

# Garimpendo Fraudadores: O Uso da Mineraço de Dados na Descoberta de Padroes de Consumidores que Causam Perdas Comerciais em uma Concessionaria de Energia Eletrica

Manoel G. de Mendonça Neto<sup>1</sup>, Ricardo E. P. Vieira<sup>2</sup>, Izabel C. Andion<sup>2</sup>, Janaina F. Ma<sup>2</sup>, Paulo V. Nascimento<sup>2</sup>, Pedro Dantas<sup>3</sup>, Monica Silveira<sup>3</sup>, Alfredo J. Brito<sup>4</sup>

<sup>1</sup>Departamento de Ciencia da Computaço – Universidade Federal da Bahia (UFBA) - Salvador, BA

<sup>2</sup>Laboratorio de Sistemas Integraveis Tecnologico (LSITEC NE) – Salvador, BA

<sup>3</sup>Consulte Energia – Salvador, BA

<sup>4</sup>Companhia de Eletricidade do Estado da Bahia (COELBA) – Salvador, BA

{manoel.g.mendonca, argollo.vieira, icandion, janainafma} @gmail.com,  
paulovnc@yahoo.com, dantas.pedro@yahoo.com.br,  
monica.silveira@consulteenergia.com.br, {ajbrito} @coelba.com.br

**Abstract.** The objective of this work was to develop a methodology supported by a software prototype to help an electricity distribution company (Coelba) to pre-select electrical energy consumers for locus inspection, improving their success in inspections. This study proposes the use of Data Mining techniques to identify suspicious behavior patterns, analyzing historical and actual data of electric energy consumption and consumers payments.

**Resumo.** O objetivo deste trabalho foi desenvolver uma metodologia apoiada por um prototipo de software que auxilie os analistas da Companhia de Eletricidade do Estado da Bahia (Coelba) em seu processo de qualificaço de consumidores para inspeço, aumentando o seu percentual de acerto na identificaço de perdas comerciais. A pesquisa envolveu a aplicaço de tecnicas de mineraço de dados para identificaço de padroes de comportamento dos clientes que causam perdas comerciais a concessionaria, analisando atributos referentes ao consumo e a inadimplencia dos consumidores.

## 1. Introduço

Os sistemas de distribuiço de energia eletrica sofrem perdas de energia que podem ser subdivididas, de acordo com sua origem, em perdas tecnicas e comerciais. As perdas tecnicas se caracterizam por ocorrer antes do ponto de entrega. As perdas comerciais sao aquelas associadas a comercializaço da energia fornecida ao usuario final. Elas sao administrativas, quando nao ha responsabilidade do consumidor e de intervenço ilicita, quando feitas por usuarios legalizados ou nao, mas nao autorizadas pela concessionaria de energia, com o objetivo de reduzir o registro do consumo [Souza e Silva, et al, 2005].

O valor das perdas comerciais no sistema elétrico nacional tem crescido desde 1980. Particularmente nos anos recentes, as perdas se intensificaram devido ao aumento dos furtos e das ligações clandestinas. Hoje, no Brasil, 16% de toda energia produzida é perdida. No estado da Bahia a perda comercial é estimada em 6%, porém em outros estados brasileiros esse valor alcança os 15% [Almeida e Dantas et al,2005].

Para mensuração exata das perdas comerciais de uma concessionária de energia elétrica seria necessário realizar inspeção nas medições de todas as unidades consumidoras, quantificar todas as ligações clandestinas, além de executar o levantamento da iluminação pública de toda a área selecionada. Diante da impossibilidade da realização de todas essas ações, as concessionárias vem desenvolvendo processos diversos para a identificação de perdas comerciais, dentre eles a utilização de novas tecnologias, intensificação da fiscalização, campanhas educativas, acompanhamento do padrão de consumo de clientes e até o registro de ocorrências policiais quando fraudes são confirmadas. Dessa forma, o processo de identificação das perdas comerciais passou a ser um procedimento indispensável e rotineiro.

Ao mesmo tempo, a revisão mais recente da Resolução 456/ANEEL/2000 reduziu os prazos para cobrança de retroativos de energia fornecida em casos de defeitos na medição, o que levou à necessidade de diminuição do período entre as inspeções de unidades consumidoras. Assim, nos últimos três anos as concessionárias têm intensificado seu programa de inspeções, buscando reforçar cada vez mais a redução do furto de energia e também a detecção de defeitos elétricos em suas áreas de concessão.

Devido aos custos elevados com a montagem e manutenção das equipes de inspeção em campo, é importante que haja um processo preliminar que permita qualificar as unidades consumidoras com maior potencial de causarem perdas comerciais à concessionária. Os processos para escolha destas unidades são particulares a cada concessionária, mas todas tem como objetivo garantir o maior percentual de acerto possível.

