ManagementSolutions
Making things happen

Fraud management in the energy industry

Energy

www.managementsolutions.com

Design and Layout
Marketing and Communication Department
Management Solutions

Photographs
Photographic archive of Management Solutions
iStock

Content

Introduction

Fraud has become a chief concern for governments and companies. In fact it is estimated that losses from fraud in organizations can be as much as 5% to 9% of their annual profit[1]. To better understand the different fraud management frameworks, it is necessary to have an understanding of what fraud is, its components and the various forms it can take.

In the business world, fraud is associated with an action that goes against truth and integrity, damaging the organization against which it is perpetrated. Fraud can compromise a company, whether it is committed externally by clients, suppliers and other parties, or internally by employees, managers or shareholders.

Some characteristics of the **current environment** and the opportunities it has to offer are as follows:

- Growing availability of **data on customers, employees, suppliers,** etc., their interaction with the company and behavior patterns. Availability of **techniques to analyze and quantify the likelihood or probability** that fraud events will occur.
- Advanced **methodologies and systems to fight internal fraud** through the **segregation of duties (SoD).**

The **value added** by these management mechanisms is reflected in economic terms (according to an ACFE[2] study, losses from fraud at the global level fell by 54% thanks to the adoption of proactive data monitoring measures[3]), and also in reputational and compliance terms. Both these two aspects are particularly relevant given the current regulatory environment, which encourages companies to invest in and implement fraud management methods.

The purpose of this document is to share some insights **on the concept of fraud,** as well as on the key elements used to manage fraud and the opportunities for optimization that arise as a result of technological advances such as Big Data and Analytics. These are based on the availability and analysis of large data volumes as well as the implementation of **profiling and segmentation** methodologies.

[1] Mark Button, Jim Gee, Graham Brooks, "Measuring the cost of fraud: an opportunity for the new competitive advantage", Journal of Financial Crime, Vol. 19.
[2] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study: "proactive data monitoring was associated with 54% lower losses and frauds detected in half the time". Analysis of a total of 2,410 occupational fraud cases across the world in 2016 (48% in the US)
[3] Including data analysis, supervision of directors/managers, setting up contacts to receive reports, surprise audits, etc.

With a particular focus on the Energy industry, this document describes **fraud events specifically for this industry** which, due to their representativeness and the fact that they drag on the resources of companies, require specific treatment and for which detection techniques and the integration of these techniques into the management process are even more relevant.

- With regard to external fraud, energy companies that distribute electricity and/or natural gas are exposed to energy theft through fraudulent network connections or access. Managing this type of fraud requires the support of methods that will quantify the probability of a meter reading not reflecting the actual amount supplied. Different methods are used (such as logistic regressions, neural networks, decision trees, etc.), which are embedded into machine learning schemes and focused on discriminating between "reasonable" and potentially fraudulent supply amounts. These techniques use variables that characterize the customer, the customer's energy consumption profile, behavior patterns, etc. in order to identify profiles or behavior that may indicate a propensity for energy theft (e.g. recurrent behavior). This document will not go into detail regarding the treatment of cyberattacks. However cyberattacks do pose a threat in relation to identity theft and communication interference, for instance generating supply interruptions.
- As for internal fraud, the main concern is the loss associated with fraud events in processes that are critical for the company, such as the commercial cycle of an energy distribution company. These events normally take place in the invoicing and collection processes, in which the possibility to alter usage, amounts, purchasing processes or bank details may result in theft of company revenues. This type of fraud is managed through the use of methodologies oriented towards the segregation of duties, controlling access to commercial and financial systems and defining indicators and reporting schemes to warn companies about any breaches to the segregation of duties.

In addition, this document will show how the modeling, profiling and segmentation methods complement the implementation of a **methodology for quantifying the economic usefulness of actions,** a methodology that discriminates the quality of the segmentation performed for fraud detection purposes (the effect of the segmentation models), from the appropriateness of implementing the actions (or the effect of the detection campaigns themselves), with the aim of separately **assessing the cost effectiveness of investing in modeling techniques and investing in theft detection inspections.** In this sense investment in fraud management is considered as just another company investment.

These techniques are supported by **modeling platforms** that combine mass data processing components with statistical software and **tools for both access control** and the management of roles, incompatibilities, etc.

Finally, this publication includes some **examples to illustrate the implementation** of energy theft detection probability modeling techniques (a specific case of external fraud). These models are based on the characterization of the point of supply using variables that identify the factors underlying fraud, such as the physical characteristics of meters, commercial and socio demographic characteristics of customers or users, usage and behavior history in relation to theft, other transactions with the customer, customer engagement, claims, the result of inspections, etc.

What is shown is therefore the **added value of data,** of information on costumers and transactions (hourly consumption, customer data, access to systems, etc.), in quantifying the probability that fraud events will occur and the use of this calculation to optimize both preventative action (e.g. segregation of duties and system access control) and mitigating action (e.g. implementation of inspection campaigns and segmentation of profiles according to their propensity to theft). Thus, **action can be prioritized under an economic and profitability rationale** based on the estimated probability that a theft event will occur or the possibility that fraud may be committed in the commercial cycle, as well as the materiality of the potential impact (energy defrauded, amounts stolen, etc.).

In fact, according to data made available by one of Europe's main electricity distribution companies, following their use of data collected from intelligent meters, the percentage of fraud cases affecting the company that were detected went from 5% to 50%[4].

[4] Fragkioudaki, A. et al. (2016). Detection of Non-technical Losses in Smart Distribution Networks: A Review.

Executive summary

Situational considerations

1. While there is no single definition, for the purposes of this document we shall define **fraud** as any intended action or omission to defraud others, resulting in a loss for the latter and/or a gain for the defrauder[5].

2. The most common fraudulent practices fall under one of two groups: **external fraud** (e.g. theft, identity theft, cyberattacks, etc.) and **internal fraud** (accounting fraud, fiscal fraud, transaction for personal benefit, etc.[6]).

3. There are three **factors** which, if they occur at the same time, will increase the probability that a person will perpetrate fraud:

    - Need or pressure, either economic or of a different nature. There must be an incentive or a need (internal) or pressure (external), that will incite or motivate an individual to perpetrate fraud.

    - Perceived opportunity. For a fraud event to take place there must be a weakness that can be exploited in a specific process. The individual perceives a way to solve his problems fraudulently or with a low risk of being discovered.

    - Rationalization/Attitude. Justification of the delinquent action. This is influenced by the individual's values, the perception the individual has of the ethical principles governing the company (on which fraud is perpetrated), and the weighing of the benefit derived the fraudulent action against the potential adverse consequences in the event of being discovered.

4. Upon analyzing the fraud **events** reported by each industry, it is significant the percentage of incidents associated with the banking/financial industry, which continues to be the industry with the largest number of cases. Nevertheless, after an analysis of the average loss per case for each industry, it is the mining and wholesale trade industries that show the largest average losses, with the banking industry in an intermediate position (with respect to the industries analyzed in a study conducted by ACFE[7]).

5. As for the **energy industry,** while it shows fewer cases according to the same study (around 5% of cases analyzed were mostly in utilities and oil&gas corporations), it has some peculiarities in relation to energy theft for which modeling techniques can provide differential value.

6. The process of **digital transformation** in which all industries are immersed implies a greater level of exposure to fraud risk, since technological advances are used by fraud perpetrators to adopt new strategies not yet included in companies' prevention, detection and action plans.

7. Due to the **changing nature of fraudulent practices,** their detection is an ongoing and dynamic process that requires organizations to define a framework for action that includes strategy, specific approaches and policies, and for all areas involved to act in a coordinated manner. In this context, companies have increasingly been building fraud management policies which assign management and control responsibilities depending on the fraudulent event.

8. This makes the implementation of a **fraud management framework** all the more relevant. A framework whose complexity can vary from simple initiatives for deploying tactical controls (authorization and validation processes, alarms, inspections etc.) to the implementation of global projects which, covering most company processes, seek to establish metrics to measure the risk of fraud in such projects and to modify the processes and systems themselves to mitigate this risk (implementation of platforms for duty segregation, access control, fraud probability modeling, information systems and fraud measurement reporting schemes, etc.).

[5] The Institute of Internal Auditors (IIA), The American Institute of Certified Public Accountants (AICPA) and Association of Certified Fraud Examiners (ACFE) (2012): Managing the Business Risk of Fraud: A Practical Guide

[6] In accordance with the Basel II framework (BCBS: International Convergence of Capital Measurement and Capital Standards), applicable to the financial industry (though this definition is of general applicability).

[7] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study.

9. Fraud management **functions** in companies tend to be **dispersed**, with the relevant business area usually being the promoter of fraud management initiatives, supported by the IT areas that provide services to the former. However, partly due to the development of compliance functions within companies (as it is the case with the CCO or Chief Compliance Officer) and also to the search for economic efficiency and process optimization, there is currently a trend to centralize fraud control mechanisms, methodologies and indicators.

10. Fraud management has been reinforced by **intelligent systems** and **statistical analysis** that help detect it. These techniques, which allow companies to identify the new strategies and patterns used by perpetrators, combine pure analysis and modeling components like data mining and machine learning with high-performance technical elements like stream computing and also full data transformation processes for the acquisition of useful knowledge, like Knowledge Discovery in Database (KDD).

Fraud management techniques applied to the energy industry

11. The increased capacity to generate, store and process **information** can be used to gain real time access to a customer's characterization, a transaction, a process, etc., which makes it possible to identify behavior that may indicate a propensity for energy theft. The use of Data Science to detect fraud in the energy industry is of great help, for instance, to differentiate what percentage of the energy lost in the distribution network relates to a technical loss (not arising from a fraud event) and what percentage relates to a non-technical loss (energy theft and therefore involving a fraud event). In addition to this, investing in advanced detection models optimizes the success rate of inspection campaigns and improves fraud detection. In any case, to ensure progress is made in the implementation of this type of models, it is advisable to previously define and implement a reference framework that will facilitate governance of the associated data, models and processes.

12. Models to be developed are expected to find patterns, trends or rules that explain the behavior of customers before fraud is detected. The techniques used vary depending on the end goal and the type of data used. The potential of Data Science techniques is taken full advantage of through two factors:

    - Real time analysis. Information collection systems make it possible to capture and track data in real time. It is also possible to program algorithms that will use that information, which facilitates and speeds up the detection of new fraud patterns and strategies not yet used by perpetrators.

    - Automatic retraining and self-learning. Fraud detection models are recalibrated automatically (with little intervention by analysts) and iteratively from large data volumes, which translates into potentially improved predictive power during subsequent retraining.

Example of modeling applied to energy theft

13. One of the most common Data Science uses in fraud management within the energy industry is maximizing the efficiency of **inspection** campaigns. Energy fraud detection in connection with the illegal use of energy from the network relies on segmenting customers based on **how likely they are to perpetrate fraud.** Modeling techniques can be used in energy fraud detection management to improve success rates in the selection of customers to be inspected.

14. Some **variables** with proven high predictive power are: meter data, sociodemographic data, historical energy consumption data, data relating to the transaction or contact made, network and meter maintenance, information on cuts and irregularities and information on customer contact or claims, etc.

15. In order to identify the **models** that best explain fraudulent customer behavior, some minimum criteria are set which the results obtained through the selected model need to meet, such as their discriminatory capacity. In addition to statistical validation, models are sanctioned via the inspections conducted over a six month period, and it has been observed that, regardless of the number of inspections, it is the machine learning techniques that make it possible to generate segments with a larger concentration of fraudsters.

16. The non-technical loss management areas of energy companies invest in human, technical and economic resources to implement customer inspections. The **cost-effectiveness** of these investments is determined by:

    - The observed success rates (the percentage of energy theft customers identified out of the inspected customers sub-group);
    - The energy gain (represented by the amount recovered per customer);
    - The number of inspected customers in the target population;
    - The unit cost associated with inspecting the customers and therefore the total cost of the campaign.

17. A **quantitative exercise** was conducted using the described methodology, and a specific algorithm was selected due to its greater discriminatory power. This model was used to develop an example of practical implementation for the configuration of inspection campaigns. In the example, the inspection success rate reaches 27%, representing a three-fold increase (over one in four inspections are successful).

Fraud management

Context

Fraud concept

There is no unified and standardized definition of what practices constitute fraud. In fact, a concern of international organizations is precisely to define a unified legal framework that will establish common criteria on the practices that should be considered fraudulent, and the sanctions to be applied in each case.

However, ACFE[8] defines fraud as ***"any purposeful action or omission to defraud others, resulting in a loss for the latter and/or a gain for the defrauder"***[9]. In the corporate context, fraud is associated with an action that goes against truth and integrity and damages the organization against which it is perpetrated. Fraud, whether it originates externally from clients, suppliers and other parties, or internally from employees, directors, civil servants or company shareholders, can compromise a company.

The most common fraudulent practices can therefore be grouped under one of two areas: external fraud (such as theft, identity theft, cyberattacks, etc.) and internal fraud (accounting fraud, fiscal fraud, transaction for personal benefit, etc.). External and internal fraud are usually considered as two different operational risk categories that companies identify, measure and manage.

External fraud is considered to be an event that causes an unexpected financial, material or reputational loss as a result of fraudulent actions perpetrated by a person outside the company. It is defined as "Losses due to acts of a type intended to defraud, misappropriate property or circumvent the law, by a third party"[10].

Fraud can be perpetrated by customers, or by competitors or third parties:

- Customers that use goods or services fraudulently - without paying for them, falsify means of payment, manipulate purchasing processes, etc.
- Competitors, suppliers and third parties in general, that manipulate tenders, invoice the company for goods or services not provided, offer bribes to employees, etc.

