UNIVERSITÉ TOULOUSE CAPITOLE

IRIT
Institut de Recherche en Informatique de Toulouse

Fairness in AI: navigating ethical challenges

José Gonzalo Domínguez - jose.dominguez@ut-capitole.fr

1. Definition

2. Criteria

3. Practical application

4. References

1. Definition

Fairness: the quality of treating people equally or in a way that is right or reasonable.

Bias: the action of supporting or opposing a particular person or thing in an unfair way, because of allowing personal opinions to influence your judgment

Cambridge Dictionary

Why is this an issue?

Not all biases are unjust, but the term is most often used to indicate an unfair advantage or disadvantage for a certain group of people.

Author: gonzalodom11 et al: Why we should trust AI?, Github, version 2024/04/20

2. Criteria

2. Criteria

We define whether an AI system is unfair in terms of its impacts on people:

➔ Fairness-related harms, including harms involving people's individual experience with AI.

Fairlearn:A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.

Examples

AI systems can unfairly allocate opportunities, resources, or information.

Instance 1: AI CV-screening system trained using the CVs of people currently employed in the tech industry, where women are already underrepresented, may conceal employment opportunities from women.

Fairlearn:A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.

Examples

AI systems can fail to provide the same quality of service to some people as they do to others.

Instance 1: a facial recognition system that has technical issues recognizing faces with particular skin tones will not provide the same quality service to every user.

Fairlearn:A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.

Examples

AI systems can under represent groups of people, or treat them as if they don't exist.

Instance 1: in response to the query "chief executive officer", a search system that mainly returns images of men in response may underrepresent non-male executive officers.

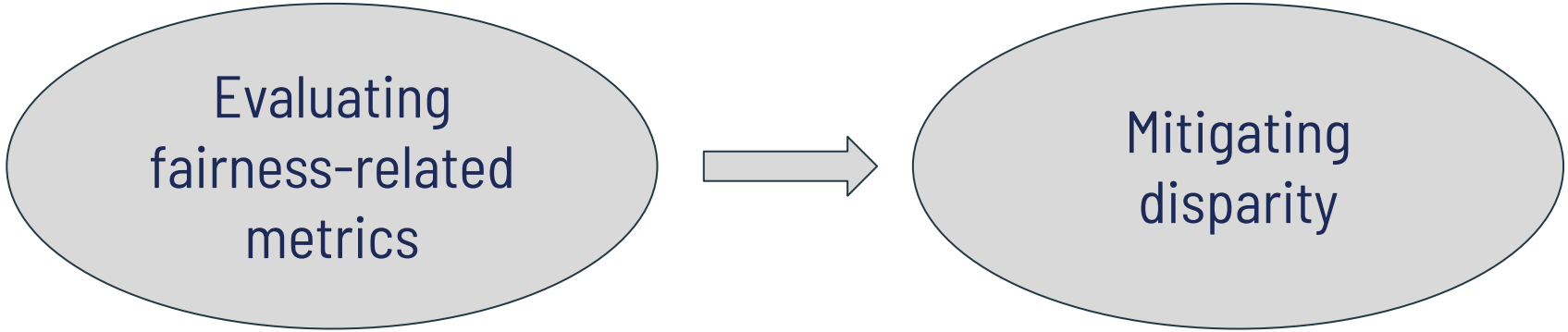
Fairlearn:A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.

3. Practical Application

3. Fairlearn

It is an open-source, community-driven project to help data scientists improve fairness of AI systems.

