*Por Fábio Nilzu, Guilherme Fernandes e Ricardo Campos**

# Capítulo IX

## Internet das coisas potencializando o resultado de modelos analíticos para prevenção de perdas não técnicas

Segundo a Agência Nacional de Energia Elétrica (Aneel), as perdas técnicas são aquelas inerentes ao transporte da energia elétrica na rede, relacionadas à transformação de energia elétrica em energia térmica nos condutores (efeito joule), perdas nos núcleos dos transformadores, perdas dielétricas, etc. Podem ser entendidas como o consumo dos equipamentos responsáveis pela distribuição de energia.

Já as perdas não técnicas correspondem à diferença entre as perdas totais e as perdas técnicas, considerando, portanto, todas as demais perdas associadas à distribuição de energia elétrica, tais como furtos de energia, erros de medição, erros no processo de faturamento, unidades consumidoras sem equipamento de medição, etc. Esse tipo de perda está diretamente associado à gestão comercial da distribuidora [1].

Pela própria definição do órgão regulador, podemos perceber a dificuldade na apuração, a definição de origens e, consequentemente, o tratamento das perdas não técnicas pelas distribuidoras, apesar de este ser o único tipo de perda que pode ser trabalhado no âmbito da gestão comercial para a melhoria da eficiência da transmissão. Dentro deste contexto, também é sabido que, dentro dos principais fatores geradores de perdas, estão os furtos de energia.

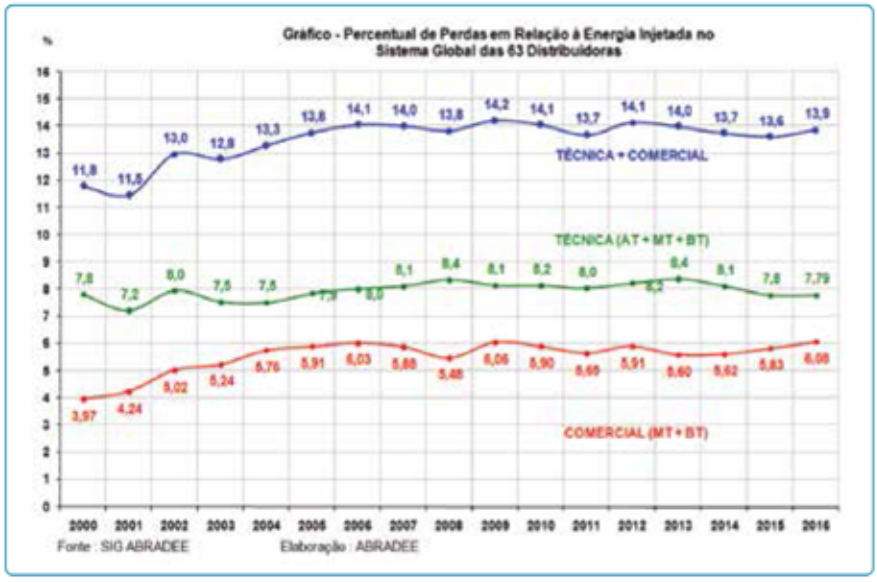


***Figura 1 – Percentual de perdas em relação à Energia Injetada.***
***Fonte: SIG Abradee; elaboração Abradee.***

Furtos e roubos de energia, seja via uma ligação direta na rede de baixa tensão ou alteração do medidor, fazem parte das chamadas perdas comerciais.

Instruções e ofertas de serviços que alteram o medidor são comuns e facilmente acessíveis. Uma busca do termo "como alterar o medidor de energia elétrica" no YouTube, em 18/08/2017, às 9h10, retorna quase 3.000 resultados. Os 15 primeiros vídeos contabilizam mais de dois milhões de reproduções e cinco desses vídeos oferecem serviços de adulteração do medidor digital via aplicativos de smartphones.

A Figura 2 mostra que as perdas comerciais variam de acordo com cada distribuidora. Um dos motivos dessa variação é a questão social. Regiões com maior concentração de aglomerados subnormais, como definido pelo IBGE, têm maior nível de irregularidade e a existência de furtos de energia é mais

***Figura 2 – Distribuidoras com maiores perdas não técnicas e a composição das perdas (fonte: SAMP/ANEEL – 2014; elaboração dos autores).***

frequente. Existe ainda um efeito de contágio por proximidade, já que, quando uma pessoa altera seu medidor, é comum os vizinhos ficarem sabendo e, ao perceberem que não há ônus financeiro ou legal, são impelidos a fazer o mesmo.