Esta pesquisa foi desenvolvida no âmbito de um projeto de Pesquisa e Desenvolvimento (P&D) firmado entre Coelba, Consulte Energia e LSITEC Nordeste, no biênio 2008/2009, com o objetivo de desenvolver uma metodologia apoiada por um protótipo de software para indicação de unidades consumidoras qualificadas para inspeção por suspeita de causar perdas comerciais à Coelba, no sentido de aumentar o percentual de acerto da seleção destas unidades.

Dada a impossibilidade técnica e financeira de se trabalhar com todo o universo de consumidores da Coelba, critérios como volume de dados e consumos mensais significativos foram estabelecidos para limitar o escopo de atuação da pesquisa. Desta forma, os consumidores Industriais do Grupo B serviram de base para o desenvolvimento deste trabalho.

Esta é uma pesquisa de cunho exploratório, portanto, de caráter preliminar. A metodologia adotada envolve uma pesquisa bibliográfica de trabalhos correlatos desenvolvidos junto a concessionárias de energia elétrica e um estudo de caso realizado na Coelba, junto ao Departamento de Inspeção de Clientes (OIN). No estudo de caso, técnicas de mineração e análises estatísticas foram utilizadas para analisar dados históricos e atuais de consumo e inadimplência dos consumidores da Coelba, a fim de buscar correlações entre atipicidade nos consumos e inadimplência na Coelba com uma

postura causadora de perdas comerciais.

## **2. Mineração de Dados**

A Mineração de Dados consiste numa abordagem de análise de dados voltada à descoberta de conhecimento. [Mendonça Neto, 2001] a define como “o processo de indução de informação não trivial, previamente desconhecida, e potencialmente útil de repositórios genéricos de dados”. Este tipo de abordagem tem se popularizado devido ao crescimento contínuo da complexidade e tamanho dos repositórios de dados disponíveis. Normalmente este processo consiste de quatro etapas. Na primeira etapa, denominada de Seleção de Dados, o analista identifica onde estão os dados desejáveis, ganha acesso a estes e os transporta do seu local original para o repositório em que ele vai trabalhar. Em seguida é feito o Pré-processamento, quando são formatados, adaptados e, algumas vezes, transformados para que possam ser usados pelo algoritmo de mineração. A Mineração objetiva extrair informação interessante dos dados. Esta etapa pode envolver técnicas muito diversas. Neste trabalho foram estudadas técnicas de aprendizagem supervisionada (classificação) e não-supervisionada (agrupamento). Por fim a Assimilação consiste em analisar a informação minerada. Quando a mineração envolve a construção de modelos, este passo consiste em avaliar a robustez e efetividade dos modelos produzidos. Se aprovados, os modelos devem ser incorporados aos processos operacionais da instituição na qual eles vão ser usados. Quando se trata da extração de padrões, este passo consiste em tentar interpretar a informação extraída pelo algoritmo de mineração de dados. Isto é geralmente feito por um perito no domínio da aplicação que vai usar seu conhecimento contextual para verificar se a informação minerada está lhe revelando algum conhecimento.

### **2.1. Técnicas de Mineração**

A aprendizagem supervisionada ou tarefa de classificação consiste em identificar padrões que possam prever o comportamento de novos itens de um banco de dados a partir de dados existentes. Para identificar esses padrões são usados conjuntos de dados de treinamento, onde é conhecida previamente a classificação dos objetos. “Estes algoritmos são preditivos, pois suas tarefas de mineração desempenham inferências nos dados com o intuito de fornecer previsões ou tendências, obtendo informações não disponíveis a partir dos dados disponíveis” [Silva, 2008]. A tarefa de classificação envolve dois passos: treinamento e teste. No primeiro passo são construídas as regras de classificação dos objetos baseadas nos dados de treinamento. No segundo, o modelo criado é utilizado para classificar novos objetos contidos num novo conjunto de dados. Uma das técnicas de indução existentes são as árvores de decisão. Elas são usadas para descobrir regras de classificação por um atributo, a partir da subdivisão sistemática dos dados contidos no repositório analisado. Esta técnica leva em consideração as regras mais relevantes e apresenta estas regras em ordem de importância, possibilitando ao usuário identificar os fatores que mais influenciam o resultado esperado.

“Na aprendizagem não-supervisionada, o rótulo da classe de cada amostra do treinamento não é conhecido e o número ou conjunto de classes a ser treinado pode não ser conhecido a priori. Estes algoritmos são descritivos, pois descrevem de forma concisa os dados disponíveis, fornecendo características das propriedades gerais dos

dados minerados” [Silva, 2008]. Agrupamento é um exemplo deste tipo de aprendizagem.

