

AnalytiikkaÄly

PedaForum 2019



AnalytiikkaÄly
AnalyticsAI

Oppimisanalytiikka opiskelun, ohjauksen
ja johtamisen tukena yliopistoissa



Koneoppiminen opetuksen kehittämisen tukena

Riku Hietaniemi, yliopisto-opettaja, CMVS / OY



Miksi älykästä analytiikkaa?

- Korkeakoulujen ja tutkinto-ohjelmien johtamiseen ja kehittämiseen on perinteisesti käytetty erilaisia tilastollisia indikaattoreita.
 - Valmistuneiden määrä, keskimääräinen opintopistekertymä jne.
- Tällaisella yksinkertaisella tarkastelulla voidaan havaita vain jäävuoden huippu.
 - Oppivia algoritmeja hyödyntämällä päästään pintaa syvemmälle!



Minkälaiseen dataan älyä voi soveltaa?

- Mihin vain, mutta kannattaa pitää mielessä koneoppimisen kultainen sääntö:

”Garbage in, garbage out”

- On siis pidettävä huoli, että käytettävä data on asiallisesti esikäsitelty.
- Tuloksia tulkintaan on myös syytä suhtautua kriittisesti.



Käytetyt aineistot

- Lukiodata – Yo-kirjoitusten arvosanat suoritetuista aineista kaikille Suomen lukiolaisille v. 2015S-2018S, n=95k.
 - Opiskelija ID puuttuu, eikä tietoa voi yhdistää muihin aineistoihin.
 - Ei ole tietoa jatko-opinnoista.
- Lukiodata – Yo-kirjoitusten arvosanat Oulun yliopistoon valituille lukiolaisille v. 2006-2015, n=13k.
 - Opiskelija ID:n avulla aineisto voidaan yhdistää opintorekisteritietoihin.
 - Ei sisällä opiskelijoita, jotka ovat tulleet yliopistoon muuta reittiä.
- Opintorekisteridata – Oulun yliopiston opiskelijadata v. 2006-2018.
 - Tutkinto-ohjelmatiedot, suoritettut kurssit (arvosana ja pvm).



Aineistojen haasteet

- Sisäänoton muutokset
 - Esim tietotekniikan tutkinto-ohjelmassa sisäänottomäärä putosi 2014 100:sta 50-60:een.
 - Tutkinto-ohjelmia ja suuntautumisvaihtoehtoja on myös poistunut ja uusia tullut tilalle.
- Muutokset tutkintorakenteissa
 - Erityisesti pakollisten kurssien muutokset.
- Omaopettajatoiminnan varsinainen alkaminen v. 2012
 - HOPS pakolliseksi 2013, jolloin osallistuminen muuttui myös pakolliseksi.
- Suhdannevaihtelut
 - Työelämään siirtyminen ennen valmistumista
- Armeija
 - Etenkin miesvaltaisilla aloilla vaikutus huomioitava.



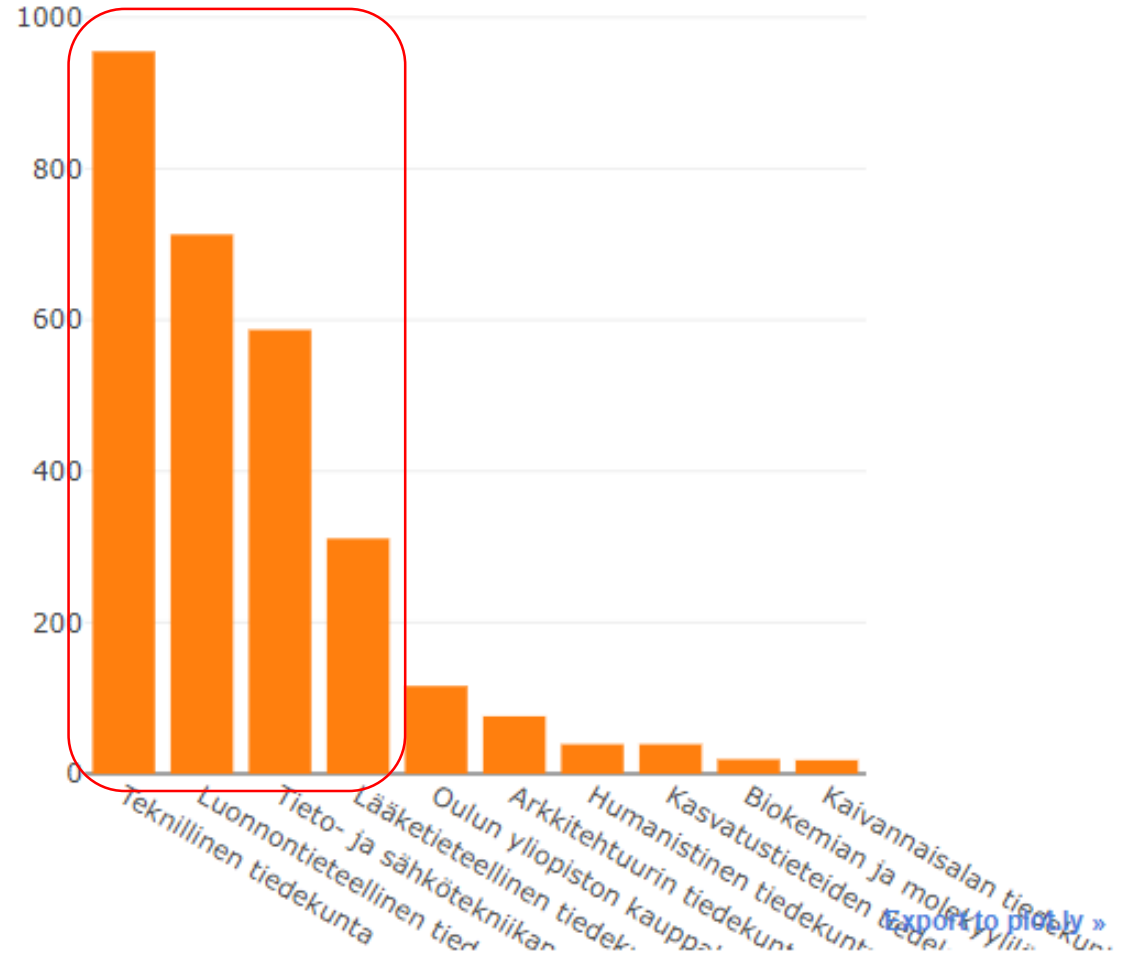
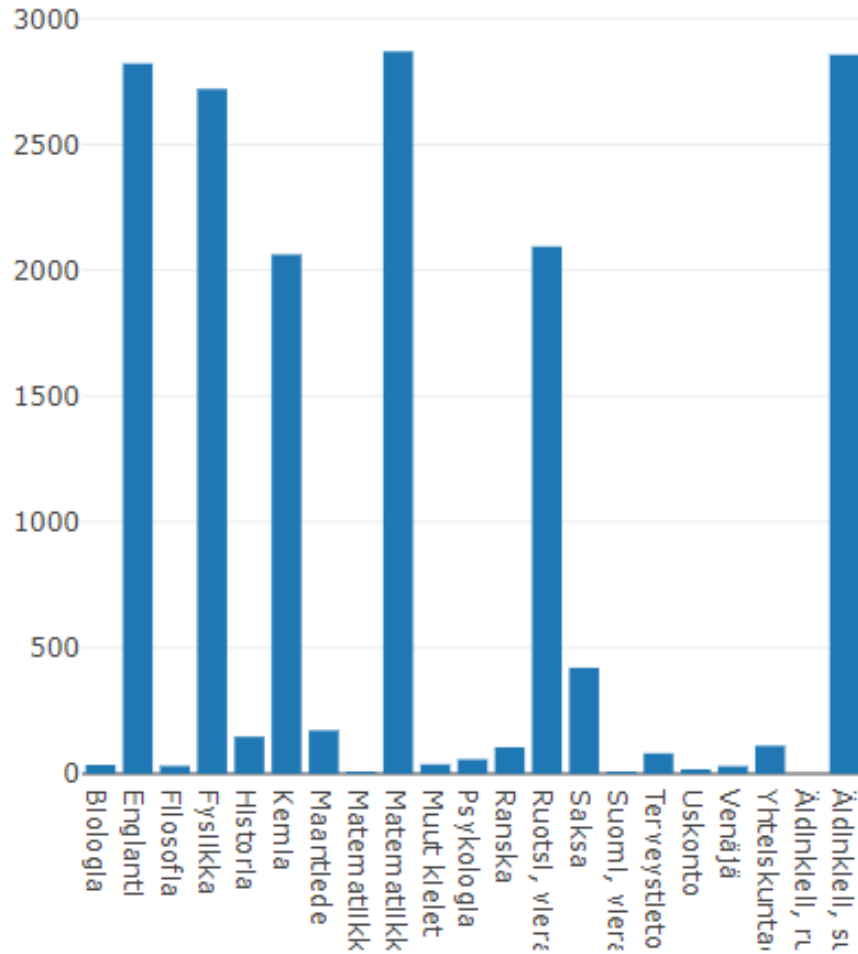


