Algoritmisen perustan ja tekoälyn etiikka: teoriasta käytäntöön -kurssi

Työkalupakki synkronisia **istuntoja** varten

CU5 | Tapaustutkimukset ja hankkeet

Avoimen vastauksen formatiivinen arviointi

Hankkeen numero | 2022-1-ES01-KA220-HED-000085257

Avoimen vastauksen formatiivinen arviointi

Avoimen vastauksen formatiivinen arviointi

	Kuvaus	Kommentit
Tehtävän kuvaus	Tutustuttaa opiskelijat IBM:n Fairness 360 -työkaluun, jonka tavoitteena on antaa tietoa mahdollisista väestörakenteisiin liittyvistä ennakkoluuloista, niiden mittaamisesta ja strategioista, joilla ennakkoluuloja voidaan lieventää, jos niitä havaitaan. Esimerkkinä käytetään Compas Propublica -sovellusta.	
Kuvaus tehtävän suorittamisesta	AI Fairness 360 - Demo (ibm.com). Valitse Compas (ProPublica recidivism) -työkalu demoa varten. Vastaa seuraaviin kysymyksiin: 1) Mihin tarkoitukseen Compasin algoritmia käytetään? 2) Millaisia ennakkoluuloja työkalussa on havaittu? 3) Mitkä ovat suojatut ominaisuudet? 4) Missä tilastollisissa mittauksissa on tässä tapauksessa havaittavissa harhaa molempien ominaisuuksien osalta? 5) Kuinka paljon keskimääräinen odds-ero muuttuu "uudelleenpunninta-algoritmin" tapauksessa molemmissa ominaisuuksissa? 6) Miksi uudelleenpunnitus on tarpeen ennen sovelluksen käyttöönottoa?	Lisätietoja ja ohjeet löytyvät seuraavalta sivulta.
Arvioitu aika tehtävän tekemiseen	40- 60 minuuttia.	
Ehdotus lähteistä tehtävän tekemistä varten	AI Fairness 360 -työkalu tarjoaa kaikki vastaukset, lukuun ottamatta numeroita 2 ja 6. Voit etsiä vastauksen vapaasti Internetistä.	
Yksityiskohtainen kuvaus tehtävän suorittamisesta	Opettajan tulisi selittää, mihin tehtävä palautetaan (esim. sähköpostitse, tiettyyn kansioon, joka on jaettu opiskelijalle aiemmin, verkko-oppimisalustan kurssirakenteeseen luotuun alueeseen...).	
Tieto tehtävän toimittamisen määräajasta	Opettajan on asetettava määräaika tehtävän toimittamiselle (huomioi kurssin rakenne asynkronisen työskentelyn ja synkronisten istuntojen osalta).	Ilmoita päivämäärä esittelyistunnossa.
Yhteystiedot tai epäselvyyksien selvittäminen	Opettajan on annettava yhteystieto.	Se voi olla sähköpostiosoite, puhelinnumero...

Avoimen vastauksen formatiivinen arviointi

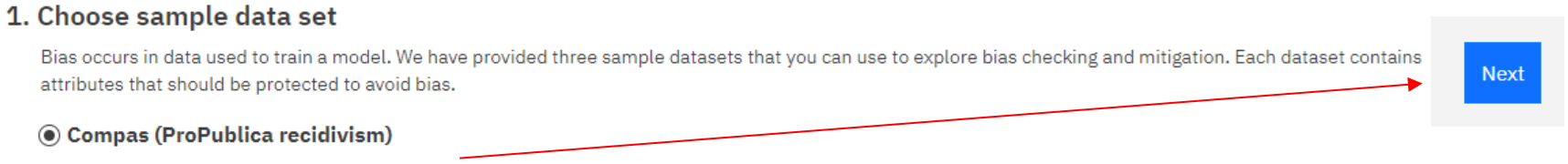
Ohjeet

IBM:n avoimen lähdekoodin AI Fairness 360 -työkalupaketin avulla voit tutkia, raportoida ja lieventää koneoppimismalleissa esiintyvää syrjintää ja ennakkoluuloja koko tekoälysovelluksen elinkaaren ajan. Tarkempia tietoja saat kotisivuilta: https://aif360.res.ibm.com/.

