

Biais et Éthique en Apprentissage Statistique

"A Survey on Bias and Fairness in Machine Learning",
MEHRABI et al.

Ema Cerezo, Axel De Montgolfier, Nicolas Leroy

Université de Montpellier

- La notion de biais
- L'éthique en machine learning
- Pistes de solutions

Types de biais

Données

Biais d'échantillonnage
Biais de regroupement

...

Modèle

Biais algorithmique

...

Utilisation

Biais d'interaction utilisateur
Biais de popularité

...

Types de biais

Données

Biais d'échantillonnage

Biais de regroupement

...

Modèle

Biais algorithmique

...

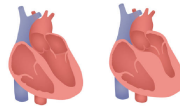
Utilisation

Biais d'interaction utilisateur

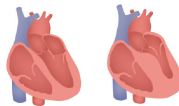
Biais de popularité

...

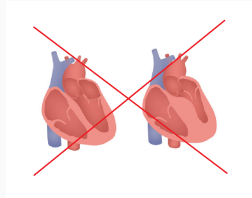
Exemple : Cardiomyopathie Hypertrophique (2016)



Exemple : Cardiomyopathie Hypertrophique (2016)



Exemple : Cardiomyopathie Hypertrophique (2016)



L'éthique en machine learning

Egalité des chances (Equalized Odds)

$$\mathbb{P}(\hat{Y} = 1 | S = 1, Y) = \mathbb{P}(\hat{Y} = 1 | S = 0, Y)$$

Parité démographique (Demographic parity)

$$\mathbb{P}(\hat{Y} = y | S = 0) = \mathbb{P}(\hat{Y} = y | S = 1)$$

Egalité des chances (Equalized Odds)

$$\mathbb{P}(\hat{Y} = 1|S = 1, Y) = \mathbb{P}(\hat{Y} = 1|S = 0, Y)$$

Parité démographique (Demographic parity)

$$\mathbb{P}(\hat{Y} = y|S = 0) = \mathbb{P}(\hat{Y} = y|S = 1)$$

Egalité des chances (Equalized Odds)

$$\mathbb{P}(\hat{Y} = 1 | S = 1, Y) = \mathbb{P}(\hat{Y} = 1 | S = 0, Y)$$

Parité démographique (Demographic parity)

$$\mathbb{P}(\hat{Y} = y | S = 0) = \mathbb{P}(\hat{Y} = y | S = 1)$$

Pré-traitement

Pondération,
Re-échantillonnage

Traitement

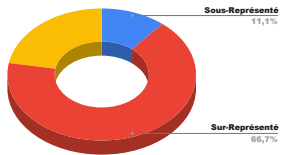
Contrainte
d'optimisation

Post-traitement

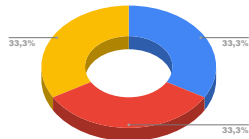
Renormalisation des
resultats

Re-échantillonnage

Distribution hétérogène



Distribution Homogène



Il reste à construire une définition plus unifiée de l'éthique algorithmique.

La mitigation de biais peut amener à des gains de précisions.

Existence de dataset biaisés pour l'entraînement.