Rocha (2008) ressalta a importância de se avaliar a correção e a confiabilidade do conhecimento produzido na mineração, já que o resultado desta etapa, muitas vezes, é uma grande quantidade de regras que podem ser irrelevantes ou não interessarem ao usuário. Para isso, deve-se analisar as informações estatísticas produzidas pelos próprios algoritmos de mineração. As medidas de avaliação de regras podem ser verificadas através de suas medidas de desempenho como: precisão, erro, sensibilidade, entre outras, que são técnicas estatísticas em que a correção e a confiabilidade do conhecimento são apuradas.

“A precisão de uma regra significa quanto uma regra é específica para o problema”, ou seja, é o percentual de acertos de uma regra em relação a todos os exemplos cobertos por ela. Isto significa que “quanto maior o valor desta medida, mais corretamente a regra cobre os exemplos de uma classe”. Já o erro é o percentual de erros de uma regra em relação aos exemplos que ela cobre. “Significa que quanto maior, menos precisamente a regra cobre exemplos de sua classe” [Rocha, 2008]. A sensibilidade é o percentual de acertos de uma regra em relação a todos os exemplos corretos da amostra. Informações detalhadas sobre medidas de desempenho de regras podem ser obtidas em Lavrac, Flach e Zupan (1999).

### **3. Aspectos Metodológicos**

Este projeto de pesquisa envolveu o desenvolvimento de uma metodologia e de um protótipo de software, com a finalidade de definir os parâmetros de maior significado para caracterização de consumidores que causam perdas comerciais à Coelba, por fraude ou por defeito, e a partir destes parâmetros, contribuir para uma identificação das unidades consumidoras mais qualificadas para inspeção em campo.

Para isto foi feita uma pesquisa bibliográfica de trabalhos correlatos e desenvolvido um estudo de caso, com o objetivo de comprovar os pressupostos adotados na pesquisa: 1. a existência de correlação entre afundamento de consumo e perdas comerciais; e 2. a existência de correlação entre inadimplência na concessionária e perda de energia elétrica por fraude. A partir dele foi possível elaborar a metodologia proposta para indicação de inspeções e desenvolver a ferramenta de apoio a esta metodologia.

#### **3.1. Pesquisa Bibliográfica**

Foram encontrados alguns trabalhos na literatura acadêmica que desenvolveram técnicas para identificar unidades consumidoras potenciais causadoras de perdas às empresas. As Centrais Elétricas de Santa Catarina S.A. junto ao Laboratório de Engenharia do Conhecimento da UFSC desenvolveram um sistema para identificação de possíveis fraudadores, empregando o processo de descoberta de conhecimento (Todesco e Morales, et al, 2005). Este estudo envolveu a análise de dados referentes ao consumo de clientes comerciais e residenciais. Foi criada uma medida chamada score, que calcula a diferença entre o consumo do mês atual e o consumo do mesmo mês no ano anterior e estabelecidas faixas de valores para estes scores. Além disso, criou-se

também um score acumulado, que representa a soma dos scores dos últimos 12 meses. Estas medidas foram importantes para neutralizar as quedas de consumo decorrentes de sazonalidade e dissipar os defeitos em equipamentos, já que a pesquisa foi voltada à identificação de possíveis fraudadores. Um fator interessante foi o desenvolvimento de um protótipo que oferece um ranking dos suspeitos através do score acumulado.

O segundo trabalho, desenvolvido na ENERSUL S.A. em parceria com as Universidades Federais de Uberlândia e do Mato Grosso do Sul (Filho, Delaiba, Manzina, Cabral e Pinto, 2004), refere-se à construção de um sistema que preseleciona clientes para inspeção por fraude ou defeito nos equipamentos de medição. Para isto, os pesquisadores utilizaram dados históricos de consumo e de inspeções realizadas para classificação através de árvores de decisão. Depois da etapa de treinamento, os testes apresentaram uma performance de 40% de acerto na seleção de fraudadores em contraposição aos 5% alcançados pela empresa.

Por fim, um terceiro trabalho, realizado na Sociedade de Abastecimento de Água e Saneamento (SANASA) em Campinas (Passini, 2002, 2008), procurou identificar em dados históricos da empresa um perfil de comportamento dos consumidores fraudadores. Foram elaborados modelos preditivos utilizando técnicas de agrupamento e de classificação por árvores de decisão.

Estes trabalhos sinalizaram a pertinência das técnicas de agrupamento e classificação para este tipo de estudo, além de confirmar o atributo variação no consumo como um indicador de ocorrência de fraude ou defeito em equipamentos de medição. Todos eles reforçam também a necessidade de bases de dados confiáveis e com qualidade e abrangência de atributos, que permitam a diversidade de análises e consequentemente aumente as chances de descoberta de conhecimentos válidos, novos e interessantes aos usuários.

