

# Tietoisku OKM:lle toteuttavasta analytiikasta

Heini Manninen & Hanna-Mari Puuska, CSC

IR-tekijäverkoston tilaisuus 11.11.2022



# Mistä kerromme tänään

- Lyhyesti CSC:n analytiikkatekemisestä
- Analytiikan askeleet -työkalusta
- OKM:n Korkeakoulu- ja tiedepolitiikan osastolle tehtävästä analytiikkaprojektista
- OKM:lle toteutettavasta analytiikkatyöstä  
Jatkuvan oppimisen digitalisaatio -hankkeessa
- Tiedejatutkimus.fi –palvelun tekoälyavusteisesta aihemallinnuksesta



# Analytiikkaa tehdään monipuolisesti CSC:lla

## Muutamia esimerkkejä

### Tutkimuksen ja koulutuksen tiedontuotanto ja analytiikka –yksikkö (Services for Research and Education Information Management and Data Analytics - READY)

Yksikössä on käynnissä tällä hetkellä useita **analytiikkaan ja tiedolla johtamiseen** liittyviä selvityksiä ja projekteja, mm.

- Analytiikka-projekti OKM:n Korkeakoulu- ja tiedepolitiikan osastolle
- Jatkuvan oppimisen digitalisaatio –hanke (JOD) ja analytiikan mahdollisuudet
- Bibliometriikka: lippulaiva-analyysi
- Tiedejatutkimus.fi ja tekoälyavusteinen aihemallinnus

### AI and Data Analytics (AIDA) - yksikkö

- Luonnollisen kielen käsittelyn työkalut (GPT, BERT, aihemallinnus)
- Suomen kielen syvät neuroverkkomallit
- Vaativat koneoppimismallit suurteholaskentaympäristössä
- Videokuvan käsittely, erityisesti ihmiskohteiden asentomallinnus
- Koneoppimisen ympäristöt ja hallintatyökalut (MLops)
- Tilannekuvien luominen ja tietojohdamisen tukeminen (AuroraAI)
- Dashboardit analytiikan käyttöliittymänä

# Analytiikan askeleet

---

↔ VAIKUTA

Vuorovaikuta ja luo uusia toimintatapoja.

📈 ENNUSTA

Ennusta tulevaa ja luokittele dataa.

🔍 YMMÄRRÄ

Ymmärrä ilmiö eli tunnista korrelaatioita ja klustereita.

📊 KUVAILE

Kuvaile data eli laske tunnuslukuja ja tee visualisointeja.

📦 KERÄÄ

Kerää dataa kiinnostava ilmiöstä.

### Kuvaile data

#### Esimerkki: Vipunen.fi

Suodattimet

Yliopistosta valmistuneiden työllistyminen

10. työllisyys jätteen %

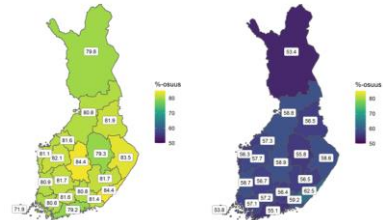
	Työllinen	Päättökoulun opiskelijat, hoidollinen	Työttö	Huut	Eläkkeellä olleet	Vuorossa
10. Kevätkoulut	87,8%	1,8%	1,8%	4,7%	3,8%	100,0%
10. Hämmentö ja koulutus	10,0%	2,3%	10,3%	3,9%	3,6%	100,0%
10. Terveystieteet	66,1%	3,0%	11,9%	11,0%	8,1%	100,0%
10. Humanistiset tieteet (paitsi kieli)	77,1%	4,0%	10,2%	5,0%	2,1%	100,0%
10. Kieli	77,4%	0,0%	0,0%	4,0%	5,0%	100,0%
10. Yhteiskunnalliset tieteet	87,4%	1,3%	4,8%	3,9%	3,4%	100,0%
10. Yhteiskuntatieteet	87,7%	1,0%	4,1%	3,0%	2,2%	100,0%
10. Vuorokausi- ja informaatiotieteet	84,1%	0,6%	2,3%	3,4%	0,0%	100,0%
10. Kauppa, hallinto ja oikeustieteet	86,2%	1,4%	2,4%	3,9%	6,1%	100,0%
10. Käsitys ja taide	92,4%	1,4%	1,1%	2,0%	7,0%	100,0%
10. Luonnontieteet	92,4%	1,3%	1,1%	4,1%	2,0%	100,0%
10. Tieteiden tutkimus (ICT)	79,0%	1,4%	2,7%	5,1%	12,7%	100,0%
10. Tieteiden tieteet	87,6%	1,0%	3,8%	2,3%	5,1%	100,0%
10. Muu ja merkittömyydet	84,0%	0,3%	1,5%	4,8%	3,8%	100,0%
10. Terveystieteiden tutkimus	86,3%	1,9%	1,3%	2,7%	8,0%	100,0%
10. Päättökoulut	82,7%	0,6%	1,0%	3,8%	2,0%	100,0%
10. Yhteensä	84,1%	1,8%	4,5%	4,8%	5,2%	100,0%

