

# Tekoäly- ja ohjelmistorobotiikkakokeilun loppuraportti - syksy 2020

Kurssi- ja tapahtumatarjonnan suunnittelun tukeminen tekoälyä hyödyntäen

Helsinki



Tukea digitalisaatiokokeiluihin kaupungin työntekijöille

# Kurssi- ja tapahtumatarjonnan suunnittelun tukeminen tekoälyä hyödyntäen

Tiimi

Helsingin kaupunki: Romeo Pulli

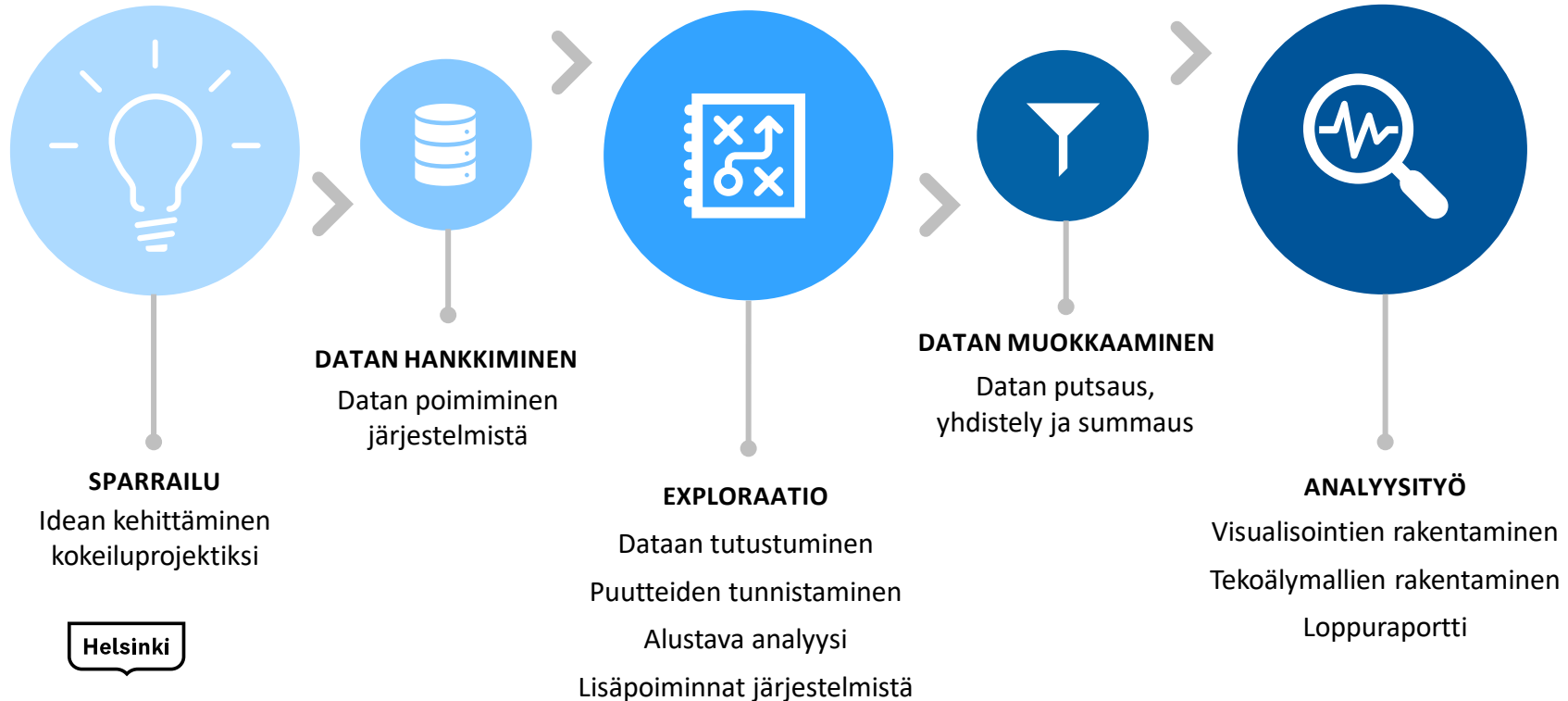
SAS: Antti Heino, Jussi Martikka, Kim Molin

Gofore: Tommi Vehviläinen

# 1. Kokeilun onnistuminen

- *Kokeilun perusteella ML ja analytiikkatyökalut voivat tarjota arvokkaan lisän kurssi- ja tapahtumatarjonnan suunnitteluun*
- *Syrjäytymisen ehkäisyyn liittyviä vaikutuksia ei käytetyllä aineistolla kyetty testaamaan*
- *Nykyisin saatavissa oleva data oli käyttökelpoista*
- *Monista relevanteista asioista dataa on saatavilla vain vähän tai ei ollenkaan*
- *Potentiaalisia henkilöstösäästöjä ei vielä tämän kokeilun perusteella kyetä arvioimaan*
- *On pidettävä todennäköisenä, että kokeilun kaltaiset työkalut lisääisivät tarjonnan houkuttelevuutta*

# 2. Kokeilun eteneminen



# 3. Kokeilun tuotokset

Kokeilun tuotoksena syntyi kolmen tasoisia asioita



Datan muokkaukset, yhdistely ja summaaminen



Interaktiiviset visualisoinnit datan tutkimiseen, joilla voidaan helpottaa kurssisuunnittelua



Tekoälymalleja, jotka toimivat esimerkkeinä siitä miten niitä voitaisiin hyödyntää kurssisuunnittelun tukena

# Kursseja voidaan verrata osallistujien ikäjakauman perusteella tai tutkia yhden kurssin ikäjakaumaa tarkemmin histogrammina

