Fraud in Electricity and Gas Consumption Dataset

Size	40 MB
Dataset Characteristics:	Multivariate
Attribute Characteristics:	Numerical, Categorical, Ordinal
Associated Tasks:	Classification
Number of Instances:	Around 21000 clients in total
Number of Attributes:	5 features in “client.csv”, 16 features in “invoice.csv”
Number of Classes	2 (“target” column of “client.csv”)
Area:	Business
Additional Details:	highly imbalanced dataset (very low number of fraud cases)

Dataset Information:

This dataset contains the client information for over 21000 people, each of them can have 3-50 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 clients are fraud. Remember that you need to classify the client and not the transaction (a client performing 3-50 transactions). You are expected to apply various innovative techniques in – (1) handling the data imbalance issue and (2) summarizing the variable number of transaction info for each client.

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