ELSEVIER

CrossMark

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 60 (2015) 413 – 421

Procedia
Computer Science

19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

An Approach to Detection of Tampering in Water Meters

Iñigo Monedero[a], Félix Biscarri[a], Juan I. Guerrero[a], Moisés Roldán[b], Carlos León[a]

[a] *Electronic Technology Department, University of Seville, - C/Virgen de África 7, 41011 - Sevilla (Spain)*
[b] *EMASESA - C/Escuelas Pías, 1 - Sevilla (Spain)*

Abstract

Meter tampering is defined as a fraudulent manipulation which implies a service that is not billed by a utility company. It is a lack of consumption control for the utility company and a main problem because they represent an important loss of income. We have developed a methodology consists of a set of three algorithms for the detection of meter tampering in the Emasesa Company (a water distribution company in Seville and one of the most important of the country). The algorithms were generated and programmed after a data mining process from the database of the company and they detect three type of consumption patterns: Progressive drops, sudden drops and abnormally low consumption. The methodology has been tested with in situ inspections of the customers of a village of the province of Seville. Once carried out the inspections by the utility, the inspectors confirmed a good success rate taking into account that the detection of this type of fraud is very difficult because it is a non-invasive technique. Besides, this type of detections is a topic that, if we take a look at the state of the art, there are few references or works.

Keywords: Meter tampering, data mining, fraud.

1. Introduction

Water is one of the most important resources of the world and, therefore, a water utility is one of the most important utilities because water is both essential for the survival of the human race and can act as an important energy resource. Governments invest a great quantity of money to ensure the distribution of potable water to the inhabitants of cities and towns; as a result, it is very important to establish a good infrastructure for water distribution, from the place in which it is stored and/or gathered to houses. This infrastructure is very expensive and, in some places, it is necessary to establish some means to control the consumption.

doi:10.1016/j.procs.2015.08.157

Meter tampering is defined as a fraudulent manipulation which implies a service that is not billed by a utility company. This type of losses represents a lack of consumption control for the utility company, as it does not allow for registering of the customer's consumption as well as a proper billing of the service [1]. Although the losses provoked by meter tampering are a major source of loss, the losses can be also due to leaks, meter malfunction or illegal water connection.

The utilities usually install a meter that registers the consumption of the customer in each house or place where the water needs exist. Thus, the utilization of new meter infrastructure, databases and information systems has provided the perfect scenario for the application of data mining and computational intelligence in the analysis of consumers' consumption for detection of meter tampering. Currently, some older water meters are in use, which are more vulnerable to different types of fraud or illegal manipulation. In addition, the new technologies related to smart grids [2], [3] will provide additional information, which will increase the amount of information regarding the consumption of customers.

Aided by the experience in power utilities [4][5], our research group has been working for the whole of 2013 and part of 2014 on the detection of meter tampering in water distribution. Thus, as a result of our work, a methodology for fraud detection in water consumption generated from a data mining process is proposed in this article. The methodology is composed of several algorithms based on the detection of customers with suspect drops in their consumption and abnormally low consumption. Besides, we have added a module of geographical location for the customers detected.

If we take a look at the state of the art, there are relatively few references regarding detection of meter tampering using data mining of the water distribution data. Thus, [6] proposed the utilization of remote systems to increase the control in the distribution network and to model the behaviour of the network from the demand and losses viewpoint; thus, the author used pressure and flow sensors. The proposed solution is applied in a real case in a water distribution network in Herzegovina (Bosnia and Herzegovina). In general, several references exist that are dedicated to water demand modelling, a good example of which is the model proposed by [7] in which the author describes a model of the different consumption levels (residential and non-residential) and the losses model, which includes well-defined examples and categories supported by statistical data from several countries. On the other hand, [8] proposed an application for fraud detection in water distribution based on a Supervisory Control And Data Acquisition (SCADA), which uses a data-driven Decision Support System (DSS), applying a water flow active control in the water distribution network.

2. Meter Tampering in Seville (Spain)

We worked with the Emasesa Company (Empresa Metropolitana de Abastecimiento y Saneamiento de Aguas de Sevilla), which is the most important water distribution company in Seville and one of the most important companies in Spain.

