Poliittisten asenteiden binääriluokittelu koneoppimismenetelmillä

Oskari Lahtinen

Pro gradu -tutkielma
Toukokuu 2026

MATEMATIIKAN JA TILASTOTIETEEN LAITOS

Tarkastajat:
Prof. Ion Petre
Dos. Yury Nikulin

Selvitys tekoälyn käytöstä

Tässä gradussa on käytetty tekoälyä aihepiirin opiskeluun, sisällön ja tilastoanalyysien suunnitteluun sekä kielenhuoltoon. Käytetyt työkalut olivat ChatGPT:n versio 5.2. ja sitä vanhemmat versiot, Clauden Sonnet 4.6 ja sitä vanhemmat versiot ja Gemini 3.0/3.1 ja niitä vanhemmat versiot. Tekoälyä käytettiin erityisesti gradun aihepiirin opiskeluun. Kielimalleilta pyydettiin selityksiä useiden gradun keskeisten käsitteiden määritelmistä, merkityksestä ja ominaisuuksista. Kehotteiden muoto oli esimerkiksi "selitä XGBoostin kaavan parametrit"tai "mikä on PR-AUC". Tekoälyä käytettiin aluksi tutkielman rakenteen suunnitteluun.

Työskentelyn aluksi kielimalleilla tuotettiin harjoitusluontoinen “gradu”, jota käytettiin aihepiirin opiskeluun ja tulevan tekstin otsikkorakenteen ja suuripiirteisen sisällön suunnitteluun. Tätä tekstiä ei kuitenkaan käytetty varsinaisena gradutekstinä ja tutkielman teksti on kirjoitettu itse. Tekoälyä käytettiin myös analyysikoodin suunnitteluun ja ohjelmoinnin harjoitteluun. Ensimmäiset analyysit tein kielimallin kokonaan tuottamalla koodilla. Varsinaista gradua varten opiskelin kuitenkin itse kaiken tarvittavan koodin ja kirjoitin sen itse. Kehotteiden muoto harjoitellessa oli esimerkiksi "koodi ROC-AUC-tulosten visualisointiin"tai "mitä parametreja gradienttitehostuksen funktio ottaa".

Tekoälyä käytettiin myös kielihuoltoon ja tekstin muotoiluun LaTeXissa. Tekoälyn tuottamien vastausten oikeellisuutta arvioitiin vertaamalla tekoälyn ehdotuksia gradun lähdekirjallisuuteen, erityisesti koneoppimisen perusteoksiin, sekä hakukoneella löydettyihin ohjelmointiresursseihin.

TURUN YLIOPISTO, Matematiikan ja tilastotieteen laitos

Pro gradu -tutkielma
Pääaine: Matematiikka
Tekijä: Oskari Lahtinen
Otsikko: Poliittisten asenteiden binääriluokittelu koneoppimismenetelmillä
Ohjaaja: Prof. Ion Petre
Sivumäärä: 32 sivua
Aika: Toukokuu 2026

Tutkielmassa verrattiin kolmen koneoppimismallin kykyä luokitella poliittisia asenteita selittävien muuttujien avulla. Ennustettavat luokittelumuuttujat olivat vastaajan kokemus onko hän "woke" (kyllä tai ei) ja kokeeko vastaaja väkivallan oikeutetuksi "poliittisesti vaarallisia" ihmisiä kohtaan. Luokittelussa käytettiin 18 selittävää muuttujaa, jotka jaettiin demografisiin (esim. ikä ja sukupuoli), poliittisiin (esim. arvostan demokratiaa) ja psykologisiin (esim. arvio omasta onnellisuudesta 1-10). Käytetyt koneoppimismallit olivat logistinen regressio, gradienttitehostus ja XGBoost.

Tutkielmassa käsiteltiin tekijän itse keräämiä poliittisen psykologian aineistoja vuosilta 2022, 2025 ja 2026, joiden otoskoot olivat $n = 4934$, $n = 626$ ja $n = 1066$. Ensimmäisellä näistä koulutettiin malli ja kahta jälkimmäistä käytettiin valmiin mallin testaamiseen. Luokittelun onnistumista tarkasteltiin standardimetriikoilla: ROC-AUC, PR-AUC, tarkkuus, tasapainotettu tarkkuus ja F1. Muuttujien permutaatiotärkeyttä tarkasteltiin tärkeimpien selittäjien löytämiseksi ja poistokokeilla tarkasteltiin vaihtoehtoisia, yksinkertaisempia malleja.

Tutkielmassa havaittiin, että koulutetut mallit toisintuivat myöhemmin kerätyissä aineistoissa hyvin. Samoin havaittiin, että yksinkertaisetkin mallit ennustivat toista luokittelutehtävää kohtalaisen hyvin: "woke"-itsearviota kyettiin ennustamaan kohtalaisesti pelkällä turvallisempiin tiloihin liittyvällä kysymyksellä koulutetulla mallilla. Tutkielmassa tarkasteltiin uudehkoa poliittisen psykologian sovellusaluetta, jolla tutkimuskirjallisuutta on rajallisesti. Tämä tutkielma on luultavasti kyseisen alan tutkimuskirjallisuudessa menetelmävalinnaltaan ensimmäinen laatuaan.

Asiasanat: poliittiset asenteet, binääriluokittelu, koneoppiminen, "woke", poliittinen väkivalta

Sisällys

1 Johdanto

Poliittisessa psykologiassa on kehitetty mittareita erilaisille asenteille ja ominaisuuksille. Tällaisia ovat tässä tutkielmassa käsiteltävät kriittisen sosiaalisen oikeudenmukaisuuden asenteet ja poliittisen väkivallan oikeuttaminen. Tutkielman tekijä on itse osallistunut tutkimukseen, jossa näitä ominaisuuksia on kartoitettu psykometrisesti ja kehittänyt ensimmäisen julkaistun kriittisen sosiaalisen oikeudenmukaisuuden asenteiden mittarin, joka on sittemmin käännetty muillekin kielille.

Psykometriikassa ominaisuuksia mitataan tyypillisesti kehittämällä erilaisia testikysymyksiä ja tarkastelemalla sitten näiden yhteisvaihtelua eksploratiivisen ja konfirmatorisen faktorianalyysin keinoin. Tutkimuksen pyrkimyksenä on tuottaa instrumentti, jonka reliabiliteetti ja validiteetti, jaettuna useisiin eri alalajeihin, on mahdollisimman hyvä.

Tämän tutkielman tarkoituksena on tutkia, missä määrin tarkasteltavia poliittisia asenteita kyetään onnistuneesti luokittelemaan muilla keinoin kuin perinteisillä faktorianalyysiin perustetuilla mittareilla. Suuriin aineistoihin kätkeytyy herkästi ilmiöitä, joiden olemassaoloa voi olla vaikea arvata etukäteen.

Koneoppimismenetelmien eräs keskeinen sovellus on antaa tilastollisille algoritmeille suuria aineistoja ja antaa mallin oppia, millä lailla aineistossa esiintyvät muuttujat ovat yhteydessä toisiinsa. Tutkijan tehtävä on tyypillisessä ohjatussa oppimisessa valmistella aineisto, nimetä selittävät ja selitettävät muuttujat ja määritellä oppimisessa käytettävä algoritmi. Algoritmi pyrkii minimoimaan virhefunktion ja muodostamaan mahdollisimman hyvin yleistyvän mallin.

Empiirisen psykologian tutkimustulokset ovat usein teoreettisesti intuitiivisia, mutta ajoittain selitykset tutkittaville ilmiöille voivat olla myös epäintuitiivisia. Eksploratiivinen analyysi koneoppimismenetelmillä voikin tuottaa ilmiöstä yhteyksiä ja selitysmalleja, joita ei ole osattu ennakoida hypoteesien muodostuksessa. Siinä missä psykometriikka pyrkii tuottamaan aihepiiriin teoriaan perustuvia mittareita, koneoppimisessa erityispainotuksina ovat ennustetarkkuus ja mallin yleistyminen muihin aineistoihin. Faktorianalyysillä tuotettu reliaabeli ja validi mittari toimii, mutta on myös aiheellista tarkastella, mitkä muut ratkaisut saattavat myös toimia.

Tässä tutkielmassa tarkasteltavat tutkimuskysymykset ovat:

1. Mikä tarkastelluista kolmesta mallista (logistinen regressio, gradienttitehostus, XGBoost) kykenee parhaiten ennustamaan tarkasteltavien poliittisten asenteiden (“woke”, poliittisen väkivallan oikeuttaminen) luokittelua?

2. Mitkä selittävät muuttujat ovat tärkeimpiä mallien ennustustarkkuuden kannalta?

3. Missä määrin koulutetut mallit yleistyvät kahteen erilliseen testiaineistoon?

4. Voiko malleja yksinkertaistaa ilman merkittävää ennustustarkkuuden häviämistä?

Näiden lisäksi tarkastellaan mielenkiintoisina lisäkysymyksinä muun muassa sitä, missä määrin kahden kohdemuuttujan ennustaminen tutkielman selittävillä muut-

tujilla eroaa toisistaan: onko toinen tehtävä vaikeampi kuin toinen? Tutkielmassa tarkastellaan myös, missä määrin binäärisen luokittelun kynnysarvoja optimoimalla voidaan helpottaa epätasapainoisen aineiston luokittelua.

Tämän tutkielman kirjoittamisessa on hyödynnetty ChatGPT-, Claude- ja Gemini-tekoälytyökaluja kielenhuoltoon, analyysien suunnitteluun, sekä tekstin rakenteen ja sisällön työstämiseen. Tekoälyn käytöstä on annettu erillinen selvitys tutkielman alussa.

2 Binäärinen luokitteluongelma

Koneoppimisella, jota nimitetään myös tilastolliseksi oppimiseksi, tarkoitetaan laajaa menetelmäjoukkoa, joilla pyritään ymmärtämään datan ominaisuuksia ja rakennetta.[1] Menetelmät voidaan jakaa ohjattuun ja ohjaamattomaan oppimiseen. Ohjatussa oppimisessa pyritään löytämään malli, joka ennustaa annetuista syötteistä annetun tulosteen. Ohjaamattomassa oppimisessa pyritään löytämään malli, joka kuvaa aineiston rakennetta ja muuttujien välisiä suhteita.[2]

Yleisessä luokitteluongelmassa on joukko havaintopareja $\{(x_i, y_i)\}_{i=1}^n$. Binäärisessä luokitteluongelmassa ennustavat tai selittävät arvot $x_i \in \mathbb{R}^p$ ja kohdearvot $y_i \in \{0, 1\}$, missä n on havaintojen lukumäärä ja p selittävien muuttujien lukumäärä. Luokittelussa voidaan myös estimoida ehdollinen todennäköisyys sille, että Y saa arvon 0 tai 1 datalla $X = x$, toisin sanoen $P(Y = 1|X = x)$.