Funded by: Microsoft TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY



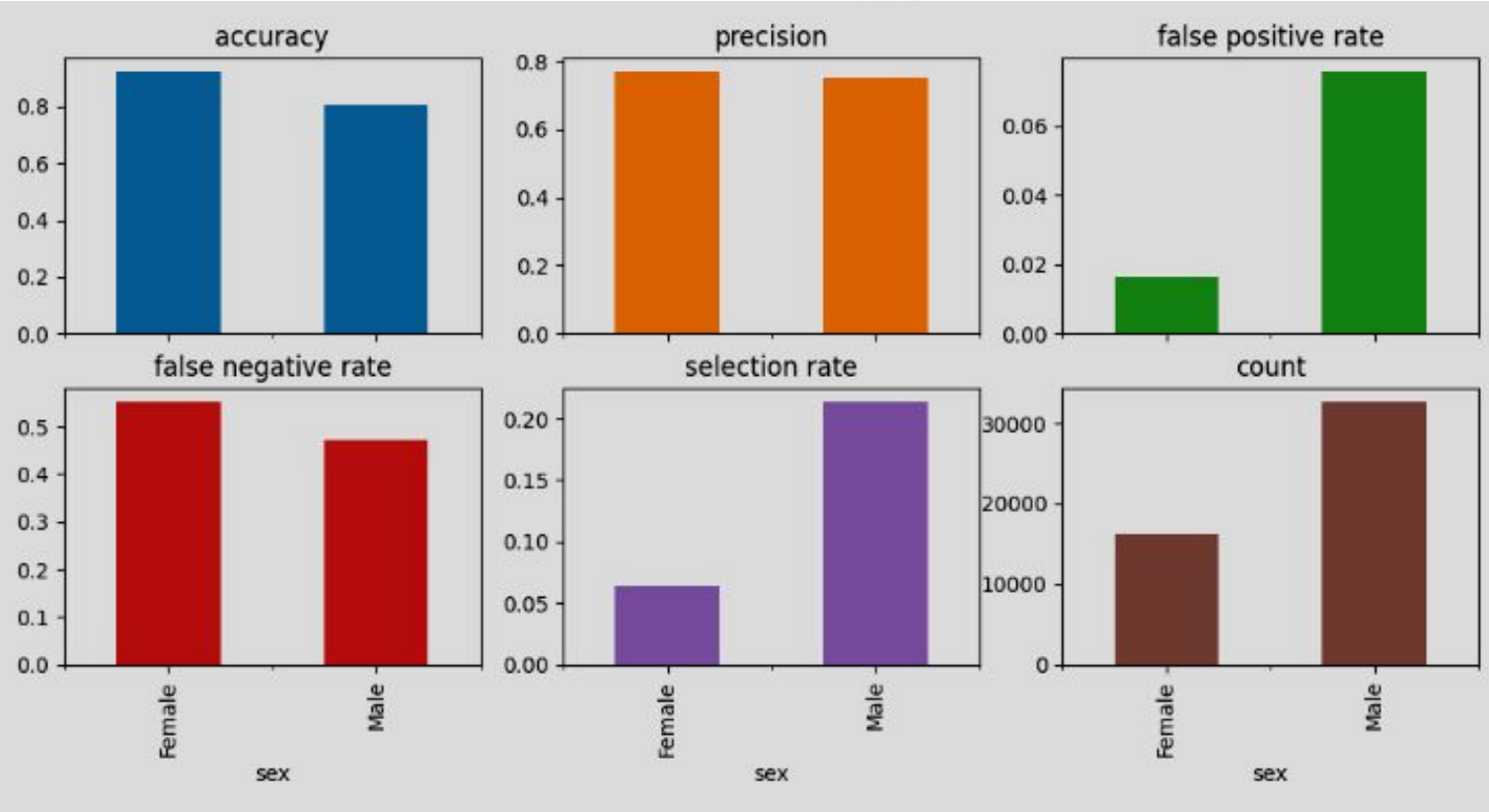
Fairlearn Homepage: https://fairlearn.org/, last updated 2024

Fairlearn

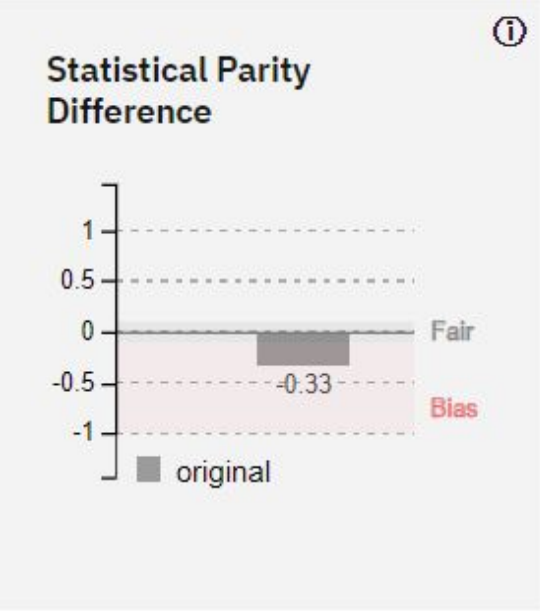
```
data = fetch_adult()

X = data.data
y_true = (data.target == ">50K") * 1
sex = X["sex"]

selection_rates = MetricFrame(
    metrics=selection_rate, y_true=y_true, y_pred=y_true, sensitive_features=sex
)

fig = selection_rates.by_group

fig.plot.bar(
    legend=False, rot=0, title="Fraction earning over $50,000"
)

plt.show()
```