The problem of meter tampering in water distribution is a well-known fact to all companies in this sector. Thus, sometimes, this illegal manipulation of the meters can be applied in a very short period of time. There are various fraudulent methods as carrying out a direct connection, removing or bypassing the meter. But the main way that the customers have for the illegal manipulation occurs because the older water meters are based on an oscillating piston or disk, which relies on the water to physically displace the moving measuring element in direct proportion to the amount of water that passes through the meter. The piston or disk moves a magnet that drives the register. This type of meter can be manipulated using an external strong magnet to slow down and, even, stop the magnet that drives the register. The detection of this type of fraud is very difficult because it is a non-invasive technique.

Currently, there are additional modules for old meters to avoid this type of fraud based on the common magnets. They include a new fraud detection system based on the detection of magnetic fields. Thus, if the register detects one, it will enable an alarm. The thing is that the implantation of such modules is very expensive, and sometimes, the installation of a smart meter is cheaper than the installation of the methods to avoid the magnet fraud. Thus, Emasesa Company is progressively implementing this type of smart meters in homes, but it is an expensive process and even most customers have the older meters.

Thus, Emasesa Company, with the old meters, registers quarterly the consumption of the residential water meters; however, for other supplies with greater consumption or for locations with smart meters, the consumption is registered more often. The proposed work was applied to consumers with quarterly reports.

There are several consumption patterns regarding meter tampering (including magnet fraud, removing the meter or bypassing the meter) and it was necessary to use several algorithms to identify these types of pattern.

Although the results for the case study of this paper is on a village in particular, for the development and programming of the algorithms we worked with the entire database of the customers of the Emasesa Company. This database covers the Seville customers and a group of villages, representing a total number of 357,920 clients (of which 197,513 were from Seville capital and the rest from villages). From all of those customers, first, we filtered the clients those contracts are relative to points relating to irrigation supplies, churches, public institutions, etc. (because such clients are unlikely to be committing fraud).

3. Methodology for Detection

The methodology is composed of three algorithms based on the detection of customers with the following incidences:

- Progressive drops in their consumption.
- Sudden drops in their consumption.
- Abnormally low consumption.

3.1. Progressive drops in their consumption

An obvious symptom of abnormal consumption for a customer (and thus a means to detect meter tampering) is a decrease in consumption. This type of decrease may be due to an actual decrease in consumption (e.g., due to a change in the number of people living in a home), equipment failure, or voluntary manipulation of the measuring equipment (fraud).

To detect decreased consumption due to fraud, we developed and applied two algorithms to detect consumption decreases. The algorithms that we programmed use as input data the following data from each customer:

- The last 24 reading values, which involve historical data of the last 6 years of consumption.
- The number of people per household.

As a first step for both algorithms, a normalization process was programmed and applied to all of the customers. The process divided each of the readings by the number of people registered in the supply point; subsequently, the value obtained was scaled between 0 and 1. Thus, a more reliable study of water consumption (water consumed per capita) was obtained using this normalized indicator.

In the first algorithm, the aim was to detect customers with progressive decreases in their water consumption, whose pattern could be due to a progressive increase in fraudulent conduct by the user. The basis of this algorithm is the Pearson correlation coefficient. In statistics, the Pearson correlation coefficient (r) is a measure that weights the degree of linear correlation between two variables X and Y. In this case, X is the time parameter (number of readings) and Y is the normalized consumption of the customer.

This coefficient is calculated using the following equation:

$$r = \frac{\sum_{i=1}^{n}\left(\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}$$

The result of the calculation of this parameter is a numerical value ranging from -1 to 1. Interpretation of this ratio is shown in Fig. 1. As observed in the above figure, the value of r approaches +1 when the correlation is positive (thus, higher values mean higher X-values) and the values approach -1 when the correlation is negative.

