



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# Tekoälyn luottamustason optimointi (valmiin työn esittely)

*Satu Salminen*

*31.08.2018*

*Ohjaaja: Jaakko Sävilammi*

*Valvoja: Ahti Salo*

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

# Motivaatio

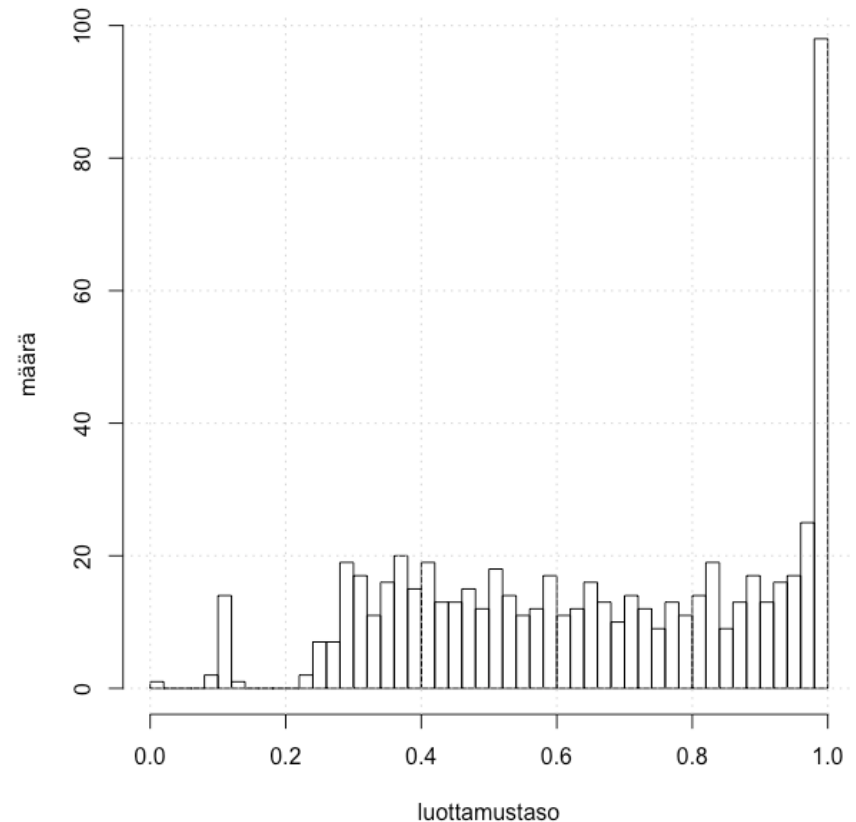
- Osa yrityksen kehitysprosessia, jossa automatisoitiin yrityksen asiakaspalvelua

# Tausta 1/2

- Tekoälyyn perustuva tietokoneohjelma – ”chatbot” siirtymässä toimimaan itsenäisesti yrityksen asiakaspalveluun
- Asiakaspalvelusta on kerätty tuhansia viestejä tekoälyn koulutusta varten

# Tausta 2/2

- Tekoäly palauttaa vastauksen jos luottamustaso ( $c$ ) ylittää asetetun kynnysarvon ( $\theta$ ).



# Tavoitteet

- Määrittää optimaalinen raja-arvo tekoälyn luottamustasolle
  - Oikeita vastauksia annetaan mahdollisimman paljon, ilman, että väorien vastausten määrä kasvaa (TP ja TN maksimointi)
- Asiakaskokemuksen parantaminen
  - Minkälainen vastaus on asiakkaan näkökulmasta väärä?
  - Asiakkaalle palvelua ilman jonotusta

# Aineisto

- Tekoälyn tutkiminen binäärisenä luokittelijana
- Otokset luokiteltiin positiivisiksi ja negatiivisiksi

	Vastaus oikein	Vastaus väärin
$c > \theta$	True positive	False positive
$c < \theta$	False negative	True negative

# Menetelmät 1/3

- Sensitiivisyys (Se)
- Spesifisyys (Sp)
- Täsmällisyys (Pr)

