovat sattumaa parempia tuloksia. SHAP-arvojen perusteella lasketut selittävien muuttujien tärkeydet on visualisoitu liitetiedoston kuvaan 3.

SHAP arvoista voidaan tulkita seuraavaa. Kysymykselle A tärkeimmät kolme selittävää muuttujaa olivat ikä ja sisun kaksi mittaria (EP ja HR). Suurempi periksiantamattomuus (EP) ennakoi positiivisempaa vastausta ja haitat ajattelulle negatiivisempaa vastausta (HR). Mitä vanhempi vastaaja oli, sitä positiivisempia olivat hänen vastauksensa. Kysymykselle B tärkeimmäksi osoittautui sisukkuuden kaksi mittaria (LP ja EP), sekä ikä. Suurempi piilevä voima (LP) ja pienempi poikkeuksellinen päättäväisyys (EP) ennakoivat positiivisempaa olotilaa. Sukupuolen ja koulutuksen merkitys olivat vähäisiä molemmissa

tapauksissa. Erityisesti ensimmäisessä kysymyksessä (A) korostuivat haitallisen sisun muuttajat ja jälkimmäisessä hyödyllisen.

Sisulla on vahva yhteys hyvinvointiin

Seuraavaksi kysyimme mitkä tekijät ennakoivat henkilön hyödyllistä ja haitallista sisua. Tässä analyysissä selitettäviä muuttujia oli kaksi: hyödyllinen sisu ja haitallinen sisu, jotka molemmat sisälsivät kolme sisun mittaria yhteenlaskettuina. Loput 35 muuttujaa olivat selittäviä, eli tutkimme näiden muuttujien ennustevoimaa sisulle. Parhaan mallimme selitysaste (R^2) oli 0.19, jota ei voi pitää erityisen korkeana, mutta kuitenkin satunnaista tulosta parempana. Liitetiedoston kuvassa 4 on mallille lasketut SHAP-arvot.

Hyödyllisen sisun osalta voimme todeta, että ylivoimaisesti vahvin selittävä muuttuja on yleinen hyvinvointi WHO-5 -mittarilla siten, että mitä parempi hyvinvointi, sitä suurempi on henkilön hyödyllinen sisu (ja toisinpäin). Muista muuttujista suurempaa hyödyllistä sisua ennakoivat parempi työsuorite (henkilön oma arvio) ja yleinen tyytyväisyys elämään. Haitallista sisua puolestaan ennakoivat eniten suurempi panostus työhön ja suuri koetun stressin määrä. Kokonaishyvinvoinnin vaikutus on käänteinen, mutta vaikutus on huomattavasti pienempi kuin hyödylliselle sisulle.

Sisu ennakoivat hyvää suoriutumista työssä

Lopuksi mallinsimme miten sisu ennakoivat hyvää suoriutumista työssä. Selitettävä muuttuja oli työssä suoriutumisen arviointi erinomaiseksi (Likert-asteikolla 5/5) verrattuna sitä huonompiin arvioihin (Likert-asteikolla 1/5-4/5) vastaajan itsensä arvioimana. Opetetun luokittelumallimme tarkkuus oli 56.7 prosenttia ja SHAP-arvot löytyvät liitetiedoston kuvasta 5. Tärkeimmät viisi selittäviä muuttujaa olivat nyt kaikki kuusi sisu mittaria. Tärkeimpänä oli haitat muille ja haitat ajattelulle, joiden suuret arvot ennakoivat huonompaa työsuoritusta. Korkea piilevä voima ja periksiantamattomuus puolestaan ennakoivat hyvän työsuorituksen arviota. Tulos osoittaa sisun hyvän ennustevoiman kokemukseen omassa työssä onnistumisessa.