### **3.2. Estudo de Caso**

O estudo de caso realizado consistiu na aplicação de processos de mineração de dados, envolvendo dados internos à concessionária, a fim de estabelecer correlações entre eles e as ocorrências de perdas comerciais na medição de energia.

Os critérios utilizados neste trabalho foram os consumos mensais de energia dos clientes da concessionária e seu comportamento de inadimplência. A metodologia utilizada envolveu duas fases de análise dos dados: 1. análise de dados de fundamento de consumo dos clientes, utilizando algoritmos de mineração de dados para detectar padrões de consumo que caracterizem unidades consumidoras que causam perdas comerciais; e 2. análise de dados de inadimplência de clientes para detectar relação com unidades consumidoras que causam perdas por fraudes.

Dada a impossibilidade técnica e financeira de se trabalhar com todo o universo de consumidores da concessionária, foi necessário limitar o escopo de atuação do estudo. Dois critérios foram estabelecidos para a seleção: 1. que os consumidores tivessem consumos mensais significativos, a fim de que as perdas comerciais detectadas representassem uma recuperação expressiva de receita e 2. que o volume de dados fosse compatível com os recursos destinados à pesquisa. Desta forma, os consumidores Industriais do Grupo B serviram de base para o desenvolvimento deste trabalho.

### 3.2.1. Análise da Correlação entre Afundamento de Consumo e Perdas Comerciais

A seleção dos dados internos dos clientes Industriais do Grupo B da concessionária foi feita pelos analistas da própria organização. Eles disponibilizaram os dados cadastrais atualizados dos contratos, as inspeções realizadas de 2004 a 2009 e os dados referentes aos consumos de 2003 a 2009. Após as devidas limpezas, o volume de dados foi de 19.103 registros no cadastro de contratos, 6.144 inspeções e 1.060.954 registros de consumo.

A transformação dos dados consistiu no cálculo do afundamento de consumo de cada contrato inspecionado. Foram calculados os seguintes atributos relativos a afundamento, definidos com os especialistas da concessionária: o percentual de afundamento do mês da inspeção em relação à média das últimas três, seis, nove e doze medições de consumo e o percentual de afundamento do mês da inspeção em relação ao mesmo mês do ano anterior. Aos cinco percentuais foram atribuídas as seguintes categorias: “NÃO” para valores menores que 20%; “MÉDIO” para valores entre 20% e 60%; ALTO para valores maiores que 60% e “NULO” para resultados sem valor. Um atributo denominado “pontuação” foi então criado através do somatório dos pontos de cada atributo de afundamento. Estes pontos foram obtidos a partir das seguintes faixas de percentual de afundamento: de 0 a 20% - 0; de 21 a 40% - 1 ponto; de 41 a 60% - 2 pontos; de 61 a 80% - 3 pontos e de 81 a 100% - 4 pontos.

Os dados foram então submetidos aos algoritmos de agrupamento e de classificação encontrados na ferramenta Weka (Waikato Environment for Knowledge Analysis). Esta ferramenta disponibiliza algoritmos que implementam diversas técnicas de mineração de dados. Dentre os algoritmos testados, os que apresentaram os melhores resultados foram o Expectation Maximization (EM) para a técnica de agrupamento e o J48 para a técnica de classificação.

A execução do agrupamento foi feita de uma única vez utilizando todas as inspeções e o indicador de perdas (fraude, defeito ou sem perda) foi a classe selecionada para a tarefa. Para a classificação, os dados foram divididos em dois conjuntos: dados de treinamento, contendo as inspeções de 2004 a 2008 e dados de teste, com as inspeções de 2009. A tabela 1 apresenta a distribuição das amostras em relação às perdas.

**Tabela 1 – Distribuição das amostras em relação a perdas comerciais.**

Amostra	Total	Fraude	Defeito	Sem perda
Treinamento	6144	856	525	4751
Teste	953	113	91	749

Os dados de treinamento foram submetidos ao algoritmo de classificação, usando como classificador o indicador de perdas, a fim de construir um modelo de predição através de regras utilizando os atributos de afundamento. Os resultados da técnica de classificação foram mais expressivos. A partir dos atributos de afundamento, várias regras foram geradas e verificou-se que algumas delas obtiveram medidas de desempenho consideráveis. Selecionou-se então aquelas cuja precisão no acerto das indicações de perdas (fraudes e defeitos) foi maior que 30%. As seguintes regras se enquadraram nesta

especificação:

- AFN\_3 e AFN\_12 = 'ALTO'
- AFN\_3 = 'ALTO'
- AFN\_3, 6, 9, 12 e 13 = 'ALTO'
- AFN\_6 = 'ALTO'
- AFN\_3, 6, 9 e 12 = 'MÉDIO' e AFN\_13 = 'ALTO'

Em seguida, iniciou-se a fase de testes das regras selecionadas. O objetivo desta fase foi aplicá-las a um novo conjunto de dados para avaliar a sua confiabilidade. Para isto, as inspeções realizadas de janeiro a agosto de 2009 foram utilizadas como amostra.

