



## XVIII Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2008 - 06 a 10 de outubro

Olinda - Pernambuco - Brasil

Clusterização e Árvores de Decisão:  
estudo de caso COELBA segmento hospitalar

<b>Sanderson Moreno Ribeiro de Oliveira</b>
<b>COELBA</b>
Sanderson Oliveira – soliveira@coelba.com.br

### Palavras-chave

Árvores de Decisão  
Clusterização  
Hospitais  
Mineração de Dados  
Modalidade tarifária

### Resumo

A resolução 456/00/ANEEL – Agência Nacional de Energia Elétrica estabelece que os clientes de alta tensão devam formalizar um contrato de fornecimento de energia elétrica com as concessionárias, definindo parâmetros como Modalidade tarifária<sup>1</sup>, demanda(s)<sup>2</sup>, entre outros. O atendimento dos clientes corporativos da Companhia de Eletricidade do Estado da Bahia - COELBA tem o objetivo de identificar de forma sistemática os clientes deste segmento com problemas na contratação destes parâmetros. Os clientes com atividade econômica hospitalar apresentam um índice de inadimplência elevado comparando com as demais atividades. Para mitigar estes problemas, pode-se utilizar a tecnologia de Mineração de Dados, este trabalho aplica os métodos de Clusterização e Árvores de Decisão com o objetivo de agrupar estes Hospitais em classes e identificar os fatores que influenciam no custo elevado das faturas de energia elétrica. No segmento hospitalar a relação entre o custo elevado com energia elétrica e a inadimplência foi verificada. Os resultados encontrados podem ser aplicados na definição de estratégias de atendimento a este segmento de clientes.

### 1. Introdução

Este trabalho tem por finalidade apresentar os resultados de uma pesquisa realizada na COELBA, onde se procurou através de um estudo de caso verificar a eficácia de tecnologias como Mineração de Dados aplicando técnicas de Clusterização e Árvores de Decisão com o objetivo de otimizar as faturas de energia elétrica e reduzir a inadimplência dos Hospitais.

<sup>1</sup> Conjunto de tarifas aplicáveis às componentes de consumo de energia elétrica e/ou demanda de potência ativas de acordo com a modalidade de fornecimento.

<sup>2</sup> Média das potências elétricas ativas ou reativas, solicitadas ao sistema elétrico pela parcela da carga instalada em operação na unidade consumidora, durante um intervalo de tempo especificado.

Reduzir a inadimplência é um dos principais desafios das distribuidoras de energia elétrica, constituindo-se num dos seus objetivos empresariais. Para manter a inadimplência em índices aceitáveis, essas empresas investem esforços e recursos consideráveis em ações de cobrança. A suspensão do fornecimento de energia elétrica por falta de pagamento, o popular corte de energia, apesar de prevista na legislação do setor, é um dos últimos recursos utilizados pelas empresas para reaver seus créditos. Além de cessar a continuidade do fornecimento de energia (a venda do produto), o corte de energia gera ainda desgaste na imagem do cliente junto à concessionária.

A aplicação da suspensão do fornecimento de energia em alguns segmentos de clientes, a exemplo do hospitalar, pelo seu caráter social e humanitário é ainda mais complexa. Para efetivação do corte do fornecimento a uma unidade hospitalar, é necessário seguir um tramite legal, com comunicação prévia aos Poderes Públicos e a Justiça. Toda essa dificuldade se justifica pela privação social do fechamento de uma unidade hospitalar, especialmente as unidades hospitalares públicas e filantrópicas.

Apesar da lei determinar a obrigatoriedade do uso de geradores em unidades hospitalares como *back up*, o que impediria a interrupção completa do seu funcionamento, o corte de energia a um hospital pode acarretar prejuízos trágicos à comunidade e a própria imagem da concessionária.

Para evitar esses transtornos e garantir o recebimento dos seus créditos nesse segmento de cliente, a COELBA tem investido na negociação, através de ofertas de soluções para otimização das faturas de energia elétrica, buscando adequá-la à capacidade de pagamento dos devedores. Essa oferta tem como ponto de partida a simulação dos parâmetros do contrato de fornecimento de energia elétrica, compreendendo os seguintes aspectos: Modalidade tarifária (convencional, horosazonal verde e horosazonal azul); demanda(s) contratada(s); opção de energia mais<sup>3</sup>.