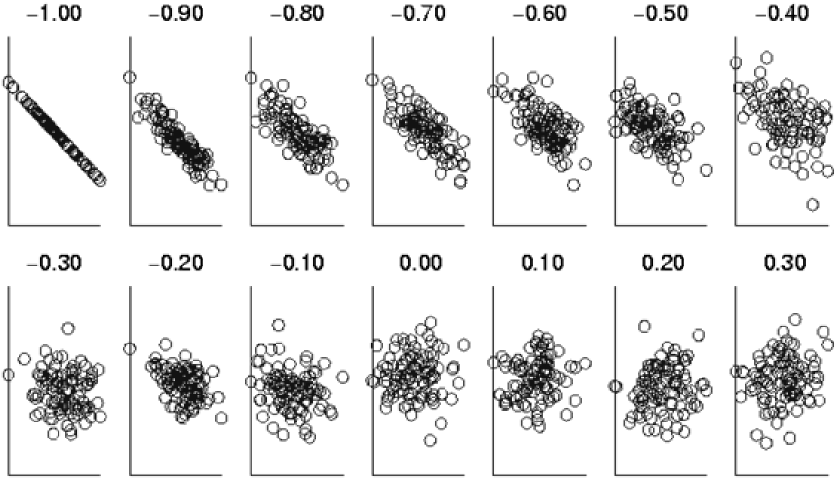

Fig. 1. Interpretation of the coefficient Pearson

We applied this algorithm to the entire sample set and obtained quite high numbers of customers with negative r values. The resulting histogram for negative values of r for the subsample relative to the city of Seville, with a total number of 35,147 customers analyzed, is shown in the following figure (Fig. 2):

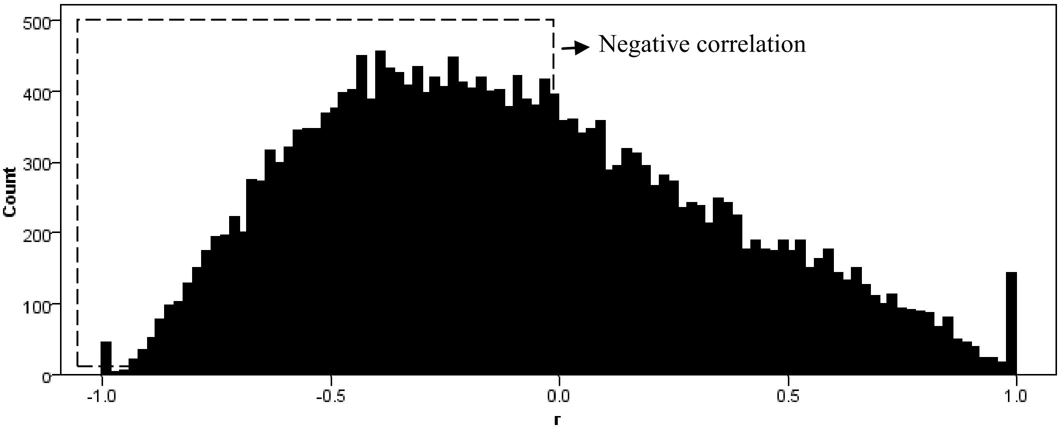

Fig. 2. Histogram of the Pearson coefficient for customers of Seville

As we see there was a majority of users had negative correlation (and therefore its water consumption decreased over time). In particular there was a 70% proportion of customers with negative correlation (24,585 customers), compared to 30% with positive correlation (10.562 customers). Besides, we observed that it was possible to set a high number of customers with strong (and some almost perfect) negative correlation. As a result, for the subsample relative to the city of Seville of 35,147 customers, 3,598 customers were found with an R value below -0.6 (which accounted for approximately 10%), 1,827 had a value of r below -0.7 (which corresponds to approximately 5%) and 626 had a value of r below -0.8 (approximately 2%).

Thus, once seen the previous ranges, we could select the customers to be inspected for the company with a range of negative correlations with a value below -0.6 or below -0.8 depending on the number of customers to be inspected by the company.

Let us consider an example of a real case of a customer of the sample set. The pattern of the following figure (Fig. 3) corresponds to the evolution of the normalized consumption of the example customer.

