Modelo para Identificação e Prevenção de Inadimplência para o Mercado de Grandes Clientes

Juan G. Lazo Lazo[1], Karla Figueiredo[2], Marley Maria B. R. Vellasco[1], Marco Aurélio C. Pacheco[1], , Carlos Hall[1], Ricardo Tanscheit[1], João R. Carrilho Jr.[1], José Eduardo N. Da Rocha[3]

[1]*ICA: Applied Computational Intelligence Laboratory*
Department of Electrical Engineering
Pontifical Catholic University of Rio de Janeiro, PUC-Rio
Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ Brasil

[2]*Departamento de Engenharia Eletrônica/Telecomunicações, UERJ*
Rua São Francisco Xavier, 524, Maracanã - CEP 20550-900, Rio de Janeiro RJ – Brasil

[3]*Light Serviços de Eletricidade S.A.*
Av. Marechal Floriano 168/ 1° Andar Centro Rio de Janeiro, CEP 20080-002, RJ Brasil

juan@ele.puc-rio.br, karlaf@uerj.br, marley@ele.puc-rio.br, marco@ele.puc-rio.br, hall@ele.puc-rio.br, ricardo@ele.puc-rio.br, joaocarrilho@globo.com, jose.rocha@light.com.br

Resumo

Este trabalho apresenta o desenvolvimento de uma metodologia capaz de identificar a possível inadimplência dos clientes da empresa Light S.A., utilizando técnicas de Mineração de Dados (*Data Mining*) a partir da análise das informações cadastrais, de freqüência de inadimplência, tempo que é cliente, faturamento de energia elétrica, e outras informações exógenas presentes na base de dados.

A metodologia é baseada em duas etapas: categorização ou agrupamento para agrupar unidades consumidoras com padrões de atividades econômicas similares; e classificação para descobrir mudanças no perfil de comportamento dos clientes, configurando a possibilidade de inadimplência.

Os resultados obtidos com modelagem utilizando técnica de inteligência computacional foi considerado muito bom por especialistas responsáveis pela área de análise de faturamento de grandes clientes na empresa Light S.A.

1. Introdução

O cenário brasileiro, inserido no processo de globalização, e as constantes mudanças nas políticas econômicas nacionais, criam situações de dificuldades financeiras para vários segmentos comerciais, industriais e de serviços. Saber identificar corretamente os clientes com alto risco de inadimplência permite otimizar o investimento da concessionária na prevenção e recuperação das perdas comerciais com o direcionamento dos recursos para estes clientes.

Os diversos fatores que devem ser levados em conta no gerenciamento da inadimplência de uma carteira de clientes não são totalmente previsíveis, de tal forma que uma tomada de decisão neste contexto é sempre uma tarefa complexa. No sentido de se direcionar a prevenção e até mesmo a recuperação da receita, muitas técnicas computacionais de programação matemática têm sido aplicadas, todavia com resultados insatisfatórios.

Este artigo descreve o desenvolvimento de um sistema inteligente para classificar a carteira de clientes em *clusters* ou grupos de nível de risco de inadimplência, de forma que seja possível direcionar as ações de pré-cobrança, cobrança de débitos efetivos e adoção de medidas preventivas, minimizando os custos associados através do direcionamento dos recursos para os grupos com maior risco e assim reduzir as perdas.

Desta forma foram avaliadas metodologias da Inteligência Computacional para modelar estas características a partir dos fatores importantes,

considerando não somente fatores quantitativos mas também qualitativos para o desenvolvimento do sistema que irá grupar e classificar os clientes da concessionária Light S.A

O sistema inteligente integra uma metodologia que possui duas etapas: categorização (agrupamento) de unidades consumidoras em grupos de mesmo perfil de faturamento e consumo de energia elétrica, e classificação para identificar quais unidades consumidoras mudaram de perfil a cada nova medição realizada pela empresa. Esta metodologia considerou as características históricas e sazonais de cada unidade consumidora.

O restante deste artigo está organizado da seguinte maneira: a seção 2 apresenta o problema de detecção de inadimplência. A seção 3 descreve a metodologia proposta, a seção 4 descreve os casos de estudo considerados, e finalmente, na seção 5, são apresentadas as conclusões do trabalho.

