ManagementSolutions
Making things happen

Gestão de fraude no setor energético

Energia

www.managementsolutions.com

Design e diagramação
Departamento de Marketing e Comunicação
Management Solutions - Espanha

Fotografias
Arquivo fotográfico da Management Solutions
iStock

Índice

Introdução

A fraude é atualmente uma das principais preocupações dos governos e empresas. De fato, estima-se que as perdas por fraude nas organizações podem oscilar entre 5% e 9% dos lucros anuais[1]. Para ter um maior entendimento dos diferentes âmbitos de gestão de fraude, é necessário conhecer em que consiste, seus componentes e as várias formas em que se pode apresentar.

No mundo empresarial, a fraude está associada a uma ação contrária à verdade e retidão, que prejudica a organização afetada. A fraude pode comprometer uma empresa, quer externamente pelos clientes, fornecedores e outras partes, ou internamente por funcionários, dirigentes ou proprietários.

O **contexto atual** apresenta, entre outras, as seguintes características e oportunidades:

- Disponibilidade crescente de **dados sobre clientes, funcionários, fornecedores,** etc., sua interação com a empresa e seus hábitos de comportamento.
- Existência de **técnicas de análise e quantificação da propensão ou probabilidade** de ocorrência de eventos de fraude.
- Avanço nas **metodologias e sistemas para combater a fraude interna** por meio da **segregação de funções (SoD, segregation of duties).**

A aportação **de valor** destes mecanismos de gestão está refletido tanto em sua vertente econômica (segundo um estudo da ACFE[2], as perdas por fraude a nível mundial reduziram 54% graças à adoção de medidas de monitoramento proativo de dados[3]), como de reputação e de conformidade. Esses últimos aspectos são especialmente relevantes dado o atual ambiente regulatório que fomenta o investimento e a implantação de meios para a gestão de fraude.

[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". Análise de um total de 2.410 casos de fraude ocupacional do mundo inteiro para o ano de 2016 (48% nos EUA).
[3] Através de, entre outros, análises de dados, supervisão dos dirigentes/managers, estabelecimento de um contato para receber denúncias, auditorias surpresa, etc.

O objetivo deste documento é compartilhar determinadas reflexões **sobre o conceito de fraude,** assim como sobre os principais elementos utilizados para a sua gestão e as oportunidades de otimização que emergem com os avanços tecnológicos, como, por exemplo, as tecnologias de Big Data e Analytics, entre outras. Estas são baseadas na disponibilidade e análise de grandes volumes de informação e na aplicação de metodologias de **perfil e segmentação.**

Particularizando no setor energético, neste documento são descritos eventos concretos de **fraude no setor energético** que, pela representatividade e consumo de recursos nas empresas, requerem um tratamento específico e onde as técnicas de detecção e sua integração na gestão têm maior relevância.

- Em relação à fraude externa, as empresas energéticas que distribuem eletricidade e/ou gás natural estão expostas ao furto de energia mediante ligações ou acessos fraudulentos à rede elétrica. A gestão deste tipo de fraude tem o apoio de métodos para quantificar a probabilidade de que uma medição não reflita o fornecimento real. Os métodos utilizados são variados (como regressões logísticas, redes neuronais, árvores de decisão, etc.), estão aplicados em esquemas de machine learning e estão orientados para distinguir fornecimentos "razoáveis" de fornecimentos potencialmente fraudulentos. Estas técnicas são apoiadas no uso de variáveis que caracterizam o cliente, seu perfil de consumo, hábitos de comportamento, etc. com o objetivo de identificar perfis ou comportamentos anômalos ou propensos ao furto de energia (por exemplo, reincidentes). Não é objeto deste documento o tratamento de ciberataques. Ainda que representem ameaças de suplantação de identidade ou intervenção das comunicações, gerando por exemplo, interrupções de fornecimento.

- No que respeita à fraude interna, a principal preocupação está centrada nas perdas associadas a eventos de fraude em processos críticos para a empresa, como pode ser o ciclo comercial de uma comercializadora energética. Estes acontecimentos estão situados principalmente nos processos de faturamento e cobrança, em que a possibilidade de alterar consumos, montantes, processos de compra ou dados bancários pode permitir subtrair receitas à empresa. A gestão deste tipo de fraude é realizadas por meio de metodologias orientadas para a segregação de funções, controle de acessos aos sistemas comerciais e econômico-financeiros e para a definição de indicadores e esquemas de reporting da existência de violações à segregação de funções.

Além disso, o presente documento mostrará como os métodos de modelagem, perfil e segmentação são complementados com a implantação de uma metodologia de **quantificação da utilidade econômica das atuações** que discrimina a qualidade da segmentação realizada para a detecção da fraude (ou efeito dos modelos de segmentação)

face ao beneficio da execução das atuações (ou efeito das próprias campanhas de detecção), com o objetivo de **avaliar a rentabilidade em separado do investimento em técnicas de modelagem do investimento em inspeções de detecção de furto.** Neste sentido, o investimento em gestão de fraude é avaliado como mais um investimento da empresa.

Estas técnicas estão suportadas em **plataformas de modelagem** que combinam componentes de tratamento massivo de dados com software estatístico e **ferramentas de controle de acessos** e gestão de funções, incompatibilidades, etc.

Por fim, são incluídos nesta publicação alguns **exemplos de aplicação** de técnicas de modelagem da probabilidade de detectar furto de energia (um caso particular de fraude externa). Estes modelos são apoiados na caracterização do ponto de fornecimento mediante variáveis que identificam os fatores subjacentes à fraude, tais como as características físicas do medidor ou contador, informação comercial e sociodemográfica do cliente ou usuário, histórico de consumo e comportamento em relação ao furto, outras formas de operar com o cliente, vinculação, reclamações ou resultado de inspeções, etc.

Portanto, é demonstrada a **aportação de valor dos dados,** da informação de clientes e operações (consumos horários, dados de clientes, acessos a sistemas, etc.), na quantificação das possibilidades de ocorrência de eventos de fraude e o seu uso para a otimização de atuações tanto preventivas (por exemplo, segregação de funções ou controle de acessos a sistemas) como mitigantes (por exemplo, execução de campanhas de inspeção e a segmentação de perfis segundo a propensão ao furto). Assim, a estimativa de probabilidades de ocorrência de um evento de furto ou a possibilidade de realizar atuações fraudulentas no ciclo comercial, combinadas com a materialidade dos potenciais impactos (energia defraudada, montantes subtraídos, etc.) permitem realizar uma **priorização das atuações sob uma fundamentação econômica e de rentabilidade.**

De fato, segundo os dados apresentados por uma das principais empresas europeias de distribuição de eletricidade, após o uso dos dados disponíveis com os medidores inteligentes, a porcentagem de casos de fraude detectados que afetavam a referida empresa passou de 5% para 50%[4].

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

Resumo executivo

Considerações de contexto

1. Embora não exista uma definição única, para efeitos do presente documento, qualificaremos como **fraude** qualquer ação ou omissão intencional concebida para enganar outros, resultando em uma perda para estes, e/ou em um ganho para o defraudador[5].

2. As práticas fraudulentas mais comuns podem ser agrupadas em dois domínios: **fraude externa** (por exemplo furto, suplantação, ataques cibernéticos, etc.) e **fraude interna** (fraude contábil, fiscal, operação em benefício próprio, etc.[6]).

3. Existem três **fatores**, que ao ocorrerem de forma simultânea implicarão um aumento da probabilidade de que uma pessoa cometa uma fraude:

    - Necessidade ou pressão, quer de índole econômica ou de outra natureza. Deve existir um incentivo ou uma necessidade (interna) ou pressões (externas), que incitem ou motivem a pessoa a cometer a ação de fraude.

    - Oportunidade percebida. Para que uma fraude ocorra deve existir uma debilidade a explorar em um determinado processo. O sujeito percebe uma forma de resolver seus problemas de forma fraudulenta com um baixo risco de ser descoberto.

    - Racionalização/Atitude. Justificação do ato delituoso. Neste sentido, influem os valores morais do sujeito, a percepção que o sujeito tem dos valores éticos que regem a empresa (vítima da fraude), assim como a valorização do benefício que implica a fraude perante as eventuais consequências negativas que pode acarretar em caso de ser descoberto.

4. Ao analisar os **eventos** de fraude reportados por cada setor, destacamos a porcentagem de incidentes associados ao setor bancário/financeiro, que continua sendo o setor que apresenta maior número de casos. Não obstante, ao analisar as perdas médias associadas a cada caso por atividade, o setor mineiro e o comércio por atacado são os que apresentam maiores perdas médias, estando o setor bancário em uma situação intermédia (face aos setores analisados em um estudo da ACFE[7]).

5. No caso do **setor energético,** embora com uma ocorrência menor segundo esse mesmo estudo (aproximadamente 5% dos casos analisados estão concentrados em utilities e corporações de oil&gas), apresenta algumas particularidades associadas ao furto de energia em que técnicas de modelagem podem fornecer um valor diferencial.

6. O processo de **transformação digital** em que estão imersos todos os setores implica uma maior exposição ao risco de fraude, porque os avanços tecnológicos são aproveitados pelos autores de fraude para adotar novas estratégias, não contempladas nos planos históricos de prevenção, detecção e atuação das empresas.

7. Devido ao **caráter mutável das práticas fraudulentas,** sua detecção é um processo contínuo e dinâmico que requer por parte das organizações ter definido um âmbito de atuação que inclua estratégias, enfoques e políticas concretas e específicas, e em que todas as áreas envolvidas atuem de forma coordenada. Neste contexto, as empresas foram estabelecendo uma política (ou políticas) para a gestão da fraude que, atendendo à origem do evento fraudulento, estabelece responsabilidades de gestão e controle.

8. Por tudo isso, tem especial interesse a implantação de um ***framework* de gestão de fraude,** cuja complexidade pode variar desde iniciativas simples de desdobramento de controles táticos (processos de autorização e validação, alarmes, inspeções, etc.) até a execução de projetos globais que, dando alcance à maioria dos processos da empresa, procuram estabelecer métricas de mensuração do risco de fraude nos mesmos e modificar os próprios processos e os sistemas para sua mitigação (implantação de plataformas de segregação de funções, controle de acessos, modelagem da propensão à fraude, sistemas informacionais e esquemas de reporting para a mensuração, etc.).

[5] 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.
[6] De acordo com o *framework* de Basileia II (BCBS: Convergência internacional de medidas e normas de capital), de aplicação no setor financeiro (embora esta definição seja de aplicabilidade geral).
[7] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study.

9. As **funções** de gestão de fraude em uma empresa estão habitualmente **dispersas**, sendo geralmente cada área de negócio afetada a responsável e promotora de iniciativas para a sua gestão, apoiada pelas áreas de tecnologia que lhe prestam serviço. Não obstante, em parte devido ao desenvolvimento de funções de conformidade nas empresas (como o CCO ou Chief Compliance Officer) e devido à busca da eficiência econômica e otimização de processos, existe uma tendência para a centralização de metodologias, indicadores e mecanismos de controle da fraude.

10. A gestão de fraude foi reforçada por **sistemas inteligentes** e de **análise estatística** para a sua detecção. Estas técnicas, que permitem detectar as novas estratégias e padrões utilizados pelos infratores, combinam elementos puros de análise e modelagem como data mining e machine learning, elementos técnicos de computação de alto rendimento como stream computing e, finalmente, processos completos de transformação de dados para a aquisição de conhecimento útil como knowledge discovery in database (KDD).

Técnicas de gestão de fraude aplicadas no setor energético

11. O aumento na capacidade de gerar, armazenar e processar **informação** pode ser aproveitado para ter acesso em tempo real à caracterização de um cliente, uma operação, um processo, etc., identificando assim comportamentos indicativos de propensão ao furto de energia. A utilização da ciência dos dados (Data Science) para a detecção da fraude no setor energético é de grande utilidade, por exemplo, para diferenciar a porcentagem de energia perdida na rede de distribuição associada a perdas técnicas (não representativas de um evento de fraude) e perdas não técnicas (furto de energia e portanto representativas de um evento de fraude). Com tudo isso, o investimento em modelos de detecção avançados otimiza as taxas de sucesso das campanhas de inspeção e melhora a detecção da fraude. Em todo o caso, para avançar com garantias na implantação deste tipo de modelos convém previamente definir e implementar um *framework* de referência que facilite a governança dos dados, modelos e processos associados.

10. Os **modelos** a desenvolver pretendem encontrar padrões, tendências ou regras que expliquem o comportamento do cliente antes da detecção da fraude. As técnicas a utilizar irão variar em função do objetivo pretendido e do tipo de dados utilizados. O aproveitamento de todo o potencial que as técnicas de Data Science representam incorpora dois elementos chave:

    - A análise em tempo real. Os sistemas de compilação de informação permitem realizar a coleta e o acompanhamento dos dados em tempo real. Além disso, é possível programar algoritmos que façam uso dessa informação, facilitando e agilizando a detecção dos novos padrões e estratégias de fraude não utilizadas até esse momento pelos infratores.

    - O re-treinamento automático e autoaprendizagem. Os modelos de detecção da fraude são recalibrados de forma automática (com escassa intervenção dos analistas) e iterativamente a partir dos grandes volumes de dados, permitindo uma potencial melhoria do poder de previsão durante os sucessivos re-treinamentos.

Exemplo de modelagem aplicado ao furto de energia

13. Um dos usos mais amplos do Data Science na gestão da fraude no setor energético consiste em maximizar a eficiência das **campanhas de inspeção.** A detecção da fraude energética, associada ao consumo ilegal de energia na rede, parte de uma segmentação dos clientes baseada em sua **probabilidade de cometer fraude.** As técnicas de modelagem aplicadas na gestão da detecção de fraude energética permitem melhorar a taxa de sucesso na seleção de clientes a serem inspecionados.

14. Foi demonstrado que algumas **variáveis** têm alto poder de previsão: dados do medidor, dados sociodemográficos, dados de consumo histórico de energia, dados da operação ou gestão realizada, manutenção da rede e dos medidores, informação de cortes e irregularidades ou informação de contatos ou reclamações do cliente, etc.

15. Com o objetivo de identificar os **modelos** que melhor explicam o comportamento dos clientes fraudulentos foram fixados alguns critérios mínimos que devem cumprir os resultados obtidos pelo modelo selecionado, tais como a sua capacidade discriminante. Adicionalmente à validação estatística, foram validados os modelos com as inspeções realizadas durante seis meses e foi observado que, seja qual for o número de inspeções, as técnicas de machine learning são as que permitem gerar segmentos com uma maior concentração de defraudadores.

16. As áreas de gestão da perda não técnica das empresas energéticas investem recursos humanos, técnicos e econômicos na execução de inspeções a clientes. A **rentabilidade** destes investimentos é determinada por:

    - as taxas de sucesso observadas (do subconjunto dos clientes inspecionados, que porcentagem de clientes com furto de energia foi identificada);

    - o ganho de energia (representado como valor econômico de recuperação por cliente);

    - o número de clientes inspecionados da população-alvo;

    - e o custo unitário associado à inspeção do cliente e, portanto, o custo total da campanha.