Existe limitação operacional do departamento de clientes corporativos para fazer a simulação manual, visando a otimização da fatura de energia de todos os Hospitais atendidos em alta tensão e a seleção daqueles que proporcionem os maiores retornos e menores riscos de inadimplência para a COELBA.

Para melhorar o gerenciamento dos dados dos clientes criou-se a necessidade de armazenar as informações relativas aos clientes corporativos da organização em Sistemas Gerenciadores de Banco de Dados (SGBD). Esta iniciativa por muitos anos foi suficiente para se ter o controle e o acesso aos dados dos grandes clientes, permitindo ao atendimento tomar decisões referentes a diversos processos. Todavia, estas bases de dados se tornaram imensas, dificultando a análise consistente dos seus dados. Essa situação contribuiu para o surgimento de novas técnicas que facilitassem o acesso e a análise desses dados, com o objetivo de auxiliar no suporte para uma tomada de decisão precisa, rápida e segura. Uma dessas novas técnicas de análise de dados é a Mineração de Dados.

O *Data Mining* também conhecido como Mineração de Dados, pode ser definido como o processo de selecionar, explorar e modelar grandes quantidades de dados em busca de padrões não triviais que tragam alguma vantagem competitiva (Cabena, 1997).

A metodologia de descoberta de conhecimento em base de dados fundamentada nos princípios de planejamento de atividades proposta por Goldschmidt e Passos (2005, p. 150-158) é utilizada neste projeto estabelecendo um conjunto de regras e tarefas a serem seguidos para produzir resultados satisfatórios.

---

<sup>3</sup> Oferta de energia elétrica excedente aos clientes horosazonais.

Ao final das tarefas de Clusterização e Árvores de Decisão foram obtidas três classes de Hospitais de acordo com o consumo e o custo com energia elétrica, regras de produção e estatísticas evidenciando alguns fatores que influenciam no valor da fatura de energia elétrica. Também se verificou a relação entre as classes de Hospitais e a inadimplência. Este conhecimento pode ser utilizado para otimizar as faturas de energia elétrica dos Hospitais, priorizando os com maiores índices de inadimplência e custos elevados com energia elétrica.

O capítulo dois traz informações sobre o projeto, as técnicas e as tecnologias de Mineração de Dados aplicados no trabalho. É apresentado, também, a avaliação dos resultados das tarefas de Clusterização e Árvores de Decisão.

O trabalho é finalizado pelo capítulo três, onde são feitas as considerações finais sobre a conclusão obtida pela análise dos resultados da aplicação.

## **2. O Projeto**

O segmento hospitalar possui um índice de inadimplência de 32 %, visando melhorar este índice o departamento de clientes corporativos iniciou em julho/07 uma série de ações incluindo: simulação de Modalidade tarifária, oferta de energia mais, orientações de uso eficiente de energia elétrica, além de projetos de eficiência energética aprovados pela ANEEL, entre outras.

A grande quantidade de dados para análise simultânea e a limitação operacional do departamento para simular manualmente os dados disponíveis no sistema comercial, criou a necessidade de buscar técnicas para obter conhecimento em banco de dados de forma sistemática e automática, com o objetivo de utilizar o resultado deste trabalho na definição de estratégias para melhorar seus indicadores empresariais. A tecnologia de Mineração de Dados é utilizada, aplicando os métodos de Clusterização e Árvores de Decisão.

A técnica de Clusterização é utilizada com o objetivo agrupar os Hospitais em classes de acordo com o custo e o consumo de energia elétrica. Posteriormente, utilizando o resultado da tarefa de Clusterização, foi aplicado o método de indução de regras de produção por Árvores de Decisão com o intuito de identificar os fatores que influenciam no custo elevado das faturas.

Na tarefa de Clusterização é utilizado o algoritmo K-MEANS implementado no banco de dados *SQL SERVER*<sup>4</sup> da *Microsoft*. Para a tarefa de indução de regras por Árvores de Decisão é utilizado o *software Decision Tree Learning* versão 4.3.1 desenvolvido pelo grupo de Inteligência Artificial<sup>5</sup> da universidade de *British Columbia* dos Estados Unidos.

Nas próximas seções são apresentadas informações sobre as técnicas de Clusterização e Árvores de Decisão, além de apresentar os resultados obtidos após a aplicação destas técnicas na base de dados da COELBA do segmento hospitalar.

