

Hyvinvointialueiden stokastinen hierarkkinen klusterointi Kansallisen terveysindeksin pohjalta (valmiin työn esittely)

Iina Manninen

06.03.2026

Ohjaaja: *Leevi Olander*

Valvoja: *Ahti Salo*

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

Työn taustaa

- Sairastavuus ja työkyvyttömyys aiheuttavat inhimillisiä ja taloudellisia menetyksiä yksilöille, yhteisöille ja yhteiskunnalle.
- Väestörakenteen muutokseen varautumiseksi sosiaali- ja terveys- ja pelastuspalveluiden järjestämisvastuu siirrettiin hyvinvointialueille vuoden 2023 alussa.
- Kansallinen terveysindeksi kuvaa väestön terveyden kehitystä ja sosiaali- ja terveydenhuollon uudistuksen tavoitteiden toteutumista.

Sairastavuus ja työkyky

- Sairastavuus kuvaa sairauksien yleisyyttä väestössä.
- Mikäli diagnosoitu sairaus estää työnteon, on mahdollista hakea
 - sairauspäivärahaa,
 - työkyvyttömyyseläkettä tai
 - ammatillista kuntoutusta.
- Työkyky rakentuu ihmisen voimavarojen ja kuormitustekijöiden tasapainosta.
- Työterveyshuollon tehtävä on työkyvyn tukeminen.

Sote-uudistus ja uudet hyvinvointialueet

1 Ahvenanmaa	7 Kainuu	13 Lappi	19 Pohjois-Savo
2 Etelä-Karjala	8 Kanta-Häme	14 Länsi-Uusimaa	20 Päijät-Häme
3 Etelä-Pohjanmaa	9 Keski-Pohjanmaa	15 Pirkanmaa	21 Satakunta
4 Etelä-Savo	10 Keski-Suomi	16 Pohjanmaa	22 Varsinais-Suomi
5 Helsinki	11 Keski-Uusimaa	17 Pohjois-Karjala	23 Vantaa ja Kerava
6 Itä-Uusimaa	12 Kymenlaakso	18 Pohjois-Pohjanmaa	

- Uudistuksen tavoitteena on kustannusten kasvun hidastaminen sekä alueellisten terveyserojen kaventaminen.
- Jatkokehitysehdotuksina on esitetty muun muassa alueiden määrän arviointia sekä hyvinvointialueiden verotusoikeutta.

Kansallinen terveysindeksi

- | | |
|-------------------------------|--|
| 1 Aivoverisuonitauti-indeksi | 6 Muistisairausindeksi |
| 2 Alkoholisairastavuusindeksi | 7 Sepelvaltimotauti-indeksi |
| 3 Diabetesindeksi | 8 Syöpäindeksi |
| 4 Keuhkosairausindeksi | 9 Tapaturmaindeksi |
| 5 Mielenterveysindeksi | 10 Tuki- ja liikuntaelinsairausindeksi |

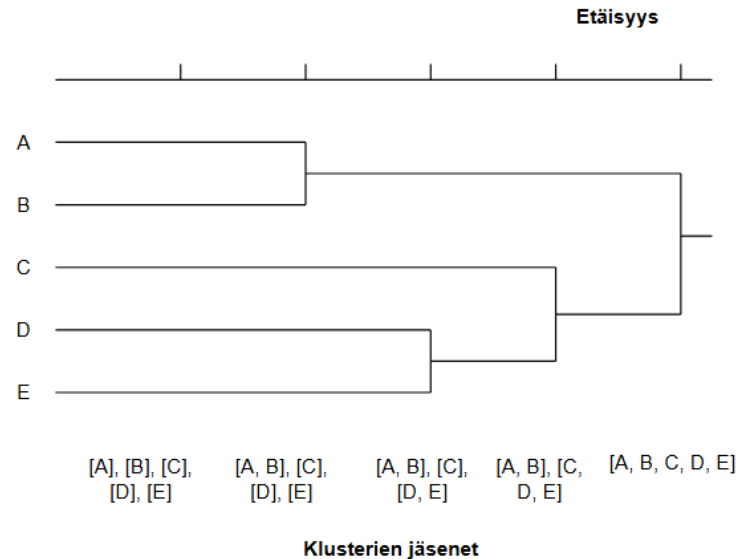
- | | |
|------------------------------|-------------------------------------|
| 1 Työkyvyttömyyseläkeindeksi | 3 Ammatillisen kuntoutuksen indeksi |
| 2 Sairauspäiväraha-indeksi | |