Ylioppilasdata ja korkeakoulutukseen hakeutuminen

Opiskelijoiden klusteroituminen ”hakijaprofiileihin” K-means ja UMAP menetelmillä



Klusteri 2/9



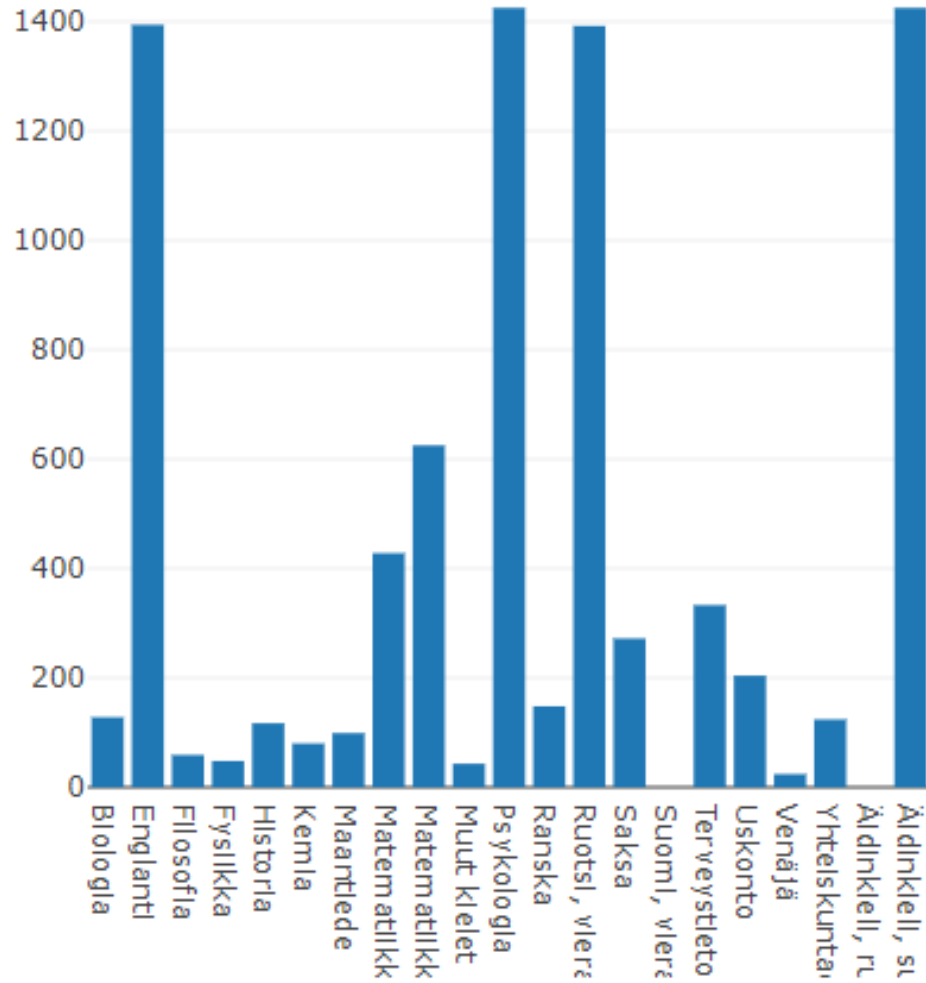
Keskiarvo: 4.943
Opiskelijoiden lkm: 2873