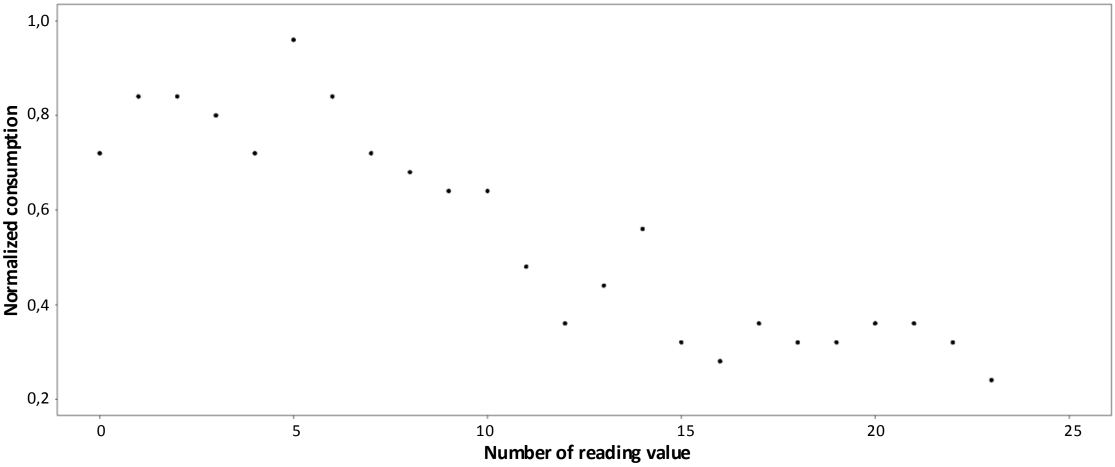

Fig. 3. Example of customer detected with the first algorithm

3.2. Sudden drops in their consumption

The detection of those customers whose consumption had stabilized after an initial drop in consumption is also of interest. Because such clients could not be fully detected using this first algorithm, we developed a second algorithm based on an analysis of windows.

Specifically, we programmed an algorithm that is based on an analysis of the normalized consumption of 6 windows for the 24 reading values of the customer (therefore, four reading values in each window, or equivalently, one year's time in each window).

The search pattern and the representation of windows taken are shown in Fig. 4. The goal, therefore, was to detect a high and constant consumption in one or more of the initial windows, followed by a sudden drop, and finally a low and stable normalized consumption in the latest consumption values.

Two configurable parameters must be configured in this algorithm to detect the sudden decrease between two windows:

- The percentage of decrease in consumption of one window with respect to the other one (Desc).
- The vertical width of the windows of the analysis or, equivalently, the limit on the deviation of the average consumption of each of the windows (Desv).

Specifically, the detections was based on the calculation for the 6 windows of maximum (Max), minimum (Min), median (Median) and mean (Mean) of the values of their normalized consumption. In turn, these window parameters (Max, Min, Median and Mean) were compared with the two previously described configurable parameters (Desc and Desv).

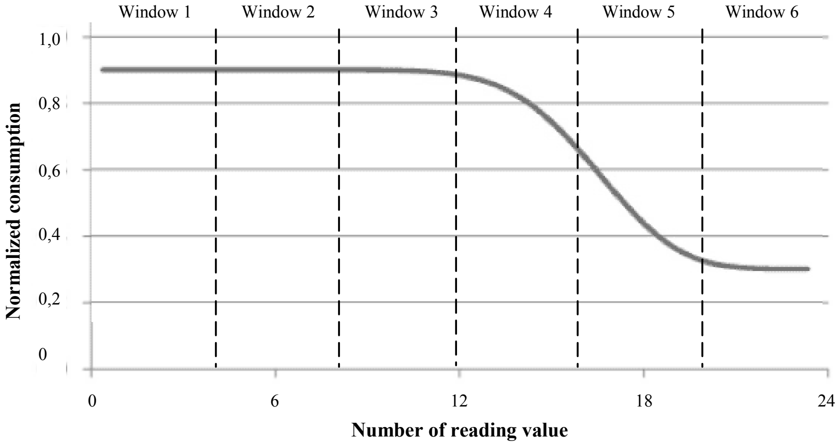

Fig. 4. Pattern of drop consumption

After a data mining process we generated a set of rules. These rules were tested with the sample of Seville Capital and, at the same time, validated by the Emasesa Company (for example, they suggested us the appropriate percentage of decrease in consumption to be detected by the rules). The rules applied were the following ones (the suffix _X of each variable identifies the number of the window that mentions the value):

*((Max_6 < (Median_6 *Desv) and Min_6 > (Median_6 /Desv) and (Max_5 < (Median_5 *Desv) and Min_5 > (Median_5 /Desv) and Mean_6 < Mean_5 / Desc))*
or
*(Max_6 < (Median_6 *Desv) and Min_6 > (Median_6 /Desv) and (Max_5 < (Median_5 *Desv) and Min_5 > (Median_5 /Desv) and (Max_4 < (Median_4 *Desv) and Min_4 > (Median_4 /Desv) and Mean_6 < Mean_4 / Desc) and Mean_5 < Mean_4 / Desc))*
or
*(Max_6 < (Median_6 *Desv) and Min_6 > (Median_6 /Desv) and (Max_5 < (Median_5 *Desv) and Min_5 > (Median_5 /Desv) and (Max_4 < (Median_4 *Desv) and Min_4 > (Median_4 /Desv) and (Max_3 < (Median_3 *Desv) and Min_3 > (Median_3 /Desv) and Mean_6 < Mean_3 / Desc) and Mean_5 < Mean_3 / Desc) and Mean_4 < Mean_3 / Desc))*
)
and
Min_3>0 and Min_4>0 and Min_5>0 and Min_6>0

Setting the parameters Desc and Desv with a value of 2 in both parameters, we covered 1,548 from 35,147 customers analyzed for Sevilla Capital (a rate of approximately 4.5%) or 3,581 with a Desc value of 1.5 (a rate of approximately 10%).