- Terveyden ja hyvinvoinnin laitos sekä Kansaneläkelaitos koostavat Kansalliseen terveysindeksiin aluekohtaista tietoa väestön sairastavuudesta ja työkyvystä.
- Tässä työssä hyödynnetään vuosien 2021 – 2023 indeksitietoja.

Klusterointi

- Klusterointi on numeerinen menetelmä, jonka avulla aineisto voidaan luokitella valittujen ominaisuuksien perusteella klustereihin eli luokkiin siten, että saman klusterin havainnoilla on mahdollisimman samankaltaiset piirteet, ja että eri klusterit erottuvat ominaisuuksiltaan mahdollisimman selkeästi.

Klusterointi



- Kokoavassa hierarkkisessa klusteroinnissa kaikki havainnot ovat aluksi omissa luokissaan. Klusteroinnin edetessä kaksi samankaltaisinta luokkaa yhdistetään yhdeksi luokaksi.

Monte Carlo -simulointi

- Monte Carlo –simuloinnissa generoidaan satunnaisotoksia tarkasteltavaa kohdetta mallintavista todennäköisyysjakaumista ja laskea näistä halutut tilastolliset suureet.
- Kokoava hierarkkinen klusterointi voidaan toteuttaa satunnaisuuselementin sisältävänä stokastisena klusterointina, jossa simuloidaan klusterien muodostumista hierarkian eri tasoilla.

Deterministinen klusterointi

- Klusteroinnin etäisyysmittana on euklidinen etäisyys.

$$d(x_i, x_j) = \left[\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right]^{1/2}, \quad x_{ik} \text{ ja } x_{jk} \text{ ovat muuttujan } k \text{ arvot hyvinvointialueille } i \text{ ja } j.$$

- Klusterien yhdistämisessä sovelletaan keskiarvolinkkifunktiota.

$$d(c_A, c_B) = \frac{1}{n_{c_A} \cdot n_{c_B}} \sum_{x_i \in c_A} \sum_{x_j \in c_B} d(x_i, x_j), \quad c_A \text{ ja } c_B \text{ ovat erilliset klusterit,}$$

n_{c_A} alkioden määrä klusterissa c_A ja n_{c_B} alkioden lukumäärä klusterissa c_B

Stokastinen hierarkkinen klusterointi

- Keskiarvolinkkifunktion sijaan yhdistettävät klusterit arvotaan todennäköisyysjakaumasta.

$$\mathbb{P}[\text{Yhdistetään } c_A \text{ ja } c_B \text{ tasolla } l] = p_{(c_A, c_B)}^{(l)} = \frac{1/d(c_A, c_B)}{\sum_{\{a < b | c_a \in y_l, c_b \in y_l\}} 1/d(c_a, c_b)}$$

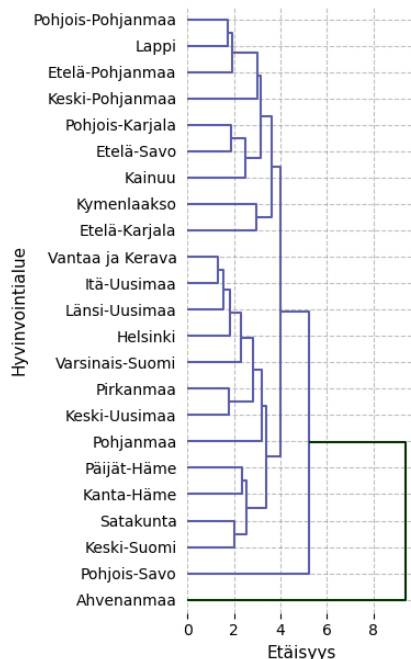
- Kun klusterit määritetään useita kertoja simuloimalla, voidaan tarkastella, kuinka usein tietyt alueet ovat samassa klusterissa.
- Kumulatiivisen klusterijäsenyysmatriisin elementtiä voidaan kuvata kaavalla

$$M_{ij}^{(l)} = \sum_{t=1}^N \mathbf{1}\{x_i \text{ ja } x_j \text{ samassa klusterissa tasolla } l \text{ toistolla } t\},$$

