**Fraud in Electricity and Gas Consumption Dataset (Team 45)**

# How to run the file:

To run the project, firstly install all the requirements needed. You can use:  
  
$ pip install -r requirements.txt

Then, you must ensure that “IT1244\_Team45\_Project.ipynb” is in the same folder as “invoice.csv” and “client.csv”. Then, ensure that you run all cells in order from the top. We have assumed that all installations have been made in the environment.

# Dataset Information:

|  |  |
| --- | --- |
| **Size** | 85 MB |
| **Dataset Characteristics:** | Multivariate |
| **Attribute Characteristics:** | Numerical, Categorical, Ordinal |
| **Associated Tasks:** | Classification |
| **Number of Instances:** | 135493 clients in total |
| **Number of Attributes:** | 5 features in “client.csv”, 16 features in “invoice.csv” |
| **Missing Values?** | Yes |
| **Number of Classes** | 2 (“target” column of “client.csv”) |
| **Area:** | Business |
| **Additional Details:** | highly imbalanced dataset (very low number of fraud cases) |

This dataset contains the client information for over 130,000 people, each of them can have many invoices (see “invoice.csv”). The common column between the two csv files is “id”. Your job is to analyze the pattern of their consumption and find out which are fraud.

# Dataset Organization:

data/

* client.csv
* invoice.csv

Attribute Information:

## client .csv

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Type** |
| id | Unique id for client | str |
| dis | The district where the client is | int |
| catg | Category client belongs to | int |
| region | Area where the client is | int |
| date | Date client joined | str |
| target | fraud:1, not fraud: 0 | int |

**invoice.csv**

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Type** |
| id | Unique id for the client | str |
| date | Date of the invoice | str |
| Tarif\_type | Type of tax | int |
| counter\_number | Number | float |
| counter\_statue | takes up to 5 values such as working fine, not working, on hold statue, ect | int |
| counter\_code |  | int |
| reading\_remarque | notes that the STEG agent takes during his visit to the client (e.g: If the counter shows something wrong, the agent gives a bad score) | int |
| counter\_coefficient | An additional coefficient to be added when standard consumption is exceeded | int |
| consommation\_level\_1 | Consumption\_level\_1 | int |
| consommation\_level\_2 | Consumption\_level\_2 | int |
| consommation\_level\_3 | Consumption\_level\_3 | int |

|  |  |  |
| --- | --- | --- |
| consommation\_level\_4: | Consumption\_level\_4 | int |
| old\_index: | Old index | int |
| new\_index: | New index | int |
| months\_number: | Month number | int |
| counter\_type: | Type of counter | str |