

Kela|Fpa 

Työkyvyttömyyseläkeriskin ennustaminen

18.11.2024

IT-innovaatio- ja kasvuyksikkö

Taustaa

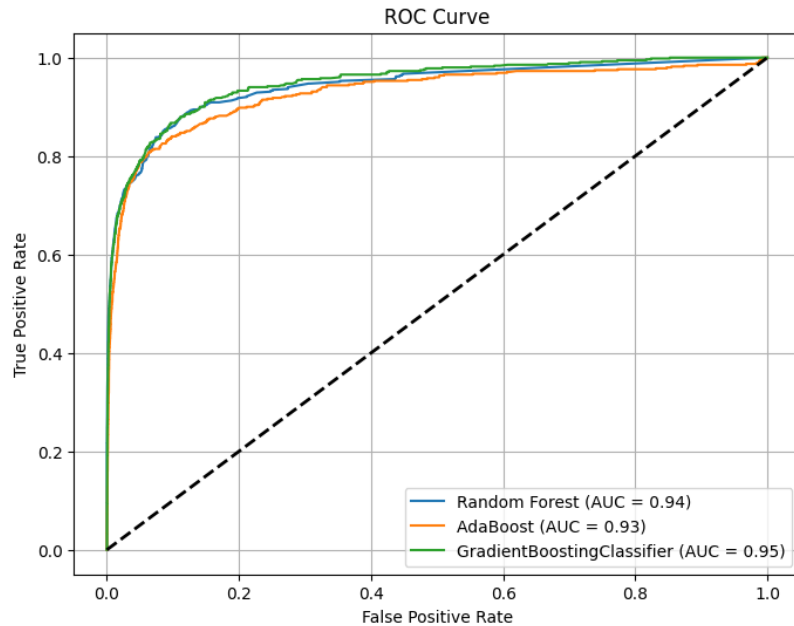
- Työkyvyttömyyden ennustemallia on kokeiltu aiemmin ETK:ssa 2018 sekä Kelassa vuonna 2021 ja 2023. Kelassa ensimmäisinä tätä kokeilivat Sauli Jäppinen ja Sanna Hemminki 2021. IT-innovaatioyksikkö toisti saman kokeilun 2023.
- Tämä kokeilu rakentaa siis edellisten kokeiluiden havaintojen pohjalle.

Vastemuuttajien selitteet

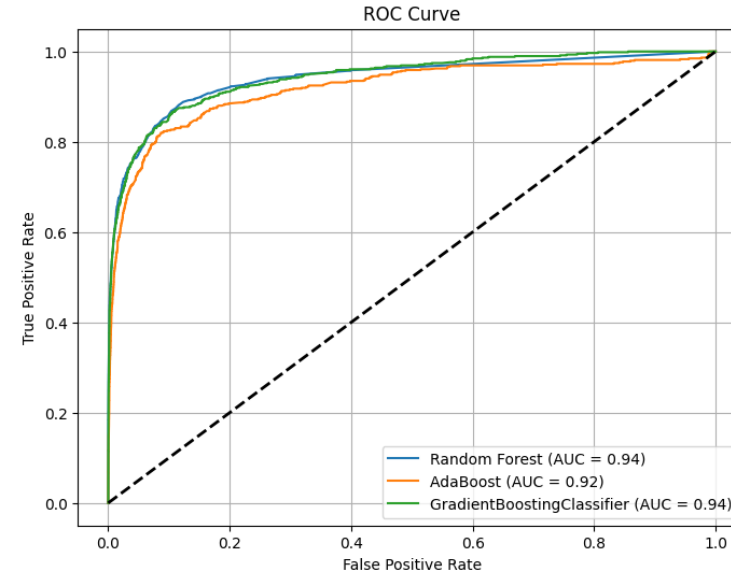
- Kokeilussa testattiin kahta vastemuuttujaa:
 - TK-eläkkeelle (tai mm. kuntoutustuelle) päätyminen ETK:n datan perusteella
 - TK-eläkkeelle tai vastaavalle työkyvyttömyysetuudelle päätyminen tulorekisterin datan perusteella
 - Jos henkilö sai vähintään 300€/kk ja vähintään 4 kuukauden ajan vuodessa, niin vastemuuttujan arvo oli 1, muutoin 0.
 - Huomioidut työkyvyttömyysetuudet olivat kuntoutustuki, työkyvyttömyyseläke (lakisääteinen työeläkevakuutus, osatyökyvyttömyyseläke (lakisääteinen työeläkevakuutus), Kuntoutustuki (lakisääteinen työeläkevakuutus), Työkyvyttömyyseläke (vapaaehtoinen työeläkevakuutus, Osatyökyvyttömyyseläke (vapaaehtoinen työeläkevakuutus)

ROC-käyrät samassa

ETK:n datasta saatu-vaste:



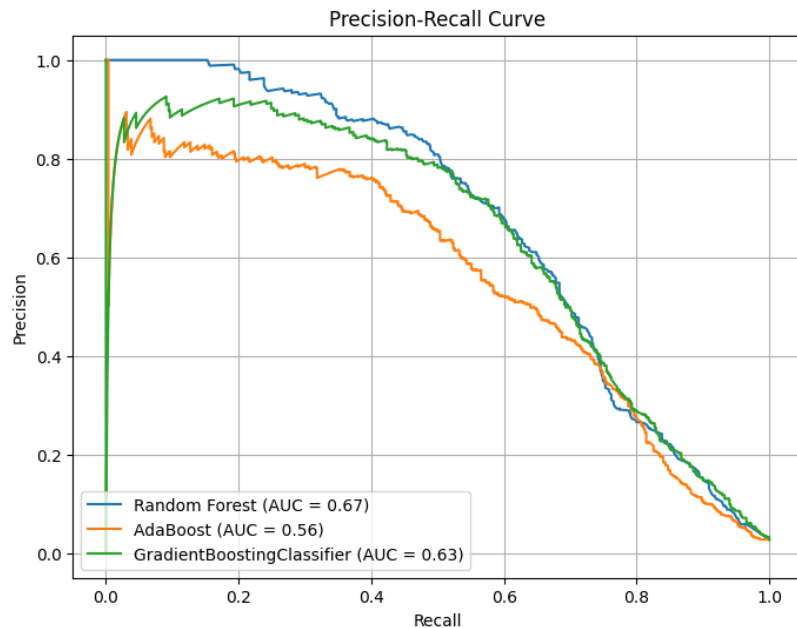
Tulorekisteri-vaste:



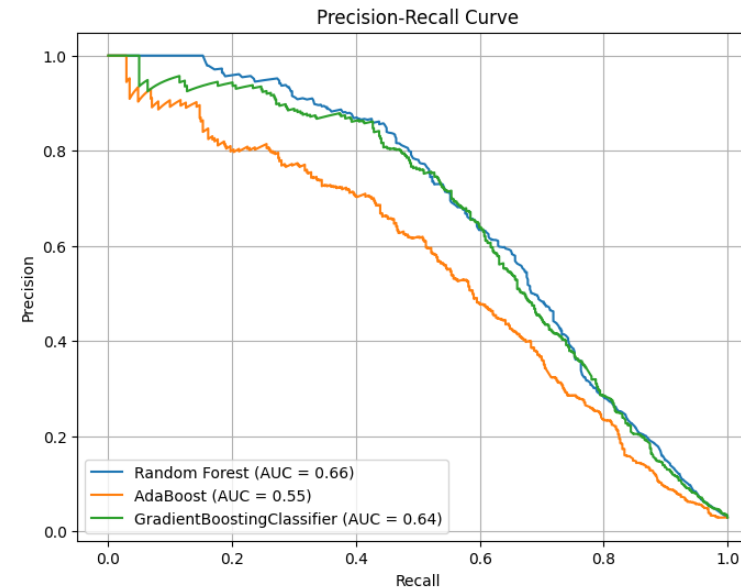
- Gradient boosting –luokittelija oli testatuista lievästi paras ROC-käyrän osalta (area under curve = 0.94).
- Tässä kokeilussa kuitenkin ennustamme harvinaista tapahtumaa, jolloin oleellisempi mittari on precision-recall –käyrä.