Klusterianalyysi: ainevalinnat teknillis-luonnontieteellisessä klusterissa



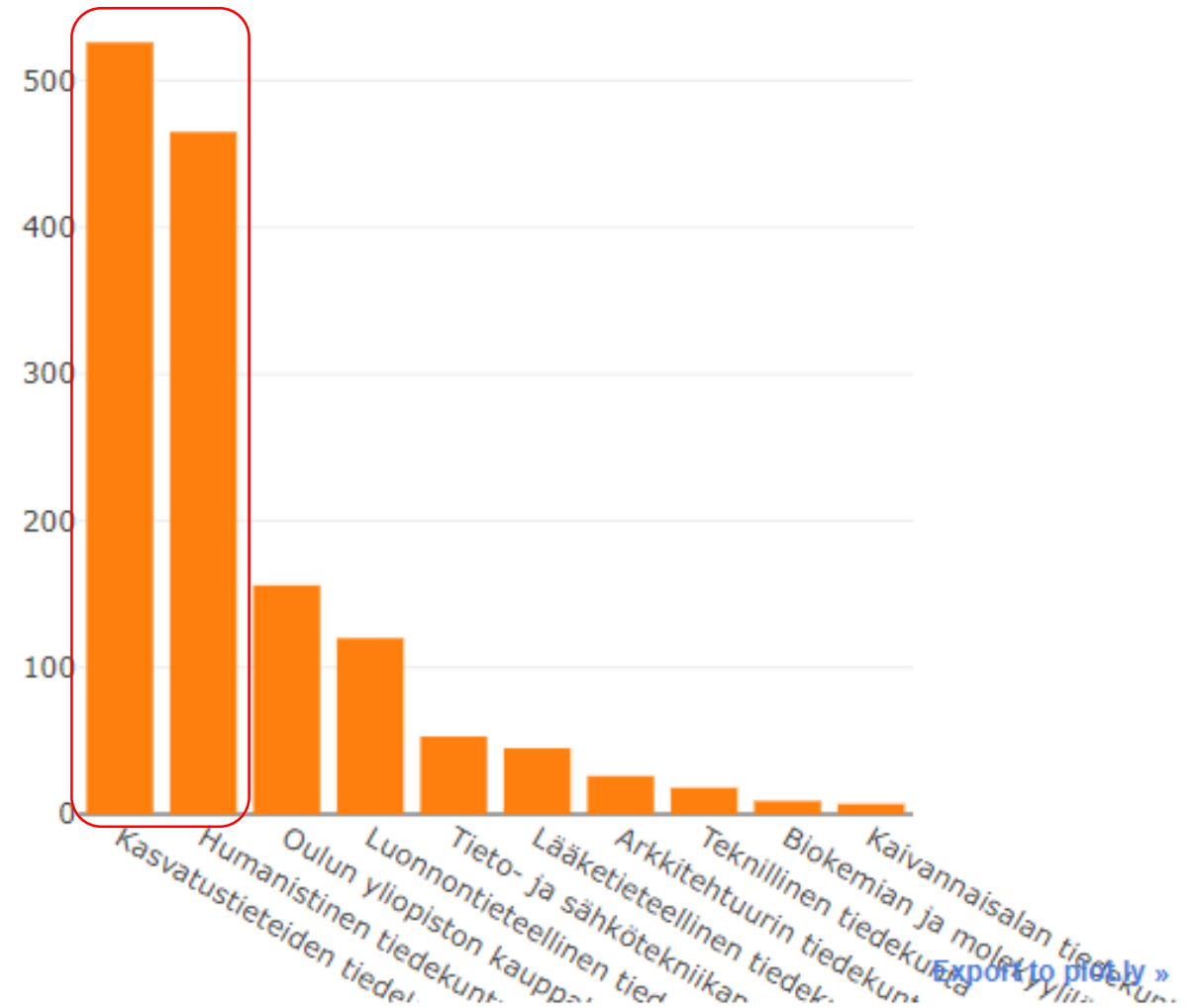
AnalytiikkaÄly
AnalyticsAI

Klusteri 6/9



Keskiarvo: 4.884

Opiskelijoiden lkm: 1425

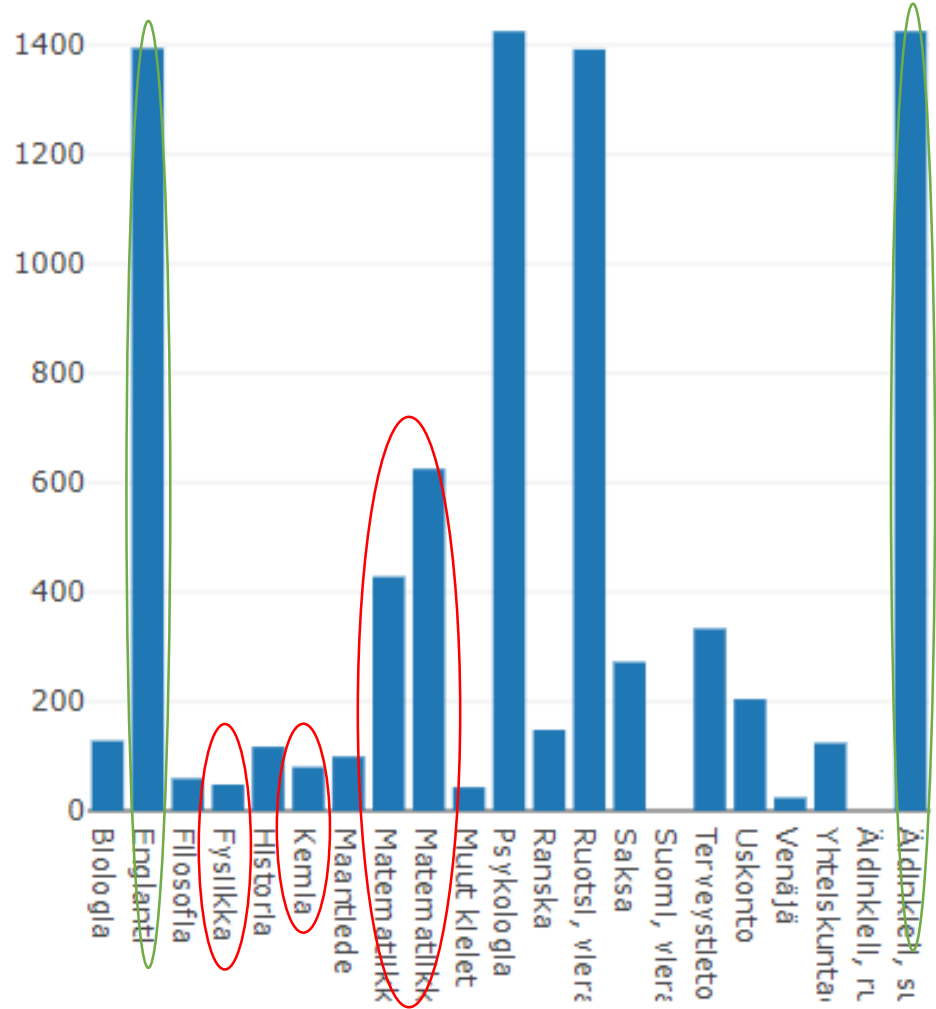


Klusterianalyysi: ainevalinnat kasv.tiede/humanisti -klusterissa



AnalytiikkaÄly
AnalyticsAI

Kasvatustieteet



Keskiarvo: 4.884
Opiskelijoiden lkm: 1425

Tekniikka



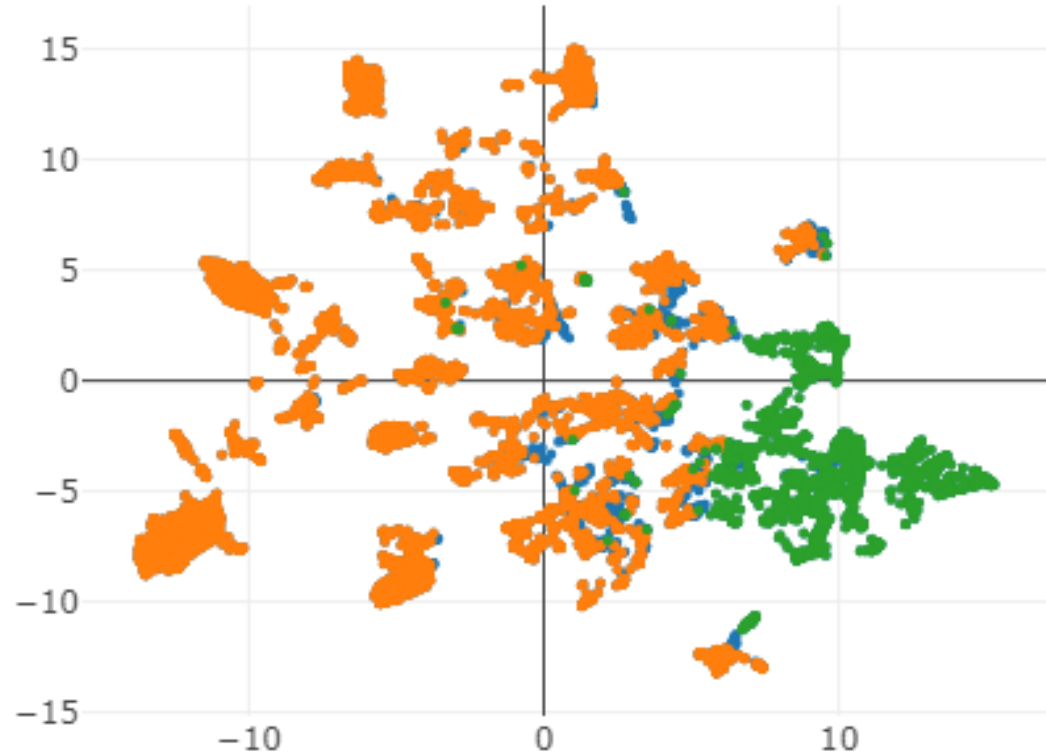
Keskiarvo: 4.943
Opiskelijoiden lkm: 2873

