

Modélisation du biais dans les recrutements : étude de l'influence d'un biais dans les données d'apprentissage de différentes procédures

Objectifs

Le marché des solutions de solution utilisant l'intelligence artificielle (IA) explose aujourd'hui sous l'impulsion de startups proposant par exemple des solutions de recommandations automatiques de candidats via une recherche sémantique sur les CV¹, des solutions d'analyse de signaux verbaux et non verbaux au cours des entretiens², ou encore des outils de prédiction de la capacité des candidats à s'entendre avec une équipe et à travailler efficacement³. Toutes ces solutions promettent des recrutements plus efficaces et surtout exempts de biais et de discrimination. Pourtant, si les biais et la discrimination sont des risques inhérents à toute décision de recrutement humain, les particularités qui caractérisent les technologies de l'IA, en termes d'opacité, de complexité, de comportement partiellement autonome ne fournissent pas de garanties pour réduire ces biais, et on peut même penser que l'IA peut engendrer d'autres types de biais (Lacroux et Martin-Lacroux, 2021). Par exemple, un algorithme entraîné à prendre des décisions à partir de données sociales biaisées reproduit ces biais, voire même les amplifie et crée des discriminations fortes fondées sur des critères comme le sexe, l'âge, l'origine des individus (pour un exemple, voir Besse (2020)).

Dans ce contexte, le projet IAB@R (Intelligence Artificielle, Biais et Acceptabilité dans le recrutement) ambitionne de questionner la validité et les impacts des outils d'IA appliqués au recrutement. Force est de constater que les études sont rares alors que la demande de données empiriques est forte (Villani et al., 2018) et que les enjeux sont importants et particulièrement sensibles notamment dans le cadre de la promotion de la diversité et la lutte anti-discrimination. Cette lutte correspond non seulement à un impératif éthique mais aussi à une obligation légale en termes d'explicabilité des décisions prises sur la base de recommandations algorithmiques (la loi précise que les recruteurs doivent être capables si besoin de justifier leurs choix).

L'objectif de ce stage est d'analyser le comportement de différents algorithmes (régression logistique, forêts aléatoire et réseaux de neurones notamment ; une phase d'apprentissage de certains algorithmes est prévue dans les 6 mois si besoin) en donnant des données biaisées en entrées et de voir l'impact sur les estimations de sortie. Pour ce faire, le ou la stagiaire analysera des plans de simulations pour estimer l'influence du biais avec des données quantitatives et qualitatives (en s'appuyant notamment sur l'article de Besse (2020) et un stage de M1 effectué l'an dernier). Dans un second temps, une analyse sur des données mimant des recrutements réels sera faite (en partenariat avec un stage de M1 proposé en parallèle). La fin du stage dépendra des résultats obtenus et du projet professionnel du ou de la stagiaire.

Aucune connaissance en science sociale n'est demandée. Ce sujet n'est pas susceptible d'aboutir sur une thèse mais des discussions sont potentiellement possibles si le ou la stagiaire serait intéressé-e.

Mots-clefs : plan d'expériences, classification supervisée, régression logistique, forêts aléatoires, réseaux de neurones, biais de réponses.

Durée : 6 mois. Indemnisé.

Niveau souhaité pour le stage : M2 ou dernière année d'école d'ingénieur en statistique appliquée et/ou théorique, science des données et/ou en machine learning.

1. <https://www.avature.net/fr/>

2. <https://cryfe.swiss/#cryfe>

3. <https://www.assessfirst.com/fr/solutions/recrutement/>

Encadrement : Vincent Brault (contact : Vincent.brault@univ-grenoble-alpes.fr) ; Christelle Martin Lacroux et Alain Lacroux

Références

- P. Besse. Détecter, évaluer les risques des impacts discriminatoires des algorithmes d'ia. 2020.
- A. Lacroux et C. Martin-Lacroux. L'intelligence artificielle au service de la lutte contre les discriminations dans le recrutement : nouvelles promesses et nouveaux risques. *Management Avenir*, (2) :121–142, 2021.
- C. Villani, Y. Bonnet, C. Berthet, F. Levin, M. Schoenauer, A. C. Cornut, et B. Rondepierre. *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*. Conseil national du numérique, 2018.