Under the second category, organizations particularly face security breach threats and intellectual property theft perpetrated by unknown third parties (e.g. through cyberattacks). Other examples of fraud are piracy, theft of confidential information, fiscal fraud, fraudulent bankruptcy, insurance-related fraud, medical assistance-related fraud, etc.

Internal fraud takes place inside an organization and is perpetrated by its employees against the organization/employer. It is defined as *"losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the law or company policy, which involves at least one internal party acting for personal benefit"*[11]. It usually arises from a conflict between an employee or group of employees' personal interest and that of the organization. It is estimated that, on average, 5% of an organization's annual profits are lost as a result of internal fraud[12].

Factors underlying fraud

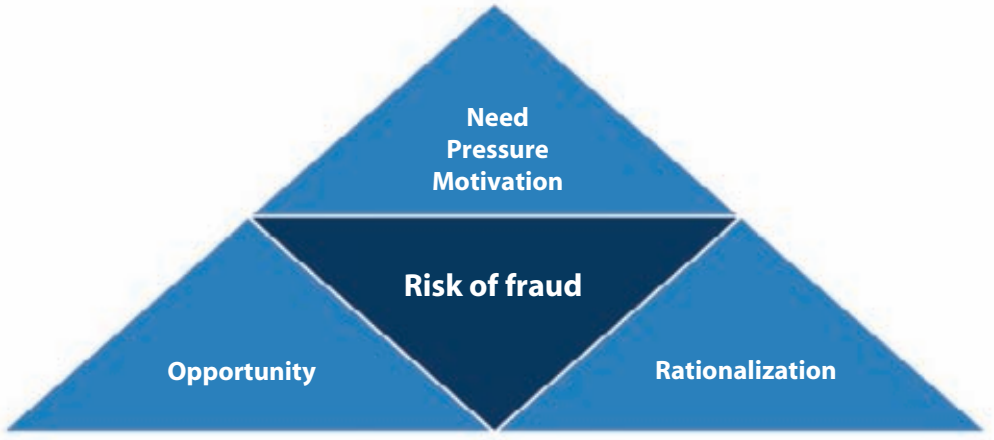
When it comes to managing fraud-related risk, companies need to identify and monitor the different factors that can trigger a fraud event. One of the tools used to assess the risk of fraud is known as the *"fraud triangle"*[13] (see fig.1). This tool is a widely accepted model for explaining the underlying factors that motivate a person to defraud.

According to this theory, there are three factors which, if occurring simultaneously, will increase the probability that a person will commit fraud:

- **Need or pressure.** Motivates the offence in the first place. It represents the pressure the person is under as a result of a problem they cannot resolve through legitimate means,

[8] Association of Certified Fraud Examiners, an organization for the provision of fraud prevention related training at the international level.
[9] Source: "Fraud Risk Management in Organizations: A Practical Guide." IIA, Institute of Internal Auditors.
[10] In accordance to the Basel definition: BCBS: 'International Convergence of Capital Measurement and Capital Standards'. June 2004.
[11] As defined by Basel II.
[12] Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study. ACFE, Association of Certified Fraud Examiners.
[13] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study

Fig. 1. Fraud triangle



which leads them to consider illegal methods as a way to solve their problem. This pressure may be economic or of a different nature. There must be an incentive or a need (internal) or pressure (external) that causes or motivates the individual to defraud.

- **Perceived opportunity.** Defines the method that will be used to commit the offence. For fraud to take place, there needs to be a weakness that can be exploited in a particular process (e.g. no controls, little segregation of duties or lack of an appropriate and updated fraud management system). The individual perceives there is a way to solve their problems fraudulently with little risk of being discovered.

- **Rationalization/Attitude.** Justifies and validates the offence; this refers to the individual's ability to internally rationalize and justify the incorrect action the fraudulent act implies. This is influenced by the individual's moral values, the perception the individual has of the ethical values governing the (defrauded) company, and how the individual weighs the benefit obtained from the defrauding action against the potential negative consequences if he is discovered. Both the individual's honesty and attitude to risk are fundamental and determining aspects.

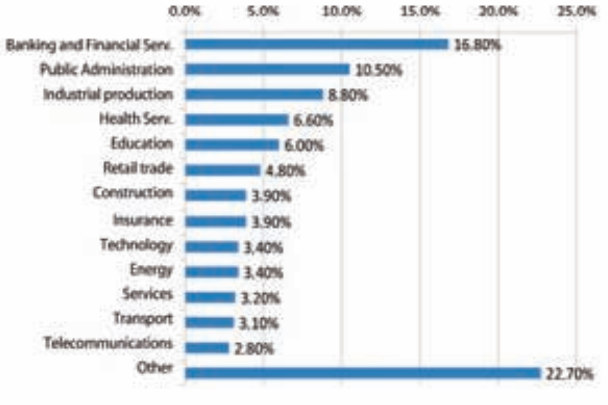
Fraud by activity sector

Fraud is a a problem common to all industries, which reinforces the need to establish controls and mechanisms to minimize these events.

An analysis of fraud events reported by each industry reveals that the banking/financial industry is still the one with the highest number of fraud cases (see fig. 2).

If we look at the average loss associated with each case by industry, the mining and wholesale trade sectors are those with the highest average losses, with the banking sector somewhere in between (with respect to the industries analyzed in the ACFE study). Though, according to the same study, the energy industry has fewer cases (around 3.4% of analyzed cases fall under utilities and oil&gas corporations), it has some peculiarities in relation to energy theft for which modeling techniques can provide differential value.

Fig. 2. Distribution of cases by activity sector[14]

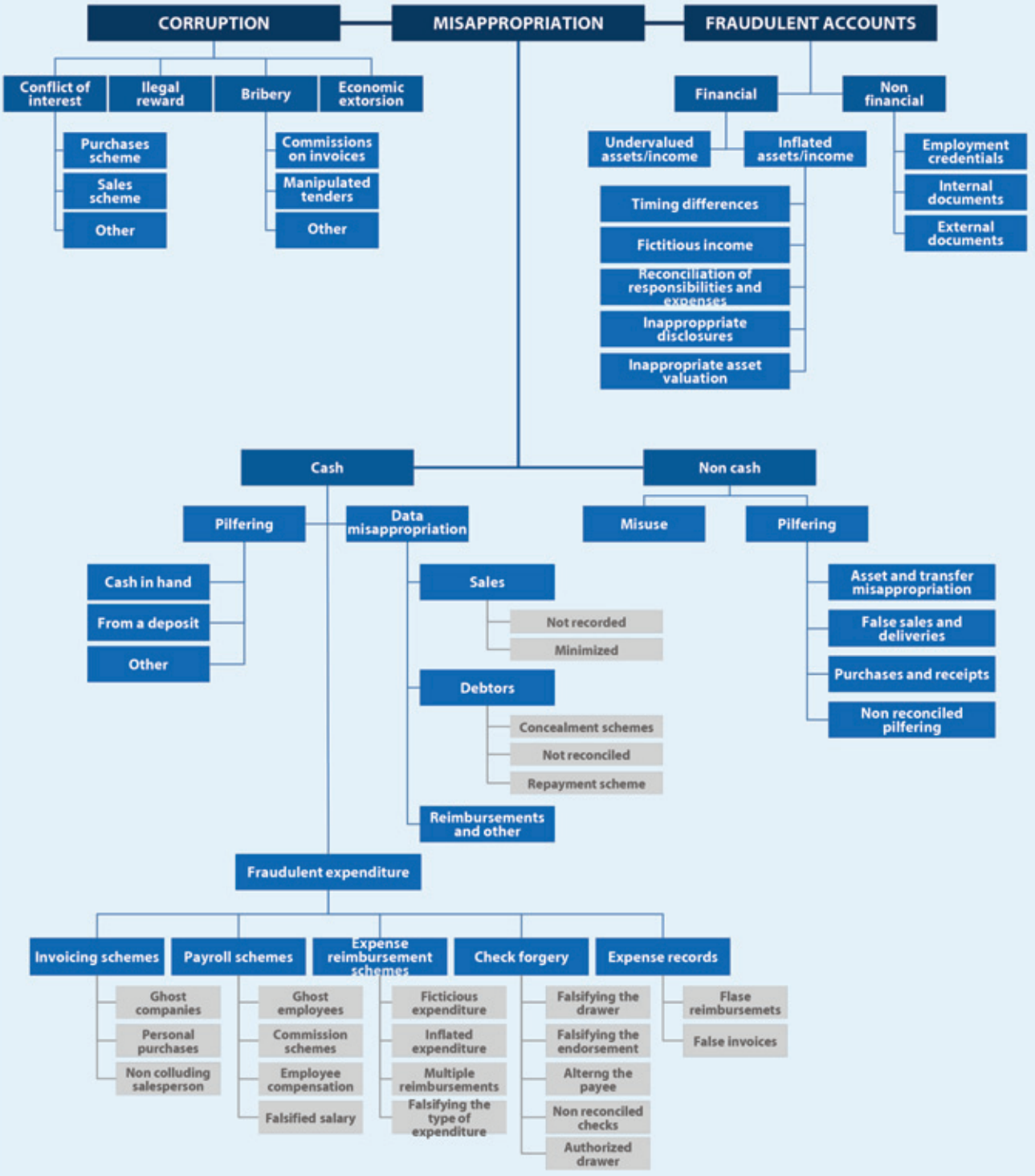


Source: 2016 Global Fraud Study: Report to the nations on occupational fraud and abuse (ACFE)

[14] Other: Professional and Social Services, Agriculture and Fishing, Real Estate, Utilities, Art and Entertainment, Wholesale trade, Mining and Communications..

Professional fraud and abuse classification system

This type of labor fraud can be classified according to the following tree or fraud scheme[15]:



[15] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study.

An element common to all industries is the process of **digital transformation** in which they are immersed. This implies a greater level of exposure to the risk of fraud, since technological advances are taken advantage of by fraudsters to adopt new strategies not yet included in the prevention, detection and action plans of companies.

This gave rise to the concept of **cybersecurity**, defined as the ability to protect or defend the use of cyberspace from cybernetic attacks[16]. In recent years, the risk associated with cybersecurity has become increasingly serious, mainly as a result of three factors: i) technological development, ii) the restructuring of industry processes and iii) the professionalization of cyber-attackers.

However, digitization across industries is also an advantage for organizations when it comes to fighting fraud, especially in relation to prevention and detection. Access to **large data volumes,** as well as the development of new techniques and models that allow **customer behavior analysis** (through advanced segmentation techniques like machine learning) constitute effective tools to combat fraud and are applicable across the different industries[17].

These solutions are designed to automatically analyze the different transactions recorded in companies' systems. They must be able to process the large data volumes available in order to detect, in real time, the different patterns and strategies designed by the defrauder. These techniques are an improvement on current mechanisms for the prevention and detection of fraudulent activity, since they are able to dynamically detect fraudulent patterns and strategies not used before.

Due to the **changing nature of fraudulent practices,** detecting them is a **continuous and dynamic process,** requiring organizations to previously define an action framework that includes specific strategies, approaches and policies, and in which all areas involved act in a coordinated manner. However, the treatment of fraud occasionally has weaknesses as a result of partial and often disperse approaches; processes that are partly outsourced and not integrated in day-to-day business management; lack of coordination between the areas responsible for preventing, detecting and responding to fraud events (anti-fraud teams, internal audit, network security and other areas); differentiation between the First Line of Defense (Management) and the Second Line of Defense (Control) not always clear, etc.

The following are a few examples of common fraud types in the banking, insurance, telecommunications and energy industries.

Banking/financial industry

As shown in the exhibit above, this is one of the industries most affected by fraud in terms of the number of occurrences. Besides, fraud events in the industry affect all products offered by banks (debit and credit cards, current accounts, checks, loans, etc.), all channels (branch offices, online/telephone banking, remote transactions), all IT support systems and all types of customers (retail, wholesale). It includes practices ranging from payments and use of fraudulent checks to phishing and identity theft, etc.

As a result, fraud mitigation and cybersecurity are among the main concerns for most companies. According to the G7 Cyber Expert Group[18], the main aim is to identify practices that represent potential fraud. Opportunity cost losses (normal transactions classified as potentially fraudulent), or false positives need to be minimized, because, when it comes to fraud, this type of error has a very high negative impact on the perception customers have of a company.

Insurance industry

Most offences in this industry are related to home insurance, life insurance, disability insurance, transport insurance and medical insurance. Here, fraud can be classified into two categories[19]:

- **"Hard fraud".** This takes place when the fraudster obtains money illegally through a planned strategy. It can involve more than one person and even a person working for the insurance company itself, acting in coordination with the policy holder or beneficiary. This type of fraud occurs, for instance, when someone deliberately causes an accident with the aim of collecting the car insurance money.

- **"Soft fraud".** This takes place when the fraudster files a legitimate claim but takes advantage of the situation to lie to the insurance company about the extent of the damage. This type of fraud occurs, for instance, when a person has a fortuitous car accident and the driver reports damages that are greater that those actually occurred. This is the most common type of fraud.

[16] As defined by the NIST: National Institute of Standards and Technology.
[17] Source: Management Solutions (2015): Data Science and Financial Industry Transformation. Management Solutions (2014): Model Risk Management: Quantitative and Qualitative aspects.
[18] Source: Fundamental Elements of Cybersecurity for the financial sector. October 2016 ("Increasing in sophistication, frequency, and persistence, cyber risks are growing more dangerous and diverse, threatening to disrupt our interconnected global financial systems and the institutions that operate and support those systems").
[19] Source: Insurance Information Institute.

