Aspects éthiques

— Slides :

— Éthique

L'analyse de données est aujourd'hui utilisé dans de nombreux domaines : santé, éducation, politique publique, etc. Les données jouent même un rôle central dans la plupart des domaines. Les analyses permettent de prendre des décisions mieux informées et parfois automatisées. Dans certains domaines, ces décisions sont inconséquentes, mais pour d'autres domaines, cela soulève des questions éthiques majeures. Par exemple, est-ce que mes données représentent fidélement ma population ? quelles décisions sont prises à partir de ces analyses ? quelles sont les conséquences pour les individus concernés ?

Les données ne sont pas neutres, elles sont le reflet d'un contexte social, institutionnel et technique. Ainsi, les modèles qui en sont issus peuvent renforcer les inégalités, reproduire des biais historiques, ou éventuellement porter atteinte à la vie privée des individus. Il est donc essentiel, lorsque l'on fait une analyse de données, d'avoir une certaine vigilance éthique, en réflechissant aux impacts sociaux, aux limites méthodologiques et aux responsabilités associées à ses choix.

1 Confidentialité des données

Lorsqu'on manipule des données individuelles, il est essentiel de protéger la vie privée des personnes concernées. Plusieurs approches existent pour minimiser les risques de ré-identification ou d'exploitation abusive.

L'**anonymisation des données** consiste à supprimer ou à transformer les identifiants directs (e.g. les noms et adresses) et indirects (e.g. les dates de naissance et les codes postaux) susceptibles de permettre l'identification d'un individu. Il convient de faire particulièrement attention à ce cas, dans certaines situations, c'est le croisement de plusieurs variables qui permet l'identification d'un individu et non une unique variable. Par exemple, deux chercheurs de l'université du Texas ont été capable d'identifier des utilisateurs de Netflix en utilisant des notes sur IMDb (c.f. lien). On peut aussi **réduire la granularité des informations** en

limitant les détails fournis (e.g. en regroupant les âges en tranches et en n'utilisant pas l'âge exact) pour réduire le risque de ré-identification.

Lorsque certaines modalités d'une variable catégorielle sont associées à très peu d'individus, on peut les **regrouper** pour éviter qu'une combinaison unique de caractéristiques ne permette de retrouver une personne. On peut aussi **ajouter du bruit** aux données ou aux résultats. Cela permet de préserver des tendances globales tout en rendant les identification individuelles plus difficiles. Il faut faire attention à ne pas ajouter trop de bruit, sinon on risque de compromettre la validité des analyses. Enfin, on peut appliquer une **approche de confidentialité différentielle** consistant à garantir qu'un individu ne peut pas être identifié, même en connaissant toutes les autres données du jeu.

2 Implications sociales

L'utilisation des résultats d'analyses statistiques dans des contextes sociaux ou décisionnels soulève d'importants enjeux éthiques. En effet, plusieurs sources de discrimination peuvent intervenir dans un processus apparemment rigoureux.

Il peut y avoir un **biais dans l'échantillonnage** si le jeu de données ne reflète pas fidèlement la population cible (e.g. la sous-représentation de certaines communautés). Dans ce cas, les modèles entraînés risquent d'être inéquitables. Il peut aussi y avoir un **biais dans la variable à expliquer**. Par exemple, pour un modèle basé sur des décisions judiciaires passées, il est possible qu'il perpétue les biais historiques des jugements.

On peut aussi se trouver avec une **validité variable selon les groupes**. Un même modèle peut très bien fonctionner pour un sous-groupe de la population, mais être très mauvais pour un autre, menant à des erreurs systématiques. Enfin, si certaines classes sont **mal représentées**, e.g. des maladies rares ou des groupes démographiques minoritaires, et donc peu présentes dans les données, le modèle peut mal les prédire ou même les ignorer.

3 Déconstruire quelques mythes

— **"La machine apprend toute seule."**

En réalité, le choix des données, des variables, de la variable à expliquer et de l'algorithme est faite par des êtres humains. L' "apprentissage" dépend entièrement des décisions humaines en amont.

— **"C'est objectif, c'est basé sur des données."**

Les données ne sont jamais neutres. Elles sont le produit d'un contexte social, institutionnel et méthodologique. Les modèles entraînés sur ces données héritent de leurs biais.

— **“Mon modèle ne peut pas être sexiste car je n’ai pas utilisé le genre pour le construire.”**

Même sans inclure explicitement une variable comme le genre, un modèle peut apprendreà en inférer à partir d’autres variables corrélées comme le profession ou le parcours scolaire. L’exclusion explicite d’une variable sensible ne garantit pas l’absence de biais.

4 Que faire pour réduire ces biais ?

Plusieurs approches peuvent être mises en oeuvre à différentes étapes du processus :

— **Agir en amont sur les données** : rééquilibrer l’échantillon, sur-représenter certains groupes ou encore corriger les biais connus dans la variable à expliquer.
— **Modifier les modèles a posteriori** : appliquer des corrections sur les résultats produits comme le recalibrage des probabilités ou l’ajustement des seuils de décision pour certaines classes.
— **Modifier la fonction de perte utilisée pour l’apprentissage** : intégrer des pénalités visant à réduire les inégalités de performance entre les groupes. On peut, par exemple, imposer une équité en terme de faux positifs pour les différents groupes dans la population.