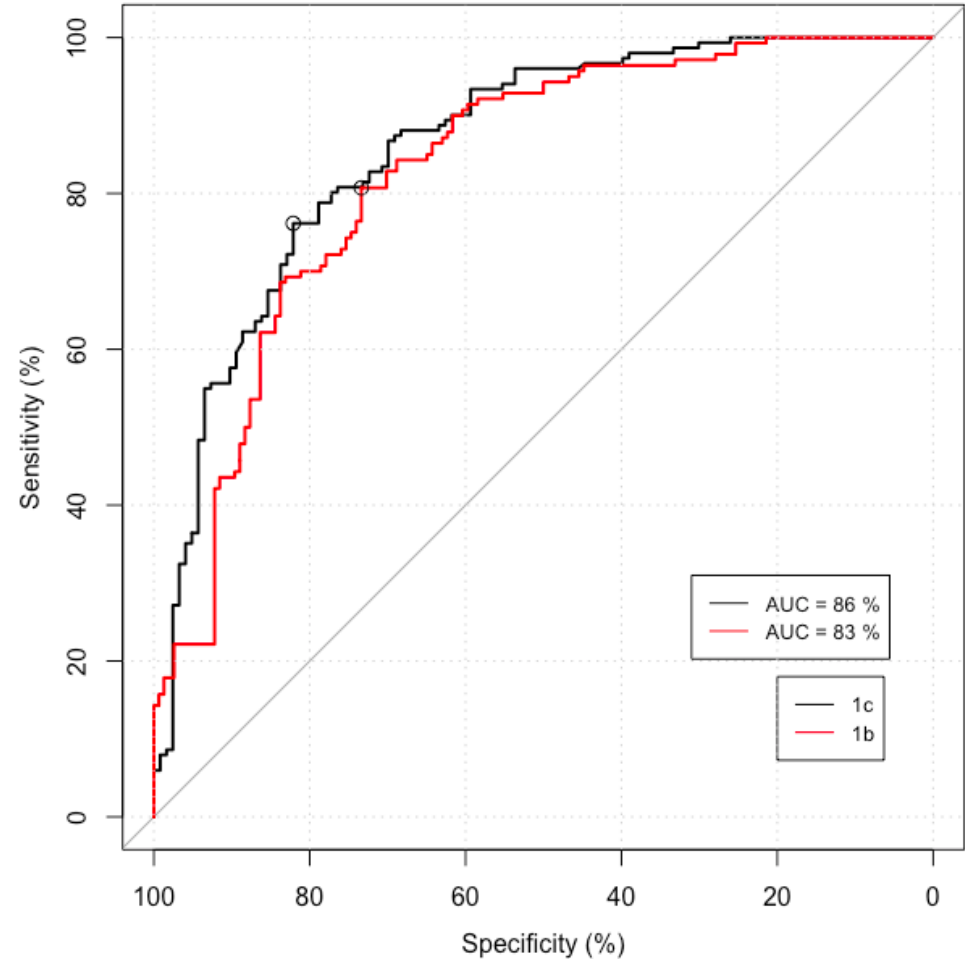
$$Se = \frac{\sum TP}{\sum TP + \sum FN}$$

$$Sp = \frac{\sum TN}{\sum TN + \sum FP}$$

$$Pr = \frac{\sum TP}{\sum TP + \sum FP}$$

# Menetelmät 2/3

- Automaattisesti ja manuaalisesti luokitellun aineiston vertailu
- ROC-käyrät
- AUC