Telecommunications industry

The main fraud categories in this industry can affect both the telecommunications company providing services and the customer entering into a legal contract with it, and they can be jointly or individually affected. A distinction can be made between three different fraud schemes depending on the fraud purpose:

- **Increased call volume.** This scheme uses network access simulation techniques to increase call traffic to certain destinations (where tariffs to be paid between operators are higher), so that illegal operators benefit from such tariffs.
- **Manipulation of the information systems of service providers.** Particularly fraud associated to systems providing network access or SIP trunking.
- **Telephony fraud.** Includes the sending of spam or text messages with the aim of obtaining personal data from users, telephone calls to financial institutions pretending to be a customer, or collapsing the network to interfere with the normal operation of a company or system.

Of particular note is the fraud known as "International Revenue Share Fraud (IRSF)"[20] , and the illegal network bypass[21] techniques.

Energy industry

In the energy industry, fraud can take place in a number of activities such as energy trading, procurement, project management, commercial management, etc., though some of the fraud in this industry is related to energy theft - the non-billing of energy that has been used (more than 80% of fraud events are related to misappropriation[22]). Here, it is worth mentioning that one of the main problems affecting efficiency and security in energy companies is the loss associated with the process of energy distribution and supply to consumers. This loss can be split into two categories[23]:

- **Technical loss.** Relates to the loss naturally occurring on the network due to phenomena such as power dissipation on transmission lines, etc.
- **Non-technical loss.** Associated with energy fraud and caused by actions external to the energy supply system. It consists mainly of energy theft, non-payment on the part of customers and loss due to errors in the invoicing process.

[20] Consists of fraud schemes where the defrauder increases the volume of calls to a high rate destination without the consent of the operator or the customer. In this case the excess costs are assigned to the operators. The gains from this type of fraud are shared between the illegal operator in the country of destination of the calls and the individual that makes the calls in order to increase the volume of calls for the illegal operator.

[21] It consists of illegally channeling international calls into a country with the sole purpose of avoiding the payment of dues. SIM Box or VoIP calls are often used to falsify call origination.

[22] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study. An analysis of the frequency of fraud events by category shows that over 80% of cases can be typified as "Asset Misappropriation"

[23] According to Pedro Antmann: Reducing Technical and Non-Technical Losses in the Power Sector. Technical report. World Bank. July 2009.

Identifying and differentiating the percentage of energy lost in distribution from non-technical causes is a challenge that companies in the industry need to face and resolve. There are three categories[24]:

- **Actions that affect the distribution company's network.** Of particular note are the **direct connections** into the distribution network without the customer having entered into a contract for the supply of electricity and the **channeling** of electrical power towards other facilities not included in the contract.

- **Actions that affect measuring and control devices such as the tampering of meters** with the aim of reporting consumption levels below actual usage.

- **Dishonest action by company employees in commercial processes.** It can take place, for instance, if there is no proper segregation of duties and the same person records or modifies transactions and authorizes the associated payments or invoicing.

These three types of fraud can constitute a breach or offence, sometimes punished by sanctions, whose amount depends on the volume of energy defrauded. As a general rule, energy distribution companies should detect and report to the authorities any network and equipment irregularities (e.g. in Spain fines are issued by the regional governments and the law provides for the re-invoicing of an amount corresponding to six power consumption hours per day in a year[25]).

Management levers

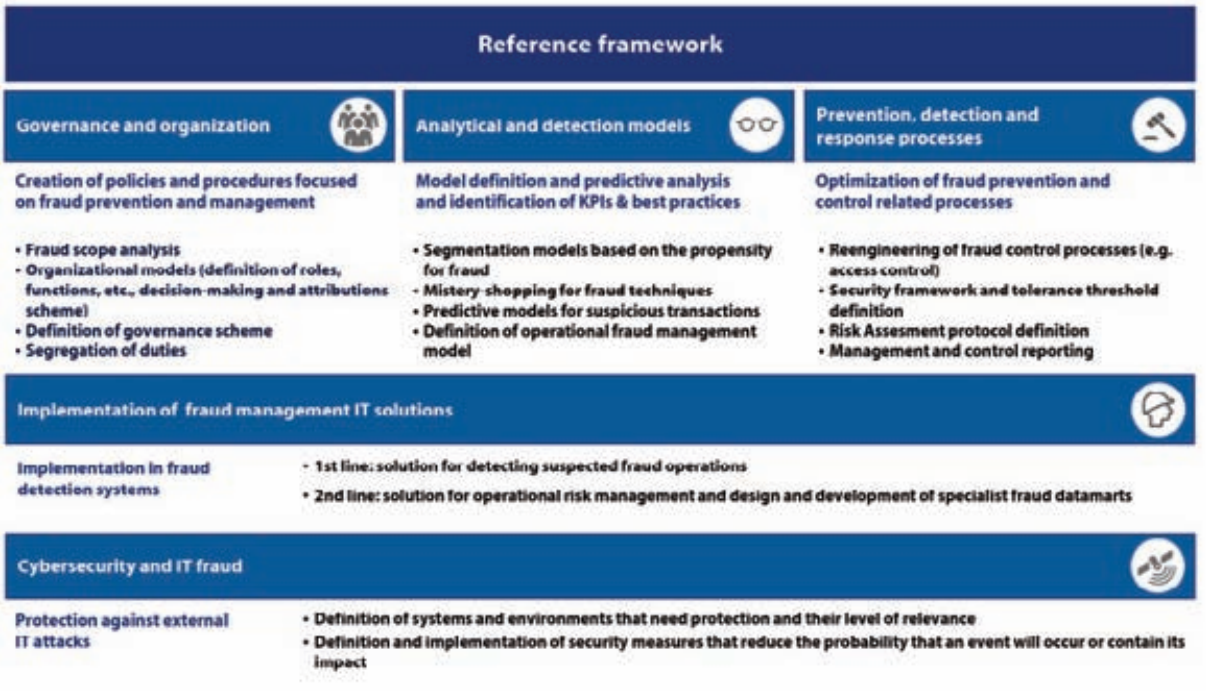
A comprehensive fraud management framework would answer questions such as:

- Should fraud management units be integrated within a single area?
- Who is responsible, as a second line of defense, for internal fraud?
- What capabilities (models/systems) are necessary to anticipate fraud?
- What overall management measures will be required for big data/machine learning?
- Etc.

Such a framework could be structured around basic principles/components[26] like (see fig. 3): definition of a governance model, establishing periodic evaluations to determine the risk of fraud, implementation of prevention and detection techniques and processes as well as corrective and response action to minimize losses once fraud has taken place. Some of the most significant across-the-board measures are implementing the aforementioned elements in the systems and integrating cybersecurity systems into the day-to-day operations of companies (in order to deal with IT fraud, one of the most significant and common types of fraud due to industry digitization).

[24] Sahoo, S., Nikovski, D., Muso, T., & Tsuru, K. (2015, February). Electricity theft detection using smart meter data. In Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society
[25] Under Law 24/2013 on the Electric Power Industry .
[26] The Institute of Internal Auditors (IIA), The American Institute of Certified Public Accountants (AICPA) y Association of Certified Fraud Examiners (ACFE) (2012): Managing the Business Risk of Fraud: A Practical Guide.

Fig 3. Comprehensive fraud prevention and management framework



Source: own elaboration

Governance and organization

As part of the organization's governance structure, companies establish fraud management policies which determine management and control responsibilities according to the type of fraud.

- For **external fraud** events, identification and management is usually carried out by the business units themselves through the implementation of anti-fraud policies for their operational processes. In parallel, and independently from the business units, control over the proper implementation of policies is carried out from the risk control, internal control and audit areas, and second and third lines of defense functions.

- For **internal fraud,** however, identification and management is usually carried out by areas that are independent from the business, such as audit or internal control, from where the required investigations are conducted in order to reach conclusions that will enable the company to take disciplinary measures. In fact, according to the Spanish Internal Auditors Institute[27], *"Internal Audit must assure the Board and Management that fraud-related controls are sufficient to cover the risks identified and to guarantee that those controls are effective"*. In addition to establishing their own monitoring processes, these functions usually rely on internal, anonymous fraud reporting channels.

When defining antifraud policies, organizations consider the level of complexity and depth they desire to reach, with company size itself being a relevant factor.

Analytical intelligence

According to the Institute of Electrical and Electronics Engineers, the probability that fraud will be detected depends on the pilfered amount and the level of investment required to detect it[28]. It is therefore necessary to **periodically assess the organization's exposure** to the risk of fraud, identifying new potential fraud strategies and events the company needs to mitigate. Besides identifying the risks, **the probability of occurrence needs to be revised,** as well as its severity if the fraud event occurs.

Based on the fraud triangle shown in section 2.1.2, risks are reviewed and analytical indicators are defined that will subsequently be monitored through reports in order to identify and anticipate the propensity for fraud.

The adaptive nature of fraud management requires intelligent and statistical analysis systems for detection purposes. Those techniques that allow the detection of new strategies and patterns used by defrauders without becoming outdated

[27] IAIE: Fraud Risk Management: Prevention, detection and research. February 2015.
[28] Source: Amin, Saurabh, Galina A. Schwartz, Álvaro A. Cardenas, and S. Shankar Sastry. "Game-Theoretic Models of Electricity Theft Detection in Smart Utility Networks: Providing New Capabilities with Advanced Metering Infrastructure." IEEE Control Systems 35, no. 1 (February 2015).

combine **pure analysis and modeling elements** such as data mining and machine learning, high performance **technical elements** such as stream computing, and **full data transformation processes** for the acquisition of useful knowledge such as Knowledge Discovery in Database (KDD). These methods need to be used prior to the implementation of internal controls (see fig. 4).

Some of the benefits of using statistical data analysis are[29]:

- **Holistic view of a company.** Portfolio of active clients for whom there are multiple internal and external data sources, with the possibility of enhancing the internal information with both data from customer interaction (payment behavior, incidents, usage, etc.) and external information provided by third parties (purchasing power, default levels, socioeconomic level, etc.).

- **Analysis of unstructured information.** Data sourced from social networks, conversations, etc. represent valuable information when it comes to detecting fraud; however, traditional databases did not make it possible to store this information in an appropriate way. New techniques allow for this data to be stored properly, and also to use it and incorporate it into predictive models.

Nevertheless, events that are not obvious and have a low incidence level need to be identified and treated using a business criterion or defined policy.

Finally, analytical fraud models need to be validated, i.e. the model supervision process needs to be defined and implemented in order to confirm that the end model performance is stable and correct, and that it meets the business requirements and, potentially also the regulatory requirements.

Prevention, detection and response mechanisms

From the operational perspective, actions are oriented towards the prevention and detection of potentially fraudulent events (implementation of techniques for the detection of fraudulent events that have not been foreseen by the company), as well as the definition of response processes to ensure that any fraud is dealt with timely and appropriately with the aim of minimizing loss associated with the fraudulent event.

- **Identification** through defining which processes, data, systems and environments require protection as well as how relevant they are.

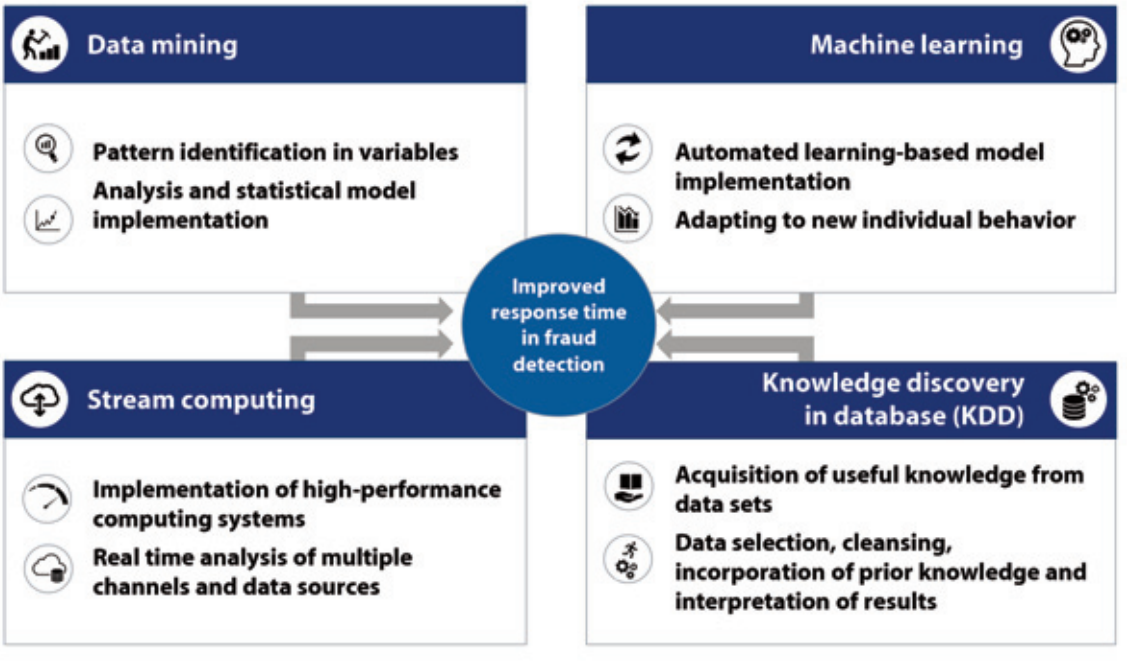
- **Protection** through the implementation of controls, whether these are control systems or procedures, that reduce ex-ante the company's risk exposure or prevent a threat from becoming fraud (e.g. access policies);

- **Detection** using early warning systems which, through monitoring or analysis, allow the identification of a fraud event or fraudulent operation (e.g. KPI/KRI dashboard); and

- **Response** and **recovery** using agile processes to implement corrective action that will minimize and/or offset the loss caused by the fraudulent event (reduce its impact).

[29] An example of the practical implementation of these benefits can be found in Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. Decision Support Systems, 46(4).