Precision-recall –käyrät

ETK:n datasta saatu-vaste:



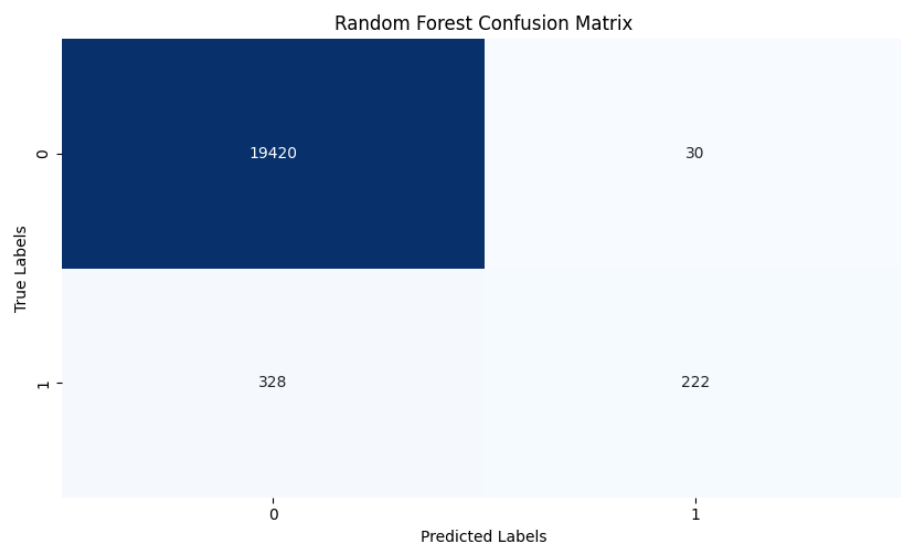
Tulorekisteri-vaste:



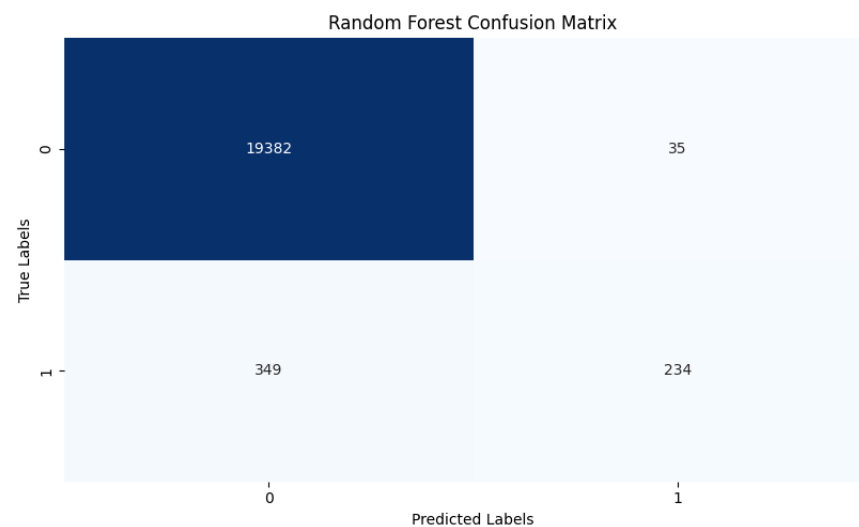
- Precision-recall –käyrien mukaan parhaaseen tulokseen päästään perinteisellä random forest – luokittelijalla (AUC=0.67 ja AUC=0.66).
- Precision-recall –käyrä on järkevä mittari, koska pyrimme ennustamaan harvinaista tapahtumaa.
- Esim. voimme löytää 40% etsittävistä asiakkaista siten, että virhepositiivisia on n. 12%.

Confusion-matriisi, random forest

ETK:n datasta saatu-vaste:



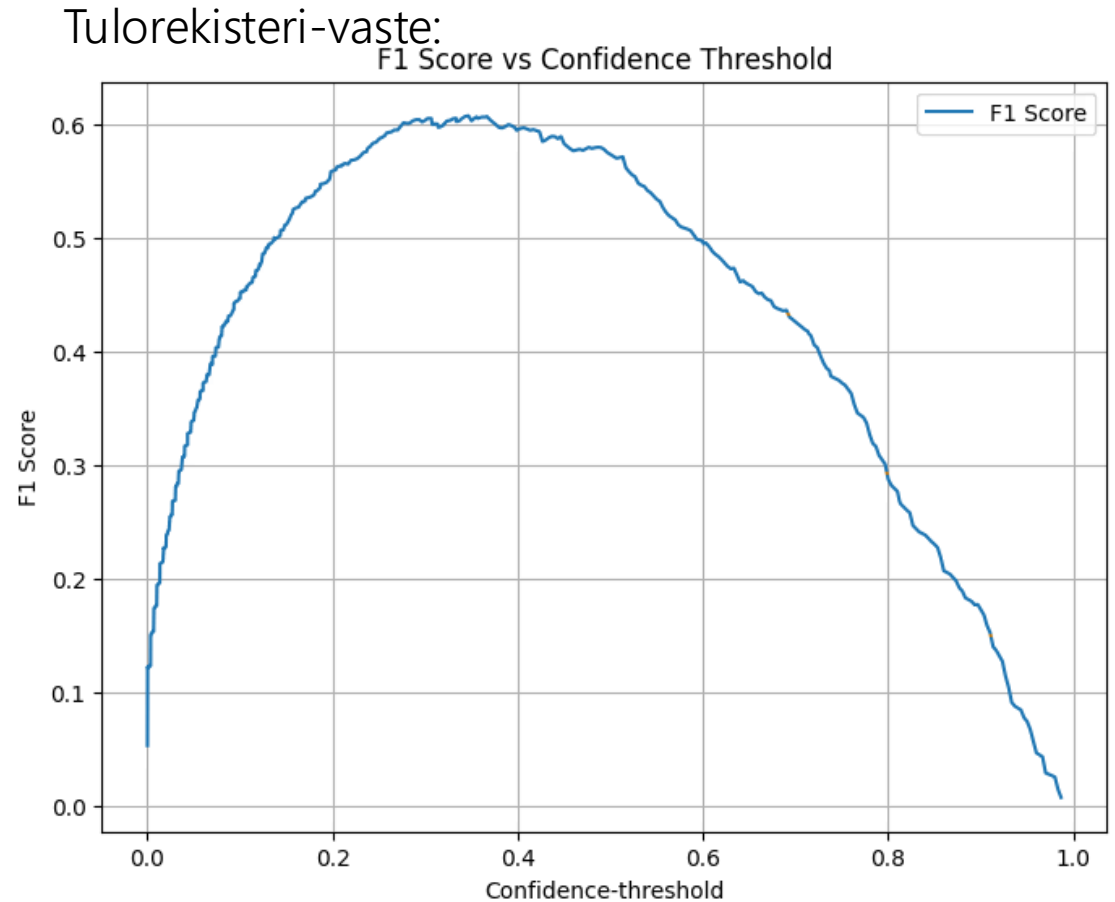
Tulorekisteri-vaste:



- 60% confidence-thresholdia käyttäen saadaan yo. confusion-matriisit.
- Esim. ETK-vastemuuttujan tapauksessa virhepositiivisia syntyy 30 kpl. Malli ennustaa työkyvyttömyyttä testiaineistossa $30+222=252$ henkilölle. Näistä $30/252 \approx 12\%$ on virhepositiivisia. Samaan aikaan löydämme aineiston oikeista työkyvyttömäksi jääneistä $222/(222+328) \approx 40\%$.

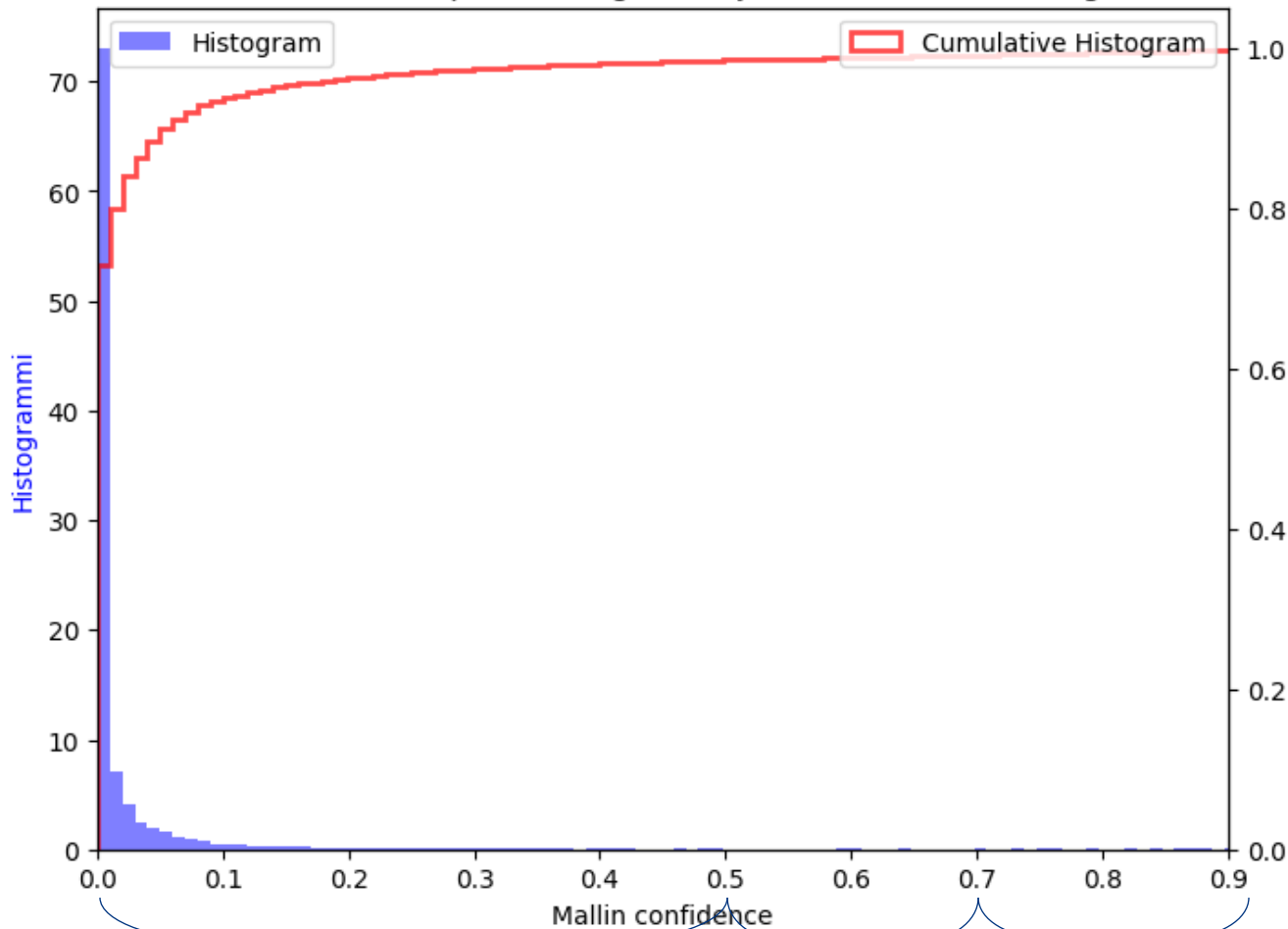
F1-score confidence-thresholdin funktiona

- F1-score on huipussaan, jos confidence-threshold asetetaan suunnilleen arvoon 0.35.
 - Tämä ei kuitenkaan välttämättä ole tässä sovelluskohteessa järkevää.

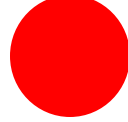
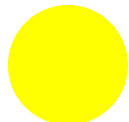


Mallin confidence-score ja "liikennevalot" ?

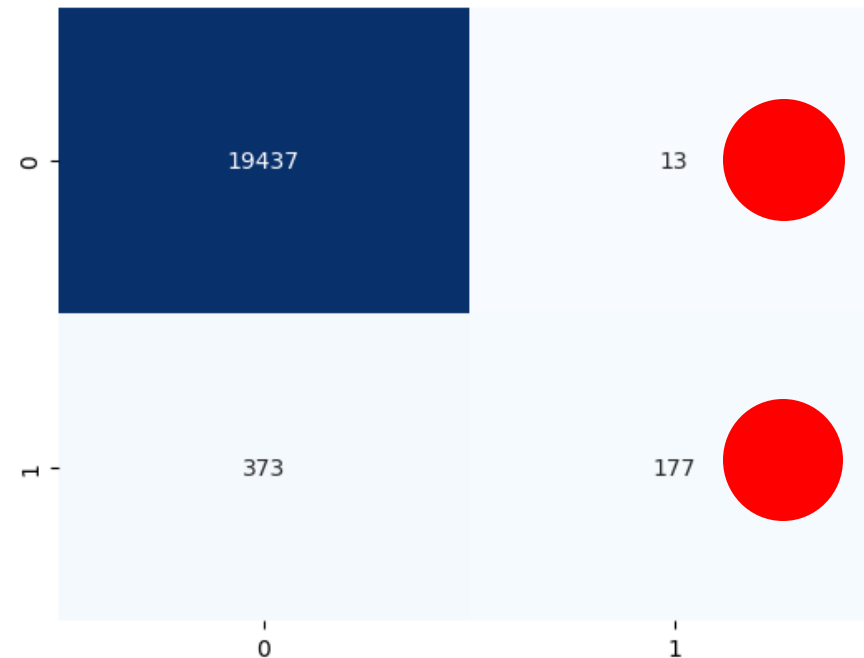
Mallin confidence-outputin histogrammi ja kumulatiivinen histogrammi