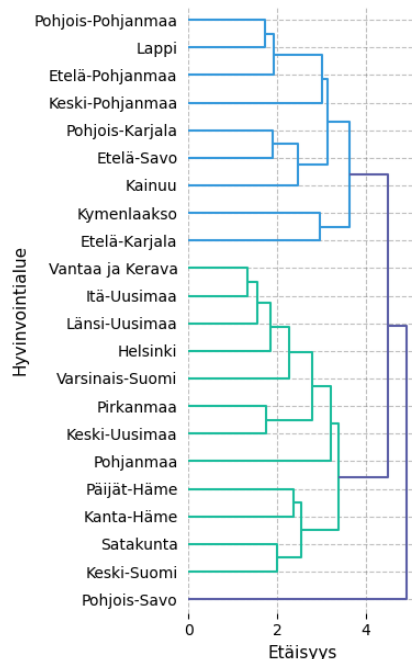
$N \in \mathbb{N}$ on simulointikertojen määrä ja $\mathbf{1}\{\cdot\}$ on indikaattorifunktio.

Tulokset – sairastavuus

Sairastavuuden osaindeksitietoihin pohjautuvan deterministisen klusteroinnin dendrogrammi, Ahvenanmaa mukana



Sairastavuuden osaindeksitietoihin pohjautuvan deterministisen klusteroinnin dendrogrammi, Ahvenanmaa ei mukana



Klusteri 1
Helsinki
Itä-Uusimaa
Kanta-Häme
Keski-Uusimaa
Keski-Suomi
Länsi-Uusimaa
Pirkanmaa
Pohjanmaa
Päijät-Häme
Satakunta
Vantaa ja Kerava
Varsinais-Suomi
Klusteri 2
Etelä-Karjala
Etelä-Pohjanmaa
Etelä-Savo
Kainuu
Keski-Pohjanmaa
Kymenlaakso
Lappi
Pohjois-Karjala
Pohjois-Pohjanmaa
Klusteri 3
Pohjois-Savo

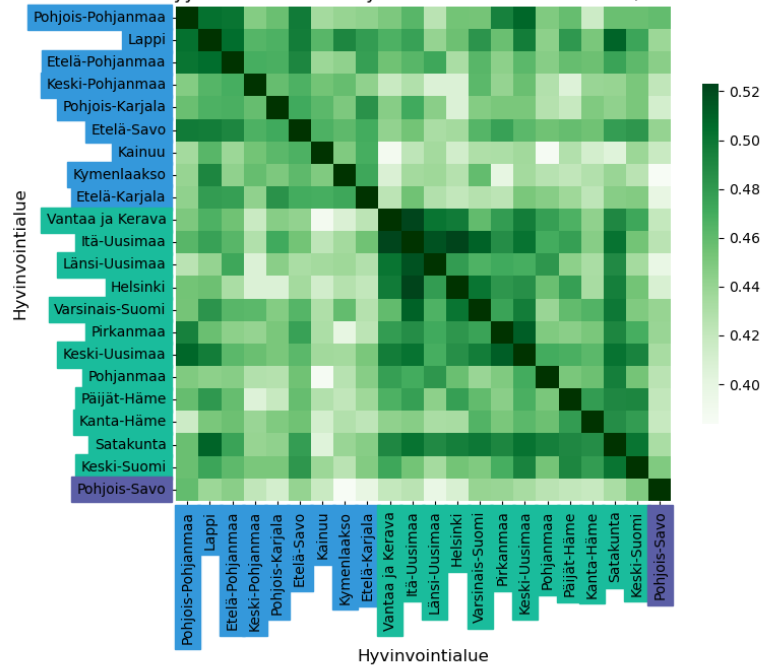
Taulukko 2: Hyvinvointialueet klusterittain sairastavuuden osaindekseihin perustuen.