Fig 4. Analytical intelligence applied to fraud management



Source: own elaboration

Support model

Implementing the above elements in the IT environments of organizations is essential to effectively managing fraud[30]:

An integrated fraud prevention and detection system makes it easier to manage all business processes affected by this risk. A fraud detection system should have the following basic characteristics:

- Having a **data repository with the organization's entire fraud reality.** It should have both the input variables to the detection models and the scores obtained from the execution of models on the organizations transaction activity and generated alerts. It should implement **data quality controls** both in the historical repository and in the capture of new variables.
- Implementing a **parameterizable fraud detection model** that can be adapted to the organization's problems. These systems usually have pre-parameterized models based on industry knowledge that need to be particularized and monitored (their performance).
- Executing **alert generation and dynamic report flows** in accordance with the review and analysis scheme implemented in the organization to define alerts either on-line or in batch mode.
- **Reporting** both the organization's fraud levels as well as the management action implemented and the loss incurred.

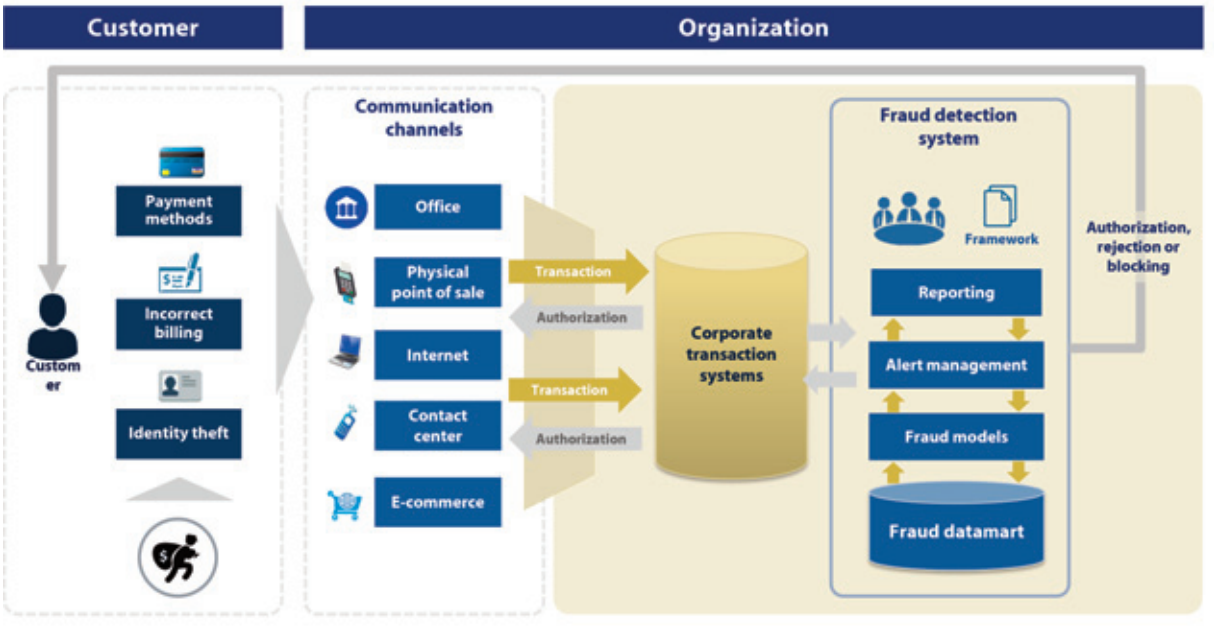
For a fraud detection system to be effectively integrated into the management process it has to facilitate the input of customer **transaction and commercial information;** its online integration (e.g. via web services) in the transaction authorization/denial processes; provide a **high technological performance level** due to its involvement in customer interaction processes; have an **authentication system that is compatible** with organizational standards (e.g. LDAP); and have a **role and user scheme** that can be defined in accordance with the organization's policies.

To illustrate the integration of a fraud detection system into an organization's IT environment, figure 5 shows an example of the life cycle for the information analyzed in the fraud detection process: from origination of the event by the customer and its input into the organization's internal processing to the final decision on the suspected fraud-related activity.

Likewise, the exhibit shows the functional hierarchy of the different internal components in that system: from storage of granular information in the data repository to the implementation of fraud detection models with the transaction data received and the subsequent alert management and reporting.

[30] *"The results of Data Analytics may be used to identify areas of key risk, fraud, errors or misuse; improve business efficiencies; verify process effectiveness; and influence business decisions."*, ISACA, Information Systems Audit and Control Association: Data Analytics – A practical approach. *White Paper. August 2011.*

Fig 5. Fraud detection system integration



Source: own elaboration

Fraud management techniques in the energy industry

The increase in information storage capacity and calculation power has driven the development of **Analytical Intelligence** as well as the different disciplines it covers, with the highlight being **Data Science,** whose purpose is to extract the maximum knowledge from data by combining big data analysis with modeling, profiling and segmentation techniques[31].

In the energy industry, these techniques are used to address problems ranging from predicting energy demand levels to identifying energy consumption patterns with the aim of personalizing the commercial offer or detecting fraud events.

Data

Collecting and subsequently processing data requires prior analysis of the type, nature and origin of the data (see fig. 6).

In any case, locating the sources, defining the extraction, storage and treatment processes, analyzing the quality of data, etc., are actions that need to be carried out under a data governance framework, approved by a company's top management level.

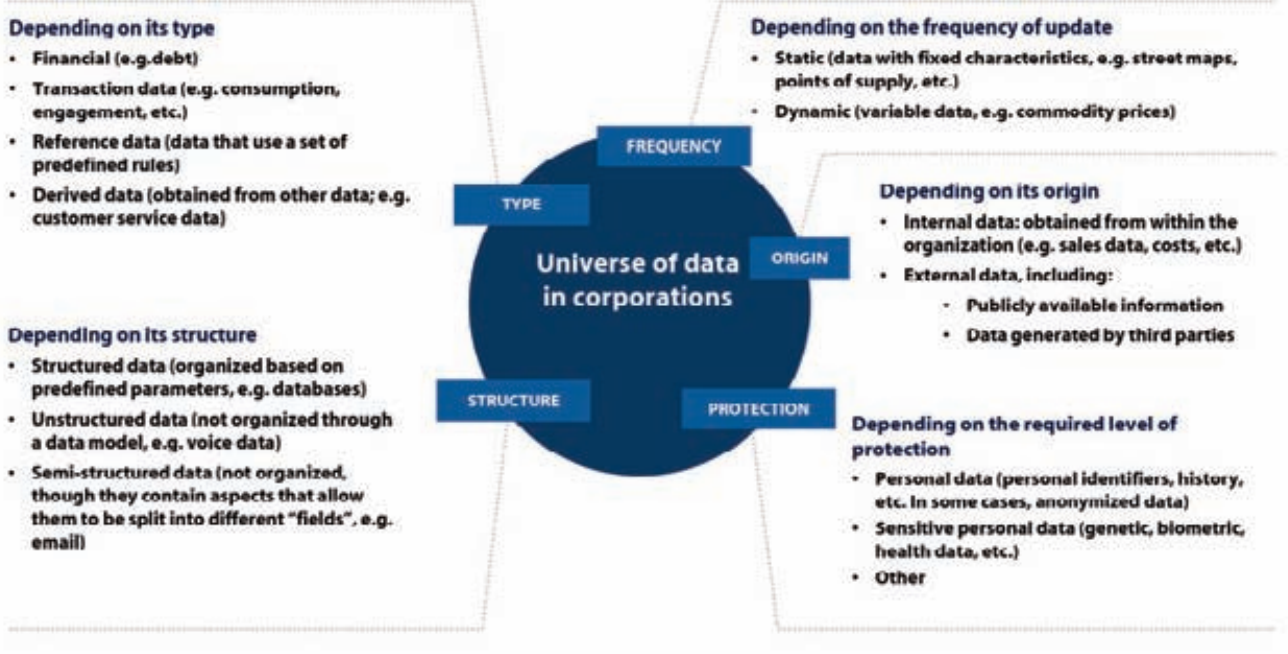
Such data governance implies the development of three different domains: i) IT Architecture, ii) People and their "Skills", iii) Processes / Instruments for effective governance (see fig. 7).

The key for data governance is to turn it into a useful management instrument. This requires first identifying and delimiting the data to be managed and then classifying the data according to its type (internal, external, structured, non-structured, etc.) and the required level of protection.

This should be used to **prioritize** the data maily accroding to the cost/benefit obtained from a greater or lesser data governance level, and finally, develop a framework focusing on **formalization** (definition of concepts, owners, information sources, etc.), **follow-up** on quality and plans (controls and indicators for quality monitoring, remediation plans, etc.) and

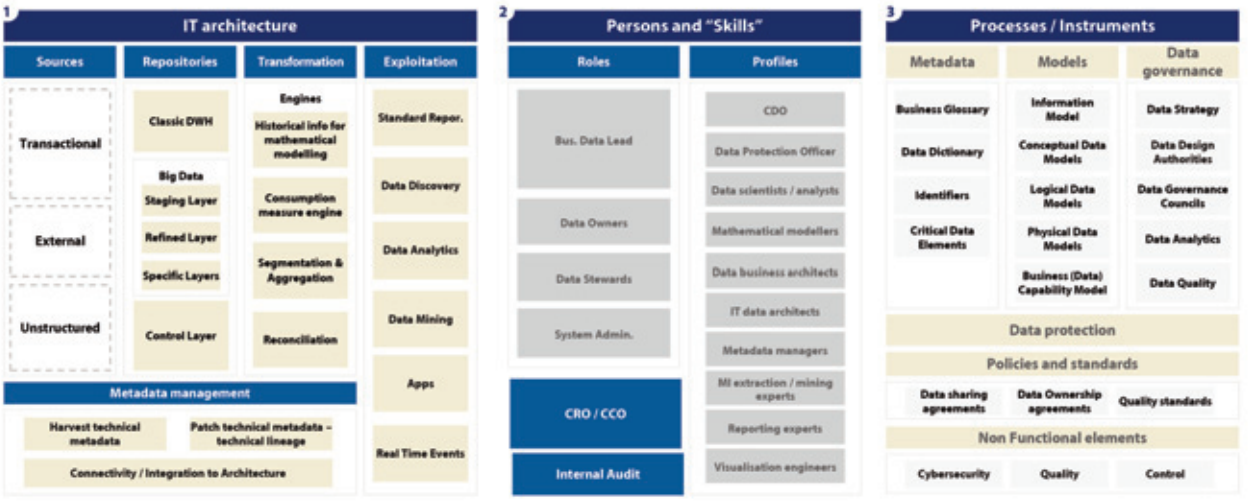
[31] Data Science and Financial Industry Transformation". Management Solutions. June 2016

Fig.6. Main data types in corporations



Source: own elaboration

Fig.7. Data governance development areas



Source: own elaboration

independent **validation** (certification or internal audit) considering the aspects relating to the protection of both internal and customer information (GDPR[32]).

Models

The models to be developed seek to find patterns, trends or rules that will explain the behavior of customers, employees or third parties before fraud is detected.

When selecting the variables to be included in the model, one needs to make sure that they meet the following conditions:

- They cover all profiles that need to be reflected.
- Together they provide the greatest predictive power and are not redundant.

This requires conducting a (univariate and multivariate) statistical analysis on the variables, including an analysis to detect multicollinearity by looking at the correlation between the explanatory variables. This achieves the greatest model simplicity.

Among the statistical and numerical tools most often used to analyze and correct the correlation are the use of statistical indicators (such as the Pearson, Kendall or Spearman correlation coefficient, depending on the type of variables) and the use of dimensionality reduction techniques such as Principal Component Analysis (PCA), which seeks a lower dimensional subspace onto which to project the data, thus minimizing projection errors.

Once the final group of explanatory variables has been delimited, the algorithm is developed. The techniques to be used to identify the profiles/observations likely to represent a fraud event will vary depending on the end goal and type of data used. Of particular note are the following methodologies:

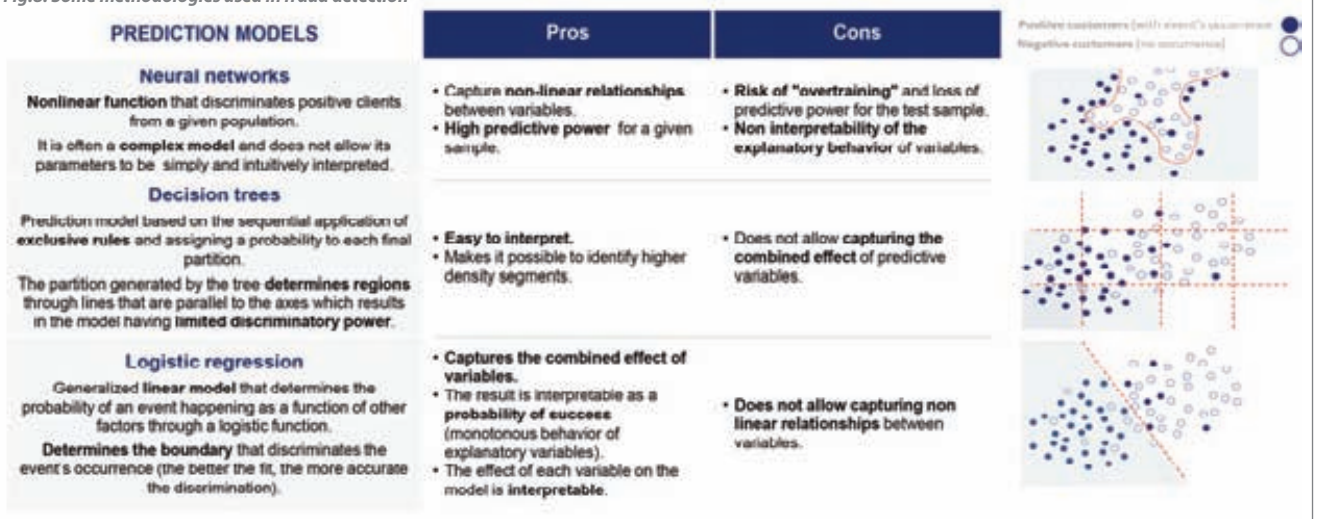
- **Classification models.** The purpose of this type of methodology is to predict the category to which each observation or entry to be analyzed belongs according to its attributes. Some of the most commonly used techniques are decision trees, which can be generated using different algorithms such as CLS, ID3, CART, etc. The **Random Forest** technique is another classification model based on aggregating several decision trees (e.g. by using the mode or the mean) to predict the category each entry belongs to.

