Fraud Patterns Detection

UPC

Marc Jené

Fraud in Distribution Grid

Distribution Grid Losses = Energy IN – Energy OUT

Energy IN: Energy entering the distribution network, from embedded generators or upstream, same level or downstream networks.

Energy OUT: Consumed and properly accounted energy.

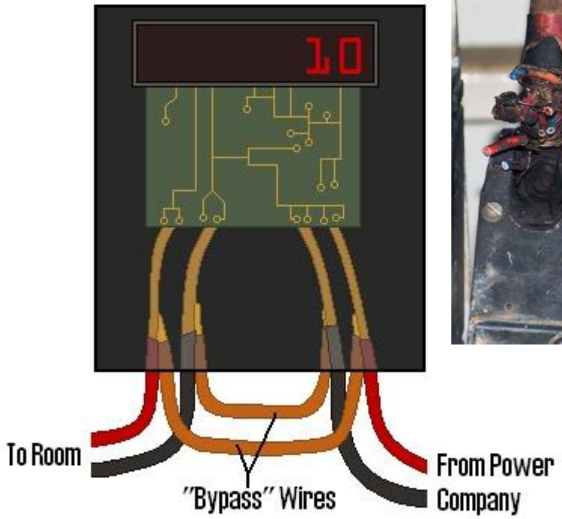
Distribution Grid Losses = Technical Losses + Non-technical Losses

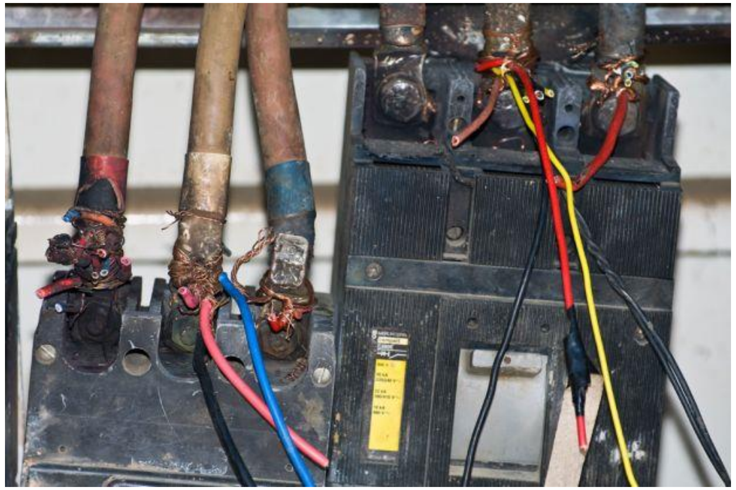
Technical Losses (TL): Variable losses (load related), fixed losses (not related to load) and network service (uncontracted consumptions of network equipment).

Non-technical Losses (NTL): Network equipment issues, network information issues and energy data processing issues.