2. Problema da Detecção de Inadimplência

Inadimplência, a rigor, é uma situação de qualquer quebra de cláusula de contrato. No mercado financeiro, no entanto, muitas vezes o termo é usado como sinônimo de situação de atraso de dívida. O inadimplente, portanto, neste sentido que vamos usar, é aquele que deixou de pagar suas dívidas em dia.

Há costumeiramente no Brasil uma significativa influência do nível de atividade econômica no grau de inadimplência. O volume de títulos protestados, um dos principais indicadores de inadimplência, tem fortíssima associação com o comportamento do Produto Interno Bruto. Quando há crescimento menor do PIB, a inadimplência sobe; quando há crescimento maior, a inadimplência cai.

Atualmente, nenhuma empresa pode manter, em seu quadro financeiro, um desequilíbrio significativo e exorbitante de inadimplentes, pois isso coloca seriamente em risco sua sobrevivência no mercado. Tentando eliminar esse tipo de fator, este trabalho buscou, através de pesquisa exploratória, descrever como as organizações estão agindo com a inadimplência, no mercado global, definindo créditos, relacionando cuidados que deverão ser tomados para que os ganhos não sejam nulos e nem sejam gerados mais custos, reduzindo a margem líquida, e, em última instância, gerando prejuízos que comprometerão sua sobrevivência. Como negociadores, verificando quais as decisões que deverão ser tomadas pelos executivos para se precaverem de perdas futuras. Preocupação com a busca por iniciativas e alternativas que fizeram tais organizações mudar a maneira de administrar sua carteira de clientes no sentido da satisfação das partes que negociam, reduzindo significativamente a inadimplência [1].

A maioria das empresas sofre com os consumidores inadimplentes, sendo que este é um dos fatores que mais atingem os setores do comércio e de serviços, fazendo com que muitos empresários sintam-se desprotegidos. Mas o que leva os consumidores à inadimplência? Abaixo foram enumerados alguns dos principais fatores:

- Dificuldades financeiras, que impossibilitam o cumprimento de obrigações;
- falta de controle nos gastos;
- atraso no recebimento de faturas;
- comprometimento de renda com outras despesas;
- queda nas vendas e/ou serviços;
- uso do dinheiro com outras compras;
- investimento dos rendimentos;
- má fé.

Vale lembrar quer a ocorrência da maioria destes fatores é mais comum em tempos de crise. Para lidar com a inadimplência é preciso antes conhecê-la bem, sabendo quais os fatores que a ocasionaram. A partir daí, deve-se utilizar práticas preventivas, podendo desta forma controlá-la. Dessa maneira, o empresário deve verificar:

- sazonalidade, ou seja, em que épocas do ano a inadimplência é maior ou menor;
- número de contas em atraso;
- valor médio das contas em atraso;

Nas empresas dos setores do comércio e de serviços, a inadimplência costuma aumentar nos três primeiros meses do ano, em decorrência das vendas efetuadas no mês de dezembro. E é justamente nestes meses iniciais que também ocorre queda nas atividades destes setores empresariais, deixando assim muitos empresários em dificuldades. Assim, o empresário, ciente do aumento da inadimplência no período após as vendas natalinas, ou seja, janeiro/fevereiro/março, deverá ser mais rigoroso na concessão de financiamentos, parcelamentos e aceitação de cheques.

Os diversos fatores que devem ser levados em conta no gerenciamento da inadimplência de uma carteira de clientes não são totalmente previsíveis, de tal forma que uma tomada de decisão neste contexto é sempre uma tarefa complexa. No sentido de se direcionar a prevenção e até mesmo a recuperação da receita, muitas técnicas computacionais de programação matemática têm sido aplicadas, todavia com resultados insatisfatórios.

