

CALIFICACIÓN

100 %

Practice Quiz

PUNTOS TOTALES DE 4

1.

Pregunta 1

Which of these methods are used by organizers to prevent overfitting to the test set using scores from the leaderboard?

1 / 1 punto



Splitting test for public/private

Correcto

The submissions are scored on public part of the test set while private score is only revealed when competition ends, thus it is not possible to overfit to private part



Shuffling order of rows in train and test sets



Adding non-relevant samples to test set

Correcto

Non-relevant samples make it much harder to do "leaderboard probing", i.e. get labels of individual objects based on leaderboard score



Limiting number of submissions per day

Correcto

This also makes "leaderboard probing" much harder

2.

Pregunta 2

What are the main reasons for participating in data science competitions? Check all that apply.

1 / 1 punto



To have an opportunity for networking with other participants

Correcto

Yes, the people on forums are very active and helpful.

☐

To easily earn a lot of money

☒

To learn basics of data science

Correcto

Yes! For example there are plenty of public datasets on [kaggle.com](https://www.kaggle.com) with a huge community of people who is also learning.

☐

To learn things about deployment/testing

3.

Pregunta 3

Imagine that you are participating in an image recognition competition and you want to use some publicly available set of images. Are you allowed to do this?

1 / 1 punto

☐

Yes, of course

☐

No, this is not the data provided by the organizers

☒

You need to check the competition rules

Correcto

If you have any doubt about competition mechanics and limitations -- check the rules or ask organizers on forums

4.

Pregunta 4

Which of these things we should take care about during competitions? Check all that apply.

1 / 1 punto

☐

Deployment issues

☐

Task formalization



Target metric value

Correcto

This one is used for leaderboard ranking



Inference speed