

# Data Science Lab

## Guide to submit solutions for the competition

Politecnico di Torino  
January 10, 2020

### Intro

This guide explains the complete procedure that has to be followed to submit solutions to a Data Science Lab Competition.

## 1 Rules and other information

**Deadline.** Each submission will have a strict deadline. Uploads later than the deadline will not be evaluated.

**Evaluation and leaderboard.** Your results will be evaluated on a specific metric. Since the goal changes from time to time, you can find which metric will be actually used in the competition text. Based on your score, you will be placed in the leaderboard. Figure 1 depicts an example of leaderboard web-page, with buttons for managing submissions.

#	User Id	Score
1	321321	0.902
2	baseline	0.902
3	123123	0.899

Figure 1: Example of leaderboard web-page

We want you to provide a label for every point of the actual dataset. Nonetheless, at evaluation time, data are split in two chunks randomly: a *public* and a *private* one. Every time you perform a submission, the first chunk (i.e. the *public* one) is used to compute your temporary score and place you into the online leaderboard.

Once the competition is closed (i.e. right after the deadline) a new evaluation pass is run on the second, *private* chunk. This second evaluation will create a new, final leaderboard to declare the winners. Consider

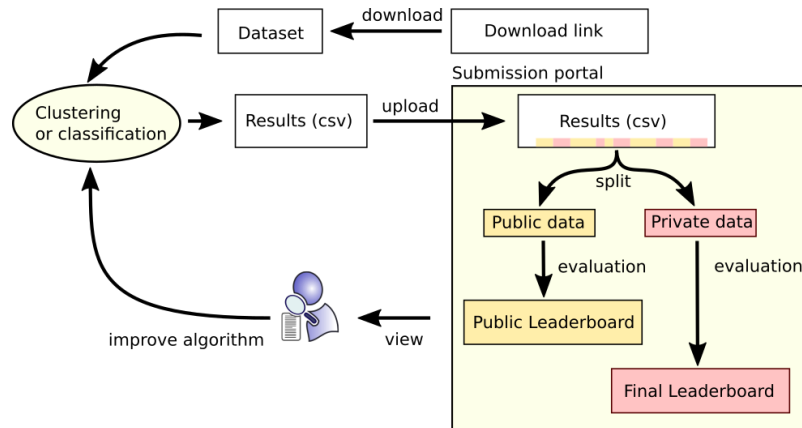


Figure 2: A graphical representation for the submission process

that the two chunks are sampled uniformly from the original dataset, hence they share the same properties and data distributions. Figure 2 shows a graphical representation for your submission pipeline.

**Baseline.** For each competition you will have a baseline score. This is a score achieved with a standard approach. You should try to achieve a better result than this value.

**Personal key.** To submit your result our system will ask for your personal key. Each student will receive its own personal key via email. These are strictly personal and must be used only in the submission form.

**Submissions limit.** In order to avoid overloading the server, you will have an upload limit (e.g. #upload per minute). Typically, we will give you this limit in the competition text and will be clearly highlighted in the platform.

**Multiple submissions and final evaluation.** During the competition you can upload multiple submissions. Your position in the *public* leaderboard will consider only the submission with the best score on public data.

You have **97** submissions left!

---

321321 submissions

#	Timestamp	Score	To evaluate
1	2020-01-09 14:08:25.466278	0.775	<input type="checkbox"/>
2	2020-01-09 14:15:17.385401	0.775	<input checked="" type="checkbox"/>
3	2020-01-09 15:31:04.306993	0.902	<input checked="" type="checkbox"/>

Click to update the submissions to be evaluated Update

© 2019-2020

Figure 3: Multiple-submission management panel.

Instead, for the final evaluation on *private* data, the system allows selecting the preferred solution(s). By clicking on the "Check submissions" button (Figure 1) and then inserting your API key you will open

the panel shown in Figure 3. This panel displays all your submissions associated with a timestamp and the score computed on the *public* data. You can select from zero to two submissions that will be evaluated. The following cases occur:

- No items selected. Your position in the *private* leaderboard will be computed on private data by considering the submission with the highest *public* score.
- One or two items selected. Your position in the *private* leaderboard will be computed on private data by considering the selected submission with the highest *private* score.