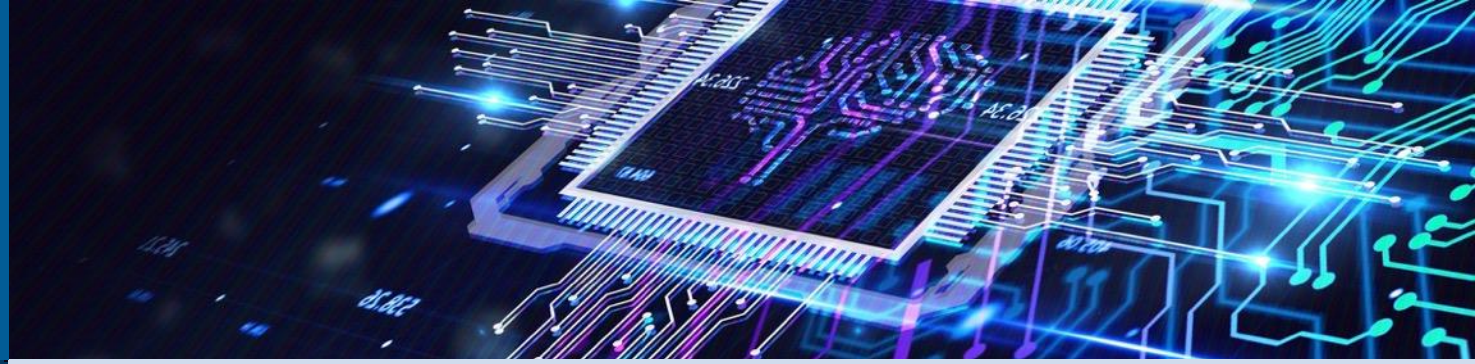
CSC

ICT Solutions for Brilliant Minds



Haku- ja valintatietojen klusterianalyysi

Markus Koskela ja Aino Ropponen
CSC – Tieteen tietotekniikan keskus

vipunen.fi/fi-fi/kkyhteiset/Sivut/Haku-ja-valinta.aspx

vipunen OPETUSHALLINNON TILASTOPALVELU

suomeksi | på svenska | in English

Tietoja palvelusta | Koulutuksen yhteiset | Varhaiskasvatus | Esi- ja perusopetus | Lukiokoulutus | Ammatillinen koulutus | Ammattikorkeakoulutus | Yliopistokoulutus | Korkeakoulutus ja t&k-toiminta | Vapaa sivistystyö | Väestö-, koulutus- ja ammattirakenne

- INDIKAATTORIT
- HAKU JA VALINTA
- OPISKELIJAT JA TUTKINNOT
- TUTKINNON SUORITTANEIDEN SIJOITTUMINEN
- OPETTAJAKELPOISUUDET
- HENKILÖSTÖ
- TIETEEN JA TEKNOLOGIAN HENKILÖVOIMAVARAT
- TUTKIMUS- JA KEHITTÄMISTOIMINTA
- BIBLIOMETRIIKKA (WEB OF SCIENCE)
- BIBLIOMETRIIKKA (SCOPUS)
- KANSAINVÄLISYYS
- JULKAISUT
- KORKEAKOULUTUS JA TUTKIMUS KANSAINVÄLISESSÄ VERTAILUSSA

Korkeakoulutuksen yhteiset ja t&k-toiminta > Haku ja valinta

Haku ja valinta

Raporteilla voi tarkastella korkeakoulutuksen haku- ja valintatietoja vuodesta 2015 lähtien.

Huom! Raporteille on tehty tietosuojaus, jos sivulla ei ole toisin mainittu. Henkilöitä koskevat lukumäärät 1–4 näytetään arvona "1-4", eikä vastaavia laskennallisia tunnuslukuja (esim. keskiarvo) ilmoiteta. Neljää suuremmat lukumäärät pyöristetään lähimpään kolmella jaolliseen lukuun. Prosenttiosuudet lasketaan suojatuilla lukumäärillä.