Em geral, os percentuais de precisão no acerto caíram, mas ainda confirmaram a relação existente entre afundamento e perdas comerciais, inclusive através do atributo pontuação de afundamento. A tabela 2 apresenta as médias das medidas obtidas pelas regras nas duas fases.

**Tabela 2 – Médias das medidas de desempenho das regras.**

Fases	Médias	
	Precisão	Sensitividade
Treinamento	40%	4%
Teste	18%	2%

### **3.2.2. Análise da Correlação entre Inadimplência na Concessionária e Fraudes**

Nesta etapa, foram verificadas: 1. a ocorrência de inadimplência nas inspeções realizadas de agosto de 2008 a agosto de 2009 que detectaram fraudes e 2. as ocorrências de fraudes, defeitos e sem perda nos contratos que obtiveram inadimplência neste período. Considerou-se inadimplente o contrato que obteve pelo menos uma fatura em aberto no mês da inspeção ou até três meses antes dela. Nas inspeções com fraude analisadas, observou-se que 58 dos 122 registros possuíam inadimplência, ou seja, 48% das fraudes obtidas em inspeções apresentaram também inadimplência.

Analisando todas as ocorrências de fraude, defeito e sem perda nas inspeções e verificando se tiveram inadimplência até três meses antes da data da inspeção, observou-se que dos 519 registros encontrados, foram detectadas 58 fraudes (11%), 46 defeitos (9%) e 415 (80%) inspeções sem perda, ou seja, nas inspeções realizadas, em 11% dos contratos inadimplentes também foi detectada a ocorrência de fraude. As análises revelaram uma presença relevante (48%) de inadimplentes entre os fraudadores. Isto significa que pelo menos metade da população de fraudadores é também inadimplente no mesmo mês da inspeção ou até três meses antes dela. Isto não significa, no entanto, que os inadimplentes sejam fraudadores. O percentual de inadimplência encontrado para fraudadores no universo foi de 11%.

## **4. Metodologia Proposta e o Protótipo MONITOR**

A metodologia resultante deste projeto tem como objetivo qualificar consumidores como suspeitos de causarem perdas comerciais à Coelba, auxiliando os analistas do OIN a gerarem listas de clientes para inspeção em campo que alcancem o maior percentual possível de acerto. Para isso, a metodologia dispõe do protótipo desenvolvido, chamado

de MONITOR, que agiliza a execução da maioria de suas etapas, garantindo maior segurança no processamento dos dados.

#### 4.1. Processo de formatação dos dados de entrada

Este processo será realizado pelo administrador com a finalidade de preparar os dados que serão carregados no MONITOR. Serão extraídos das bases de dados corporativas e preparados quatro arquivos em formatos csv (contrato, consumo, inspeção e inadimplência) de acordo com um padrão pré-definido.

#### 4.2. Processo de carga dos dados

O MONITOR disponibiliza a funcionalidade de carga dos dados, conforme apresentado na figura 1. O usuário poderá atualizar a tabela de contratos, inserir novos registros de consumo, de inspeções e informações de inadimplência. Cada tabela possui uma data que servirá de controle sobre os dados a serem utilizados na análise.