Kohderyhmä

Työkäiset

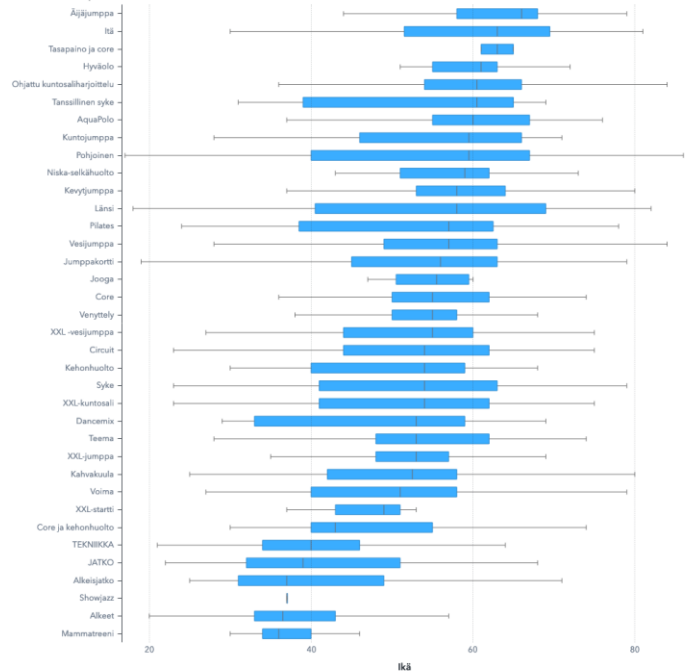
Kurssityyppi

Ohjattu kuntosaliharjoittelu

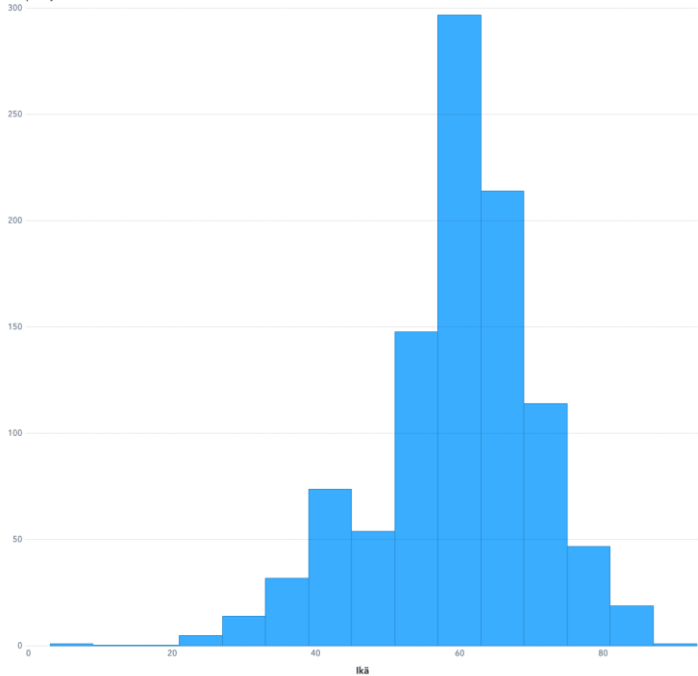
Kurssi\_ID

Osallistujien ikä kurssityypeittäin

Reservation Group



Frequency



# Uutta kurssia suunniteltaessa voidaan tarkastella yksittäisen kurssin historiatietoja ja saada käsitys kurssin suosioista aiemmilla kerroilla