  Other commonly used classification techniques are the **Bayesian Classifiers, Neural Networks or the k-Nearest Neighbors** technique**.**

- **Regression models.** The main aim is the numerical estimation of the relationship between an independent variable and a group of explanatory variables. There are two types of regression, **linear and non-linear.** Examples of these types of models are **Bayesian linear regressions** and the **Generalized Linear Models (GLM). Logistic Regressions** are regression models whose purpose is to

[32] General Data Protection Regulation. New EU regulatory framework (for relationships between member states and with third parties) in the area of data protection with the aim of harmonizing and unifying criteria for the implementation and enforcement of data protection and privacy rights, adapting standards to the digital environment.

Fig.8. Some methodologies used in fraud detection



Source: own elaboration

classify observations. Regression analysis is also used to predict how a variable will change over time, i.e. for time series analysis. In particular, **ARIMA models (Autoregressive Integrated Moving Average)** are used for predicting future values of time series, based on past values.

- **Segmentation or clustering models.** The aim of this methodology is to group the different observations into uniform groups or clusters, based on their similarities. These are non-supervised models, since data is adjusted through a sample in which the target groups are unknown a priori, i.e. there is no previous knowledge of the categories to which the observations might be assigned. Depending on the criteria used to determine the degree of similarity between observations, there are different cluster analysis types: **centric models** (e.g. **k-means** or **k-medians**), **statistical distribution models** (e.g. **expectation-maximization algorithm, Gaussian Mixture Models**), **hierarchical models** in which clusters are merged or successively subdivided according to a priority ranking or hierarchy, etc. Clustering techniques are usually the first step when it comes to addressing classification issues when there is not enough information on the categories that are to be differentiated.

Other techniques used in fraud detection are **associative rules** or **ARs, sequence identification (Motif Mining, Autocorrelation Function)** or the **detection of anomalies or outliers.** Each type of model shows a different predictive power, stability and interpretability level.

Taking advantage of the full potential of Data Science techniques implies the implementation of the following characteristics in some of the processes that make up the model life cycle:

- **Real time analysis.** Data collection systems allow the capturing and follow-up of data in real time. Besides, it is possible to program algorithms that will use such data, which facilitates and speeds up the detection of new fraud patterns and strategies not yet used by the defrauders. In addition, this characteristic makes it possible for models not to lose predictive power overtime, and to always remain updated and sensitive to new fraud patterns.

- **Automatic retraining and self-learning.** Fraud detection models are recalibrated automatically (with little input from analysts) and iteratively from large data volumes,

which translates into predictive power potentially improving during the successive re-trainings.

Real time analysis and automated model retraining and self-learning reduce time-to-market for models. Besides, these features make it possible to search for and detect up-to-date patterns and relationships without predefined restrictions, as well as to identify and incorporate new relevant variables into models. It is also possible to program models so they can automatically recalibrate through variations in the relative weight each variable contributes to the detection of fraud events.

The downside is that this process sophistication implies more complex model risk management, requiring the implementation of internal controls and warning systems to detect any deviation from the model, in addition to controls over the degree of freedom in process automation. All this should be supported by a model management framework.

Main areas of fraud management applicability in the energy industry

The next section discusses two specific areas of applicability in the **energy industry: energy theft in the distribution network** and **fraudulent activities in the commercial cycle.** These areas are relevant due to both their economic impact (it is estimated that utilities companies lose up to 95 billion dollars a year worldwide from non-technical energy loss[33]) and reputational impact (e.g. increased difference between the demand for electricity measured at consumption points and the energy measured at power stations, resulting in excess costs for the system which are passed on to the consumers[34]). The first impact revolves around measuring the propensity or probability that a customer or user will use energy without the company knowing about it (external fraud). The second impact has to do with identifying incompatibilities in processes, such as the commercial cycle, which might result in gains for employees (internal fraud).

External fraud: energy theft

Increased capacity to generate and store information can be taken advantage of to gain real time access to a customer's characterization, a transaction, a process, etc., which is used to identify behavior indicating a propensity for energy theft.

For years now, detecting non-technical losses has been a main concern for companies. However, the historically available solutions implied high costs (inspections by engineers).

Today, many countries are developing and implementing what is known as Advanced Metering Infrastructures or AMI[35] on Smart Grids[36], which include data collection systems and real time data monitoring, as well as the use of analysis techniques based on artificial intelligence, game theory, etc.

According to findings from a study conducted by the US Energy Department[37], using these Data Science techniques to detect fraud in the energy industry is very useful when it comes to telling apart the percentage of energy lost in the distribution network for technical reasons (not reflecting fraud

[33] Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors. May 2017, Northeast Group, LLC.
[34] Report on regulatory alternatives for loss reduction and treatment of fraud in the electricity supply. Report of July 16, 2015 by the CNMC - Spain's National Commission on Markets and Competition.
[35] Advanced Metering Infrastructure.
[36] Intelligent electrical networks.
[37] US Department of Energy – Office of Electricity Delivery and Energy Reliability: AMI and Customer Systems: Results from the SGIG Program. September 2016.

events) and from non-technical lossess (energy theft - therefore reflecting fraud events).

Likewise, energy companies are undertaking campaigns to implement SM's[38] that help to reduce non-technical losses in the distribution network[39]. According to the US Department of Energy[40], collecting energy consumption data through these devices, and then analyzing it using large data volume processing techniques, brings new possibilities for developing effective and efficient fraud detection methods, improving income recovery. Some of the benefits of SM usage are the possibility to read energy consumption data remotely, greater measurement resolution, and the **detection of cuts or disconnections not foreseen** in the data collection process by the meter.

However, their use also presents challenges:

- Collecting data for long periods of time due to **storage** capacity limitations.
- The existence of compression processes that reduce **data quality** and limit the possibility of using this data at a later stage.
- The computational cost of **processing data in real time.**
- The protection of personal data governed by **confidentiality** restrictions[41].

Investing in advanced detection models **optimizes the success rates of inspection campaigns** and improves fraud detection. However, making progress in the implementation of these types of models requires previously defining and implementing a reference framework that will facilitate the governance of data as well as of models and related processes (see fig. 9)

The process under analysis is the **execution of inspections**. As seen in previous sections, integrating analytical segmentation techniques into the management process requires a scheme for processing large data volumes as well as profiling customers.

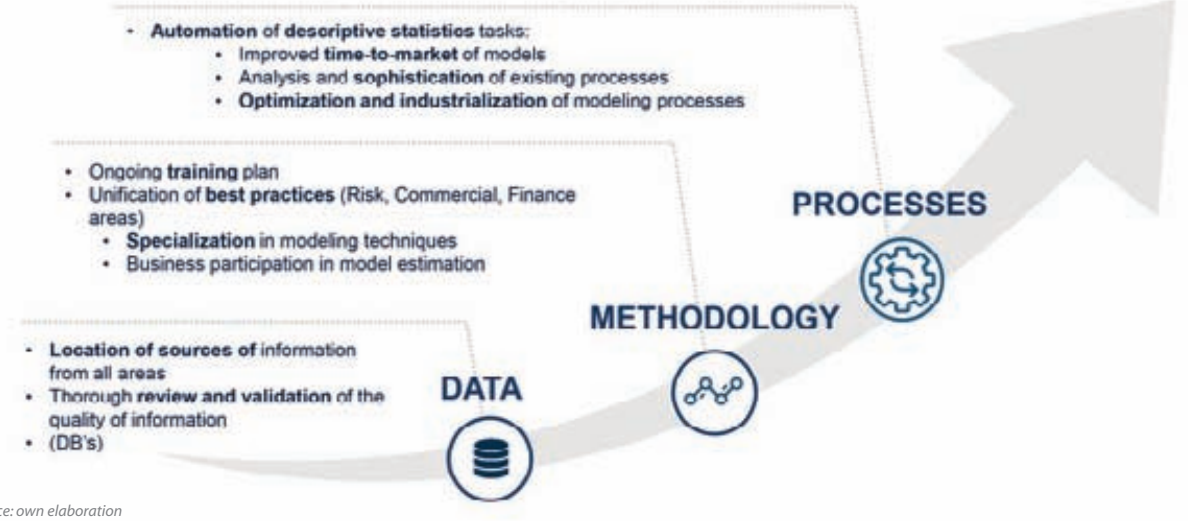
Once segmentation models have determined the probability of occurrence for the event (in this case the existence of energy theft), these probabilities, combined with the expected benefit from the action (e.g. the amount of pilfered energy) help to predict the expected outcome from an inspection.

Therefore, for the operationalization of a scheme of inspections of this type, a support model or campaign management tool is often used in order to apply the filters and prioritizations generated by the model, and to compile the inspection campaign results that will be fed back into the models.

This last point is extremely important, since the result of an inspection will be a valuable input for the characterization of a particular supply in the future, helping to typify both repeat offenders and false positives.

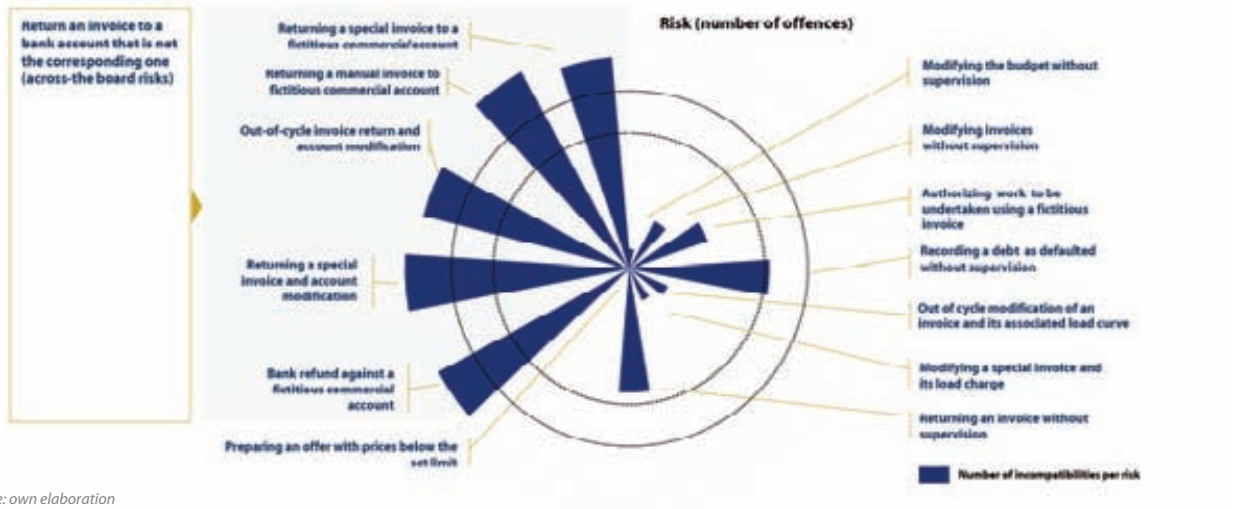
[38] Smart Meters.
[39] Reference:" AMI (advanced metering infrastructure) provides powerful tools to reduce total losses and increase collection rates", published by World Bank: Reducing Technical and Non-Technical Losses in the Power Sector. July 2009.
[40] US Department of Energy – Office of Electricity Delivery and Energy Reliability: AMI and Customer Systems: Results from the SGIG Program. September 2016.
[41] These data are considered to be of a personal nature and are governed by Spain's Organic Law on Data Protection (LOPD) and its implementing regulation (Royal Decree 1720/2007). According to article 6 of this law, the express consent of the individual in question (the customer in this case) is required for the use of this information. The business terms and conditions of the main distributors' contracts provide for the use of personal data by the distributor and by third parties of companies they have a contractual relationship with, and only for commercial prospecting. These requirements are reinforced by the new General Regulation on Data Protection (RGPD), which reinforces the requirements for the management of customer consent (explicit grant of use for a specific purpose and period of time).

Fig.9. The use of Data Science techniques impacts on data, methodology and processes



Source: own elaboration

Fig.10. Illustrative taxonomy of risks in commercial systems



Source: own elaboration

Internal fraud: segregation of duties in the commercial cycle

Energy companies have many systems to support their operational processes. These systems are accessed by both their own employees and external professionals.

Employees changing positions or the existence of generic users in some systems (mainly as a result of functions such as invoicing and collections or call centers being outsourced) make it more **complex to monitor the functions** performed by the different participants.

The possibility that employees may perform actions for personal benefit (e.g. replacing a customer's bank account with the employee's own bank account to reimburse an invoice, or modifying the amount of the employee's own invoice and its related energy consumption) constitutes **internal fraud risk.**

Implementing access control solutions makes it possible to identify weaknesses, especially in commercial systems where many functions are outsourced, which leads to many "segregation of duties" related risks. An illustrative taxonomy of risks (fig. 10) is shown above.

Access control projects seek to: i) review roles to identify functions that are incompatible, ii) reassign tasks to eliminate such incompatibilities or iii) if the incompatibility cannot be avoided, establish mitigation controls.