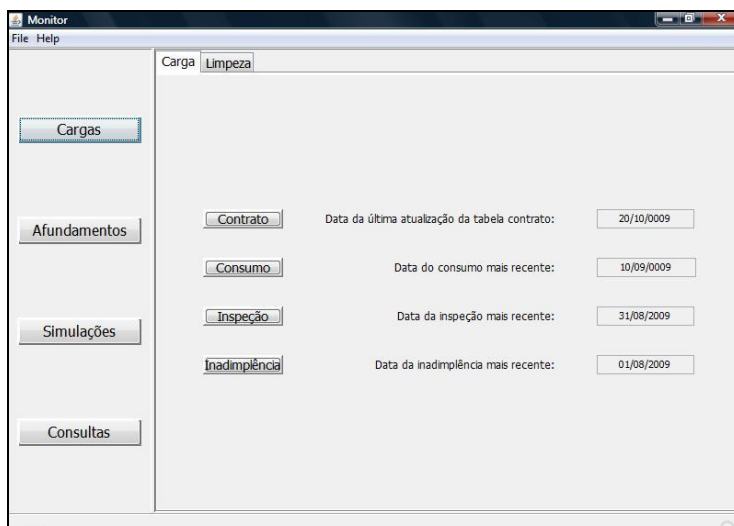
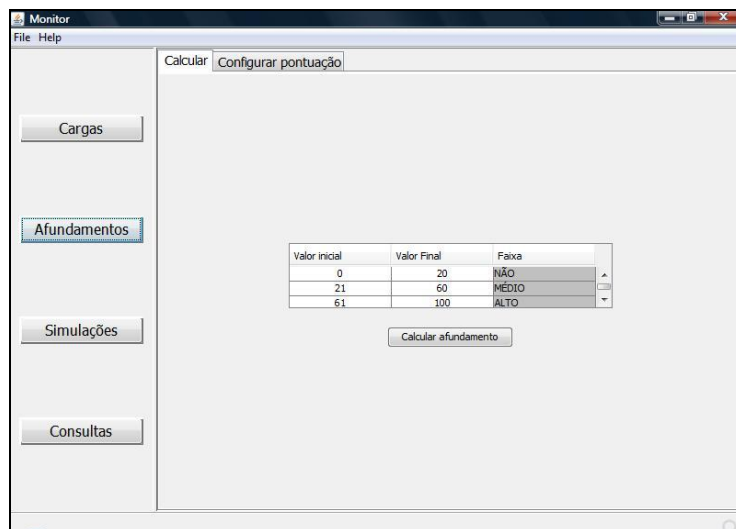


Figura 1. Tela para carga dos dados

#### 4.3. Processo de cálculo do afundamento

Este procedimento representa a automatização do processo de cálculo dos atributos de afundamento e de sua respectiva pontuação para cada consumidor, que de maneira ágil e segura processa o grande volume de dados utilizados nos cálculos. Conforme demonstrado na figura 2, as faixas das categorias de afundamento e de pontuação poderão ser configuradas antes de comandar o cálculo.

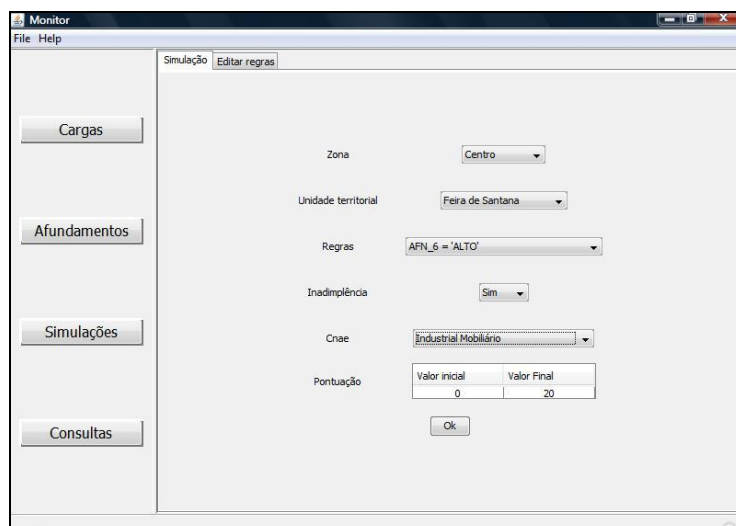




**Figura 2 – Tela para cálculo do afundamento e da pontuação**

#### **4.4. Processo de simulação das listas para inspeção**

Este passo permite, através da escolha de diversos parâmetros, a simulação de uma lista de consumidores qualificados para inspeção. O usuário poderá filtrar os consumidores por zona, unidade territorial e atividade econômica. Ele deverá escolher uma regra de afundamento, definir o corte para pontuação e um valor para o indicador de inadimplência (Sim ou Não). Caso o usuário deseje criar uma nova regra ou excluir regras existentes, o protótipo disponibiliza esta opção. A figura 3 apresenta uma ilustração em que foram selecionados contratos da zona ‘Centro’, unidade territorial ‘Feira de Santana’, CNAE ‘Indústria Mobiliária’, com indicador de inadimplência ‘Sim’ e utilizando todo o intervalo de pontuação. A regra de afundamento escolhida foi AFN\_6 = ‘ALTO’.



**Figura 3 – Tela para simulação de listas**

Após a efetivação da seleção, o protótipo gera uma lista em planilha Excel dos consumidores potencialmente suspeitos de perda comercial. A figura 4 exibe a lista resultante da seleção da figura 3.