---

<sup>4</sup> MS SQL Server é um SGBD - Sistema Gerenciador de Banco de Dados relacional criado pela Microsoft.

<sup>5</sup> É uma área de pesquisa da ciência da computação dedicada a buscar métodos ou dispositivos computacionais que possuam ou simulem a capacidade humana de resolver problemas, pensar ou, de forma ampla, ser inteligente.

## 2.1 Clusterização

Esta técnica consiste em classificar objetos em grupos diferentes. Clusterização é a classificação não-supervisionada de dados<sup>6</sup>, formando agrupamentos ou clusters.

Um exemplo de algoritmo de Clusterização é o K-MEANS que forma clusters no domínio numérico, particionando registros em grupos disjuntos. O K-MEANS foi utilizado neste trabalho.

### 2.1.1 Algoritmo K-MEANS

É um método de agrupamento que classifica um conjunto de dados em subconjuntos ou cluster, a partir da similaridade do valor da média dos atributos numéricos.

De acordo com Goldschmidt e Passos (2005), o algoritmo mais popular para realização da tarefa de agrupamento é o K-MEANS proposto em 1967 por J. MacQueen ele utiliza o método de partição.

O algoritmo K-MEANS faz uso de dados numéricos, selecionando k elementos para formação inicial dos centróides (elementos centrais) do grupo.

Após a seleção dos centróides, é calculada a distância de cada elemento (registro da base) em relação aos centróides, sendo considerada a menor distância encontrada para efetuar o agrupamento. O processo termina somente quando todos os elementos estejam agrupados. A figura 1 apresenta um resumo do funcionamento do algoritmo K-MEANS.

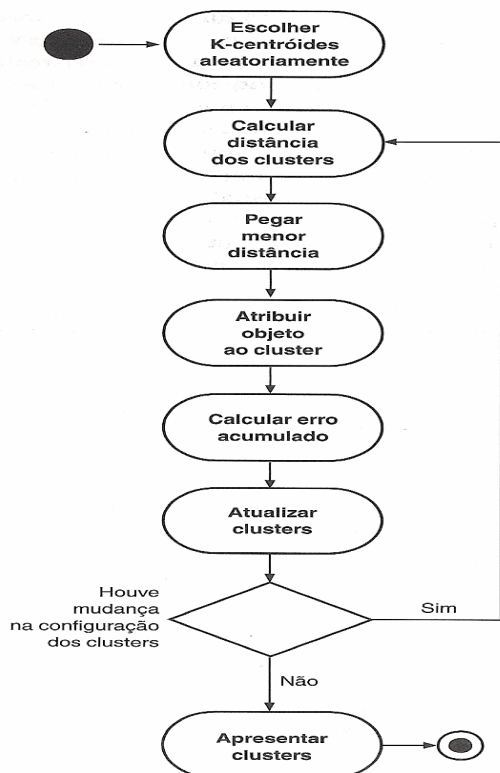


FIGURA 1: Algoritmo K-MEANS (Goldschmidt & Passos, 2005, p. 104).

<sup>6</sup> Extrai as propriedades estatísticas do conjunto de dados e agrupa os vetores similares em classes.

## 2.2 Árvores de Decisão

É uma forma de representar resultados a partir de um modelo escolhido fazendo-se associações de objetos através de um *software*, com essa técnica do *Data Mining* o usuário poderá se aprofundar nas suas pesquisas abrindo cada nó da árvore nas suas diversas ramificações.

As Árvores de Decisão utilizam a estratégia de dividir para conquistar: um problema complexo é decomposto em sub-problemas mais simples e recursivamente esta técnica é aplicada a cada sub-problema (Gama, 2004).

Só após ser analisada por especialista, a Árvores de Decisão poderá ser convertida em regras para formar a base de conhecimento de um sistema. As regras são formadas pelo caminho partindo do nó pai até nó filho.

As vantagens principais das Árvores de Decisão são: o método considera as regras que são mais relevantes, além de serem compreensíveis para a maioria das pessoas. Ao escolher e apresentar as regras em ordem de importância, as Árvores de Decisão permitem aos usuários perceber com clareza, quais fatores que mais influenciam os seus trabalhos.

