

Live-EO Data Science Challenge

Taking the provided Sentinel-2 (S-2) satellite image scene and labelled data (see links below), the task is to:

1. explore the dataset a bit, and
2. train a classifier that recognizes the different labelled objects (forest, water, urban, etc) on the S-2 image.

The labelled data is already rasterized at the same resolution, and contains the different land cover objects encoded as integers (see link to image below, pixel labels are stored in the first band of image).

The bands of the S-2 scene are preprocessed for the analysis: separated by band, clipped to the labelled data, warped to 10m resolution, and are on the same pixel grid as the rasterized labelled image.

To handle the images, we recommend using [rasterio](#). Make sure you exclude nodata value from model training and evaluation!

The choice of classifier is up to you, it can be very simple or something more complex. The goal here is just to provide a minimal solution, not to over-optimize classification accuracy. We care more about how you approach the problem, rather than the actual results.

Please send us

- all the code you have written (e.g. a Jupyter Notebook)
- the distribution of class labels in the labelled image
- results of a statistical test for relations between each class and each S-2 band, using a statistical test of your choice
- the trained classifier
- your evaluation of the classifier (e.g. accuracy measures, confusion matrix, etc)

If you have any questions or run into any problems, feel free to write us an email (david@live-eo.com).

Good luck!

Links to images

S-2 images: [S2_17SLD_2019-09-15_masked_warped](#)

Labelled image (use 1st band only!): [rasterize_on_sentinel_2_labelled_17SLD.tif](#)

Class encoding

0: urban

1: treerow

2: forest

3: single tree

4: agriculture

5: grassland

6: water

4294967295: nodata value (leave out from training & evaluation!)