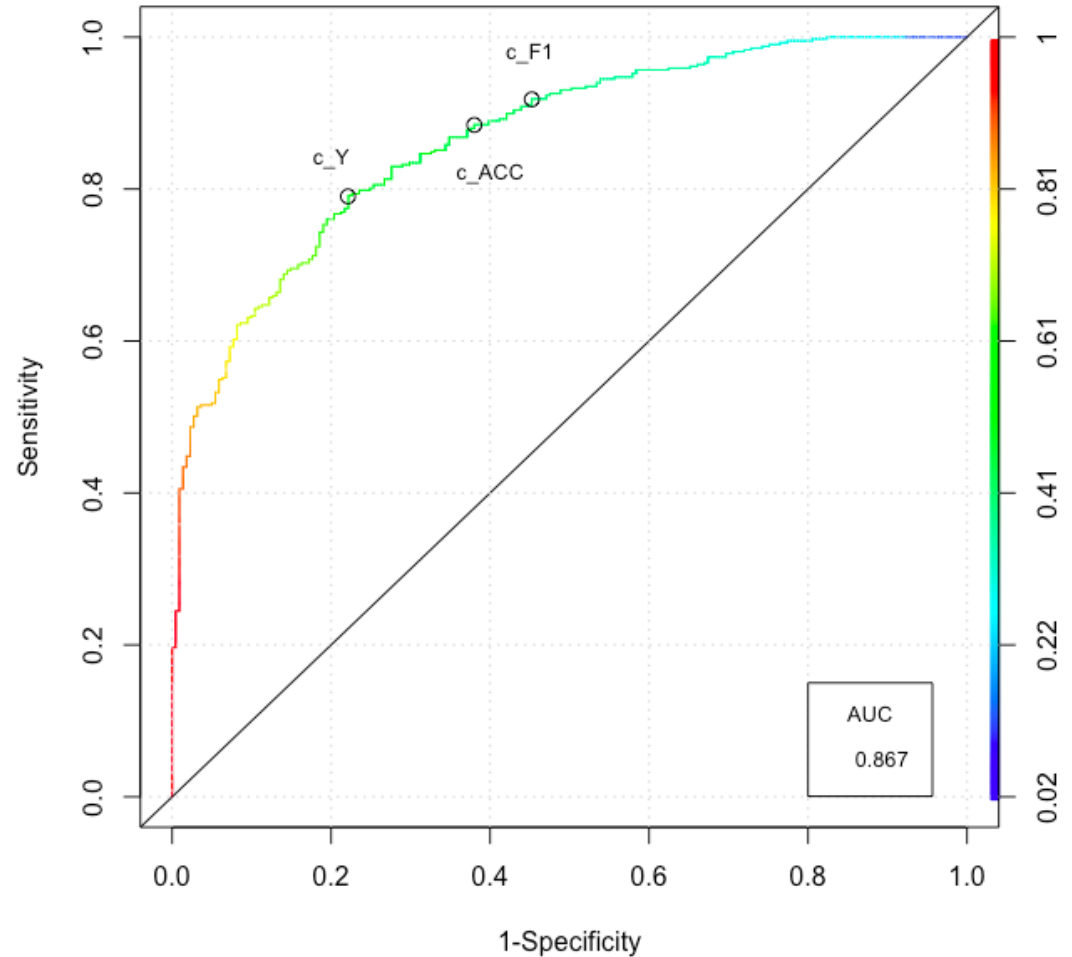


# Menetelmät 3/3

$$Youden = Se + Sp - 1$$

$$ACC = \frac{\sum TP + \sum TN}{\sum TotalPopulation}$$

$$F_{\beta} = (1 + \beta^2) \frac{Se \cdot Pr}{Se + \beta^2 \cdot Pr}$$



# Tulokset 1/4

- Tuloksia kerättiin ROC –analyysistä ja automaatiotestauksesta

Optimoinnin kriteeri	Sensitiivisyys	Spesifisyys	Täsmällisyys
Ei optimointia	93.5 %	47.5 %	77.0 %
F1	91.8 %	54.8 %	79.2 %
ACC	88.4 %	62.0 %	81.4 %
Youden	79.0 %	77.8 %	87.0 %

# Tulokset 2/4

Optimoinnin kriteeri	Distance	Youden	ACC	F1
Ei optimointia	0.28	41.0 %	77.5 %	84.4 %
F1	0.22	46.1 %	78.8 %	84.9 %
ACC	0.16	50.4 %	79.2 %	84.8 %
Youden	0.10	56.0 %	78.5 %	82.8 %