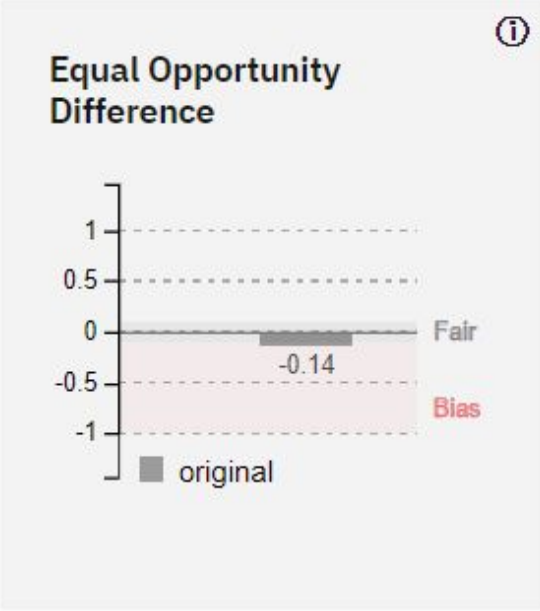
Evaluating fairness-related metrics

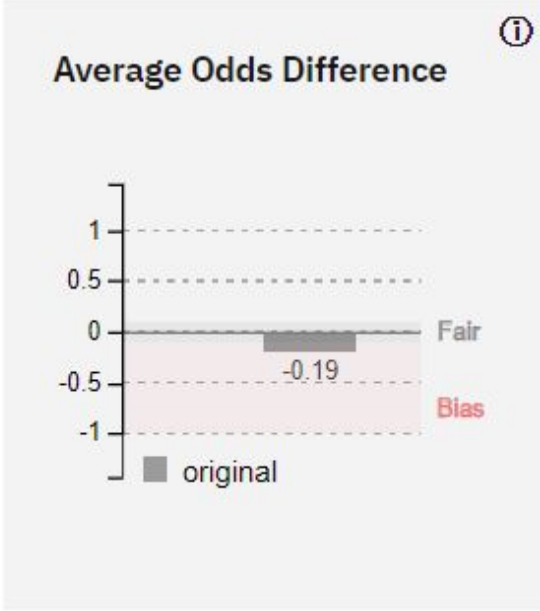
Fairlearn


accuracy
precision
false positive rate
false negative rate
selection rate
count
0.0
0.2
0.4
0.6
0.8
0.00
0.02
0.04
0.06
0.1
0.3
0.5
0.05
0.10
0.15
0.20
0
10000
20000
30000
Female
Male
sex

AI Fairness 360







Ai Fairness 360 Homepage: https://aif360.res.ibm.com/, last updated 04/2024

References

- Fairlearn:A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.
- Fairlearn Homepage: https://fairlearn.org/, last updated 2024
- Author: gonzalodom11 et al: Why we should trust AI?, Github, version 2024/04/20
- Ai Fairness 360 Homepage: https://aif360.res.ibm.com/, last updated 2024/04

Recommendations to read:

- Fairlearn: A toolkit for assessing and improving fairness in AI, Microsoft, Microsoft, version 2020/09/22.
- Caltech. (n.d.). Can we trust artificial intelligence? Retrieved from https://scienceexchange.caltech.edu/topics/artificial-intelligence-research/
- Ruf, B., & Detyniecki, M. (2021, September 30). Towards the right kind of fairness in ai. arXiv.org. https://arxiv.org/abs/2102.08453.


Thank
You