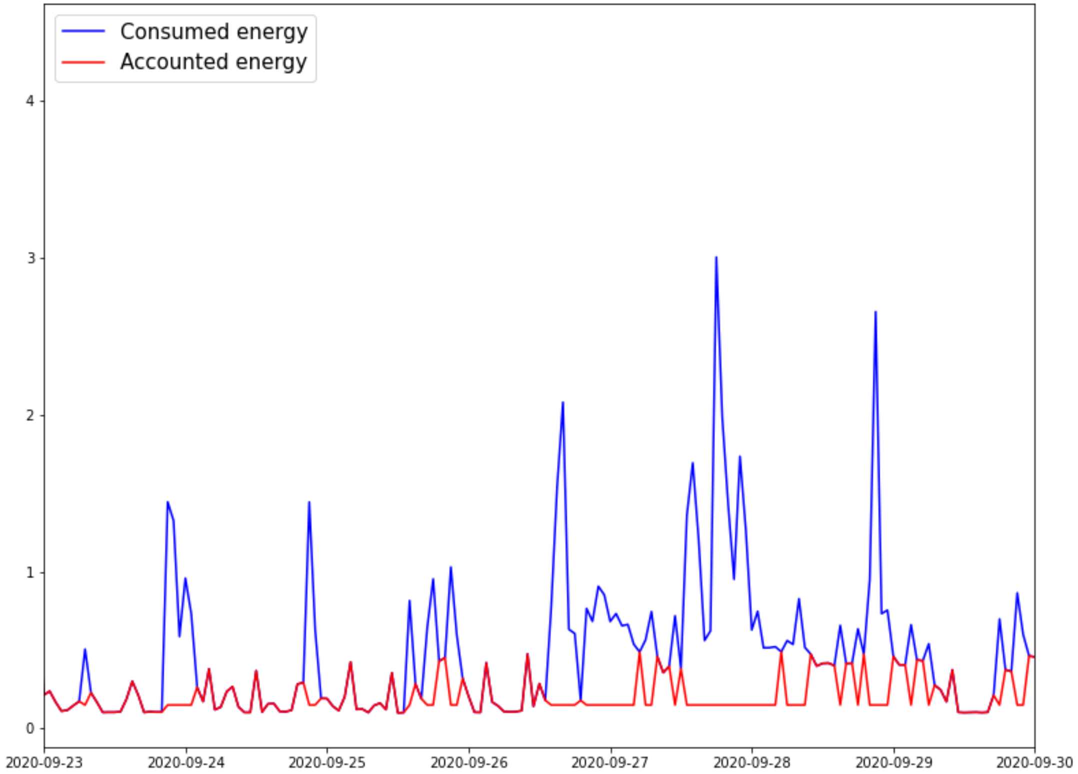
Mainly theft and fraud

Fraud in Distribution Grid





Direct Connection



False Data Injection

False Data Injection

How to detect this type of fraud?

- Upward trend due to SM deployment.
- Many different possibilities.
- Changes in consumption are recorded in the SM, but are more difficult to detect by the operator.
- Possible approaches:
    - Short-pattern clustering.
    - Clustering of all SMs.
- Problem: subtle changes are difficult to detect.

Types	Modification
FDI1	$\widetilde{x}_t \leftarrow \alpha \,.\, x_t$ where $0 < \alpha < 1$ α : same for all reports
FDI2	$\widetilde{x}_t \leftarrow f(x_t)$ $f(x_t) = \begin{cases} x_t & \text{if} \quad x_t \leq c \\ \widetilde{c} & \text{if} \quad x_t > c \end{cases}$ c : cut-off point $c_{min} < \widetilde{c} < c$: randomly defined
FDI3	$\widetilde{x}_t \leftarrow \max(x_t - c, 0)$ c: fixed value in kWh
FDI4	$\widetilde{x}_t \leftarrow f(t) \,.\, x_t$ $f(t) = \begin{cases} 0 & \text{if } t_i < t < t_f \\ 1 & \text{otherwise} \end{cases}$ $t_f - t_f$: randomly defined each day
FDI5	$\widetilde{x}_t \leftarrow \alpha_t \,.\, x_t$ where $0 < \alpha_t < 1$ α_t : randomly defined for each report
FDI6	$\widetilde{x}_t \leftarrow \alpha_t \,.\, \bar{x}_t$ where $0 < \alpha_t < 1$ $\bar{x}$: average consumption of previous month α_t : randomly defined for each report

FDI examples [2]

Correlation between SM consumption and NTL curve can help with this.