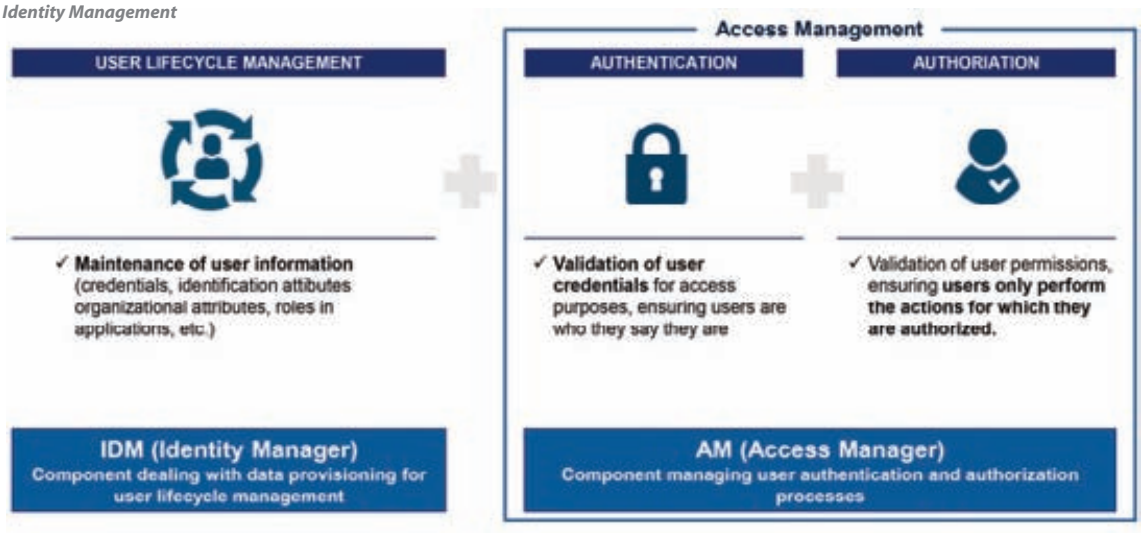
Many energy companies have implemented Access Control solutions (also known as Identity Management solutions) to fulfill three objectives:

- **Reducing the risk of unauthorized access** to systems thorugh the definition of user provisioning models that will perform ex ante control with a view to anticipate incompatible situations (e.g. before assigning a role to a user, verify that such an assignment does not imply non-compliance with the SoD).

- **Ensuring data confidentiality, integrity and availability through the segregation of duties,** ensuring that access to critical company information is controlled and only the right people have access (e.g. payroll systems or accounting systems).

- **Automating user provisioning,** ensuring that employees have the required permissions in the shortest possible time (e.g. automating access to systems according to the employee's position and removing access when employees leave the company or change areas within the company).

Fig.11. Identity Management



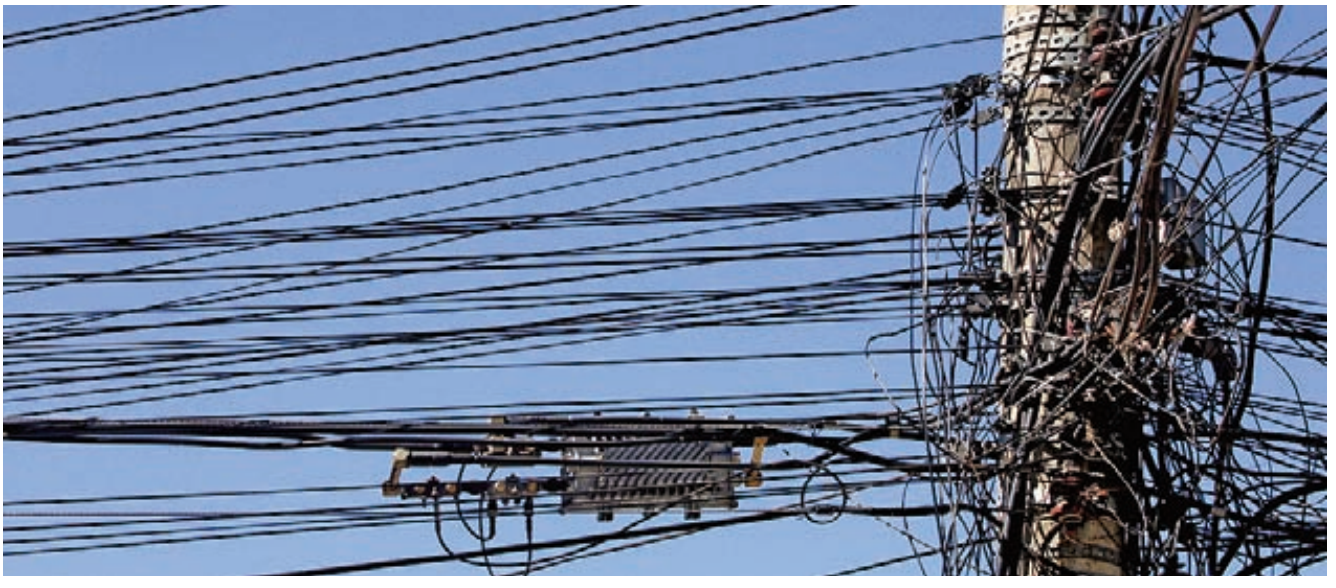
Source: own elaboration

Identity and Access control (known as Identity Management) comprises three main processes (User Lifecycle Management, Authentication and Authorization) ensuring that users i) are who they say they are and ii) access the right application with the required permissions.

These processes are supported by IT solutions known as IGA (Identity Governance and Administration), whose implementation allows companies to increase the **level of control** over access to systems, thereby ensuring the segregation of functions and reducing the company's internal risk. However, these solutions require continuous monitoring to ensure proper control over the segregation of duties, as well as the availability of analysis tools to define:

- **A plan to measure or control the mitigation of incompatibilities,** such as role/user remediation or the establishment of exceptions to be monitored with mitigating control (e.g. defining reports or processes to ensure risk is controlled).
- An **action protocol** so that risks and .incompatibilities appearing in future reviews are quickly resolved or mitigated.

In any case, effective access control can be jeopardized if the company does not have a global view of employees' overall access rights and independent authorizations by application, as well as the processes in place to remove such access (e.g. when employees change position within the company, the permissions that were required for their previous position are not always removed). It is equally important to carry out a risk analysis by position before cancelling access rights.



Example of the implementation of modeling techniques: energy theft

Detecting energy fraud, in connection with the illegal use of energy from the network, starts with customer segmentation based on the customers' individual probability of committing fraud. This requires interpreting historical data on fraud and its impact on the business, as well as the historical behavior of energy pilferers (including the outcome from previous inspections, invoicing and collections).

The modeling techniques used in energy fraud detection seek to improve success rates in the selection of customers that should be inspected.

Detection model life-cycle

Models used in fraud detection comprise four stages: data collection, statistical data analysis, construction, and validation and certification (see fig. 12).

Data collection and processing

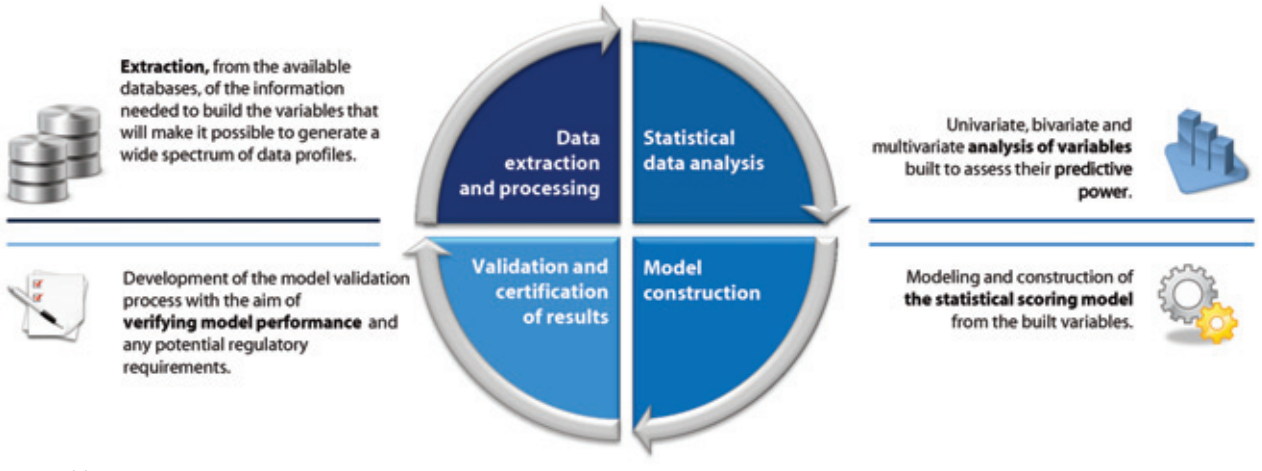
The first stage in the collection and processing of data from energy customers comprises various phases in turn: requesting and collecting the data, analyzing the quality of data and building new variables (see fig. 13).

The information collected will be recorded in a single table. Also, new variables will be created from the collected data. Selecting relevant variables with good predictive power for energy fraud detection purposes requires in-depth statistical analysis. Some variables that have shown good predictive power are:

- **Meter data:** the type of connection (single-phase, two-phase, three-phase), brand and model of the meter may be relevant due to the complexity of manipulation for fraud and because they represent an opportunity perceived by the customer. Other relevant variables may be the type of network and the installed meter power that defines the customer's maximum potential energy consumption, since actual consumption values should be similar to the consumption values for customers that do not commit fraud.

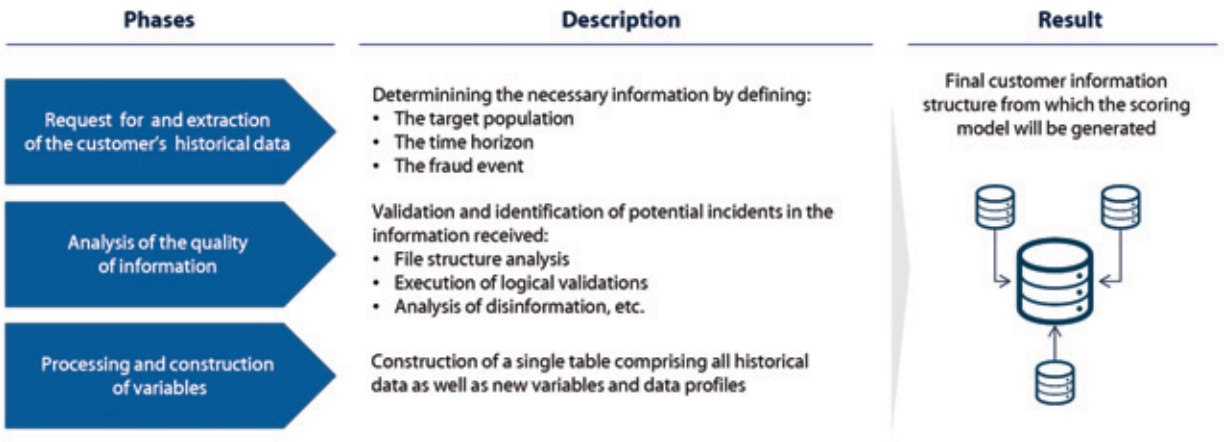
- **Sociodemographic data:** this data can be useful for segmentation purposes. For instance, location attributes (region, province, municipality, post code, neighborhood or area) show strong explanatory factors for the fraud event. To supplement the customer data, it is possible to obtain external information enhanced by area, such as average income, average consumption, type of housing, weather, etc.

Fig.12. Model development stages



Source: own elaboration

Fig.13. Data treatment in model development



Source: own elaboration

- **Customer data:** energy companies usually have useful customer information (length of customer relationship with the company, individual or collective consumption, type of customer: residential, business, industrial, public service, rural, public lighting, low-income customer, etc.). In the case of large clients the volume of data available is increased with attributes such as activity, industry, etc.

- **Historical energy consumption data:** this information plays a fundamental role in customer behavior analysis and, thanks to smart metering, the quality of these variables is improved. Worthy of note is the analysis of deviations in expected consumption for cyclical, climate, energy cuts, and other reasons, average consumption with respect to similar customers or the same customer in the past, meter reading irregularities, frequency of customer readings or customer debt.

  The frequency and variety of this information increases every year (the historical information available could be scant, such as quarterly consumption from a collective supply point, but today smart meters provide detailed information on currents, phase, tension, power, etc. in real time).

- **Operation or interaction data:** data on network and meter maintenance operation, information on cuts and irregularities or information on customer contact or claims, etc. can be useful.

- **Inspection data:** the detail of the inspection outcome cannot be used as a variable for the model, but it does help to conduct an initial analysis of what causes fraud. The repeat offenders subgroup may provide some data to quickly identify the new fraud.

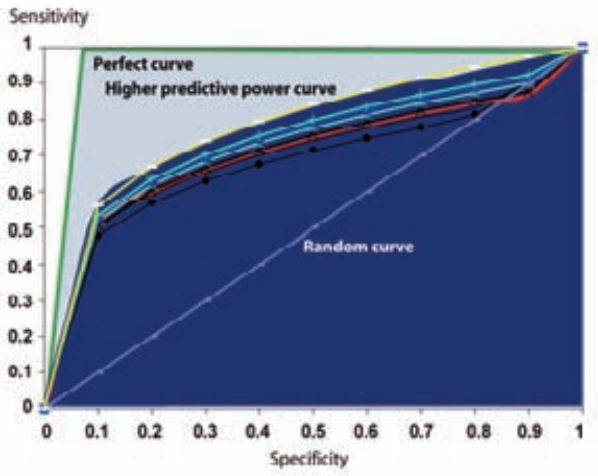
Statistical data analysis and model construction

First, a statistical analysis of the information is carried out to detect the most relevant variables, as described in section 3.1.2. At this stage, there is a set of variables available which have been debugged and can potentially predict customer fraud, therefore model construction (calibration) can take place[42]. Next, the most appropriate **algorithm is selected** for the model that is to be built on the target population and the model's parameters are determined. As an option, assigning weights to the variables (selected using statistical techniques) allows customer profiling as well as **identifying customers** with behavior comparable to fraud.

