



## Guide pour les créateurs de challenges

Un challenge doit tout d'abord être proposé en envoyant un courriel à [challenge.data@ens.fr](mailto:challenge.data@ens.fr). Le projet de challenge sera relu par l'équipe du Challenge Data, afin de faciliter son intégration dans la plateforme. Une fois que le projet est validé, vous devez créer un compte et un challenge sur le site en suivant la procédure décrite ci-dessous. Le challenge sera testé par l'équipe Challenge Data avant d'être validé définitivement et mis à la disposition des participants.

Vous pouvez présenter votre challenge et de rencontrer des étudiants et des chercheurs à l'Ecole Normale Supérieure (Amphi Rataud, 45, rue d'Ulm, Paris) le vendredi 18 Novembre 2016, de 9h30 à 12h30. Au cours de la cérémonie de clôture organisée à la fin de la saison, les gagnants des challenges auront l'opportunité de présenter leurs algorithmes.

### 1. Création d'un compte

- Cliquez sur le lien INSCRIPTIONS :
- Entrez votre adresse email ainsi qu'un mot de passe contenant des lettres et des chiffres ;
- Cochez la case « Créateur de challenge » ;
- Entrez le nom de l'entreprise ainsi que les nom et prénom du contact dans cet entreprise ;
- En cas de problème avec le formulaire, un message rouge apparaît en haut de la page ;
- Ouvrez votre boîte mail (y compris le dossier spam) et trouvez un courriel de Challenge Data ;
- Cliquez sur le lien contenu dans l'email afin d'activer votre compte ;
- Votre inscription est maintenant en attente : les administrateurs vont tâcher de valider manuellement votre compte dans les plus brefs délais. Une fois que votre compte sera validé, vous recevrez un courriel confirmant votre inscription en tant que fournisseur de challenge.

### 2. Votre espace personnel (« Mon espace »)

- Sous l'onglet « Description », vous pouvez insérer et modifier la description de votre entreprise, ajouter un lien vers un site web, ainsi qu'un logo pour votre entreprise.
- Une fois que votre compte a été validé par les administrateurs, vous pouvez créer un challenge dans l'onglet « Vos challenges ».

### 3. Création d'un challenge

**Avant de créer un challenge, vous devez préparer quatre fichiers différents :**

1. Un fichier .zip contenant les **entrées d'entraînement** (taille maximale : 4Go)
2. Un fichier .zip contenant les **entrées de test** (taille maximale : 4Go)
3. Un fichier .csv contenant les **sorties d'entraînement** (taille maximale : 200Mo), qui a typiquement le format suivant :

ID, Target  
ID0000, 0  
ID0001, 1  
ID0002, 0  
...  
ID10000, 1

4. Un fichier .csv contenant les **sorties de test** (taille maximale : 10Mo), c'est-à-dire le « fichier de solutions » : c'est l'unique fichier non accessible par les participants, qui sera utilisé pour comparer leurs prédictions avec une valeur de référence. Ce fichier doit avoir le même format que les sorties d'entraînement (avec les ID continuant dans l'ordre) :

ID, Target  
ID10001, 1  
ID10002, 0  
ID10003, 1  
...  
ID20000, 0

Une fois que ces fichiers ont été créés, vous pouvez créer le challenge :

- Pour créer un challenge, cliquez sur MON ESPACE, allez dans l'onglet « Vos challenges » et cliquez sur le lien « Créer un challenge ».
- Un formulaire pop-up va s'ouvrir : vous devez le remplir avec la description du challenge (cf tableau en page suivante) et télécharger les données.
- Terminez la création.

#### 4. Tableau de bord du challenge

- Une fois que votre challenge est validé par les administrateurs, il devient visible sur la page CHALLENGES.
- Sur la page de votre challenge, l'onglet « Introduction » affiche la description de votre challenge et de votre entreprise. Sur cet onglet, à droite, le lien « Modifier le benchmark » vous permet de télécharger un fichier au même format celui qui est attendu des participants. Ce fichier donnera un score affiché au classement général comme étant l'étalon de votre entreprise.
- L'onglet « Classement en cours » affiche le classement actuel des participants, calculé sur la moitié publique des données d'entraînement, y compris votre étalon. Il est visible par tous les participants au challenge.
- L'onglet « Informations Participants » est un tableau de bord présentant des informations plus détaillées sur les concurrents. En particulier, il vous permet de contacter les participants qui ont accepté d'être contactés, ainsi que de télécharger les rapports publics.
- Au moment de la date de clôture intermédiaire, un nouvel onglet intitulé « Classement intermédiaire » apparaît. Il contient les classements des concurrents sur la moitié privée des données de test à la date en question, et ne sera plus modifié par après.
- A la fin de la saison, un dernier onglet intitulé « Classement final » affiche les scores de tous les concurrents sur la moitié privée des données de test, avec le classement approprié. Des différences qui apparaîtraient par rapport au classement en cours sont la marque d'un surapprentissage de la part de certains concurrents.

| Champ                   | Explication  | Exemple   |
|-------------------------|--|---|
| Saison                  | Sélectionnez la saison actuelle  | 2- Challenge Data 16-17   |
| Métrique                | Sélectionnez la métrique utilisée pour évaluer la performance de prédiction. Merci de nous contacter si la métrique que vous souhaitez utiliser n'est pas dans la liste, ainsi que de préparer un script python la calculant.  | Classification<br>- Accuracy<br><br>Regression<br>- Mean-square error   |
| Nom                     | Nom du challenge   | Prédire les marchés boursiers   |
| Description courte      | Il s'agit d'une description courte visible par tout le monde et visant à attirer les participants sur votre challenge.   | Peut-on prédire la bourse à l'aide des données massives et de l'apprentissage supervisé ?   |
| Contexte                | Expliquez le contexte du challenge.  | L'augmentation des capacités de stockage des données, combinée aux progrès récents en apprentissage statistique, etc.   |
| Objectifs               | Expliquez les objectifs du challenge, ce que vous voulez prédire et pourquoi c'est important   | L'objectif de ce challenge est de prédire les prix futurs de la bourse car ...  |
| Description des données | <p>Il s'agit de la section la plus importante et vous devez la remplir attentivement. Il s'agit de décrire ici les données. Trois fichiers seront mis à la disposition des concurrents :</p> <ul style="list-style-type: none"> <li>- Fichiers d'entrée d'entraînement et de test : décrivez l'organisation des données (format des fichiers, nombre d'échantillons, nombre et rôle des colonnes, etc.)</li> <li>- Sortie d'entraînement : ce sont les valeurs cibles correspondant aux données d'entrée pour l'entraînement.</li> </ul> <p>Les concurrents doivent prédire les valeurs du fichier de sortie de test mais n'y ont bien évidemment pas accès. Vous devez préciser le format du fichier .csv qui est attendu en donnant un exemple (faux).</p> | <p>Les entrées d'entraînement sont données dans le pack d'entraînement comme un fichier .zip. Les entrées de test sont fournies dans le pack de test comme un fichier .zip</p> <p>Chaque ligne correspond à un échantillon, chaque colonne à un feature.</p> <p>Les sorties d'entraînement sont fournies dans un fichier .csv . Chaque ligne correspond à un échantillon.</p> <p>Le fichier de prédiction devrait être un fichier .csv au format suivant :</p> <p>ID, Target<br/>ID5001, 45.99<br/>ID5002, 44.53<br/>...<br/>ID10000, 36.65</p> |
| Logo                    | Il s'agit d'une image pour le challenge  |   |