



Aalto-yliopisto
Perustieteiden
korkeakoulu

Hyvinvointialueiden stokastinen hierarkkinen klusterointi Kansallisen terveystiedon pohjalta (aihe-esittely)

Iina Manninen

22.11.2024

Ohjaaja: *Leevi Olander*

Valvoja: *Ahti Salo*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Tausta

- Työkyvyttömyys ja sairauspoissaolot aiheuttavat merkittäviä inhimillisiä ja taloudellisia menetyksiä.
- Väestön ikääntyminen lisää painetta vähentää poissaoloja.
- Sairastavuus ja työkyvyttömyys vaihtelevat alueittain.

Yhä useampi nuori aikuinen jää työkyvyttömyyseläkkeelle

Kela | Yhteensä alle 35-vuotiaita työkyvyttömyyseläkettä saavia oli vuoden 2023 lopussa noin 25 500.



Yhteensä alle 35-vuotiaita työkyvyttömyyseläkettä saavia oli vuoden 2023 lopussa noin 25 500.
Kuva: Heikki Saukkomaa / Lehtikuva

HS (2024)

Mielenterveysongelmat veivät jo yli 100 000 suomalaista pitkälle sairauspoissaololle vuonna 2023

Julkaisu 19.1.2024

Sairauspäiväraha mielenterveysyhtiä saaneiden määrä kasvoi vuonna 2023 ennätyslukemiin: yli sataanuhanteen henkilöön. Kasvu johtuu erityisesti ahdistuneisuushäiriöistä. Sairauspäiväraha saaneiden kokonaismäärä vähenyi.

Mielenterveysongelmat veivät yli 100 000 suomalaista sairauspäivärahalle

Sairauspäiväraha saaneiden lukumäärä yksittäisissä sairausjärjestelmissä vuosina 2010–2023



Työkyvyttömyys kohtaa useammin pohjois- ja itäsuomalaisia kuin etelässä ja lännessä asuvia suomalaisia

Terveys ja sairastavuus jakautuvat Suomessa maantieteellisesti eriarvoisesti, ilmenee uudesta kansallisesta terveysindeksistä. Pohjoisessa ja idässä asuvat sairaimmat ja etelässä ja lännessä terveimmät suomalaiset.



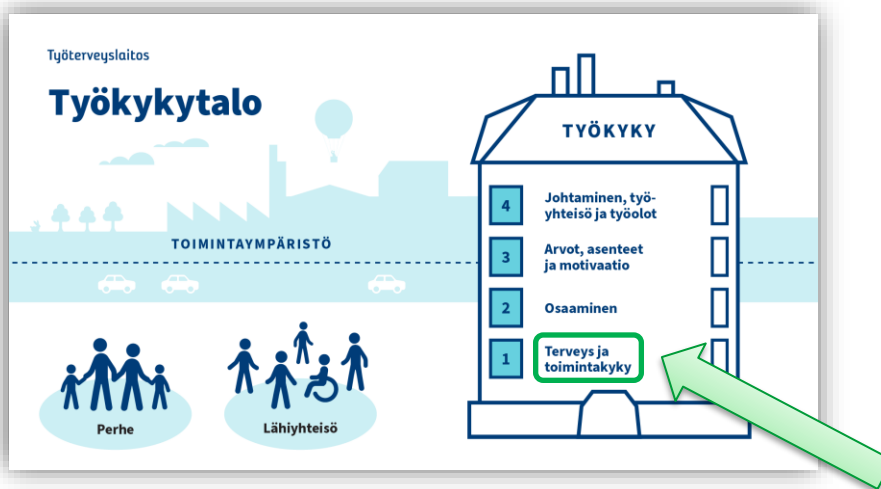
Työkyvyttömyys on yleisintä muualla Suomessa Pohjois-Pohjanmaalla ja Kainuussa. Kuvasse ilmakuva Oulun keskustan aluetta. Kuva: Paulus Märkkö / Yle

Yle (2023)

Kela (2024)

Työkyky ja sairastavuus

- Väestön ikärakenne, työllisyystilanne sekä koulutus- ja tulotaso vaikuttavat sairastavuuteen.



Työterveyslaitos TTL (2023)

- Myönnetyn työkyvyttömyyseläkkeen taustalla on aina sairauden lääketieteellinen diagnoosi.