Osaindeksi	Klusteri 1	Klusteri 2	Klusteri 3
<i>Keskiarvo</i>			
Aivoverisuonitauti-indeksi	101.1	104.2	95.2
Alkoholisairastavuusindeksi	94.7	117.7	114.1
Diabetesindeksi	97.3	106.7	112.1
Keuhkosairausindeksi	98.3	106.7	110.0
Mielenterveysindeksi	91.3	105.5	144.3
Muistisairausindeksi	95.0	108.8	110.2
Sepelvaltimotauti-indeksi	94.4	116.6	103.4
Syöpäindeksi	100.2	99.3	94.9
Tapaturmaindeksi	98.0	113.3	73.1
Tuki- ja liikuntaelinsairausindeksi	93.3	111.5	132.9
<i>Mediaani</i>			
Aivoverisuonitauti-indeksi	93.4	103.1	95.2
Alkoholisairastavuusindeksi	95.2	118.2	114.1
Diabetesindeksi	98.4	105.1	112.1
Keuhkosairausindeksi	96.6	106.9	110.0
Mielenterveysindeksi	91.2	104.2	144.3
Muistisairausindeksi	96.9	110.3	110.2
Sepelvaltimotauti-indeksi	93.1	116.3	103.4
Syöpäindeksi	100.7	101.4	94.9
Tapaturmaindeksi	98.2	111.3	73.1
Tuki- ja liikuntaelinsairausindeksi	94.9	113.4	132.9
<i>Minimi</i>			
Aivoverisuonitauti-indeksi	89.6	93.8	95.2
Alkoholisairastavuusindeksi	60.1	93.8	114.1
Diabetesindeksi	83.6	98.7	112.1
Keuhkosairausindeksi	84.6	90.6	110.0
Mielenterveysindeksi	74.7	86.3	144.3
Muistisairausindeksi	80.5	89.3	110.2
Sepelvaltimotauti-indeksi	80.9	105.1	103.4
Syöpäindeksi	93.3	88.2	94.9
Tapaturmaindeksi	86.6	93.5	73.1
Tuki- ja liikuntaelinsairausindeksi	72.5	94.5	132.9
<i>Maksimi</i>			
Aivoverisuonitauti-indeksi	123.6	114.4	95.2
Alkoholisairastavuusindeksi	119.1	142.3	114.1
Diabetesindeksi	104.6	116.1	112.1
Keuhkosairausindeksi	111.9	119.6	110.0
Mielenterveysindeksi	106.0	122.8	144.3
Muistisairausindeksi	101.6	119.7	110.2
Sepelvaltimotauti-indeksi	109.4	126.0	103.4
Syöpäindeksi	105.3	104.7	94.9
Tapaturmaindeksi	109.4	136.3	73.1
Tuki- ja liikuntaelinsairausindeksi	110.4	124.1	132.9

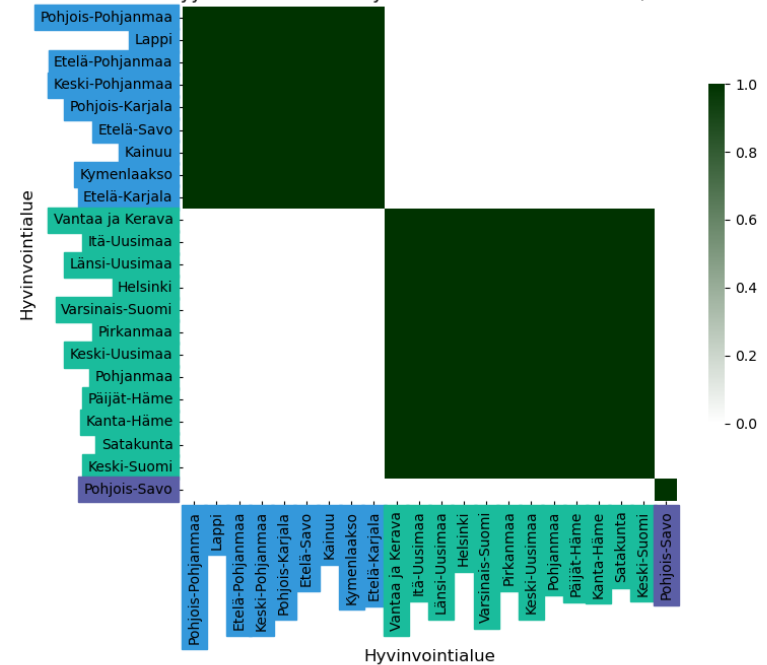
Taulukko 3: Sairastavuuden osaindeksien tilastollisia tunnuslukuja klusterittain.

