



Challenge data  
[challengedata.ens.fr](http://challengedata.ens.fr)



## Guidelines for Challenge Providers

A challenge project must first be submitted by sending an email to [challenge.data@ens.fr](mailto:challenge.data@ens.fr). The challenge project will be reviewed by the Challenge Data team, to help its incorporation in the platform. Once the project is validated, you must create an account and a challenge according to the procedure described below. The challenge will be tested by the Challenge Data team before it is fully validated and available to competitors.

You can present your challenge and meet students and scientists at Ecole Normale Supérieure (Amphi Rataud, 45, rue d'Ulm, Paris) on Friday 18<sup>th</sup> November 2016, from 9:30 am to 12:30 pm. Challenge winners will be given the opportunity to present their algorithm in an oral session, at the end of the season in June.

### 1. Account creation

- Click on the SIGN UP link;
- Enter your email address and a password with letters and numbers;
- Select the case “Company”;
- Enter the company name and the contact's last and first names;
- If there is a problem with your form, a red message will appear at the top of the page;
- Check your mailbox (including your spam folder) for an email from Challenge Data;
- Click on the link contained in the email to activate your account;
- Your registration is now pending; your account will be validated shortly by the administrators of the platform. Once your account has been validated, you will receive an email message that confirms your registration as a Challenge Provider.

### 2. Your personal space (My Space)

- Under Description tab, you can insert and modify your company description, a website url, and add a company logo.
- Once your account has been validated by the administrators, you can proceed to challenge creation under the tab “Your challenges”;

### 3. Challenge creation

**Before creating a challenge, you must prepare four different files:**

1. A .zip file containing the **training inputs** (maximal size: 4GB)
2. A .zip file containing the **testing inputs** (maximal size: 4GB)
3. A .csv file containing the **training outputs** (maximal size: 200 MB), which typically looks like:

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

4. A .csv file containing the **testing outputs** (maximal size: 10MB), in other words the “solution file”: this is the only file which will be kept secret from the participants and will be used to compare their predictions with the ground truth. This file must have the same format as the training outputs file (with IDs following up the list):

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

Once these files have been created, you can proceed to the challenge creation:

- To create a challenge, click on MY SPACE, go under the “Your Challenges” tab and click on “Create a challenge”
- A pop-up form will open, and you have to fill it with the challenge description (cf table next page) and upload the data
- Finish the creation

#### 4. Challenge Dashboard

- Once your challenge is validated by the administrators, it is visible on the CHALLENGES page.
- On the page of your challenge, the tab “Introduction” displays the description of the challenge and of your company. In this tab, on the right, the link “Modify benchmark” allows you to upload a file in the same format as the testing output for the competitors. This file will result in a score visible to the participants as the benchmark of your company.
- The tab “Current ranking” shows the current ranking of the participants on the public half of the testing dataset, including your benchmark. It is visible by all participants to the challenge.
- The tab “Participants informations” is a dashboard showing more refined information about the competitors. In particular, it allows you to send messages to participants which have agreed to be contacted, as well as downloading the reports which have been made public.
- The tab “Current ranking” gives you access to the ranking of the competitors obtained on the “public” half of the testing dataset, along with their scores.
- When the season reaches the intermediate closing date, a new tab entitled “Intermediate ranking” appears. It contains the rankings of the competitors obtained on the “private” half of the testing dataset, at the date of the closing date, and will not be updated afterwards.
- At the end of the season, a last tab entitled “Final ranking” contains the scores of all the competitors on the “private” half of the testing dataset, along with the relevant ranking. Differences with the current ranking might show some overfitting which occurred during the competition.

Field	Explanation	Example
Season	Select the current season	3- Challenge Data 2016-2017
Metric	Select the metric used to evaluate the prediction performance. Please contact us if the metric you want is not in the list and prepare a Python script computing your metric.	Classification - Accuracy  Regression - Mean-square error
Name	Name of the challenge	Predict the stock market
Short description	This is a short teaser which will be accessible to anyone and must be written to attract participants	Can we predict the stock market using big data and machine learning?
Context	Explain the context of the challenge	Our increasing access to large amounts of data, combined with recent advances in machine learning, etc.
Goals	Explain the goals of the challenge, what you want to predict and why it matters	The aim of this challenge is to predict the future price of the stock market because ...
Data description	<p>This is the most important section and you should fill it very carefully. You should describe here the dataset. Three files will be available to the competitors:</p> <ul style="list-style-type: none"> <li>– Input training and input testing: describe how the dataset is organized (file format, number of samples, features, etc)</li> <li>- Output training : these are the target values corresponding to the input training set.</li> </ul> <p>The competitors must predict the values of the output testing file but are obviously not allowed to download it. You should specify the format of the .csv file you expect them to furnish by giving a (fake) example.</p>	<p>The training inputs are given in the training package as a .zip file. The testing inputs are given in the testing package as a .zip file. Each line corresponds to a sample, each column to a feature.</p> <p>The training outputs are given in a .csv file. Each line corresponds to a sample.</p> <p>The prediction file should be a .csv file in the format:</p> <p>ID, Target  ID5001, 45.99  ID5002, 44.53  ...  ID10000, 36.65</p>
Logo	This is an image for the challenge	