Tässä tehtävässä tarkastelemme yhtä tapausta web-demossa.
Avaa demotyökalu tämän linkin kautta: AI Fairness 360 - Demo (ibm.com).

Tässä tapauksessa käytämme esimerkkinä Compasia (ProPublica recidivism).

1) Mihin tarkoitukseen Compas-algoritmia käytetään? Se mainitaan suoraan työkalussa.
2) Millaisia harhoja Compas Probulica -ratkaisussa on havaittu? Etsi vastaus internetistä.
3) Mitkä ovat suojatut ominaisuudet ratkaisussa? Se näkyy suoraan työkalussa.
4) Mitkä tilastolliset mittaukset osoittavat tässä tapauksessa harhaa molempien attribuuttien osalta? Valitse työkalu klikkaamalla sitä ja ota seuraava -> näet vastauksen seuraavalla sivulla.



5) Kuinka paljon keskimääräinen odds-ero muuttuu "uudelleenpunninta-algoritmin" tapauksessa molemmissa attribuuteissa? Näet vastauksen tähän, kun otat seuraavan sivun ja painat "reweighing" -vaihtoehtoa.
6) Miksi uudelleenpunnitus on tarpeen ennen sovelluksen käyttöönottoa? Katso esim. tästä vastaus Reweighting: Maxine | Medium).

Avoimen vastauksen formatiivinen arviointi

Formatiivisen arvioinnin oikeat vastaukset

Voit kerätä vastauksia sinulle parhaiten sopivalla tavalla. Kaikkien muiden kysymysten paitsi kysymysten 2 ja 6 vastaukset ovat suoraan työkalusta.

1. Mihin tarkoitukseen Compas-algoritmia käytetään?

Rikosoikeudellinen tuomitseminen.

2. Millaisia ennakkoluuloja työkalussa on havaittu?

Mustien syytettyjen riski joutua uusintarikoksen uhriksi oli usein ennustettu suuremmaksi kuin he todellisuudessa olivat.
Valkoisten syytettyjen riskin ennustettiin usein olevan pienempi kuin he olivat.
Analyysi osoitti myös, että vaikka aiemmat rikokset, tuleva uusintarikollisuus, ikä ja sukupuoli otettiin huomioon, mustille vastaajille annettiin 45 prosenttia todennäköisemmin korkeammat riskipisteet kuin valkoisille vastaajille.

Mustat vastaajat luokiteltiin myös kaksi kertaa todennäköisemmin kuin valkoiset vastaajat virheellisesti suuremmaksi väkivaltaisen uusintarikollisuuden riskiksi. Väkivaltaisen uusintarikollisuuden analyysi osoitti myös, että vaikka aiemmat rikokset, tuleva uusintarikollisuus, ikä ja sukupuoli otettiin huomioon, mustat vastaajat saivat 77 prosenttia todennäköisemmin korkeamman riskipisteytyksen kuin valkoiset vastaajat. (lähde Miten analysoimme COMPAS-ohjelman rikoksen uusintarikollisuuden algoritmin - ProPublica).