### Kuvaile data

#### Esimerkki: datan visualisointi

Tyytyväinen elämäänsä tässä hetkellä, pojat 8-9 ik.

Tyytyväinen elämäänsä tässä hetkellä, tytöt 8-9 ik.



Pojista 80,8 % on tyytyväisiä elämäänsä. Tytöistä 56,8 %.

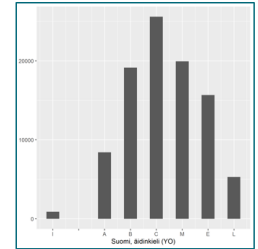
Lähde: Kouluterveyskysely 2021



Kuvaile data

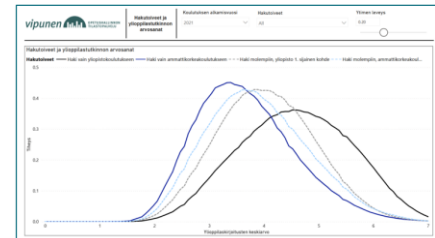
### Esimerkki: histogrammi

Ylioppilaskirjoitusten arvosanojen histogrammi kuvaa kunkin arvosanan saaneiden lukumäärää palkkikaaviona. Kuvassa vasemmalla pitkän matematiikan arvosanojen histogrammit.



Kuvaile data

### Esimerkki: ydinestimointi



Ydinestimointi (engl. kernel density estimation) on yleistyshistogrammista.

Kuvassa ylioppilaskirjoitusten keskiarvon jakauma hakutoiveittain.

Lähde: vipunen.fi

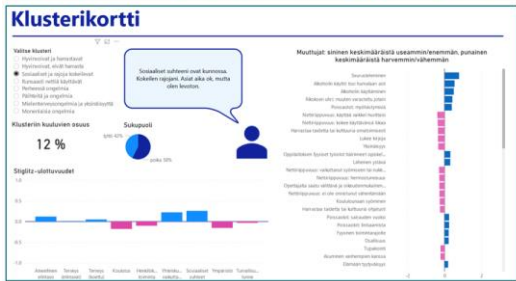
# Analytiikan askeleet

## Ymmärrä ilmiö



### Ymmärrä ilmiö

#### Esimerkki: AuroraAI:n klusterikortit



### Ymmärrä ilmiö

#### Esimerkki: tasa-arvoindikaattori



Kuvassa Biden sosioekonomisen aseman ja tulojen perusteella laskettu tasa-arvoindikaattori (SE5) uusilla opiskelijoilla.

Lähde: <https://data.biden.gov/Socioeconomic-Status>

VAIKUTA

ENNUSTA

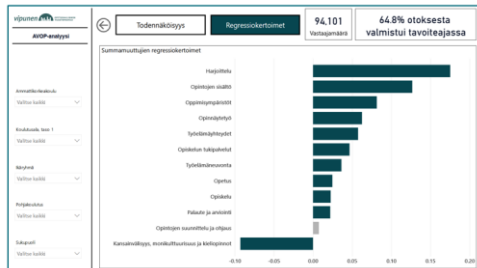
YMMÄRRÄ

KUVAILE

KERÄÄ

### Ymmärrä ilmiö

#### Esimerkki: tavoiteajassa valmistuminen



Valterri Vainio tutki vuonna 2021 diplomityössäään tavoiteajassa valmistumiseen ja opiskelijatytyvääsyteen vaikuttavia tekijöitä valmistumisvaiheen opiskelijapalauttekyseilyn perusteella.