17. Foi realizado um **exercício quantitativo** aplicando a metodologia exposta, e foi selecionado um algoritmo específico pelo seu maior poder discriminante. Com este modelo foi desenvolvido um exemplo de aplicação prática na configuração de campanhas de inspeção. No exemplo, a taxa de sucesso das inspeções é multiplicada por 3, chegando a 27% (mais de uma de cada quatro inspeções têm sucesso).

Gestão da fraude

Contexto

Conceito de fraude

Não existe uma definição unificada e homogênea das práticas que são consideradas fraude. De fato, uma das preocupações dos organismos internacionais é precisamente definir um *framework* regulatório unificado, em que sejam estabelecidos critérios comuns sobre as práticas que devem ser consideradas como fraudulentas, assim como as sanções a aplicar em cada caso.

Não obstante, de acordo com a ACFE[8] é definida fraude como ***"qualquer ação ou omissão intencionada concebida para enganar os outros, resultando em uma perda para estes, e/ou em um ganho para o defraudador"***[9] . A fraude no contexto empresarial está associada a uma ação contrária à verdade e retidão, que prejudica a organização afetada. A fraude pode comprometer uma empresa, seja originada externamente nos clientes, fornecedores e outras partes, ou internamente por empregados, dirigentes ou proprietários.

Portanto, as várias práticas fraudulentas mais comuns podem ser agrupadas nestes dois domínios: fraude externa (por exemplo furto, suplantação, ataques cibernéticos, etc.) e fraude interna (fraude contábil, fiscal, operação em benefício próprio, etc.). É habitual considerar tanto a fraude externa como interna como duas categorias de risco operacional que as empresas identificam, mensuram e gerenciam.

É consideradas fraude externa o evento que faz sofrer uma perda inesperada de tipo financeiro, material ou de reputação devido a atos fraudulentos realizados por uma pessoa externa à empresa. São definidas como *"perdas decorrentes de atuações por um terceiro destinadas a defraudar, a apropriação de bens indevidamente ou a contornar a legislação"*[10].

Isto pode ser realizado por clientes, ou por concorrentes ou terceiros:

- Clientes que utilizam bens ou serviços de forma fraudulenta, sem pagar, falsificam meios de pagamento, manipulam processos de compras, etc.
- Concorrentes, fornecedores, terceiros em geral, que manipulam licitações, faturam à empresa por bens ou serviços não prestados, oferecem subornos a funcionários, etc.

Nesta segunda categoria, as organizações enfrentam em particular ameaças de violações da segurança e roubos da propriedade intelectual cometidos por terceiros desconhecidos (por exemplo, por meio de ataques cibernéticos). Outros exemplos de fraudes são a pirataria, o roubo de informação confidencial, a fraude fiscal, a falência fraudulenta, a fraude associada a seguros, a fraude associada ao atendimento médico, etc.

A **fraude interna** é a originada no interior de uma organização e é cometida pelos seus funcionários contra essa organização/empregador. Define-se como *"perdas decorrentes de atuações destinadas a defraudar, apropriar bens indevidamente ou a contornar regulações, leis ou políticas empresariais em que está envolvida, pelo menos, uma parte interna da empresa em benefício próprio"*[11]. Costumar ter origem em um conflito existente entre os interesses pessoais de um funcionário ou grupo de funcionários e os da organização. Estima-se que, em média, uma organização perde 5% dos seus lucros anuais só como resultado da fraude originada internamente[12].

Fatores subjacentes à fraude

Na hora de gerenciar o risco de fraude, as empresas devem identificar e monitor os diversos fatores que podem motivar um evento de fraude. Uma das ferramentas utilizadas para a avaliação do risco de fraude é o conhecido como *"triângulo da fraude"*[13] (ver fig.1). Esta ferramenta constitui um modelo amplamente aceito no momento de explicar os fatores subjacentes que motivam uma pessoa a cometer fraude.

Segundo esta teoria, existem três fatores, que ocorrerem de forma simultânea implicarão um aumento da probabilidade de que uma pessoa cometa uma fraude:

- **Necessidade ou pressão.** Motiva o delito em primeiro lugar. Representa a pressão a que está submetido o sujeito devido à existência de um problema que não é capaz de

[8] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study
[9] Fonte: "Gestão do Risco de Fraude nas Organizações: Um Guia Prático." IIA, Institute of Internal Auditors
[10] Segundo a definição de Basileia: BCBS: Convergência internacional de medidas e normas de capital. Junho 2004.
[11] Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study. ACFE, Association of Certified Fraud Examiners.
[12] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study.
[13] Conceito desenvolvido pelo Dr. Donald R. Cressey, sociólogo e criminalista, cuja investigação foi centrada em malversadores a quem chamou "violadores de confiança" (Other People´s Money, Montclair: Patterson Smith, 1973).

Fig. 1. Triângulo da fraude

resolver por meios legítimos, levando a considerar atos ilegais como meio para solucionar o referido problema. Estas pressões podem ser de índole econômica ou de outra natureza. Deve existir um incentivo ou uma necessidade (interna) ou pressões (externas), que incitem ou motivem o indivíduo a cometer a ação de fraude.

- **Oportunidade percebida.** Define o método pelo qual será cometido o delito. Para que uma fraude tenha lugar deve existir uma debilidade a explorar em um determinado processo (por exemplo, ausência de controles, pouca segregação de funções ou ausência de um *framework* de gestão da fraude correto e atualizado). O sujeito percebe uma maneira de resolver os seus problemas de forma fraudulenta com uma baixa assunção do risco de ser descoberto.

- **Racionalização/Atitude.** Justifica e valida o ato delituoso; ou seja, refere-se à habilidade do indivíduo para racionalizar e justificar internamente os atos incorretos que implicam o ato fraudulento. Neste sentido, têm influência valores morais do sujeito, a percepção que tem dos valores éticos que regem a empresa (vítima da fraude), assim como a valorização do benefício que implica a fraude face às possíveis consequências negativas que pode acarretar no caso de ser descoberto. A posição individual face ao risco e à honestidade são aspectos fundamentais e determinantes.

O valor do triângulo da fraude identifica os fatores alvo que têm que estar presentes para que ocorra um evento de fraude. O reconhecimento destes fatores objetivos ajuda na definição das ações a tomar para prevenir, detectar e dar resposta à fraude.

Fraude por setores de atividade

Os casos de fraude são um problema comum em todos os setores, reforçando a necessidade de estabelecer controles e mecanismos que minimizem a existência desses acontecimentos.

Ao analisar os eventos de fraude reportados por cada setor, destacamos a porcentagem de incidentes associados ao setor bancário/financeiro, que continua sendo o setor que apresenta maior número de casos (ver fig. 2).

Ao analisar as perdas médias associadas a cada caso por atividade, o setor mineiro e o comércio por atacado são os que apresentam maiores perdas médias, tendo o setor bancário em uma situação intermediária (face aos setores analisados em um estudo da ACFE). O setor energético, embora com uma ocorrência menor segundo esse mesmo estudo (aproximadamente 3,4% dos casos analisados estão concentrados em utilities e corporações de oil&gas), apresenta algumas particularidades associadas ao furto de energia em que técnicas de modelagem podem proporcionar um valor diferencial.

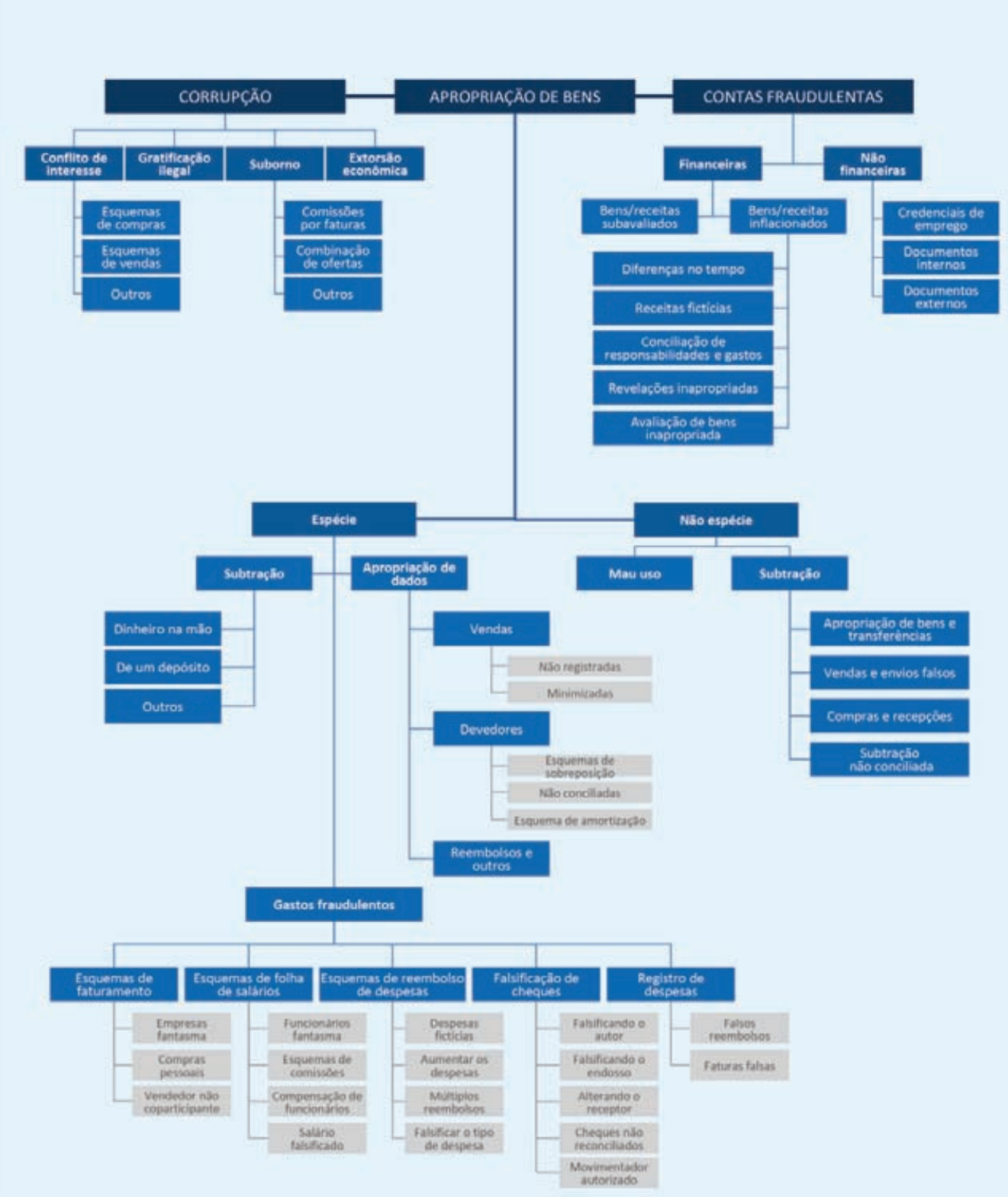
Fig. 2. Distribuição de casos por setores de atividade[14]

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

[14] Outros: Serviços profissionais e sociais, Agricultura e Pesca, Imobiliário, Utilities, Arte e Entretenimento, Comércio por Atacado, Prospecção Mineira e Comunicações.

Sistema de classificação da fraude e de abuso profissional

Este tipo de fraude pode ser classificada de acordo com a seguinte "árvore" ou esquema de fraude[15]:



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

Um elemento comum a todos os setores é o **processo de transformação digital** em que estão imersos. Tal implica uma maior exposição ao risco de fraude, porque os avanços tecnológicos são aproveitados pelos autores de fraude para adotar novas estratégias, não contempladas nos planos históricos de prevenção, detecção e atuação das empresas.

Aparece assim o conceito de **cibersegurança**, definido como a capacidade para proteger ou defender o uso do ciberespaço dos ataques cibernéticos[16]. Nos últimos anos, foi agravado o risco associado à cibersegurança decorrente principalmente de três fatores: i) o desenvolvimento tecnológico, ii) os processos de reestruturação do setor e iii) a profissionalização dos atacantes.

Não obstante, a digitalização dos vários setores é também uma vantagem para as organizações no momento de lutar contra a fraude, sobretudo relativamente à prevenção e detecção. O acesso a **grandes volumes de informação,** assim como o desenvolvimento de novas técnicas e modelos que permitem a **análise do comportamento dos clientes** (mediante técnicas de segmentação avançadas por exemplo, machine learning) constituem ferramentas eficazes para combater a fraude e são aplicáveis nos vários setores[17].

Estas soluções foram concebidas para analisar automaticamente as várias operações registradas nos sistemas das empresas. Devem ser capazes de processar os grandes volumes de dados disponíveis para detectar em tempo real os vários padrões e estratégias concebidas pelo defraudador. Estas técnicas melhoram os mecanismos atuais de prevenção e detecção das atividades fraudulentas, porque são capazes de detectar de forma dinâmica padrões e estratégias fraudulentas não utilizados anteriormente.

Devido ao **caráter mutável das práticas fraudulentas,** a detecção é um **processo contínuo e dinâmico,** requerendo por parte das organizações ter definido um âmbito de atuação que inclua estratégias, enfoques e políticas concretas e específicas, e em que todas as áreas envolvidas atuem de forma coordenada. Não obstante, o tratamento da fraude apresenta por vezes debilidades como consequência da existência de enfoques parciais e habitualmente dispersos; processos em parte externalizados e não integrados na gestão diária do negócio; falta de coordenação entre as áreas responsáveis da prevenção, detecção e resposta face a eventos de fraude (equipe antifraude, auditoria interna, segurança da rede, segurança de sistemas, etc.); diferenciação nem sempre clara entre Primeira Linha de Defesa (Gestão) e Segunda Linha de Defesa (Controle), etc.

A seguir são apresentados alguns exemplos de tipos de fraude comuns nos setores bancário, segurador, telecomunicações e energético.

Setor bancário/financeiro

Segundo foi apresentado na figura anterior, este pode ser um dos setores mais afetados pela fraude em relação ao número de ocorrências. Além disso, os eventos de fraude no setor afetam todos os produtos oferecidos pelos bancos (cartões de crédito e débito, contas correntes, cheques, empréstimos, etc.), a todos os canais (agências, banco on-line/telefônico, transações remotas), a todos os sistemas de suporte tecnológico e a todo tipo de clientes (varejista, por atacado). Inclui práticas desde pagamentos e utilização de cheques fraudulentos até phishing ou usurpação de identidade, etc.

Portanto, a mitigação da fraude e a melhoria da cibersegurança estão entre uma das principais preocupações das empresas. De acordo com o G7 Cyber Expert Group[18], o objetivo principal é identificar as práticas que representam uma eventual fraude. As perdas por custo de oportunidade (transações normais classificadas como potencialmente fraudulentas), ou falsos positivos devem ser minimizados, pois no caso da fraude este tipo de erro apresenta um impacto negativo muito elevado na percepção que os clientes têm sobre a empresa.

