

## SITE CHALLENGE DATA POUR L'ENSEIGNEMENT ET LA RECHERCHE

Mathieu ANDREUX<sup>1</sup>, Stéphane MALLAT<sup>2</sup> et Gilles WAINRIB<sup>3</sup>

Le site Web [challengedata.ens.fr](http://challengedata.ens.fr) met à disposition des challenges de traitement de données, destinés aux étudiants de licence ou de master, mais aussi aux chercheurs. Ces challenges sont proposés par des entreprises ou des scientifiques, et sont issus de problématiques concrètes qu'ils rencontrent dans leur activité. Ils s'inscrivent dans un esprit d'échange scientifique, avec un partage de données et de résultats : les données mises à disposition par les entreprises sont non-confidentielles, et les rapports algorithmiques des étudiants et des chercheurs sont mis à disposition de tous.

Les challenges sont des problèmes de classifications supervisées, de régressions ou de prédictions, avec des données réelles, mises à disposition par des start-ups, des entreprises, des hôpitaux, ou des scientifiques. Les performances des algorithmes des participants sont évaluées sur des données de tests, avec une métrique permettant de quantifier l'importance des erreurs. Le site Web peut ainsi calculer un score et un classement pour chaque participant. Cela permet aux étudiants, professeurs et chercheurs d'évaluer leurs résultats dans une communauté très large, et comparer des approches très différentes.

Ces challenges couvrent un grand spectre d'applications : données médicales, images/sons, marketing, recherches Internet, politiques publiques... Deux saisons de challenges sont déjà disponibles : 2015/16 et 2016/17. La sélection des projets est organisée par une équipe de l'École Normale Supérieure, soutenue par la Chaire *Capital Fund Management* et la Fondation des Sciences Mathématiques de Paris. Cette année, il y a eu 11 challenges, avec environ 2000 participants incluant des étudiants mais aussi des chercheurs et des ingénieurs. Plus d'une dizaine d'enseignants ont utilisé ces challenges pour leurs cours, à travers toute la France et dans des pays étrangers (Danemark, Maroc, Inde...).

Les challenges sont formatés pour simplifier la définition du problème, mais ils correspondent tous à la résolution d'une question importante. Le site inclut toute l'information nécessaire pour comprendre les enjeux, y compris une vidéo de présentation par les auteurs. Dans plus de la moitié des cas, les étudiants ont obtenu des résultats remarquables, nettement meilleurs que ceux des entreprises ou scientifiques qui ont proposé ces challenges. Une remise de prix pour les gagnants a lieu en Juin à l'Institut Henri Poincaré à Paris. C'est l'occasion pour les élèves de rencontrer les entreprises qui ont proposé ces challenges. De nombreux stages et embauches résultent de ces prises de contact.

Les Professeurs peuvent inscrire leur cours sur le site Web et spécifier une liste de projets pouvant être traités par les élèves du cours. Le site inclut un tableau de bord pour les enseignants, leur permettant de suivre l'activité et les performances de leurs élèves, et récupérer automatiquement leurs rapports.

---

<sup>1</sup> École Normale Supérieure, 45 rue d'Ulm, 75005 Paris, [Mathieu.andreux@ens.fr](mailto:Mathieu.andreux@ens.fr)

<sup>2</sup> École Normale Supérieure, [stephane.mallat@ens.fr](mailto:stephane.mallat@ens.fr)

<sup>3</sup> École Normale Supérieure, [gilles.wainrib@ens.fr](mailto:gilles.wainrib@ens.fr)

*Enseigner le recueil de données : étude de cas en épidémiologie*

La saison actuelle se finira en Décembre 2017. Une nouvelle série de challenges seront disponibles en Janvier 2018, et présentés dans le cadre du cours de Stéphane Mallat de la nouvelle Chaire de « Sciences des Données » du Collège de France.