In the next figure (Fig. 5), it is possible observe the pattern of a customer detected using the second algorithm for the sample of Seville Capital.

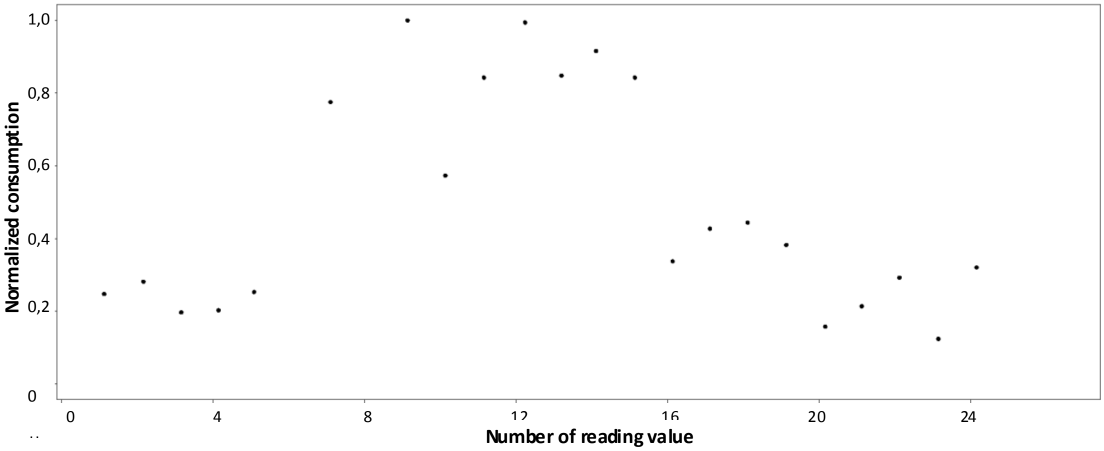

Fig. 5. Example of customer detected with the second algorithm

3.3. Abnormally low consumption

Experts from companies that supply water have estimated a minimum amount of daily water consumption per capita. For Emasesa Company, this minimum value is approximately 114 to 116 litters of water per person per day. This amount is equivalent to a minimum annual consumption of approximately 41 m3 per person (114 litters*person*day extrapolated to 30 days per month and 12 months per year). Although the consumption of water can be very uneven and dependent on multiple factors (closed homes, change of business, spending habits, used appliances, the use of irrigation) and, as previously stated, a low consumption does not necessarily correspond to a meter tampering case or some type of measurement equipment failure, this number (114 to 116 litters of water per person and day) was reached after years of experience of the company, and therefore, it was a very relevant data point to be used in the detections.

Thus, we generated an algorithm based on a rule that covered customers with annual consumption values of less than 10 m^3 in the last year (calculated as the sum of the last four quarterly reading values). This threshold value (10 m3 in the last year) was set after discussions with the company and based on the statistical research of frauds detected in the past. For the calculation of the number of people per household, we used the information collected by the company on the population census and the information registered by the customer in the company (not always consistent).

Once configured, applying the algorithm to the subsample corresponding to Seville Capital resulted in the detection of a ratio of approximately 5% of households, which meant a total number of 1,653 customers from the 35,147 customers analyzed.

4. Results and Conclusions

After performing some initial tests, we performed the first actual execution of the algorithms, with the objective that the Emasesa Company sends inspectors to investigate a set of clients. Emasesa Company selected for this first test of the algorithms a small sample of customers of a village of Seville Province. Thus, the number total of customers of this village used as inputs in the algorithms after the previous filters was 859.

In regard to the configuration parameters of the algorithms: For the first algorithm, we selected the customers to be inspected for the company with a correlation value below -0.8 and also with a number of real readings greater than 10 during the 6 years of the analysis. This second filter was applied to avoid negative correlations, and therefore fraud detections, relative to consumption estimates for the company due that it is common practice for the company when they cannot access to the measuring device of the customer to perform an estimation of the consumption. By other hand, for the second algorithm, we set the parameters Desc and Desv with a value of 2 in both parameters.