Kun on tarpeen tarkastella luokkiin jakautumisen todennäköisyyksiä, voidaan päättelyn apuna käyttää Bayesin luokittelijaa. Sillä tarkoitetaan binäärisen luokittelun tapauksessa sääntöä, joka antaa ensimmäisen luokan kun $P(Y = 1|X = x) \geq 0,5$ ja toisen luokan, kun $P(Y = 1|X = x) < 0,5$.[3]

Bayesin luokittelijaan ja luokitteluongelmaan laajemmin liittyy virhefunktion käsite. Virhefunktiolla tarkoitetaan Y:n todellisen arvon sekä ennustetun Y:n arvon välistä eroa tai poikkeamaa eli virhettä. Kun Y:llä merkitään kaikkien kohdearvojen joukkoa ja Y':llä kohdemuuttujan ennustettujen arvojen joukkoa, virhefunktio L on kuvaus $L : Y \times Y' \to \mathbb{R}_+$.[4]

Määritelmä 2.1. Bayesin luokittelijan tapauksessa virhefunktio on[3]

$$P(\text{virhe}) = 1 - \max\big(P(Y = 1 \mid x), P(Y = 0 \mid x)\big) \tag{1}$$

Mallin parantaminen tarkoittaa käytännössä tämän virhefunktion minimointia. Koska kyseinen virhefunktio saa ainoastaan arvoja 0 ja 1, se on ei-jatkuva ja ei-konveksi. Käytännön sovelluksissa sitä approksimoidaan konveksilla funktiolla, kuten logistisella tappiofunktiolla[4]:

$$\Phi(u) = \log(1 + e^u) \tag{2}$$

Mallin **yleistymisellä** tarkoitetaan sen kykyä ennustaa kohdemuuttujan arvoja eri datalla, kuin millä malli on alunperin koulutettu.[3] Mikäli malli on kompleksisuudeltaan yksinkertaisempi kuin aineistossa selittäjien ja kohdemuuttujien välillä vallitseva yhteys, mallin sanotaan olevan **alisovitettu**. Mikäli malli on kompleksisuudeltaan monimutkaisempi kuin selittäjien ja kohdemuuttujien yhteys, mallin sanotaan olevan **ylisovitettu**.[3]

Mallin kokonaisvirhe hajoitetaan yleensä kahteen komponenttiin: systemaattinen virhe (bias) ja varianssi. Kun mallin kompleksisuus kasvaa, varianssi kasvaa, mutta systemaattinen virhe pienenee.[3] Vastaavasti kompleksisuuden pienentyessä varianssi pienenee ja systemaattinen virhe kasvaa. Täten mallin kokonaisvirheen riippuvuus kompleksisuudesta noudattaa U-mallista käyrää.[5]

Kompleksisuuden vaikutusta malliin säädellään **säännöllistämisellä**, mikä tarkoittaa rangaistustermin lisäämistä optimointilausekkeeseen.

Määritelmä 2.2. Yleisimmin käytetyt säännöllistämismenettelyt ovat[4]

$$L1(\text{Lasso}) : \Omega(\theta) = \|\theta\|_1$$
$$L2(\text{Ridge}) : \Omega(\theta) = \|\theta\|_2^2$$

L1 korjaa useiden parametrien arvoja nolliksi, kun taas L2 pienentää arvoja kohti nollaa, mutta ei aseta niitä täsmälleen nollaan.[3] Koneoppimisalgoritmi koulutetaan yhdellä datalla ja mallia testataan erikseen toisella, koskemattomalla datalla. Tämän lisäksi koulutusdatan sisällä voidaan tehdä **ristiinvalidointia**, missä data jaetaan k osaan, koulutetaan malli $k-1$:llä osalla ja testataan jäljelle jääneellä osalla. Menettelyä toistetaan k kertaa.

3 Luokittelumenetelmät

3.1 Luokittelumittarit ja mallin arviointi

Määritelmä 3.1.1. Sekoitusmatriisi muodostaa esityksen luokittelumallin onnistumiselle. Binäärisellä mallilla on neljä mahdollista tapausta: oikeat positiiviset (TP), väärät positiiviset (FP), oikeat negatiiviset (TN) ja väärät negatiiviset (FN).[3]

Esitellään hyödyllisimpiä metriikkoja luokittelumallin tarkkuuden arviointia varten.[6]

Määritelmä 3.1.2. Mallin kokonaistarkkuus on

$$\frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

Vastaavasti mallin virhe on

$$\frac{FP + FN}{TP + FP + TN + FN} \tag{4}$$

Kokonaistarkkuus ja virhe voidaan muodostaa myös toisistaan laskemalla virheelle $1 -$ kokonaistarkkuus ja kokonaistarkkuudelle $1 -$ virhe.

Määritelmä 3.1.3.

$$\text{Täsmällisyys (precision)} = \frac{TP}{TP + FP} \tag{5}$$

Täsmällisyydellä kuvataan siis kuinka moni positiiviseen luokkaan kuuluvaksi ennustetuista arvoista oli tosiasiassa positiivinen.

$$\text{Sensitiivisyys (recall)} = \frac{TP}{TP + FN} \tag{6}$$

Sensitiivisyydellä tarkoitetaan siis samaa kuin oikeiden positiivisten osuudella (TPR, true positive rate).

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

Määritelmä 3.1.4. ROC-käyrä on lyhenne sanoista ”receiver operating characteristics”.[7] Luvussa 2 käytettiin luokkiin jakamisen kriteerinä sitä, onko todennäköisyys että tapaus kuuluu luokkaan $\geq 0,5$. Merkitään tätä rajaa θ:lla. ROC-käyrän x-akselina toimii väärien positiivisten suhdeluku ja y-akselina oikeiden positiivisten suhdeluku. ROC-käyrä muodostuu, kun käydään läpi kaikki parametrin θ arvot välillä $[0, 1]$.

ROC-käyrään liittyy läheisesti myös metriikka AUC (area under curve), jolla tarkoitetaan ROC-käyrän alle jäävää aluetta ja joka voi saada arvoja väliltä $[0, 1]$. Malli

on sitä parempi, mitä suurempi AUC-arvo on. ROC-käyrään liittyvää AUC-suuretta kutsutaan ROC-AUC, muista AUC-suureista erottamisen vuoksi.

Aineiston luokkien jakautumisen ollessa tasapainoinen sekoitusmatriisi ja ROC-AUC voivat olla melko informatiivisia mallin ennustustarkkuuden suhteen. Kuitenkin epätasapainoisten luokkien tapauksessa on hyödyllistä katsoa epätasapainon huomioivia metriikoita.[8] Tärkein tällainen metriikka on täsmällisyys-sensitiivisyys-käyrä (precision-recall curve), eli PR-käyrä.

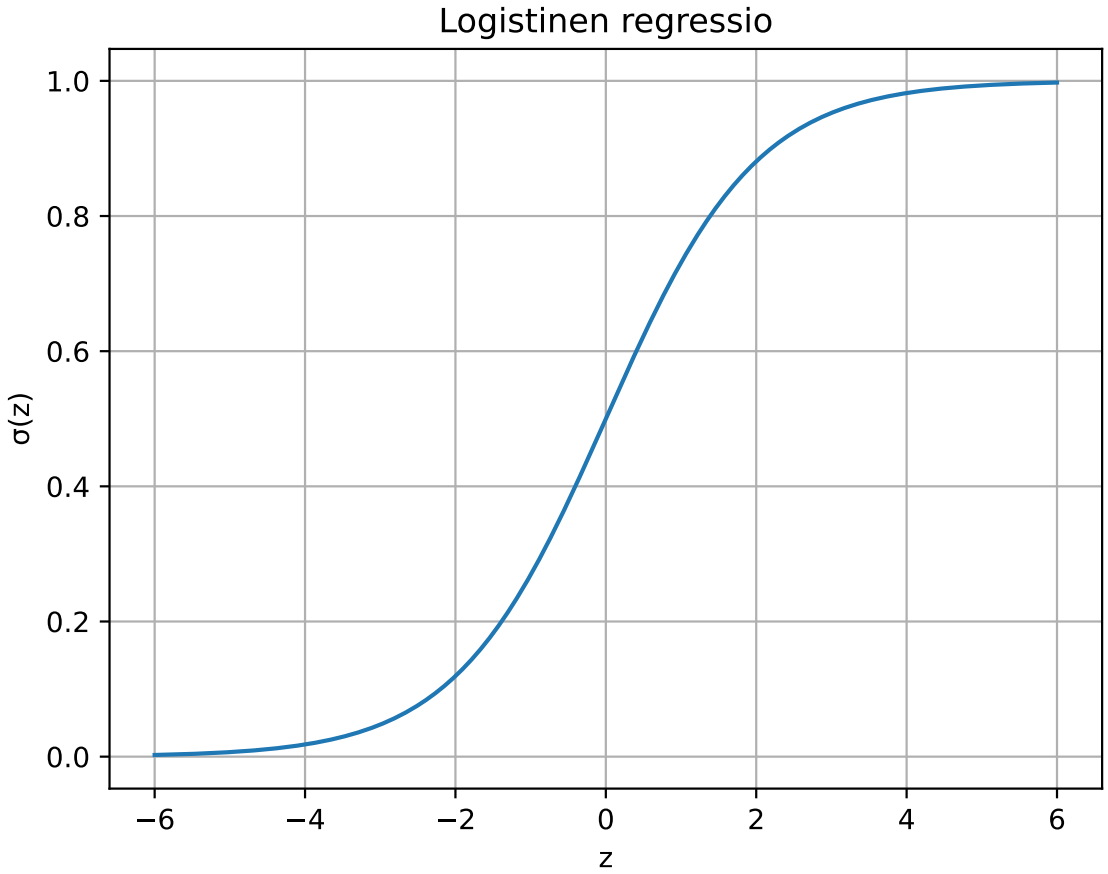
Määritelmä 3.1.5. PR-käyrän y-akseli kuvaa täsmällisyyttä ja x-akseli sensitiivisyyttä θ:n käydessä läpi arvot väliltä $[0, 1]$. PR-käyrään liittyy, kuten ROC-käyrään, vastaava suure PR-AUC.

PR-käyrän lisäksi epätasapainoisissa aineistoissa voidaan käyttää **tasapainotettua tarkkuutta**, joka lasketaan kaavalla $\frac{1}{2}(TPR + TNR)$, sekä kynnysoptimoinnilla löydettyä päätöskynnystä.

3.2 Logistinen regressio

Logistinen regressio (kuva 1) mallintaa todennäköisyyttä sille, että Y:n arvo on 0 tai 1, kun selittäjät ovat X.[5][9][10] Tehtävänä on siis löytää tapa mallintaa X:n ja $p(X) = P(Y = 1 \mid X)$:n välistä suhdetta. Yksi tapa lähestyä tällaisen suhteen mallinnusta on tarkastella lineaarista regressiota $p(X) = \beta_0 + \beta_1 X$. Tämän funktion vaihteluväli kuitenkin alittaa 0:n ja ylittää 1:n, jonka vuoksi se ei sellaisenaan sovi mallintamaan välille $[0, 1]$ asettuvaa todennäköisyyttä. Ongelma ratkeaa siirtymällä käyttämään **logistista funktiota**:

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \tag{8}$$



Kuva 1: Logistinen regressio

Malli sovitetaan käyttämällä suurimman uskottavuuden menettelyä.

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} (1 - p(x_{i'})) \tag{9}$$

Uskottavuusfunktio maksimoidaan etsimällä sopivat $\hat{\beta}_0$ ja $\hat{\beta}_1$. Logistisen regression mallia optimoidaan gradienttilaskeutumisella, Newtonin menetelmällä tai kvasi-Newtonin menetelmällä, kuten L-BFGS, jota tässä tutkielmassa käytettiin.[11] Muokkaamalla kaavaa 8 saadaan:

$$\log\left(\frac{p(X)}{1 - p(X)}\right) = \beta_0 + \beta_1 X \tag{10}$$

Yhtälön vasenta puolta kutsutaan nimillä log-riskisuhde tai logit. Se on lineaarisessa suhteessa X:ään. Funktiota $p(X) = \sigma(z) = \frac{1}{1+e^{-z}}$ nimitetään **sigmoidiksi**.

Lopuksi esitetään logistisen regression säännöllistetty muoto

$$\hat{\beta} = \arg\min_{\beta} \{-\ell(\beta) + \lambda\omega(\beta)\} \tag{11}$$

jossa lambdan paikalle sijoitetaan L_1 tai L_2 säännöllistystermi.