From a statistical standpoint, there are two different approaches to the classification problem. Under the first approach, the groups are clearly defined and the aim is to determine the criteria to be used for labelling each individual as belonging to any of these groups, based on values from a limited set of parameters. In this case, the techniques most commonly used are known as **discriminant analysis,** though there are other alternatives such as the use of logistic regression, neural networks or decision trees. The second approach relates to cases where the groups are not known a priori, and the aim is precisely to establish them based on the available data. The statistical techniques most often used in this area are known as **cluster analysis.**

[42] Model calibration through machine learning methods is supported by Bootstrap Aggregating, or "Bagging techniques" (models are trained simultaneously and the combined predictions are used as the final prediction) or Boosting techniques (models are trained sequentially so that the next model can focus on correctly predicting errors in the previous models). See Breiman, Leo (1996). Bagging predictors 24 (2), or Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems. Springer Berlin Heidelberg.

ROC curve



A number of models are compared below (ranked from lowest to highest in terms of their level of sophistication):

- **Decision tree:** calibration of an automatic decision tree by selecting the variables with greater predictive power according to the "Chi square" or "Cramer" test (each criterion gives rise to a different tree).
- **Logistic regression:** selection of a subset of variables with greater predictive power (according to the Chi-square test) and calibration of a logistic regression.
- **Neural network:** training of a neural network with raw data.
- **Transformation + Neural network:** combination of simple transformations (logarithms, inverse functions, roots, powers, translations, etc.) of continuous variables with the original variables for the training of a neural network.
- **Regression + Neural network:** combination of a logistic regression for the aggregation of variables with a neural network.
- ***Random Forest:*** sequential combination of predictor trees that give rise to the so-called "forest", which will provide a prediction of the event by linking the predictions of all trees in the process.
- ***Gradient Boosting:*** combination of predictor trees to obtain a more robust classifier through the implementation of machine learning algorithms.

Validation

With the aim of identifying the models that best explain the behavior of fraudulent customers, some minimum criteria are established that need to be met by the results obtained through the selected model.

These criteria are based on both the **model's discriminating ability,** e.g. its AUC index[43], and **reasonableness**. The latter is measured using trend analysis (which confirms that the trend for each variable's estimator, with respect to fraud, matches expectations in economic terms) and the relative weights of variables according to their expected contribution.

A quantitative exercise was conducted, implementing the models described in the previous paragraph independently on the same training basis. The following is a comparison of the models' discriminating ability based on the area under the ROC curve (AUC):

Model	ROC obtained in the validation phase
Decision tree	0.78
Logistic regression	0.74
Neural network	0.82
Transformation + Neural network	0.83
Regression + Neural network	0.79
Random Forest	0.81
Gradient Boosting	0.86

[43] AUC: Area Under the Curve. The curve used is the ROC (Receiving Operating Characteristics) curve, obtained from a percentage of individuals duly sorted by the model based on their propensity for theft, i.e. if a model identifies a potential pilferer as having a higher propensity than another, the probability that this is correct corresponds to the AUC.

As can be observed from the analysis conducted, the **Gradient Boosting** model shows a better fit in ROC terms.

In addition to statistical validation, the models were validated with the inspections conducted over a six-month period. The groups of 12, 25 and 50 customers with the greatest propensity in each model were compared with the inspections made during that time to determine how effective the model was in identifying fraudsters. The following table shows the percentage of pilferers the model would find in each of these groups (of 12, 25 or 50 customers).

Model	12 clients	25 clients	50 clients
Decision tree	58%	40%	30%
Logistic regression	50%	52%	44%
Neural network	75%	64%	38%
Transformation + Neural network	75%	68%	46%
Regression + Neural network	67%	60%	42%
Random Forest	75%	64%	52%
Gradient Boosting	92%	72%	50%

It is observed that, regardless of the number of inspections, the **Gradient Boosting** model is the one that shows the greatest concentration of pilferers (except the **Random Forest** model under the 50 inspections premise, though the 2% improvement is not significant). Therefore, this is the model that will be used in the practical example.

Energy gain

Energy gain is the concept used to represent the **economic value that is recovered for each client after energy fraud is identified.** This concept corresponds to the economic value of the energy that starts to be invoiced to normalize consumption after the inspection, plus the energy not invoiced in the past. The concept is analogous to the Customer Lifetime Value concept used in commercial segmentation techniques. While its calculation is a combination of business criteria defined by each company, we will describe the main factors used to calculate it:

- **Country's legislation:** regulations in each country establish criteria for penalizing energy fraud and determine how companies should proceed. An example of this could be immediately stopping supply and a sanction for the offence based on an estimate of the energy consumption corresponding to the product that was either contracted or should have been contracted, for the supply where the fraud took place, for a number of energy consumption hours per year.

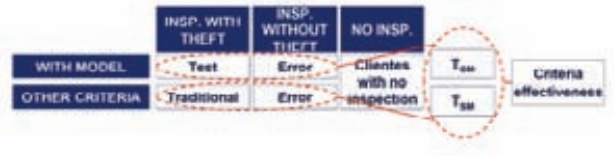
- **Energy consumption calculation methodology:** according to the information provided in the meter, the type of contract, the invoice cycles and the historical consumption. Different criteria can be used for different customers.

Measuring effectiveness

The non-technical loss management areas of energy companies invest in human, technical and economic resources for the execution of customer inspections.

The cost-effectiveness of these investments is determined by i) the observed success rates (in the inspected customers subgroup, the percentage of energy pilfering customers identified), ii) the energy gain (represented by the economic value of the recovery per customer), iii) the number of customers inspected in the target population, and iv) the unit cost associated with the customer inspection and therefore the total cost of the campaign.

Comparing the effectiveness of an inspection campaign determined by a model with another defined using different business criteria (e.g. report from the maintenance or meter reading officers, etc.) requires that conditions are similar in both populations, i.e. the structural propensity to steal needs to be very similar for both populations (e.g. areas with different economic backgrounds should not be compared).



Practical implementation

As was indicated in the validation selection, the algorithm selected was the one with the greatest discriminatory power. Using this model, we will next run through an example of practical implementation in the configuration of inspection campaigns.

The exercise is framed within an electricity distribution company with 5 million customers, where the **non-technical loss** amounts to 10%. The loss team aims to reduce non-technical losses by means of **inspections** with a unit cost 30 USD. Due to operational and economic restrictions, the number of annual inspections is limited to 100,000 (well below the number of energy pilfering customers, which amounts to around 500,000).

So far the prioritization criterion used has been based on the e**xpert judgement of the loss team,** which prioritizes the inspection of energy consumption levels that are close to 0, historically achieving inspection success rates of 9% (level of expected hits for randomly conducted campaigns).

The loss area starts to develop a discriminatory model by identifying all historical customer data available prior to the inspections conducted. As explained in section 4.2.4., after the information from the different data sources is unified, a quality data sample is selected, variables derived using business rules are built and the model selection methodology is developed.

The selected model is based on decision trees combined with machine learning techniques; specifically, neural networks for deep learning, based on variables mainly relating to i) **energy consumption pattern** – historical and recent moving averages, previous inspections, ii) **technical characteristics of the point of supply** – meters, connections, etc., and iii) **behavior** –repeat offences and other recovery-related variables such as willingness and ability to pay, etc. This last group of variables is especially relevant due to the **relationship between theft and default**. As a general rule, these are two closely related problems, since solving a theft problem sometimes leads to a default problem, and vice-versa, i.e. the inability or unwillingness to pay results in non-performance which, when dealt with e.g. though a supply cut, creates a theft incentive. Likewise, solving a theft problem can create a non-payment problem.

This model **increases the inspection success rate by up to 27%** (over one in four inspections are successful). The following table shows the cost-effectiveness of inspections before and after implementing the model.

Parameter	Before…(expert judgment campaign)	After…(campaign using model)
Cost per inspection	30 usd/inspection	30 usd/inspection
Capacity	100,000 inspections/year	100,000 inspections/year
Average energy gain	300 usd/theft	300 usd/theft
Succes rate	9%	27%
Cost of campaign	3,000,000 usd	3,000,000 usd
Income from campaign	2,700,000 usd	8,100,000 usd
Gain/loss from campaign	-300,000 usd	5,100,000 usd
Profitability	-10%	170%

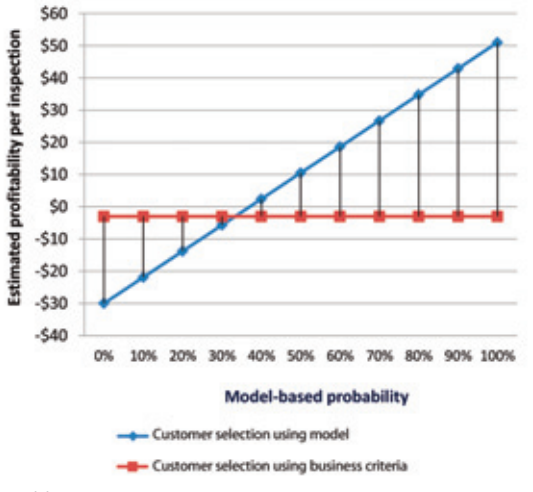
To summarize, trebling the inspection success rate and selecting the right cut-off point for conducting the inspections **increases campaign profitability and generates profits of more than 5 million USD for the company,** simply by defining inspection campaigns through techniques for prioritizing visits that are based on an analytical propensity to theft model.

The estimated profitability per inspection is defined as:

$$\textit{Profitability per inspection} \approx \textit{Success rate}*300 - 30 \quad \frac{USD}{inspection}$$

For instance, by conducting inspections on customers where **the expected success rate is above 50%, the expected profitability per inspection would always exceed the minimum required profitability level (10 USD).**

Fig.14. Impact on inspection profitability



Source: own elaboration

Conclusions

Now with an understanding of fraud (both internal and external) and its implications in terms of organization, processes and systems, we will focus on fraud management optimization techniques in the energy industry for cases of energy theft and fraud in the commercial process.

Among other **initiatives**, energy companies make substantial efforts to manage fraud through:

1. **Customer profiling and segmentation** that allow them to direct their inspection or mitigation actions.

2. Defining and implementing schemes to **quantify the usefulness of actions and measure their profitability** (e.g. profitability analysis on the acquisition of variables from external sources).

The use of **new modeling and machine learning techniques** in these processes can be a useful fraud detection tool to increase the success rate, industrialize the fraud detection process and reduce the cost of traditional inspections.

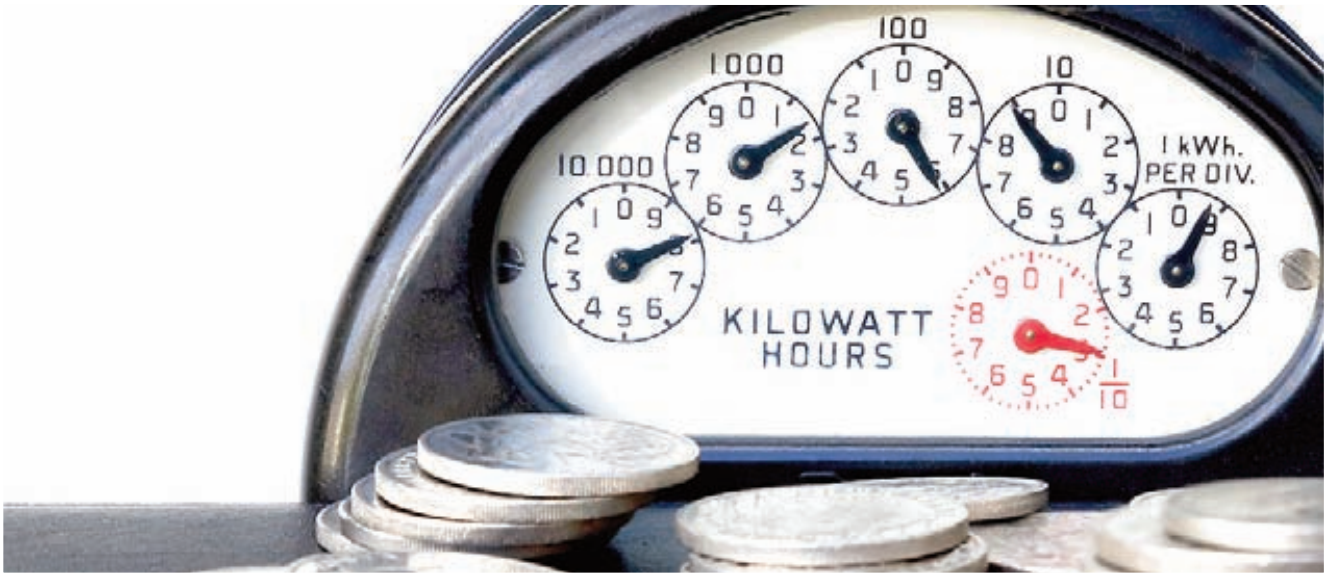
Thus, the **value contributed by data** and information on customers and transactions in determining the probability of occurrence of fraud is of particular importance. The availability of data (hourly consumption, customer data, access to systems, etc.) allows the use of both customer and employee profiling techniques and segmentation techniques for personalized management.

This will generate annual savings over one million US dollars thanks to data analysis on small companies (~2 million customers)[44]. For this reason, data **governance and data quality control** mechanisms are being implemented.