Lähde: <https://app.pomeri.com/raportit/raportti/otintykytyvyyden-vaikuttavat-tekijot-2021>

### Ymmärrä ilmiö

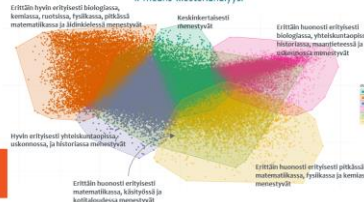
#### K-means-klusterointi

K-means on klusterointimenetelmä, joka pyrkii ryhmittelämään aineiston  $k$ :hon klusteriin.

- Jokainen havainto liittyy lähimpään klusteriin.
- Ihminen voi yrittää löytää selityksiä klusterille.

Huom. Tämä on erittäin yksinkertainen esimerkki, luotettavan analyysin tekeminen vaatii huolellisempaa dataa soveltuvuutta.

Esimerkki ylioppilaskirjoitus- ja päättötodistusarvosanojen mukaisesti tehty k-means-klusterianalyysi

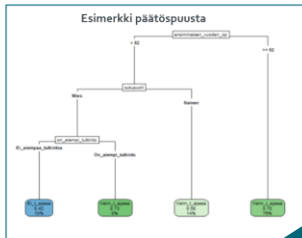


### Ennusta

#### Esimerkki: tavoiteajassa valmistuminen

CSC:n Annamari Koskiniemi ja Arttu Kottonen tekivät kesällä 2020 selvityksen tekijöistä, jotka vaikuttavat tavoiteajassa valmistumiseen. Analyysseissä käytettiin seuraavia menetelmiä:

- Korrelaatiokerroin
- Logistinen regressio
- Päätöspuu



Linkki: <https://wiki.euro.fi/pages/viewpage.action?pageId=67908025>

Millainen on tarkasteltavan nuoren syrjäytymisriski?

### Ennusta

#### Esimerkki: Annif ja Jyx

Annif on koneoppimiseen perustuva automaattisen sisällönkuvailun työkalu.

Se on käytössä muun muassa Jyväskylän yliopiston JYX-julkaisuarkistossa, jossa sitä käytetään opinnäytetöiden aiheannoittamisen apuna. Käytettävä sanasto on YSO (yleinen suomalainen ontologia).



Kenelle käsittelijälle avustushakemus pitäisi ohjata?

VAIKUTA

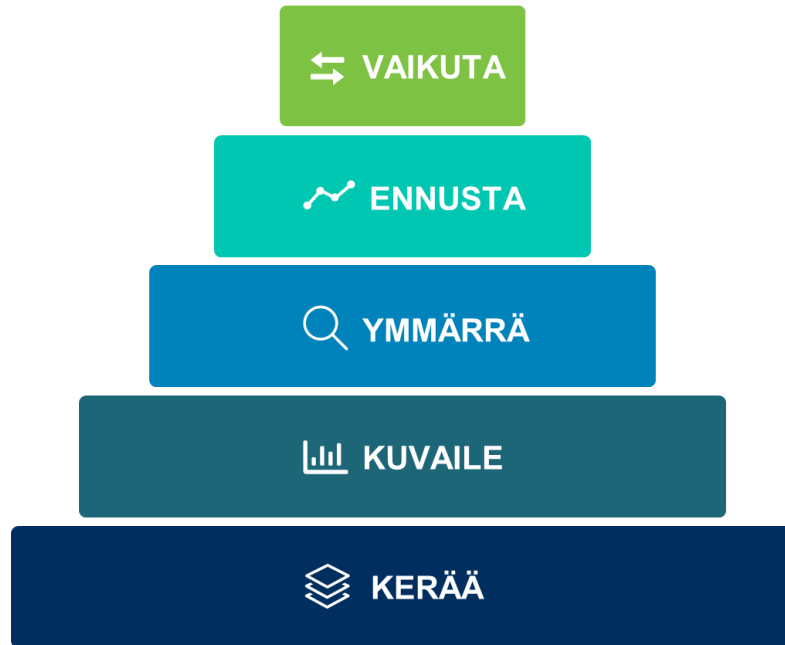
ENNUSTA

YMMÄRRÄ

KUVAILE

KERÄÄ

# Vuorovaikuta ja luo uutta



Tavoitteena **vaikuttaa ilmiöön ja luoda uusia toimintatapoja**.