Haun ja valinnan koontiraportit

Koontiraporteihin on koottu haun ja valinnan keskeisiä tietoja helposti luettavaan muotoon. Hakukohteen profiili – raporttiin on koottu hakukohdetason keskeisiä tietoja yhteen näkymään. Raportti on suunnattu esimerkiksi hakukohteiden välistä vertailua tekeville käyttäjille. Yhteenvetoraportille on koottu haun ja valinnan keskeisiä koontitietoja esimerkiksi hakijoiden ja paikan vastaanottaneiden volyymien kehittymisestä. Mukana on myös koontinäkymä hakukohdetason tiedoista. Raporteissa on oletuksena suodatettu esiin I aloitussyklin hakujen tiedot (tarkoittaa koulutusta johon haetaan toisen asteen tutkinnon jälkeen). Tarvittaessa suodattimista saa valittua myös II aloitussyklin hakujen tiedot (esim. yliopistojen maisterivaihe).

Hakukohteen profiili
Haun ja valinnan yhteenveto

Korkeakoulujen hakeneet ja paikan vastaanottaneet

Raporteilla on tietoja korkeakoulujen yhteishaun ja erillishaun hakijatiedoista. Kaikki hakijat -mittarin tiedot on nettoutettu siten että yhteensä-rivillä hakija voi esiintyä vain yhden kerran, muilla riveillä useaan kertaan.

Vuoden 2022 koulutusten koulutusalaan ja OKM:n ohjauksen alaan on tehty päivitys 5.7.2022, jonka seurauksena ala on saattanut muuttua aikaisemmasta. Päivitys ei koske kaikkia koulutuksia. Lue lisää Tilastoneuvos-blogista.

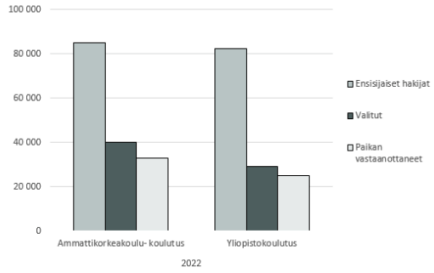
Hakukohde
Hakukohteen maakunta
Ikäryhmä
Korkeakoulu
Koulutuksen alkamiskausi
Sektori

Raportti ilman henkilömuuttujia (ei tietosuojausta)

Korkeakoulujen rekrytointialueet

Hakijat hakijoiden kotimaakunnan ja hakukohteen maakunnan mukaan. Raportin avausnäkymässä koulutuksen maakunta on riveillä ja hakijan asuinmaakunta sarakkeilla.

Hakijan kotimaakunta (%)



Ensisijaiset hakijat ja paikan vastaanottaneet kevään yhteishaussa

Visualisoinnit

Hakutoiveet ja ylioppilastutkinnon arvosanat
Korkeakoulupaikan vastaanottaneiden ylioppilaiden arvosanat
Ammatillisen tutkinnon suorittaneiden korkeakoulutukseen sijoittuminen (0,5-1 vuotta)
Ammatillisen tutkinnon suorittaneiden korkeakoulutukseen sijoittuminen (0,5-5 vuotta)
Ylioppilastutkinnon suorittaneiden koulutukseen sijoittuminen (0-1 vuotta)
Ylioppilastutkinnon suorittaneiden koulutukseen sijoittuminen (0-5 vuotta)

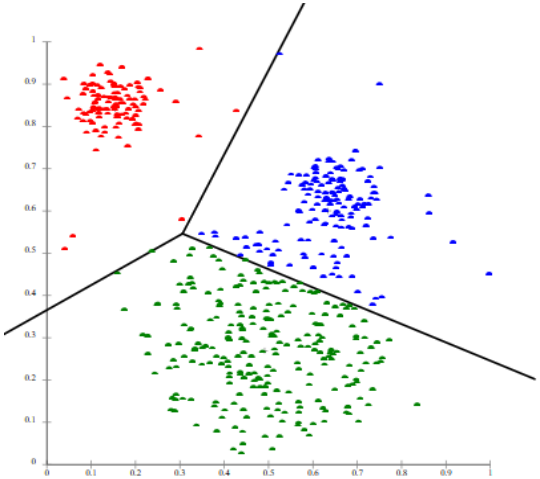
Analyysiraportit

Laajennettu tietosisältö analyysikäyttöön.