Filters: ALKEET YLI 7 -VUOTIAAT x

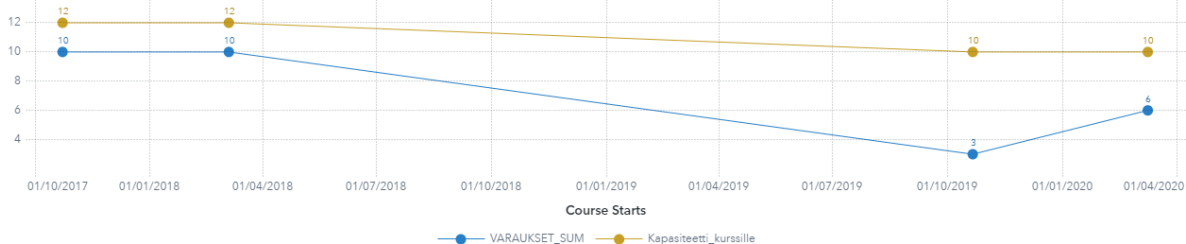
ALKEET YLI 7 -VUOTIAAT

Paikka

Course\_tags

Varaukset ja kapasiteetti

VARAUKSET\_SUM / Kapasiteetti\_kurssille

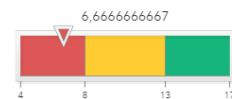


Ikä keskiarvo

10

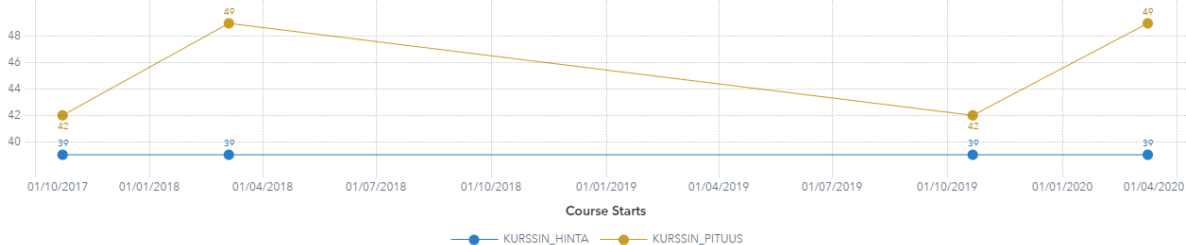
Course\_tags:Lapset, nuoret ja perheet

Varauksia keskimäärin

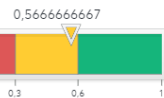


Kurssien hinta ja pituus

KURSSIN\_HINTA / KURSSIN\_PITUUS



%-osuus joka varataan päivässä

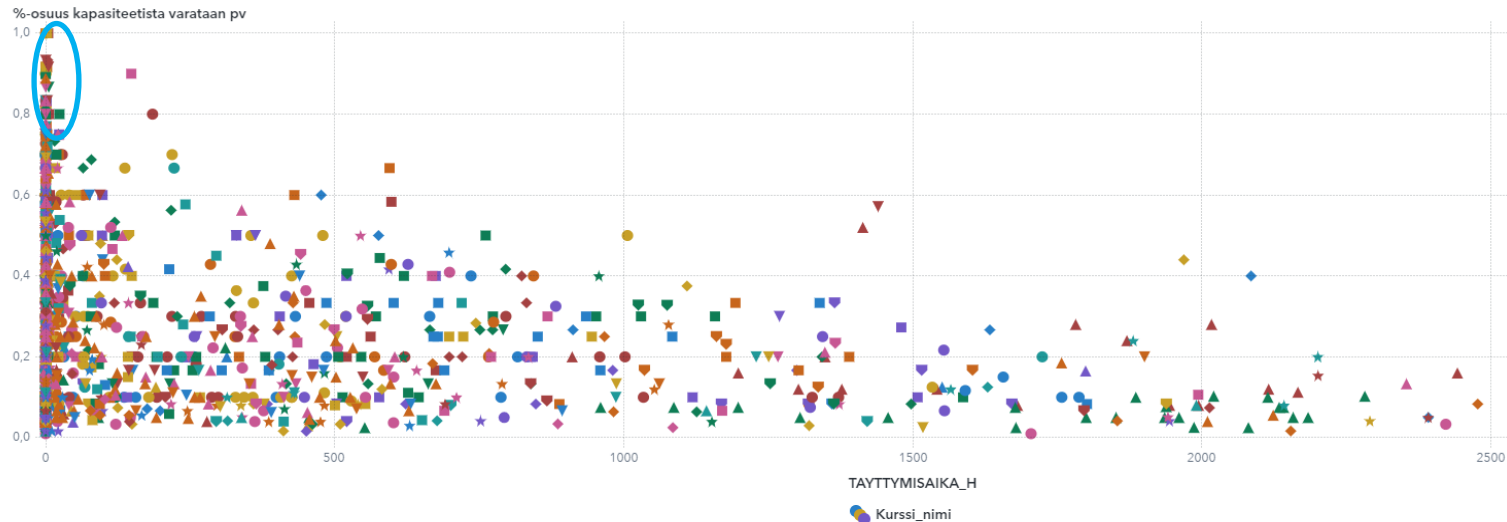


Täyttymisaika (h)

48

Erittäin suosittu kurssit hahmottuvat helposti kun kursseille on laskettu täyttymisaika ja kuinka suuri %-osuus niistä varataan päivässä. Osa kursseista menee melkein heti täyteen, joten niitä voisi järjestää lisää.

Nopeasti täyttyvät kurssit



TAYTTYMISAIKA_H	%-osuus kapasiteetista varataan pv	Kurssi_nimi
0,08		1 ALKEET 5-6 -VUOTIAAT
0,13		1 ALKEISJATKO 5-6 -VUOTIAAT
4,36		1 ALKEISJATKO yli 7 -vuotiaat
2,36		1 JATKO 1
3,57	0,9333333333	EasySport -Parkour, akrobatia ja sirkuslajit
0,22	0,9333333333	Lapsi-ikäinen temppujumppa 2-3 -vuotiaat
0,11	0,9166666667	ALKEISJATKO 5-6 -VUOTIAAT

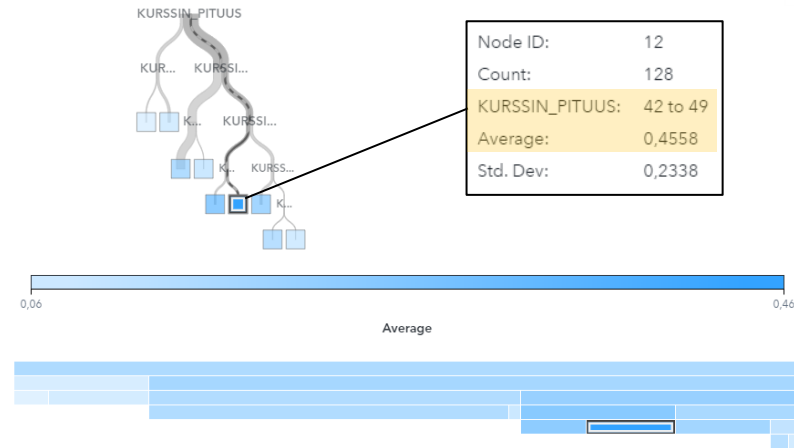


Tekoälymalleilla voidaan analysoida mitkä tekijät vaikuttavat kurssien suosioon. Yksinkertaisimmillaan ne voivat olla esim. päätöspuualgoritmeja, joissa käytetään vain yhtä ennustavaa tekijää.

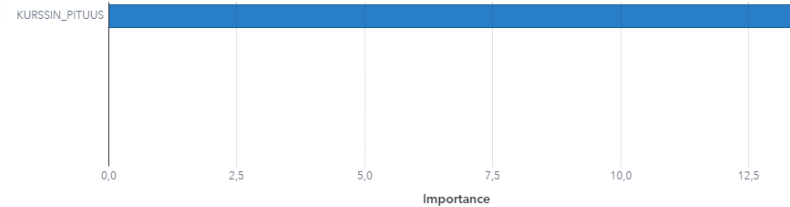
Esim. päivässä varattavaa kapasiteetin osuutta ennustetaan kurssin pituudella. Mallin avulla nähdään, että 42-49 päivän kurssit näyttävät olevan suosituimpia keskimäärin. Tietoa voidaan hyödyntää tulevien kurssien suunnittelussa.