## 2.3 Avaliação dos resultados da tarefa de Clusterização

O resultado da tarefa de Clusterização é simplificado através de um gráfico de dispersão demonstrado na figura 2. O eixo X do gráfico é representado pelo o consumo de energia elétrica (MWh) e o eixo Y pelo custo unitário com energia elétrica (R\$/MWh).

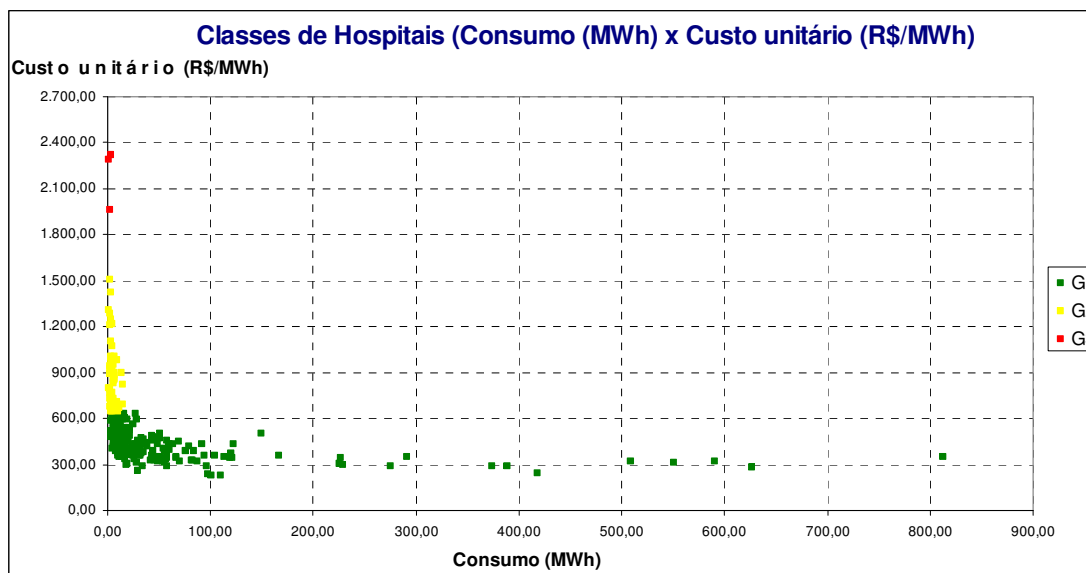


FIGURA 2 – Gráfico das classes encontradas ao final da Clusterização.

No gráfico, os valores foram divididos em três clusters denominados: cluster G1, cluster G2 e cluster G3.

Cada ponto X (MWh) e Y (R\$/MWh) ao final da Clusterização é associado a um cluster, de acordo com a similaridade do ponto com os centróides.

Os centróides são demonstrados na tabela 1.

TABELA 1 – Centróides encontrados após a Clusterização

Centróides		
Classe	MWh	R\$/ MWh
G1	54,68	436,09
G2	5,10	852,01
G3	2,36	2.189,67

Os valores dos clusters são: G1 – (230 a 640), G2 – (640 a 1.960) e G3 – (> 1.960) R\$/MWh. Os clusters foram denominados classes de custo unitário.

Novos clientes da COELBA do segmento hospitalar poderão ser classificados em uma dessas classes. Por exemplo, um hospital com custo unitário de 850 R\$/MWh pertence a classe – G2.

A tarefa de Clusterização despertou nos especialistas do domínio da aplicação, a necessidade de buscar novos conhecimentos, sendo levantadas as seguintes perguntas: “Quais são as características em comum dessas classes?” e “Qual é a relação dessas classes com a inadimplência dos Hospitais?”.

As características das classes serão analisadas com a aplicação de indução de Árvores de Decisão.

A relação entre as classes de custo unitário e a inadimplência, foi realizada através de instruções SQL<sup>7</sup> - Structured Query Language de agregação de dados.

Foram considerados inadimplentes os Hospitais com mais de uma fatura pendente de pagamento. O resultado é demonstrado na tabela 2.

TABELA 2 – Relação das classes de consumo unitário (R\$/MWh) e a inadimplência.