https://vipunen.fi/fi-fi/kkyhteiset/Sivut/Haku-ja-valinta.aspx

Klusterointi

- Perusperiaate: ryhmitellään aineistoa tietämättä luokkia
- Paljon eri menetelmiä, joiden tuottamat klusteroinnit erilaisia samalle aineistolle
- Tarkoituksena löytää datasta ryhmiä eli klustereita
  - Klusterin sisällä näytteet jollain tapaa samanlaisia
  - Klustereiden välillä mahdollisimman erilaisia
- Ihminen voi yrittää löytää selityksiä klustereille



https://upload.wikimedia.org/wikipedia/commons/e/e5/KMeans-Gaussian-data.svg

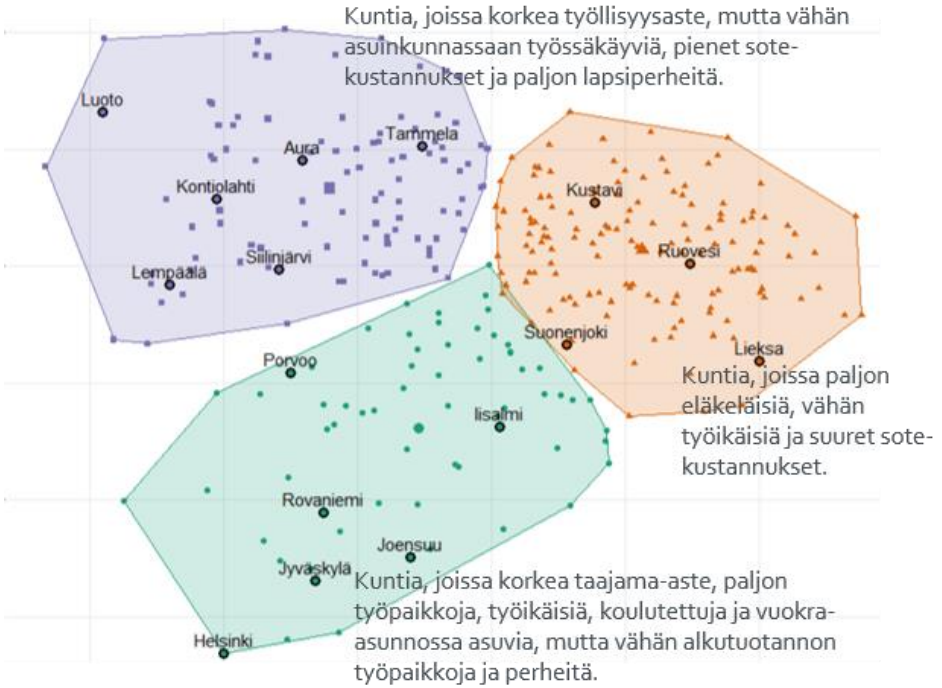
Esimerkkejä klusteroinnista

Samankaltaisten kuntien tunnistaminen

Nuorten hyvinvointierojen tunnistaminen Kouluterveyskyselyn perusteella

Samankaltaisten käyttäjien/asiakkaiden ryhmittely ja ominaisuuksien tunnistaminen