Decision Tree %-osuus kapasiteetista varataan pv Validation ASE 0,0439 Observations Used 1 608 Unused 1 Create Pipeline

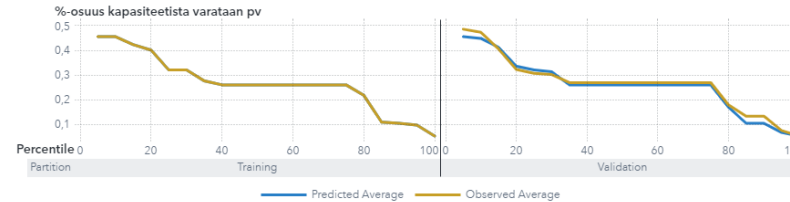
Tree



Variable Importance



Assessment



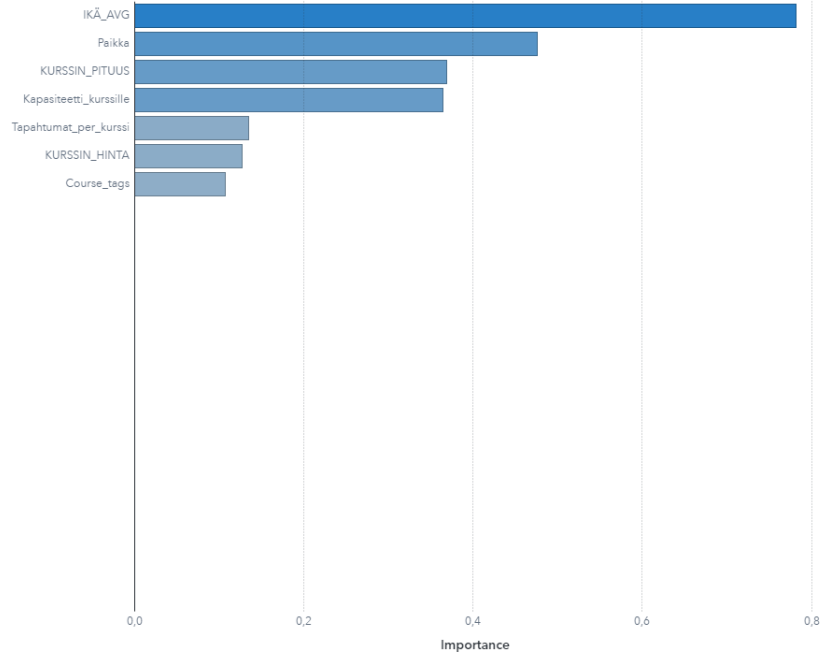
Node Statistics Node Rules Variable Importance Cost-Complexity Pruning Assessment Assessment Statistics

Node ID	Depth	Parent ID	N Children	Type	Observations	% Observations	Percent of Parent	N Missing	Gain	Predicted Value	Average	Std. Dev
12	4	9	0	Leaf	128	11,38 %	57,92 %	0	0,0000	0,4558	0,4558	0,2338
13	4	10	0	Leaf	136	12,09 %	76,40 %	0	0,0000	0,3221	0,3221	0,2275
14	4	10	2	Class	42	3,73 %	23,60 %	0	0,1316	0,1874	0,1874	0,1168

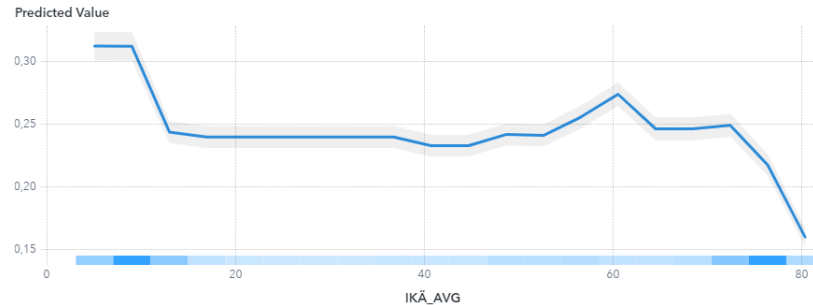
Tekoälymalleja on myös monimutkaisempia (esim. gradient boosting), jolloin niiden ennusteita täytyy tulkita esim. Partial Dependence – menetelmällä. Alla olevaan malliin on kurssin pituuden lisäksi lisätty muitakin tekijöitä, joilla voi olla vaikutusta kurssin suosioon.

Gradient Boosting %-osuus kapasiteetista varataan pv Validation ASE 0,0307 Observations Used 1 590 Unused 1 Create Pipeline