Yhteenveto

Sovelsimme koneoppimismenetelmiä tutkiessamme miten sisu liittyy laajaan joukkoon erilaisia työtä, terveyttä ja hyvinvointia mittaavia piirteitä. Analyysimme paljasti, että hyödyllisen sisun tekijät ennakoivat tyytyväisyyttä elämään ja parempaa terveydentilaa, kun taas haitallinen sisu yleensä heikentää näitä. Voimme päätellä, että sisu on käyttökelpoinen mittari ennakoimaan henkilön yleistä hyvinvointia ja suoriutumista työssä. Löysimme sisulle myös selkeän yhteyden henkilön tuottamiin kirjallisiin vastauksiin liittyen periksiantamiseen ja omaan hyvinvointiin. Sisun kaikki kuusi mittaria olivat mallien kannalta informatiivisia kukin eri yhteyksissä.

Tulokset myös havainnollistavat miten koneoppimista, erityisesti regressiota, luokittelua ja tekstianalytiikkaa, voidaan soveltaa kyselydatan analyysissä perinteisen tilastoanalyysin rinnalla. SHAP ja vastaavat menetelmät ovat tärkeä osa koneoppimista ja mallien tulosten

tulkintaa. Ilmiön tarkan ennustamisen lisäksi on vähintään yhtä tärkeä tietää miksi malli tekee juuri tietyn ennusteen.

Tutkimuksemme havainnollistaa, että koneoppimisella (sekä laajemmin tekoälyllä), voidaan tutkia ilmiötä laajasta näkökulmasta ottamalla mukaan runsaasti muuttujia. Tämä voi auttaa pääsemään eroon ns. litteän maan harhasta, jossa monimutkaista ilmiötä yritetään selittää liian pienellä määrällä muuttujia. Tämä on edelleen haaste etenkin monissa käyttäytymistieteen tutkimuksissa, jotka soveltavat perinteistä tilastoanalyysiä (Jolly & Chang 2019; Kauttonen & Suomala 2021). Toisaalta koneoppiminen, sen paremmin kuin perinteinen tilastollinen analyysikään, eivät pysty vastaamaan siihen mikä on terveyden, työhyvinvoinnin ja stressitason kausaalinen suhde sisuun. Miten henkilön raportoima sisu riippuu esimerkiksi elämäntilanteesta, terveydentilasta, iästä ja ajankohdasta? Tämän selvittäminen vaatii laajempaa pitkittäisaineistoa, jolloin päästään kunnolla kiinni myös riippuvuussuhteisiin.

Tutkimus on osa Euroopan sosiaalirahaston (ESR) rahoittamaa Sisätyöelämässä -hanketta, jonka tavoitteena on parantaa työntekijöiden hyvinvointia ja jaksamista. Hankkeessa seurataan ihmisten hyvinvointia ja hyvinvointiin vaikuttavien piirteiden muutoksia ja etsitään keinoja vaikuttaa myönteisesti hyvinvoinnin kehitykseen.

Lähteet

Dorogush, A. V., Ershov, V. and Gulin, A. (2018). CatBoost: gradient Boosting with categorical features support. arXiv preprint arXiv:1810.11363 .

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics: 1189-1232.

Henttonen, P., Määttä, I., Makkonen, E., Honka, A., Seppälä, V., Närväinen, J., ... & Lahti, E. E. (2022). A measure for assessment of beneficial and harmful fortitude. Heliyon 8(11), e11483.

Jolly, E., and Chang, L. J. (2019). The fatland fallacy: moving beyond low-dimensional thinking. Top. Cogn. Sci. 11, 433–454. doi: 10.1111/tops.12404

Kauttonen, J. & Suomala, J. (2021). [Tekoäly parantaa käyttäytymisen ennustettavuutta](#). Laurea Journal.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., ... & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1), 56-67.

McInnes, L., Healy, J. and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

Stekhoven, D. J., and Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28.1: 112-118.

LIITETIEDOSTO: [Koneoppimisella tietoa sisusta: Hyödyllinen sisu ennakoi hyvinvointia ja työssä onnistumista](#)