Tätä kutsutaan takaisinkytkennäksi, eli toimintaa muokataan ennusteen perusteella.

Parhaimmillaan toimintaprosessia voidaan muokata joustavammaksi tai kehittää uusia tapoja toimia.

Aloitetaan tukitoimet syrjäytymisriskissä olevan nuoren tukemiseen.

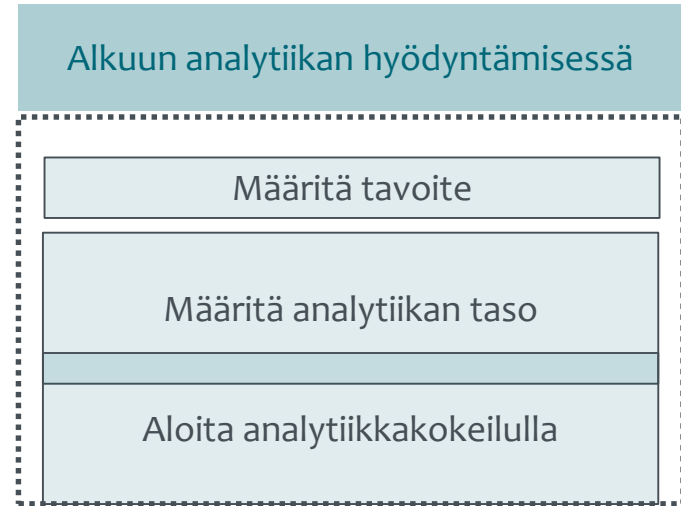
Automaattinen päätös avustushakemukseen.

Itseohjautuva auto: *auton pysäyttäminen tai esteen kiertäminen*



# Määritä tavoite

## Tunnista ilmiö eli määritä tavoite



Mitä selkeämpi tavoite, sitä todennäköisemmin data-analytiikasta on hyötyä päätöksenteossa.

# Määritä analytiikan taso

Millaista analytiikkaa ongelman ratkaisemiseksi tarvitaan?

**Kuvaile**-askeleella voidaan esim. laskea kuntakohtaisia keskiarvoja tai visualisoida viiden vuoden kehitystä

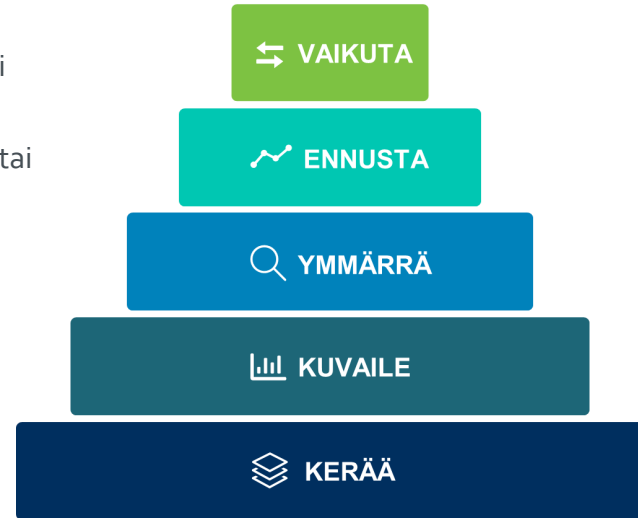
**Ymmärrä**-askeleella voidaan esim. tunnistaa datasta korrelaatioita tai klustereita.

**Ennusta**-askeleella voidaan esim. luokitella hakemuksia aihepiireihin tai ennustaa harvemmin mitattavaa muuttujaa.

**Vaikuta**-askeleella voidaan esim. automaattisesti toimia ennusteen määrittämällä tavalla.

Ylemmälle askeleelle ei voi yleensä siirtyä ennen kuin alemmat askeleet on käsitelty.

Seuraavalla askeleelle siirtyminen voi olla kallista eikä onnistuminen ole taattua.