## Características do furto de energia

Algumas características específicas do furto de energia elétrica fazem com que as características da modelagem analítica para este setor tenha que obedecer a algumas premissas para que possa ser efetiva, entre elas, podemos destacar:

O potencial de retorno financeiro de um furto de energia é relativamente baixo em relação ao furto de informações de cartão de crédito. Este tipo de furto tem retorno a longo prazo, mas, apesar disso, o

risco jurídico de pena decorrente deste tipo de ato também é menor.

A regularização das unidades consumidoras depende de ação física na residência do usuário e, após a correção na medição, o ambiente continua vulnerável a novas ações por parte de fraudadores e, em algumas localidades específicas, se torna inviável qualquer ação por parte da distribuidora devido à falta de segurança física a seus colaboradores.

Além disso, fatores regulatórios associados à revisão tarifária não favorecem ações mais drásticas na redução das perdas não técnicas, em vez disso, tornam-se mais atraentes ações de melhoria contínua, porém sem grandes alterações de números ano a ano, limitando o recurso disponível para as inspeções (muitas vezes compartilhado com outras ações de campo) e inviabilizando a criação de uma estratégia que aumente o esforço operacional em campo para grandes ações.

Também devemos considerar que o roubo de energia no setor tem maior influência da questão regional, tanto de uma forma mais macro (de acordo com componentes climáticos e culturais de cada região ou estado) e de uma forma mais micro (com forte componente de utilização dos mesmos procedimentos em uma mesma vizinhança).

Todas estas características somadas fazem com que as fraudes deste setor não sejam tão atraentes para "fraudadores profissionais" (a menos que estes sejam fornecedores de insumos para que outras unidades consumidoras realizem o furto), o que torna as ações muito mais pulverizadas, apesar disso, o baixo custo de execução, aliado ao baixo risco jurídico fazem com que a fraude aconteça em grandes volumes. A oferta de serviços para viabilizar o furto de energia está facilmente acessível à maioria das pessoas.

Devemos também considerar que não existe grande clareza em relação à totalidade das unidades consumidoras que cometem fraudes, pois a vítima da fraude é a distribuidora de energia elétrica, diferentemente do setor financeiro, por exemplo, em que, na grande maioria dos casos, a vítima é o cliente final. Esta característica do setor financeiro permite que, na maioria dos tipos de fraudes, praticamente a totalidade dos casos seja identificado, já que o cliente sempre acaba reclamando do valor debitado indevidamente.

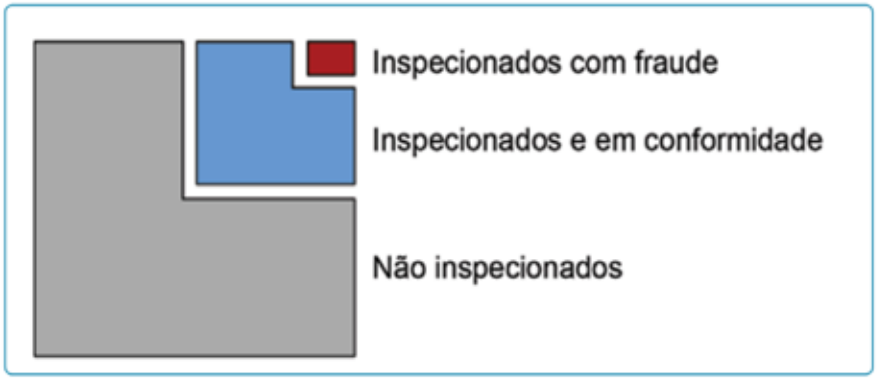


*Figura 3 – Distribuição da fraude de acordo com conformidade e inspeção.*

No caso das distribuidoras, as equipes de inspeção têm como responsabilidade o levantamento de instalações com potencial de não conformidade. Isso passa por uma avaliação julgamental, analisando o logradouro em si, mas cria um viés de seleção de casos inspecionados. Assim, somente são identificadas fraudes nos domicílios indicados pela estratégia de inspeção das distribuidoras, violando uma premissa importante em amostragem: a presença da aleatoriedade (completa ou parcial), conforme representado na Figura 3.