We were covering about 4% customers with each of the two algorithms. Thus, from this subsample and with this configuration, the algorithms marked 85 customers as possible fraudulent customers. From those 85 customers, after completing the inspections, the total number of fraudulent customers detected by Emasesa Company with the use of the algorithms was 6. The data obtained with the algorithms correspond to a success rate of approximately 7%.

When inspectors went to their home and asked customers for their consumption drop, the main reason why they justified their drop was the death of a family member. This problem was difficult to control and ascertain by Emasesa Company because the update of the number of people in a household was a data that Emasesa Company received updated once a year.

On the other hand, although this result might appear to be low, these results are good for Emasesa Company conditioned by two facts:

- To date, Emasesa Company in its inspections swept an entire geographical area without making any processing for the selection of customers. Thus, it implies a poor success rate. In addition to improving this success rate, these algorithms could also save hours of inspection to the company.
- The ease of the customers to remove the magnet from the measuring equipment, which reduces the probability that the user is surprised while performing the fraud.
- The algorithms indicated that the analyzed area corresponding to the particular village was not a particularly fraudulent area.

The conclusions obtained after performing the work and the analyzing the results achieved by the inspections are as follows:

- The fraud in this type of business leaves no sign of manipulation, making it quite difficult to detect frauds via inspection. During working hours, fraudulent customers usually remain vigilant because an inspector can detect fraud by surprise. Thus, the inspections by Emasesa Company were performed during non-business hours (usually, early in the morning) in order that the customers were not located in the business premises and they could remove the magnet.
- The algorithms developed detect approximately 7% of the meter tampering. The obtained model provided the Emasesa Company with good support to identify different patterns of customers.
- Having a higher frequency of reading values (currently, four for each year) significantly improves the results. This is due to that sometimes the effect of this manipulation is imperceptible in water meters read quarterly in which detecting drops of consumption is not possible.
- The main solution to improve the results, which today is investing Emasesa Compan, is the installation of smart meters. This type of meters has got two advantages: On the one hand, they are not affected by magnets and, on the other hand, they provide measurement time interval of an hour (And therefore, better resolution for the algorithms).

Acknowledgements

This work has been supported by the Ministerio de Ciencia y Tecnología (Spain) under project "Sistema Inteligente Inalámbrico para Análisis y Monitorización de Líneas de Tensión Subterráneas en Smart Grids" (reference number TEC2013-40767-R).

References

1. Mutikanga, H., Sharma, S.K. and Vairavamoorthy, K. Assessment of Apparent Losses in Urban Water Systems. Water and Environment Journal, 25 (3), 327-335, 2011.
2. Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. Smart meters for power grid: Challenges, issues, advantages and status. Renewable and Sustainable Energy Reviews, 15(6), 2736–2742, 2011
3. Bayliss, C. R., & Hardy, B. J. Chapter 27 - Smart Grids. In Transmission and Distribution Electrical Engineering (Fourth Edition) (pp. 1059–1074), 2012.
4. León, C., Biscarri, F., Monedero, I., Guerrero, J.I., Biscarri, J. Variability and Trend-Based Generalized Rule Induction Model to NTL Detection in Power Companies. IEEE Transactions on Power Systems, Vol. 26. Núm. 4. Pag. 1798-1807, 2011.

5. Monedero, I., Biscarri, F., León, C., Guerrero, J. I., Biscarri, J., & Millán, R. Detection of frauds and other non-technical losses in a power utility using Pearson coefficient, Bayesian networks and decision trees. International Journal of Electrical Power & Energy Systems, 34(1), 90–98, 2012.
6. Obradović, D. Modelling of demand and losses in real-life water distribution systems. Urban Water, 2(2), 131–139, 2000.
7. Ratnayaka, D. D., Brandt, M. J., & Johnson, K. M. CHAPTER 1 - The Demand for Public Water Supplies. In Water Supply (Sixth Edition) (pp. 1–35), 2009.
8. Gouthaman, J., Bharathwajanprabhu, R., & Srikanth, A. Automated urban drinking water supply control and water theft identification system (pp. 87–91). Presented at the 2011 IEEE Students' Technology Symposium (TechSym), 2011.