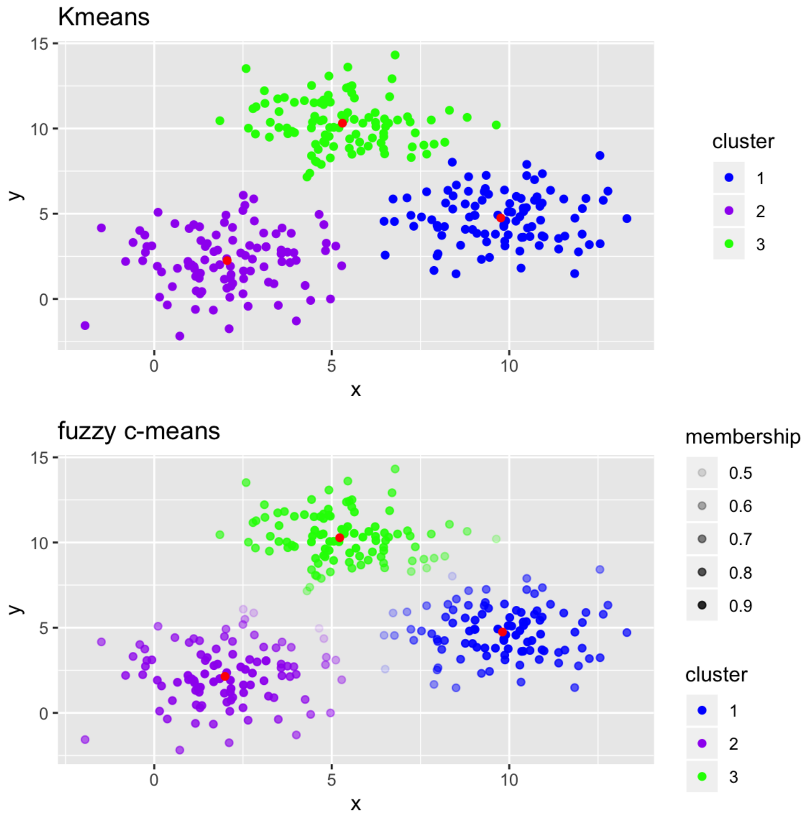
False Data Injection

Clustering

Division of the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. Unsupervised learning.

Features used: Mean, Median, Std, Max, Min, Mean P1 (0-8h), Mean P2 (8-16h), Mean P3 (16-24h), number of times the consumption crosses the average, steep rise, steep descent, peaks average.

Algorithms used: K-Means and Fuzzy C-Means with Euclidean Distance.



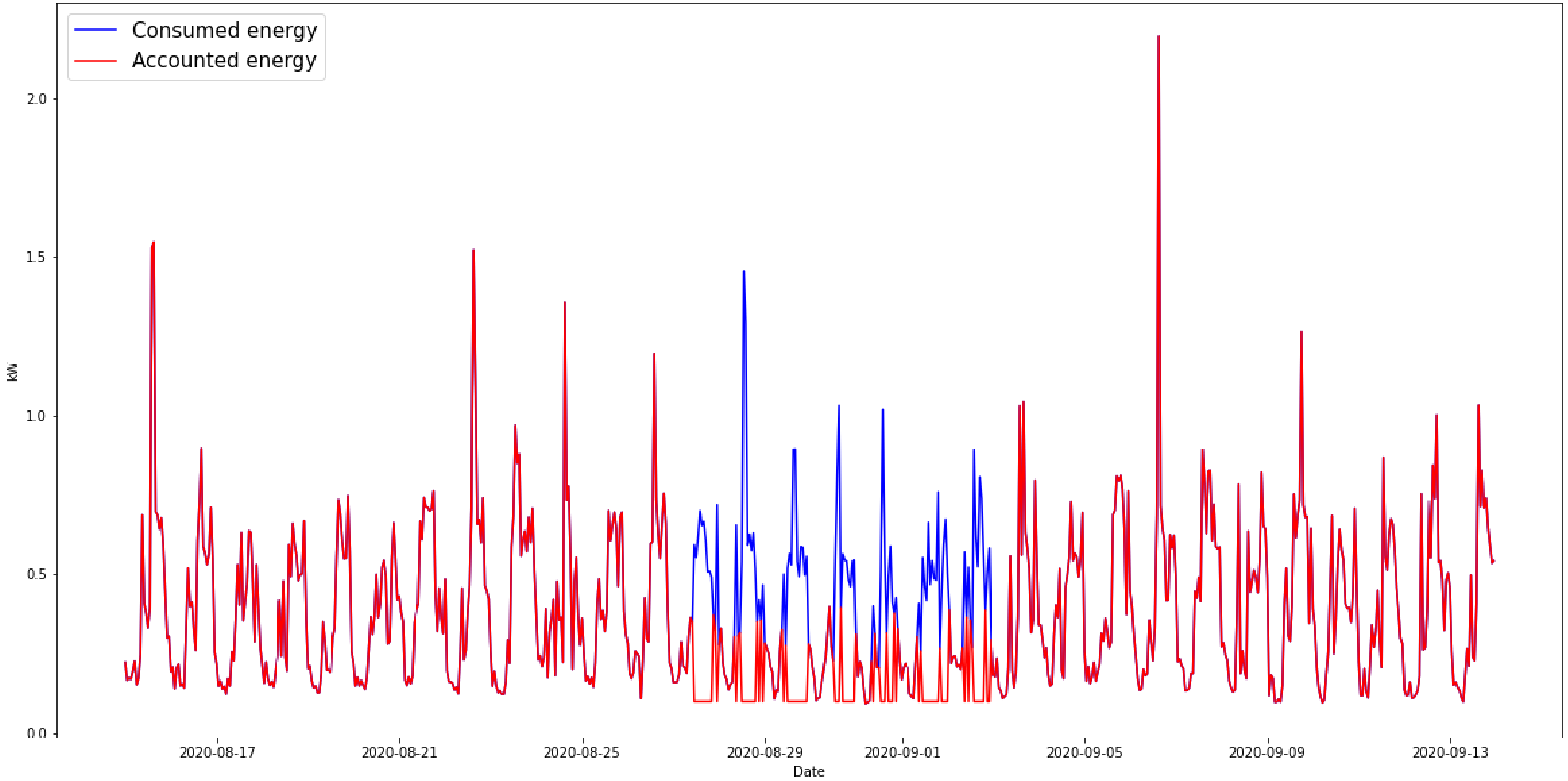
K-Means vs. Fuzzy C-Means. [3]