**Custo Unitário (R\$/MWh) x Inadimplência**

Grupo Custo Unitário	Custo Unitário (R\$/MWh)	Qtd Contratos	Qtd Inadimplentes	(%)
G1	230 a 640	213	61	28,64%
G2	640 a 1.960	64	26	40,63%
G3	acima 1.960	3	2	66,67%
<b>Total</b>		<b>280</b>	<b>89</b>	<b>31,79%</b>

A pesquisa evidenciou que quanto maior o custo unitário da classe, maior é o percentual de inadimplência. Diante deste cenário, os especialistas identificaram a necessidade de estabelecer estratégias de atendimento que levem em consideração estes aspectos, tentando adequar a fatura de energia elétrica a capacidade de pagamento do cliente.

<sup>7</sup> Linguagem de Consulta Estruturada é uma linguagem de pesquisa declarativa para banco de dados relacional (base de dados relacional).

## 2.4 Avaliação dos resultados de indução por Árvores de Decisão

A expectativa estabelecida na fase de definição de objetivos, era obter um modelo de conhecimento transparente para ser utilizado no atendimento dos clientes do segmento hospitalar, podendo assim fazer uma previsão dos custos com energia elétrica.

Para possibilitar a análise destas características, os especialistas recomendaram a utilização de indicadores extraídos das faturas de energia elétrica, analisados em projetos de eficiência energética.

Os atributos selecionados como parâmetros para indução de Árvores de Decisão são: Modalidade tarifária, FC<sup>8</sup> (Fator de Carga), IDC (Índice da Demanda Contratada) e a classe de custo unitário.

Para obtenção de regras de produção mais concisas, é necessária a redução de valores através de codificação dos valores numéricos em categorias. Os domínios dos atributos utilizados na tarefa são:

- Modalidade tarifária - GRA (convencional), HSV (horosazonal verde) ou HSA (horosazonal azul);
- FC: “0 a 0,25”, “0,26 a 0,46” ou “>= 0,47”;
- IDC: “SIM” ou “NÃO”;
- Classes de custo unitário: G1 (230 a 640), G2 (640 a 1.960) e G3 (> 1.960) R\$/MWh.

O resultado obtido é demonstrado na figura 3. A Árvore de Decisão obtida permite a representação do conhecimento através de regras de produção e a extração de estatísticas do relacionamento entre os dados.

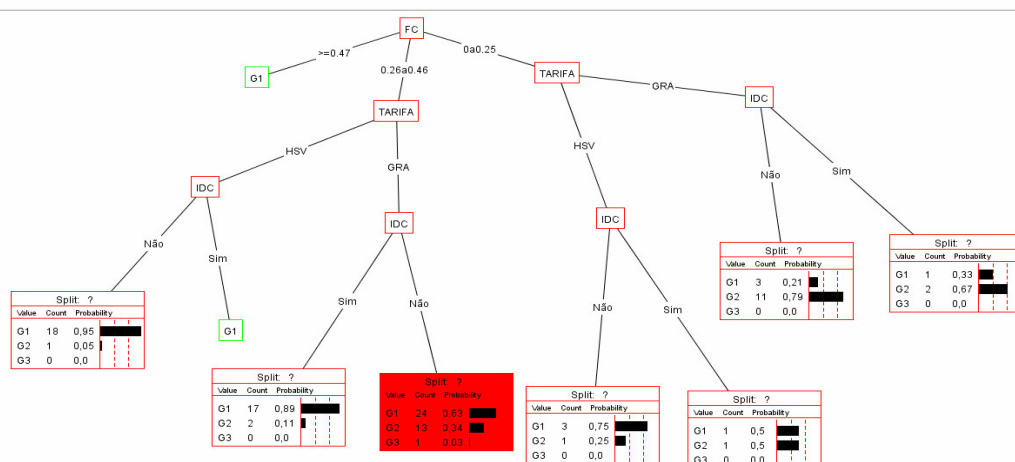


FIGURA 3 – Resultado do pós-processamento do *Decision Tree*

As regras de produção encontradas são apresentadas a seguir:

REGRA 1:

SE (FC “ $\geq 0,47$ ”) ENTÃO

Classe de custo unitário = G1

FIM SE

<sup>8</sup> Razão entre a demanda média e a demanda máxima da unidade consumidora, ocorridas no mesmo intervalo de tempo especificado.