Tulokset – sairastavuus

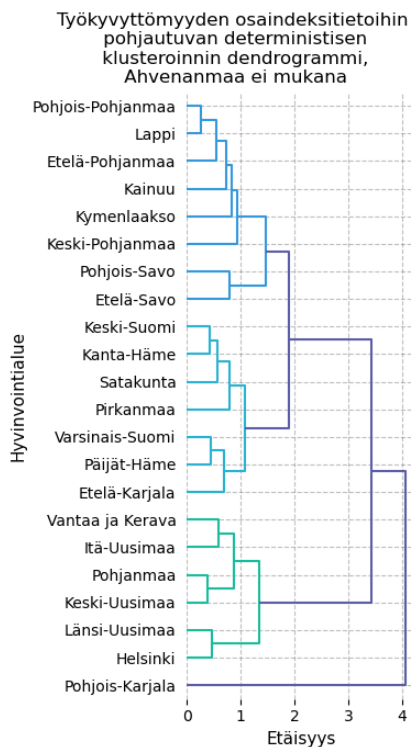
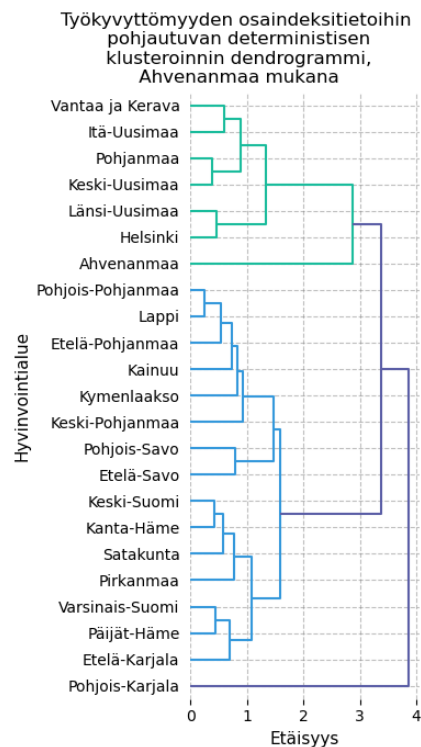
Sairastavuuden osaindeksitietoihin pohjautuvan stokastisen klusteroinnin todennäköisyydet alueiden esiintymiseen samassa klusterissa, 3 klusteria



Sairastavuuden osaindeksitietoihin pohjautuvan deterministisen klusteroinnin todennäköisyydet alueiden esiintymiseen samassa klusterissa, 3 klusteria