Fyysinen, psyykinen ja sosiaalinen toimintakyky sekä terveys muodostavat perustan työkyvylle.

Tavoitteet ja rajaukset

- Klusteroidaan Suomen hyvinvointialueet sairastavuuden ja työkyvyttömyyden perusteella.
- Alueiden erot ovat pieniä, joten simuloidaan klusterointia useita kertoja.
- Hyödynnetään Kansallisen terveystieteiden sairausryhmäkohtaisia indeksejä sekä työkyvyttömyysindeksiä vuodelta 2022.

Suomen hyvinvointialueet



Sosiaali- ja
terveysministeriö (2024)

Kansallinen terveysindeksi

- Kansallinen terveysindeksi kokoaa tietoa väestön terveydestä sekä toiminta- ja työkyvystä eri alueilla.
- Indeksien suurempi arvo tietyllä alueella viittaa maan keskiarvoa yleisempään sairastavuuteen tai työkyvyttömyyteen.

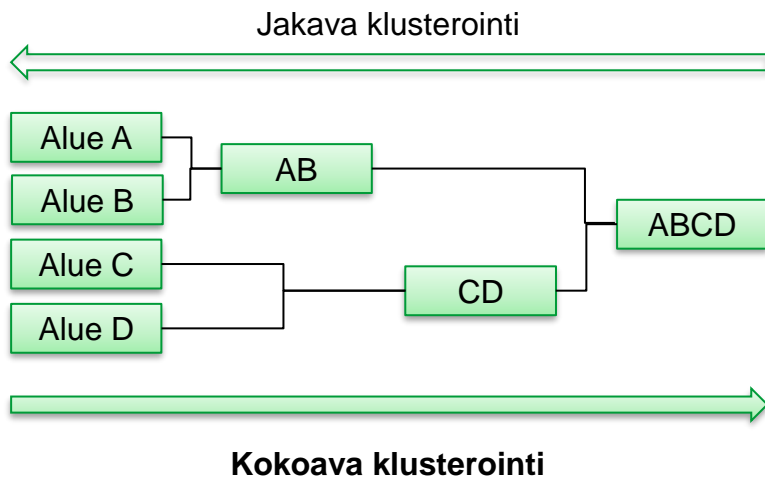
Kansallinen terveysindeksi



Mukaiillen: Terveyden ja hyvinvoinninlaitos THL (2024)

Stokastinen hierarkkinen klusterointi

- Klusterointi on ohjaamattoman oppimisen menetelmä, jonka avulla objektit jaetaan ryhmiin samankaltaisten piirteiden perusteella.



- Hyödynnetään kokoavaa hierarkkista klusterointia, jossa jokaisella iteraatiolla kaksi klusteria yhdistetään toisiinsa.

Stokastinen hierarkkinen klusterointi

Hyödynnetään esimerkiksi euklidista etäisyyttä:

$$d^2(x_i, x_j) = (x_i - x_j)^T (x_i - x_j)$$

Mahdollisia linkkifunktioita:

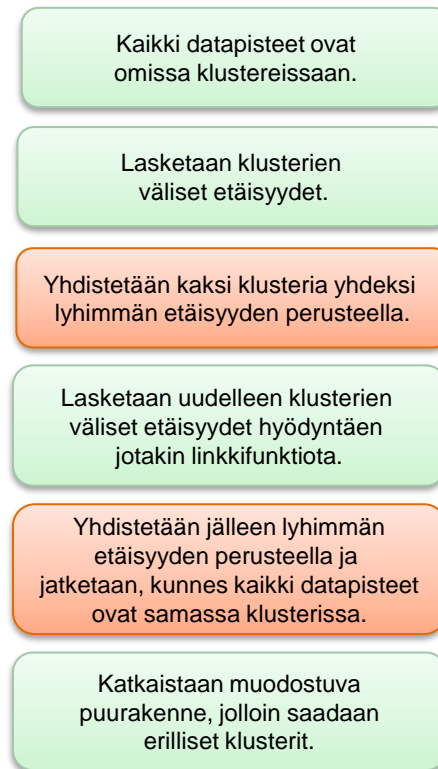
$$d(A, B) = \min d(x_i, x_j)$$

$$d(A, B) = \max d(x_i, x_j)$$

$$d(A, B) = \frac{1}{n_A n_B} \sum_{x_i \in A} \sum_{x_j \in B} d(x_i, x_j)$$

$$x_i \in A, i = 1, \dots, n_A \text{ ja}$$
$$x_j \in B, j = 1, \dots, n_B$$

Klusterissa A on n_A alkiota ja klusterissa B n_B alkiota.