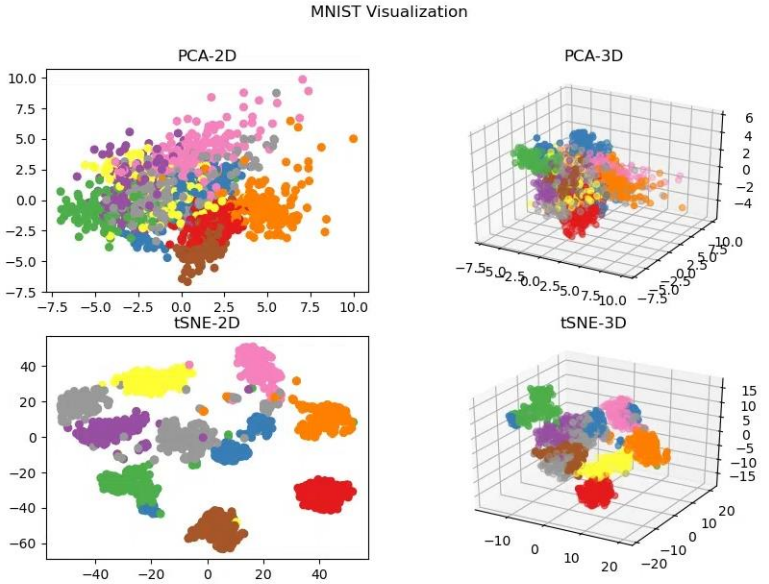
Eduskuntavaaliehdokkaiden ryhmittely vaalikonevastausten perusteella



Datalähde: Tilastokeskus

Visualisointi

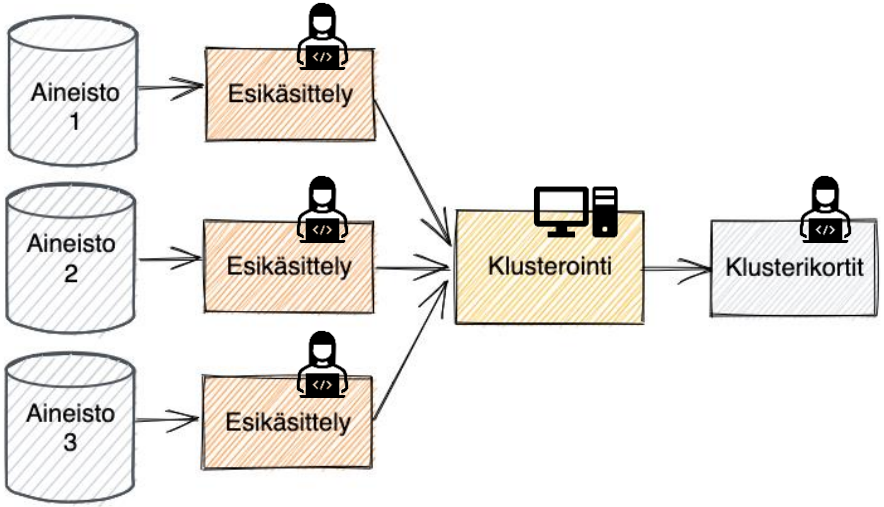
- Klusteroinnin tulosta eli saatua aineiston ryhmittelyä voidaan hyödyntää monin tavoin
- Usein saadut klusterit halutaan visualisoida, johon voidaan käyttää erilaisia ulotteisuuden pienentämismenetelmiä
  - esimerkiksi pääkomponenttianalyysi (PCA)



https://in2techs.com/mnist-visualization-using-pca-and-tsne-in-python/

Klusterikortit

- Klusterikortit on visuaalinen tapa esittää kunkin klusterin tilannekuva
- Menetelmää voidaan soveltaa monenlaisille aineistoille, sillä syötedata voidaan muokata esikäsittelyvaiheessa sopivaan muotoon
- Voidaan toteuttaa esimerkiksi interaktiivisena tietotyöpöytänä tai PowerPoint-tiedostona



Haku- ja valintatietojen klusterikortit

- Aineistona korkeakoulujen haku- ja valintatiedot 2019-2021
  - Koko aineisto tai vuosikohtaiset tiedot
- 25 muuttujaa (+ 17 taustamuuttujaa)
- Interaktiivinen tietotyöpöytä (dashboard)
- K-means ja pääkomponenttianalyysi
- 8 klusteria:
  1. “Aidosti ensikertalaiset”
  2. “Pitkä matematiikka”
  3. “Lyhyt matematiikka”
  4. “Ulkomaan kansalaiset”
  5. “Usealle alalle hakeneet”
  6. “Yhdelle alalle hakeneet”
  7. “Aiemmin valitut”
  8. “Vain yhteen kohteeseen hakeneet”