3.3 Päätöspuut

Päätöspuu on hierarkinen, ohjatun oppimisen malli. Siinä jaetaan laajempi alue lokaaleihin ala-alueisiin rekursiivisten päätösten sarjalla. Puu koostuu solmuista, joissa tehdään päätökset ja lehdistä, jotka toimivat terminaaleina.[3]

Luokittelupuun jaot tehdään perustuen epäpuhtausmittaan (impurity measure). Tällaisia mittoja ovat esimerkiksi entropia ja Gini-indeksi. Puiden kompleksisuutta rajoitetaan karsimalla (pruning), missä solmujen määrää rajoitetaan joko ennen puun muodostusta (prepruning) tai sen jälkeen (postpruning).[3]

3.4 Gradienttitehostus

Gradienttitehostus (boosting) on ensemble-menetelmä, jolla tarkoitetaan menetelmiä, joissa yhdistetään toisiinsa useita perusoppijoita (base-learners), joita kutsutaan myös heikoiksi oppijoiksi (weak learners)[5][3]. Gradienttitehostuksessa perusoppijana toimivia puita koulutetaan edellisten puiden virheiden perusteella[3][13][2]. Toisin kuin yksittäistä päätöspuuta käytettäessä, lukuisia puita toisiinsa yhdistävä malli rakentuu hitaasti iteraatioiden myötä[5]. Uusi puu sovitetaan edellisten puiden yhdistelmän residuaaleihin. Puiden ominaisuuksia säädellään määrittämällä algoritmin hyperparametrit, joita ovat puiden lukumäärä, mallin askelkoko ja puiden maksimisyvyys.

Määritelmä 3.1.6. Gradienttitehostuksen algoritmi määritellään

1. Asetetaan $\hat{f}(x) = 0$ ja $r_i = y_i$ kaikille koulutusdatan i.

2. $b = 1, ..., B$:

   (a) Sovitetaan puu $\hat{f}^b$, jolla d jakoa ($d + 1$ terminaalisolmua) koulutusdataan (X, r)

   (b) Päivitetään $\hat{f}$ lisäämällä siihen kutistunut versio uudesta puusta: $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda\hat{f}^b$

   (c) Päivitetään residuaalit $r_i \leftarrow r_i - \lambda\hat{f}^b(x_i)$

3. Saadaan tehostettu malli $\hat{f}(x) = \sum_{b=1}^{B} \lambda\hat{f}^b(x)$.

Kun tappiofunktiona on neliövirhe, gradientit ovat residuaaleja r_i ja niiden muoto on $r_{im} = y_i - f_{m-1}(x_i)$, missä m on tehostusiteraatioiden järjestysnumero.[12] Määritelmän 3.1.6. algoritmi kuvaa tällaisen mallin funktionaaliseen gradienttilaskeutumiseen perustuvaa optimointia.

3.5 XGBoost

Gradienttitehostuksen paranneltu versio XGBoost lisää edellisessä kohdassa kuvattuun menettelyyn kolme lisäominaisuutta.[14] Ensinnäkin se ehkäisee ylisovittamis-

ta lisäämällä tappiofunktioon $\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i)$ säännöllistämistermin $\Omega(f) = \gamma T + \frac{1}{2}\lambda\|w\|^2$. Toiseksi XGBoost hyödyntää gradienttien lisäksi myös toisen kertaluvun derivaattoja Taylor-approksimaation muodossa.

$$\tilde{\mathcal{L}}^t = \sum_i \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

Kolmanneksi XGBoost tekee gradienttitehostukseen nähden teknisiä parannuksia: rinnakkaislaskenta, harvan datan hyödyntäminen, sekä välimuistin käyttö laskennan nopeuttamiseksi.

3.6 Muuttujien merkitys

Muuttujien tärkeyttä mallille voidaan arvioida mallista riippuvilla (kuten logistisen regression β-kertoimet) ja mallista riippumattomilla menetelmillä. Mallista riippumattomia menetelmiä ovat esimerkiksi permutaatiotärkeyden tarkastelu[15] ja poistokokeet (ablation analysis)[16].

Permutaatiotärkeydellä tarkoitetaan toimenpidettä, jossa jonkin selittäjän arvovektorin komponentit permutoidaan ja tarkastellaan mallia ilman tuon muuttujan tavallista vaikutusta. Poistokokeissa mallista poistetaan muuttujia kokonaan ja tarkastellaan mallia ilman niitä.

4 Tutkimusasetelma

4.1 Aineiston kuvaus

Tutkimuksen datana käsitellään kolmea aineistoa, jotka graduntekijä on kerännyt tutkijan ominaisuudessa vuosina 2022 (data 1), 2025 (data 2) ja 2026 (data 3). Aineistojen otoskoot ovat järjestyksessä $n = 4934$, $n = 626$ ja $n = 1066$. Aineistot data 1 ja data 3 on kerätty Helsingin Sanomien lukijakunnasta ja niiden otoksia voidaan siis pitää jossain määrin vertailukelpoisina. Aineisto data 2 on kerätty osana suomalaista äärioikeistoa ja äärivasemmistoa käsitellyttä tutkimusta ja sen aineiston voidaan otaksua jossain määrin poikkeavan kahdesta muusta. Aineistoista data 1 ja data 2 on julkaistu aiempia tutkimuksia.[17][18]

Näistä ensimmäistä ja suurinta aineistoa (data 1) käytetään koneoppimismallien kouluttamiseen ja arviointiin. Toiseksi suurinta aineistoa käytetään mallien lopulliseen testaamiseen koulutusvaiheen päätyttyä. Pienintä aineistoa (data 2) käytetään toisena testaamisdatana tutkimuksen lopuksi. Tuloksien hyvää yleistymistä dataan 2 pidetään vähemmän todennäköisenä kuin Helsingin Sanomien lukijoita käsittelevien otoksien välillä. Aineistot pidettiin tutkimuksen aikana täysin erillään omissa tiedostoissaan.

4.2 Vastemuuttujat ja selittävät muuttujat

Alkuperäiset aineistot käsittivät useita muuttujia. Tähän tutkimukseen otettiin mukaan muuttujat, jotka pysyivät kolmessa aineistossa täsmälleen samoina. Muut muuttujat poistettiin. Aineistojen 2 ja 3 muuttujat olivat lähes identtiset, mikä lisäsi aineistojen yhteensopivuutta. Muuttujia, jotka pysyivät samoina läpi aineistojen oli kaikkiaan 20. Näistä selittäviksi muuttujiksi valittiin 18 ja vastemuuttujiksi 2. Vastemuuttujat voidaan jakaa **demografisiin**, **poliittisiin** ja **psykologisiin**.

Demografisia muuttujia olivat sukupuoli, ikä, koulutustaso ja onko henkilö opiskelijana tai töissä yliopistolla. Poliittisia muuttujia olivat itsemääritelty sijoittuminen vasemmisto-oikeisto-akselilla, itsemääritelty sijoittuminen liberaali-konservatiivi-akselilla, mitä puoluetta henkilö sanoi äänestävänsä, missä määrin henkilö arvostaa demokratiaa, missä määrin henkilö arvostaa hyvinvointivaltiota sekä missä määrin henkilö oli samaa mieltä seuraavista väitteistä (R = käänteinen pisteytys):

(1) Korkeakoulujen kurssikirjallisuuden tulisi sisältää vähemmän valkoihoisia tai eurooppalaisia kirjailijoita.
(2) Yhteiskunnassa tulisi olla enemmän turvallisia tiloja (=tila, josta on sovituin säännöin pyritty poistamaan ennakkoluulot, konflikti, kritiikki tai potentiaalisesti loukkaavat teot, ideat tai keskustelut).
(3) R Etuoikeutetun ryhmän jäsen saa ottaa käyttöönsä vähemmän etuoikeutetun identiteettiryhmän piirteitä tai kulttuurituotteita.
(4) Mikroaggressioihin (=verbaalinen viesti tai käytös, jonka voi tulkita viestivän negatiivisia asenteita vähemmistöryhmää kohtaan, riippumatta alkuperäisestä tarkoituksesta) tulee puuttua usein ja aktiivisesti.
(5) Transnaiset ovat naisia. ja

(6) R Ihmislajilla on kaksi biologista sukupuolta.

Psykologisia muuttujia olivat henkilön ilmoittamat kokemukset sorrosta, ulkoinen kontrollikäsitys (”Muut ihmiset tai rakenteet ovat hyvinvoinnistani vastuussa itseäni enemmän.”) sekä onnellisuus asteikolla 0-10.

Vastemuuttujia olivat missä määrin henkilö koki olevansa “woke” (”Jos ystäväni kutsuisi minua hyvää tarkoittaen sanalla “woke”, olisin arviosta samaa mieltä - riippumatta siitä hyväksynkö itse termin.”) ja missä määrin hän hyväksyi poliittisen väkivallan (”Väkivalta poliittisesti vaarallisia ihmisiä kohtaan on mielestäni oikeutettua.”).

4.3 Esikäsittely ja mallinnusstrategia

Koska kyseessä oli binäärinen luokittelu, molemmat vastemuuttujat binärisoitiin. Ne olivat alunperin viisipaikkaisia Likert-muuttujia (1 = täysin eri mieltä, 2 = jokseenkin eri mieltä, 3 = ei samaa eikä eri mieltä, 4 = jokseenkin samaa mieltä ja 5 = täysin samaa mieltä), joten luokat 1-2 yhdistettiin luokaksi 0 ja luokat 4-5 luokaksi 1. Luokka 3 poistettiin analyyseistä. Tämä datan esikäsittely tehtiin SPSS-tilasto-ohjelmalla, minkä jälkeen siirryttiin jatkamaan analyysejä Jupyter notebookeja käsittelevässä Google Colab -ympäristössä. Ympäristö sopii erityisen hyvin koneoppimismallien toteuttamiseen ja ohjelmointi toteutetaan Python-kielellä. Ohjelmoinnin tukena käytettiin erityisesti pandas, NumPy, Scikit-learn ja Matplotlib-kirjastoja.

Selittävien muuttujien valmistelu Colabissa aloitettiin tarkastelemalla puuttuvia arvoja.

Taulukko 1: Puuttuvat arvot muuttujittain

Muuttuja	Puuttuvia
Woke	1292
Liberaali-konservatiivi-akseli	1067
Mitä puoluetta äänestää	907
Vasemmisto-oikeisto-akseli	892
Ikä	697
Poliittisen väkivallan oikeuttaminen	429
Koulutus	209
Kulttuurinen appropriaatio	182
Sukupuoli	161
Korkeakoulujen kurssikirjallisuus	157
Turvalliset tilat	142
Arvostan demokratiaa	131
Työt/opinnot yliopistolla	113
Mikroaggressiot	97
Transnaiset ovat naisia	94
Arvostan hyvinvointivaltiota	88
Kokenut sortoa	63
Muut ihmiset tai rakenteet	51
Onnellisuus	51
Kaksi biologista sukupuolta	45

Tutkimuksen aineistossa esiintyi epätasaisuutta erityisesti toisen vastemuuttujan osalta: poliittisen väkivallan oikeuttamisen vastauksista koulutusdatassa 12,6% ja testidatan vastauksista vielä harvempi kuului luokkaan 1, joka tarkoitti vastaajan pitävän poliittista väkivaltaa oikeutettuna. Mediaani-imputoinnin katsottiin sopivan kohtalaisen hyvin epätasapainoisen datan käsittelyyn.

Kaikki muuttujat olivat diskreettejä (ikä oli kategorisoitu ikäluokiksi). Selittävät muuttujat olivat joko järjestysasteikollisia, tai kategorisia muuttujia, jotka oli mahdollista korjata järjestysasteikollisiksi. Esimerkiksi puoluekanta koodattiin niin, että vasemmistoliberaalein puolue sai arvon 1 ja oikeistokonservatiivisin puolue arvon 7.