Klusterianalyysi: Vertailu keskenään

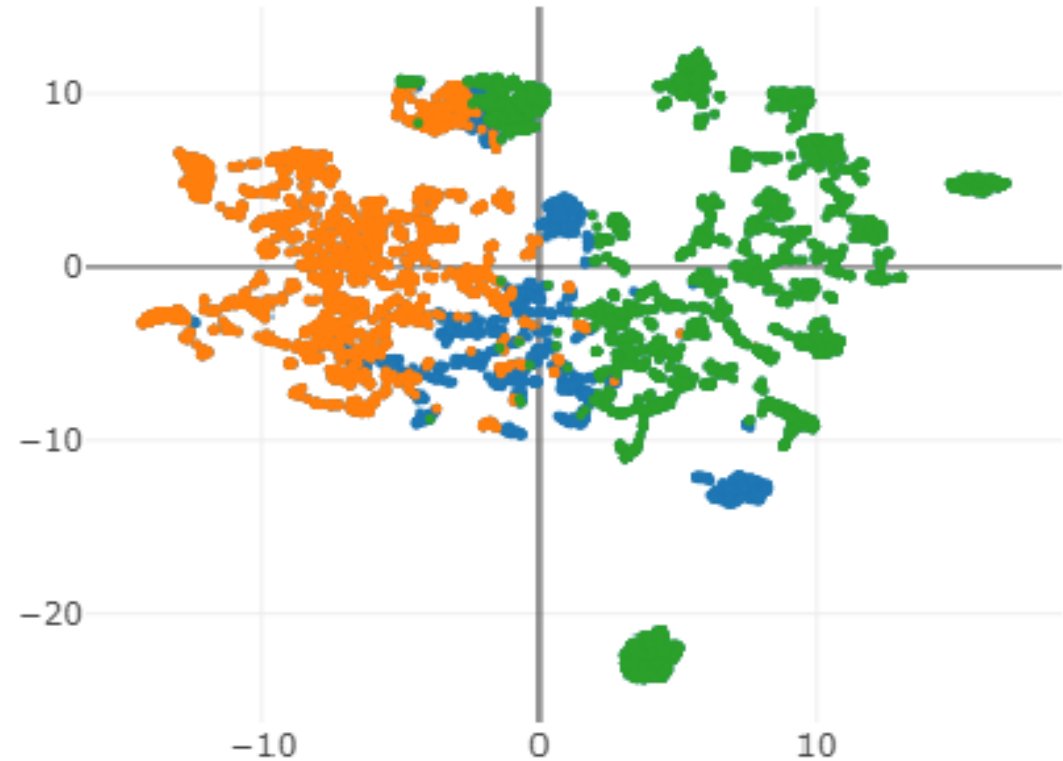


AnalytiikkaÄly
AnalyticsAI

Oulun yliopisto 2006 - 2015 (n=12767)



Ylioppilaat 2015S-2018K otos (n=10000)



- Ei kumpaakaan
- Matematiikka, pitkä oppimäärä
- Matematiikka, lyhyt oppimäärä

Unifold Matrix Approximation and Projection (UMAP)-esitys
YO-menestyksestä visualisoituna matematiikan suhteen





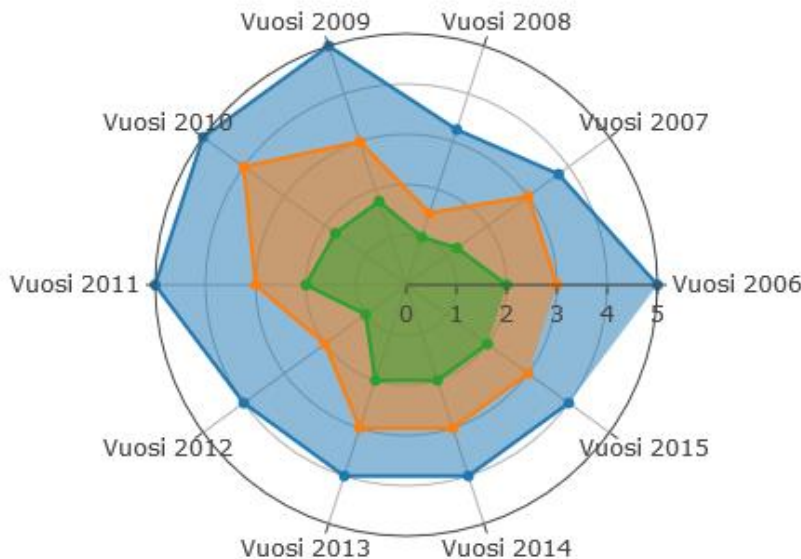
Opintorekisteridata

Mikä voisi vaikuttaa opintojen sujuvuuteen?