A identificação das ações de furto de energia não precisa necessariamente ser em tempo próximo ao real, pois a regulação permite que seja recuperada receita referente a até três meses retroativos em que a fraude estava sendo executada (energia recuperada). Entretanto, após a detecção da fraude, é possível que o domicílio altere seu perfil de consumo, mascarando o efeito do volume consumido indevidamente (energia agregada). A Figura 4 exemplifica essa ideia.

O volume e a precisão das informações disponíveis, quando consideradas somente as informações internas da distribuidora, para a criação de modelos de identificação de furtos são muito escassos e, na grande maioria das vezes, depende exclusivamente da leitura mensal executada de forma manual e de indicações de violação dos medidores, também feita através dos leituristas.

No cenário atual, em que a maioria das ações são embasadas somente nos dados existentes nas distribuidoras, a baixa quantidade de informações associada à alta

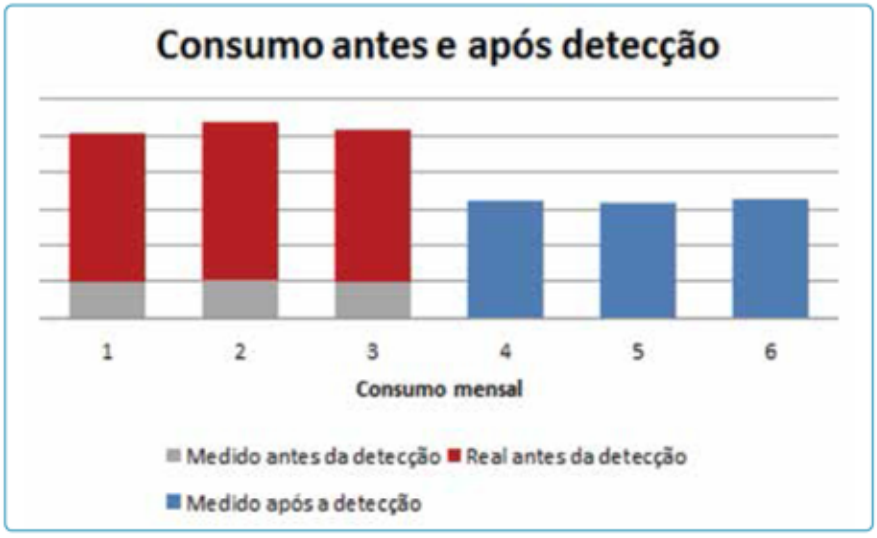


*Figura 4 – Consumo antes e após a detecção.*

quantidade de falso negativos, fazem com que os modelos criados tenham um baixo potencial de identificação de possíveis fraudadores, bem como não seja possível identificar padrões de fraudes diferentes dos já conhecidos, diminuindo em muito o valor dos mesmos.

## Ganhos com IoT:

Uma grande expectativa foi criada em torno do smart grid, que traria a possibilidade de receber uma série de informações adicionais e em tempo muito mais próximo ao real, permitindo, inclusive, que fosse possível identificar o tipo de equipamento que estava consumindo energia elétrica. Atualmente, ganha força também o uso de tecnologias não específicas do setor, que permite um desenvolvimento compartilhado entre as diversas verticais de mercado. O conceito de Internet das Coisas pode ser um grande viabilizador de modelos mais precisos.

O principal ponto a ser considerado é a viabilidade de medição remota em larga escala, o que permite fazer a aquisição de um maior volume e precisão nas informações a respeito do comportamento de cada consumidor individualmente e a capacidade de correlacionar esta informação com a de outros consumidores. Mesmo considerando que o medidor envie dez medidas por dia, já seriam 300 vezes mais do que um leiturista.

A medição remota, por outro lado, diminui a eficiência das denúncias dos leituristas, pois eles não terão o contato direto com o equipamento para detectar possíveis violações, porém, além de liberá-lo para outros serviços prioritários, o medidor conectado na distribuidora através de uma rede de comunicação permite que possam ser enviados alarmes de violação que possam indicar a tentativa de fraude física.