8



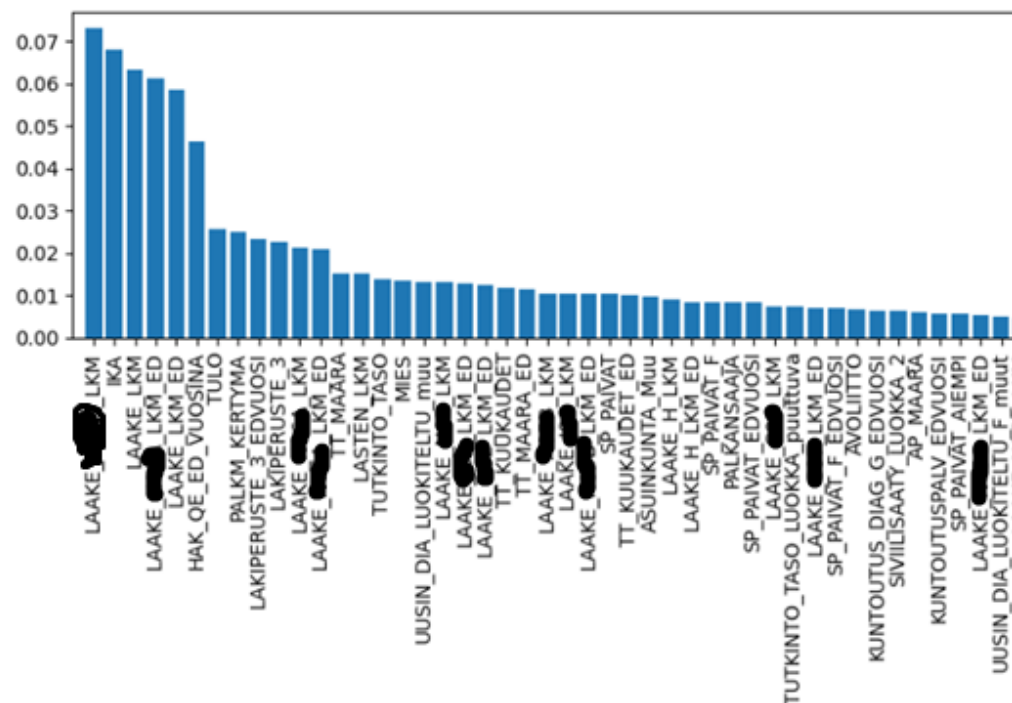
Confusion-matriisi, kun confidence-threshold = 0.7



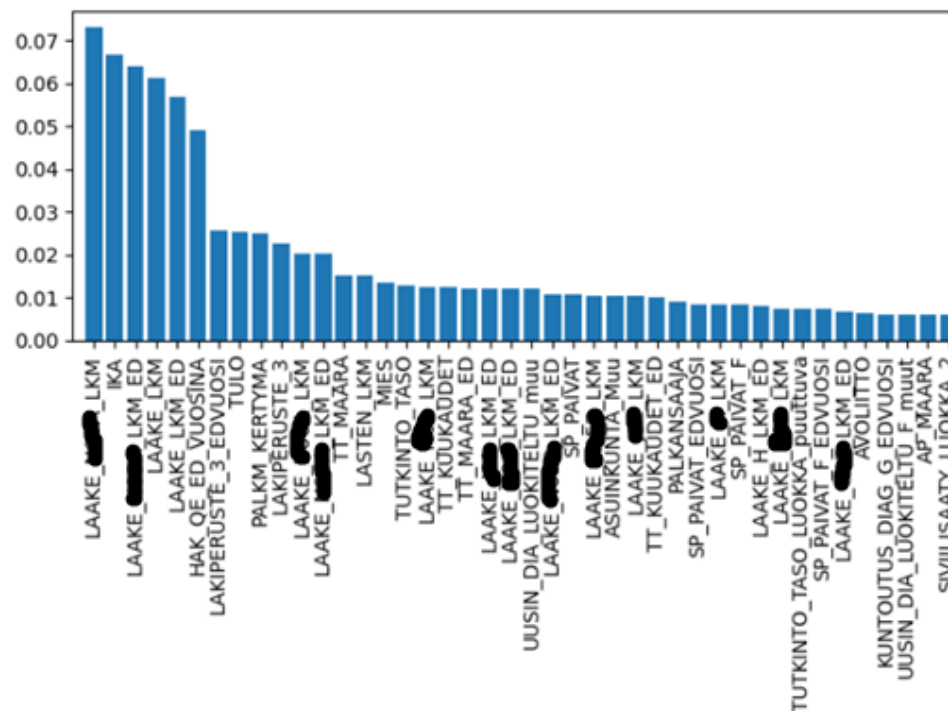
- Malli antaa binäärisen ennusteen lisäksi aina n.s. confidence-scoren, joka on reaaliluku väliltä $[0,1]$. Tämä voidaan tulkita todennäköisyydeksi.
- Confidence-scoren perusteella voisi olla mahdollista luoda järkevä "liikennevalosysteemi".

Top –muuttujat (random-forest -luokittelija)

ETK:n datasta saatu-vaste:



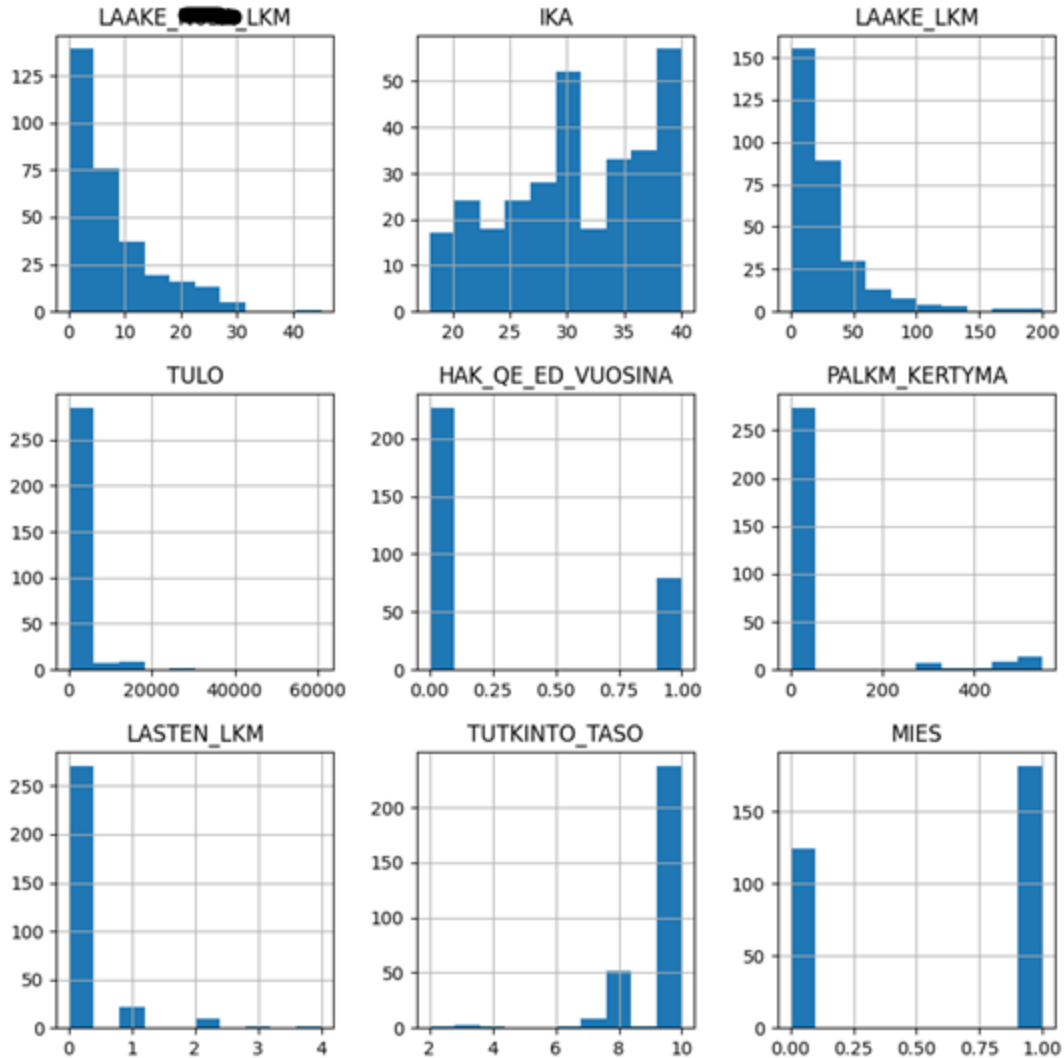
Tulorekisteri-vaste:



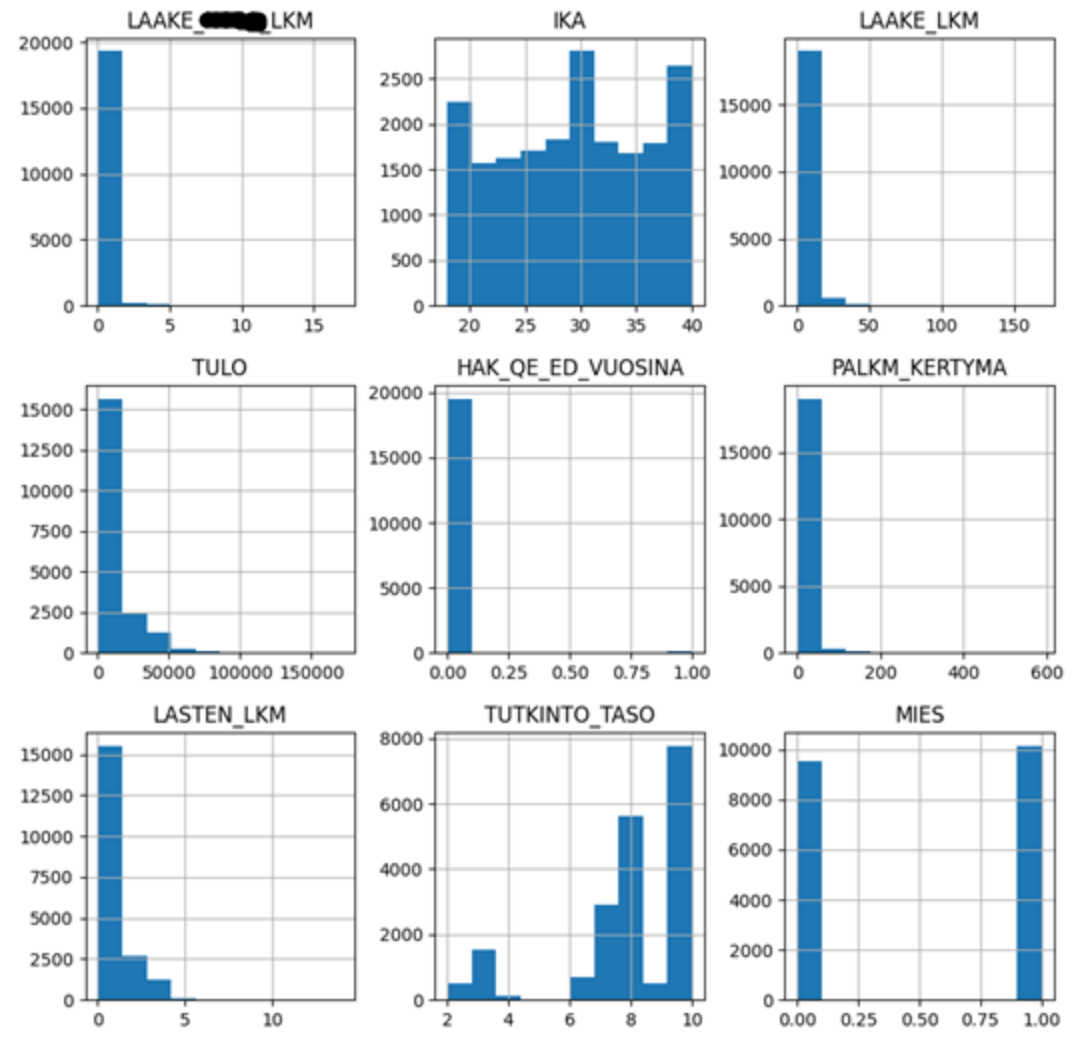
- Yllä on listattu koulutetun random forest –mallin oleellisimmat muuttujat (feature importance).