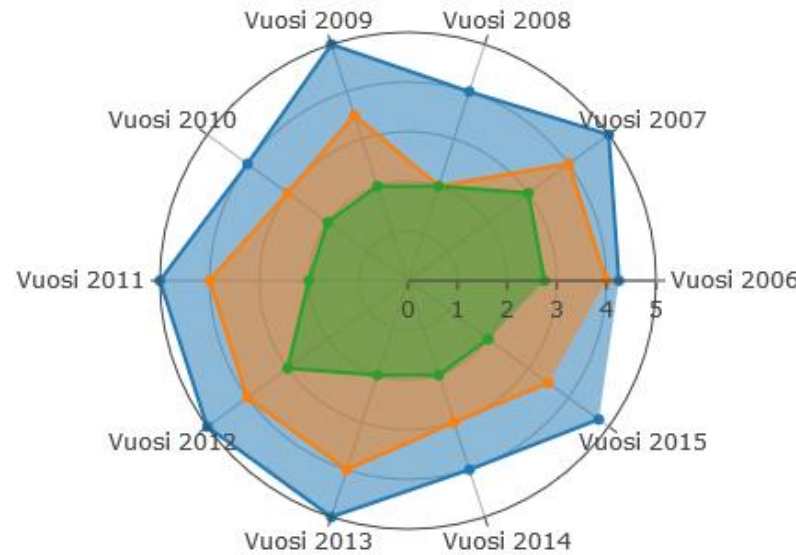


Arvosanat vuosikursseittain, miltä trendi näyttää?

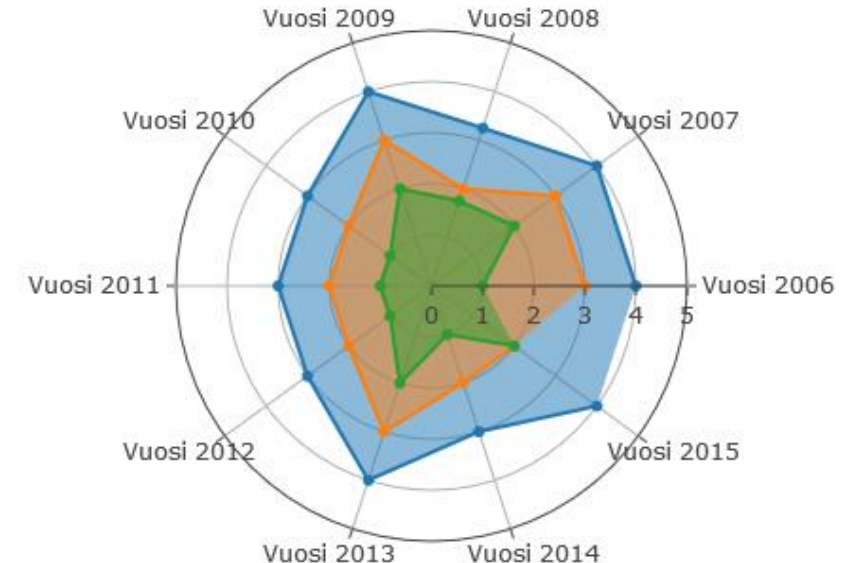
Matematiikan peruskurssi I



Matematiikan peruskurssi II



Matriisialgebra

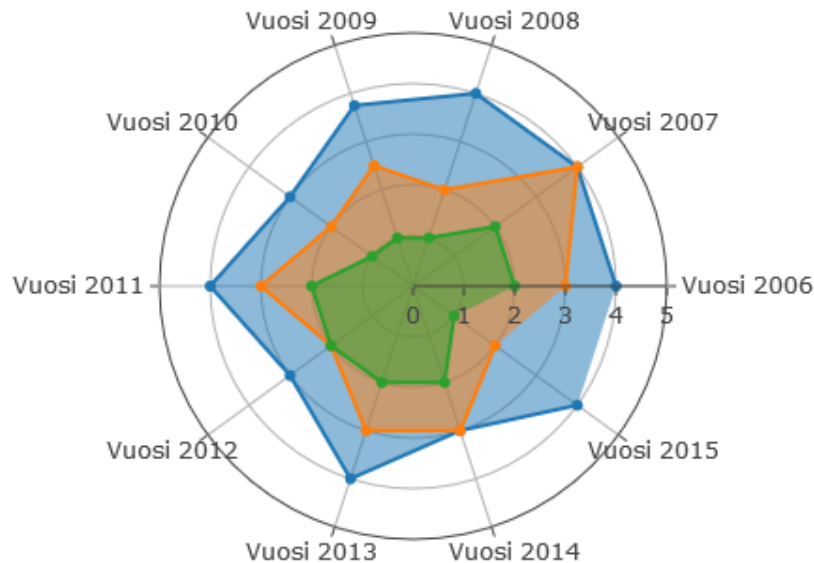


Kurssiarvosanan yläkvartiili (sin), mediaani (rusk) ja alakvartiili (vihr), mukana kaikki opiskelijat, joilla kurssista arvosana.

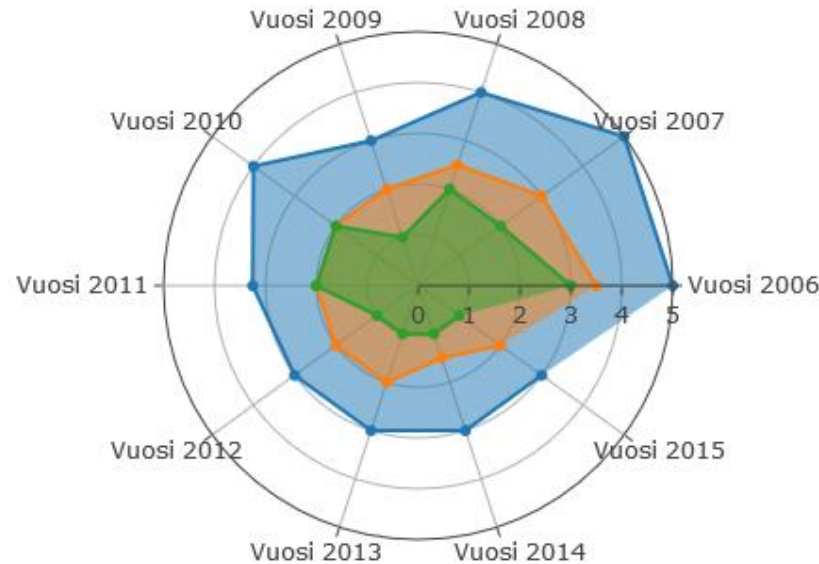


Arvosanat vuosikursseittain, miltä trendi näyttää?

