Algorithmic foundations and ethics in AI: from theory to practice course

Toolkit for synchronous sessions

CU5 | Case studies and projects

Open answer formative assessment

Project number | 2022-1-ES01-KA220-HED-000085257

Open answer formative assessment

Open answer formative assessment

	Description	Comments
Task description	Introduces students to IBM's Fairness 360 tool, aiming to provide insight into potential biases within demographics, their measurement, and strategies for mitigation if biases are identified. We use Compas Propublica application as an example.	
Description of how to do the task	AI Fairness 360 - Demo (ibm.com) Choose Compas (ProPublica recidivism) tool for demo. Answer for the following questions: 1) For what purpose is Compas algorithm used for? 2) What kind of biases has been noticed in the tool? 3) What are the protected attributes? 4) Which statistical measurements show bias in this case for both attributes? 5) How much does the average odds difference change in case of «reweighing algorithm» in both attributes? 6) Why is reweighing necessary before bringing the application into use?	More details and instructions can be found from the next page.
Estimated time to do the task	40- 60 minutes.	
Suggestion of sources for doing the task	AI Fairness 360 tool provides all the answers, except for number 2 and 6, you can freely search internet for the answer.	
Detailed description of how to deliver the task	The teacher should explain on where to return the assignment (e.g. by email, in a certain folder previously shared with the student, in an area created in the course structure on the e-learning platform...).	
Information on the deadline for the task delivery	The teacher should set a deadline for the submission of this assignment (please note the structure of the course in terms of asynchronous work and synchronous sessions).	Give the date in the introduction session.
Contact information or how to clarify doubts	The teacher must provide a form of contact.	It could be an email address, a telephone number...

Open answer formative assessment

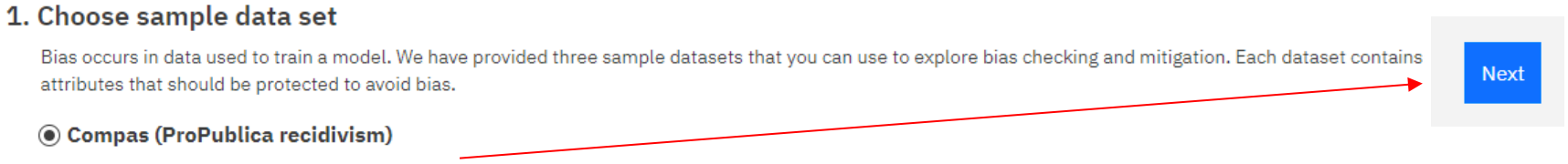
Instructions

IBM's AI Fairness 360 open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. Check the homepage for detailed information: https://aif360.res.ibm.com/

In this task, we'll examine one case in the web demo.
Open the demo tool via this link: AI Fairness 360 - Demo (ibm.com)

In this case, we use the Compas (ProPublica recidivism) as an example.

1) For what purpose is Compas algorithm used for? ⓘ It is mentioned directly in the tool.
2) What kind of biases has been noticed in the Compas Probulica solutionl? ⓘ Search the internet for an answer.
3) What are the protected attributes in the solution? ⓘ It can be seen directly in the tool.
4) Which statistical measurements show bias in this case for both attributes? ⓘ Select the tool by clicking it and take next -> you can see the answer in the next page.



5) How much does the average odds difference change in case of «reweighing algorithm» in both attributes? ⓘ You can see the answer for this, when taking the next page and pressing the "reweighing" option.
6) Why is reweighing necessary before bringing the application into use? ⓘ See eg. this for answer Reweighting: Refining AI with Precision and Efficiency | by maxine | Medium)

Open answer formative assessment

Correct answers for the formative assessment

You can collect answers the way that suits you the best. Answers are directly from the tool itself for all other questions except for 2 and 6.

1. For what purpose is Compas algorithm used for?

Criminal justice sentencing.

2. What kind of biases has been noticed in the tool?

Black defendants were often predicted to be at a higher risk of recidivism than they actually were.
White defendants were often predicted to be less risky than they were.
The analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 45 percent more likely to be assigned higher risk scores than white defendants.

Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism. The violent recidivism analysis also showed that even when controlling for prior crimes, future recidivism, age, and gender, black defendants were 77 percent more likely to be assigned higher risk scores than white defendants. (source How We Analyzed the COMPAS Recidivism Algorithm — ProPublica)

3. What are the protected attributes?

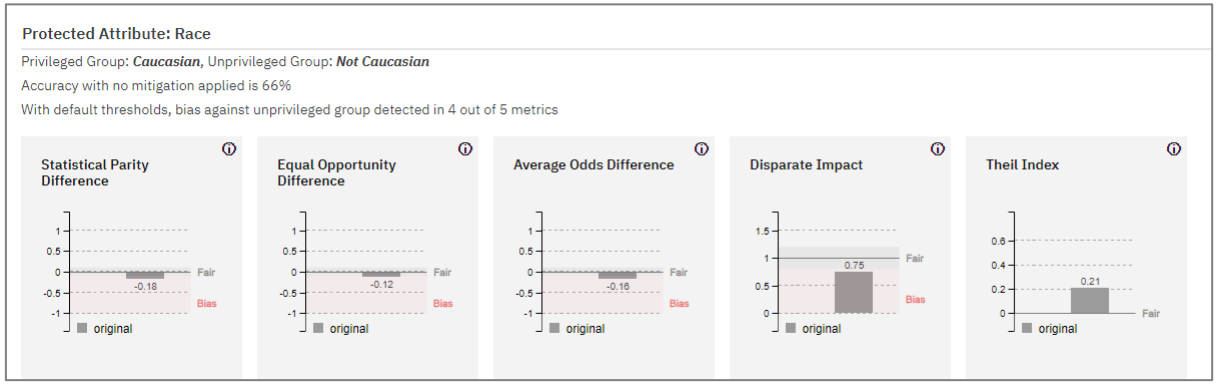
(Answer from the tool directly)
Sex, privileged: ***Female***, unprivileged: ***Male***
Race, privileged: ***Caucasian***, unprivileged: ***Not Caucasian***

Open answer formative assessment

Correct answers for the formative assessment

4. Which statistical measurements show bias in this case for both attributes?

Answer seen from the pictures below (from the tool) ie. for both attributes statistical parity difference, equal opportunity difference, average odds difference and disparate impact all show bias.

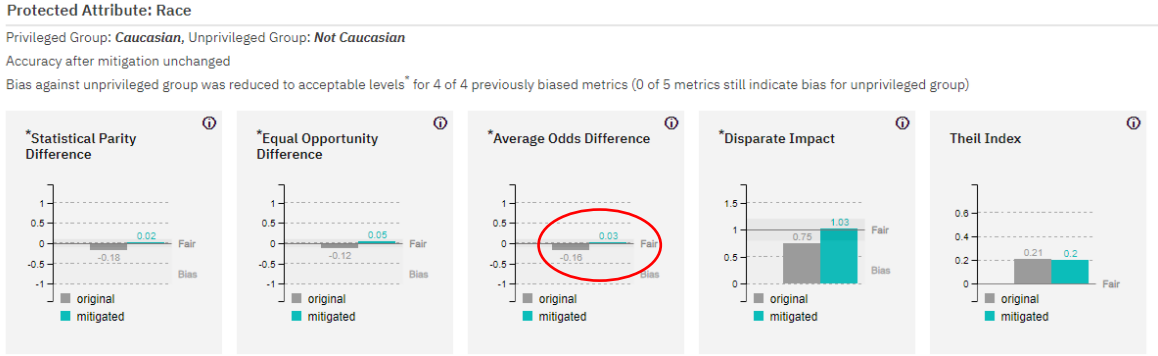


Open answer formative assessment

Correct answers for the formative assessment

5. How much does the average odds difference change in case of «reweighing algorithm» in both attributes?

Sex: 0,33 units
Race: 0,19 units



6) Why is reweighing necessary before bringing the application into use.

See the link here [Reweighting: Refining AI with Precision and Efficiency | by maxine | Medium]

In short: it allows algorithm to be recalibrated without using more data.

THANK YOU

Project number: 2022-1-ES01-KA220-HED-000085257

Co-funded by the European Union