Estas inadimplências encontram-se em parte registradas e ao mesmo tempo ocultas no enorme volume de dados históricos dos clientes, depositados na base de dados da empresa Light S.A.. Para extrair da base de dados a identificação das unidades consumidoras que apresentam inadimplências, tem-se utilizado técnicas computacionais clássicas de

programação matemática e estatística, comumente aplicadas na análise da variação percentual de consumo, todavia com resultados pouco precisos. As técnicas de mineração de dados da inteligência computacional apresentam-se como uma alternativa viável para realizar uma análise abrangente de todas as classes econômicas da base de dados. Este trabalho propõe a categorização natural do espaço amostral de parte dos clientes da Light pelo algoritmo Kohonen [2], em função das características elétricas de consumo destas unidades nos meses de janeiro a dezembro, independentemente de suas atividades econômicas e considerando o comportamento sazonal das faturas, e a classificação do restante das atividades utilizando redes neurais com aprendizado do tipo back propagation.

3. Metodologia Proposta

A metodologia proposta foi construída de forma que o modelo gerado consegue-se capturar as características dos clientes por suas atividades econômicas e a influencia desta atividade econômica no pagamento das suas contas de energia elétrica. Por esse motivo foram usadas informações de índices econômicos de diferentes setores de atividades econômicas, além das informações cadastrais, de freqüência de inadimplência, tempo que é cliente, faturamento entre outras. Este foi um fator fundamental para se obter os bons resultados atingidos com o modelo proposto.

A modelagem foi feita a partir da divisão do problema em de duas etapas: agrupamento da base de dados histórica de faturamento dos clientes utilizando técnicas estatísticas e uma rede neural auto-organizável do tipo mapa de Kohonen e classificação dos clientes inadimplentes segundo estes grupos.

A categorização (*clustering* ou agrupamento) é uma das técnicas de mineração de dados (*data mining*) usada para descobrir padrões ocultos em uma base de dados.

A técnica de categorização particiona um conjunto de padrões de entrada em grupos (*clusters*) homogêneos [3], de tal forma que os dados de um mesmo *cluster* são mais similares entre si que os dados encontrados em diferentes *clusters* [4].

A similaridade destes pontos é definida de acordo com critérios pré-estabelecidos. Por exemplo, considerando os registros de um banco de dados de uma empresa que contém os itens comprados pelos seus clientes, o procedimento de categorização poderá grupar clientes de forma que aqueles com padrões similares de compras estejam no mesmo *cluster*. Logo, o principal objetivo do processo de categorização é revelar a organização dos padrões em grupos que nos permitam descobrir similaridades e diferenças, bem como permitir extrair conclusões sobre esta distribuição.

No processo de categorização não há classes pré-definidas, nem exemplos que possam mostrar que tipos de relações desejáveis devem estar válidos entre os dados. Por esta característica o algoritmo utilizado neste processo é dito não supervisionado [5].

A categorização produz categorias iniciais nas quais os valores dos padrões encontrados nos dados são classificados. Este processo pode resultar em diferentes particionamentos dos dados dependendo dos critérios específicos utilizados. Logo, existe a necessidade de pré-processamento antes de se iniciar a tarefa de categorização. A seção 3.1 a seguir descreve o pré-processamento.

3.1. Pré-processamento dos dados

A base de dados foi fornecida pela Light em um arquivo texto contendo as informações dos clientes. A **Tabela 1** apresenta a organização dos dados neste arquivo, o qual contem 35 atributos por linha, isto é, cada coluna representa um atributo.

A esta base de dados foi adicionada um a conjunto de índices econômicos que afetam as atividades econômicas dos cliente. A **Tabela 2** apresenta a relação de índices econômicos utilizados.