Setor segurador

Neste setor destacamos os delitos associados aos seguros do lar, seguros de vida, seguros trabalhistas, seguros sobre meios de transporte e seguros médicos. Podem ser classificadas as fraudes em duas categorias[19]:

- "***Hard fraud***". Tem lugar quando o infrator obtém dinheiro ilegalmente por meio de uma estratégia premeditada. Pode envolver mais de uma pessoa e inclusive uma pessoa que trabalhe para a própria empresa de seguros, que atue de forma coordenada com o titular ou o beneficiário do seguro. Este tipo de fraude ocorre, por exemplo, quando alguém provoca um acidente rodoviário de forma deliberada com o objetivo de cobrar o seguro do carro.

- "***Soft fraud***". Tem lugar quando o infrator interpõe uma reclamação legítima, mas aproveita a conjuntura para mentir à empresa seguradora sobre o dano sofrido. Este tipo de fraude ocorre, por exemplo, quando há um acidente de carro de forma fortuita e o condutor reporta danos maiores do que aqueles realmente sofridos. É o tipo de fraude que ocorre com maior frequência.

[16] Segundo o NIST: National Institute of Standards and Technology.
[17] Fonte: Management Solutions (2015): Data science e a transformação do setor financeiro. Management Solutions (2014): Model Risk Management. Aspectos quantitativos e qualitativos da gestão do risco de modelo.
[18] Fonte: Fundamental Elements of Cybersecurity for the financial sector. Outubro de 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] Fonte: Insurance Information Institute.

Setor das telecomunicações

As principais categorias de fraude neste setor podem afetar tanto a empresa de telecomunicações fornecedora do serviço como o cliente que subscreve um contrato legal com a empresa de telecomunicações, podendo afetar tanto de forma conjunta como individualmente. Podem ser distinguidos três esquemas de fraude em função do objetivo da fraude:

- **Aumento do tráfego de chamadas.** Este esquema aplica técnicas de simulação de acesso à rede para aumentar o tráfego de chamadas para determinados destinos (onde as tarifas a pagar entre operadoras são mais elevadas), de modo que operadoras ilegais beneficiam dessas tarifas.

- **Manipulação dos sistemas de informação das empresas fornecedoras de serviços.** Destacamos a fraude associada aos sistemas que fornecem acesso à rede ou trunking SIP.

- **Fraude telefônica.** Inclui o envio de spam ou de mensagens de texto com o objetivo de obter dados pessoais dos usuários, chamadas telefônicas a instituições financeiras usurpando a identidade de um cliente, ou o colapso da rede para impedir o funcionamento normal de uma empresa ou sistema.

Menção especial merecem as fraudes conhecidas como "International Revenue Share Fraud (IRSF)"[20] e as técnicas de "by pass"[21] ilegais das redes.

Setor energético

No setor energético a fraude pode ocorrer em várias atividades tais como o trading de energia, os aprovisionamentos, o gerenciamento de projetos, a gestão comercial, etc.; contudo, uma parte da fraude existente neste setor é relativo ao furto de energia, ao não faturamento de energia consumida (mais de 80% dos eventos de fraude estão relacionados com a apropriação indevida[22]). Neste sentido, devemos mencionar que um dos principais problemas que impacta a eficiência e segurança das empresas de energia são as perdas associadas ao processo de distribuição e fornecimento aos consumidores. Estas perdas podem ser decompostas em duas categorias[23]:

- **Perdas técnicas**. Associadas a perdas que ocorrem de forma natural na rede, devido a fenômenos como a dissipação de potência nas linhas de transmissão, etc. e portanto não podem ser consideradas como perdas devido a eventos de fraude.

- **Perdas não técnicas.** Associadas à fraude energética, causadas por ações externas ao sistema de fornecimento de energia, que consistem principalmente em furtos de energia, inadimplência de clientes e perdas por erros nos processos de faturamento.

[20] Consiste em esquemas de fraude, na qual o defraudador aumenta o tráfego de chamadas para um destino com tarifas elevadas, sem o consentimento da operadora nem do cliente. Neste caso, os sobrecustos são imputados às operadoras. O lucro deste tipo de fraude é partilhado pela operadora ilegal do país de destino das chamadas e pelo sujeito que gera as chamadas, aumentando o tráfego associado à operadora ilegal.

[21] Consiste na receita de tráfego de chamadas internacionais para um país de forma ilegal, com o único propósito de evitar pagar as devidas taxas contábeis entre operadores. Tendem a utilizar SIM Box ou chamadas VoIP para falsificar o registro do origem da chamada.

[22] Association of Certified Fraud Examiners (ACFE): Report to the nations on occupational fraud and abuse. 2016 Global Fraud Study. A análise de frequência de eventos de fraude por categoria demonstra que mais de 80% dos casos se podem tipificar como "Asset Misappropriation".

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

A identificação e diferenciação da porcentagem de energia perdida na distribuição por motivos não técnicos é um desafio que as empresas do setor devem enfrentar e resolver. Podem ser distinguidas três categorias[24]:

- As ações que incidem sobre a rede da empresa distribuidora. Destacamos as ligações diretas à rede de distribuição sem ter subscrito nenhum contrato de fornecimento de energia elétrica e as derivações no fornecimento de energia para outros pontos ou instalações não abrangidos no contrato.

- As ações que incidem sobre os equipamentos de medida e controle tais como a manipulação dos medidores, com o objetivo de falsificar os registros de consumo e comunicar um consumo inferior ao real.

- As ações desonestas realizadas pelos funcionários das empresas e que afetam o ciclo comercial. Podem ocorrer, por exemplo, quando não há uma correta segregação de funções e uma mesma pessoa registra ou modifica as operações e autoriza os pagamentos ou faturamento associados às mesmas.

Estes três tipos de fraude podem constituir uma falta ou delito, penalizado muitas vezes com sanções, cujo montante depende da quantidade de energia defraudada. De modo geral, as empresas distribuidoras devem detectar e dar conhecimento às autoridades as irregularidades na rede e nos equipamentos (por exemplo, na Espanha, as multas são da competência das Comunidades Autônomas e é estabelecido por lei o novo faturamento do montante correspondente a seis horas[25]).

Alavancas de gestão

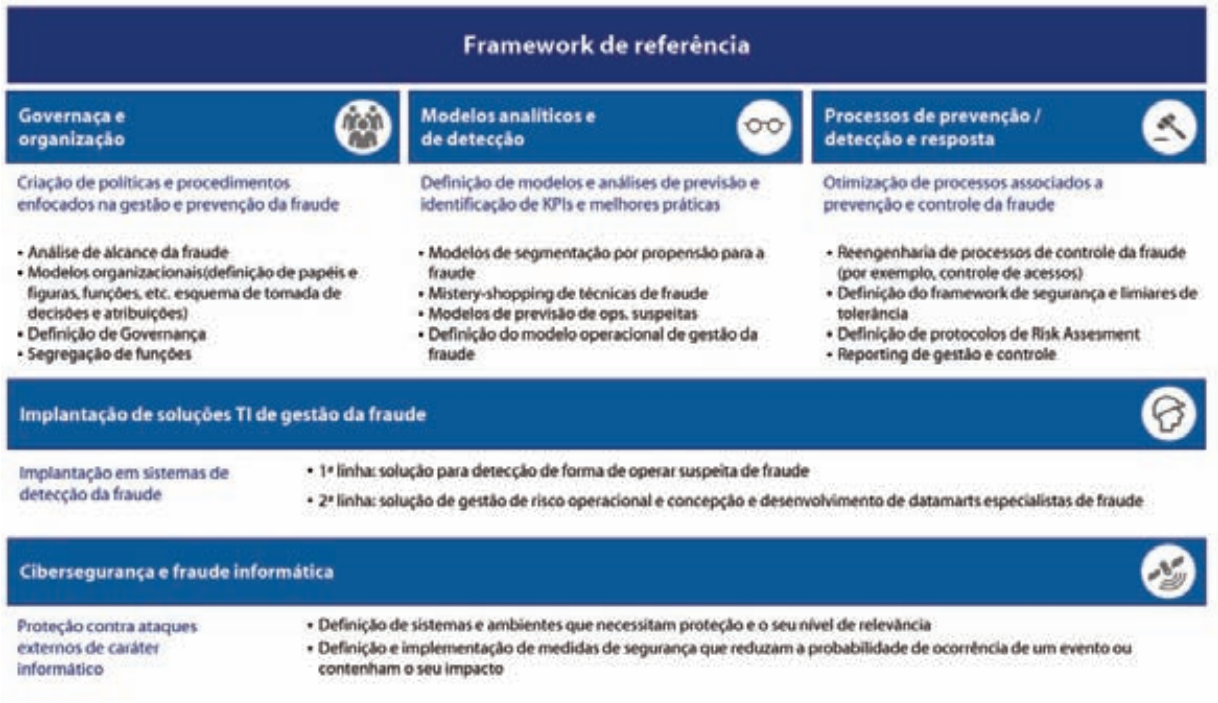
Um âmbito integral para a gestão da fraude daria resposta a perguntas como:

- As unidades de gestão da fraude devem ser integradas em uma área?
- Quem é o responsável como segunda linha de defesa da fraude interna?
- Que capacidades (modelos/sistemas) são necessárias para anteciparmos a fraude?
- Que percursos seguem as novas tendências como big data/machine learning em sua gestão?
- Etc.

O referido âmbito poderia ser estruturado em torno de princípios/elementos básicos[26] tais como (ver fig. 3): definição de um modelo de governança, estabelecimento de avaliações periódicas do risco de fraude, implementação de técnicas e processos de prevenção, detecção e ações corretivas e de resposta com que atuar para minimizar as perdas uma vez que a fraude ocorreu. Como ações transversais, destacamos a implementação dos elementos anteriores nos sistemas e a integração dos sistemas de cibersegurança na operação diária das empresas (para enfrentar a fraude informática, um dos tipos de fraude mais importantes e comuns, devido à digitalização dos setores).

[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] Segundo a Lei 24/2013 do Setor Elétrico.
[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. Âmbito integrado de referência para prevenir e gerir a fraude



Fonte: elaboração própria.

Governança e organização

Como parte da estrutura de governança de uma organização, as empresas estabelecem uma política (ou políticas) para a gestão da fraude que, atendendo à origem do evento fraudulento, estabelece responsabilidades de gestão e controle.

- Para eventos de fraude com **origem externa** é habitual que a identificação e gestão seja feita a partir das próprias unidades de negócio por meio da aplicação das políticas antifraude nos seus processos operacionais. Paralelamente, e sem uma dependência do negócio, a supervisão da correta aplicação destas políticas é realizada a partir das áreas de controle de riscos, controle interno e auditoria, funções de segunda e terceira linhas de defesa.

- Para eventos de fraude com **origem interna,** no entanto, a sua identificação e gestão é feita por meio de áreas independentes do negócio, como controle interno ou auditoria, a partir de onde são realizadas as correspondentes investigações para obter conclusões que permitam a tomada de medidas disciplinares. De fato, segundo o Instituto de Auditores Internos da Espanha[27], a *"Auditoria Interna deve assegurar ao Conselho e à Administração que os controles em matéria de fraude são suficientes para cobrir os riscos identificados e garantir que esses controles funcionam de forma eficaz"*. Estas funções, além de estabelecerem os seus processos de monitoramento, contam habitualmente com canais de denúncia interna anônima.

No momento de definir as políticas antifraude, as organizações consideram o nível de complexidade e profundidade que querem alcançar, sendo um fator relevante o tamanho da própria empresa.

Inteligência analítica

Segundo o Institute of Electrical and Electronics Engineers, a probabilidade de detecção da fraude depende da quantidade roubada e do nível de investimento realizado na referida detecção[28]. Devemos, portanto, **avaliar periodicamente a exposição** da organização ao risco de fraude, identificando eventuais novas estratégias de fraude e eventos que devem ser mitigados pela empresa. Além de identificar os riscos, devemos **rever a probabilidade de ocorrência e a sua gravidade** em caso de ocorrência de um evento de fraude.

Baseados no triângulo da fraude apresentado na figura 1 é realizada uma revisão dos riscos e uma definição dos indicadores analíticos para acompanhamento em relatórios que permitam identificar e antecipar a propensão à fraude.

A gestão da fraude, devido ao seu caráter adaptativo, requer sistemas inteligentes e análise estatística para a sua detecção. As técnicas que permitem detectar as novas estratégias e padrões utilizados pelos infratores sem perder vigência no tempo combinam **elementos puros de análise e modelagem** como

[27] IAIE: Gestão do Risco de Fraude: Prevenção, detecção e investigação. Fevereiro 2015.
[28] Fonte: 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).

data mining e *machine learning,* **elementos técnicos** de computação de alto rendimento como *stream computing* e, finalmente, **processos completos de transformação de dados** para a aquisição de conhecimento útil como *knowledge discovery in database*" (KDD). Estes métodos devem ser utilizados de forma anterior à aplicação dos controles internos (ver fig. 4).

Alguns benefícios de utilizar análise estatística dos dados são[29]:

- **Visão holística de uma empresa.** Carteira de clientes ativos para os quais são integradas diversas fontes de dados (internas e externas) com a possibilidade de enriquecer a informação interna fruto da relação com o cliente (comportamento de pagamento, incidentes, consumo, etc.) com informação externa proporcionada por terceiros (poder de compra, inadimplência, nível socioeconômico, etc.).

- **Análise de informação desestruturada.** Dados provenientes de redes sociais, conversações, etc. representam uma informação valiosa para detectar a fraude; no entanto, as bases de dados tradicionais não permitiam o seu armazenamento de forma apropriada. As novas técnicas permitem um correto armazenamento destes dados, assim como a sua exploração e inclusão nos modelos preditivos.

Não obstante, os eventos não óbvios e com escasso número de incidentes devem ser identificados e tratados com um critério de negócio ou política definida.

Em último lugar devemos realizar a validação dos modelos analíticos de fraude, ou seja, deve ser definido e implementado o processo de supervisão dos modelos com o objetivo de confirmar que o modelo final tem um desempenho constante e correto, além de cumprir os requisitos de negócio e, potencialmente, regulatórios.

Mecanismos de prevenção, detecção e resposta

Do ponto de vista operacional, as atuações estão orientadas para a prevenção e detecção de potenciais eventos fraudulentos (implantação de técnicas de detecção dos eventos fraudulentos que não foram previstos pelos sistemas da empresa), assim como para a definição de processos de resposta para assegurar que a eventual fraude é abordada de forma apropriada e atempada com o objetivo de minimizar a perda associada ao evento fraudulento.

- **Identificação** por meio da definição de que processos, dados, sistemas e ambientes requerem proteção e o seu nível de relevância.

