BENCHATHON DATACARFT
25/06/2021

AI FAIRNESS AND ALGORITHM BIAS

Objective : comparison of fairness libraries identified to date
- Aequitas
- AIF360
- Fairlearn
- Shapash
- What if tool

Disclaimer : the results presented in this document are the result of a datacraft workshop whose goal was to get an overview of the identified fairness libraries to date. It is not mean to be exhaustive or based on a specific audit approach, but simply reflect the feedback of data expert users.

		Fairlearn		Shapash		Aequitas		AIF360		What if tool	
		Grade	Comment	Grade	Comment	Grade	Comment	Grade	Comment	Grade	Comment
Description	*What can it do?*		metrics calculation and bias mitigation (Open source project - Fairlearn Oarganization)		Interpretability and transparency of models (webapp) (MAIF)		Bias and audit toolkit (Open source project developed by the Center for Data Science and Public Policy at University of Chicago) Detect bias : - Biased actions or interventions that are not allocated in a way that's representative of the population. - Biased outcomes through actions or interventions that are a result of your system being wrong about certain groups of people.		Library built by IBM. Two main aspects: - Given a training set with ground truth labels and the same set with predicted labels, compute a number of fairness metrics. - Several mitigation of bias methods (pre-proc, in-proc, post-proc)		Web application for conterfactual analysis (what-if) - Handles a wide range of use cases (not just tabular data) - Compare different models - Conterfactual (nearest conterfactual, predict contefactual example) Developed by Google Research PAIR
Access	*Online tool, Python tool, …*		Python tool opening Jupyter dashboards https://fairlearn.org https://github.com/fairlearn/fairlearn		https://github.com/MAIF/shapash		http://aequitas.dssg.io (webapp for audit) http://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/ https://github.com/dssg/aequitas		https://github.com/Trusted-AI/AIF360		https://pair-code.github.io/what-if-tool/ https://github.com/pair-code/what-if-tool Python notebook
Installation	*How easy is it? Scale 1-5*	3	Works in 50% of the cases, even in a blank environment. Jupyter or Python 3.8 dependency issue	5	pip install and go	5	pip install aequitas	5	pip install aif360	5	pip install witwidget jupyter nbextension install --py --symlink --sys-prefix witwidget jupyter nbextension enable --py --sys-prefix witwidget
Usability	*Is it user friendly? Scale 1-5*	4	Not so easy to operate	5	maybe too much user friendly. Generates a local webapp from your notebook with a full-fledged interface. A very nice effort. Let see what we can get from code directly	5	rather user friendly	5	yes	2	Very tied to Google Ecosystem (tensforflow-friendly only) Helper functions available in notebooks, not directly in the source code. No real proper documentation
Documentation	*Is it well documented? Scale 1-5*	5	The documentation is well done and gives desire at first sight, but you don't really know where to start.	5	there is everything: product documentation, dev guide, tutorial notebooks...	5	well documented	3	Poor documentation (in particular, no tutorial)	4	Document looks complete on https://pair-code.github.io/
Completeness	*Does it perform all it is supposed to? Scale 1-5*	2	Not clear		it does assessment correctly	3	simple tool but seems effective	3	Very few (if no) tools for studying the data itself	5	
Reliability	*Does it seem reliable (code quality, code tested…)? Scale 1-5*	1	Not really			4	seems ok	5	Yes	4	seems ok
Legitimacy	*Is it popular within the community? (# stars on Github, …) Scale 1-5*	3	960 stars			3	424 stars	5	1.4K stars	3	537 stars
Future	*Gut instinct: would you trust it and use it in your real projects? Y/N*	5	Honestly, this is not ready. Micrsoft uses raiwidgets (Responsible AI Widgets) that provides responsible AI user interfaces for Fairlearn, that is not yet mature.	5	gut feeling: in short, yes.	5	yes	5	yes	0	- For business users, yes, even though it looks a little bit restrictive - For Data Scientists, probably a little too restrictive
Weaknesses	*What is missing?*		The bias mitigation technique is interesting (gridsearch on the sample weight with a reweighting method) that could be reproduced with AIF360. There are many methods in the library but they are not implemented. To be tested on a regression		It seems quite ok as a tool to 'assess' but there is nothing to 'act'. So you will need another tool		we were unable to get the online audit tool to work		- Vizualisation tools - Regression - Abstract the choice of the penalized population		Fairness. This tool is one block within the X/ethical AI ecosystem

	Fairlearn	Shapash	Aequitas	AIF360	What if tool
Installation	3	5	5	5	5
Usability	4	5	5	5	2
Documentation	5	5	5	3	4
Completeness	2		3	3	5
Reliability	1		4	5	4
Legitimacy	3		3	5	3
Future	5	5	5	5	0