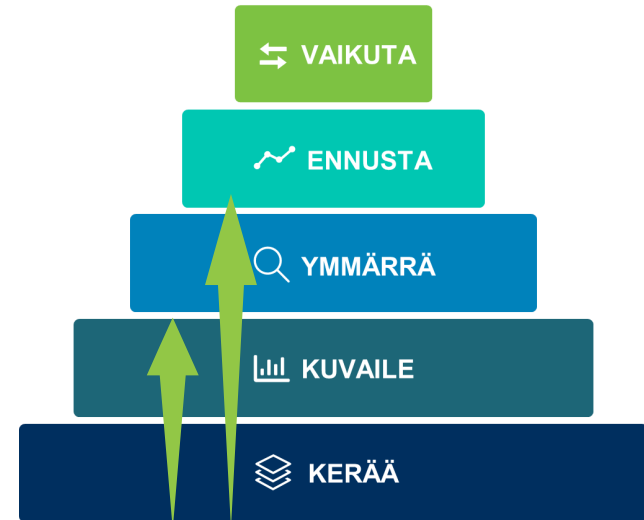
# Aloita analytiikkakokeilulla

Pienet analytiikkakokeilut ovat hyvä tapa kartoittaa eri tasoisten analyysien hyötyjä ja mahdollisuuksia

- Niiden avulla kasvatetaan osaamista analytiikan mahdollisuuksista ja riskeistä.
- Lisätään ymmärrystä omista tarpeista.

Muutaman kuukauden hankkeen aikana ehditään yleensä:

- Selvittää soveltuuko olemassa oleva data analyysitarpeisiin.
- Tehdä kuvaile- ja ymmärrä-askelien alustavia analyyskejä, mutta ei selvittää tarkemmin analyysitulosten taustalla olevia syitä.
- Kokeilla jotain ennustusmenetelmää, jos ennustustehtävä on hyvin määritelty ja dataa on riittävästi.



# OKM:n Korkeakoulu- ja tiedepolitiikan osaston kanssa aloitettu analytiikkaprojekti

1/2

## TUTKIMUSKYSYMYKSET

11.5.2022 palaverissa sovitut tutkimuskysymykset

**Tutkimuskysymys 1: mikä on lukioarvosanojen vaikutus siihen että pääsee korkeakouluun? Onko vaikutus muuttunut viimeisen viiden vuoden aikana?**

- Tarkastellaan arvosanoittain millä todennäköisyydellä saa opiskelupaikan (=pääsee johonkin korkeakouluun).
- Erikseen yo ja amk
- Vertailu miten muuttunut viiden vuoden aikana
- Kaikkien aineiden keskiarvo, alajoukkona matematiikan arvosana (lyhyt/pitkä)
- Lopputuloksena voisi olla tarkastelu, jossa lukion arvosanojen keskiarvot, esim. keskiarvot yli M mikä todennäköisyys eri vuosina, keskiarvo yli E:n jne.
- Tarkastellaan kaikkia lukion päättäneitä (todistusvalinta ja pääsykoe)
- Voiko ottaa huomioon että johonkin on hankalampi päästä kuin toisiin?
- Välivuosien merkitys?

**Tutkimuskysymys 2: Opiskelualojen eriytyminen esim. hajontaa tarkastelemalla?**

**Tutkimuskysymys 3: Ketkä eivät ole päässeet opiskelemaan ollenkaan? Esim. ei ole hakenut ollenkaan, haki ulkomaille, ei saanut paikkaa?)**

--> Tästä voisi tehdä klusterointia ja katsoa toisiko lisäymmärrystä

Näitä ryhmiä voisivat olla esim.

- ne jotka ei ole kunnolla yrittäneetkään
- ne jotka lähteneet suoraan ulkomaille opiskelemaan
- hakeneet sekä Suomeen että ulkomaille, ja menneet ulkomaille, kun ei päässeet Suomeen
- ei ole pärjänneet todistusvalinnassa / haussa

**Tutkimuskysymys 4: Minkälainen porukka ei ota paikkaa vastaan**

--> Tästä voisi tehdä klusterointia ja katsoa toisiko lisäymmärrystä

- Hakutaktikointi: ensikertalaisuus (lisäpisteitä haussa), kuinka monta paikkaa hakee,

Tavoitteena

tunnistaa ja vahvistaa yhteisesti analytiikan tekemisen tapoja

tuottaa päätöksentekoa tukevaa tietoa valtakunnallisella tasolla, erityisesti opiskelijavalinnan uudistuksen vaikutuksista

Tutkimuskysymyksiä tarkastellaan erilaisilla analytiikan menetelmillä

# OKM:n Korkeakoulu- ja tiedepolitiikan osaston kanssa aloitettu analytiikkaprojekti

2/2



## Valituksi tuleminen ja ylioppilastutkinnon arvosanat

Tämän raportin avulla voi tarkastella ylioppilastutkinnon arvosanojen sekä korkeakouluun valituksi tulemisen välistä suhdetta. Raportilla on saatavilla kolme arvosanatieta: kaikkien arvosanojen keskiarvo, lyhyen matematiikan arvosana ja pitkän matematiikan arvosana.