Top-muuttujien histogrammeja

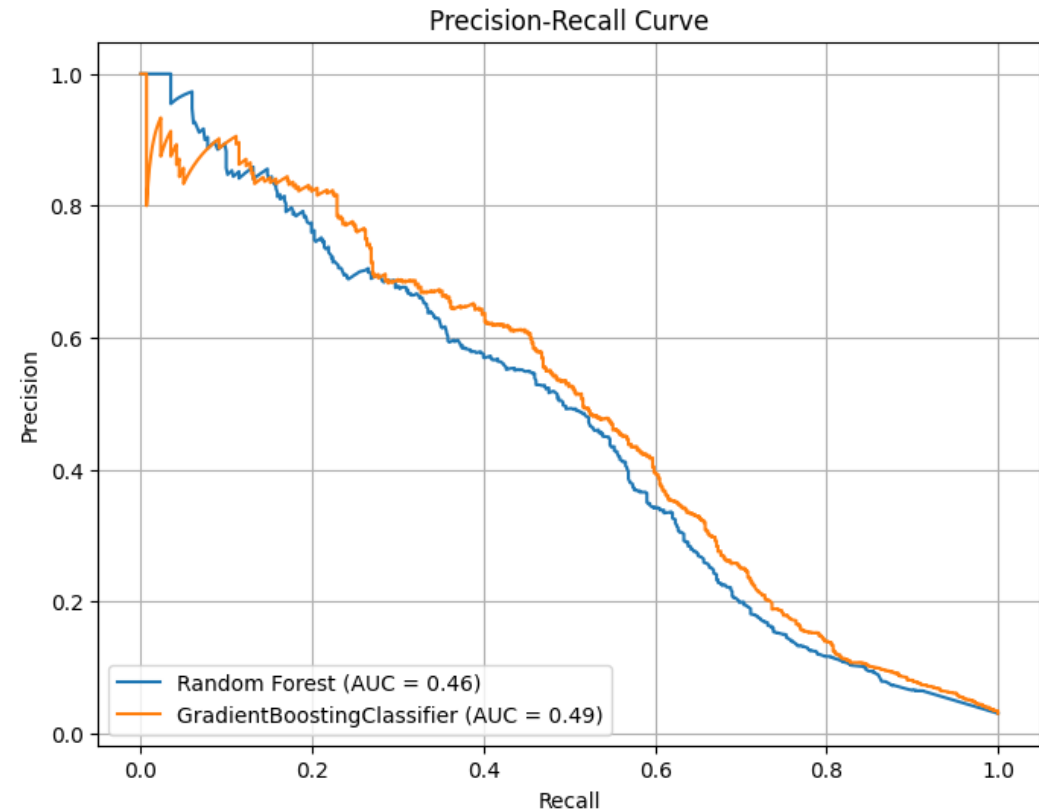
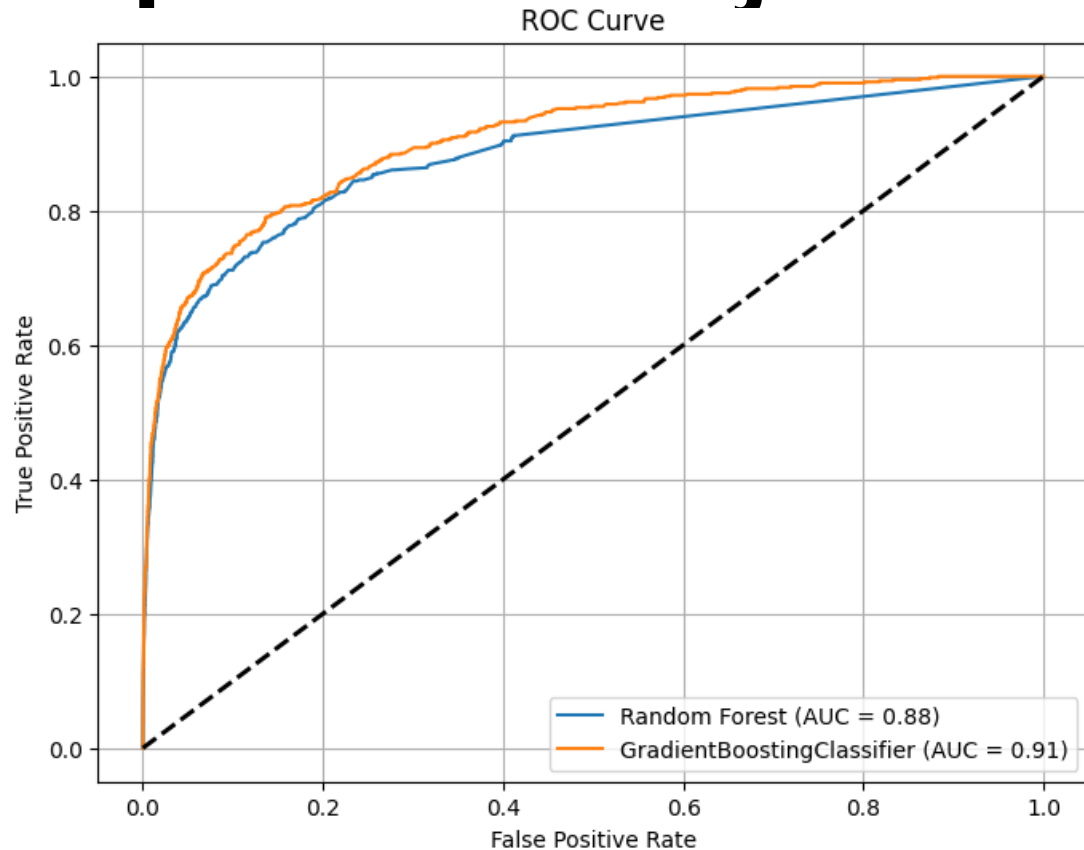
Kun malli ennustaa työkyvyttömyyttä