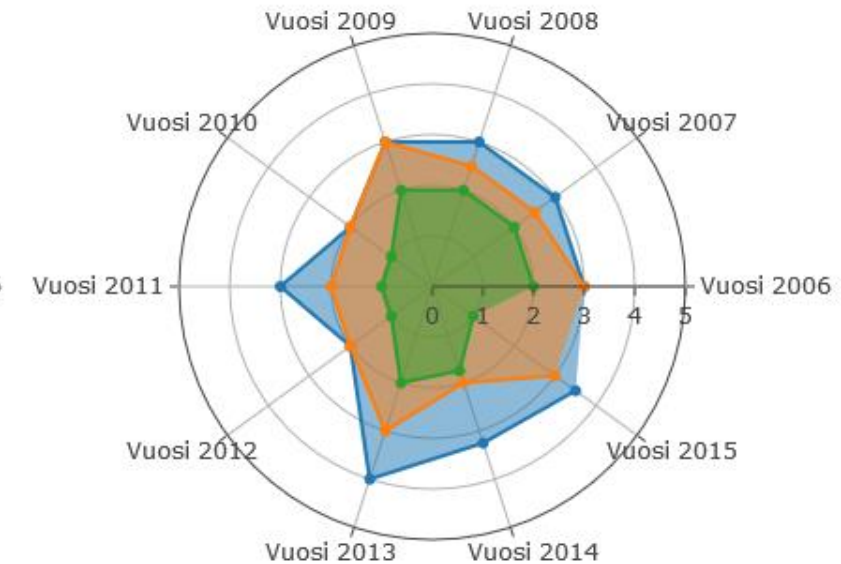
Digitaalitekniikka 1



Tilastomatematiikka



Sähkömittaustekniikan perusteet



Esimerkit pakollisista ensimmäisen vuoden kursseista Oulun yliopiston Tietotekniikan tutkinto-ohjelmasta. HUOM! Pelkkää tilastollista analyysiä.



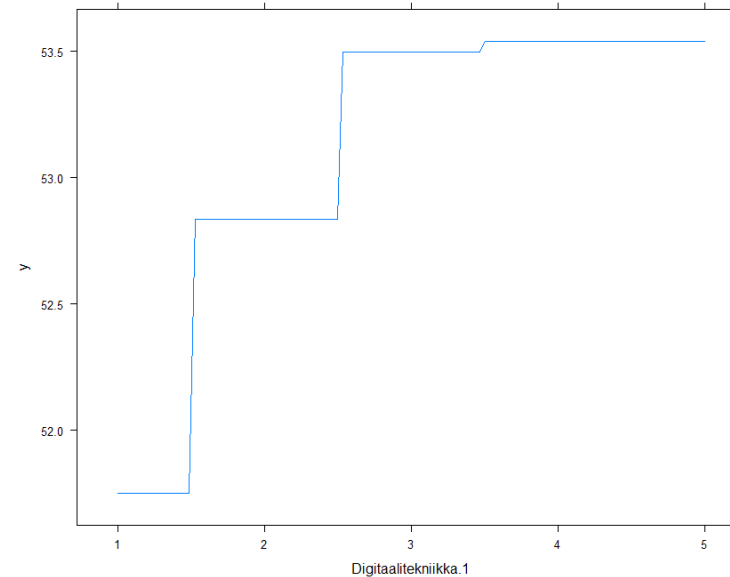
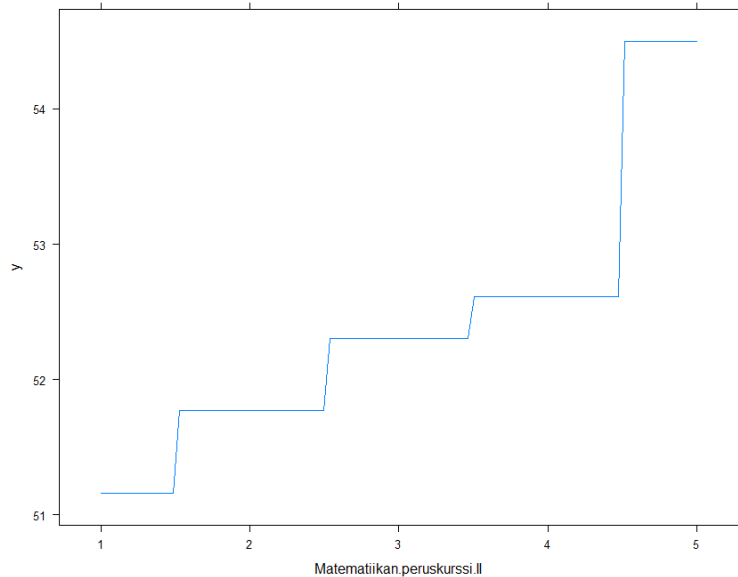
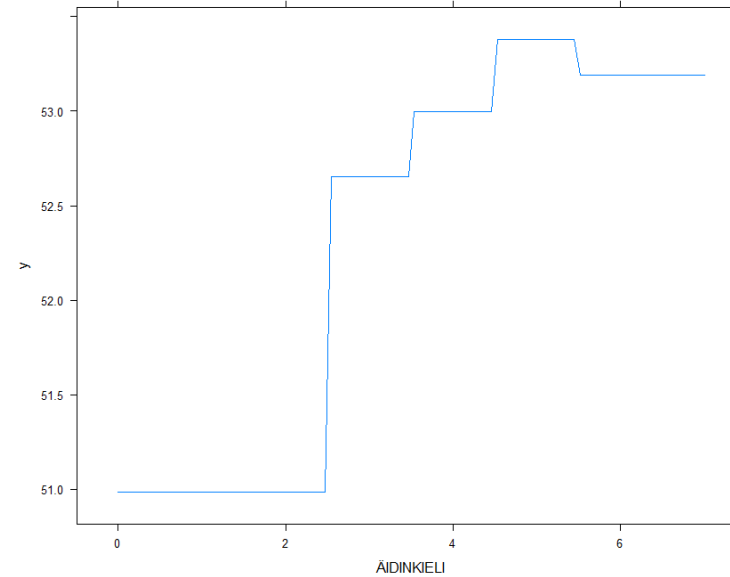


Esimerkkejä analyysistä yksittäiselle tutkinto-ohjelmalle

- Ennustemalli (Gradient Boosting Machine, GBM) **ensimmäisen vuoden opintopistekertymälle** Tietotekniikan tutkinto-ohjelmassa Oulun yliopistossa.
- Syötteinä
 - opintosuoritteet ensimmäisen vuoden syksyltä.
 - Lukioaineiden arvosanat yo-kirjoituksissa.
 - Mallissa yli 300 opiskelijaa.