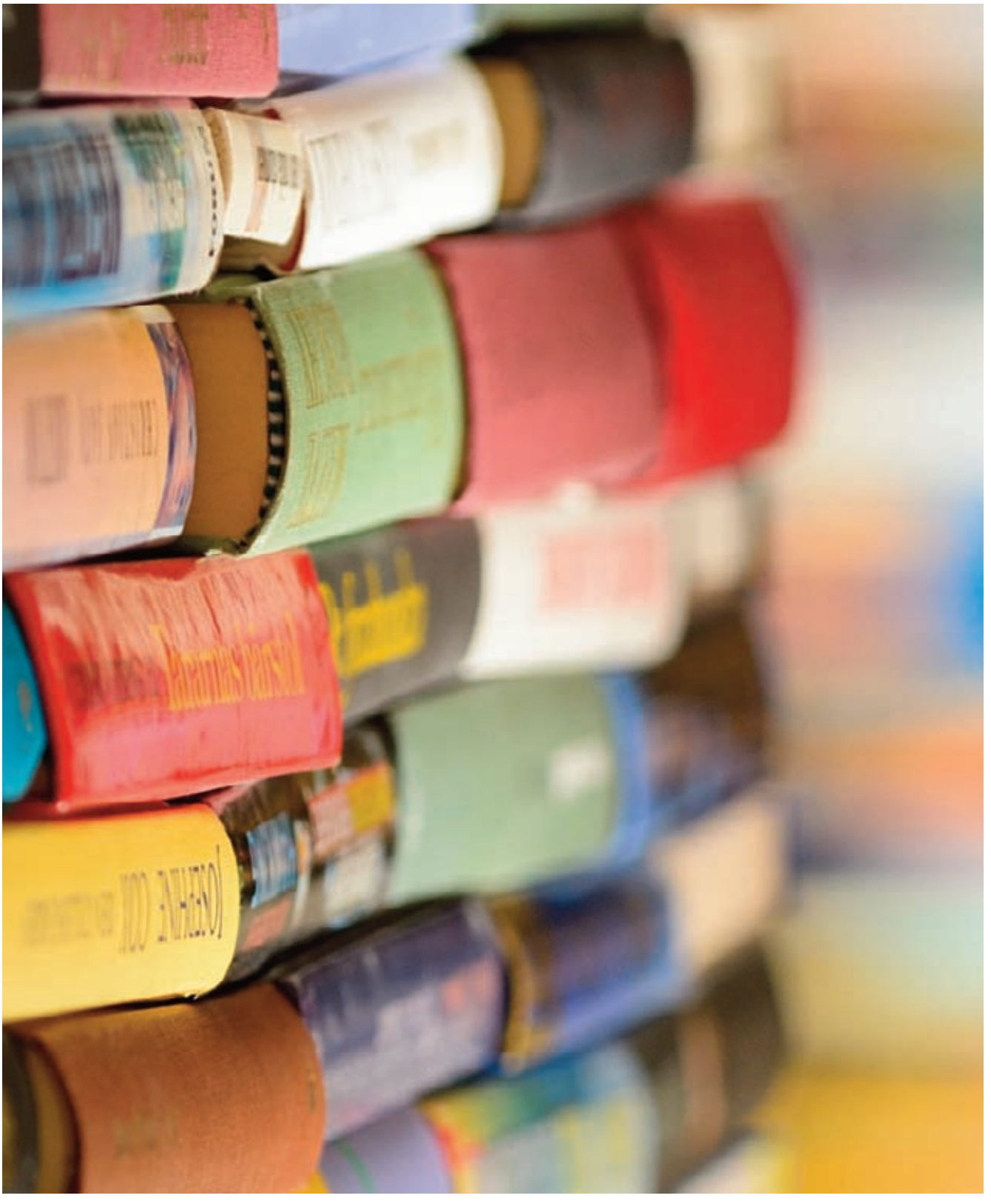
Furthermore, actions to detect and mitigate fraudulent events are in competition with other company investments. Which is why integrating them into the management process allows companies to measure their profitability.

All this is set within an **overall fraud management framework** to ensure the objectives from implementing statistical methods are accomplished.

[44] Advanced Metering Infrastructure and Customer Systems. Results from the Smart Grid investment grant program. September 2016, United States Department of Energy.



References



Report to the nations on occupational fraud and abuse. Global Fraud Study. ACFE (2016).

Fraud Risk Management in Organizations: A Practical Guide. IIA, Institute of Internal Auditors (2015).

International convergence of measures and capital standards. Basel: BCBS (2004).

Other People´s Money. Donald R. Cressey (1973).

Fundamental Elements of Cybersecurity for the financial sector. G7 Cyber Expert Group (2016).

Reducing Technical and Non Technical Losses in the Power Sector. Technical report. World Bank (2009).

Law 24/2013 on the Electric Power industry (Spain)

Data science and financial industry transformation. Management Solutions (2015).

Model Risk Management. Quantitative and qualitative aspects of model risk management. Management Solutions (2014).

Managing the Business Risk of Fraud: A Practical Guide. ACFE (2012).

Game-Theoretic Models of Electricity Theft Detection in Smart Utility Networks: Providing New Capabilities with Advanced Metering Infrastructure. IEEE Control Systems 35, no. 1 (2015).

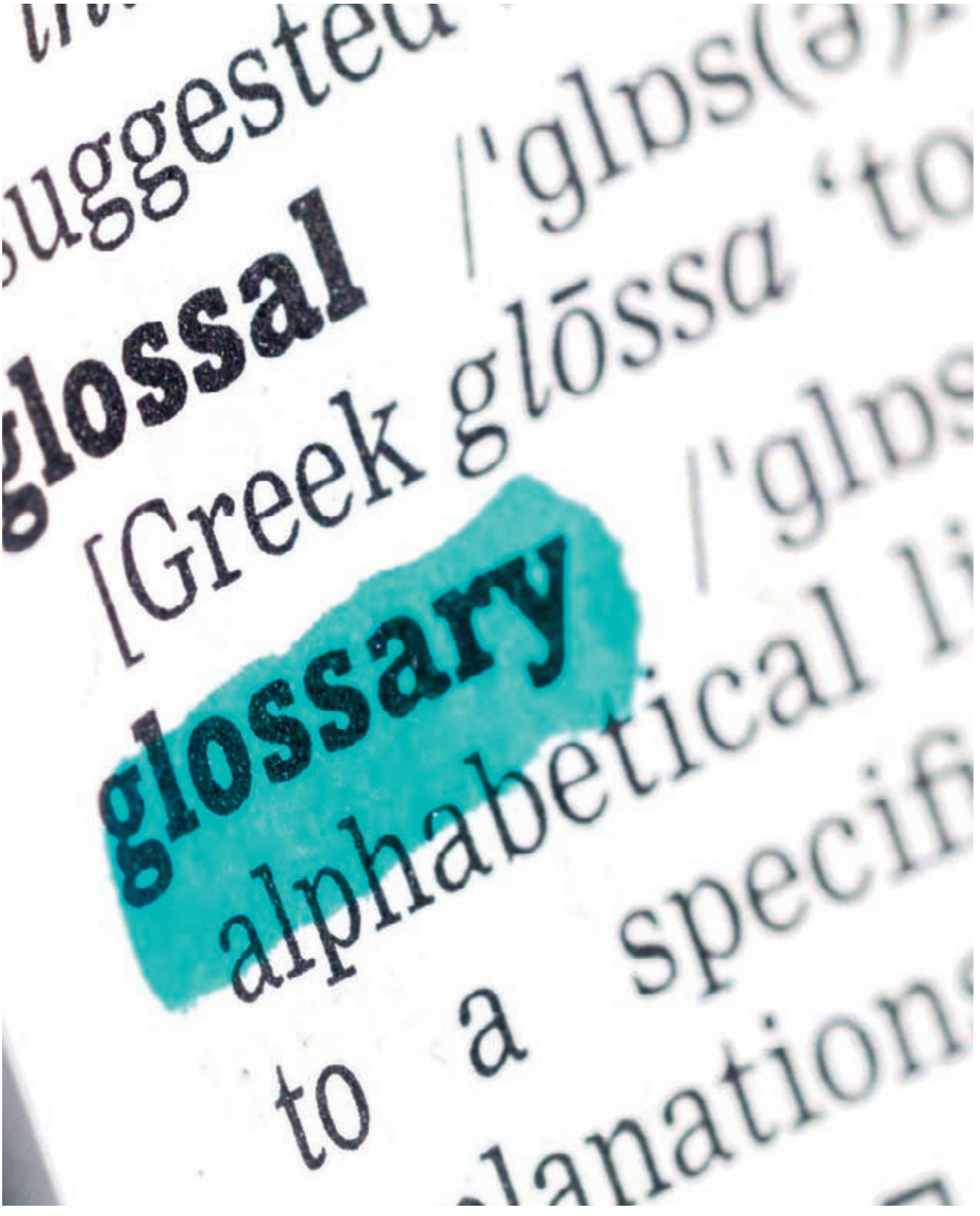
Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors. Northeast Group, LLC (2017).

Report on regulatory alternatives for loss reduction and treatment of fraud in the electricity supply. CNMC - National Commission on Markets and Competition (Report of 16 July 2015).

Spanish Organic Law on Data Protection (LOPD) and its implementing regulation (Royal Decree 1720/2007).

Smart Grid investment grant program. United States Department of Energy. (2016).

Glossary



ACFE (Association of Certified Fraud Examiners): established in 1988, it is a professional organization of fraud examiners. Its activities consist of creating fraud management tools, providing training and managing a knowledge database.

AMI (Advanced Metering Infrastructure): systems that are capable of measuring, collecting and analyzing energy usage and also interact with other devices such as smart meters for electricity, gas or water. They have the capacity to manage the information collected and make decisions. The difference between them and automatic reading systems is that with AMI there is two-way communication between the meter and the company's control center.

Backtest: term that refers to the testing of a predictive model using historical information to determine and/or ensure its profitability.

Chi squared: statistical test to verify the existence of a relationship between variables.

CFCA (Communications Fraud Control Association): global not-for-profit educational organization which focuses on fraud prevention in the telecommunications industry.

COSO (The Committee of Sponsoring Organizations of the Treadway Commission): a committee created by a joint initiative of private sector organizations which is dedicated to providing knowledge through the development of frameworks and guidelines on internal control, fraud prevention and risk management in companies.

Cramer's V: measure the intensity of the relationship between two or more categorical variables when at least one of the variables can take at least two possible values.

Data Lineage: is defined as the information life cycle including its source, movement and transformations. Describes what happens to information as it goes through different processes providing visibility so that errors and their source can be detected.

Deep Learning: set of automatic learning algorithms that attempt to learn data representations. An observation (e.g. an image) can be represented in many ways (e.g. a vector of pixels), but some representations make it easier to learn tasks of interest (e.g. "is this image a human face?").

Financial Fraud Action UK: body responsible for leading the collective fight against fraud on behalf of the UK financial iindustry. Its primary function is to facilitate the activity between the different actors involved in the fight against fraud.

IIA (Institute of Internal Auditors): Professional internal audit and risk management association established in 1941.

KDD (Knowledge Discovery in Database): process of extracting potentially useful information from a database. It is an iterative process that exhaustively explores very large data volumes to determine relationships.

KPI (Key Performance Indicator): metric used by organizations to measure the results of a specific action or strategy against predetermined objectives.

Machine Learning: method of data analysis that automates the process of creating analytical models. It uses an algorithm that iteratively learns from information, allowing tools to find hidden patterns without having to be explicitly programmed for it.

Data Mining: a computational process for discovering hidden patterns, trends and correlations through the extraction of a large amount of data.

NIST (National Institute of Standards and Technology): an agency of the US Department of Commerce's Technology Administration whose mission is to promote innovation and industrial competence in the United States through advances in metrology, standards and technology.

Phishing: attempt to obtain sensitive information such as user IDs, passwords, credit card information, etc., generally with malicious intent, by deceiving legitimate entities through electronic communication.

SM (Smart Meters): electronic equipment that captures energy consumption in intervals of one hour or less and in turn communicates the collected information to the company's control center for monitoring or electricity invoicing purposes.

SoD (Segregation of Duties): the concept of assigning more than one person to a task. It is a control measure that divides a task into sub-processes, assigning them to different parties, to prevent fraud.

Stream Computing: a computer system that analyzes multiple data streams from various sources, processing the information and transmitting it back in a single stream.

Our aim is to exceed our clients' expectations, and become their trusted partners

Management Solutions is an international consulting services company focused on consulting for business, risks, organization and processes, in both their functional components and in the implementation of their related technologies.

With its multi-disciplinary team (functional, mathematicians, technicians, etc.) of over 2,000 professionals, Management Solutions operates through its 24 offices (11 in Europe, 12 in the Americas and 1 in Asia).

To cover its clients' needs, Management Solutions has structured its practices by sectors (Financial Institutions, Energy and Telecommunications) and by lines of activity (FCRC, RBC, NT), covering a broad range of skills -Strategy, Commercial Management and Marketing, Organization and Processes, Risk Management and Control, Management and Financial Information, and Applied Technologies.

In the Energy industry, Management Solutions provides services to all types of companies - electricity, gas, petrochemical, etc. - both in global corporations and in local enterprises and public bodies.

Jesús Martínez
Socio de Management Solutions
jesus.martinez.gimenez@msspain.com

Manuel Ángel Guzmán
Gerente de I+D de Management Solutions
manuel.guzman@msspain.com

Javier Salcedo
Supervisor de Management Solutions
javier.salcedo@msbrazil.com

Diseño y Maquetación
Dpto. Marketing y Comunicación
Management Solutions - España

www.managementsolutions.com