Outro ponto importante em ações tecnológicas coordenadas com a ação de outras verticais de mercado permite o barateamento de equipamentos a partir do momento em que a escala em que eles serão produzidos atinge outro patamar, bem como possibilita a utilização de infraestrutura compartilhada para a conectividade dos medidores (um dos exemplos pode ser visto no artigo "iluminação pública como habilitador de cidades inteligentes" desta mesma série de artigos sobre IoT), diminuindo custos de manutenção.

## Fatores preditivos das irregularidades

Não é fato novo que a presença de aglomerados subnormais (definição formal do IBGE para as chamadas favelas, comunidades, mocambos, etc.) aumenta o risco de haver perda não técnica. O Gráfico 1 mostra essa relação por meio da comparação do percentual de população moradora em aglomerados subnormais e as perdas não técnicas das 15 distribuidoras com maiores perdas. Os dados de perdas não técnicas foram coletados de estudo da Samp/Aneel de 2015 combinados com dados do Censo2010 do IBGE e da Serasa Experian. O valor do coeficiente de correlação de Pearson da ordem de 0,73 indica uma correlação positiva forte.

Nessas regiões, a ausência da infraestrutura de acesso e, muitas vezes, a falta de segurança tornam difíceis as inspeções dessas áreas, mas existe perda técnica mesmo nas demais áreas.

Atualmente, existem diversas maneiras de minimizar os efeitos das perdas comerciais, sendo duas delas uma estratégia de inspeções mais frequentes ou uma estratégia de inspeções mais direcionadas. A primeira é bastante custosa: tem uma taxa de captura altíssima das irregularidades, mas, ao mesmo tempo, também prevê um alto índice de inspeções de locais sem fraude. A segunda estratégia faz uso de técnicas estatísticas e matemáticas de aprendizado de máquinas (machine learning) para predizer quais clientes têm maior chance de apresentar irregularidades nos seus medidores. Essas técnicas utilizam informações como histórico de uso e características sociodemográficas da localização como fatores com poder preditivo relevante. Além disso, a existência de instalações irregulares pode estar associada às restrições orçamentárias dos consumidores e, por esse motivo, é natural entender que informações relativas aos endividamentos possam ser um relevante fator preditivo também.

O histórico de consumo da instalação tem um forte poder preditivo, pois é possível verificar a existência de degraus na energia consumida antes e depois da fraude, como exemplificado na Figura 6. Após a detecção da irregularidade, é possível confirmar o efeito da fraude, pois há também um

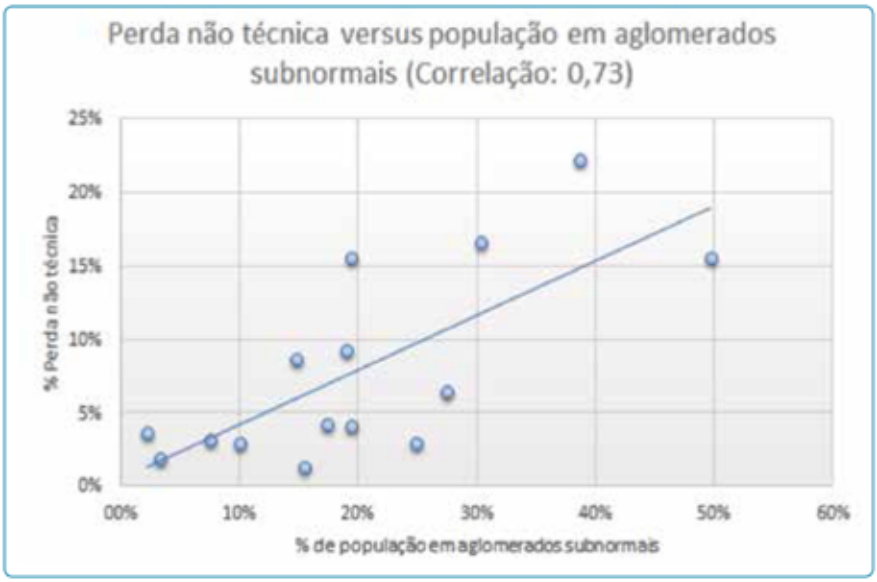


***Figura 5 – Relação da perda não técnica e do percentual de população em aglomerados subnormais (Fonte: elaboração dos autores a partir dos dados de Samp/Aneel 2015, IBGE/ CENSO2010 e Serasa Experian).***