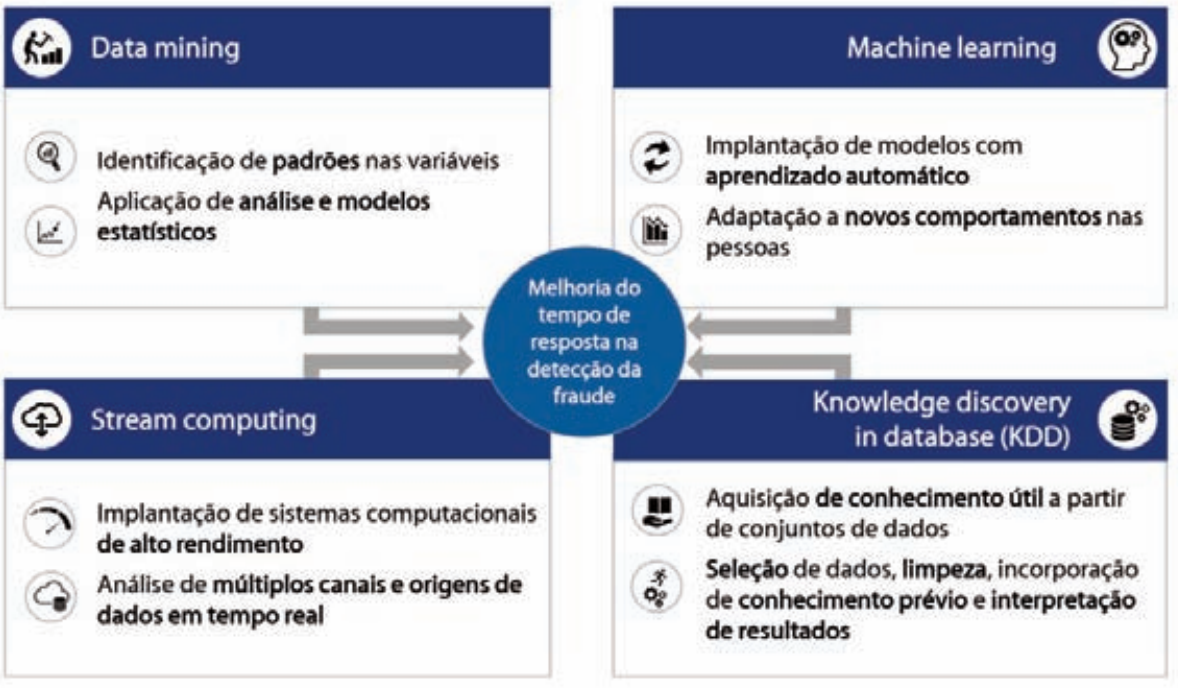
- **Proteção** por meio da implantação de controles, quer sejam sistemas ou formas de operar de controle, que reduzam ex-ante a exposição ao risco ou impeçam que uma ameaça seja convertida em fraude (por exemplo, políticas de acessos);

- **Detecção** por meio de sistemas de alerta precoce que, por meio da monitoramento ou análise, permitam a identificação da ocorrência de uma fraude ou forma de operar fraudulenta (por exemplo, *dashborad* de KPIs/KRIs); e

- **Resposta e recuperação** utilizando processos ágeis de colocação em funcionamento de medidas corretivas para minimizar e/ou resolver as perdas causadas pelo evento fraudulento (reduzir o impacto).

[29] Um exemplo de aplicação prática destes benefícios pode ser encontrado em 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. Inteligência analítica aplicada à gestão da fraude



Fonte: elaboração própria.

Modelo de suporte tecnológico

A implementação dos elementos anteriores nos ambientes tecnológicos das organizações é critica para a gestão da fraude[30]:

Um sistema integrado de prevenção e detecção da fraude facilita a gestão de todos os processos de negócio impactados por este risco. As características básicas de um sistema de detecção da fraude são as seguintes:

- Dispor de um **repositório de dados com toda a realidade de fraude da organização.** Deve contar tanto com as variáveis de entrada nos modelos de detecção como com as pontuações obtidas na execução dos modelos sobre a atividade transacional da organização e os alertas gerados. Implementa **controles de qualidade de dados** tanto no histórico como na captura de novas variáveis.

- Implementar um **modelo de detecção da fraude parametrizável e adaptável** à problemática da organização. Geralmente, estes sistemas têm modelos pré-parametrizados baseados no conhecimento do setor que devem ser particularizados e monitorados (desempenho).

- Implementar **fluxos de geração de alertas e relatórios dinâmicos** de acordo com o esquema de revisão e análise implantado na organização para a definição de alertas em modo on-line ou batch.

- **Reportar** tanto os níveis de fraude da organização como a gestão realizada e perda sofrida.

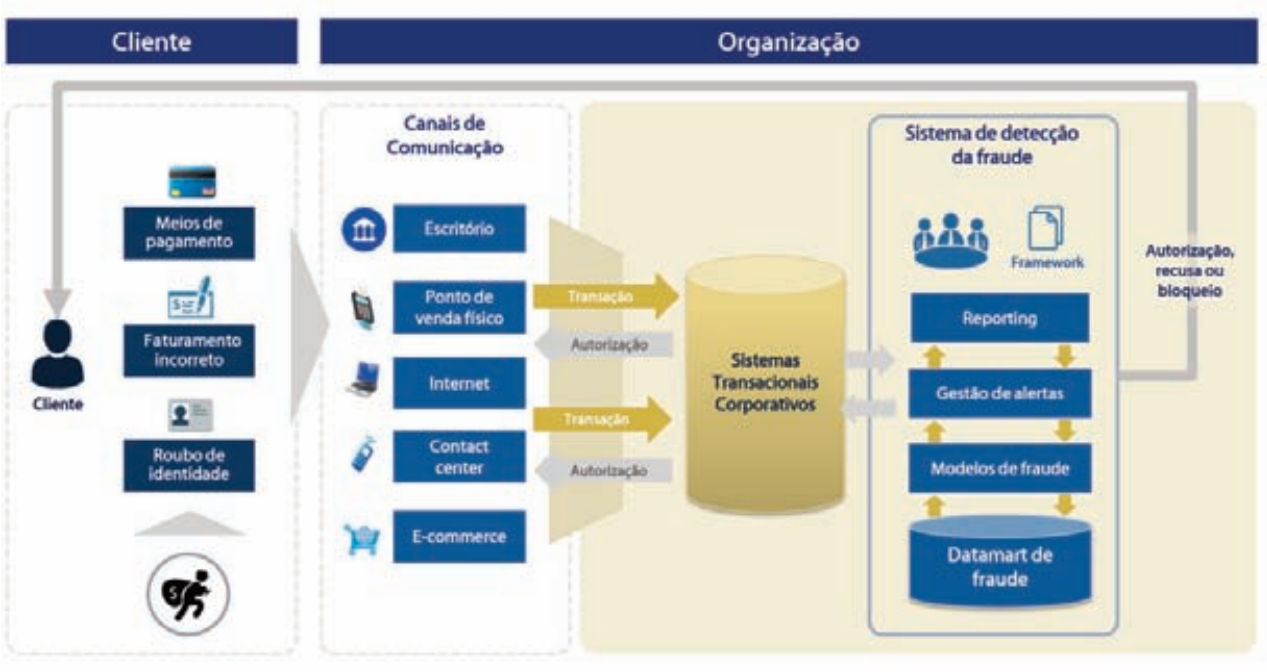
Para que um sistema de detecção da fraude seja integrado eficazmente na gestão tem que facilitar o aprovisionamento da **informação transacional e comercial** dos clientes; sua integração on-line (por exemplo, via serviços web) nos processos de autorização / recusa de atividade transacional; apresentar um **alto rendimento tecnológico** pelo envolvimento em processos com interação com o cliente; dispor de um **sistema de autenticação compatível** com os standards organizacionais (por exemplo, LDAP); e ter um **esquema de funções e usuários definível** de acordo com as políticas da organização.

Com o objetivo de ilustrar a integração de um sistema de detecção da fraude no ambiente tecnológico de uma organização, na figura 5 é apresentado um exemplo do ciclo de vida da informação analisada no processo de detecção da fraude: desde o lançamento do evento de origem pelo cliente e da entrada no processamento interno da organização até a decisão final relativa à suspeita de atividade relacionada com fraude.

De igual forma, é apesentada a hierarquia funcional dos diversos componentes internos do referido sistema: desde o armazenamento da informação granular no repositório de dados até a execução dos modelos de detecção da fraude com a informação transacional recebida e a posterior gestão dos alertas e o seu 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. Integração de um sistema de detecção da fraude



Fonte: elaboração própria.

Técnicas de gestão da fraude no setor energético

O aumento da capacidade de armazenamento de informação e da potência de cálculo fomentou o desenvolvimento da **Inteligência Analítica,** assim como das diversas disciplinas que engloba, entre as quais destacamos ***Data Science,*** cujo objetivo é extrair o máximo conhecimento dos dados, combinando análise de informação maciça com técnicas de modelagem, perfil e segmentação[31].

No setor energético, estas técnicas são utilizadas para enfrentar problemáticas que vão desde a previsão da procura energética até a identificação de padrões de consumo com o objetivo de realizar ofertas comerciais personalizadas ou detectar eventos de fraude.

Dados

A compilação e posterior tratamento de dados implicam uma análise prévia da sua tipologia, natureza e origem dos mesmos (ver fig. 6).

Em todo o caso, localizar as fontes, determinar os processos de extração, armazenamento e processamento, analisar a qualidade dos dados, etc. são atuações que requerem ser realizadas dentro de um *framework* de governança de dados, aprovado pelo primeiro nível da empresa.

Essa governança de dados implica desenvolver três domínios: i) Arquitetura tecnológica, ii) Pessoas e as suas capacidades, iii) Processos / Instrumentos para realizar uma governança efetiva (ver fig. 7).

A chave na governança dos dados está em fazer dele um instrumento útil para a gestão. Neste sentido, é preciso primeiro identificar e delimitar os dados a gerir e depois classificar em função da sua tipologia (internos, externos, estruturados, não estruturados, etc.) e do nível de proteção requerido.

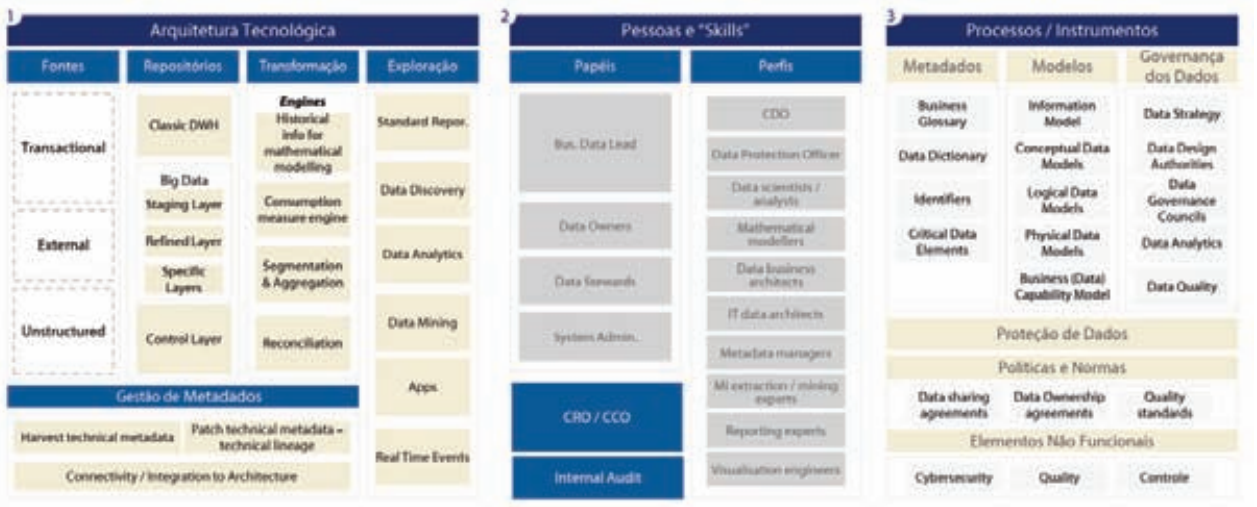
Com tudo isso, deve ser estabelecida uma **priorização** dos dados atendendo principalmente ao custo /benefício de um maior ou menor governança sobre a informação e, finalmente, desenvolver um *framework* focado na **formalização** (definição

[31] *Data Science e a transformação do setor financeiro". Management Solutions. Junho 2016*

Fig.6. Principais blocos de informação em empresas

Fonte: elaboração própria.

Fig 7. Domínios de desenvolvimento para a governança dos dados



Fonte: elaboração própria

dos conceitos, owners, fontes de informação, etc.), **acompanhamento** de qualidade e planos (controles e indicadores para acompanhamento de qualidade, planos de remediação, etc.) e **validação** independente (certificação ou auditoria interna) considerando os aspectos relacionados com a proteção da informação tanto interna, como de clientes (GDPR[32]).

Modelos

Os modelos a desenvolver pretendem encontrar padrões, tendências ou regras que expliquem o comportamento do cliente, empregado ou um terceiro antes da detecção da fraude.

No momento de selecionar as variáveis a incluir no modelo, deve ser garantido que estas cumpram as seguintes condições:

- Cubram todos os perfis que funcionalmente pretendemos refletir.
- Proporcionem o maior poder previsional conjunto e não sejam redundantes

Para isso, é realizada uma análise estatística das variáveis (univariante e multivariante), incluindo análise para detectar a multicolinearidade, por meio do estudo da correlação entre as variáveis explicativas. Com isto conseguimos maior simplicidade no modelo.

Entre as ferramentas estatísticas e numéricas mais utilizadas para analisar e corrigir a correlação estão o uso de indicadores estatísticos (como o coeficiente de correlação de Pearson, Kendall ou Spearman, segundo o tipo de variáveis), ou a utilização de técnicas de redução da dimensionalidade, como a análise de componentes principais (PCA na sigla em inglês), em que é procurado um subespaço dimensional inferior sobre o qual projetar os dados, minimizando os erros de projeção.

Uma vez balizado o conjunto final de variáveis explicativas, procedemos à construção do algoritmo. As técnicas a utilizar para a identificação dos perfis/observações suscetíveis de representar um evento de fraude variam em função do fim pretendido e do tipo de dados utilizados. Destacamos as seguintes metodologias:

- **Modelos de classificação.** O objetivo deste tipo de metodologias é prever a classe a que pertence cada uma das observações ou registros que pretendemos analisar em função de seus atributos. Algumas das técnicas mais utilizadas são as árvores de decisão, as quais podem ser geradas por meio de diferentes algoritmos como CLS, ID3, CART, etc. A técnica de ***Random Forest*** constitui outro tipo de modelo de classificação que é baseado na agregação de várias árvores de decisão (por exemplo, usando a moda ou a média) para prever a classe a que pertence cada registro.

  Outras técnicas de classificação habitualmente utilizadas são os **Classificadores Bayesianos,** as **Redes Neuronais** ou a técnica **k-Nearest Neighbours.**

- **Modelos de regressão.** O principal objetivo é a estimativa numérica da relação entre uma variável dependente e um conjunto de variáveis explicativas. Existem dois tipos de regressões, **lineares e não-lineares.** Exemplos deste tipo de modelos são as **regressões lineares Bayesianas** e os **modelos lineares generalizados** (GLM). As r**egressões logísticas** constituem um modelo de regressão cuja finalidade é a classificação das observações. A análise por meio de regressões também é utilizada para realizar previsões sobre a evolução temporal de uma variável, ou

[32] General Data Protection Regulation. Novo *framework* regulatório para a UE (tanto para as relações entre estados-membros como com terceiros) em termos de proteção da informação com o objetivo de harmonizar e unificar critérios na aplicação e garantia dos direitos em matéria de privacidade e proteção de dados, adaptando os padrões ao ambiente digital.