Muuttujat mediaani-imputoitiin.[21][19][20] Kummallekin vastemuuttujalle muodostettiin oma aineisto, johon otettiin mukaan vastausrivit, joissa vastemuuttuja sai arvon 0 tai 1, eli ei ollut puuttuva. Vastausrivit, joissa selittäjän arvo puuttui, poistettiin. Selittäjät standardoitiin, jolloin jokaisen selittäjän keskiarvoksi tuli 0 ja keskihajonnaksi 1.

Esikäsittely toteutettiin Scikit-learn-kirjaston pipeline-toimintoa käyttäen, jossa skaalaus ja imputointi yhdistettiin logistisen mallin osalta jatkumoksi. Selittäjät skaalattiin logistista regressiota varten, koska se ei ole robusti menetelmä eri asteikolla oleville muuttujille. Gradienttitehostusmalleissa jatkumo koostui ainoastaan imputoinnista, sillä ne ovat binääriluokittelussa robusteja skaalauksesta riippumatta.[22] Mallit määriteltiin samoin pipeline-toiminnolla niin, että samaan jatkumoon kuuluivat esikäsittely ja mallin määrittely.

Mallien hyperparametreille ei tehty laajaa optimointia (esim. grid search), vaan

tähdättiin riittävän yksinkertaisiin malleihin, joissa ylisovittamisen riski pysyi hallittavana. Logistisen regression jatkumo sisälsi esikäsittelyn ja Scikit-learnin LogisticRegression-toiminnon. Iteraatioiden määräksi valittiin 2000, koska haluttiin varmistua, ettei iteraatioiden loppuminen estä mallin suppenemista ennenaikaisesti. Ratkaisijana (solver) käytettiin kvasi-newtonilaista L-BFGS-menetelmää, säännöllistämiseen L2-termiä ja $C = 1.0$-standardiasetusta.

Gradienttitehostus toteutettiin 200:lla puulla ja oppimisnopeudeksi valittiin varovainen $0,05$. Jotta malli pysyisi verrattain yksinkertaisena, puiden maksimisyvyydeksi asetettiin 3. Tappiofunktiona oli log-tappio. XGBoostissa oppimisnopeus ja maksimisyvyys pidettiin samoina, mutta puita sallittiin 300 algoritmin vahvemman säännöllistämisen vuoksi. Mallissa käytettiin L2-säännöllistämistä ja jokainen malli käytti 80% sekä muuttujista että havainnoista. Molempien puumallien parametreista päätettäessä pyrittiin välttämään ylisovittamista, minkä vuoksi arvot pidettiin maltillisina. Kuten logistisen regression, tehostusmallien tappiofunktioiden muoto oli log-tappio.

Analyysikoodin sisältävät Google Colab -tiedostot sekä kaikki tutkimuksessa käytetty data ladattiin avoimesti saataville Turun yliopiston Gitlab-palvelimelle avoin data ja koodi.

5 Mallin kouluttaminen

5.1 Mallien vertailu koulutusdatalla

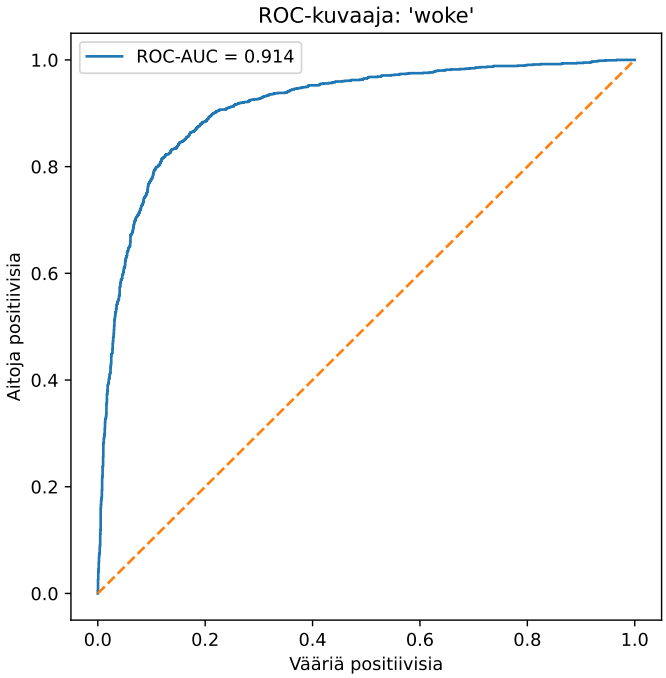
Mallien keskinäinen vertailu toteutettiin Scikit-learn-kirjaston StratifiedKFold-toiminnon avulla. Toiminto mahdollistaa ristiinvalidoinnin, eli aineiston jakamisen erikseen analysoitaviin osioihin säilyttäen samalla kohdemuuttujan luokkajakauman samana osioiden yli. Jakojen määräksi valittiin $n = 5$. Ristiinvalidoinnissa neljää osiota käytettiin aineiston kouluttamiseen ja viidettä validointiin ja toimenpide toistettiin viisi kertaa aina eri yhdistelmällä aineistoja, minkä jälkeen tulokset yhdistettiin. Tulokseksi saatuja todennäköisyyksiä y:n arvolle, kun x tunnetaan, verrattiin kynnykseen, joka oli tässä $\theta = 0,5$. Kun todennäköisyys ylitti kynnyksen, sen katsottiin ennustavan arvon 1 ja muulloin arvon 0.

Malleja arvioitiin lukuisilla luvussa 3. esitellyillä metriikoilla. Koulutusdatan päätulokset on esitetty taulukossa 2.

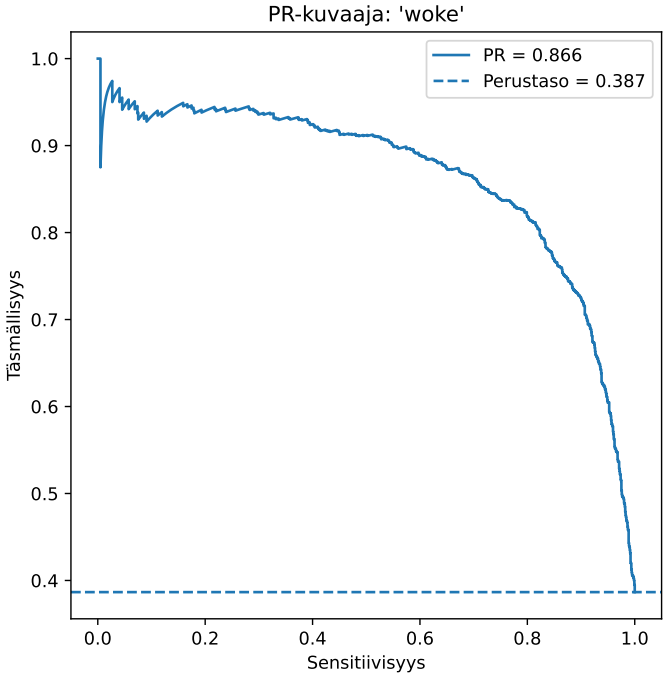
Taulukko 2: Koulutusdatan mallien arviointi

Kohde	Malli	n	ROC-AUC	PR-AUC	Tarkkuus	Tasapainotettu tarkkuus	F1
“Woke”	Logistinen regressio	3642	0,91	0,87	0,86	0,84	0,81
“Woke”	Gradienttitehostus	3642	0,91	0,87	0,85	0,84	0,80
“Woke”	XGBoost	3642	0,91	0,86	0,85	0,84	0,81
Pol. väkivalta	Logistinen regressio	4505	0,68	0,28	0,88	0,52	0,07
Pol. väkivalta	Gradienttitehostus	4505	0,68	0,28	0,88	0,53	0,13
Pol. väkivalta	XGBoost	4505	0,68	0,28	0,88	0,53	0,13

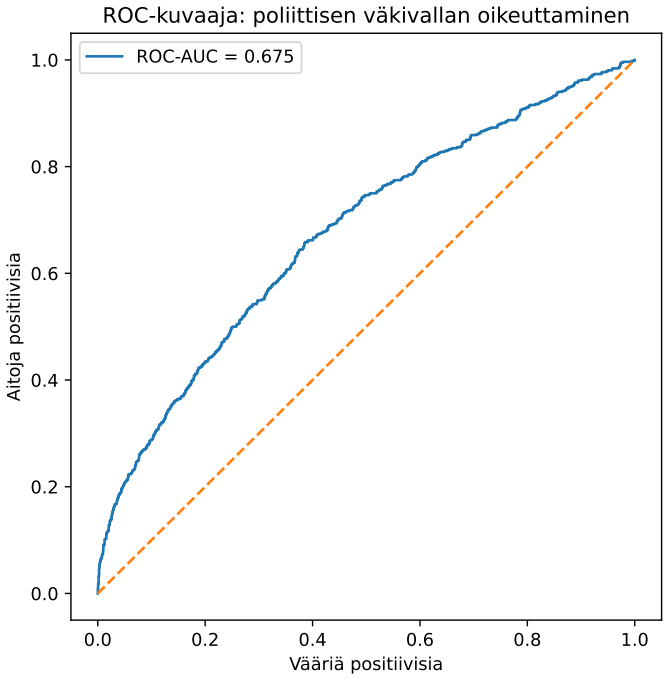
Yleisesti ottaen erot mallien välillä olivat pieniä. Tarkastellessa “woke”-luokittelua ennustavia malleja havaittiin kaikkien mallien saaneen kaikilla metriikoilla melko hyviä arvoja (Taulukko 2). Kaikkien mallien ROC-AUC-luvut ylittivät 0,90, PR-AUC-arvot 0.85 ja tarkkuus- sekä F1-arvot 0,80. Korkeimman tarkkuusarvon (ja ROC-AUC-arvon) sai logistisen regression malli, jota päätettiin jatkossa käsitellä ”parhaana” mallina, vaikka toisenlainenkin valinta oli saatavilla olleilla kriteereillä osin mielivaltainen. Parhaan mallin ROC-AUC-arvo oli erinomainen 0,91 (Kuva 2) ja PR-AUC-arvo sekin korkea 0,87 (Kuva 3). PR-kuvaajiin on lisätty perustasoa kuvaava katkoviiva osoittamaan kuinka moni 1-luokittelu menisi oikein, mikäli luokittelu tehtäisiin arvaamalla puhtaan sattuman varassa.



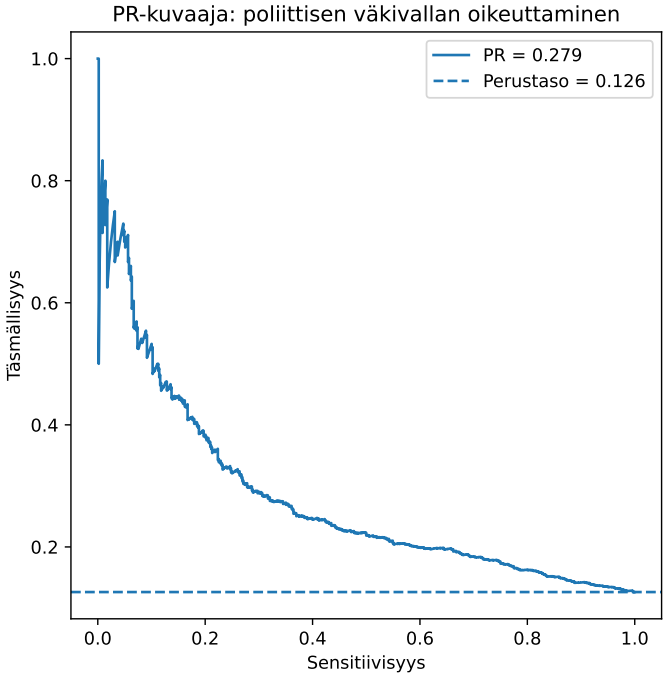
Kuva 2: ROC-kuvaaja “woke”-muuttujan ennustamiselle



Kuva 3: PR-kuvaaja “woke”-muuttujan ennustamiselle



Kuva 4: ROC-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle



Kuva 5: PR-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle

Poliittisen väkivallan hyväksymisen osalta mallit toimivat huomattavasti heikommin kaikkien mallien ROC-AUC-lukujen ollessa 0,68 (Kuva 4) ja PR-AUC-arvojen 0,28 (Kuva 5) kaikille malleille. Erityisen huonoja olivat F1-arvot, erityisesti logistisen regression osalta (0,07). Kuten “woken” luokittelun, myös poliittisen väkivallan

oikeuttamista luokittelevat mallit olivat jokseenkin tasavertaisia muiden suureiden kuin F1-arvon ja mahdollisesti tasapainotetun tarkkuuden osalta. Parhaaksi malliksi valittiin XGBoost-malli korkeimman tarkkuuden perusteella, myös PR-AUC-arvon ollessa hieman parempi kuin tavallisen gradienttitehostuksen mallissa.