Klusterikortit - OKM

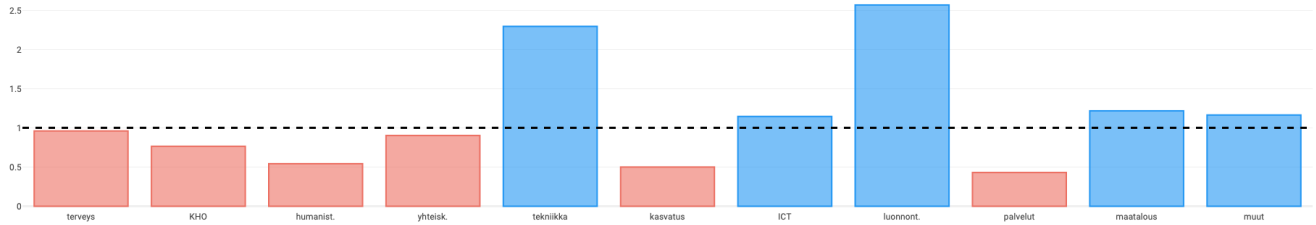
Koko aineisto | Klusterit

- 1: "Aidosti ensikertalaiset"
- 2: "Pitkä matematiikka"
- 3: "Lyhyt matematiikka"
- 4: "Ulkomaan kansalaiset"
- 5: "Usealle alalle hakeneet"
- 6: "Yhdelle alalle hakeneet"
- 7: "Aiemmin valitut"
- 8: "Vain yhteen kohteeseen hakeneet"

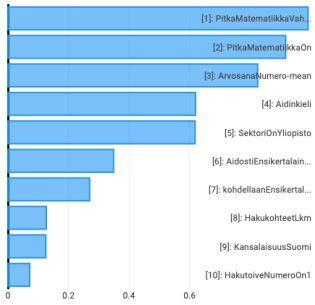
Kirjoittaneet pitkän matematiikan, usein hyvillä arvosanoilla, ja muutenkin hyvät arvosanat. Hakevat yliopistoon, usein tekniikan tai luonnontieteelliselle alalle. Nuoria ja vain vähän aikaa toisen asteen koulutuksesta.

Osuus:	Keskim. ikä:	Ensikertalaisia:	YO/AMK:	Valittu:	Paikan vastaanotto:
9 %	21 v	91 %	89 / 11	28 %	89 %

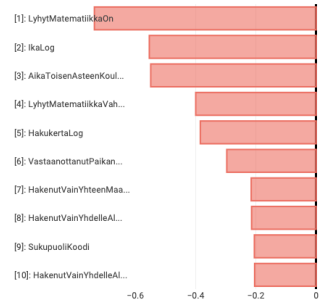
Koulutusalojen suhteellinen yleisyys tässä klusterissa



