
Compétitions d'apprentissage automatique avec le package rchallenge

Robin Genuer*^{†1,2} and Adrien Todeschini³

¹Institut de Santé Publique, d'Epidémiologie et de Développement (ISPED) – Université Victor Segalen
- Bordeaux II – 146 rue Léo Signat 33076 Bordeaux Cedex, France

²SISTM (INRIA Bordeaux - Sud-Ouest) – Université de Bordeaux, Inserm : U897, INRIA – France

³CQFD (INRIA Bordeaux - Sud-Ouest) – Université de Bordeaux, CNRS : UMR5251, INRIA – France

Résumé

En apprentissage automatique, les performances empiriques obtenues sur données réelles sont déterminantes dans le succès d'une méthode. Ces dernières années ont vu l'apparition d'un grand nombre de compétitions d'apprentissage automatique. Ces challenges sont motivés par des applications industrielles (prix Netflix) ou académiques (challenge HiggsML) et mettent en compétition chercheurs et data scientists pour obtenir les meilleures performances. Nous avons souhaité confronter les étudiants à cette réalité en leur soumettant un challenge dans le cadre du cours d'apprentissage automatique. Leur classement est affiché sur une page web mise à jour automatiquement permettant une émulation parmi les étudiants. L'historique des résultats leur permet également de visualiser leur progression au fil des soumissions. De plus, le challenge peut se poursuivre en dehors des sessions encadrées favorisant l'autonomie et l'exploration de nouvelles techniques d'apprentissage et outils informatiques. Le système que nous avons mis en oeuvre est disponible sous forme de package R afin d'être réutilisé par d'autres enseignants. S'appuyant sur les outils R Markdown et Dropbox, il ne nécessite aucune configuration réseau et peut être déployé très facilement sur un ordinateur personnel.

*Intervenant

[†]Auteur correspondant: robin.genuer@isped.u-bordeaux2.fr