Tulokset – työkyvyttömyys



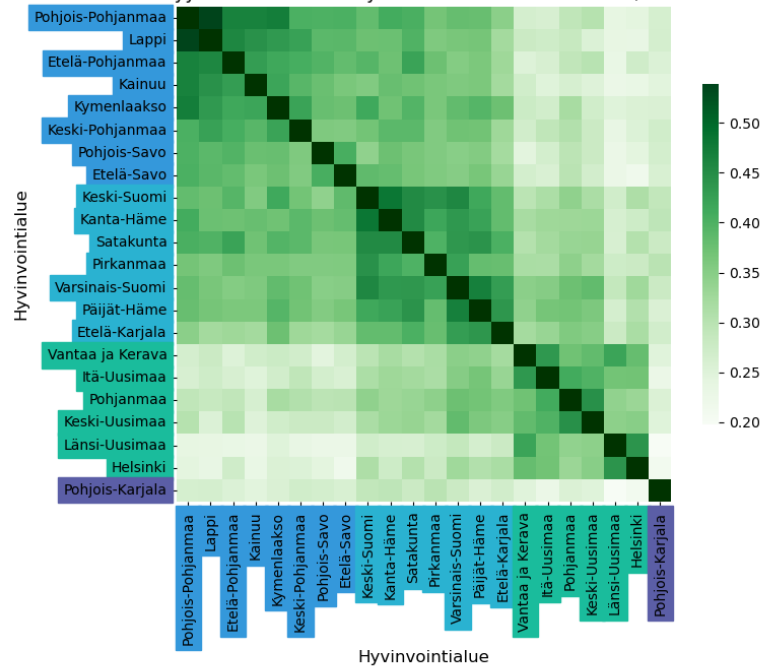
Klusteri 1	Osaaindeksi	Klusteri 1	Klusteri 2	Klusteri 3	Klusteri 4
Helsinki	<i>Keskiarvo</i>				
Itä-Uusimaa	Työkyvyttömyyseläkkeet	75.3	107.0	128.5	120.0
Keski-Uusimaa	Sairauspäivärahat	81.5	108.4	125.0	104.0
Länsi-Uusimaa	Ammatillinen kuntoutus	63.8	109.4	121.4	222.0
Pohjanmaa	<i>Mediaani</i>				
Vantaa ja Kerava	Työkyvyttömyyseläkkeet	74.00	107.00	129.0	120.0
Klusteri 2	Sairauspäivärahat	83.00	108.0	126.00	104.0
Etelä-Karjala	Ammatillinen kuntoutus	62.50	110.00	114.00	222.0
Kanta-Häme	<i>Minimi</i>				
Keski-Suomi	Työkyvyttömyyseläkkeet	67.00	100.00	113.00	120.0
Pirkanmaa	Sairauspäivärahat	63.00	97.00	112.0	104.0
Päijät-Häme	Ammatillinen kuntoutus	52.00	82.00	96.00	222.0
Satakunta	<i>Maksimi</i>				
Varsinais-Suomi	Työkyvyttömyyseläkkeet	84.00	116.00	139.0	120.0
Klusteri 3	Sairauspäivärahat	97.00	115.00	132.0	104.0
Etelä-Pohjanmaa	Ammatillinen kuntoutus	82.00	143.00	166.00	222.0
Etelä-Savo					
Kainuu					
Keski-Pohjanmaa					
Kymenlaakso					
Lappi					
Pohjois-Pohjanmaa					
Pohjois-Savo					
Klusteri 4					
Pohjois-Karjala					

Taulukko 5: Työkyvyttömyyden osaindeksien tilastollisia tunnuslukuja klustereittain.

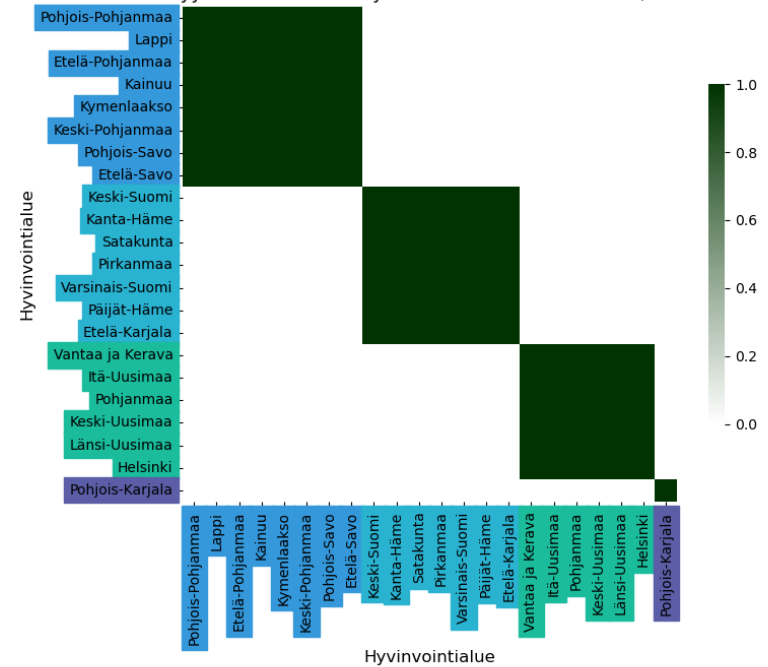
Taulukko 4: Hyvinvointialueet klustereittain työkyvyttömyyden osaindekseihin perustuen.