Matematiikan.peruskurssi.II	36.80071233
Digitaalitekniikka.1	19.83193718
Matriisialgebra	15.00141568
Matematiikan.peruskurssi.I	9.91378818
Ohjelmoinnin.alkeet_suur	4.07443327
aloitusvuosiryhma	3.74200110
ÄIDINKIELI	2.49091409
Digitaalitekniikka.1_suur	1.81787480
ENGLANTI_PITKÄ	1.14392564





Esimerkkejä analyysistä yksittäiselle tutkinto-ohjelmalle

- Ennustemalli (GBM) **yksittäisen kurssin läpäisylle** Tietotekniikan tutkinto-ohjelmassa Oulun yliopistossa.
 - Syötteinä opintojen arvosanat.
 - Mallissa yli 300 opiskelijaa.



```
In [334]: gbm = gbm_fit(X_tite3, df_tite["KOMPAN"] >= 1)
```

```
GBM score: 0.8738095238095238
```

```
Cross val score: [0.87058824 0.79761905 0.85714286 0.75          0.85542169]
```

```
Mean: 0.826154365360602
```

```
Feature importances:
```

```
PK2      0.494566
```

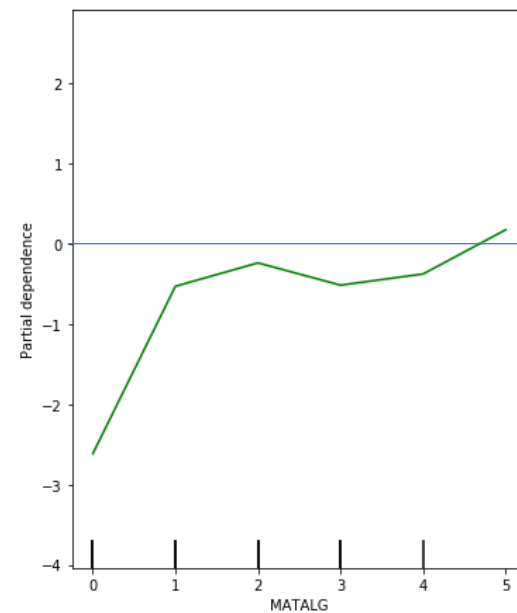
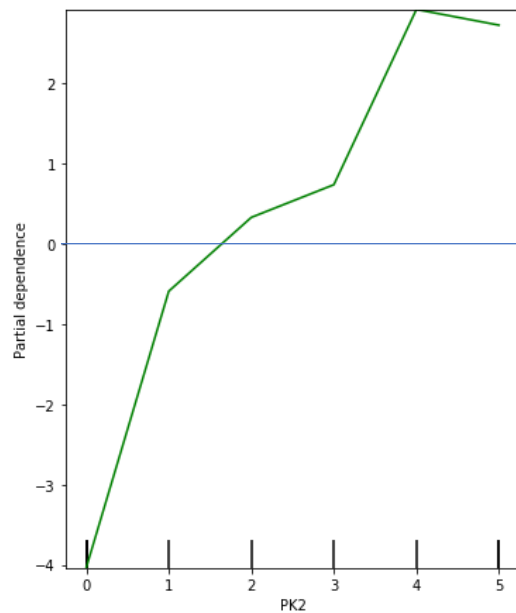
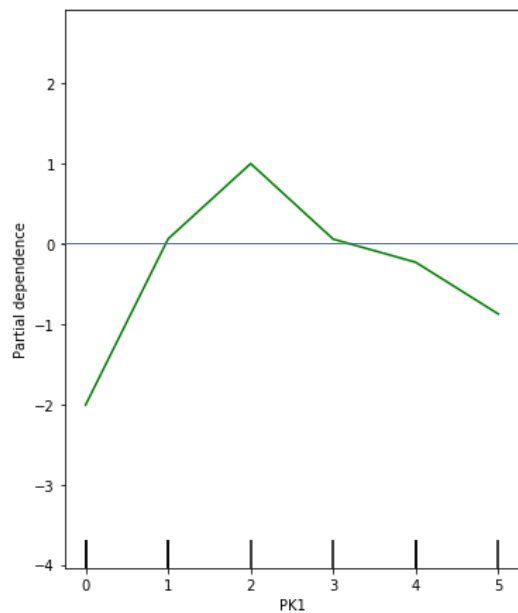
```
PK1      0.324674
```

```
MATALG   0.180760
```

```
dtype: float64
```

Esimerkki 1. vuoden
matematiikan opintojen
vaikutuksesta erääseen 2.
vuosikurssin matematiikan
opintojaksoon.

```
In [336]: fig, ax = plt.subplots(3, 3, figsize=(20, 10))
plot_partial_dependence(gbm, X_tite3, [0, 1, 2], feature_names=X_tite3.columns, ax=ax[0, 0])
plt.show()
```



```
In [348]: gbm = gbm_fit(X_tite2, df_tite["DIGSU"] > 0)
```

```
GBM score: 0.9261904761904762
```

```
Cross val score: [0.81176471 0.8452381 0.71428571 0.77380952 0.84337349]
```

```
Mean: 0.797694306638318
```