# Tulokset 3/4

- Prosentuaaliset muutokset ROC-analyysin ja automaatiotestauksen palauttamissa luokitteluissa kun luottamustasoa muutettiin alkuperäisestä

## ROC

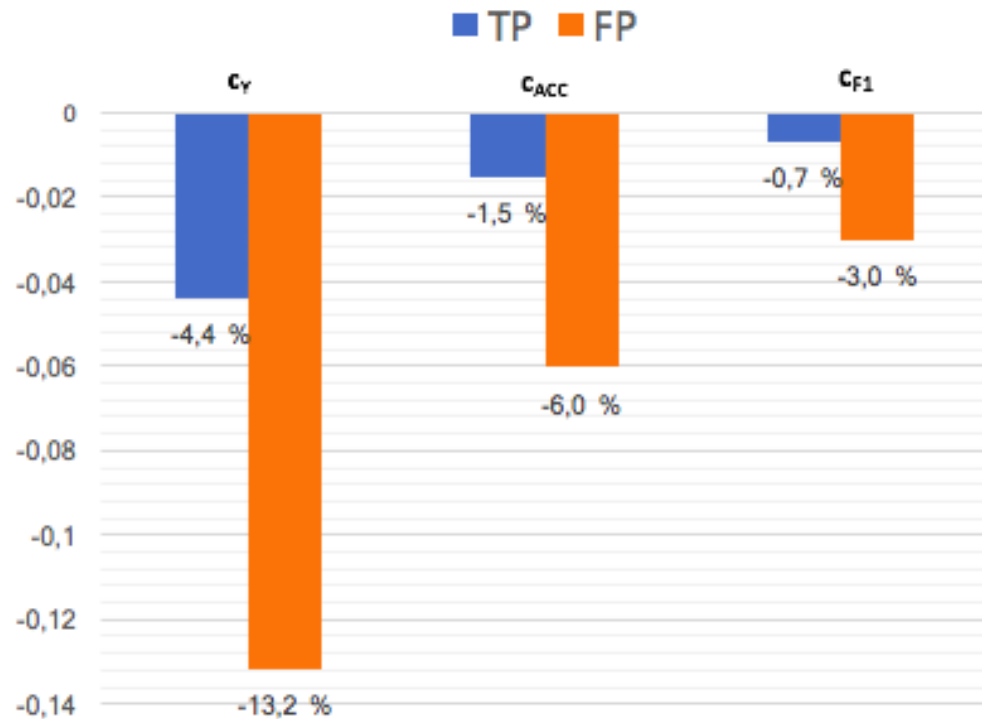
Kriteeri	TP	FP
F1	-1.8 %	-12.9 %
ACC	-5.4 %	-27.6 %
Youden	-14.9 %	-55.2 %

## Automaatiotesti

Kriteeri	TP	FP
F1	-1.3 %	-10.6 %
ACC	-2.8 %	-21.2 %
Youden	-8.0 %	-46.2 %

# Tulokset 4/4

- Muutokset prosenttiyksiköissä automaatiotestauksen palauttamissa tuloksissa
- ”Optimoinnilla saavutettiin merkittävää liiketoiminta-hyötyä.”



# Yhteenveto

- Optimoinnilla onnistuttiin tehostamaan tekoälyn suoriutumista
- Aineiston rakenne ja tulkinnan haastaavuus vaikutti tuloksiin ja rajasi tulosten yleispätevyyttä
- Tutkimuksen jatkokehityksen kannalta on useampi näkökulma, josta lähestyä ongelmaa.
  - Suoriutumisen metriikka
  - Päätöksentekomalli
  - Vastausavainten erittely

# Lähteet

- Tom Fawcett. An Introduction to ROC-analysis, *Pattern recognition letters*, 2006.
- Marina Sokolova, Nathalia Japkowicz & Stan Szpakowicz. Beyond accuracy, f1-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, sivut: 1015-1021, Springer 2006.
- Jennifer Hill, W Randolph For, and Ingrid G Farreras. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behaviour*, 2015.