REGRA 2:

SE (FC = “0,26 a 0,46”) E (Tarifa = “HSV”) E (IDC = “SIM”) ENTÃO

Classe de custo unitário = G1

FIM SE

A análise da REGRA 1, apontou que os Hospitais com fator de carga igual ou superior a 0,47 pertencem à classe de custo unitário G1.

O fator de carga é um indicador preponderante para determinação das classes de custo unitário, confirmando a hipótese identificada no mapeamento do conhecimento prévio: “Quanto mais próximo o valor do fator de carga de 1 mais racional é a utilização de energia elétrica pelo cliente”.

A REGRA 2, indicou que outros fatores como: Modalidade tarifária e IDC podem prever os custos com energia elétrica. Por exemplo, os Hospitais com FC entre “0,26 a 0,46”, Modalidade tarifária igual à “HSV” e o IDC igual a “SIM” também pertencem à classe G1.

As estatísticas relevantes extraídas da Árvore de Decisão são listadas a seguir:

ESTATÍSTICA 1:

SE (FC = “0,26 a 0,46”) E (TARIFA = “GRA”) E (IDC = “SIM”) ENTÃO

Têm 89% de probabilidade de pertencer a classe de unitário G2

FIM SE

ESTATÍSTICA 2:

SE (FC = “0,26 a 0,46”) E (TARIFA = “GRA”) E (IDC = “NÃO”) ENTÃO

Têm 63% de probabilidade de pertencer a classe de unitário G2

FIM SE

ESTATÍSTICA 3:

SE (FC = “0 a 0,25”) E (TARIFA = “GRA”) E (IDC = “NÃO”) ENTÃO

Têm 79% de probabilidade de pertencer a classe de unitário G2

FIM SE

ESTATÍSTICA 4:

SE (FC = 0 a 0,25) E (TARIFA = “GRA”) E (IDC = “SIM”) ENTÃO

Tem 67% de probabilidade de pertencer a classe de unitário G2

FIM SE

Algumas estatísticas relevantes foram identificadas pelos especialistas. O relacionamento de determinadas características indicam a predominância de um grupo sobre os outros. Por exemplo, a ESTATÍSTICA 1 revelou que os Hospitais com fator de carga entre “0 a 0,25”, Modalidade tarifária “GRA” e o IDC igual a “NÃO” têm 79% de probabilidade de pertencer a classe de custo unitário G2.

O conhecimento obtido através das regras de produção e estatísticas pode auxiliar os especialistas na definição de estratégias de atendimento aos clientes do segmento hospitalar.

Neste projeto a relação existente entre inadimplência e o custo unitário elevado foi constatada. Por exemplo, os Hospitais que pertencem à classe G2 possuem uma inadimplência de 40,63%. A



ESTATÍSTICA 1 revelou que Hospitais com FC entre “0 a 0,25”, Modalidade tarifária “GRA” e o IDC igual a “NÃO” têm 79% de probabilidade de pertencer à classe de custo unitário G2. Relacionando os resultados obtidos na Clusterização e Árvores de Decisão, constatou-se que o conhecimento pode ser utilizado para priorizar as ações de atendimento sobre este grupo de Hospitais, priorizando os inadimplentes e custo unitário elevado. Nestes casos, algumas ações informadas pelos especialistas são: simulação de Modalidade tarifária e demanda(s) contratada(s); orientação aos clientes sobre o uso eficiente de energia elétrica; adequação custo unitário a capacidade de pagamento através da melhoria dos indicadores.

A seguir são apresentadas algumas recomendações dos especialistas para melhoria dos indicadores:

### MELHORIA DO FC

O FC (Fator de Carga) é obtido através do cálculo:

$$FC = kWh / (kW * 730 h)$$

onde

kWh é o consumo ativo lido no ciclo de faturamento,

kW é a demanda ativa lida no ciclo de faturamento.

Segundo a distribuidora de energia elétrica AES Eletropaulo, dentre as práticas que merecem mais atenção no gerenciamento de contas de energia elétrica está à melhoria do FC. Pode ser resumida em conservar o consumo (kWh) e reduzir a demanda (kW) ou aumentar o consumo e manter a demanda.

### MELHORIA DO IDC

O IDC (Índice da Demanda Contratada) é calculado da seguinte forma:

$$IDC = MDLD / DCTR$$

onde

MDLD é a Maior Demanda Lida nos últimos 12 (doze) ciclos de faturamento,

DCTR é a Demanda Contratada nos últimos 12 (doze) ciclos de faturamento.