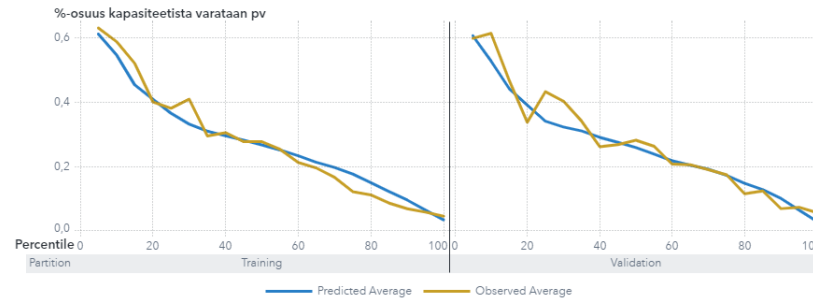
Variable Importance



Partial Dependence

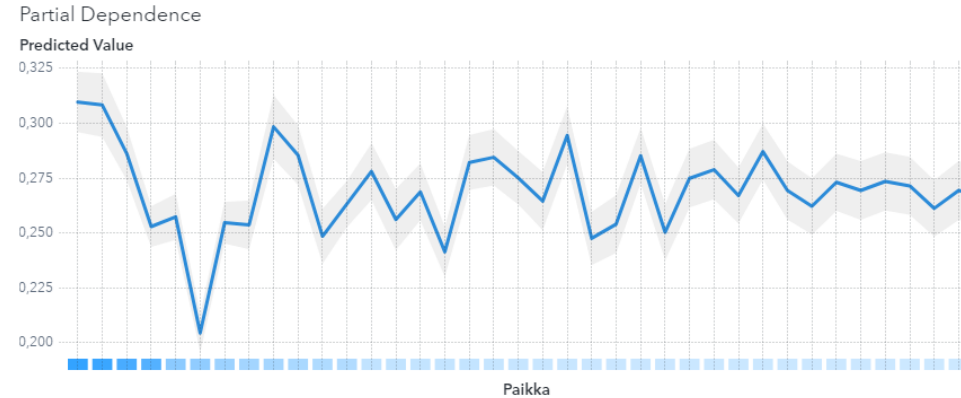
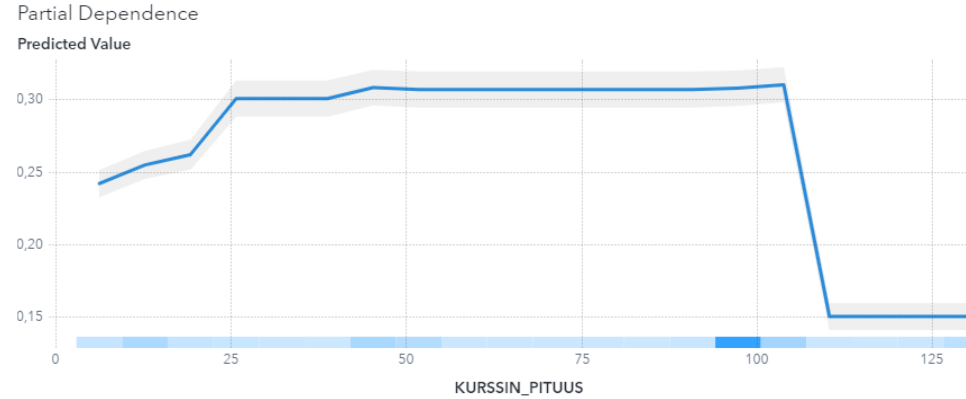
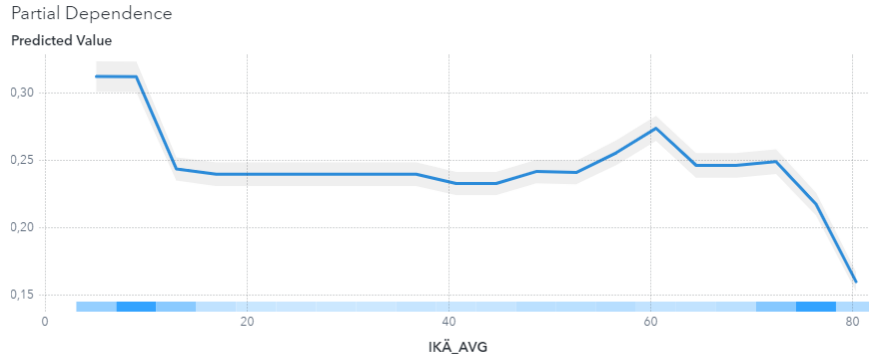


Assessment



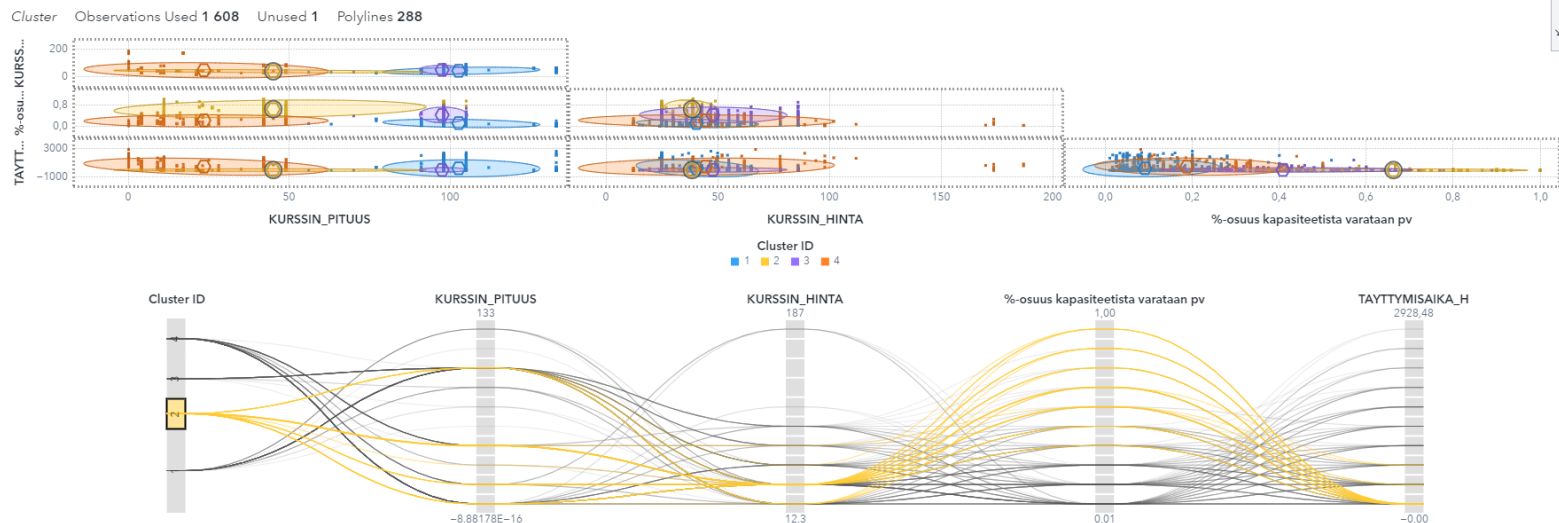
## Suosituimmat (nopeimmin täyttyvät) kurssit näyttäisivät olevan

- 25-105 pv mittaisia
- Suunnattu 5-9 vuotiaille
- Paikkana esim. parkour akatemia tai uimastadion



Klusteroinnilla voidaan ohjaamattoman oppimisen keinoin löytää datasta samanlaisia kursseja ja tehdä johtopäätöksiä suosituista kursseista.