Faturas		
	Campo	**Descrição do Campo**
1.	etapa	Lote de Faturamento do Cliente
2.	orgao	Sigla do Órgão responsável pela documentação do Cliente
3.	tipoDocumento	Informa o tipo do Documento que identifica o Ciente
4.	numDocumento	Número do Documento do Cliente
5.	Uc	Unidade Consumidora
6.	cliente	Nome do Cliente
7.	endereco	Endereço
8.	numImovel	Número do Imóvel
9.	codLogradouro	Código do Logradouro
10.	trechoLogradouro	Trecho do Logradouro
11.	codLocalidade	Código da Localidade
12.	codMunicipio	Código do Município
13.	municipio	Nome do Município
14.	codClassePrincipal	Código da Classe Principal do Cliente
15.	codClasseConsumo	Código da Classe de Consumo
16.	classeConsumo	Nome da Classe de Consumo
17.	grupoTensao	Grupo de Tensão Faturamento
18.	subGrupoTensao	SubGrupo Tensão Faturamento
19.	codTipoTarifa	Código do Tipo da Tarifa
20.	codTipoArrecadacao	Código do Tipo de Arrecadação
21.	descTipoArrecadacao	Descrição do Tipo de Arrecadação
22.	codSituacao	Código de Situação da Fatura
23.	dataSituacao	Data da Situação
24.	statusReaviso	Indica se ocorreu reaviso
25.	numFatura	Número da Fatura
26.	dataReferencia	Data de Referência da Fatura
27.	dataCompetenciaFatura	Data de Competência da Fatura
28.	codOrigemFatura	Código de Origem da Fatura
29.	codSituacaoFatura	Código de Situação da Fatura
30.	dataSituacaoFatura	Data de Situação da Fatura
31.	dataApresentacao	Data de Apresentação
32.	dataVencimento	Data de Vencimento da Fatura
33.	dataPagamento	Data de Pagamento da Fatura
34.	dataBaixa	Data de Baixa
35.	valorEmissao	Valor da Fatura

Tabela 1 – Campos do arquivo txt de faturas.

	Campo	Descrição
1	Data	Data relativa ao mês di índice
2	Índice 1	Indústria Extrativa - RJ (2002=100)
3	Índice 2	Indústria Geral - RJ (2002=100)
4	Índice 3	Indústria de Transformação - RJ (2002=100)
5	Índice 4	Alimentos - RJ (2002=100)
6	Índice 5	Bebidas - RJ (2002=100)
7	Índice 6	Têxtil - RJ (2002=100)
8	Índice 7	Edição, Impres.e Reprod.de Grav. - RJ (2002=100)
9	Índice 8	Refino de Petróleo e Álcool - RJ (2002=100)
10	Índice 9	Farmacêutica - RJ (2002=100)
11	Índice 10	Perf.,Sabões,Deterg.e Prod.de Limpeza - RJ (2002=100)
12	Índice 11	Outros Produtos Químicos - RJ (2002=100)
13	Índice 12	Borracha e Plástico - RJ (2002=100)
14	Índice 13	Minerais Não Metálicos - RJ (2002=100)
15	Índice 14	Metalurgia Básica - RJ (2002=100)
16	Índice 15	Veículos Automotores - RJ (2002=100)
17	Índice 16	Taxa de Cambio (R$/US$) media mensal,compra
18	Índice 17	IBOVESPA - Média Mensal
19	Índice 18	Vestuário e Acessórios - sem Ajuste (2002=100)
20	Índice 19	Nominal Vendas no Varejo - RJ (03=100)
21	Índice 20	Nom. Vendas no Varejo (Comb.e Lubr.) - RJ (03=100)
22	Índice 21	Nom.Vendas Var.(Hip,Sup.,Prod.Alim.,Beb.Fumo)-RJ(03=100)
23	Índice 22	Nom. Vendas no Varejo (Tec.,Vest.e Calçados)-RJ(03=100)
24	Índice 23	Nom. Vendas no Varejo (Móveis e Eletrod.) - RJ (03=100)
25	Índice 24	Nom. Vendas no Varejo (Hiperm. e Superm.) - RJ (03=100)
26	Índice 25	Nom.Vendas Var.(Veíc.,Motos,Partes e Peças) - RJ(03=100)
27	Índice 26	Máqs.p/Escrit. e Equip. de Inform. - sem Ajuste (2002=100)
28	Índice 27	Máquinas e Equipamentos - sem Ajuste (2002=100)
29	Índice 28	Celulose, Papel e Prod. de Papel - sem Ajuste (2002=100)
30	Índice 29	Calçados e Artigos de Couro - sem Ajuste (2002=100)

Tabela 2 – Índices econômicos.