Os especialistas destacaram que a melhoria do IDC pode ser obtida através da análise da série histórica da demanda contratada e a demanda lida. Um exemplo é ilustrado na tabela 3.

TABELA 3 – Exemplo de série histórica demanda contratada (DCTR) x demanda lida (DLD).

Demanda Contratada				Série	DCTR	DLD	IDC (%)
DCTR	MDLD	IDC	IDC				
65	44,01	0,68	NÃO	1	65,00	37,45	57,62%
				2	65,00	25,85	39,77%
				3	65,00	14,63	22,51%
				4	65,00	0,00	0,00%
				5	65,00	28,47	43,80%
				6	65,00	27,36	42,09%
				7	65,00	44,01	67,71%
				8	65,00	41,52	63,88%
				9	65,00	30,18	46,43%
				10	65,00	26,43	40,66%

A legislação do setor elétrico através da Res. 456/00/ANEEL determina que o cliente do Grupo A contrate uma demanda para fins de faturamento mensal. O maior valor entre a demanda contratada (DCTR) e a demanda lida (DLD) é utilizado para fins de faturamento. Por exemplo, a DLD = 26,43 kW e a DCTR = 65,00 kW, a demanda utilizada para fins de faturamento é 65,00 kW.

A regra estabelece ainda uma penalidade para ultrapassagem da demanda lida pela contratada observando uma tolerância de 5 ou 10% sobre o valor da demanda contratada. Nos casos de ultrapassagem é aplicada a parcela excedente uma tarifa três vezes maior que a normal. Por exemplo, a DLD = 90 kW e DCTR = 65,00 kW, é aplicada a tarifa normal até o valor da demanda contratada, e uma tarifa três vezes o valor excedido de 25 kW.

O valor da demanda contratada pode onerar o custo com energia elétrica. No exemplo ilustrado na tabela 7 os valores da demanda lida estão bem abaixo da contratada.

O valor do IDC apurado neste exemplo, é igual a 67,71%. Os especialistas do domínio da aplicação definiram que os valores entre 90% e 110% são classificados como “SIM”, porque em geral os valores dentro deste intervalo estão bem contratados. Os demais valores são classificados como “NÃO”.

Os especialistas fizeram uma ressalva. A avaliação do IDC pode ser prejudicada por clientes com consumo de energia elétrica, apresenta grandes variações entre o mínimo e o máximo durante um período. Porém, a eficácia do indicador foi demonstrada ao sinalizar clientes com problemas na demanda contratada.

## MODALIDADE TARIFÁRIA

O atributo de Modalidade tarifária determina os valores das tarifas aplicadas a cada conceito que compõe a fatura de energia elétrica como: consumo de energia elétrica (kWh) e a(s) demanda(s) (kW).

Os especialistas informaram que a opção pelo cliente de uma Modalidade tarifária depende de uma série de aspectos como: carga instalada, nível de tensão, características de consumo de energia elétrica em horários durante o dia, entre outros.

A legislação do setor elétrico nas condições gerais de fornecimento de energia elétrica estabelece que a concessionária a exemplo da COELBA deve prestar por escrito informações necessárias e adequadas a cada caso, cabendo o consumidor formular sua opção também por escrito.

Os especialistas realizaram a simulação de outras opções de Modalidade tarifária para um cliente da classe de custo unitário G2, o resultado é demonstrado na tabela 4.

TABELA 4 – Simulação de opções de Modalidade tarifária.

### Simulações de Estrutura Tarifária

Tarifa	Demanda(s) Proposta (kW)	Importe R\$	MWh	Custo Unitário (R\$/MWh)	(%)	
Convencional (GRA)	30	2.166	2,47	877		<b>Opção Atual</b>
Horo Verde (HSV)	30	1.175	2,47	476	-45,78%	
Horo Azul (HSA)	30	2.816	2,47	1.140	29,99%	
Grupo B	-	1.028	2,47	416	-52,55%	

A simulação realizada demonstra que mantendo as características de consumo atuais e fazendo opção pela Modalidade tarifária (tarifa) de “Grupo B” ou “HSV”, o cliente pode obter reduções com o custo unitário com energia elétrica maiores que 45%.

É importante ressaltar que os atributos selecionados neste estudo de caso não esgotam os aspectos que determinam as características das classes de custo unitário.