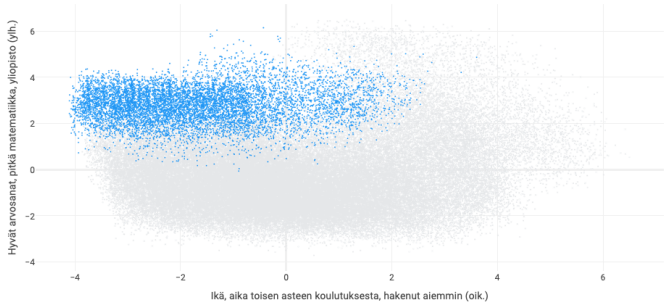
Yleistä tässä klusterissa



Harvinaista tässä klusterissa



Pääkomponenttivisualisointi



Klusterikortit - OKM

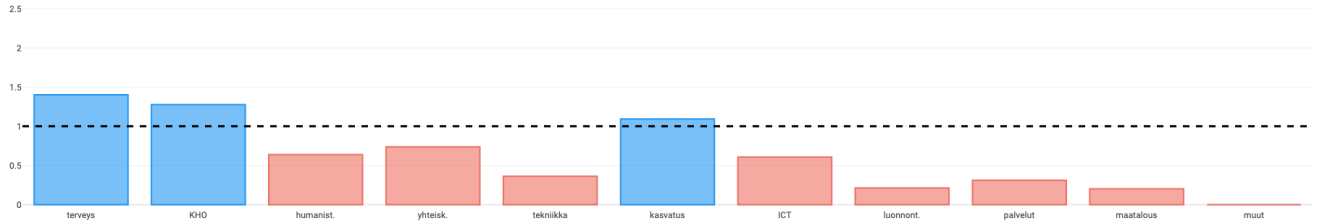
Koko aineisto | Klusterit

- 1: "Aidosti ensikertalaiset"
- 2: "Pitkä matematiikka"
- 3: "Lyhyt matematiikka"
- 4: "Ulkomaan kansalaiset"
- 5: "Usealle alalle hakeneet"
- 6: "Yhdelle alalle hakeneet"
- 7: "Aiemmin valitut"
- 8: "Vain yhteen kohteeseen hakeneet"

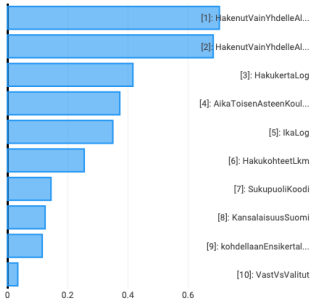
Hakeneet vain yhdelle alalle tai yhteen kohteeseen. Useita aiempia hakukertoja. Keskimääräistä vanhempia. Arvosanat keskimääräistä huonommat. Ottavat paikan vastaan keskimääräistä useammin.

Osuus:	Keskim. ikä:	Ensikertalaisia:	YO/AMK:	Valittu:	Paikan vastaanotto:
17 %	25 v	84 %	39 / 61	10 %	94 %

