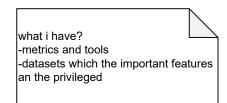
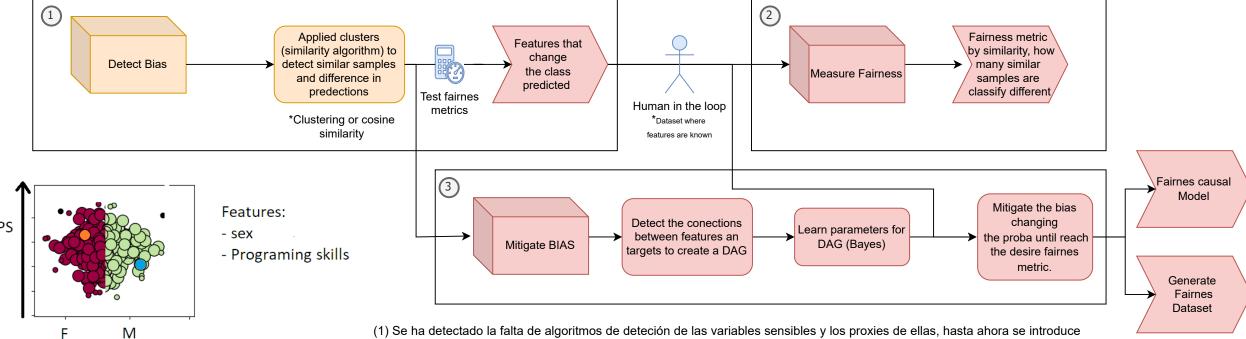


- *The detection of the features right now is manual, and in the tools tested are hardcode.
- **No sense calculate fairnes over all dataset cause this metrics no should affect all.
- *** The base of improve fairness is use causality and no correlation, for this reason for scoring and classification problem causal algoritms can get improvements in this area. Is important maintain or store the bias features to monitor and continuos analizing the problem. For other problem could be possible use bayes layer?? In this case for bayes could be adapt the reject option algorithm use for mitigation that change probabilities in the output class.





PS

Features

- Programing skills
- studies
- (1) Se ha detectado la falta de algoritmos de deteción de las variables sensibles y los proxies de ellas, hasta ahora se introduce en los algoritmos de medida estas features a mano habiendo sido selecionadas despues de auditorias de expertos. Con una deteción automatica de las variables y sus proxies se podría tener una representación de las variables que podrían estar siendo sesgadas. Siempre va a necesitar ser revisadas por un experto, ya que podría haber una seleción erronea de features, pero en problemas con una dimensionalidad muy grande podría ayudar a auditar los algoritmos y detectar problemas que estan siendo obviados.
- (2) Hasta ahora en los algoritmos se tiende a generalizar en la seleción de variables, y utilizar variables que correlan mucho con el target y pueden reducir la dimensionalidad, como es el caso del sexo, ya que esta variable lleva intrinsecos muchos factores, pero al igual que podría ser injusto tratar igual a dos hombres es injusto segregar una mujer y un hombre con caracteristicas similares en las variables que realmente nos importan. Con esta metrica se trataría de reducir dicha generalización en los algoritmos y penalizar variables altamente correladas.
- (3) Aportar un problema de mitigacion en el que la causalidad es la base del todo reflejada en el grafo y sobre ello mitigar el efecto de aquellas variables que no deberían estar alli. Este algoritmo permitiria una alta transparencia e interpretabilidad como explicabilidad del problema para ser auditado y entendido, aumentando la confianza en los procesos.