Os especialistas do domínio da aplicação listaram outra série de aspectos que podem contribuir para determinar as características das classes de custo unitário como: fator de potência abaixo do estabelecido pela legislação do setor elétrico; pagamento de juros e multas por atraso no pagamento das faturas; estado de conservação, características técnicas e a utilização dos equipamentos de energia elétrica, entre outras.

Estas características são avaliadas em projetos de eficiência energética, sendo assim o conhecimento obtido pode auxiliar na identificação de clientes para participação nestes projetos, que são aprovados e têm verba assegurada pelo governo federal de acordo com um planejamento e limites de verbas estabelecidos anualmente.

### **3. Conclusões**

A descoberta de conhecimento em base de dados busca desenvolver ferramentas para auxiliar o homem, de forma automática e inteligente, auxiliando no processo de tomada de decisão e definição de estratégias em diversas áreas (Goldschmidt & Passos, 2005).

A aplicação da tarefa de Mineração de Dados composta pelas técnicas de Clusterização e indução de Árvores de Decisão, contribui para a descoberta de conhecimentos novos e não triviais, além da confirmação de conhecimentos prévios identificados no levantamento inicial do projeto.

A relação entre as classes de custo unitário encontradas na tarefa de Clusterização e a inadimplência foi evidenciada, constatando que quanto maior o custo com energia elétrica maior é a dificuldade de pagamentos dos Hospitais, os índices de inadimplência das classes com maior custo são superiores a 40%. O estudo também possibilitou a confirmação da relação existente entre o uso racional de energia elétrica e o fator de carga (FC), sendo identificado que para pertencer à classe com o menor custo unitário encontrado, os Hospitais necessitam ter o FC acima de 0,47. Estes são exemplos de conhecimentos obtidos ao final da tarefa de Mineração de Dados composta.

As regras de produção e estatísticas geradas ao final da tarefa de indução de Árvores de Decisão evidenciaram alguns fatores que influenciam no custo com energia elétrica são eles: FC, Modalidade tarifária e demanda(s) contratada(s). O mapeamento de todas as causas do custo elevado com energia elétrica pode ser realizado pelos especialistas com a utilização de uma série de técnicas da área de engenharia elétrica aplicados em projetos de eficiência energética.

Os resultados obtidos podem ser utilizados para definir uma metodologia de classificação dos Hospitais, com base nos custos e consumo de energia elétrica, além de permitir análise e identificação de forma sistemática dos fatores que influenciam no custo unitário com energia elétrica através de indicadores e regras. O conhecimento obtido também pode ser aplicado na definição de estratégias de atendimento priorizando as classes de clientes com custos elevados com energia elétrica e os maiores

índices de inadimplência, além de permitir a seleção de clientes através da análise dos seus indicadores para participação em projetos de eficiência energética.

#### **4. Referências bibliográficas e/ou bibliografia**

AES Eletropaulo. Fator de carga. Acesso em 14/10/2007, disponível em: [http://www.eletropaulo.com.br/portal/page.cfm?conteudo\\_id=713&desc=fator%20de%20carga&origem\\_id=709](http://www.eletropaulo.com.br/portal/page.cfm?conteudo_id=713&desc=fator%20de%20carga&origem_id=709)

CABENA, Peter; HADJINIAN, Pablo; STANDLER, Rolf; VERHEES, Jaap; ZANASI, Alessandro. Discovering Data Mining from Concept Implementation. New Jersey, Ed. Prentice Hall, 1997.

Condições gerais de fornecimento de energia elétrica, Res. 456/00/ANEEL – Agência Nacional de Energia Elétrica, 29 de novembro de 2000. Acesso em 14/08/2007, disponível em: <http://www.aneel.gov.br/>

FAYYAD, Usama; GRINSTEIN, Geoges G; WIERSE, Andreas. Information Visualizing in Data mining and Knowledge Discovery. Oxford, Ed. Hardcover, 2001

GAMA, João. Árvores de Decisão. Acesso em 20/09/2007, disponível em: [www.liacc.up.pt/~jgama/Aulas\\_ECD/arv.pdf](http://www.liacc.up.pt/~jgama/Aulas_ECD/arv.pdf)

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. Data Mining – Um guia prático. Rio de Janeiro, Ed. Campus, 2005.