
Impact de l’anonymisation sur la fouille de données

Benjamin Nguyen^{*†1}

¹Laboratoire d’Informatique Fondamentale d’Orléans – Université d’Orléans : EA4022, Institut National des Sciences Appliquées - Centre Val de Loire : EA4022 – France

Résumé

Dans les pays européens, le traitement de données personnelles est régi depuis 2018 par le Règlement Général sur la Protection des Données (RGPD). Afin de protéger la vie privée des individus, tout traitement de telles données est contrôlé attentivement par les autorités de régulation (CNIL). Toutefois, si une donnée est anonyme, c’est-à-dire qu’il est difficile, voire impossible de la relier avec un individu du monde physique, alors cette donnée est moins risquée, et donc son traitement peut se faire librement. Dans ce tutoriel, nous présentons les modèles d’anonymisation classiques pour des données tabulaire (modèles à base de k -anonymat et confidentialité différentielle). Nous proposons en illustration, la mise en pratique de ces notions sur machine, avec le logiciel d’anonymisation open source ARX, développé par l’Université Technologique de Munich, pour générer une base de données anonyme exploitée avec des algorithmes de classification (sous WEKA).

*Intervenant

†Auteur correspondant: benjamin.nguyen@insa-cvl.fr