- Stokastisessa klusteroinnissa yhdistäminen tapahtuu esimerkiksi etäisyysfunktion käännteislukuun pohjautuvan todennäköisyyden perusteella.

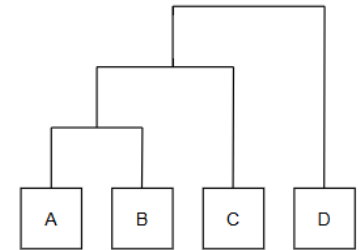
Stokastinen hierarkkinen klusterointi

Deterministinen ja stokastinen klusterointi:

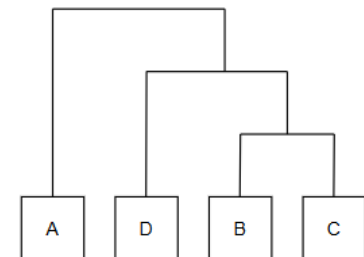
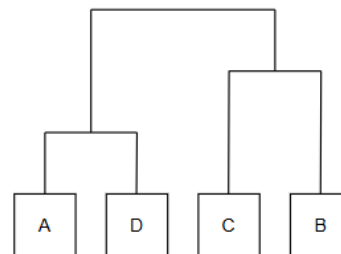
Etäisyysmatriisi alussa:

	Alue A	Alue B	Alue C	Alue D
Alue A	0	2	4	10
Alue B	2	0	3	9
Alue C	4	3	0	6
Alue D	10	9	6	0

Deterministinen klusterointi päättyy aina samaan lopputulokseen, joka riippuu ainoastaan linkkifunktion valinnasta.



Kaksi esimerkkiä **stokastisen klusteroinnin** lopputuloksista:



Aikataulu

- Aihe-esittely 11/2024
- Lähdeaineistoon perehtyminen sekä klusteroinnin tekninen toteutus 11-12/2024
- Työn kirjoittaminen 01-02/2025
- Valmis työ 02/2025

Viitteet

- Helsingin Sanomat. (2024). Yhä useampi nuori aikuinen jää työkyvyttömyyseläkkeelle. <https://www.hs.fi/politiikka/art-2000010827146.html>
- Kela. (2024). Mielen terveysongelmat veivät jo yli 100 000 suomalaista pitkälle sairauspoissaololle vuonna 2023. <https://www.kela.fi/ajankohtaista/mielen-terveysongelmat-veivat-jo-yli-100-000-suomalaista-pitkalle-sairauspoissaololle-vuonna-2023>
- Sosiaali- ja terveysministeriö. (2024). Hyvinvointialueet kartalla. <https://stm.fi/hyvinvointialueet-kartalla>
- Terveyden ja hyvinvoinnin laitos. (2024). Kansallinen terveysindeksi – THL. <https://thl.fi/tutkimus-ja-kehittaminen/tutkimukset-ja-hankkeet/kansallinen-terveysindeksi>
- Työterveyslaitos TTL. (2023). Työkyky | Työterveyslaitos. <https://www.ttl.fi/teemat/tyohyvinvointi-ja-tyokyky/tyokyky>
- Virtanen, M., Heikkilä, K., Vahtera, J., Kivimäki, M., Halonen, J. I., Alexanderson, K., Rautiainen, S., Lallukka, T., & Mittendorfer-Rutz, E. (2022). Clustering of disability pension and socioeconomic disadvantage in Sweden: A geospatial analysis. *European Journal of Public Health*, 32(5), 703–708. <https://doi.org/10.1093/eurpub/ckac096>
- Yle. (2023). Työkyvyttömyys kohtaa useammin pohjois- ja itäsuomalaisia kuin etelässä ja lännessä asuvia suomalaisia. <https://yle.fi/a/74-20034330>
- Zeshui, X. (2009). Intuitionistic fuzzy hierarchical clustering algorithms. *Journal of Systems Engineering and Electronics*, vol. 20, no. 1, pp. 90-97.