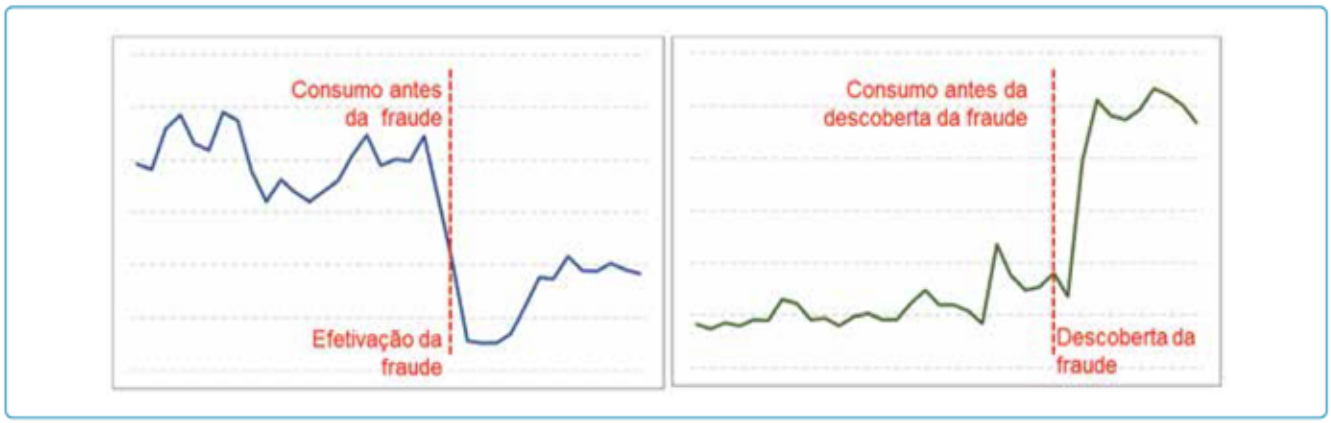


***Figura 6 – Exemplos do efeito da irregularidade no consumo.***

degrau no consumo. Entretanto, é possível que haja uma mudança de comportamento após a detecção da fraude: em outras palavras, o cliente passa a consumir energia de forma consciente, já que agora ele será devidamente cobrado por ela. Nesses casos, ao analisar o consumo de energia, o degrau após a detecção pode não ser percebido ou ser subestimado.

Além do histórico de uso, é possível utilizar dados públicos e de bureau como fatores preditivos na identificação das irregularidades. A análise de desvio de um comportamento esperado, dadas as características do cliente, também pode ser um indício de problemas no medidor. Dados do setor censitário do IBGE (como, por exemplo, quantidade de eletrônicos, PIB, etc.) podem ser uma medida de potencial do setor e, portanto, espera-se um intervalo aceitável de consumo por setor e os comportamentos anômalos devem ser analisados.

Com os dados de bureau é possível analisar o comportamento dos clientes no mercado e classificar os que possuem característica de alta renda (luxo). Ao cruzar essa informação com o consumo, espera-se não encontrar consumos muito baixos, a menos que a casa/apartamento possua uma fonte alternativa de energia, possível de observar também através do modo de vida do cliente e de análises de imagens aéreas da residência. Portanto, se o consumo for mais baixo que o esperado para um perfil que geralmente gasta mais, compreende-se outro indicativo de problemas no medidor.

Outro ponto interessante é a influência

de pessoas próximas no risco de cada cliente: se muitos vizinhos cometem a fraude, a probabilidade de uma nova irregularidade na região pode ser maior. Essa medida de distância pode estar relacionada a questões espaciais, como no caso de vizinhos, ou à proximidade das pessoas através de uma rede de influência familiar. Em outras palavras, se alguém da família cometeu uma fraude no medidor, os outros familiares podem ser influenciados a cometê-la.

