

MACHINE LEARNING PROJECT · NEUEFISCHE DATA SCIENCE

Finding **meter fraud** at STEG

A fraud priority list for Tunisia's electricity and gas company, built on real billing data from 2005 to 2019.

Tim Pfeiffer · Irem Baykal · Dhanya Rajendran

THE PROBLEM

Tampered meters cost STEG **about 200 million dinars**

200 M

Tunisian dinars lost because consumers tamper with their meters.

Few

STEG can only inspect a limited number of homes. Every wasted visit costs money.

| The question: **whom should STEG visit first?**

OUR TASK

Give every client a **fraud risk score**

1 Billing history

of every client



2 Computer model

learns what fraud looks like



3 Priority list

highest risk first

The list tells STEG where to send its **limited inspections**.

THE DATA WE LOOKED AT

135,493 clients and 4.5 million bills

135,493

clients with a known fraud result (training data)

4.5 M

bills, about 30 per client

22

facts: 6 per client (region, client category, start date, ...) and 16 per bill (date, price plan, meter, consumption, ...)

58,069

more clients (1.9 M bills) without fraud result, for the final test

| The model needs **one row per client**, so the bills are summarised per client.

FIRST FINDING

Fraud is rare: 5.6% of clients



1 in 18

clients is fraud (5.6%).

94%

right if we said "no fraud" for everyone. And we would find nobody.

| So we do not ask "how often are we right?", but "**how well do we rank the clients?**"

A few signals stand out

Share of clients with fraud, by signal

Number of meters (replaced meter): 4 or more



Different meter codes used: 3 or more



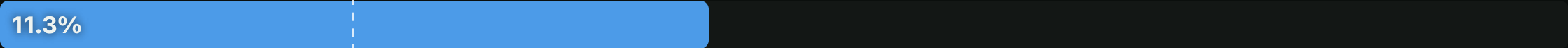
Client category 51



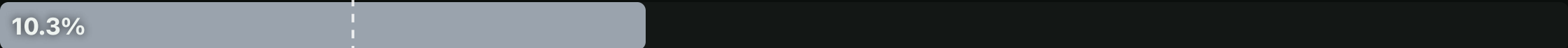
Biggest single bill: 10,000 to 50,000



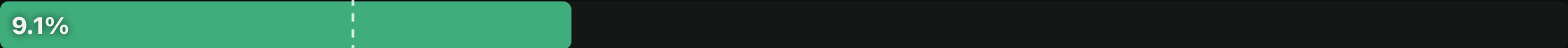
Meter reading goes backward



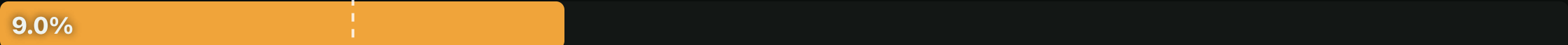
Region 103



Gas: ups and downs in use (most unsteady)



Number of bills: 51 or more



THREE SIGNALS UP CLOSE

What these signals look like

Number of meters

1 meter

2.1%

2 meters

6.2%

3 meters

14.8%

4 or more meters

20.2%

Like a car whose odometer was swapped several times. Something may be going on.

Gas: ups and downs in use

Steadiest gas use

5.1%

Most unsteady gas use

9.1%

One household heats steadily all winter, another one's gas use jumps up and down.

Gas and electricity **behave differently**

7.2%

of gas clients are fraud, against 5.6% of all clients.

Same

result on both meters for region and client category.

9.2%

of electricity clients whose readings do not add up to the billed consumption are fraud (5.5% for the others). On gas: 7.9% vs 7.1%.

Some signals are mainly an **electricity** signal. Many gas groups are small, so we read them with care

OUR APPROACH

From raw bills to a **priority list**

1 Clean and explore

same rules for training and test data

2 One row per client

summarise about 30 bills each

3 Baseline

a simple model as the score to beat

4 Compare models

four models, same fair test

5 Tune

search for better settings

6 Learn from mistakes

which fraud cases do we miss, and why?

7 Test new signals

keep only what really helps

8 Final model

trained on all labelled clients, one submission

An **early split** and fair tests keep the result honest.

RESULT: RANKING QUALITY

From 0.757 to 0.881

Ranking quality (ROC-AUC) on held-out clients. 0.5 = coin flip, 1.0 = perfect.

Simple baseline (logistic regression)

0.757

LightGBM, default settings

0.767

LightGBM, tuned

0.774

+ time and meter features

0.858

+ meter setup variety (final model)

0.881

Pick one fraud client and one honest client at random: the better the score, the more often the model puts the fraud client higher on the list. Bars start at 0.5 (coin flip). Hover or tap for details.

KEY LESSON

Better **signals** beat better settings

Gain in ranking quality (ROC-AUC) per step

Time and meter features

+0.084

Meter setup variety features

+0.022

Tuning the settings

+0.007

Tuning again

+0.001

+0.106

from new signals (+0.084 and +0.022)

+0.008

WHAT IT MEANS FOR STEG

Visit the top 10%, find **57% of the fraud**

Random checks (flip a coin)

10%

of all fraud cases found, about 150 of 1,513

1 in 18

visits is a hit (5.6%)

0.5

ranking quality (ROC-AUC): a coin flip

Our model

5.7x

57%

of all fraud cases found: 864 of 1,513

5.7x

1 in 3

visits is a hit (32%)

0.881

ranking quality (ROC-AUC); Zindi test: 0.874

Same 2,710 visits (top 10% of 27,099 held-out clients, 1,513 fraud). Random checks: expected values from the 5.6% fraud rate, not a measured test.

Zindi score **0.874**, and what to be careful about

Ranking quality, public leaderboard

4th place, original challenge (public leaderboard)

0.883

Our final model (public leaderboard)

0.874

Top 25% (self-reported)

0.864

Zindi benchmark (coin flip)

0.500

Our own check (0.876) was close to the leaderboard (0.874), so the validation was honest. Comparison is approximate: different competition rounds use different parts of the test set.

Be careful with two strong signals

Number of meters and **time since the last bill** may partly show what STEG did *after* it found fraud: replace a meter, stop visiting closed accounts.

Our data has no date of detection. **Next step: ask STEG** how these columns come about.

**Thank you for your
attention.
Questions?**

Tim Pfeiffer · Irem Baykal · Dhanya Rajendran