Why is Fuzzy C-Means interesting?

False Data Injection

FDI 2* – Examples

A. LGZ003XXX648



$$f\,(x_t) = \begin{cases} x_t & if \quad x_t \leq TH \\ c & if \quad x_t > TH \end{cases}$$

TH = 0,4 kW

C = 0,1 kW

Period of fraud:

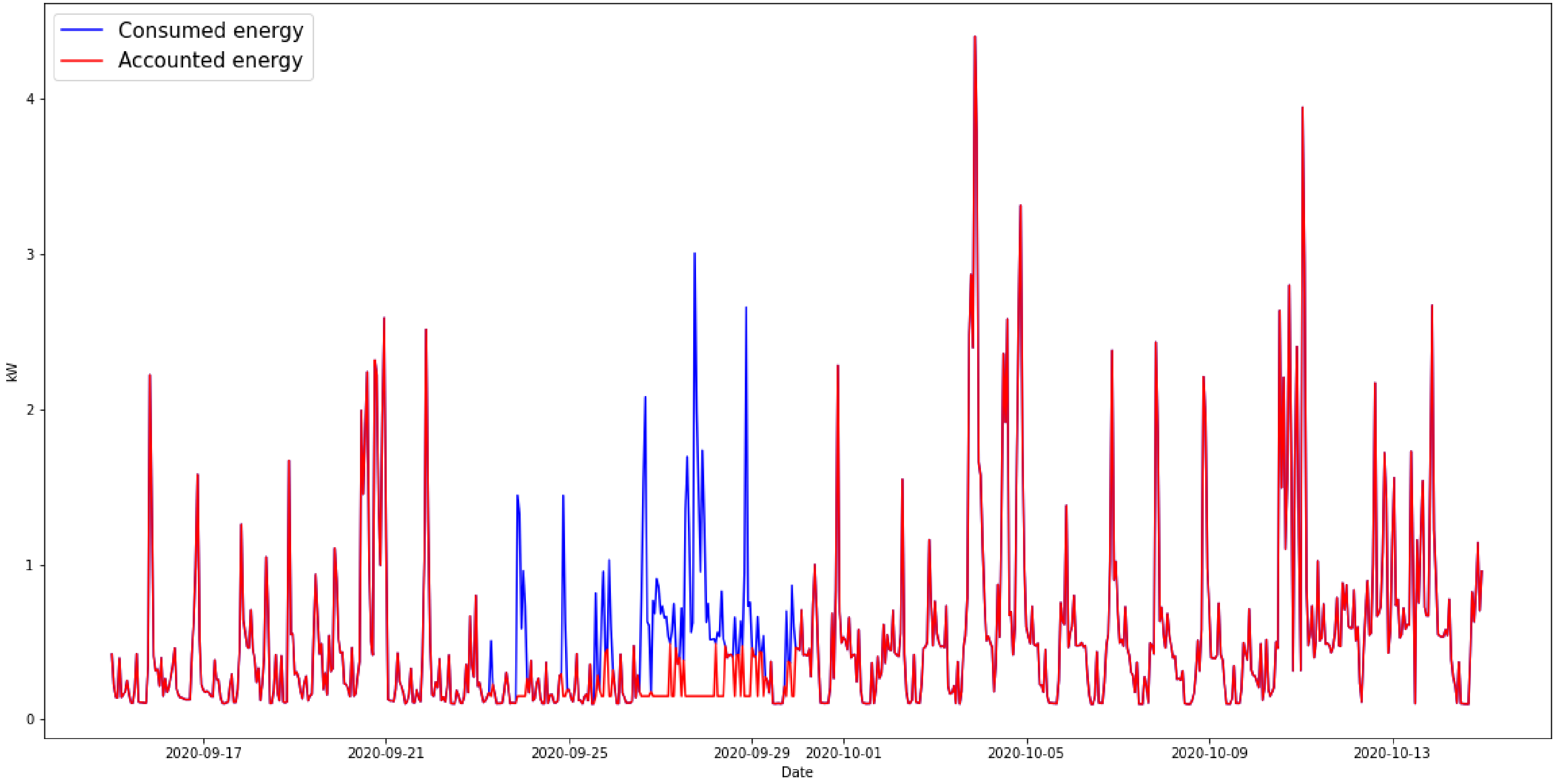