Por fim, o desvio de perfil é mais um fator preditivo potencial: se a instalação deveria pagar uma categoria de tarifa diferente, como, por exemplo, uma categoria de PJ e o cadastro está como PF, essa é uma fraude de cadastro, que altera o valor da tarifa e não a quantidade de KWh consumidos. Através do cruzamento de endereço das empresas da região e as instalações dos clientes da distribuidora, é possível encontrar os perfis problemáticos. A Figura 7 – ilustra as fontes dos fatores preditivos necessários na detecção das irregularidades.

## Modelos de machine learning na estratégia de inspeção direcionada

Os fatores preditivos são usados em modelos de aprendizado de máquina (machine learning) para determinar a chance de uma instalação apresentar irregularidades. Esses fatores são usados como features ou covariáveis, nomes dados às informações mais relevantes já tratadas para serem usadas nos modelos de aprendizado de máquina. Existem muitas técnicas na literatura especializada, mas aqui apresentamos um case para comparar três das técnicas mais utilizadas: regressão logística, gradient boosting e redes neurais artificiais MLP (Multi-layer perceptron).

*Figura 7 – Fontes de fatores preditivos na detecção de irregularidades*

A regressão logística se baseia na combinação linear dos fatores para ser usada como argumento em uma função 'logito'. Essa função tem uma propriedade interessante e faz com que a combinação linear seja transformada em probabilidade de irregularidade. Assim, dependendo do perfil do cliente e das suas variáveis, a probabilidade de irregularidade pode aumentar ou diminuir.

A regressão por árvore trabalha com uma ideia de avaliação de sucessivas variáveis e a tomada de decisão para cada avaliação. Por exemplo, se o cliente não apresentou um degrau de consumo nos últimos seis meses, avalia-se a existência de pendências financeiras no bureau de crédito. Se isso se confirmar, avalia-se a coerência entre padrão de renda e consumo. Se esse último está abaixo do esperado, esse caso tem uma chance maior de ser uma irregularidade. O gradient boosting é uma técnica que combina uma sequência de regressões por árvore: essa sequência é definida de maneira que a próxima árvore busque identificar padrões que as árvores anteriores ainda não identificaram.

Por fim, a rede neural MLP é um modelo que emula o sistema nervoso humano, fazendo múltiplas combinações lineares aplicadas a funções de ativação numa primeira camada, seguidas por combinações lineares das resultantes dessas funções e assim sucessivamente. O resultado pode ser visto como uma rede que captura um alto nível de interação das features (variáveis).

A métrica de comparação das três técnicas usadas nesse artigo é calculada a partir do ganho na taxa de detecção de fraudes ($GDF_k$) em relação à taxa de fraude sem utilizar uma estratégia direcionada ($TF_0$). A métrica é definida por:

$$GDF_k = \frac{TF_k}{TF_0}$$

Em que:

$TF_k$ é a taxa de fraude detectada sob a estratégia k. No caso da nossa comparação, temos três estratégias ($k \in \{GB, RL, RN_MLP\}$) utilizadas: Gradient boosting, regressão logística e rede neural MLP. A comparação foi feita em uma amostra de teste, não usadas no treinamento das três.

A comparação foi feita utilizando os dados de uma distribuidora brasileira e conta com uma amostra de cerca de 10 mil domicílios. Os dados foram cedidos pela Serasa Experian, que conta com o maior bureau de informações de empresas e consumidores do Brasil. A Figura 8 – mostra o ganho das técnicas com relação ao benchmark de uma seleção ao acaso. Assim, a técnica de gradient boosting mostrou um desempenho 3,44 vezes superior ao benchmark, seguido pela regressão logística, com ganho de 3,04 vezes e, por fim, a rede neural MLP com 2,96 vezes. Note que, independentemente do resultado, os ganhos de ter uma estratégia baseada também em modelos de machine learning são importantes para a redução das perdas comerciais.

## Utilização de técnicas aplicadas a outros setores:

Com o objetivo de potencializar ainda mais os resultados através da diminuição da influência dos casos de falso negativo, verificamos que este cenário tem similaridade com outro caso específico em que se utiliza em larga escala a modelagem estatística no setor financeiro: o caso de concessão de crédito.