Esim. klusteri 2: korkeintaan keskipituiset, halvat kurssit ovat saaneet hyvin osallistujia ja täyttymisaika on ollut erittäin lyhyt



Centroids Cluster Summary Model Information Within Cluster Statistics Iteration History Standardization Interval information Parallel Coordinates Plot

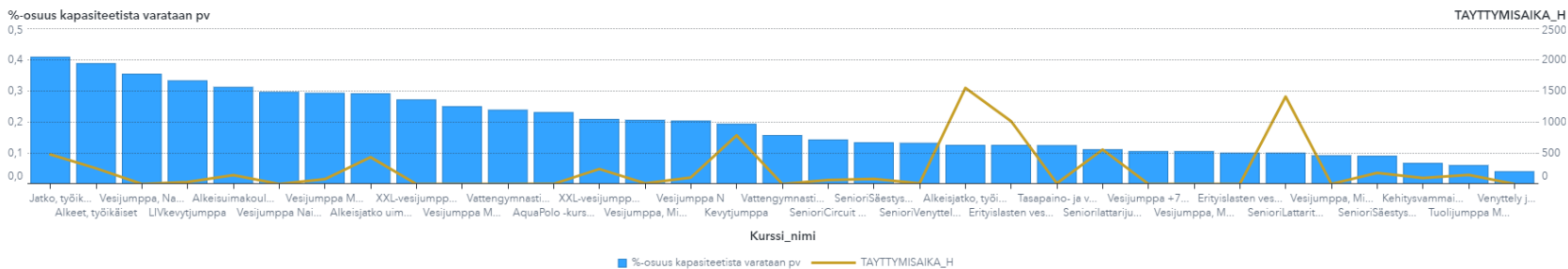
Cluster ID	KURSSIN_PITUUS	KURSSIN_HINTA	%-osuus kapasiteetista varataan pv	TAYTTYMISAIKA_H
1	103,47345768	40,934863702	0,10	275,14
2	43,943548387	37,97983871	0,64	26,12
3	97,936170213	48,063829787	0,42	51,30
4	24,092814371	45,001796407	0,20	508,06

Kursseja voidaan tutkia myös paikkakohtaisesti jos halutaan suunnitella tietyssä paikassa järjestettävää kurssitarjontaa.

Pylväsdiagrammista ja taulukosta nähdään esim. mitkä ovat menneet nopeimmin täyteen ja mihin on jäänyt vähiten/eniten tyhjiä paikkoja. Tiedon perusteella voidaan valita mitä kannattaa järjestää uudelleen.

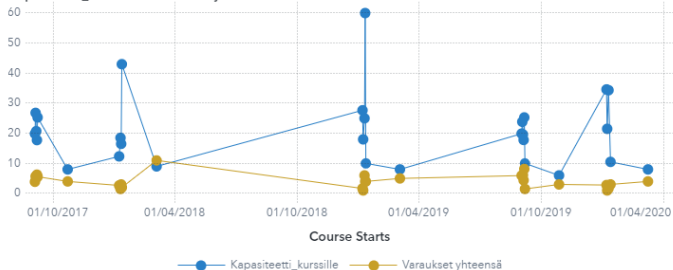
Kampin liikuntakeskus

Kapasiteetti ja täyttymisaika



Varausten ja kapasiteetin KA

Kapasiteetti\_kurssille / Varaukset yhteensä



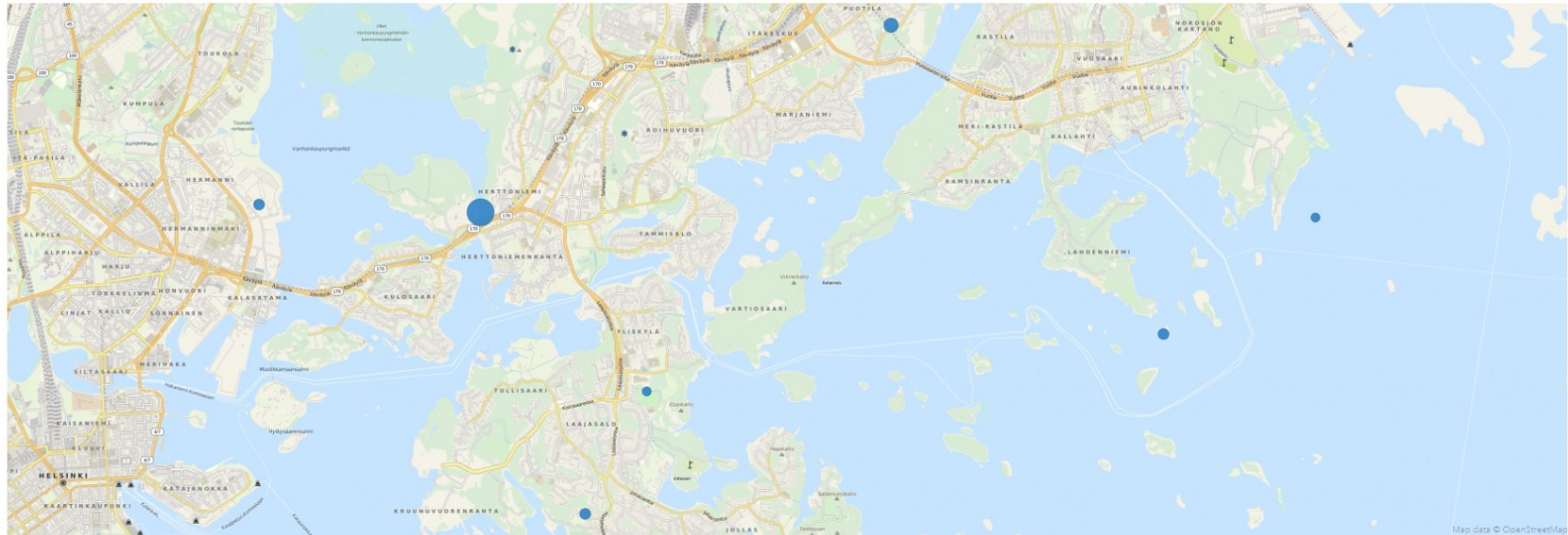
Täynnä olevat kurssit

Kurssi_nimi	Tyhjät paikat	Tyhjät paikat suht kapa ▲	Varaukset yhteensä	Kapasiteetti_kurssille
Jatko, työkäiset	1,00	0,14	7,5	8,5
Alkeisjatko, työkäiset	3,00	0,38	5,0	8,0
Alkeet, työkäiset	3,50	0,39	5,5	9,0
Alkeisuimakoulu, työkäiset	4,00	0,50	4,0	8,0
Alkeisjatko uimakoulu, työkäiset	3,50	0,50	3,5	7,0
LIVkevytjumppa	8,50	0,57	6,5	15,0
Vesijumppa, Naiset	12,31	0,68	6,4	18,2
Vesijumppa Naiset	12,75	0,70	5,4	18,1
XXL-vesijumppa M + N	11,00	0,71	4,6	15,6
Vesijumppa M + N	13,17	0,72	5,4	18,2
XXL-vesijumppa Naiset	13,50	0,73	5,0	18,5
Vesijumppa N	13,33	0,74	4,7	18,0

Kurssin järjestämispaikan (tai vaikka yksittäisen kurssin) perusteella voidaan myös katsoa miltä alueilta osallistujat tyypillisesti tulevat. Näkymän muodostamiseen käytettiin Enkoran postinumerodataa.