3. Mitkä ovat suojatut ominaisuudet?

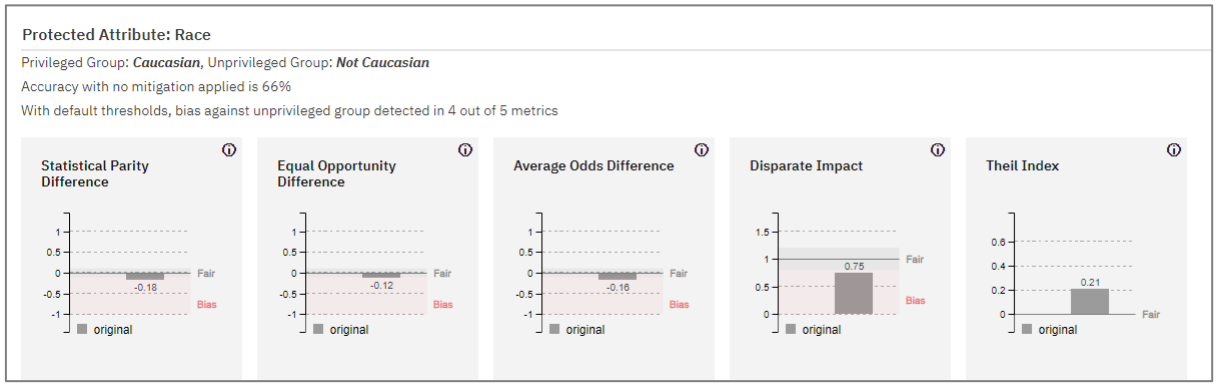
(Vastaus suoraan työkalusta)
Sukupuoli, etuoikeutettu: *Nainen*, etuoikeudeton: *Mies*
Rotu, etuoikeutettu: Etuoikeutettu: *Valkoihoinen*, etuoikeudeton: *Ei valkoihoinen*

Avoimen vastauksen formatiivinen arviointi

Formatiivisen arvioinnin oikeat vastaukset

4. Mitkä tilastolliset mittaukset osoittavat tässä tapauksessa harhaa molempien ominaisuuksien osalta?

Vastaus näkyy alla olevista kuvista (työkalusta) eli molempien ominaisuuksien osalta tilastollinen tasavertaisuusero, yhtäläisten mahdollisuuksien ero, keskimääräinen todennäköisyysero ja eriarvoinen vaikutus osoittavat kaikki puolueellisuutta.

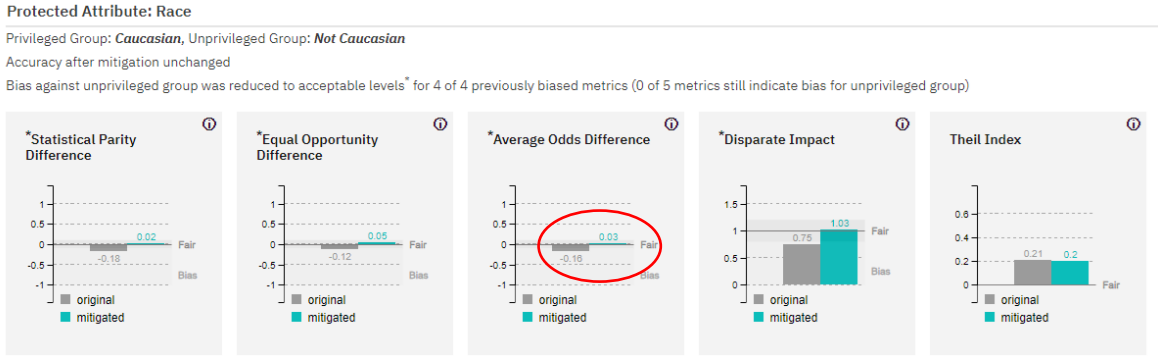


Avoimen vastauksen formatiivinen arviointi

Formatiivisen arvioinnin oikeat vastaukset

5. Kuinka paljon keskimääräinen kertoimien ero muuttuu, jos käytetään "uudelleenpunninta-algoritmia" molemmissa ominaisuuksissa?

Sukupuoli: 0,33 yksikköä
Rotu: 0,19 yksikköä



6) Miksi uudelleenpunnitus on tarpeen ennen sovelluksen käyttöönottoa.

Katso linkki tästä [Uudelleen punnitus: Maxine | Medium | Medium.]

Lyhyesti: sen avulla algoritmi voidaan kalibroida uudelleen ilman, että käytetään lisää dataa.

KIITOS

Hankkeen numero: 2022-1-ES01-KA220-HED-000085257.

Co-funded by the European Union

Euroopankomission tuki tämän julkaisun tuottamiselle ei tarkoita sitä, että sen sisältö vastaa ainoastaan sen kirjoittajien näkemyksiä, eikä komissio ole vastuussa sen sisältämien tietojen mahdollisesta käytöstä.