5.2 Koulutusdatan luokittelutulokset ja kynnysarvon vaikutus

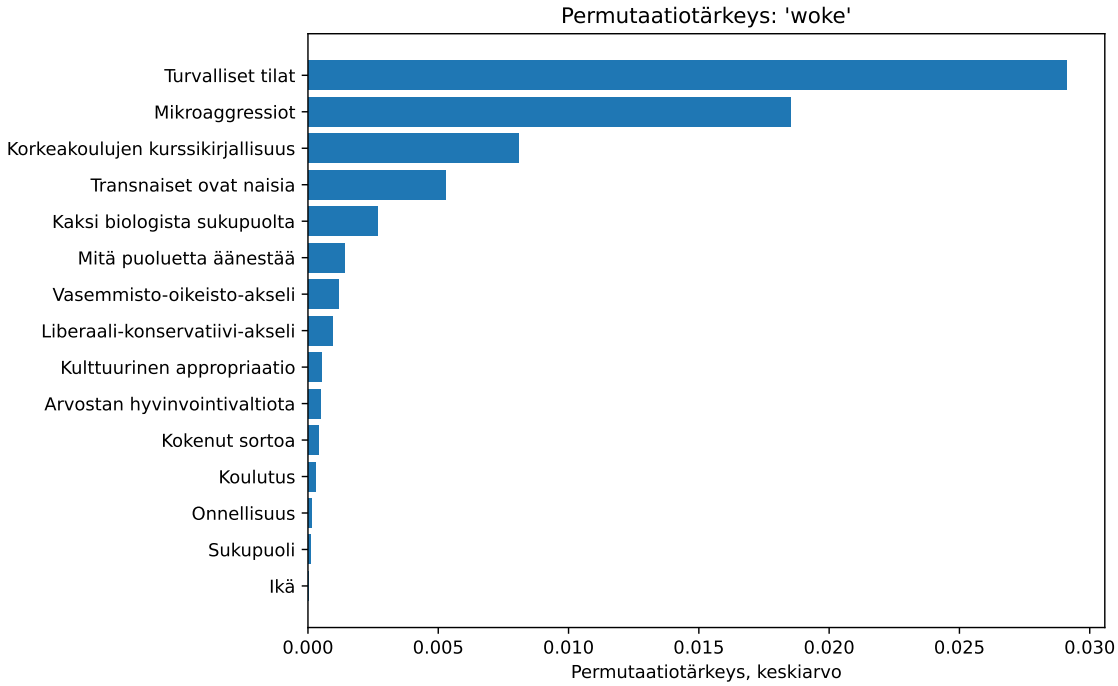
"Woke" muuttujan arvoista nollia oli 61,3% (2234/3642) ja 1-arvoja 38,7% (1408/3642). Paras "woke"-muuttujaa ennustava malli (logistinen regressio) luokitteli 0-havainnoista oikein 89,2% (1993/2234) ja 1-havainnoista 79,5% (1120/1408). "woke"-malli luokitteli oikein kaikista havainnoista 85,5%. Puuttuvia ja siten poistettavia vastauksia oli siis 1292. Poliittisen väkivallan oikeuttamisen muuttujan arvoista oli nollia 87,4% (3937/4505) ja 1-arvoja 12,6% (568/4505), kun 429 arvoa puuttui. Paras poliittisen väkivallan oikeuttamista ennustava malli (XGBoost) luokitteli 0-havainnoista oikein 99,2% (3904/3937), mutta 1-havainnoista vain 7,2% (41/568). Poliittisen väkivallan oikeuttamisen malli luokitteli oikein 87,6% kaikista havainnoista.

Koska erityisesti poliittisen väkivallan oikeuttamisen vastauksia luokitellut malli kykeni niin huonosti ennustamaan 1-vastauksia, tarkasteltiin luokittelukynnyksen vaikutusta luokittelun tarkkuuteen. Analyysi toteutettiin katsomalla Scikit-learnin PR-ROC-toiminnon avulla, millä kynnysarvolla kohdemuuttuja sai suurimman F1-arvon. Tällaiseksi kynnykseksi valikoitui 0.165, jolloin F1 sai huomattavasti paremman, mutta silti keskinkertaisen, arvon 0,31. "woke"-muuttujaan F1-optimointi ei sanottavasti vaikuttanut: uusi kynnys olisi ollut 0,46 ja F1-arvo pysyi käytännössä samana (0,81).

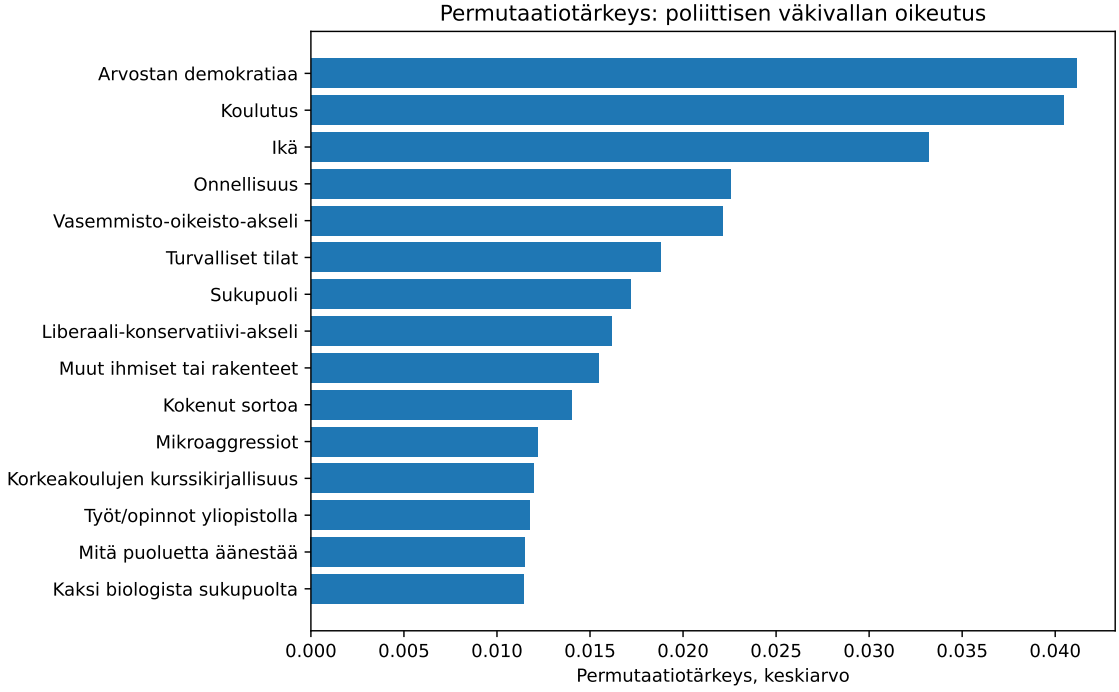
5.3 Koulutusdatan muuttujien tärkeys ja poistokokeet

Mallien selittäjien tärkeyttä tarkasteltiin luomalla funktio, joka permutoi satunnaisesti yksi kerrallaan yhden muuttujan arvot ja tarkastelee sitten mallia ilman tämän muuttujan aiempaa selitysvaikutusta. Tämä toimenpide toistettiin 20 kertaa ja tuloksista laskettiin keskiarvot. Muuttujien permutaatiotärkeydet vastemuuttujille näkyvät kuvissa 6 ja 7. Tärkeimmät selittävät muuttujat "woke"-luokittelun ennustamisessa olivat turvallisiin tiloihin (permutaatiotärkeys = 0,03) ja mikroaggressioihin (0,02) liittyvät kriittisen sosiaalisen oikeudenmukaisuuden uskomukset. Poliittiset muuttujat olivat luokittelussa selvästi tärkeimpiä ja niihin verrattuna demografisten ja psykologisten muuttujien vaikutus oli vähäinen.

Poliittisen väkivallan oikeuttamisen luokittelua ennustavien selittävien muuttujien permutaatiotärkeydet sen sijaan sisälsivät heterogeenisemmin selittäjiä niin poliittisista, demografisista kuin psykologisistakin muuttujista. Korkeimman permutaatiotärkeyden selittäjät olivat "arvostan demokratiaa-uskomus (0,04) sekä henkilön koulutus (0,04).



Kuva 6: “Wokea” ennustavien muuttujien permutaatiotärkeys



Kuva 7: Poliittisen väkivallan oikeuttamista ennustavien muuttujien permutaatiotärkeys

Permutaatiotärkeyksien lisäksi tarkasteltiin poistokokeita (ablation analysis), jossa kriittisen sosiaalisen oikeudenmukaisuuden uskomukset poistettiin mallista (Taulukot 3 ja 4). Tämän lisäksi tarkasteltiin riisuttua mallia, jossa selittäjinä olivat

ainoastaan kolme demografista muuttujaa ikä, sukupuoli ja koulutus sekä sijoittuminen vasemmisto-oikeisto-akselille. Näiden lisäksi tarkasteltiin vielä yhden selittäjän mallia, jossa ainoa selittäjä oli turvallisten tilojen uskomus, jolla oli korkein permutaatiotärkeys "woke"-luokittelun selittäjänä. Poistokokeissa päämielenkiinnonkohteena oli "woke"-luokittelua selittävä malli. Huomionarvoisesti turvallisten tilojen uskomus oli "woke"-luokittelun ainoana selittäjänä ROC-AUC-arvoltaan (0,86) verrattain lähellä täyttä mallia (0,91). Se oli selitysvoimaltaan parempi kuin riisuttu malli, jossa oli mukana 12 demografista, psykologista ja poliittista selittäjää.

Taulukko 3: Poistokokeet: woke

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,91	0,87	0,81
Ei kriit.	0,85	0,79	0,72
Riisuttu	0,82	0,74	0,66
Vain turv.	0,86	0,77	0,75

Taulukko 4: Poistokokeet: poliittisen väkivallan oikeuttaminen

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,68	0,28	0,13
Ei kriit.	0,68	0,28	0,13
Riisuttu	0,64	0,21	0,06
Vain turv.	0,48	0,12	0,00

6 Testidata

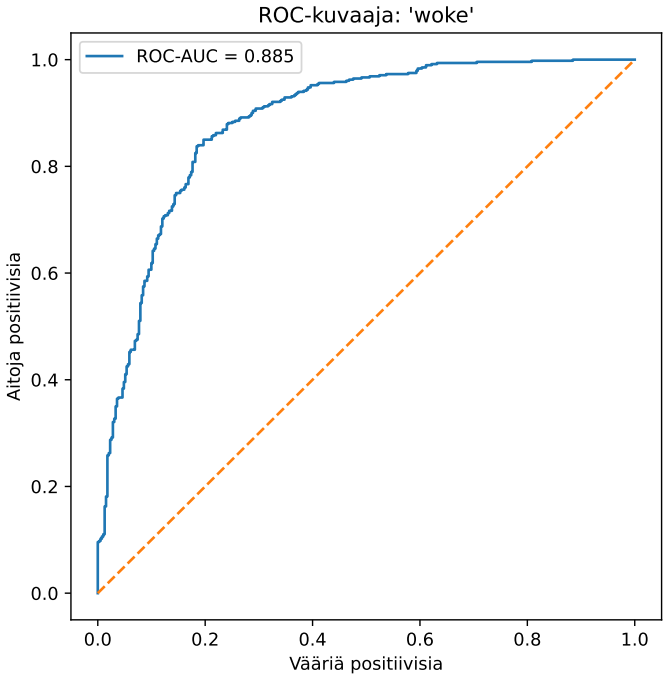
6.1 Testidata 1: tulokset

"Woke"-muuttujan osalta oli testidatassa 1 vastauksia 871, joista 44,8% (391/871) kuului luokkaan 0 ja 55,1% (480/871) luokkaan 1. Puuttuvia ja siten poistettavia vastauksia oli siis 195. Poliittisen väkivallan oikeuttamisen osalta 92,1% (916/995) kuului luokkaan 0 ja 7,9% (79/995) luokkaan 1, kun 71 arvoa puuttui.