Herttoniemenrannan liikuntahalli

Frequency of Coordinates



11  
1  
Frequency

Helsinki

Yksityiskohtaista tarkastelua varten dataa voidaan myös katsoa perinteisemmin taulukkomuodossa.

Dataa yhdistelemällä ja summaamalla saadaan tietoon esim. kurssit joilla on kaikkein vähiten tyhjiä paikkoja suhteessa kapasiteettiin. Näille kursseille voisi siis olla enemmänkin kysyntää muissakin paikoissa.

Kurssi_nimi	Paikka	Tyhjät paikat	Tyhjät paikat suht kapa ▲	%-osuus kapasiteetista varataan pv	Varaukset yhteensä	Kapasiteetti_kurssille
ALKEISJATKO, 5-6 -vuotiaat		0,00	0,00	0,90	10,0	10,0
Jatko, työikäiset	Kampin liikuntakeskus	1,00	0,14	0,41	7,5	8,5
JATKO 1, YLI 6 -VUOTIAAT		2,25	0,19	0,79	9,8	12,0
EasySport -Parkour, akrobatia ja sirkuslajit	Kauppakeskus Redi, Gym...	3,00	0,20	0,20	12,0	15,0
EASYSPOORT MELONTA 12-15 -vuotiaat	Helsingin Melontakeskus	4,00	0,20	0,20	16,0	20,0
ALKEET 5-6-VUOTIAAT	Jakomäen uimahalli	2,00	0,20	0,50	8,0	10,0
JATKO 2	Uimastadion	2,50	0,21	0,54	9,5	12,0
JATKO 1	Kumpulan maauimala	2,50	0,21	0,46	9,5	12,0
Easy Hockey	Hernesaaren jäähalli	5,67	0,23	0,27	19,3	25,0
JATKO 1	Uimastadion	2,75	0,23	0,75	9,5	12,0

# 4. Opit tekoälyn mahdollisuuksista

- Tarjonnan parempi kohdistaminen keskeisiksi koetuille asiakasryhmille
- Tulosten parempi ennustettavuus
  - Budjetointi
  - Vaikutukset tavoitteiden toteutumiseen (esim. Syrjäytyminen tai liikuntaan aktivointi)



# 5. Opit tekoälyn kehittämisestä

- Määriteltävä nykyistä selkeämmin tavoitteet ja mittarit
- Puutteet datan laadussa
- Puutteet datan saatavuudessa

# 6. Opit kokeilemisesta

- Hankintaprosessi
  - Raskas hankinnan suuruuteen suhteutettuna
  - Keskitetympi hankinta kokeiluihin; mahdollisesti jo ennen kokeiluhakuja
- Ympäristö
  - Sopimusasiat saatava tehtyä mahdollisimman nopeasti, koska tunnuksia voidaan tilata vasta sen jälkeen
  - Mitä paremmin voidaan ennakoida sopimusten valmistumisen aikataulua niin sitä helpommin resurssit saadaan varattua oikealle hetkelle

# 7. Opit resursoinnista

- Datat keräämiseen hyvä varata riittävästi aikaa ja varautua siihen, että joitakin poimintoja täytyy tehdä uudestaan tai niitä täytyy tehdä lisää
- Tekeminen jakaantui suhteellisen tasaisesti, painottuen hiukan enemmän loppuun
- Yhdistyi osittain KuVa:n tietopyyntöön

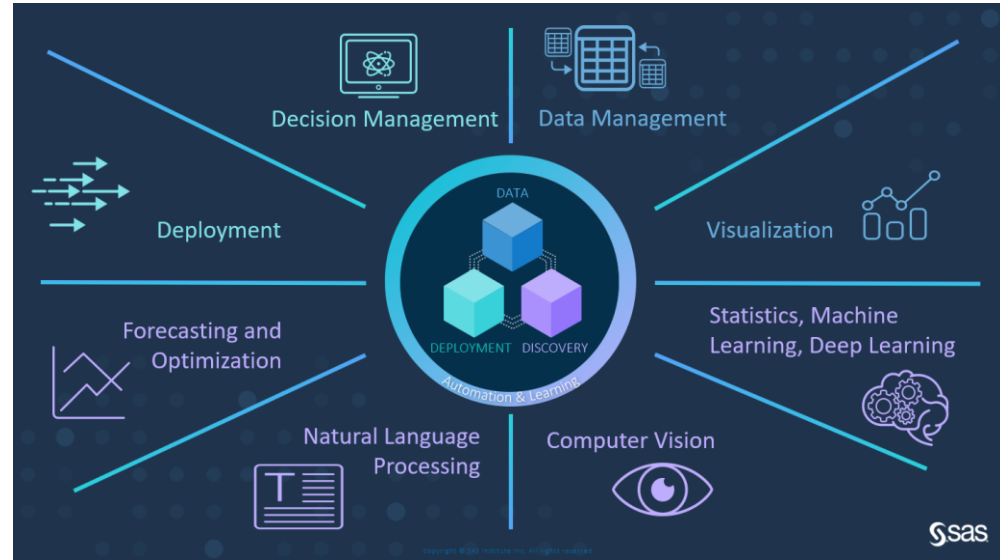
# 8. Kokeilun tekninen ympäristö

- Kokeilussa käytettiin SAS Viya analytiikka- ja tekoäly alustaa
- Alusta on asennettuna Microsoft Azureen
- Samaa alustaa hyödynnettiin jo viime kokeilukierroksella ja se otettiin käyttöön nyt uudestaan