Tulokset – työkyvyttömyys

Työkyvyttömyyden osaindeksitietoihin pohjautuvan stokastisen klusteroinnin todennäköisyydet alueiden esiintymiseen samassa klusterissa, 4 klusteria



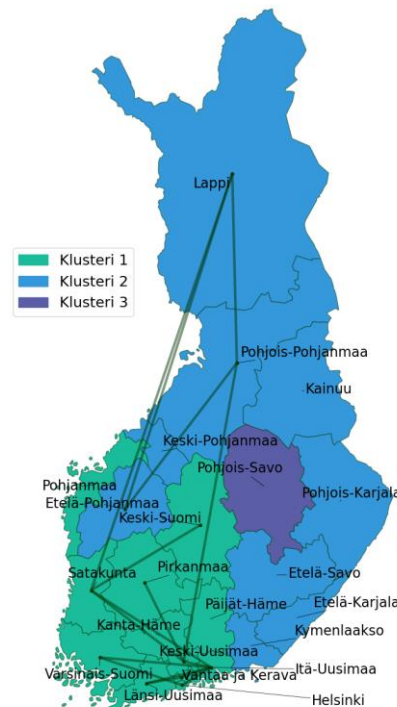
Työkyvyttömyyden osaindeksitietoihin pohjautuvan deterministisen klusteroinnin todennäköisyydet alueiden esiintymiseen samassa klusterissa, 4 klusteria



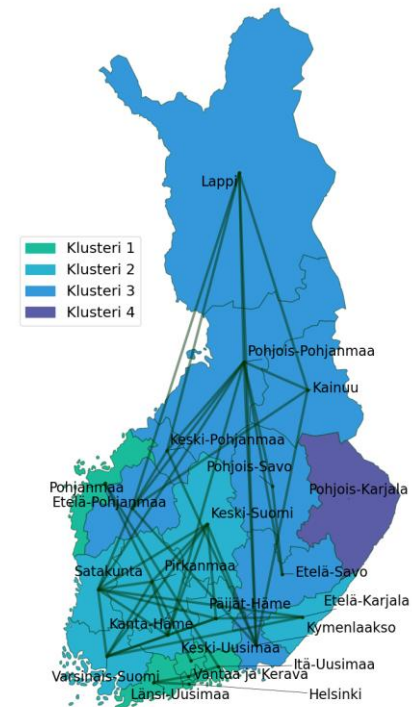
Sairastavuus ja työkyvyttömyys hyvinvointialueilla

- Klusteroinnin tuloksena nähdään jako vähäisemmän sairastavuuden ja työkyvyttömyyden Länsi- ja Etelä-Suomeen sekä suuremman sairastavuuden ja työkyvyttömyyden Itä- ja Pohjois-Suomeen.

Sairastavuuden osaindeksitietoihin pohjautuvan klusteroinnin tulokset kartalla, 3 klusteria, todennäköisyyskynnys 0.5



Työkyvyttömyyden osaindeksitietoihin pohjautuvan klusteroinnin tulokset kartalla, 4 klusteria, todennäköisyyskynnys 0.4



Jatkotutkimuksen näkökulmia

- Stokastisen klusteroinnin tuloksena saadaan kahden alueen todennäköisyys samassa klusterista esiintymisestä.
- Hyvinvointialueiden ohella alueellista vertailua voidaan tehdä myös kuntatasolla.
- Alueiden sairastavuuden kirjauskäytännöt voivat vaikuttaa indeksitietoihin.