Fig.8. Algumas metodologias aplicadas à detecção da fraude

Fonte: elaboração própria

seja, para a análise de séries temporárias. Em particular, os modelos **ARIMA (Autoregressive Integrated Moving Average)** são utilizados para a previsão de valores futuros das séries temporárias, com base nos valores passados.

- **Modelos de segmentação ou clustering.** O objetivo desta metodologia é agrupar diversas observações em grupos homogêneos ou clusters, em função do grau de semelhança que apresentam. São modelos não fiscalizados, dado que os dados são ajustados a partir de uma amostra na qual os grupos objetivos são desconhecidos a priori, ou seja, não existe conhecimento prévio sobre as classes às quais as observações podem ser atribuídas. Dependendo do critério utilizado para determinar o grau de semelhança entre observações, são distinguidos diferentes tipos de análise de cluster: **modelos de centróides (por exemplo, k-means ou k-medians), modelos de distribuições estatísticas (por exemplo, expectation-maximization algorithm, Gaussian Mixture Models), modelos hierárquicos** nos quais os clusters são fundidos ou subdivididos sucessivamente, de acordo com uma prioridade ou hierarquia, etc. As técnicas de clustering constituem um primeiro passo no momento de abordar os problemas de classificação, quando não há informação suficiente sobre as classes que pretendemos diferenciar.

Outras técnicas usadas na detecção da fraude, são as **regras associativas (associative rules ou AR), a identificação de sequências (Motif Mininig, Autocorrelation Function) ou a detecção de anomalias ou outliers.** Cada um dos tipos de modelo demonstra um diferente grau de capacidade preditiva, estabilidade e interpretabilidade.

O aproveitamento completo de todo o potencial representado pelas técnicas de Data Science envolve a implementação das seguintes caraterísticas em alguns dos processos que compõem o ciclo de vida dos modelos:

- **Análise em tempo real.** Os sistemas de coleta de informação permitem realizar a captura e acompanhamento dos dados em tempo real. Além disso, é possível programar algoritmos que utilizem a referida informação, facilitando e agilizando a detecção dos novos padrões e estratégias de fraude não utilizadas até a data pelos infratores. Esta caraterística faz com que os modelos não percam o poder de previsão com a passagem do tempo, continuando sempre atualizados e sensíveis aos novos padrões de fraude.

- **Treinamento automático e autoaprendizagem.** Os modelos de detecção de fraude são recalibrados de forma automática (com escassa intervenção dos analistas) e interativamente a partir dos grandes volumes de dados, permitindo uma potencial melhoria do poder preditivo durante os diversos treinamentos.

A análise em tempo real, o treinamento automático e a autoaprendizagem dos modelos reduz o *time-to-market* dos mesmos. Além disso, estas caraterísticas possibilitam a procura e detecção de padrões e relações sem restrições predefinidas e de forma atualizada, bem como a identificação e incorporação nos modelos de novas variáveis relevantes. Também é possível programar os modelos para que sejam recalibrados automaticamente por meio da variação do peso relativo da contribuição de cada variável para a detecção dos eventos de fraude.

Em contrapartida, esta sofisticação dos processos envolve uma maior complexidade na gestão do risco do modelo, sendo necessário implementar controles internos e sistemas de alertas que permitam detectar qualquer desvio do modelo, bem como controles sobre os graus de liberdade na automatização dos processos. Tudo isto deve ser acompanhado por uma estrutura de gestão dos modelos.

Principais âmbitos de aplicação na gestão da fraude no setor energético

A seguir abordamos os âmbitos específicos de aplicação no setor energético: os relacionados com o **roubo de energia na rede de distribuição** e os associados às **atividades fraudulentas no ciclo comercial.** Estes âmbitos são relevantes tanto pelo impacto econômico (é estimado que, a nível mundial, as utilities cheguem a sofrer mais de 95 bilhões de dólares por perdas não técnicas de energia por ano[33]) como pelo impacto na reputação (por exemplo, aumento da diferença entre a procura elétrica medida nos pontos de consumo e a energia medida nas centrais de geração, provocando excessos de custos para o sistema que são repercutidos nos consumidores[34]). O primeiro destes gira em torno da quantificação da propensão ou probabilidade de utilização de energia por um cliente ou usuário, sem indicação pela companhia (fraude externa). O segundo é relativo à identificação de incompatibilidades em processos, como o ciclo comercial, que podem gerar lucros para funcionários (fraude interna).

Fraude externa: roubo de energia

O aumento na **capacidade de gerar e armazenar informação** pode ser aproveitado para acessar em tempo real à caraterização de um cliente, uma operação, um processo, etc., que seja utilizado para **identificar comportamentos indicativos de propensão para a existência de furto de energia.**

Desde há vários anos que a detecção de **perdas não técnicas** constitui uma das principais preocupações das companhias. Não obstante, as soluções implicavam elevados custos (inspeções pelos técnicos).

Atualmente, diversos países estão desenvolvendo e implementando as chamadas Infraestruturas de Mensuração Avançada, ou AMI[35], nas Smart Grids[36], que incluem sistemas de coleta de dados e monitoramento em tempo real, bem como a utilização de técnicas de análise baseadas na inteligência artificial, teoria de jogos, etc.

De acordo com os resultados obtidos no estudo desenvolvido pelo Departamento de Energia dos Estados Unidos[37], a utilização destas técnicas de Data Science para a detecção da fraude no setor energético é de grande utilidade para diferenciar a porcentagem de energia perdida na rede de distribuição associada a perdas técnicas (não representativas de evento de fraude) e a perdas não técnicas (roubo de energia e, portanto, representativas de um evento de fraude).

[33] Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors. Maio de 2017, Northeast Group, LLC.
[34] Relatório sobre as alternativas de regulação em matéria de redução de perdas e tratamento da fraude no fornecimento eléctrico. Relatório de 16 de julho de 2015 da CNMC, Comissão Nacional dos Mercados e da Concorrência.
[35] Advanced Metering Infrastructure.
[36] Redes elétricas inteligentes.
[37] US Department of Energy – Office of Electricity Delivery and Energy Reliability: AMI and Customer Systems: Results from the SGIG Program. Setembro 2016.

Desta forma, as companhias energéticas estão realizando campanhas de implementação de medidores inteligentes SM[38] que ajudam a reduzir as perdas não técnicas na respectiva rede de distribuição[39]. De acordo com o Departamento de Energia dos Estados Unidos[40], a coleta dos dados de consumo por meio destes dispositivos e a posterior análise por meio da utilização de técnicas de processamento de grandes volumes de dados oferecem novas possibilidades para o desenvolvimento de métodos eficientes e efetivos de detecção da fraude, melhorando a recuperação das receitas. Alguns dos benefícios da utilização de SM incluem a possibilidade de leitura dos dados de consumo remotamente, a maior resolução das medições e a detecção de cortes ou desconexões não previstos no processo de coleta de dados no medidor.

Não obstante, sua utilização também apresenta diversos desafios:

- A coleta de dados durante longos períodos de tempo devido a limites na capacidade de **armazenamento**.
- A existência de processos de compressão que reduzem a **qualidade dos dados** e limitam as possibilidades de utilização posterior.
- O custo computacional do **processamento em tempo real.**
- A proteção da informação privada sujeita a restrições de **confidencialidade**[41].

O investimento em modelos avançados de detecção **otimiza as taxas de êxito nas campanhas de inspeção** e melhora a detecção da fraude. O avanço na implementação deste tipo de modelos requer a definição prévia e a implementação de uma estrutura de referência que facilite a governança dos dados, modelo e processos associados (ver a fig. 9)

O processo objeto de análise consiste na **execução de inspeções.** Integrar as técnicas analíticas de segmentação na gestão do referido processo, como se viu nos parágrafos anteriores, requer um esquema de tratamento de grandes volumes de informação, bem como de estabelecimento de perfis dos clientes.

Uma vez que os modelos de segmentação geraram as probabilidades de ocorrência do evento (neste caso, a existência de roubo de energia), estas probabilidades conjugadas com o benefício esperado da atuação (por exemplo, a quantidade de energia roubada) ajudam a prever o resultado esperado de uma inspeção.

Portanto, para a implementação de um esquema de inspeções deste tipo é habitual dispor de um modelo de apoio ou gestor de campanhas que permita aplicar os filtros e prioridades gerados pelo modelo, bem como a coleta dos resultados das campanhas de inspeção que permitam realimentar os próprios modelos.

Esta última pergunta é de grande importância; o resultado de uma inspeção constituirá um valioso input para a caracterização do fornecimento no futuro, ajudando a caraterizar tanto reincidentes como falsos positivos.

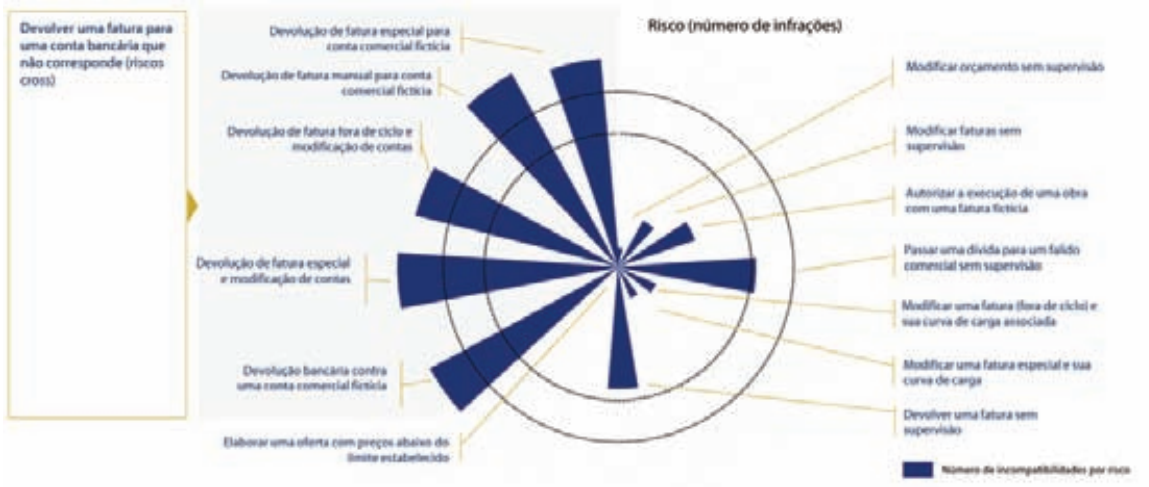
.

[38] Smart Meters.
[39] Referência:" AMI (advanced metering infrastructure) provides powerful tools to reduce total losses and increase collection rates", publicado en World Bank: Reducing Technical and Non-Technical Losses in the Power Sector. Julho 2009.
[40] US Department of Energy – Office of Electricity Delivery and Energy Reliability: AMI and Customer Systems: Results from the SGIG Program. Setembro 2016.
[41] São dados considerados de carácter pessoal, e ficam sujeitos à Lei Orgânica de Proteção de Dados (LOPD) e ao seu regulamento de desenvolvimento (Decreto Real espanhol 1720/2007). Nesse sentido, para a exploração seria necessário o consentimento expresso do interessado (se for o caso, do cliente) em conformidade com o respectivo artigo 6. Nas condições gerais dos contratos das principais distribuidoras é permitido o uso dos dados pela distribuidora, e por terceiros de empresas que tenham uma relação contratual e apenas para efeitos de prospecção comercial. Estes requisitos são reforçados pelo novo Regulamento Geral de Proteção de Dados (RGPD), no qual se reforçam os requisitos para a gestão de consentimentos pelo cliente (concessão explícita do uso para uma finalidade e períodos específicos).

Fig.9. O emprego das técnicas de Data Science afeta os dados, metodologia e processos

Fonte: elaboração própria

Fig.10. Taxonomia ilustrativa de riscos em sistemas comerciais



Fonte: elaboração própria.

Fraude interna: segregação de funções no ciclo comercial

As companhias energéticas dispõe de diversos sistemas para suportar os respectivos processos operacionais. Tanto empregados quanto profissionais externos têm acesso aos referidos sistemas.

As alterações de posição dos funcionários ou a existência de usuários genéricos em alguns sistemas (principalmente decorrentes da externalização de funções como as áreas de faturamento e cobranças ou call centers) acrescentam **complexidade ao acompanhamento das funções** desempenhadas pelos diversos intervenientes.

A possibilidade de um trabalhador poder realizar ações que permitam obter um benefício próprio (por exemplo, alterando a conta corrente de um cliente pela do funcionário e realizando uma devolução da fatura ou alterando o valor de uma fatura do funcionário no consumo associado), implica um risco de fraude interno.

A implementação de soluções de controle de acessos permite identificar debilidades, especialmente nos sistemas comerciais, nos quais diversas funções são externalizadas, fazendo com que surja uma variedade de riscos relacionados com a segregação de funções. Consulte uma taxonomia ilustrativa de riscos em sistemas comerciais (fig. 10).

Os projetos de controle de acessos têm por objetivo i) rever os cargos para identificar funções incompatíveis, ii) reatribuir tarefas para eliminar as referidas incompatibilidades ou iii) caso não seja possível evitar a incompatibilidade, estabelecer controles de mitigação.

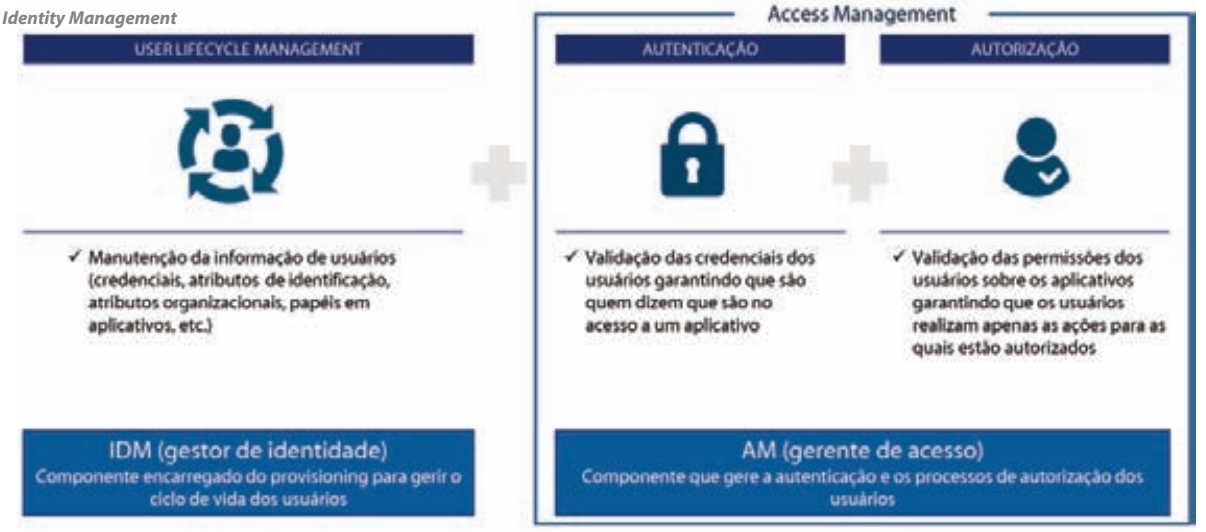
Diversas empresas energéticas implementaram soluções de controle de acessos (também conhecidos como Access Control ou Identity Management) com um objetivo triplo:

- **Reduzir o risco de acessos não autorizados** aos sistemas por meio da definição de um modelo de user provisioning que proceda a um controle ex ante que permita antecipar situações de incompatibilidade (antes de atribuir um cargo a um usuário é verificado se a referida concessão não implica um incumprimento da SoD).