### Sisällysluettelo

1. Etusivu
2. Hakijat ja arvosanaryhmät
3. Valittujen osuus arvosanaryhmittäin
4. Korrelaatiokertoimet
5. Raporttiselle

Vipunen-portaalin osio:  
Korkeakoulutus ja t&k-toiminta

Lähteet:  
Opetushallituksen ylläpitämä opiskelijavalintarekisteri



Analysia on jo visualisoitu

Tutkimuskysymysten tarkastelu jatkuu

# Jatkuvan oppimisen digitalisaatio -hanke (JOD) ja analytiikan mahdollisuudet

---

1/2

- OKM tilaa CSC:ltä asiantuntijatyötä Jatkuvan oppimisen digitalisaatio (JOD) -hankkeelle 1.6.2022-31.5.2023
- Tavoitteena analytiikan, koneoppimisen ja tekoälyn mahdollisuuksien selvittäminen, toteutussuunnitelman laadinta ja kokeilut palvelukokonaisuudessa.

# Jatkuvan oppimisen digitalisaatio -hanke (JOD) ja analytiikan mahdollisuudet



Korkeakoulusta juuri valmistunut Mikael etsii oman alan työpaikkaa

Filosofian maisteriksi valmistunut Mikael on opiskellut englannin kieltä ja kirjallisuutta. Opintojen aikana hän on tehnyt lyhyitä työharjoitteluja sekä työskennellyt asiakaspalvelutehtävissä. Valmistumisen jälkeen Mikael etsii työpaikkaa, mutta ei täysin tiedä mitä työmahdollisuuksia omalla tutkinnollaan voi löytää.

Mikael haluaa selvittää minkälaisiin työtehtäviin muut saman tutkinnon suorittaneet ovat päätyneet. Hän haluaa myös selvittää minkälaisia työtehtäviä työmarkkinoilla ennakoidaan tarvittavan. Mikael hyödyntää JOD7, JOD8, JOD9, JOD10 tietoja sekä JOD12 analyyseja tilannekuvan muodostamiseksi.

Mikael ei ole varma minkälainen työ häntä kiinnostaa. Hän päättää kirjautua palveluun, jotta voi luoda itselleen osaamisprofiilin yksilön osaamisen, kiinnostusten ja reunaehtojen kartoituksella (JOD14). Hän hyödyntää koulutus- ja urasuunnittelu-työkaluja (JOD26) osaamisprofiilin taustaksi.

Osaamisprofiilin luomisen jälkeen hän suorittaa kohtaannon yksilön kohtaantotyökalulla (JOD16). Mikael saa tuloksena ehdotuksia hänelle sopivista urapoluista.

Hän haluaa vertailla erilaisia mahdollisuuksien vertailun työkalulla (JOD20). Mikael vertaa kohtaantotyökalun ehdottamia urapolkua yleisesti tietolustan ammattikuvauksiin. Tämän pohjalta hän määrittää itselleen uratoiveen. Palvelu tunnistaa urapolun mahdollistavia työtehtäviä.

Mikael tekee itselleen koulutus- ja urapolkusuunnitelman (JOD21).

- Tutkimustietovarantoon on toteutettu koneoppimismenetelmällä hankkeiden aihemallinnus, joka mahdollistaa Tiedejatutkimus.fi:ssä tutkimushankkeiden tarkastelun aiheiden mukaan.
- Aihemallinnusta tarkoitus laajentaa myöhemmin kattamaan esim. tutkimusjulkaisut.
- Lisäksi on toteutettu ns. älykäs hakutoiminnallisuus, joka tuottaa laajan hakutuloksen hakusanaan liittyvistä julkaisuista, hankkeista jne.