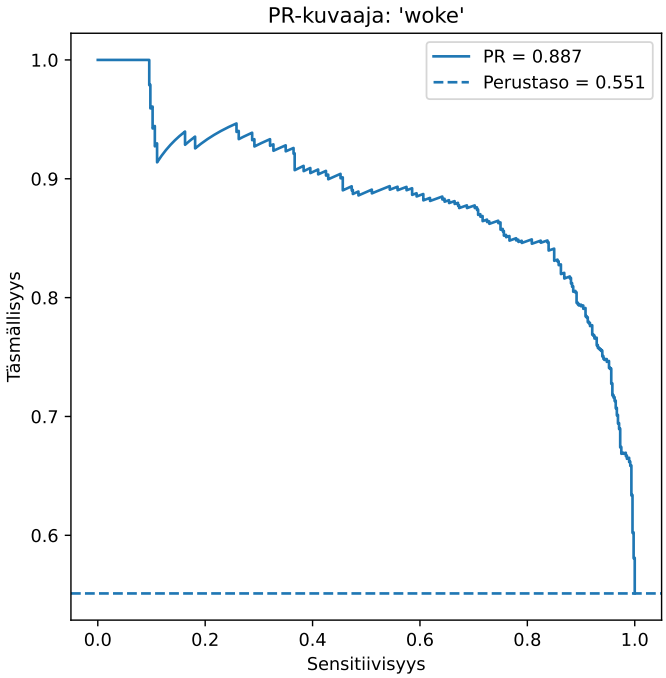
Päätuloksien osalta testidata 1 toisinsi ja siten yleisti koulutusdatalla koulutetut mallit: kaikki pääindeksit (ROC-AUC, PR-AUC, tarkkuus, tasapainotettu tarkkuus ja F1) olivat samaa luokkaa kuin koulutusdatalla (Taulukko 5). Testidatan mallien ROC-AUC ja PR-AUC kuvaajat on esitetty kuvissa 8, 9, 10 ja 11).

Taulukko 5: Testidata 1: mallien arviointi

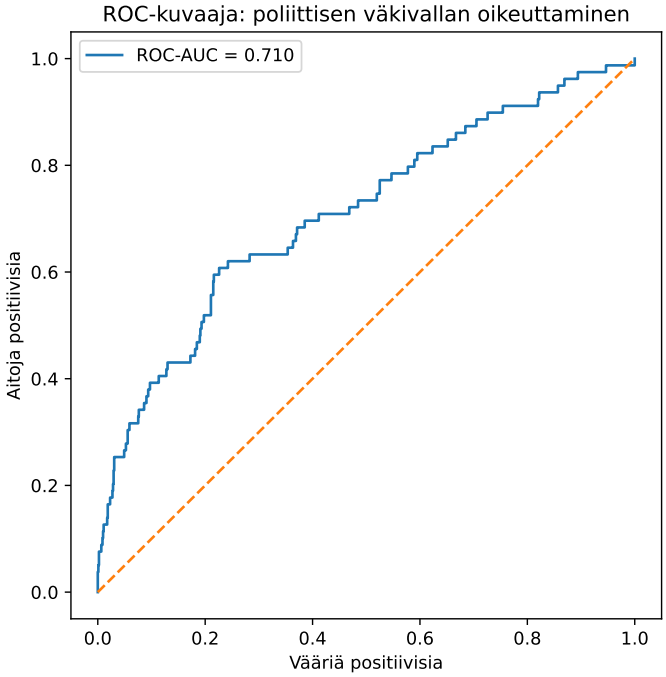
Kohde	Malli	n	ROC-AUC	PR-AUC	Tarkkuus	Tasapainotettu tarkkuus	F1
"Woke"	Tavallinen logistinen regressio	871	0,89	0,89	0,82	0,82	0,84
"Woke"	Kynnysoptimoitu logistinen regressio	871	0,89	0,89	0,82	0,81	0,84
Pol. väkivalta	Tavallinen XGBoost	995	0,71	0,27	0,92	0,53	0,10
Pol. väkivalta	Kynnysoptimoitu XGBoost	995	0,71	0,27	0,84	0,64	0,28



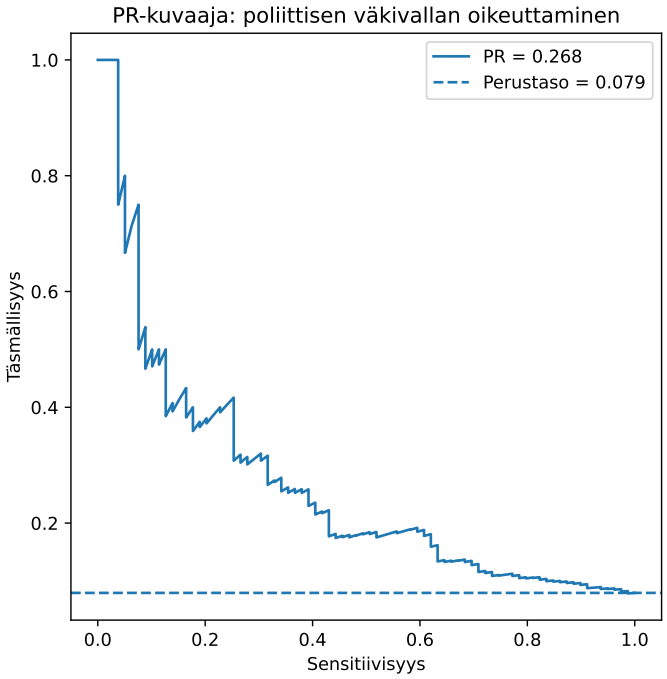
Kuva 8: ROC-kuvaaja "woke"-muuttujan ennustamiselle



Kuva 9: PR-kuvaaja “woke”-muuttujan ennustamiselle

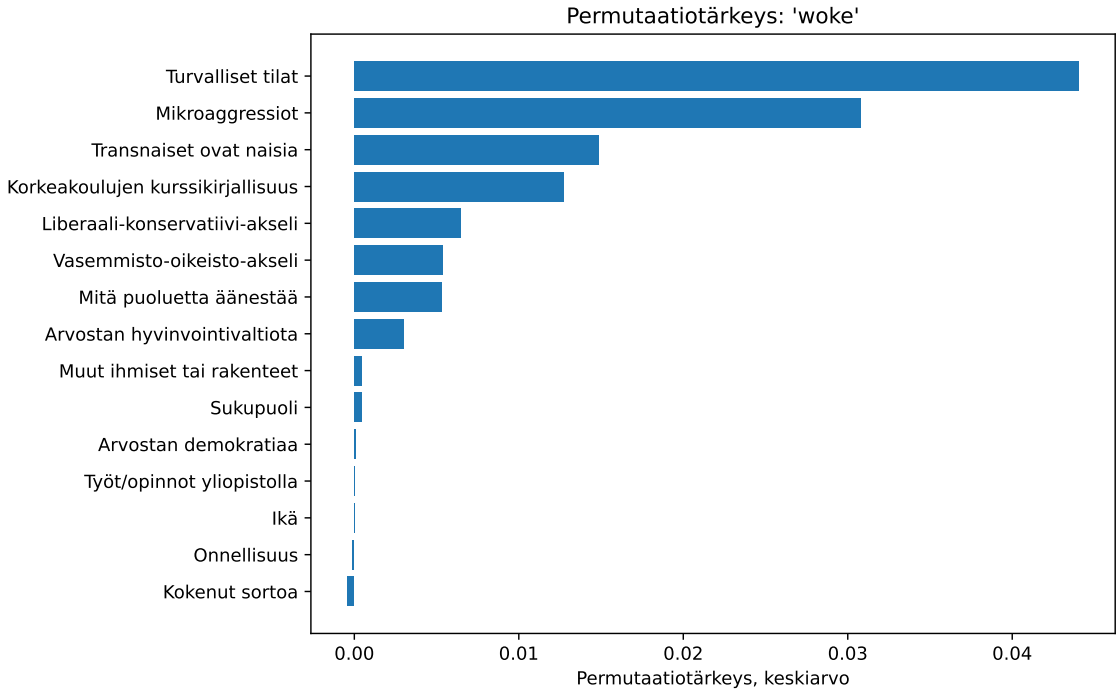


Kuva 10: ROC-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle

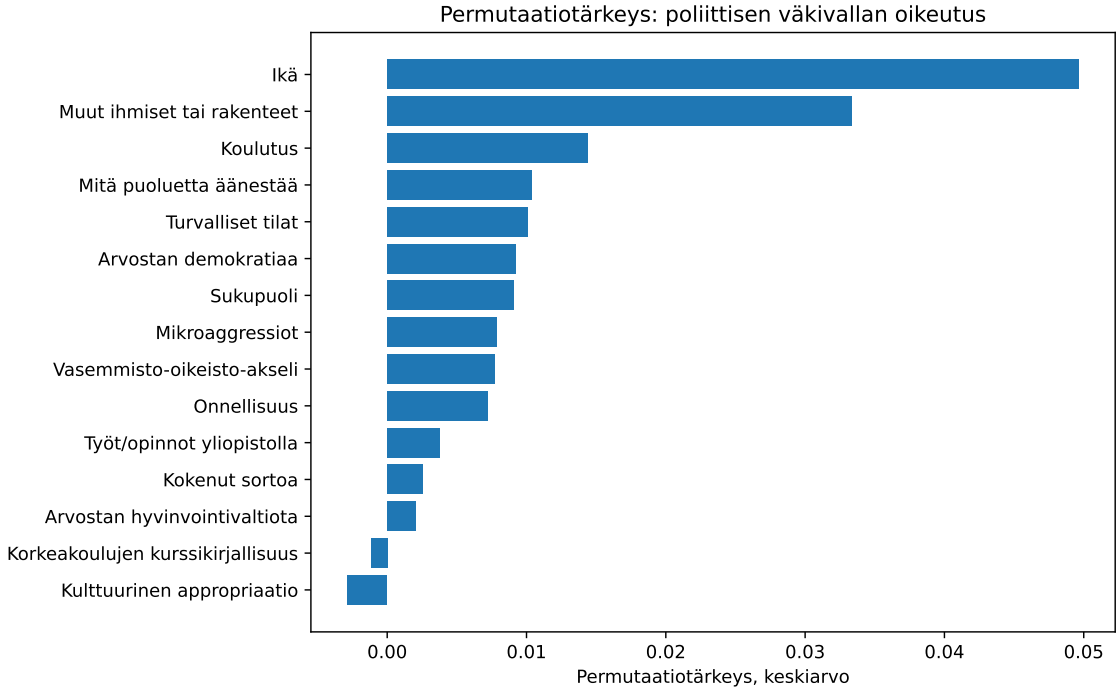


Kuva 11: PR-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle

Permutaatiotärkeyden osalta testidata 1 toisinsi ja yleisti koulutusdatan tuloksen “woke”-luokittelun ennustamisen osalta: turvalliset tilat (0,04) ja mikroaggressiot (0,03) olivat jälleen tärkeimmät selittäjät. Sen sijaan poliittisen väkivallan oikeuttamisen osalta tärkein muuttuja oli tällä kertaa ikä (0,05) ja toiseksi tärkein uskomus, että muut ihmiset tai rakenteet ovat henkilön hyvinvoinnissa vastuussa tätä itseä enemmän (0,03).