27/08/2020 – 02/09/2020

BD4OPEM

False Data Injection

FDI 2* – Examples

B. SAG01XXX64



$$f\,(x_t) = \begin{cases} x_t & if \quad x_t \leq TH \\ c & if \quad x_t > TH \end{cases}$$

TH = 0,5 kW

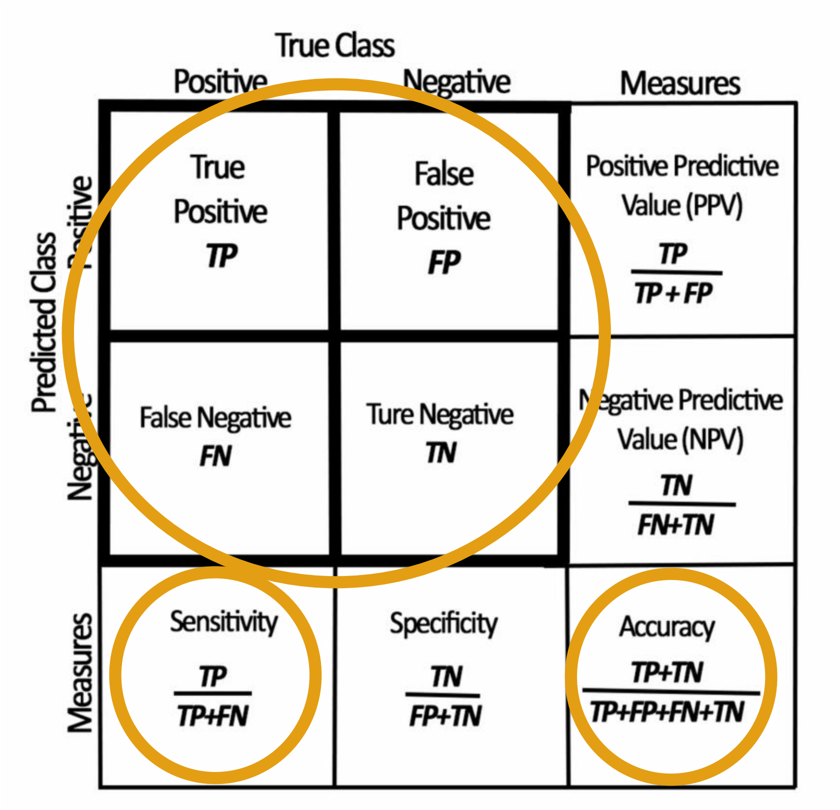
C = 0,15 kW

Period of fraud:

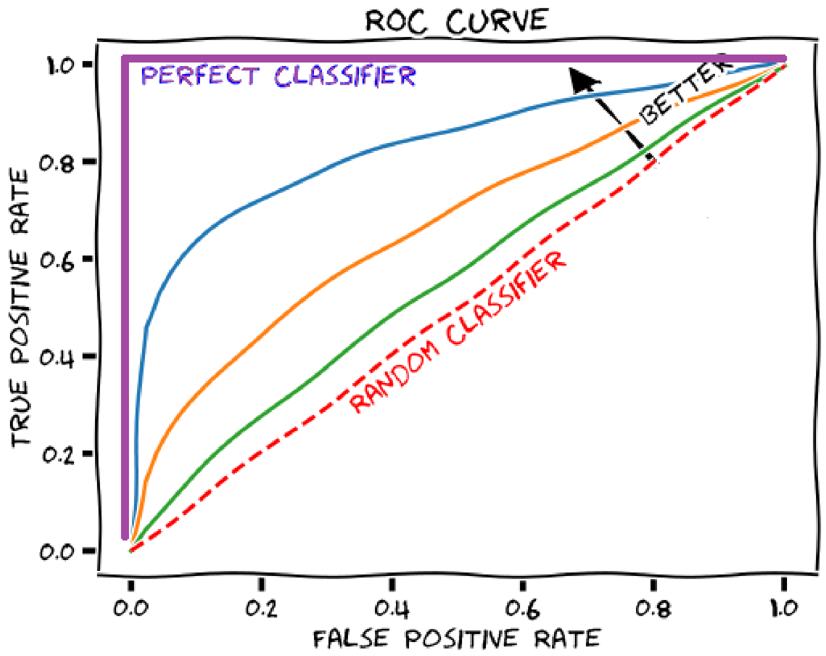
23/09/2020 – 30/09/2020

False Data Injection

KPIs Used



Binary classification KPIs. [4]



AUC-ROC curve. [5]

