The training set is composed by 16762 samples, each sample contains information about a listing on our platform. Goal of the classification is distinguishing between scam and legit listings at creation time, where the majority of these features are actually available. A brief explanation of the available features is provided in the subsequent table.

The final goal is to provide a prediction for some unseen samples using the trained model. The unseen samples might have been available (and therefore collected) in a different moment from the ones in the training set, and some features might be missing, being related to the labeling process itself.

| Feature Name | Description |
|---|---|
| LISTING_KIND | 0 - entire place; 1 - private room; 2 - shared room |
| LISTING_CITY | The city where the listing is located |
| LISTING_PRICE | The monthly rent (€) of the listing |
| IS_ARCHIVED | If the advertiser creating the listing has been |

| | archived or not. |
|---|---|
| **ARCHIVE_REASON** | The reason why the advertiser creating the listing has been (possibly) archived. |
| **LOGIN_COUNTRY_CODE** | The (last) country where the advertiser logged in from. |
| **LISTING_COUNTRY_CODE** | The country where the listing is located. |
| **LISTING_REGISTRATION_POSSIBLE** | If it's possible to use listing's address for registering at the city's municipality. |
| **ADVERTISER_COMPLETENESS_SCORE** | Percentage of completeness of the advertiser's profile. |
| **MANAGED_ACCOUNT** | If the advertiser creating the listing is managed by our employees or not. |
| **HAS_PROFILE_PIC** | If the advertiser creating the listing has a profile pic or not. |
| **BROWSER** | The browser used to create the listing. |
| **OS** | The operating system used to create the listing. |
| **ANONYMISED_EMAIL** | The email address (anonymised) of the advertiser. Letters have been changed with random letters, numbers have been changed with random numbers, all the other characters have been maintained. The email domain has been maintained. |
| **IS_SCAMMER** | Whether the listing is a scam (1 - bad) or not (0). |

پاسخهای مورد نیاز(به همراه کامنتها برای توضیحات)

- A csv file named HA_Data_Science_labeled.csv containing all the samples in HA_Data_Science_test.csv. The values for the column IS_SCAMMER (0 or 1) have to be populated using the trained model.

● The Jupyter Notebook (.ipynb) or the Python script (.py) developed for populating the dataset of the previous step.