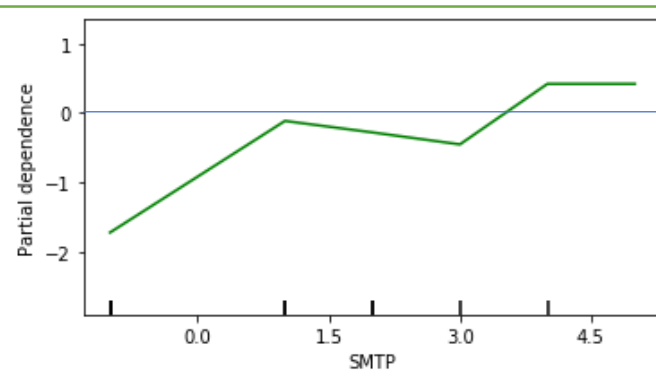
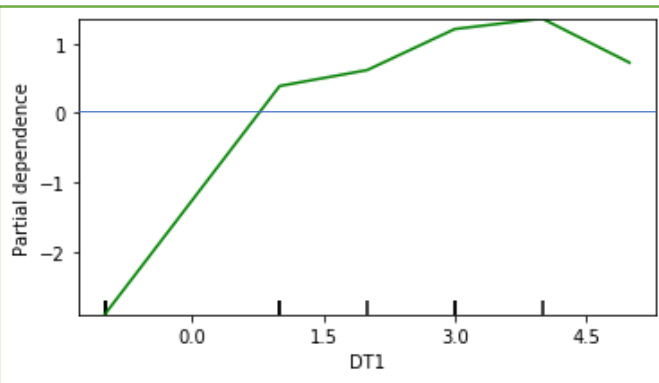
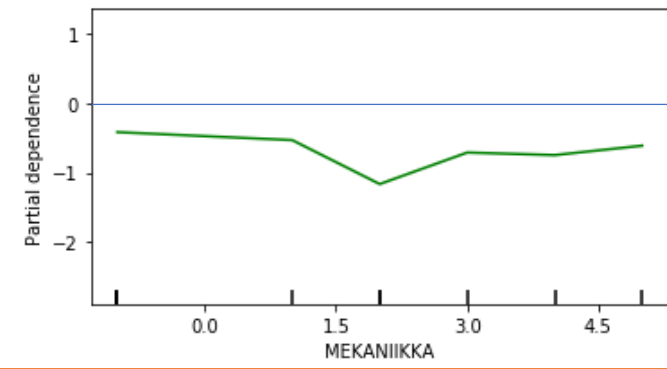
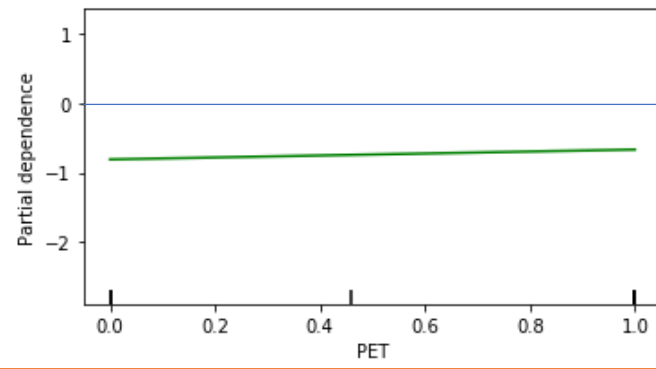
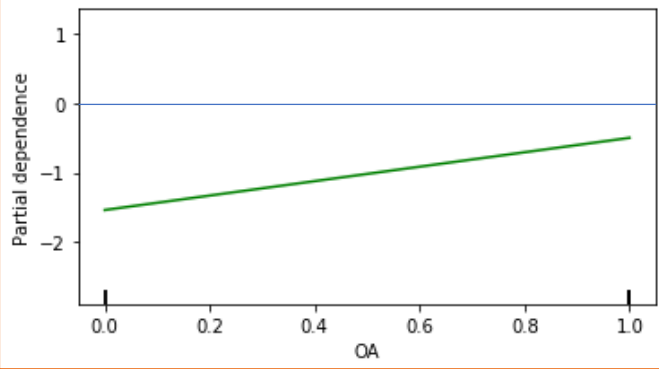
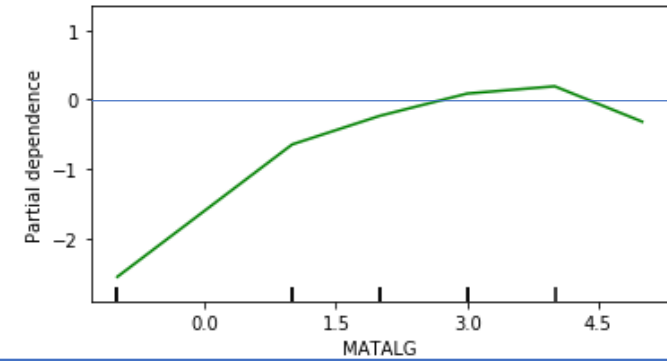
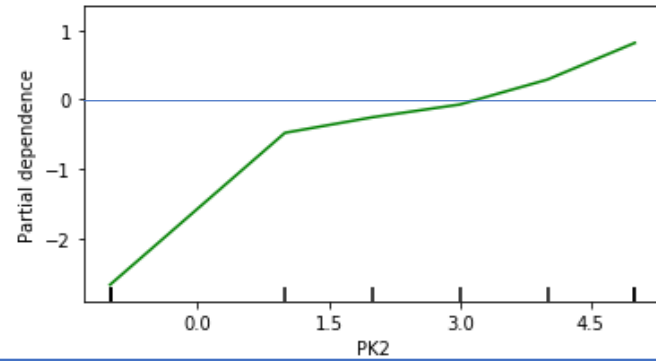
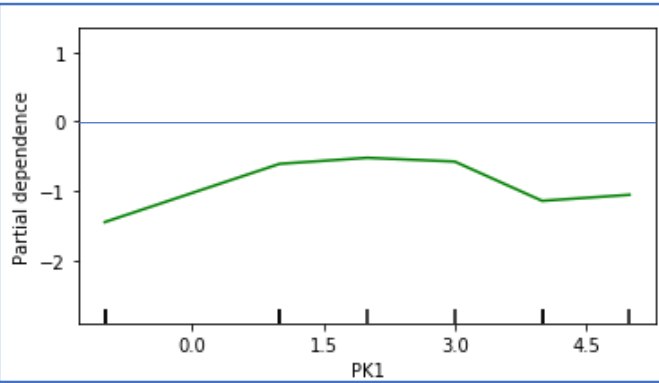
```
Feature importances:
```

```
DT1          0.234385  
MATALG      0.164682  
SMTP        0.163687  
PK2         0.139033  
MEKANIikka  0.126305  
PK1         0.095257  
OA          0.055529  
PET         0.021122
```

```
dtype: float64
```

Esimerkki 1. vuoden
opintojaksojen
vaikutuksesta erääseen 2.
vuosikurssin opintojaksoon.





Tulevaisuuden mahdollisuudet

- Tällä hetkellä keskitetysti dataa saatavilla YO-kirjoituksista sekä opintojaksojen suorittamisesta.
- Tulevaisuudessa toivottavasti mahdollisuus tarkastella, mitä kurssien sisällä tapahtuu.
 - Mikä on opiskelijan osaamistaso opintojakson eri osaamistavoitteiden suhteen vs. yksi arvosana koko kurssista.
 - Voidaan käyttää piirteinä, kun pyritään hahmottamaan paremmin tutkinnon suorittamisaikaa tai yksittäisten kurssien läpäisyä.
- Johtaa siihen, että pystytään entistä paremmin löytämään koulutuksen pullonkaulat ja varmistamaan opintojen sujuva eteneminen.



Yhteenveto

- Älykkäällä analyysillä voidaan saada selville mm.
 - Millaisia opiskelijoita alalle hakeutuu.
 - Kenen kanssa ”kisailaan” samoista opiskelijoista.
 - Mitkä seikat ovat potentiaalisia vaikuttajia ensimmäisen syksyn / ensimmäisen vuoden / kandiopintojen / koko opintopolun keston ja sujuvuuteen.
 - Tekijöitä, joita ihminen ei voi havaita tilastoja silmämääräisesti tarkastelemalla.





Kiitos!

Lisätietoja:

Riku Hietaniemi, riku.hietaniemi@oulu.fi

Satu Tamminen, satu.tamminen@oulu.fi