A similaridade se dá pelo fato de os indicadores de risco de crédito utilizados ainda serem muito dependentes do histórico do tomador de crédito no ambiente bancário. Em países em desenvolvimento como o Brasil, grande parte da população não possui conta bancária, o que faz com que estes modelos estejam "cegos" para grande parte da população que requisita crédito financeiro.

Uma das técnicas utilizadas para esta situação é a inferência de rejeitados. Esta técnica, grosso modo, expande a abrangência do modelo através da aplicação do modelo criado com a base das unidades consumidoras inspecionadas para os não inspecionados, posterior seleção aleatória das instalações com menor risco de fraude e criação de um novo modelo considerando os selecionados como baixo risco como não fraudadores. Isso permite que o modelo criado passe a ter maior representatividade dentro da base total de clientes, fazendo com que o modelo melhore a sua performance. Se utilizarmos a Figura 1 como referência, podemos ver que o modelo deixará de contemplar somente as figuras em azul e vermelho, mas passará a contemplar também a parte cinza.

A aplicação de técnicas de inferência de rejeitados nessa situação é importante para garantir o desempenho das soluções analíticas de detecção das fraudes. Um caso de uso prático foi realizado em um projeto de Pesquisa e Desenvolvimento da Aneel realizado em conjunto entre CPqD e Elektro. A abordagem inovadora, inclusive, foi alvo de um pedido de patente feito pelo CPqD e pela Elektro e depositado no INPI através do código BR 10 2013 009314 9.

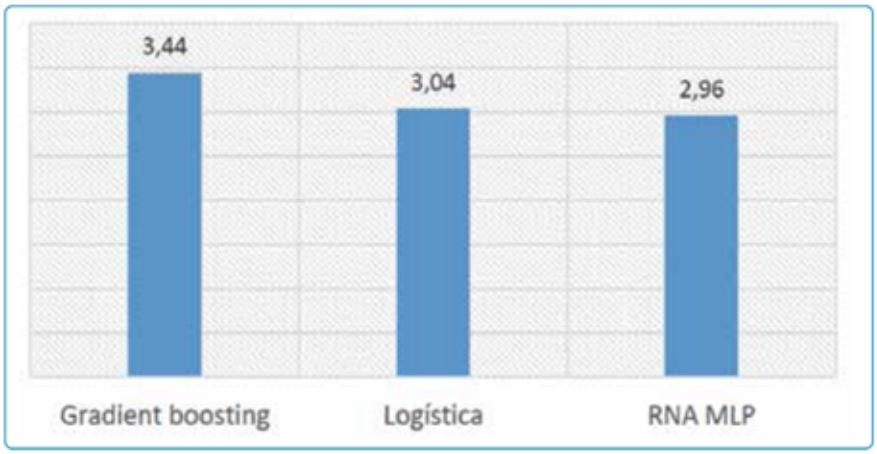


*Figura 8 – Ganho na detecção de fraude em relação ao benchmark.*

## Operacionalização dos modelos preditivos e a capacidade de utilização de regras de negócio:

No contexto colocado anteriormente, a maneira que possibilita maximizar a recuperação dos valores comerciais perdidos é uma aplicação que direcione as ações de inspeção considerando os seguintes critérios: precisão na escolha das UCs fraudadoras e que possuam maior potencial de perda e maior potencial de recuperação.

Esse diferencial é possível através do uso de uma plataforma de decisão que possibilita a correlação de diversas informações internas da distribuidora (faturamento, leitura, perfil de equipe, efetividade de equipe, dados cadastrais, etc.), informações externas (bureaus de crédito, listas negativas e positivas, etc.), aplicação de técnicas de machine learning, inteligência artificial, otimização, dentre outros, combinadas com aplicação de regras de negócio para trazer inteligência, padronização, agilidade e maior precisão para área de combate a perdas a partir do tratamento dos alertas que irão bloquear as principais perdas e maximizar a recuperação de receita.

Para fechar o ciclo de otimização do combate às perdas não técnicas, é importante a capacidade de automação na geração de listas de inspeção através de uma plataforma de gerenciamento de decisão, que permite a criação de regras e utilização de modelos de escore para que a lista de unidades consumidoras a serem inspecionadas de acordo com o cenário vigente seja gerada de maneira automática.

