Qualification du biais de données dans le processus de la science des données

Ginel Dorleon[1], Nathalie Bricon-Souf[1],
Imen Megdiche[1], Olivier Teste[1]

[1]Institut de Recherche en Informatique de Toulouse, France
prenom.nom@irit.fr

Dans le contexte de l'apprentissage machine, les données constituent la principale ressource pour guider les prises de décisions. Cependant, lorsque des biais existent dans les données, cela affecte de façon significative l'interprétation des décisions. Le biais est défini comme *une distorsion systématique d'une évaluation ou d'un échantillon statistique choisi de façon défectueuse* [1]. Notre travail consiste à définir et qualifier les biais dans les données utilisées au cours du processus d'apprentissage. Sur un processus d'apprentissage, nous avons identifié les étapes où les biais peuvent survenir. Nous définissons les biais selon Mehrabi et al. (2019) en les qualifiant tout au long du processus d'apprentissage comme illustré à la Fig.1. Nous soulignons dans la suite les défis liés à ces biais.

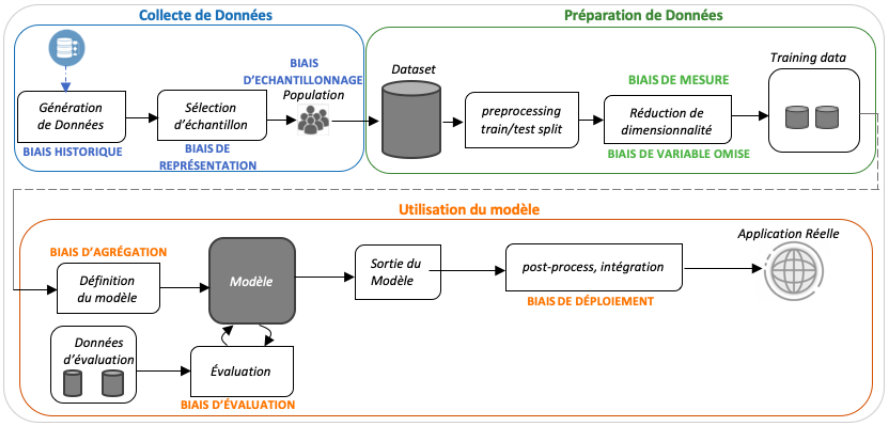

FIG. 1 – *Identification des biais au cours des différentes étapes d'un processus d'apprentissage automatique.*

Les biais dans le processus d'apprentissage. Nous présentons ce processus en trois étapes afin de catégoriser les biais. Chaque étape est dédiée à un objectif spécifique :

1. Le Petit LAROUSSE Illustré 2020, page 155

1. Collecte de Données : cette étape consiste à recueillir des données du monde réel par un échantillon de la population décrite par les données. Nous identifions trois types de biais à cette étape : (i) le biais historique qui concerne les problèmes socio-techniques déjà présents dans les données collectées ; (ii) le biais d'échantillonage qui est dû à la manière de réaliser l'échantillonnage (par exemple en modifiant la représentativité de certaines catégories dans les données) et (iii) le biais de représentation qui provient des critères utilisés pour échantillonner la population provoquant une sous-représentation des différents sous-groupes.
2. Préparation de Données : cette étape consiste à mesurer et à sélectionner les caractéristiques pour construire un ensemble de données d'apprentissage. Deux types de biais sont identifiés à cette étape : (i) le biais de mesure qui provient de la façon dont les caractéristiques sont choisies et (ii) le biais de variables (ou caractéristiques) omises qui se produit lorsqu'une ou plusieurs caractéristiques importantes sont omises pendant la phase d'apprentissage du modèle.
3. Utilisation du Modèle : cette étape regroupe les actions liées à l'entraînement, à l'évaluation et au déploiement du modèle. Nous identifions trois types de biais : (i) le biais d'agrégation qui apparaît lorsque le modèle est utilisé sur des sous-groupes avec des distributions conditionnelles différentes ; (ii) le biais d'évaluation qui se produit lorsque les données d'évaluation utilisées pour évaluer le modèle ne représentent pas la population cible initiale et (iii) le biais de déploiement qui surgit s'il y a un décalage entre le problème pour lequel le modèle est conçu et la façon dont il est réellement utilisé.

Les défis soulevés par les biais du processus d'apprentissage. Nous considérons trois principales catégories de défis : « Imbalanced Data », « Feature Selection », « Model Deployment ». Ces défis correspondent aux différents types de biais ; dans le tableau 1, nous résumons ces défis et les biais respectifs susceptibles de découler de ces défis.

Étape	**Collecte de Données**			**Préparation de Données**		**Utilisation du Modèle**		
Biais	Historique	Échantillonnage	Représentation	Mesure	Var. Omise	Agrégation	Évaluation	Déploiement
Défis	Imbalanced Data			Feature Selection		Model Deployment		
Solution Existante	-	Resampling	Reweighting	Ensemble Feature Selection	Double Cross-validation	Multi-task Learning	Subgroup Evaluation	-

TAB. 1 – *Défis et biais associés - solutions existantes*

Bien que nous ne proposions pas de nouvelles méthodes pour atténuer ces biais, nous identifions des solutions existantes pouvant aider à déjouer les effets négatifs des biais. Cependant, ces solutions doivent être utilisées dans leur contexte de biais respectif. Leur applicabilité à des contextes différents sera étudiée dans un prochain travail. Cette étude vise à sensibiliser les lecteurs sur les risques sous-jacents à la mise en place de modèles d'apprentissages sur des données biaisées.

Références

Mehrabi, N., F. Morstatter, N. Saxena, K. Lerman, et A. Galstyan (2019). A Survey on Bias and Fairness in Machine Learning. arXiv e-prints, arXiv :1908.09635.