## Hankkeiden jakautuminen vuosittain [Lisätietoa](#)

Rajaukset (4):

2020 ✕

2019 ✕

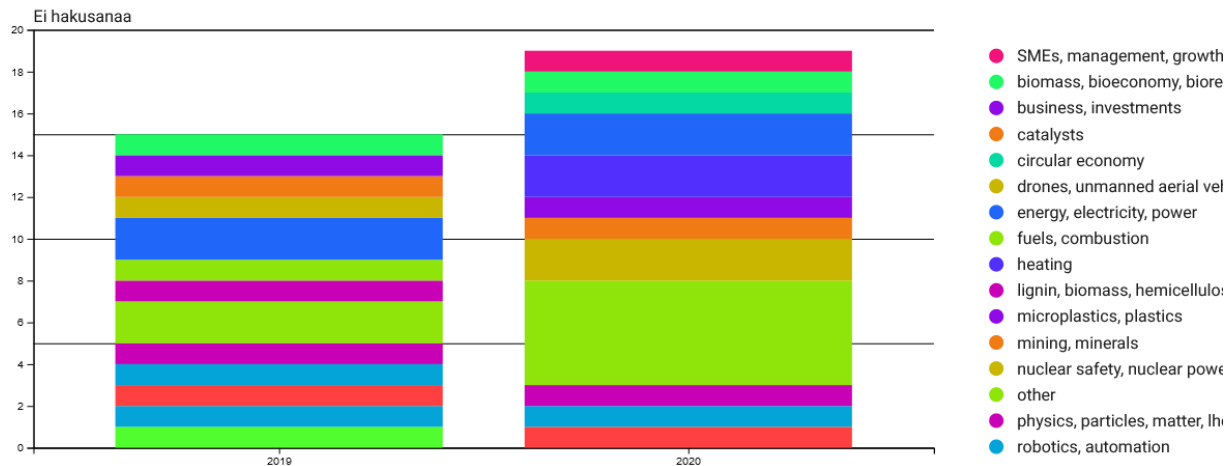
Lappeenrannan–Lahden teknillinen yliopisto LUT ✕

Euroopan unioni ✕

Tyhjennä rajaukset

Esimerkki:

Koneoppimismenetelmällä tunnistetut aiheet yhden yliopiston EU-rahoitetuille hankkeille



## Esimerkki:

Ontologioihin perustuva sisällön automaattinen annotointi tuottaa hakusanaan ”climate change” liittyvät julkaisut tunnistamalla läheiset tai liittyvät sanat



The screenshot shows the Tiedejatutkimus.fi website interface. At the top left is the logo and name 'Tiedejatutkimus.fi'. To the right are navigation links: 'Etusivu', 'Haku', and 'Tiede- ja innovaa'. Below this is a search bar with a dropdown menu set to 'Kohdistu haku' and the search term 'climate change'. A link '< Takaisin hakutuloksiin' is visible. Below the search results, there is a breadcrumb trail: 'Julkaisut' > 'Elements in the policy landscape for ac...'. The main result is a publication titled 'Elements in the policy landscape for action on black carbon in the Arctic. Supporting material to the EUA-BCA report. Enhancing the reduction of black carbon emissions to protect the Arctic: mapping the policy landscape of national, regional and international actions'. The publication date is listed as 'Julkaisuvuosi 2021' and the authors as 'Tekijät Åström, Stefan; Hildén, Mikael; Matthews, Bradley'.

Uutinen

TIVI 19.8.2022  
**Tieteen tietotekniikan keskus lähtee  
kehittämään vaihtoehtoa Googlelle**



**OpenWebSearch.EU** -hankkeessa (2022-25) kehitetään avointa eurooppalaista verkkohakuinfrastruktuuria.

CSC osallistuu projektissa mm. hakuindeksin, tallennusinfrastruktuurin, kielimallien sekä avoimen verkkohaun ekosysteemin kehittämiseen.

Yhtenä osa-alueena kehitetään suomen- ja ruotsinkielisten tieteellisten tekstiaineistojen hakua, pohjalla mm. Tiedejatutkimus.fi:ssä tehty työ.



Services for  
Research and Education Information  
Management and Data Analytics



**Hanna-Mari Puuska**



**Heini Manninen**



[facebook.com/CSCfi](https://facebook.com/CSCfi)



[twitter.com/CSCfi](https://twitter.com/CSCfi)



[linkedin.com/company/csc--it-center-for-science](https://linkedin.com/company/csc--it-center-for-science)



[github.com/CSCfi](https://github.com/CSCfi)