- **Garantir a confidencialidade, a integridade e a disponibilidade da informação por meio da segregação de funções**, assegurando que os acessos às informações críticas da empresa estão controlados e só são acessados por pessoas adequadas (por exemplo, sistemas de folhas de pagamento ou sistemas contábeis).

- **Automatizar o user provisioning,** garantindo que os funcionários dispõem das permissões requeridas no menor tempo possível (por exemplo, automatização do

Fig.11. Identity Management



Fonte: elaboração própria

acesso a sistemas em função do posto de trabalho e eliminação de acessos quando ocorrem baixas ou alterações de área na companhia).

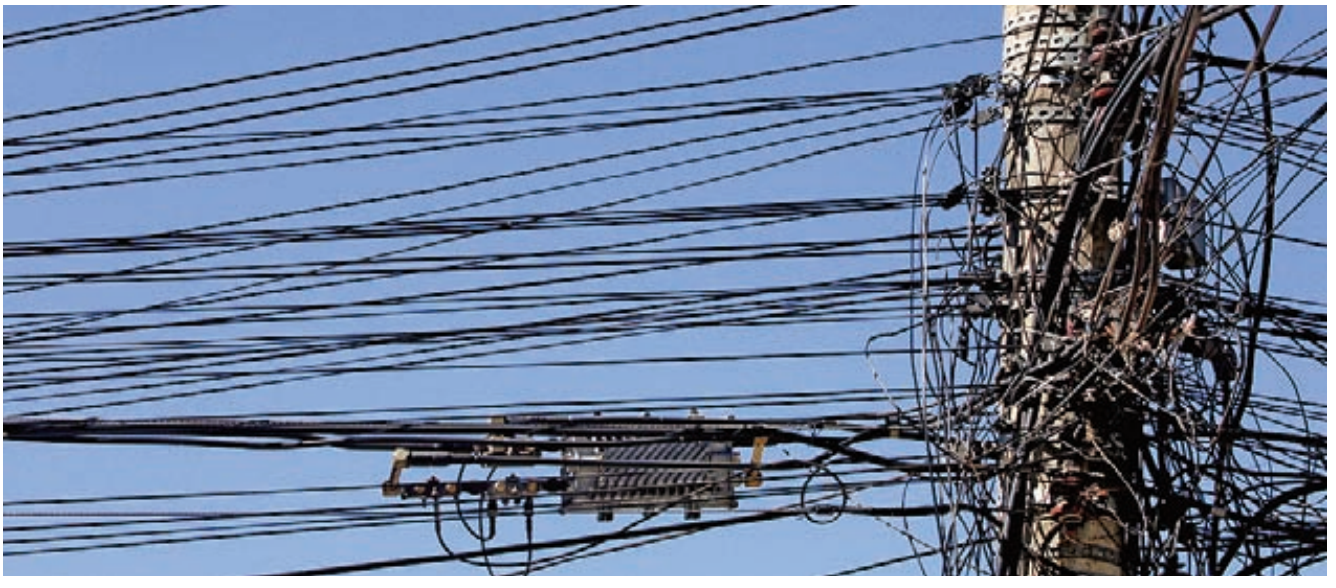
O controle de identidade e acesso (denominado Identity Management) engloba três processos principais (User Lifecycle Management, Autenticação e Autorização) que garantem que os usuários i) são quem afirmam ser e ii) acessam as aplicações adequadas com os direitos necessários.

Estes processos são sustentados em soluções tecnológicas (conhecidas como IGA: Identity Governance and Administration) cuja implementação permite às empresas aumentar o nível de controle sobre os acessos aos sistemas garantindo a segregação de funções, diminuindo, portanto, o risco interno da empresa. Estas soluções exijam um monitoramento contínuo para garantir o controle correto da segregação de funções, bem como a disponibilidade de ferramentas de análise que permitam definir:

- **Um plano de medidas ou controles de mitigação para as incompatibilidades,** como a remediação de cargos/usuários ou o estabelecimento de exceções que sejam monitoradas por meio de controles mitigantes (por exemplo, definir informações ou processos para garantir o controle sobre o risco).

- **Um protocolo de atuação** para que os riscos e incompatibilidades manifestados em revisões futuras sejam resolvidos ou minimizados com rapidez.

Em qualquer caso, o controle efetivo dos acessos poder ser comprometido caso não haja uma visão global dos acessos do funcionário e autorizações independentes por aplicação, bem como de processos para a supressão dos referidos acessos (por exemplo, quando um funcionário muda de cargo na empresa, nem sempre são eliminadas as permissões exigidas para o cargo ocupado anteriormente). É igualmente importante realizar análise de risco por posto, antes de cancelar um acesso.



Exemplo de aplicação de técnicas de modelagem: roubo de energia

A detecção da fraude energética, associada ao consumo ilegal de energia na rede, parte de uma segmentação dos clientes baseada em sua probabilidade de cometer fraude. Para isso, são interpretados os dados históricos de fraude e seu impacto no negócio, assim como o comportamento histórico dos infratores (entre outros, o resultado de inspeções prévias, faturamento e cobrança).

As técnicas de modelagem aplicadas na gestão da detecção da fraude energética visam melhorar a taxa de sucesso na seleção de clientes a inspecionar.

Ciclo de vida dos modelos de detecção

Os modelos utilizados na detecção da fraude passam por quatro etapas: extração da informação, análise estatística dos dados, construção, e validação e certificação (ver fig. 12).

Extração e tratamento dos dados

Na primeira etapa de extração e tratamento da informação dos clientes de energia são distinguidas várias fases: o pedido e a extração dos dados, a análise da qualidade da informação e a construção de novas variáveis (ver fig. 13).

A informação coletada ficará registrada em uma tabela única. Além disso, serão criadas novas variáveis a partir dos dados coletados. A seleção de variáveis relevantes com bom poder de previsão para a detecção da fraude energética requer uma profunda análise estatística. Algumas variáveis que demonstraram um alto poder de previsão são:

- **Dados do medidor:** o tipo de conexão (monofásica, bifásica ou trifásica), a marca e modelo do medidor podem ser relevantes pela complexidade de manipulação para a fraude e representar uma oportunidade percebida pelo cliente. Outras variáveis relevantes podem ser o tipo de rede e a potência instalada do medidor que define o possível consumo máximo do cliente, porque a sua relação com o consumo real deveria ter valores próximos em um cliente sem fraude.

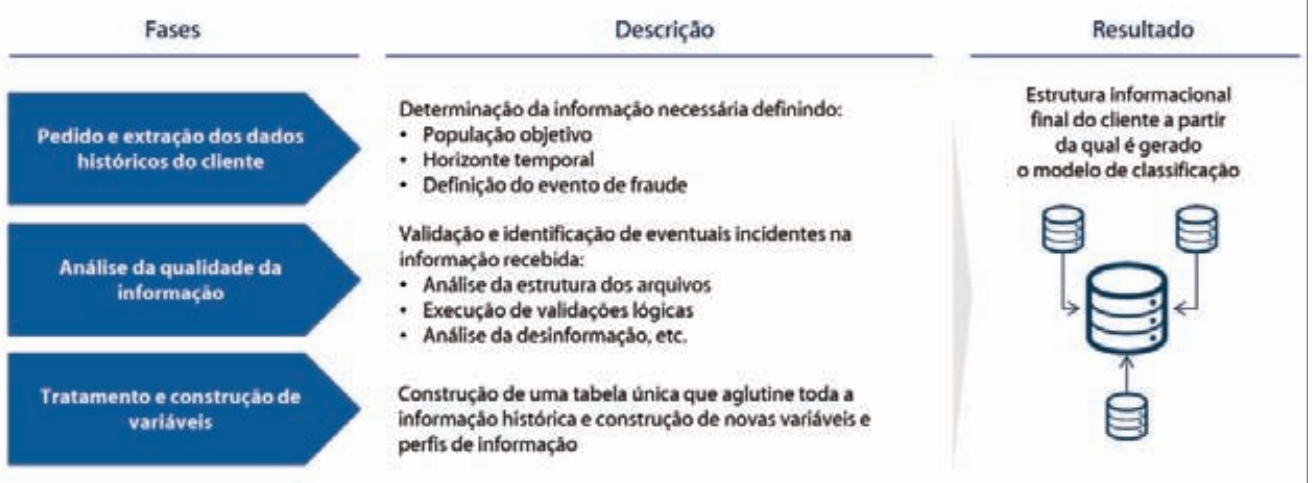
- **Dados sociodemográficos:** estes dados podem ser úteis para segmentar a população. Por exemplo, os atributos de localização (comunidade, província, cidade, município, código postal, bairro ou região), onde existe um alto poder explicativo do evento de fraude. Para complementar a informação do cliente, é possível obter informação externa enriquecida por zonas, como renda média, consumo médio, tipos de casa, clima, etc.

- **Dados do cliente:** em geral, as empresas de energia dispõem de informação útil do cliente (antiguidade, consumo coletivo ou individual, classe do cliente: residencial, empresa, industrial, serviço público, rural, luz

Fig.12. Etapas na criação de um modelo

Fonte: elaboração própria

Fig.13. Tratamento dos dados na construção de um modelo



Fonte: elaboração própria

pública, cliente de baixa renda, etc.). No caso de grandes clientes há maior volume de informação disponível com atributos como a atividade, setor, etc.

- **Dados de consumo** histórico da energia: esta informação desempenha um papel fundamental para analisar o comportamento do cliente e graças à telemedida, a qualidade destas variáveis são muito importantes. Devemos destacar a análise do desvio no consumo esperado por motivos cíclicos, de clima, cortes de energia ou outros, média do consumo face a clientes semelhantes ou ao mesmo cliente no passado, irregularidades nas leituras do medidor, periodicidade nas leituras do cliente ou dívida do cliente.

  A frequência e variedade desta informação são enriquecidas a cada ano (historicamente, poderia estar disponível uma escassa informação, como consumo trimestral de um ponto de fornecimento comunitário, embora hoje em dia seja possível obter, com os telemedidores inteligentes, informação pormenorizada de correntes, fases, tensão, potência, etc. em tempo real).

- **Dados da operação ou gerenciamento realizado:** Dados da operação de manutenção da rede e dos medidores, informação de cortes e irregularidades ou informação de contatos ou reclamações com o cliente, etc. são informações que podem ser úteis.

- **Dados da inspeção:** o pormenor do resultado da inspeção não pode ser usado como variável para o modelo, mas ajuda na realização de uma análise prévia das causas que originam uma fraude. No subconjunto de clientes reincidentes, é possível usar alguma informação que rapidamente identifique a nova fraude.

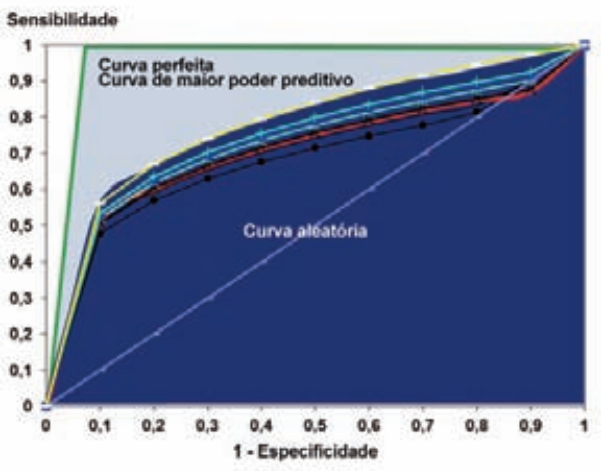
Análise estatística da informação e construção do modelo

Em primeiro lugar, é realizada uma análise estatística da informação para detectar as variáveis mais relevantes, segundo o exposto no ponto 3.1.2. Ao chegar a esta etapa dispomos de uma bateria de variáveis depuradas de erros e potencialmente previsíveis da fraude do cliente, pelo que é possível proceder à construção (calibragem) do modelo[42]. A seguir, é **selecionado o algoritmo** mais apropriado para o modelo que pretendemos construir sobre a população de estudo e são determinados os seus parâmetros. Opcionalmente, a atribuição de pesos às variáveis (selecionadas por técnicas estatísticas) permite perfilar e **discriminar clientes** de comportamentos assimiláveis à fraude.

Do ponto de vista estatístico, podemos distinguir dois enfoques diferentes no problema de classificação. No primeiro, os grupos estão bem definidos e trata-se de determinar um critério para etiquetar cada indivíduo como pertencente a algum dos grupos, a partir dos valores de uma série limitada de parâmetros. Neste caso, as técnicas mais utilizadas são conhecidas com o nome de **análise discriminante,** embora existam outras possíveis alternativas, tais como a utilização da regressão logística, redes neuronais ou árvores de decisão. O segundo enfoque corresponde ao caso em que, a priori, não são conhecidos os grupos e o que precisamente é pretendido é estabelecer os grupos a partir dos dados que detemos. As técnicas estatísticas mais utilizadas nessa área são conhecidas pelo termo **análise cluster,** que podemos traduzir como análise de agrupamentos e também como análise de conglomerados.

[42] Sua calibragem por meio métodos de machine learning é apoiada em técnicas de Bootstrap Aggregating ou Bagging (os modelos são treinados em paralelo e é utilizada a combinação das previsões como previsão final) ou Boosting (os modelos são treinados sequencialmente para que o modelo seguinte esteja concentrado em prever corretamente as falhas dos anteriores). Ver Breiman, Leo (1996). Bagging predictors 24 (2), ou também Dietterich, T. G. (2000, June). Ensemble methods in machine learning. In International workshop on multiple classifier systems. Springer Berlin Heidelberg.