False Data Injection

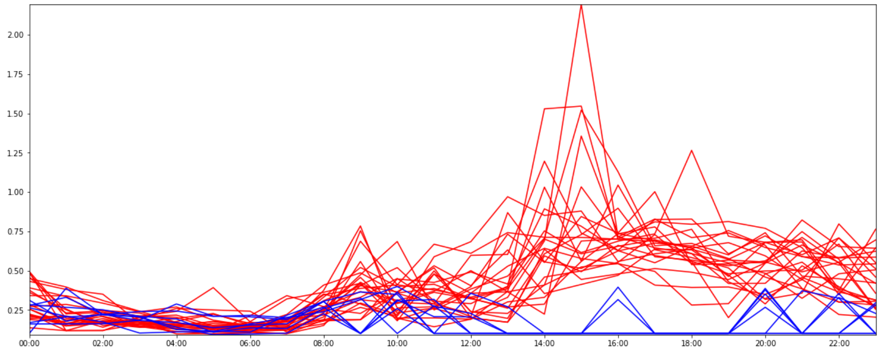
Results - A. LGZ003XXX648

K-Means

		True Class	
		Positive	Negative
Predicted Class	Positive	7	0
	Negative	0	23

Sensitivity = 1

Accuracy = 1



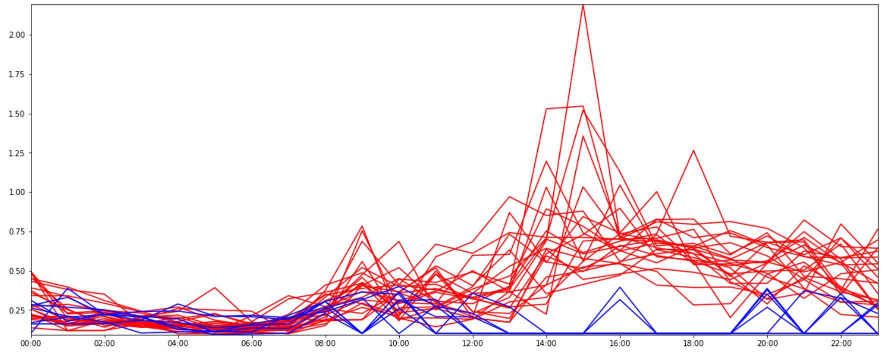
Fuzzy C-Means

		True Class	
		Positive	Negative
Predicted Class	Positive	7	0
	Negative	0	23

Fraud Membership = 0,5

Sensitivity = 1

Accuracy = 1



BD4OPEM

False Data Injection

OQ: How should we define the fraud membership degree in order to optimize the clustering?

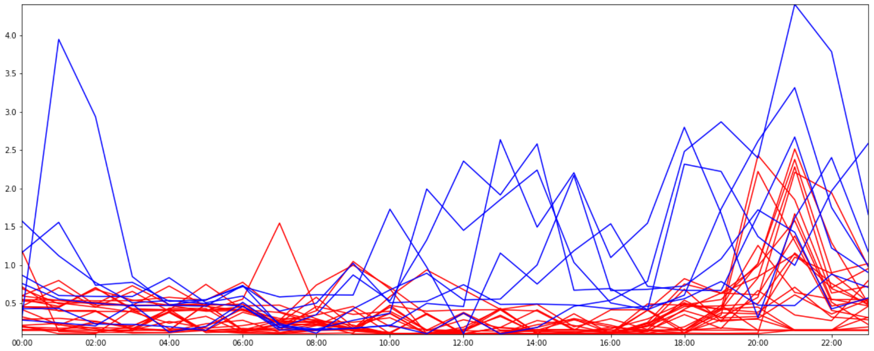
Results - B. SAG01XXX64

K-Means

		True Class	
		Positive	Negative
Predicted Class	Positive	7	16
	Negative	0	7

Sensitivity = 1

Accuracy = 0,47



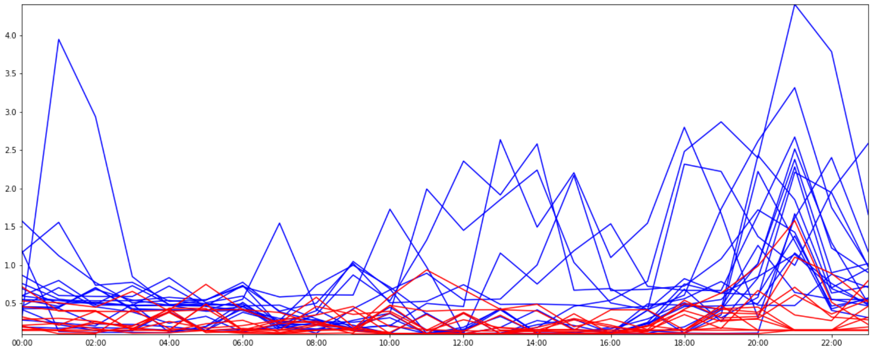
Fuzzy C-Means

		True Class	
		Positive	Negative
Predicted Class	Positive	7	5
	Negative	0	18

Fraud Membership = 0,85

Sensitivity = 1

Accuracy = 0,83



Ethical Questions of the Service

- AI ethics is a field of increasing relevance.
- Generally, in technical services, it might not be taken into account.
- Fraud Detection involves people, so social aspects are also part of the service.

OQ: Should geospatial data be included to predict frauds? Location is closely related to wealth and race - danger of malicious feedback loop.

OQ: Why are these frauds happening? Not all of them should be considered the same – energy poverty shouldn't be criminalized.

Thank you for your attention

BD4OPEM

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 872525