
Trajectoires professionnelles de docteur.es au sein de l'enseignement supérieur : l'intérêt des algorithmes de classification de prénom

Matthieu Cisel^{*1}, Pauline Antognelli^{*2}, and Sami Nakab^{*2}

¹Laboratoire AGORA – CY Cergy Paris Université : EA7392, CY Cergy Paris Université – France

²Institut des Humanités numériques – CY Cergy Paris Université – France

Résumé

Au cours des dix dernières années, sont apparus des outils informatiques visant à associer à des prénoms d'origines variées, d'une part un genre, et d'autre part une origine ethnique. Malgré les limites consubstantielles aux modèles d'apprentissage machine sous-jacents, un certain nombre de recherches en sciences humaines se sont saisies de ces techniques pour produire des résultats prometteurs. Dans la lignée de ces travaux, nous proposons une étude fondée sur la base de données theses.fr et visant à analyser les trajectoires de centaines de milliers de docteur.es au sein de l'enseignement supérieur français. En mobilisant des régressions logistiques, nous cherchons en particulier à identifier des liens statistiques qui existent entre certaines caractéristiques des docteur.es – disciplines d'appartenance, genre et origine ethnique prédite, et la probabilité de devenir soi-même un jour encadrant.e, et présenter ainsi son/sa premier.e doctorant.e. Eu égard à cette progression dans la carrière académique, nous constatons des écarts significatifs entre prénoms prédits comme masculins et ceux prédits comme féminins, écarts amplifiés lorsque l'on croise le genre avec l'origine ethnique du prénom. Laissant de côté l'interprétation des contrastes ainsi révélés, nous nous basons sur ces résultats pour mener une réflexion méthodologique sur les limites et opportunités des bibliothèques visant à classifier les prénoms, et sur celles des techniques d'apprentissage machine en général. Si celles-ci ne sont pas suffisamment précises pour produire des statistiques précises, elles ont néanmoins permis ici de dégager rapidement des tendances qu'il convient de confirmer par la suite, tout en respectant les contraintes associées à l'interdiction de la production de statistiques ethniques.

*Intervenant