Kun malli ei ennusta työkyvyttömyyttä



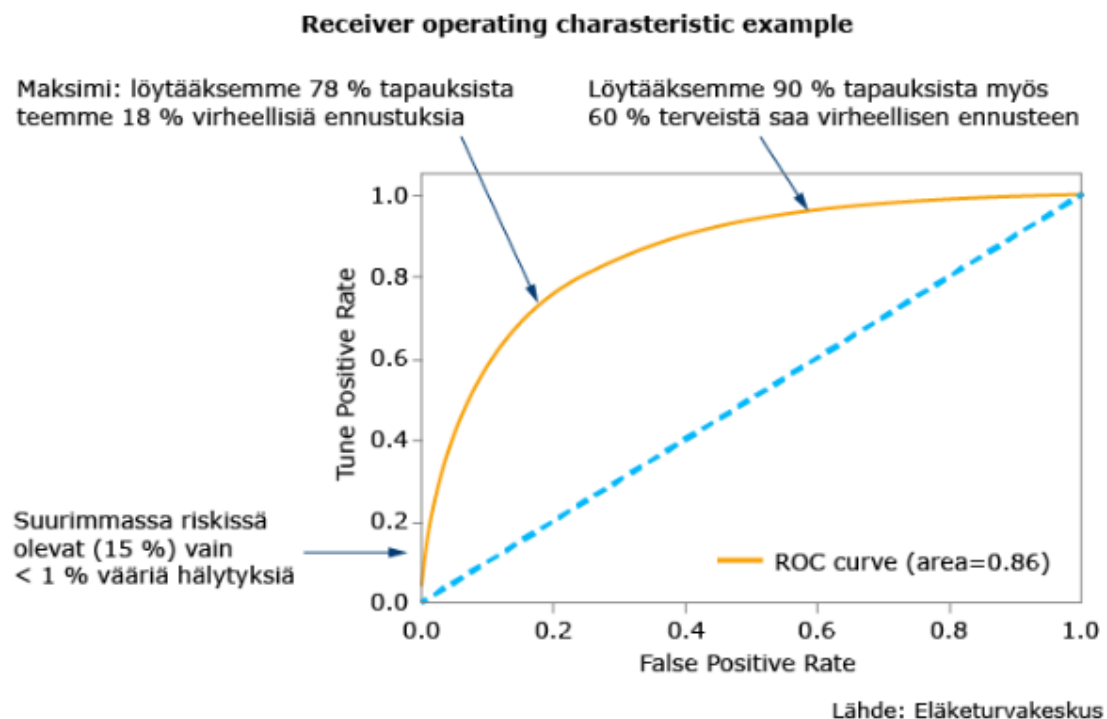
Top 10 -muuttujilla tehty malli



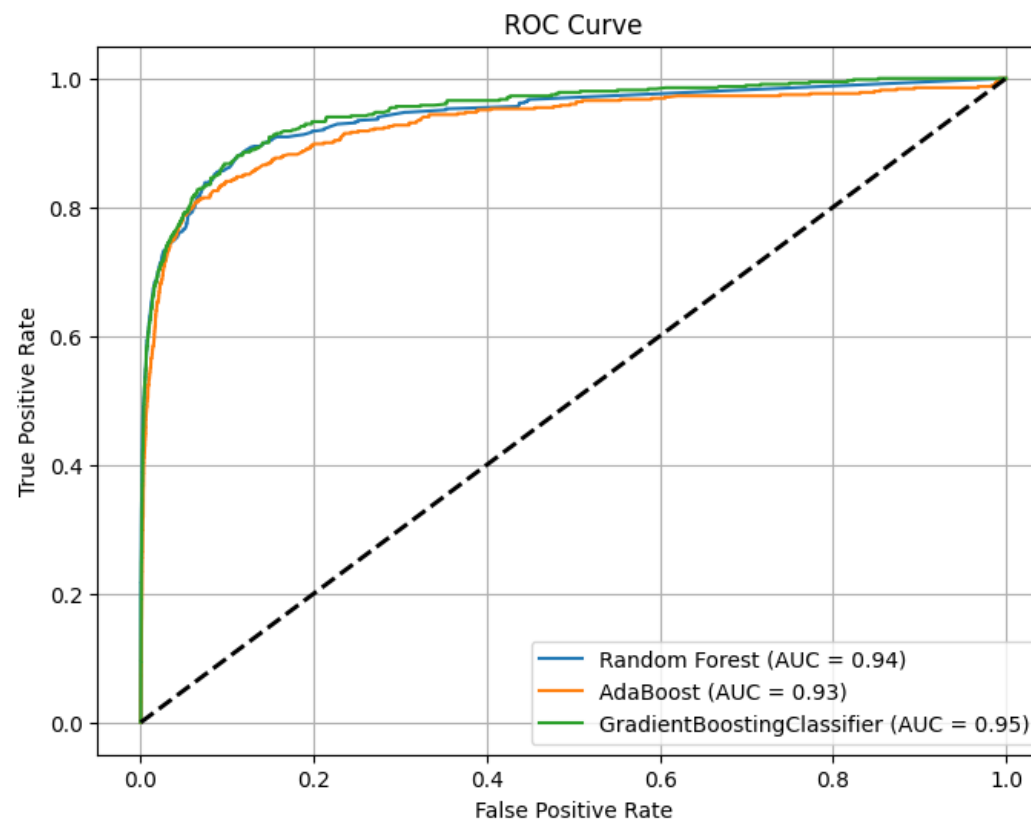
- Testasimme kouluttaa pienemmän mallin käyttäen ison mallin merkitsevimpiä muuttujia.
- Lopputulos ei yllä ison mallin ennustetarkkuuden tasolle.

Vertailu edellisiin kokeiluihin (1/3)

ETK:n ROC-käyrä vuoden 2018 kokeilussa:

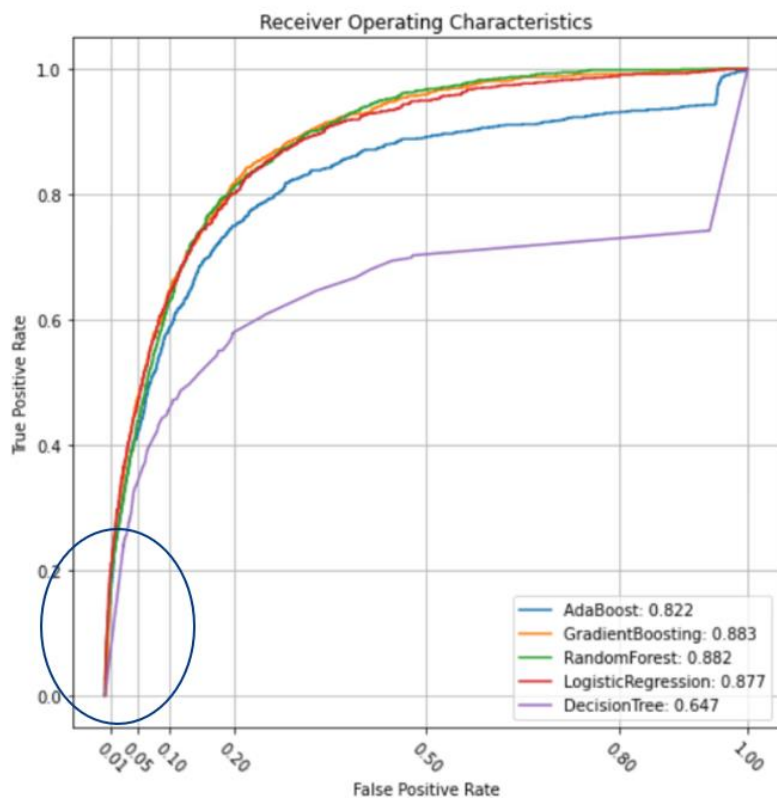


ROC-käyrät tässä kokeilussa:

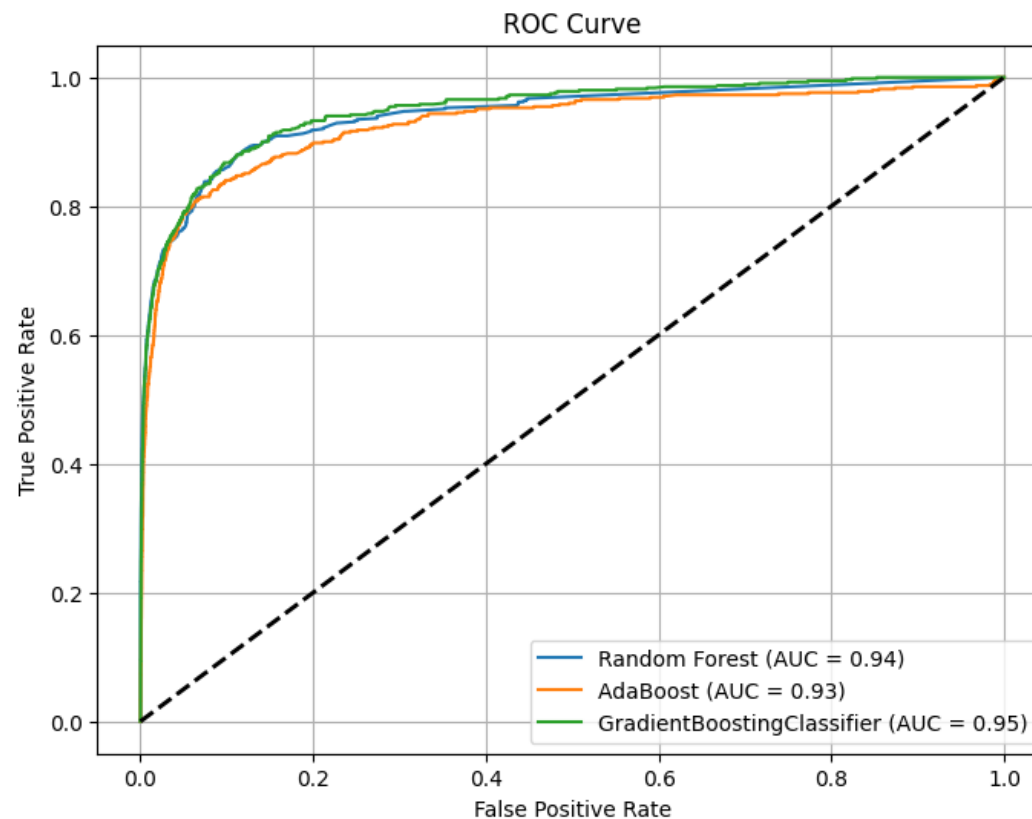


Vertailu edellisiin kokeiluihin (2/3)

ROC-käyrät Kelan vuoden 2023 kokeilussa:

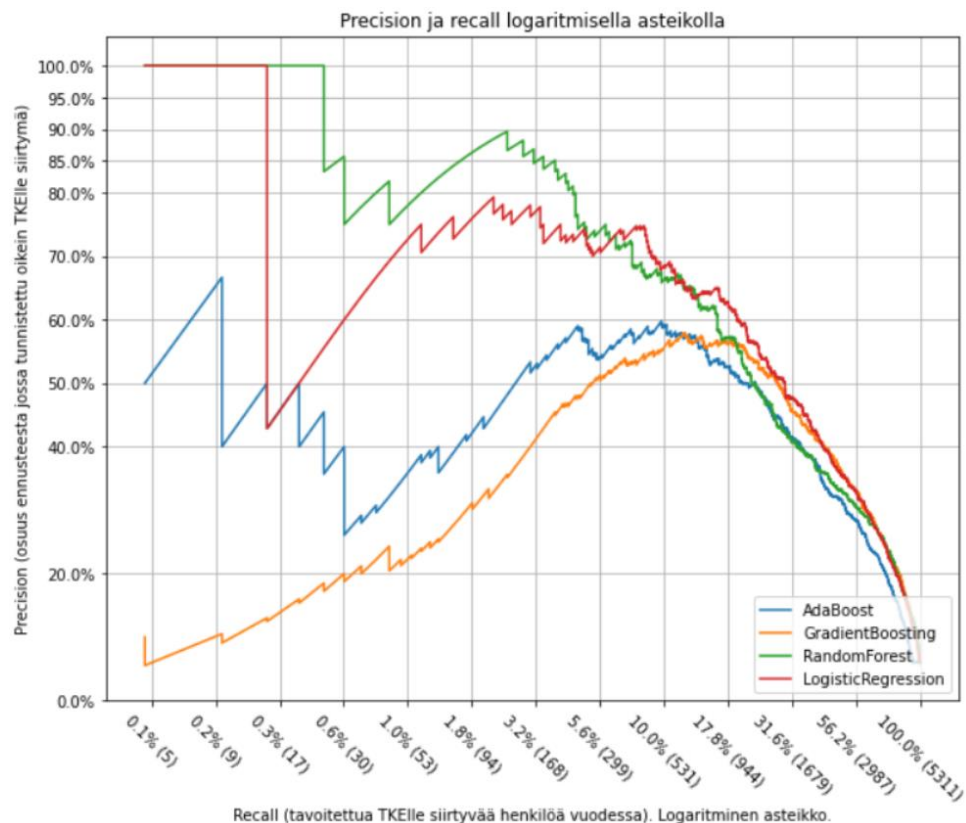


ROC-käyrät tässä kokeilussa:

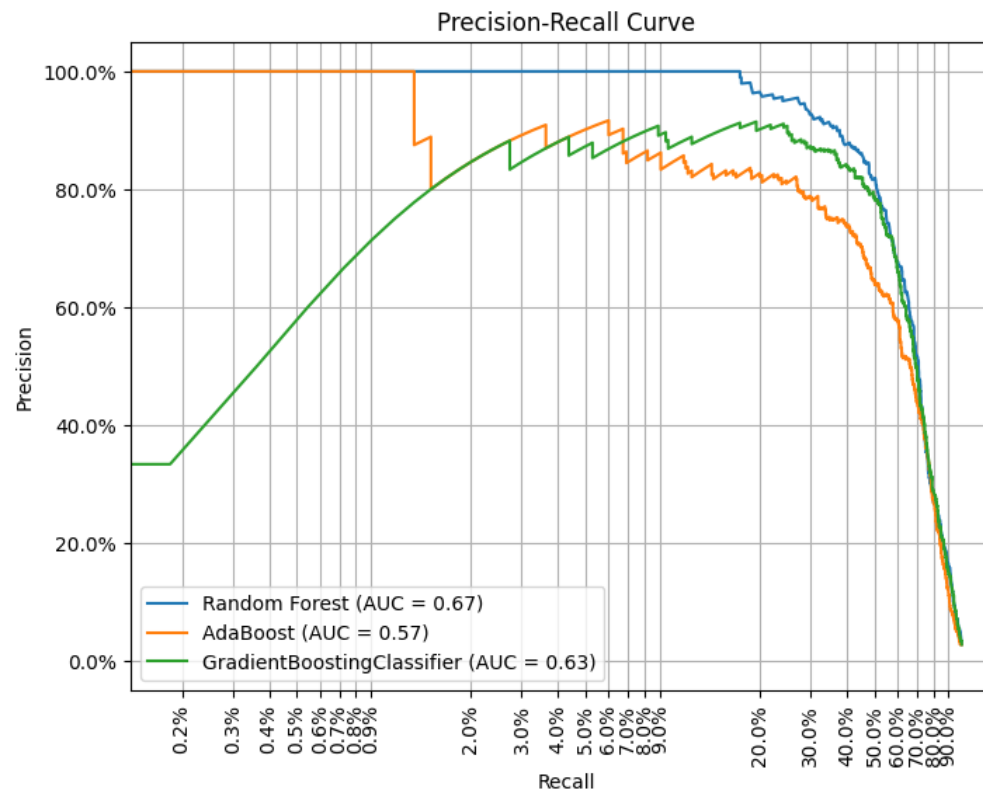


Vertailu edellisiin kokeiluihin (3/3)

Precision-recall -käyrät Kelan vuoden 2023 kokeilussa (logaritmiset x-akselit):



Precision-recall -käyrät tässä kokeilussa (logaritmiset x-akselit)



Johtopäätökset

- Mallin työstöaikataulu oli tiukka, mutta tulokset näyttäisivät huomattavasti paremmilta kuin aiemmissa kokeiluissa.
 - Yksi syy tähän on varmastikin uusien muuttujien tuominen osaksi koulutusdataa.
- Mallilla pyritään ennustamaan n.s. harvinaista eventtiä, jolloin precision-recall –tarkastelu järkevämpää kuin ROC.
- Malli voisi olla hyödynnettävissä esim. "liikennevalotyyppisesti" siten, että ihminen saa mallin ennusteen signaalina (human in the loop).
 - Sääten confidence-thresholdin arvoa
 - voitaisiin saada kiinni esim. 40% työkyvyttömyysetuudelle päätyvistä siten, että virhepositiivisia tulisi n. 13%
 - TAI 30% työkyvyttömyysetuudelle päätyvistä siten, että virhepositiivisia tulisi n. 7-8%
 - TAI 20% työkyvyttömyysetuudelle päätyvistä siten, että virhepositiivisia tulisi n. 2%.
- Olisi hyvä testata käytännössä, jotta saataisiin paremmin ymmärrystä operatiivisesta hyödynnettävyydestä.

Mitä seuraavaksi?

- Tässä kokeilussa uusista muuttujista on selvää hyötyä. Voitaasiinko mallin recall-arvoja parantaa vielä lisää lisämuuttujilla – esim. eri toimijoiden datasettejä yhdistelemällä?
- Työkyvyttömiä tunnistaminen jotenkin työmarkkinatuen saajien joukosta ja tämän ymppääminen vastemuuttujiin?
- Tarkasteltavan datasetin kasvattaminen? Tässä kokeilussa datasetti rajattiin 100,000 havaintoon.