Em grande parte das distribuidoras, a geração das listas de inspeção é dependente de uma lista inicial gerada a partir de regras codificadas diretamente nos sistemas transacionais, o que faz com que estas devam ser genéricas e com poucos filtros, pois qualquer alteração demanda uma alteração no sistema, o que é custoso tanto para a área de negócio que perde em agilidade, quanto para a área de tecnologia que gasta muito tempo executando as alterações. Uma solução de gerenciamento de decisão baseada em um motor de regras permite que as regras fiquem fora dos sistemas e possam ser alteradas de maneira intuitiva pela própria área de negócio.

Outra vantagem observada na utilização deste tipo de tecnologia é a independência de processos manuais que podem não estar alinhados com uma diretriz estratégica da empresa ou sofrer variação de acordo com o recurso humano que gera a lista de inspeção. Com o trabalho automatizado, os usuários de negócio têm mais tempo para executar análises ao invés de realizar trabalhos repetitivos e que requerem pouco esforço intelectual.

## Conclusão

As distribuidoras de energia elétrica têm apresentado importantes perdas comerciais, ou não técnicas, devido a irregularidades, em particular, as alterações de medidores. As 11 distribuidoras com maiores perdas dessa natureza somaram quase três bilhões de reais em 2014 e, ainda que existam medidores digitais que dificultem a fraude, existe um fácil acesso a serviços de alteração de medidores.

A inspeção presencial é, na maioria das vezes, a melhor maneira de corrigir instalações com irregularidades, mas apresenta um alto custo associado. A associação de dispositivos de medição inteligente, conectividade, unificação de todas as informações disponíveis (de dentro ou de fora da distribuidora) quebrando os silos de informação, aliada a modelos de machine learning são uma alternativa para identificar instalações com maiores chances de fraude. Existem diversos fatores preditivos, desde histórico de consumo até análise de imagens da residência ou estabelecimento, incluindo o uso de uma Plataforma de Decisão como a oferecida pelo CPqD e o uso de dados do bureau da Serasa Experian.

Quando utilizamos tecnologias de internet das coisas para permitir uma maior quantidade e melhor qualidade nas informações, potencializando os resultados dos modelos analíticos e viabilizamos ações mais próximas ao tempo real, abrindo o leque com relação às estratégias de combate a perdas.

Por fim, o uso de uma estratégia de inspeção direcionada, que une o poder dos dados com a capacidade preditiva dos modelos de machine learning, possibilitou melhorias na eficiência do processo de detecção de fraudes, quando comparada ao desempenho do benchmark. Quando esta capacidade adicional de tratamento de dados é aliada a técnicas utilizadas em outros setores, se soma a regras de negócio preestabelecidas e ainda pode ser automatizada diretamente nos sistemas operacionais das distribuidoras de uma forma amigável para o usuário de negócio, é criado um novo patamar na capacidade das distribuidoras coibirem fraudes.

## Referência:

[1] Referência em http://www2.aneel.gov.br/area.cfm?idArea=801&idPerfil=4

*Fábio Yuasa Niizu possui graduação em Engenharia de Computação pela Universidade Estadual de Campinas (2001) e MBA em Gestão Estratégica em Marketing pela ESAMC (2012). Atua com pesquisa e desenvolvimento desde 1994, passando por grandes empresas como IPT, Alcatel, Nortel e VIVO. Atualmente trabalha como Especialista em Inovação na Fundação CPqD.

Guilherme Fernandes é bacharel em estatística e mestre em Economia. Hoje, atua como Data Scientist no Datalab Latam da Serasa Experian com foco de pesquisa em soluções antifraude e modelos de crédito.

Ricardo De Rosso Campos é graduado em Engenharia de Computação (PUCCAMP) com MBA em Gestão de Negócio (FGV). Liderou implantações de projetos de Business Intelligence e gerenciou equipe de desenvolvimento do produto CPqD Antifraude, implantando este produto em três grandes bancos do país. Atualmente é responsável pela equipe de Business Analytics e pelo produto CPqD Plataforma de Gerenciamento de Decisão.

**CONTINUA NA PRÓXIMA EDIÇÃO**
Acompanhe todos os artigos deste fascículo em www.osetoreletrico.com.br
Dúvidas, sugestões e outros comentários podem ser encaminhados para redacao@atitudeeditorial.com.br