Kuva 12: “Wokea” ennustavien muuttujien permutaatiotärkeys



Kuva 13: Poliittisen väkivallan oikeuttamista ennustavien muuttujien permutaatiotärkeys

Poistokokeissa tulokset toisinsivat ja yleistivät koulutusdatan tulokset melko hyvin (Taulukot 6 ja 7).

Taulukko 6: Poistokokeet: woke

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,89	0,89	0,84
Ei kriit.	0,85	0,86	0,80
Riisuttu	0,79	0,78	0,76
Vain turv.	0,81	0,79	0,73

Taulukko 7: Poistokokeet: poliittisen väkivallan oikeuttaminen

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,71	0,27	0,10
Ei kriit.	0,73	0,27	0,09
Riisuttu	0,69	0,22	0,09
Vain turv.	0,63	0,13	0,0

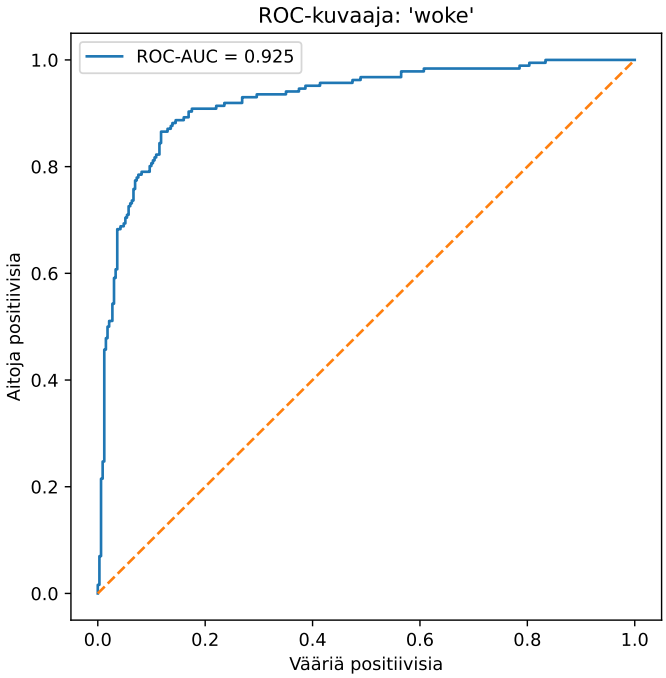
6.2 Testidata 2: tulokset

"Woke"-muuttujan osalta oli testidatassa 2 vastauksia 517, joista 64,0% (331/517) kuului luokkaan 0 ja 36,0% (186/517) luokkaan 1. Puuttuvia ja siten poistettavia vastauksia oli siis 109. Poliittisen väkivallan oikeuttamisen osalta 92,8% (539/581) kuului luokkaan 0 ja 7,2% (42/581) luokkaan 1, kun 45 arvoa puuttui.

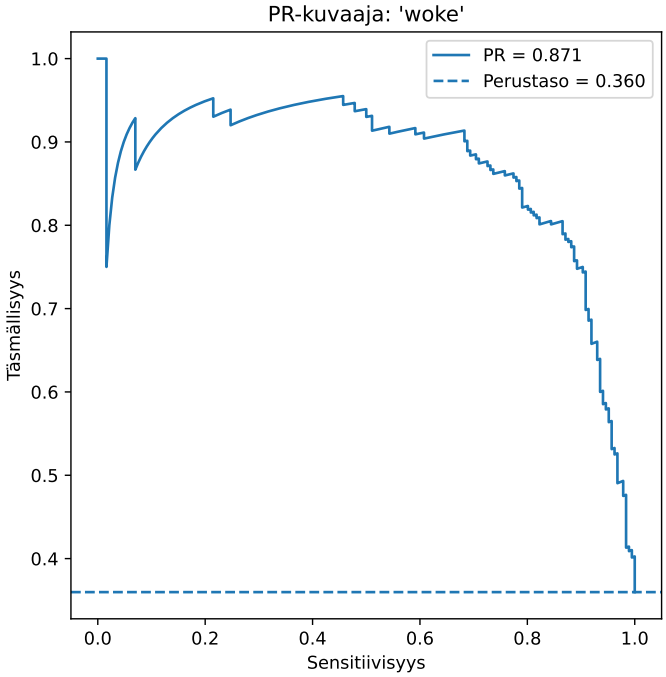
Päätuloksien osalta testidata 2 samoin toimi toisintavana ja yleistävänä näyttönä koulutusdatan malleille: pääindeksit (ROC-AUC, PR-AUC, tarkkuus, tasapainotettu tarkkuus ja F1) olivat jälleen melko samanlaisia kuin koulutusdatalla (Taulukko 8). Testidatan mallien ROC-AUC ja PR-AUC kuvaajat on esitetty kuvissa 14, 15, 16 ja 17).

Taulukko 8: Testidata 2: mallien arviointi

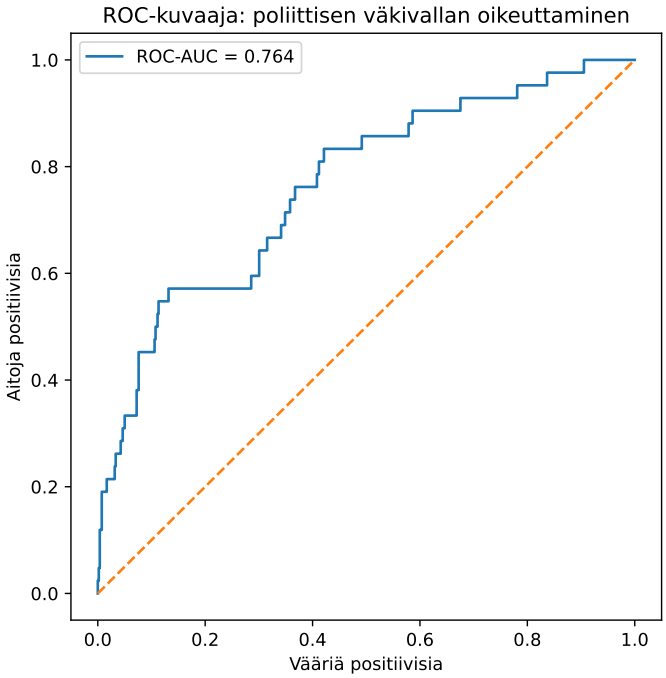
Kohde	Malli	n	ROC-AUC	PR-AUC	Tarkkuus	Tasapainotettu tarkkuus	F1
"Woke"	Tavallinen logistinen regressio	517	0,93	0,87	0,87	0,84	0,81
"Woke"	Kynnysoptimoitu logistinen regressio	517	0,93	0,87	0,87	0,85	0,81
Pol. väkivalta	Tavallinen XGBoost	581	0,76	0,31	0,92	0,59	0,27
Pol. väkivalta	Kynnysoptimoitu XGBoost	581	0,76	0,31	0,63	0,69	0,23



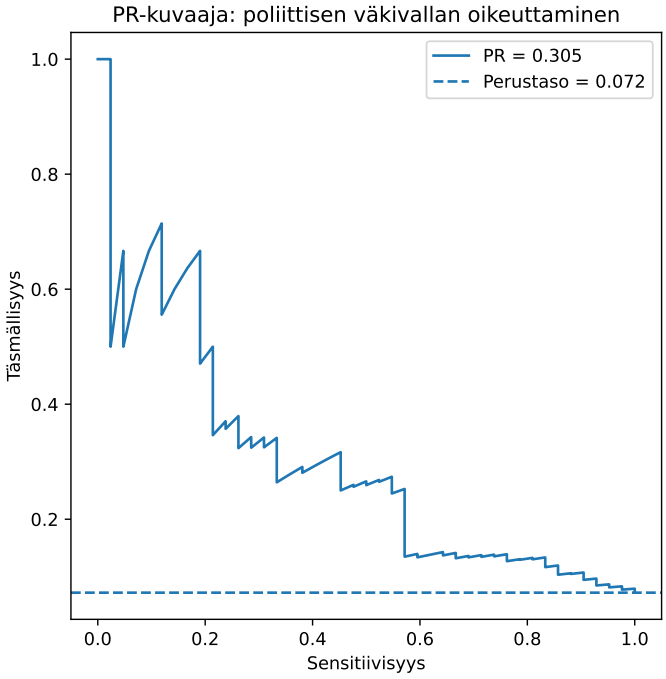
Kuva 14: ROC-kuvaaja “woke”-muuttujan ennustamiselle



Kuva 15: PR-kuvaaja “woke”-muuttujan ennustamiselle



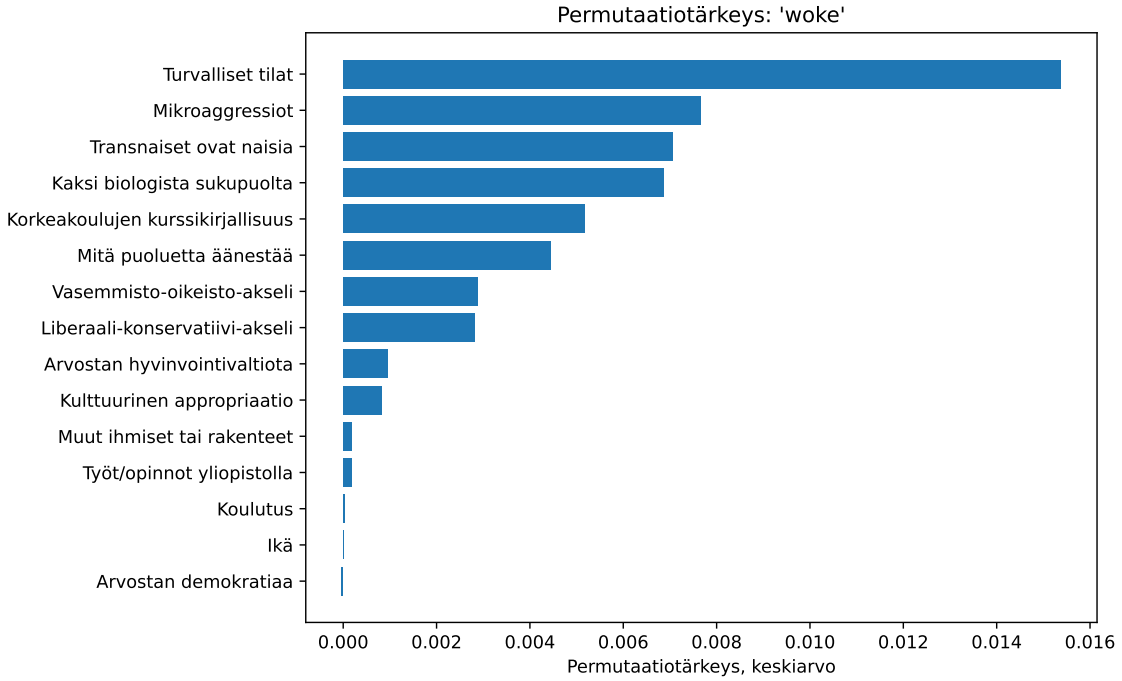
Kuva 16: ROC-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle



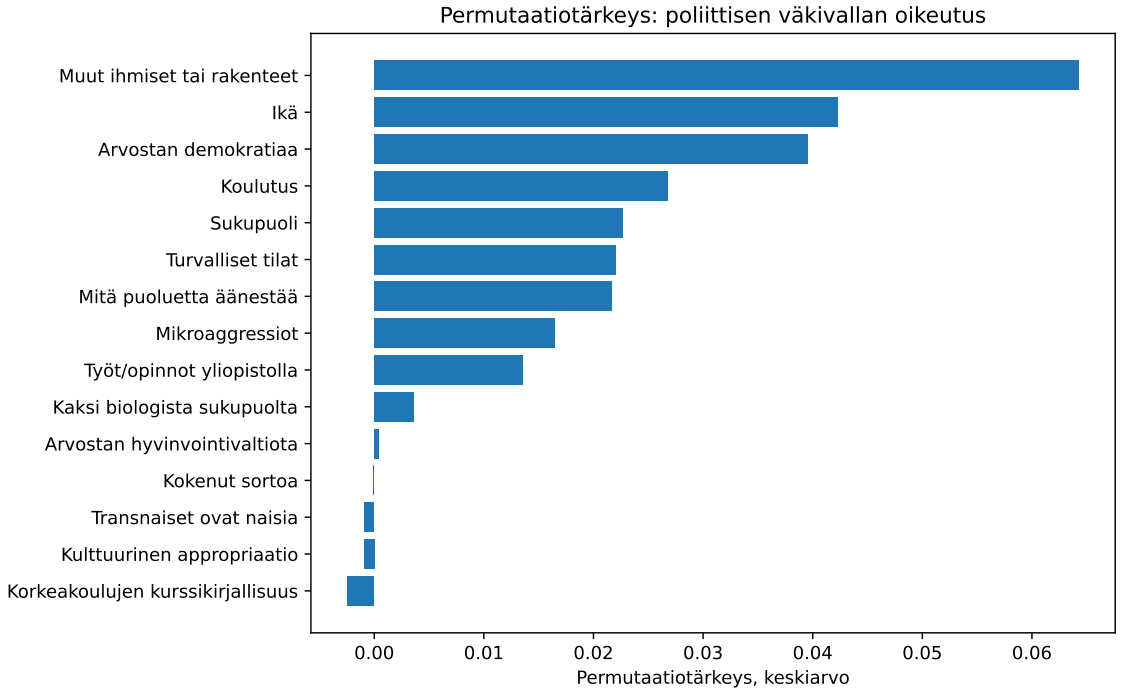
Kuva 17: PR-kuvaaja poliittisen väkivallan oikeuttamisen muuttujan ennustamiselle

Katsottaessa permutaatiotärkeyttä havaittiin, että myös testidata 2 toisinsi ja yleisti koulutusdatan tuloksen “woke”-luokittelun ennustamisen osalta, joskin turvallisten tilojen permutaatiotärkeyden arvo (0,02) oli hieman matalampi kuin muissa

tutkimuksen aineistoissa. Poliittisen väkivallan oikeuttamisen osalta tärkeimpiä selittäjiä olivat uskomus, että muut ihmiset tai rakenteet ovat henkilön hyvinvoinnissa vastuussa tätä itseä enemmän (0,06), ikä (0,04) ja demokratian arvostus (0,04).



Kuva 18: "Wokea" ennustavien muuttujien permutaatiotärkeys



Kuva 19: Poliittisen väkivallan oikeuttamista ennustavien muuttujien permutaatiotärkeys

Poistokokeissa tulokset toisinsivat ja yleistivät koulutusdatan tulokset melko hyvin (Taulukot 9 ja 10).

Taulukko 9: Poistokokeet: woke

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,93	0,87	0,81
Ei kriit.	0,91	0,86	0,79
Riisuttu	0,85	0,81	0,72
Vain turv.	0,84	0,71	0,70

Taulukko 10: Poistokokeet: poliittisen väkivallan oikeuttaminen

Malli	ROC-AUC	PR-AUC	F1
Täysi	0,76	0,31	0,27
Ei kriit.	0,76	0,29	0,25
Riisuttu	0,65	0,12	0,0
Vain turv.	0,59	0,11	0,0

7 Yhteenveto

Tutkielmassa tarkasteltiin koneoppimismallien käyttöä poliittisten asenteiden binääriluokittelussa. Luokittelukohteena oli kaksi poliittista asennetta: kuinka "woke" henkilö koki olevansa ja missä määrin hän koki poliittisen väkivallan oikeutetuksi. Tehtävää varten koulutettiin kolmenlaisia koneoppimismalleja: logistiseen regressioon, gradienttitehostukseen, sekä XGBoostiin perustuvia malleja. Koulutusdatassa näistä "woke"-mallille parhaita ennustuksia tekeväksi valittiin logistiseen regressioon perustuva malli ja poliittisen väkivallan oikeuttamiselle XGBoostiin perustuva malli, joskin erot mallien välillä olivat pieniä ja valinta niiden välillä jossain määrin mielivaltainen.

Tutkielmassa käytettiin kolmea tutkielman tekijän keräämää aineistoa vuosilta 2022, 2025 ja 2026, joissa oli yhteensä 6626 validia vastausta. Ensimmäistä, ja suurinta, aineistoa käytettiin mallien koulutuksen ristiinvalidointimenettelyllä ja kahta pienempää aineistoa käytettiin koulutetun mallin testaamiseen. Malleja testatessa huomattiin niiden toisintuvan testausaineistoissa hyvin. Erityisenä lisähuomiona tutkimuksessa havaittiin, että turvallisiin tiloihin liittyvä uskomus riitti yksin koulutusaineistoksi mallille, jonka suoriutuminen oli vain kohtalaisesti heikompaa kuin täyden, 18 selittäjää käyttävän mallin. Tutkimuksen voidaan katsoa onnistuneen hyvin ja tuottaneen, moderneilla menetelmillä, uutta tietoa poliittisten asenteiden tutkimukseen. Turvallisten tilojen uskomukseen liittyvä tulos on uusi ja tutkimuksellisesti kiinnostava.

Lopuksi on hyvä todeta joitakin tutkimuksen rajoitteita. Ensinnäkin malli ei kyennyt kovin hyvin luokittelemaan poliittisen väkivallan oikeuttamista. Vuosien 2022 ja 2025 välillä vaikutti tutkimuksen aineistojen valossa tapahtuneen muutos poliittisen väkivallan oikeuttamisessa siten, että vasemmistossa kannatus pysyi samana ja oikeistossa laski. Näin vuoden 2022 aineistoon koulutettu malli ei välttämättä kyennyt optimaalisesti ennustamaan vuosien 2025 ja 2026 asenteita. Jatkotutkimuksessa saattaisi olla hyödyllistä kouluttaa mallit tuoreeseen aineistoon ja tarkastella, olisiko poliittisen väkivallan oikeuttamisen ennustaminen tarkempaa nykyistä suhdannetta kuvaavilla malleilla.

Toiseksi, kaikki muuttujat eivät sopineet mitta-asteikoltaan täydellisesti tämän tutkielman tarkasteluihin. Esimerkiksi puoluekanta oli kyllä koodattu ordinaalisesti vasemmistoliberaaleista puolueista oikeistokonservatiivisiin, mutta erot puolueiden välillä eivät vastaa täydellisesti kokonaislukuja. Suurin osa muuttujista sopi tarkasteluihin hyvin, mutta neljään muuttujaan (sukupuoli, koulutus, "äänestän tyypillisesti"ja "opiskelen/työskentelen yliopistolla") oli jäänyt pienehköjä luokkia, kuten "muu", "en äänestä", "työskentelen korkeakoulussa, en opetus- tai tutkimushenkilökuntaa", joilla ei ollut selvää järjestystä. Tämä todennäköisesti lisäsi virhettä näiden muuttujien ennustusvoimaan. Ordinaalisia ja käytännössä ordinaalisia muuttujia haluttiin käyttää tutkielmassa sellaisinaan järjestystiedon säilyttämiseksi. One hot encoding -tyyppisten binäärimuuttujien käyttö olisi kadottanut järjestysinformaatiota ja monimutkaistanut selittäjiä.

Kolmanneksi, henkilön "woke"-itsearvion ennustaminen kriittisen sosiaalisen oikeudenmukaisuuden mittarin kysymyksillä sisältää mahdollisen tautologian riskin. Mitkään kysymyksistä eivät kuitenkaan olleet yksi yhteen samoja "woke"-itsearvion

kanssa ja mukana oli vain osa kokonaisen mittarin kysymyksistä: 4/7 kysymystä CSJAS-mittarista ja 2/6 uusimmasta CSJAS-R-mittarista.[17][18] Tautologisuuden riski huomioitiin myös poistokokeissa, joissa havaittiin, että malli toimi kohtalaisesti myös ilman kriittisen sosiaalisen oikeudenmukaisuuden uskomuksia.

Kaiken kaikkiaan tutkimusta voidaan pitää onnistuneena ja uutta tavoittelevana: tutkielman tekijä ei ole tietoinen, että kriittisen sosiaalisen oikeudenmukaisuuden tai poliittisen väkivallan oikeuttamisen asenteita olisi aiemmin tutkittu koneoppimismenetelmillä. Suurten aineistojen ja koneoppimismenetelmien käyttö mahdollistaa kattavien analyysien tekemisen ja yllättävienkin tulosten löytymisen. Tätä tutkielmaa voidaankin pitää ensimmäisenä askeleena mainittuun tutkimussuuntaan.

Viitteet

[1] Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning. Cambridge University Press.

[2] Duda, R. O., Hart, P. E., & Stork, D. (2001). Pattern classification. John Wiley.

[3] Alpaydin, E. Introduction to Machine Learning. 4. painos. MIT Press, 2020.

[4] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. MIT press.

[5] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: with applications in R. New York: springer.

[6] Christen, P., Hand, D. J., & Kirielle, N. (2023). A review of the F-measure: its history, properties, criticism, and alternatives. ACM Computing Surveys, 56(3), 1-24.

[7] Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.

[8] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PloS One, 10(3), e0118432.

[9] Cramer, J. S. (2002). The origins of logistic regression (No. 02-119/4). Tinbergen Institute discussion paper.

[10] Dey, D., Haque, M. S., Islam, M. M., Aishi, U. I., Shammy, S. S., Mayen, M. S. A., ... & Uddin, M. J. (2025). The proper application of logistic regression model in complex survey data: a systematic review. BMC Medical Research Methodology, 25(1), 15.

[11] Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. Mathematical programming, 45(1), 503-528.

[12] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning. 2nd ed. Springer.

[13] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

[14] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[15] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[16] Meyes, R., Lu, M., De Puiseau, C. W., & Meisen, T. (2019). Ablation studies in artificial neural networks. arXiv preprint arXiv:1901.08644.

[17] Lahtinen, O. (2024). Construction and validation of a scale for assessing critical social justice attitudes. Scandinavian Journal of Psychology, 65(4), 693-705.

[18] Lahtinen, O. (2025). Two Kinds of "Woke"? Psychometric Validation of the Critical Right Scale and Revised Critical Social Justice Attitudes Scale. Scandinavian Journal of Psychology.

[19] Jadhav, A., Pramod, D., & Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. Applied Artificial Intelligence, 33(10), 913-933.

[20] Kang, H. (2013). The prevention and handling of the missing data. Korean Journal of Anesthesiology, 64(5), 402-406.

[21] Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. Journal of Big data, 8(1), 140.

[22] Pinheiro, J. M. H., de Oliveira, S. V. B., Silva, T. H. S., Saraiva, P. A. R., de Souza, E. F., Godoy, R. V., ... & Becker, M. (2025). The impact of feature scaling in machine learning: Effects on regression and classification tasks. IEEE Access, 13, 199903-199931.