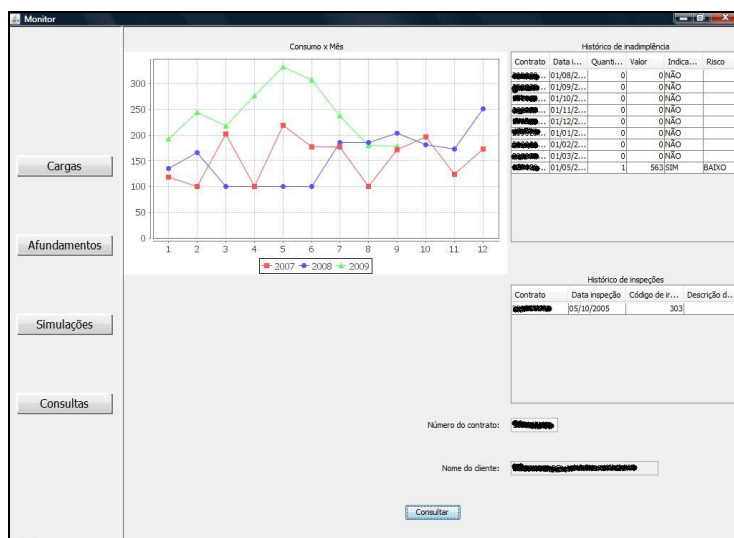
Contrato	Nome	Cons.ref.	Data ref.	Média3	Média6	Média9	Média12	Pont.	Inad.
111111111	JOAO SANTANA	100	12/08/2009	100,00	636,83	810,33	897,08	16	S
222222222	MARCOS REIS	50	31/08/2009	169,00	169,50	157,11	151,09	16	S
333333333	MARTA QUEIROS	100	28/08/2009	194,00	585,83	760,67	810,50	18	S
444444444	MOBILIARIA PADRAO	100	26/08/2009	166,00	588,83	608,78	481,58	15	S
555555555	MOBILIARIA RUSTICA	50	01/09/2009	50,00	130,33	103,56	71,50	5	S
666666666	LUCIO SANTOS	30	28/08/2009	39,00	95,67	118,44	132,00	14	S

**Figura 4 – Lista de consumidores suspeitos de causarem perdas comerciais**

Por questões de segurança e privacidade, as informações constantes na figura 4 são fictícias, usadas apenas em caráter ilustrativo.

#### 4.5. Processo de Análise dos Especialistas

A última etapa da metodologia corresponde às análises dos especialistas no domínio, que a partir de seus conhecimentos e experiência podem filtrar ainda mais a lista gerada. Para auxiliar em suas análises, os especialistas podem recorrer a informações disponíveis em outros meios ou até mesmo em outros sistemas informatizados. Neste sentido, o próprio MONITOR oferece a opção de Consultas, que permite analisar em detalhe determinado consumidor. A partir do código do contrato, são apresentadas informações históricas de valores de consumo por ano, de inspeções realizadas e de posições de inadimplência, conforme mostrado na figura 5. Desta forma o especialista poderá avaliar e se certificar se o consumidor será ou não candidato a inspeção.



**Figura 5 – Tela para consulta a partir de um contrato.**

### 5. Considerações Finais

A metodologia proposta traz como contribuição a descoberta de regras que classificam clientes causadores de perdas comerciais a partir do afundamento de consumo. Estas regras, obtidas a partir da mineração de dados históricos de inspeções de 2004 a 2008 alcançaram um acerto médio de 18% quando testadas nas inspeções de 2009 e foram incorporadas no protótipo desenvolvido.

O protótipo, tendo como base principal as informações de afundamento de consumo dos clientes, amplia a capacidade de análise dos gestores ao permitir o cálculo

seguro e ágil de diversos percentuais de afundamento, de todos os clientes Industriais do grupo B, a partir do software desenvolvido. Com isso, ela permite a simulação de diversas listas de clientes para inspeção, a partir da configuração desejada pelo usuário, que dispõe de atributos como atividade econômica, zona, unidade territorial e inadimplência do cliente, além das regras descobertas na etapa de mineração, cuja precisão no acerto na fase de treinamento foi superior a 30%.

O diferencial de análise disponibilizado por esta variedade de atributos num ambiente integrado é a possibilidade de combiná-los, simulando listas conforme a demanda dos analistas. Vale ressaltar que este modelo disponibiliza também a variável inadimplência, que pode ser combinada às análises de afundamento para fortalecer as indicações de clientes para inspeção.

Estas listas simuladas, no entanto, não devem ser consideradas como um produto final da análise. Elas foram concebidas como um primeiro filtro no universo de aproximadamente 20.000 clientes, para então sofrer análises mais aprofundadas por parte dos especialistas no domínio. Neste sentido, a ferramenta também pode auxiliar os gestores, através das consultas por contrato, para obter informações históricas sobre consumos, inspeções e inadimplências. Espera-se que os gestores sejam capazes de calcular o afundamento deste total de contratos periodicamente com agilidade e segurança, concentrando seus esforços em análises mais estratégicas.