Curva ROC



Para isso, podem ser comparados distintos modelos, entre os quais destacamos (ordenados de menor para maior nível de sofisticação):

- **Árvore de decisão:** calibrado de uma árvore automática de decisão selecionando as variáveis com maior poder de previsão segundo o teste de "Qui-quadrado" ou "Cramer" (a aplicação de cada critério dá lugar a uma árvore diferente).
- **Regressão logística:** seleção de um subconjunto de variáveis com maior poder de previsão (segundo o teste de "Qui-quadrado") e calibrado de uma regressão logística.
- **Rede Neuronal:** treinamento de uma rede neuronal com as variáveis originais da base de treinamento.
- **Transformação + Rede Neuronal:** combinação de transformações simples (logaritmos, inversas, raízes, potências, traslações, etc.) de variáveis contínuas com as variáveis originais para o treinamento de uma rede neuronal.
- **Regressão + Rede Neuronal:** combinação de regressão logística para a agregação de variáveis com uma rede neuronal.
- **Random Forest:** combinação sequencial de árvores de previsão que dão lugar ao chamado "floresta", que proporcionará uma previsão do evento encadeando as previsões de todas as árvores do processo.
- **Gradient Boosting:** combinação de árvores de previsão para obter um classificador mais sólido mediante a aplicação de algoritmos de machine learning.

Validação

Com o objetivo de identificar os modelos que melhor explicam o comportamento dos clientes fraudulentos foram fixados alguns critérios mínimos que devem cumprir os resultados obtidos pelo modelo selecionado.

Estes critérios são baseados tanto na c**apacidade discriminante do modelo,** por exemplo, o seu índice AUC[43], como na sua **razoabilidade**. Este último é quantificado por meio da análise de tendências (que corrobora que a tendência do estimador de cada variável, em relação à fraude, corresponde ao esperado em termos econômicos) e os pesos relativos das variáveis segundo a sua contribuição esperada.

Foi realizado um exercício quantitativo, executando os modelos descritos no ponto anterior de forma independente sobre a mesma base de treinamento. A seguir, é apresentada uma comparação da sua capacidade discriminante em função da área sob a curva ROC (AUC):

Modelo	ROC sobre base de validação
Árvore de decisão	0.78
Regressão logística	0.74
Rede Neuronal	0.82
Transformação + Rede Neuronal	0.83
Regressão + Rede Neuronal	0.79
Random Forest	0.81
Gradient Boosting	0.86

[43] AUC: Area Under the Curve. A curva utilizada é a ROC (Receiving Operating Characteristics), que é obtido por meio da porcentagem de indivíduos corretamente ordenados pelo modelo em função de sua propensão para o furto. Ou seja, se um modelo identifica um potencial autor do furto com uma propensão maior que outro, a probabilidade de que este esteja correto corresponde com o AUC.

Como se pode observar na análise realizada, o modelo de Gradient Boosting apresenta um maior ajustamento em termos de ROC.

Adicionalmente à validação estatística, foram validados os modelos com as inspeções realizadas durante seis meses. Foram comparados os 12, 25 e 50 clientes com maior propensão de cada um dos modelos com as inspeções realizadas durante esse tempo para calcular a eficácia dos clientes infratores que seria obtida. A seguinte tabela mostra as percentagens de infratores que o modelo encontraria em cada um destes grupos (de 12, 25 ou 50 clientes).

Modelo	12 clientes	25 clientes	50 clientes
Árvore de decisão	58%	40%	30%
Regressão logística	50%	52%	44%
Rede Neuronal	75%	64%	38%
Transformação + Rede Neuronal	75%	68%	46%
Regressão + Rede Neuronal	67%	60%	42%
Random Forest	75%	64%	52%
Gradient Boosting	92%	72%	50%

Observamos que, seja qual for o número de inspeções, o modelo de Gradient Boosting é o que reúne uma maior concentração de infratores (exceto o Random Forest ao propor 50 inspeções; embora a melhoria não seja significativa, 2%). Portanto, este modelo é o que será utilizado no exemplo prático.

Ganho de energia

O ganho de energia é o conceito usado para representar o valor econômico que é recuperado em cada cliente após a identificação da fraude energética. Este conceito corresponde ao valor econômico de energia que começa sendo faturado pela normalização no consumo após realizar a inspeção mais a recuperação da energia histórica não faturada. Seria o conceito análogo ao Customer Lifetime Value ou valor do cliente usado nas técnicas de segmentação comercial. Seu cálculo é uma combinação de critérios de negócio definidos por cada empresa, embora apresentemos os principais fatores que são considerados para o seu cálculo:

- Legislação do país: as normas de cada país estabelecem um critério de penalização pela fraude energética e como devemos proceder. A título ilustrativo, pode consistir na interrupção do serviço de forma imediata e em uma sanção pelo delito baseada em uma estimativa do consumo correspondente ao produto da potência contratada, ou que deveria ter sido contratada, do fornecimento onde ocorreu a defraudação, por uma quantidade de horas de utilização diárias durante um ano.

- Metodologia da estimativa do consumo: segundo a informação que proporciona o medidor, o tipo de contrato, os ciclos de faturamento e o histórico de consumo usado podem ser aplicados critérios diferentes para cada cliente.

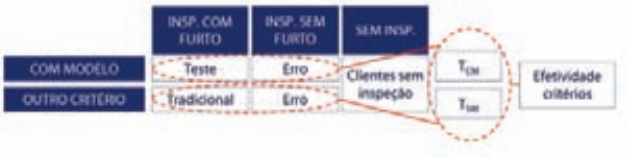
Mensuração da eficácia

As áreas de gestão da perda não técnica das empresas energéticas investem recursos humanos, técnicos e econômicos na execução de inspeções a clientes.

A rentabilidade destes investimentos é determinada por i) as taxas de sucesso observadas (do subconjunto dos clientes inspecionados, que porcentagem de clientes com furto de energia foi identificada), ii) o ganho de energia (representado como valor econômico de recuperação por cliente), iii) o número de clientes inspecionados da população-alvo, e iv) o custo unitário associado à inspeção do cliente e, portanto, o custo total da campanha.

Comparar a eficácia de uma campanha de inspeção determinada por um modelo com outra definida mediante outros critérios de negócio (por exemplo, denúncia dos operários da manutenção ou leitura, etc.) requer condições similares em ambas as populações. Ou seja, a propensão estrutural ao furto em ambas as populações deve ser muito similar (por exemplo, não devem ser comparadas zonas de diferente nível socioeconômico).

O comportamento destes modelos é supervisionado por meio de um acompanhamento estatístico do poder de previsão de cada uma das variáveis usadas e a verificação de sua qualidade, estabilidade populacional e capacidade discriminante.



Aplicação prática

Segundo indicado no ponto de validação, foi selecionado o algoritmo de maior poder discriminante. Com este modelo, a seguir desenvolvemos um exemplo de aplicação prática na configuração de campanhas de inspeção.

O exercício é enquadrado em uma distribuidora de eletricidade com 5 milhões de clientes, onde as **perdas não técnicas** representam 10%. A equipe de perdas tem o objetivo de reduzir as perdas não técnicas por meio de inspeções que têm um custo unitário de 30 USD. Por restrições operativas e econômicas dispõe-se de 100.000 inspeções anuais (sensivelmente abaixo do número de clientes infratores, que seria na ordem dos 500.000).

Até o momento foi utilizado um critério de priorização baseado no **juizo especialista da equipe de perdas,** que prioriza a inspeção de consumos próximos a 0, alcançando umas taxas de sucesso históricas nas inspeções de 9% (nível de acerto esperado para campanhas realizadas aleatoriamente).

A área de perdas começa o desenvolvimento de um modelo discriminante com a identificação de toda a informação histórica disponível dos clientes em um período anterior às inspeções realizadas. Segundo foi exposto no ponto 4.2.4, é realizada uma homogeneização das diferentes fontes de dados, uma seleção de informação de qualidade, a construção de variáveis derivadas mediante regras de negócio e é desenvolvida a metodologia de seleção do modelo.

O modelo selecionado é baseado em árvores de decisão combinadas com técnicas de machine learning; em concreto, redes neuronais para deep learning, com o apoio de variáveis principalmente de i) **padrão de consumo** – históricos e médias móveis recentes, inspeções prévias-, ii) **características técnicas do ponto de fornecimento** – medidores, conexões, etc.- e iii) **comportamento** – reincidência e outras associadas com a cobrança, como capacidade e vontade de pagamento, etc. Este último grupo de variáveis é especialmente relevante devido à **vinculação existente entre o furto e a inadimplência.** Em geral, trata-se de dois problemas intimamente relacionados, dado que, por vezes, a resolução de um problema de furto dá lugar a um problema de não pagamento, e vice-versa. Ou seja, a falta de capacidade ou vontade de pagamento deriva em incumprimentos que, quando são geridos, por exemplo por meio do corte de fornecimento, geram um incentivo ao furto. Igualmente, a resolução de um problema de furto pode gerar um problema de não pagamento.

Este modelo aumenta a taxa de sucesso das inspeções até 27% (mais de uma de cada quatro inspeções são de sucesso). A seguinte tabela reúne a rentabilidade das atuações antes e depois da implantação do modelo.

Parâmetro	Antes... (campanha juízo especialista)	Depois... (campanha com modelo)
Custo por inspeção	30 usd/inspeção	30 usd/inspeção
Capacidade	100.000 inspeções/ ano	100.000 inspeções/año
Ganho médio de energia	300 usd/furto	300 usd/furto
Taxa de sucesso	9%	27%
Custo campanha	3.000.000 usd	3.000.000 usd
Receita campanha	2.700.000 usd	8.100.000 usd
Resultado campanha	-300.000 usd	5.100.000 usd
Rentabilidade	-10%	170%

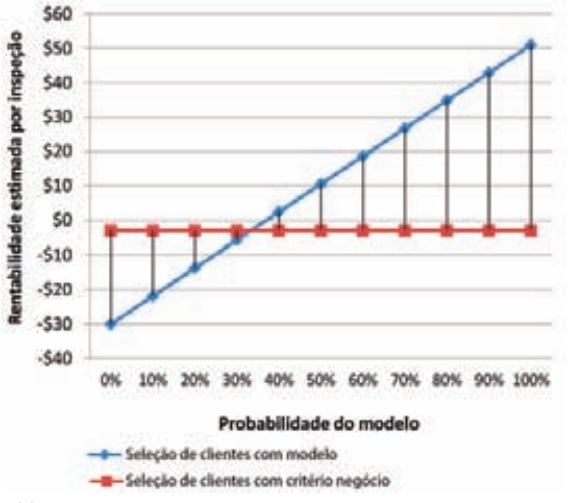
Em resumo, o fato de triplicar a taxa de sucesso nas inspeções e selecionar o ponto correto de corte para fazer as mesmas **aumenta a rentabilidade das campanhas e gera lucros de mais de 5 milhões de USD para a empresa,** simplesmente pelo fato de definir campanhas de inspeção com técnicas de priorização das visitas baseadas em um modelo analítico de propensão ao furto.

A rentabilidade estimada de cada inspeção é definida como:

$$\textit{Rentabilidade por inspeções} \approx \textit{Taxa de sucesso} * 300 - 30 \frac{USD}{inspeção}$$

Por exemplo, o envio de inspeções a todos os clientes com **uma taxa de sucesso esperada superior a 50% proporcionaria uma rentabilidade esperada por inspeção positiva e sempre superior à rentabilidade mínima exigida (10 USD).**

Fig.14. Impacto na rentabilidade das inspeções



Fonte: elaboração própria

Conclusões

Após expor o conceito de fraude, externa e interna, e suas implicações em termos de organização, processos e sistemas, centramos as técnicas de otimização da gestão da fraude no setor energético para os casos de furto de energia e de fraude no ciclo comercial.

Entre outras **iniciativas**, as empresas energéticas realizam esforços relevantes para a gestão da fraude mediante:

1. O **perfil de clientes e a segmentação** que permita orientar as atuações de inspeção ou mitigação.

2. A definição e implantação de esquemas de **quantificação da utilidade e mensuração da rentabilidade** das atuações (por exemplo, análise da rentabilidade da aquisição de variáveis externas de fornecedores).

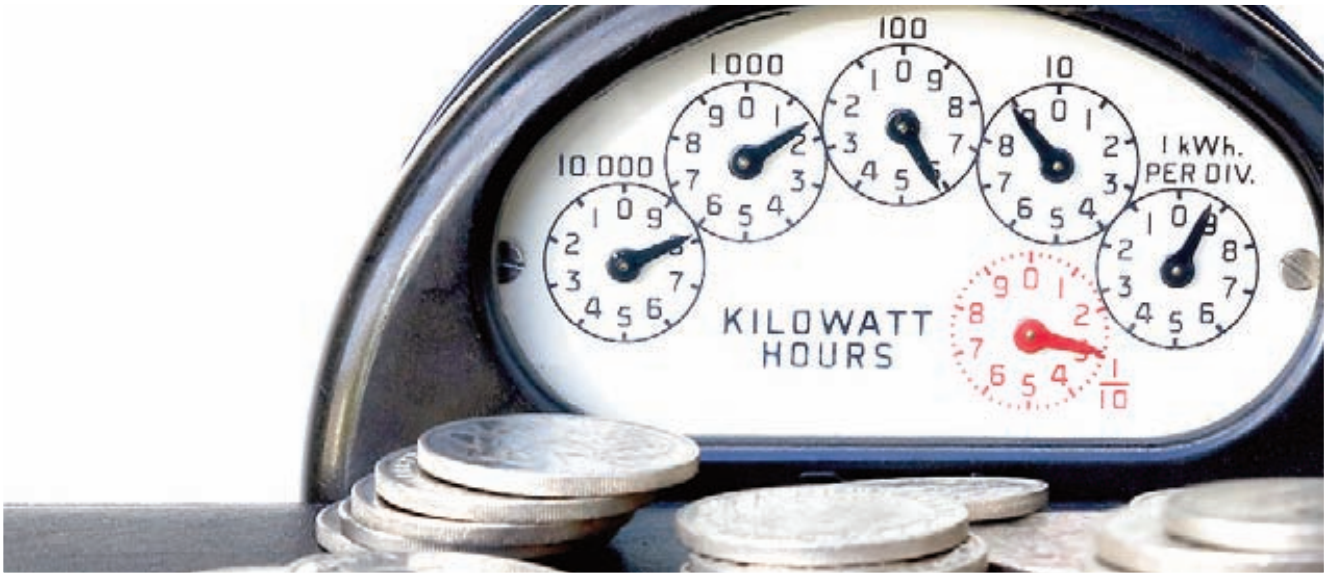
O uso de **novas técnicas de modelagem e *machine learning*** nestes processos pode ser uma ferramenta eficaz na detecção da fraude, aumentando a taxa de sucesso, industrializando o processo de detecção da fraude e reduzindo o custo de inspeções tradicionais.