# 8. SAS Viyalla pystytään toteuttamaan valtava määrä erilaisia tekoälyn käyttötapauksia

- Samaa alustaa pystyy hyödyntämään moneen tekoälyn käyttötapaukseen
  - Koneoppiminen
  - Syväoppiminen
  - Konenäkö
  - Tekstianalytiikka
  - Aikasarjaennustaminen
  - Optimointi
  - Tilastollinen analyysi
  - Visualisointi
- Alustassa myös vahvat kyvykkyudet tuotannollistamiseen, jotta projektit eivät jää kokeiluksi ja niistä saadaan jatkuvia hyötyjä
- Välttää vaikeasti hallittavalta joukolta pisteratkaisuja kun yhdellä alustalla voi ratkaista monta haastetta



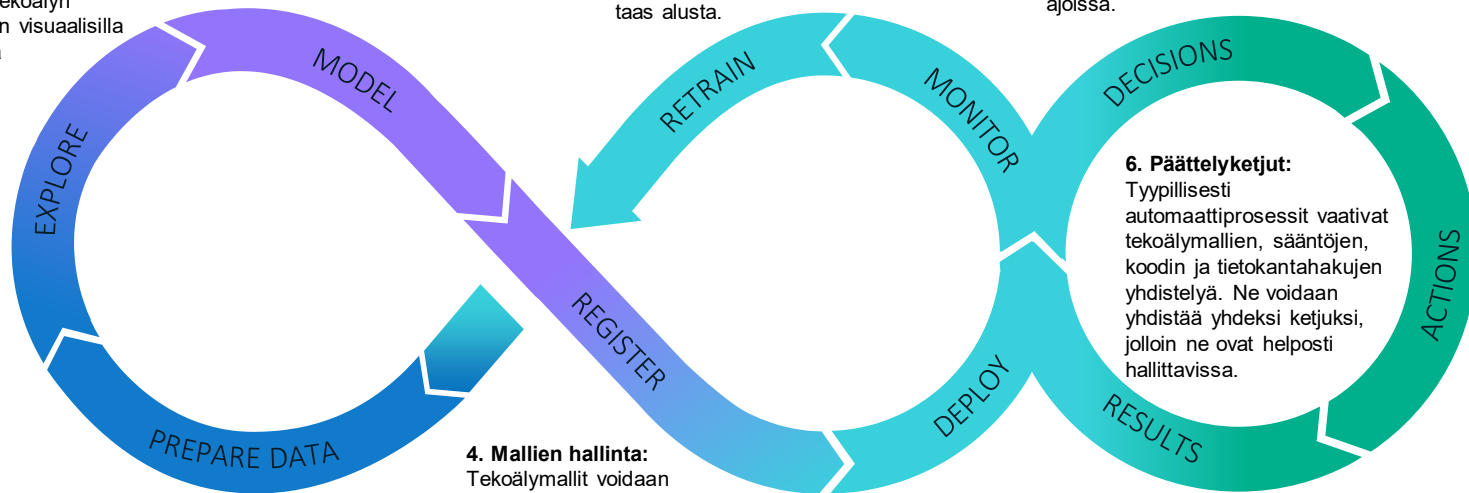
# 8. SAS Viya mahdollistaa tekoälyn elinkaaren kaikki vaiheet ja varsinkin tuotannollistamisen

**2. Explorointi & visualisointi:** Datan nopea tutkiminen ja tekoälyn hyödyntäminen visuaalisilla käyttöliittymillä

**3. Tekoölymallien kehitys ja hienosäätö:** Erilaisia algoritmeja on nopeaa verrata toisiaan vastaan ja saada paras mahdollinen malli datalle

**8. Uudelleen koulutus:** Data ja ymmärrys siitä harvoin pysyvät samanlaisena ikuisesti, joten jossain vaiheessa on syytä uudelleen kouluttaa mallit, jolloin elinkaari alkaa taas alusta.

**7. Monitorointi:** Tekoölymallien performanssia voidaan seurata, jotta mahdollinen uudelleen kouluttamisen tarve huomataan ajoissa.



**1. Datan valmistelu:** yhdistely, muutokset, uusien muuttujien luonti etc.

**4. Mallien hallinta:** Tekoölymallit voidaan rekisteröidä keskitettyyn mallien hallintaan, jotta tiedetään missä ne ovat, kuka ne on tehnyt ja mikä versio on käytössä.

**5. Tuotannollistaminen:** Tekoölymallit voidaan tuotannollistaa sellaisenaan (eräajo / reaaliaikainen) tai lisätä osaksi päättelyketjuja. Näin ne voivat päivittää esim. näkymiä tasaisin väliajoin tai palvella automaattiprosesseja.

## 6. Päättelyketjut:

Tyypillisesti automaattiprosessit vaativat tekoölymallien, sääntöjen, koodin ja tietokantahakujen yhdistelyä. Ne voidaan yhdistää yhdeksi ketjuksi, jolloin ne ovat helposti hallittavissa.

# 9. Kokeilun data

- Käytetyt datalähteet
  - Myynti: kursseille ilmoittautumiset ja peruutukset
  - Kurssitiedot: kurssin perustiedot
  - Enkora: kurssille osallistuneiden ikä ja postinumero
  - Koordinatit postinumeroille
- Data poimittiin järjestelmistä tiedostoiksi ja vietiin SAS Viyan muistiin analysoitavaksi
- Opit
  - Tärkeää olla eri datataulujen yhdistämiseen tarvittava avain
  - Täytyy olla tarkkana, että saadaan koko aineisto ja rivejä ei puutu
  - Datan standardointi on tärkeää sen hyödyntämisen kannalta (esim. kurssien nimet vapaamuotoisia ja siksi voivat olla kirjoitettu eri tavalla vaikka olisi kyse samasta kurssista)

# 10. Jatkopäätökset ja -ideat

- Huomioidaan kokeilu uudistettaessa KuVa:n kurssi- ja tapahtumasuunnittelun kokonaisuutta
- Kiinnitetään huomiota datan keräämiseen
- Tietomallien ja käsitteiden yhtenäistäminen