A automatização do processo de cálculo do afundamento e a possibilidade de simulação de diversas listas de clientes para inspeção, a partir de diferentes variáveis, podem ser estendidas a outros grupos de clientes. O protótipo foi desenvolvido tendo como base a entrada de dados de clientes Industriais do Grupo B, mas, uma vez validado pelos especialistas no domínio e incorporado às suas rotinas de trabalho, é possível adaptá-lo para que suas funcionalidades sejam utilizadas também com outros grupos de clientes.

Sendo assim, a finalização desse projeto traz expectativas de ampliação da amostra das bases de dados para análise, da possibilidade de realização de novos testes, a partir da incorporação da ferramenta à rotina dos analistas e de novas demandas que surgirão no sentido de melhorar continuamente os processos de qualificação de clientes para inspeção.

## **6. Referências Bibliográficas**

Almeida, M.A.S., Dantas, P.R.P., Valente, A.L.C., Silveira, M., Strauch, M., Araújo, R.J.P., Bastos, P.R.F. (2005), “Sistema de Auxílio à Identificação de Perdas Comerciais” In: III Congresso de Inovação Tecnológica em Energia Elétrica, Florianópolis. Anais do III CITENEL.

JFreeChart: Disponível em: <<http://www.jfree.org/jfreechart/>>. Acesso em: out/2009.

J. R. Filho, A. C. Delaiba, E. Mazina, J. E. Cabral e J. O. P. Pinto (2004), “Fraud Identification in Electricity Company Customers Using Decision Tree”, In IEEE International Conference on Systems, Man and Cybernetics on Publication, pp. 3730- 3734, vol.4.

JXL: Disponível em: <<http://jexcelapi.sourceforge.net>>. Acesso em: out/ 2009.

- Lavrac, N., Flach, P. e Zupan B. (1999) "Rule Evaluation Measures: a unifying view", In: Proceedings of the Ninth International Workshop on Inductive Logic Programming, v.1634, pp. 74-185.
- M. G. Mendonça Neto (2001), "Mineração de Dados", In: Escola Regional de Informática da SBC Regional de São Paulo, 6, 2001, São Carlos. Anais: Minicursos - Coletâneas e textos. São Carlos: USP-ICMC, p.189-218.
- M. R. M. Rocha (2008), "Uso de Medidas de Desempenho e de Grau de Interesse para Análise de Regras Descobertas nos Classificadores", Dissertação de Mestrado, Programa de Pós-graduação em Engenharia Elétrica, Universidade Presbiteriana Mackenzie, São Paulo.
- Rocha, M. R. M. (2008) "Uso de Medidas de Desempenho e de Grau de Interesse para Análise de Regras Descobertas nos Classificadores", Dissertação de Mestrado, Universidade Presbiteriana Mackenzie, Programa de Pós-graduação em Engenharia Elétrica, São Paulo.
- Santana, C.C. (2004) "Uma Ferramenta para Construção Interativa de Árvores de Classificação Usando Mapas em Árvore", Dissertação de Mestrado, Programa de Pós-graduação em Redes de Computadores, Universidade Salvador - UNIFACS, Salvador.
- Silva, M. P. S (2008) "Mineração de Dados - Conceitos, Aplicações e Experimentos com Weka", Disponível em: [www.sbc.org.br/bibliotecadigital/download.php?paper=35](http://www.sbc.org.br/bibliotecadigital/download.php?paper=35), Acesso em Outubro.
- Souza, A.N., Silva, P.S., Oltremari, A.M., Zago, M.G., Amaral, F.L., Costa Jr., P., Costa Jr., B.C., Martins, L.C., Moreira, T.D. "Desenvolvimento de um Sistema Especialista para Detecção de Pontos Potenciais de Perdas Comerciais".
- S. R. R Passini (2008) "Detecção de Fraudes em Ligações de Água, uma experiência utilizando a tecnologia de Mineração de Dados", Disponível em: [http://www.semasa.sp.gov.br/Documentos/ASSEMAE/Trab\\_35](http://www.semasa.sp.gov.br/Documentos/ASSEMAE/Trab_35). Pdf. Acesso em Outubro.
- Todesco, J. L., A. B. T. Morales, S. Rautenberg, L. A. Garbelotto e E. D. Atahyde (2005) "Aplicação de Técnicas de Mineração de Dados para Detecção de Fraudes de Energia", Laboratório de Engenharia do Conhecimento, Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina.
- Weka 3: Datamining Software in Java, In: <http://www.cs.waikato.ac.nz/ml/weka>.