Koulutusalojen suhteellinen yleisyys tässä klusterissa

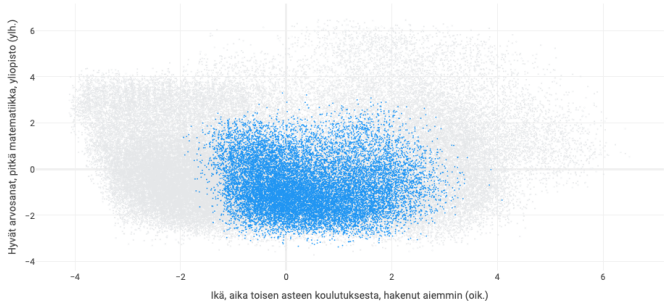


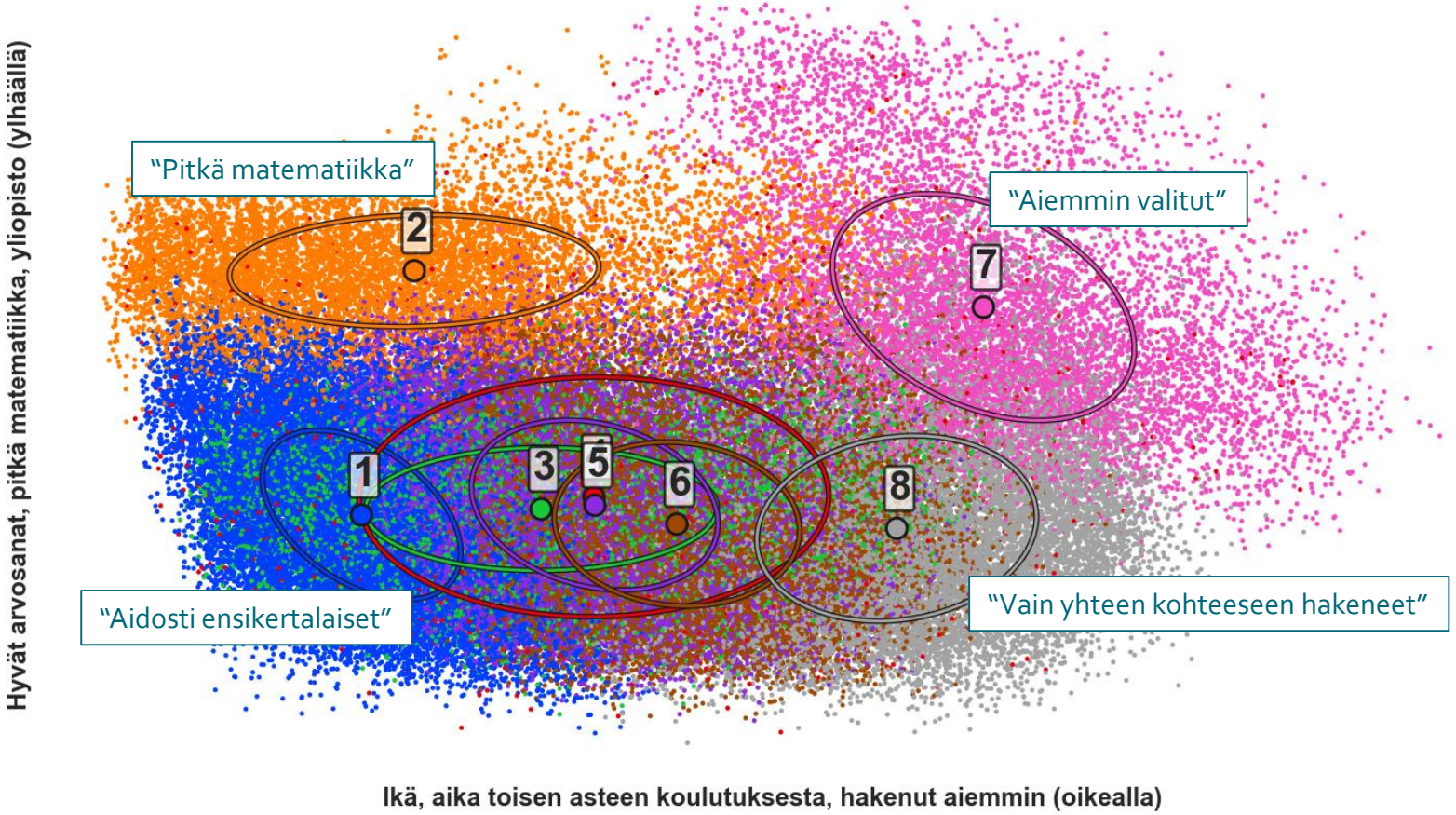
Yleistä tässä klusterissa



Harvinaista tässä klusterissa

Pääkomponenttivisualisointi




Hyvät arvosanat, pitkä matematiikka, yliopisto (ylhäällä)
“Pitkä matematiikka”
“Aiemmin valitut”
“Aidosti ensikertalaiset”
“Vain yhteen kohteeseen hakeneet”
1
2
3
5
6
7
8
Ikä, aika toisen asteen koulutuksesta, hakenut aiemmin (oikealla)

Lopuksi

- Klusterointi on kuvaileva tilastollinen menetelmä eikä se tuota tilastollisia tunnuslukuja
- Klusterointimenetelmät tuottavat aina jonkinlaisen ryhmittelyn, vaikka aineistossa ei mitään ryhmiä olisikaan
- Klusterikorttien muodostaminen on aina subjektiivista
  - substanssiosaamista tarvitaan mm. muuttujien valinnassa, esikäsittelyssä ja tulosten tulkitsemisessa
- Klusterikortit tiivistävät moniulotteisen aineiston ja saattavat auttaa esimerkiksi aineiston hahmottamisessa, hypoteesien muodostamisessa tai pienten osajoukkojen identifioinnissa

Markus Koskela

TkT, koneoppimisasiantuntija
CSC – Tieteen tietotekniikan keskus

markus.koskela@csc.fi
analytics@csc.fi

facebook.com/CSCfi

twitter.com/CSCfi

linkedin.com/company/csc---it-center-for-science

github.com/CSCfi