Assim, tem uma grande relevância a **contribuição de valor dos dados** e da informação de clientes e operações na quantificação das possibilidades de ocorrência de eventos de fraude. A disponibilidade de informação (consumos horários, dados de clientes, acessos a sistemas, etc.) permite aplicar técnicas tanto de perfil de clientes e funcionários como de segmentação para a gestão personalizada, sendo quantificadas poupanças superiores a um milhão de dólares anuais graças à análise de dados em empresas de pequena dimensão (~2 milhões de clientes)[44]. Por isso, estão sendo implantados mecanismos de **governança dos dados e controle da sua qualidade.**

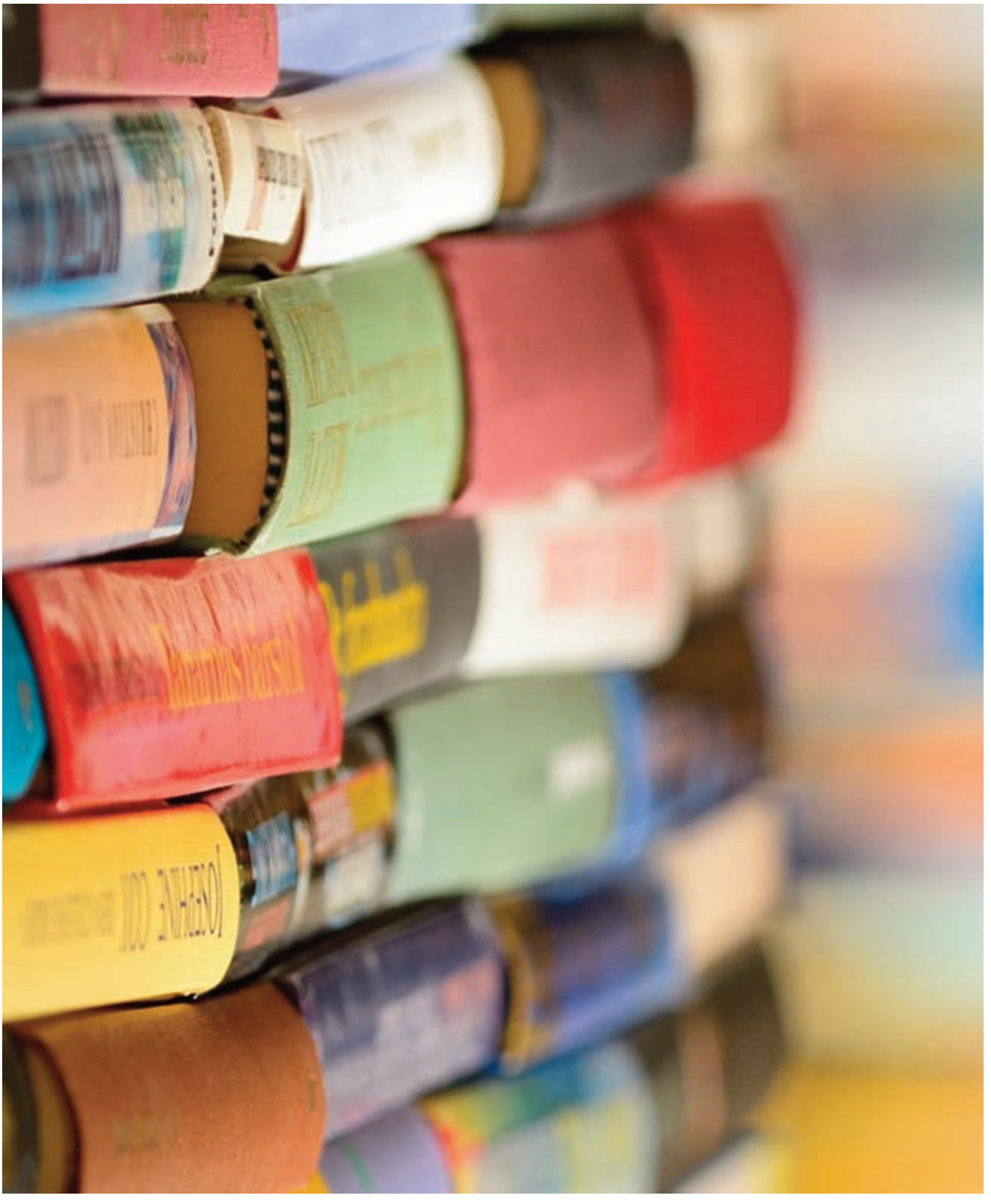
Por outro lado, as atuações de detecção e mitigação de eventos fraudulentos competem com os restantes investimentos da empresa. Por este motivo, sua integração na gestão incorpora a **mensuração da sua rentabilidade.**

Tudo isso é englobado em um ***framework* integrado de gestão da fraude**, que pode ser desenvolvido nas organizações para garantir a consecução dos objetivos da aplicação de métodos estatísticos.

[44] Advanced Metering Infrastructure and Customer Systems. Results from the Smart Grid investment grant program. Setembro 2016, Departamento de Energia dos Estados Unidos.



Referências



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

Gestão do Risco de Fraude nas Organizações: Um Guia Prático. IIA, Institute of Internal Auditors (2015).

Convergência internacional de medidas e normas de capital. Basileia: 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).

Lei 24/2013 do Setor Elétrico Espanhol.

Data science e a transformação do setor financeiro. Management Solutions (2015).

Model Risk Management. Aspetos quantitativos e qualitativos da gestão do risco de modelo. 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).

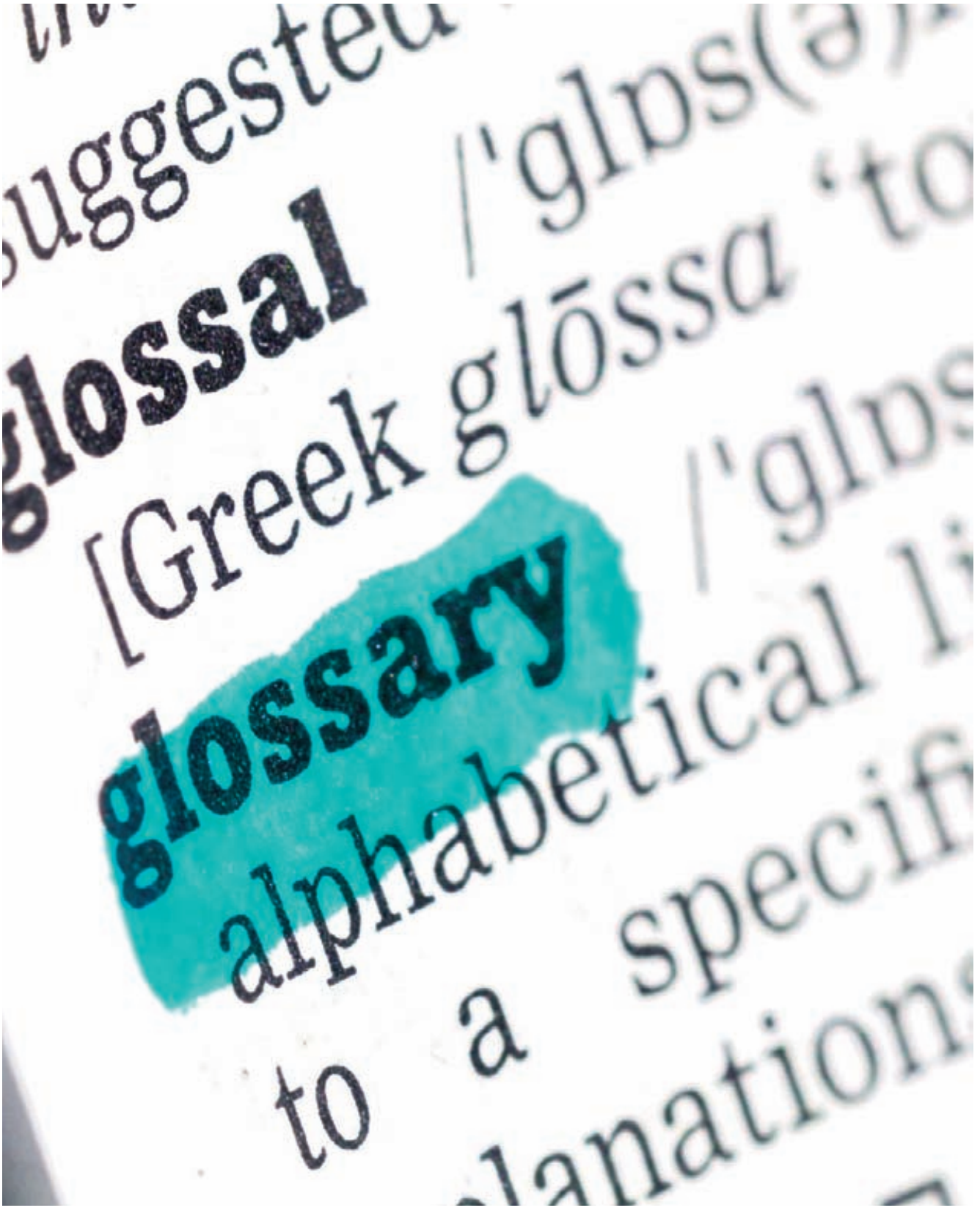
Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors. Northeast Group, LLC (2017).

Relatório sobre as alternativas de regulação em matéria de redução de perdas e tratamento da fraude no fornecimento eléctrico. CNMC, Comissão Nacional dos Mercados e da Concorrência (Relatório de 16 de julho de 2015).

Lei Orgânica de Proteção de Dados (LOPD) e o seu regulamento de desenvolvimento (Decreto Real Espanhol 1720/2007).

Smart Grid investment grant program. Departamento de Energia dos Estados Unidos. (2016).

Glossário



ACFE (Association of Certified Fraud Examiners): constituída no ano de 1988, é uma organização profissional de examinadores de fraude. Suas atividades consistem em criar ferramentas de gestão da fraude, ministrar formação e gerir uma base de dados de conhecimento.

AMI (Advanced Metering Infrastructure): sistemas que são capazes de medir, recolher e analisar o uso da energia e por sua vez interagir com outros dispositivos como os medidores inteligentes de eletricidade, gás ou água. Dispõem da capacidade de gerir a informação recolhida e tomar decisões. Estes sistemas são diferenciados dos sistemas de leitura automática, em que com os AMI existe uma comunicação bidirecional entre o medidor e o centro de controle da empresa.

Backtest: termo que é relativo ao teste de um modelo de previsão utilizando informação histórica para determinar e/ou assegurar a sua rentabilidade.

Qui-quadrado: teste estatístico para verificar a existência de uma relação entre variáveis.

CFCA (Communications Fraud Control Association): Associação global de educação sem fins lucrativos, que está focada na prevenção da fraude no setor de telecomunicações.

COSO (The Committee of Sponsoring Organizations of the Treadway Commission): comitê criado a partir de uma iniciativa conjunta entre organizações pertencentes ao setor privado, dedicado a fornecer conhecimento por meio do desenvolvimento de frameworks e guias sobre o controle interno, prevenção da fraude e gestão do risco nas empresas.

Cramer's V: medida da intensidade da relação entre duas ou mais variáveis categóricas quando, pelo menos, uma das variáveis pode tomar pelo menos dois valores possíveis.

Data Lineage: É definido como o ciclo de vida da informação que inclui a sua origem, movimento e transformações. Descreve o que ocorre com a informação à medida que é submetida a diversos processos proporcionando visibilidade para poder apagar os erros e as suas fontes.

Deep Learning: conjunto de algoritmos de aprendizagem automática que tentam aprender representações de dados. Uma observação (por exemplo, uma imagem) pode ser representada de muitas formas (por exemplo, um vetor de pixéis), mas algumas representações tornam mais fácil aprender tarefas de interesse (por exemplo, "esta imagem é uma face humana?").

Financial Fraud Action UK: organismo responsável por liderar a luta coletiva contra a fraude em nome do setor financeiro do Reino Unido; sendo sua função primordial facilitar a atividade entre os diversos atores envolvidos na luta contra a fraude.

IIA (Institute of Internal Auditors): associação internacional profissional de auditoria interna e gestão de riscos estabelecida no ano 1941.

KDD (Knowledge Discovery in Database): processo de extração de informação potencialmente útil de uma base de dados. Processo interativo que exaustivamente explora volumes muito grandes de dados para determinar relações.

KPI (Key Performance Indicator): métrica que as empresas utilizam para medir os resultados de uma determinada ação ou estratégia em função de objetivos predeterminados.

Machine Learning: método de análise de dados que automatiza o processo da criação de modelos analíticos. Utiliza um algoritmo que aprende interativamente a informação, permitindo às ferramentas encontrar padrões escondidos sem terem que estar explicitamente programadas para tal.

Prospecção de dados (Data Mining): processo computacional para descobrir padrões ocultos, tendências e correlações por meio da extração de uma grande quantidade de dados.

NIST (National Institute of Standards and Technology): agência da Administração de Tecnologia do Departamento de Comércio dos Estados Unidos cuja missão é promover a inovação e a competência industrial nos Estados Unidos mediante avanços em metrologia, normas e tecnologia.

Phishing: tentativa de obter informação sensível como usuários, palavras-chave, informação de cartões de crédito, etc., geralmente com intenções maliciosas, enganando empresas legítimas mediante uma comunicação eletrônica.

SM (Smart Meters): equipamento eletrônico que captura o consumo energético em intervalos de uma hora ou menos e, por sua vez, comunica a informação coletada ao centro de controle da empresa para o monitoramento ou faturamento da eletricidade.

SoD (Segregation of Duties): conceito de dedicar mais de uma pessoa para realizar uma tarefa. É uma medida de controle que divide uma tarefa em subprocessos, atribuindo a diferentes responsáveis, para prevenir a fraude.

Stream Computing: sistema informático que analisa múltiplos fluxos de dados de diversas fontes, processando a informação, transmitindo de volta em um só fluxo.

Nosso objetivo é superar as expectativas de nossos clientes, nos convertendo em parceiros de confiança

A Management Solutions é uma empresa internacional de serviços de consultoria com foco em assessoria de negócios, riscos, organização e processos, tanto sobre seus componentes funcionais como na implementação de tecnologias relacionadas.

Com uma equipe multidisciplinar (funcionais, matemáticos, técnicos, etc.) de cerca de 2.000 profissionais, a Management Solutions desenvolve suas atividades em 24 escritórios (11 na Europa, 12 nas Américas e um na Ásia).

Para atender às necessidades de seus clientes, a Management Solutions estruturou suas práticas por setores (Instituições Financeiras, Energia e Telecomunicações) e por linha de negócio (FCRC, RBC, NT), reunindo uma ampla gama de competências de Estratégia, Gestão Comercial e *Marketing*, Organização e Processos, Gerenciamento e Controle de Riscos, Informação Gerencial e Financeira e Tecnologias Aplicadas.

Na indústria de energia, a Management Solutions presta serviços a todos os tipos de sociedades - elétricas, óleo e gás, petroquímicas, etc. - tanto em corporações globais como companhias locais e órgãos públicos.

Jesús Martínez
Sócio
jesus.martinez.gimenez@msspain.com

Manuel Ángel Guzmán
Gerente de P&D
manuel.guzman@msspain.com

Javier Salcedo
Supervisor
javier.salcedo@msbrazil.com

Design e diagramação
Departamento de Marketing e Comunicação
Management Solutions - Espanha

www.managementsolutions.com

Madrid Barcelona Bilbao London Frankfurt Paris Warszawa Zürich Milano Roma Lisboa Beijing New York Boston Atlanta Birmingham San Juan de Puerto Rico Ciudad de México Medellín Bogotá São Paulo Lima Santiago de Chile Buenos Aires