Numa primeira etapa foi realizada a transformação da base de dados original em uma base de dados que possua integridade relacional. Esta transformação foi feita a partir da análise da base de dados retirando-se os dados repetidos, ambíguos, outliers e variáveis com informações redundantes.

Como resultado deste processo obteve-se uma base de dados capaz de ser tratada pelos processos de agrupamento e classificação.

3.2. Agrupamento por rede neural auto-organizável do tipo mapa de Kohonen

As Redes Neurais cujo treinamento não requer o conhecimento dos valores de saída da rede são denominadas Redes Neurais não supervisionadas [2], [6] e [7] e são indicadas para problemas de agrupamento. As Redes Neurais não-supervisionadas são geralmente baseadas em competição, onde o neurônio vencedor tem seu vetor de pesos atualizado, sendo o mapa de Kohonen [2] o principal tipo de algoritmo utilizado.

Self-Organizing Maps (SOM) [2] são redes baseadas em mapas de Kohonen que podem aprender a detectar regularidades e correlações de suas entradas e adaptar futuras respostas a um novo padrão de entrada. Os neurônios de redes competitivas aprendem a reconhecer grupos similares de vetores de entrada, nos casos de SOM os neurônios estão fisicamente próximos.

Em SOM, os neurônios (*clusters*) são organizados normalmente em um grid bidimensional, podendo ser também unidimensional e raramente tri ou multi-dimensional. Dessa forma, um padrão de entrada de qualquer dimensão poderá ser visualizado como um mapa de uma ou duas dimensões.

A principal característica deste algoritmo é que não apenas o neurônio vencedor tem seu vetor de pesos atualizado, mas também os neurônios dentro de uma certa vizinhança (**Figura 1**). Esta vizinhança deve ser grande inicialmente e decrescer até zero durante o treinamento. Além deste parâmetro, a taxa de aprendizado que é usada na equação de atualização dos pesos também deve decrescer com o treinamento.

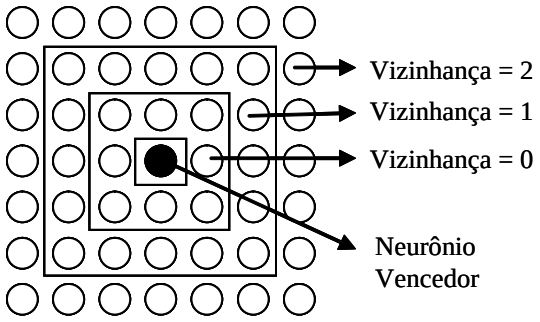

Figura 1 - Mapa com 7x7 neurônios com diferentes vizinhanças em torno do neurônio vencedor

O processo de aprendizado tem natureza estocástica, isto significa que a precisão do mapa depende do número de iterações do algoritmo, das taxas de aprendizado e de vizinhança. No entanto, não existe base teórica para seleção destes parâmetros, eles são determinados através de processos heurísticos.

Os pesos da Rede Neural são inicializados com a média dos valores da matriz de entrada.

Para determinar o neurônio vencedor foi utilizado o método de distância Euclidiana entre os padrões de entrada e os vetores de pesos, como pode se visto abaixo.

$$i(x) = \text{distância mínima} \parallel x(n) - w(j) \parallel \qquad \text{eq. 1}$$

Atualização dos pesos:

$$W(n+1)=\begin{cases} W(n)+\eta(n)[x(n)-w_j(n)]. & j\in\Lambda_{i(x)} \\ w_j(n), \text{ outros casos} \end{cases} \qquad \text{eq. 2}$$

onde:

η (n): é a taxa de aprendizado, e

$\Lambda_{i(x)}$ (n): é a vizinhança centrada no neurônio vencedor.

3.3. Classificação por Rede neural Back propagation

Redes neurais Back-Propagation (BP) [7] são redes *feed-forward* de uma ou mais camadas escondidas. Este algoritmo define uma maneira sistemática de atualização dos pesos das diversas camadas baseada na idéia que os erros dos neurônios

das camadas escondidas são determinados pela retropropagação reversa dos erros dos neurônios da camada de saída.

O treinamento, supervisionado, é baseado no método do gradiente decrescente (*gradient descent*), buscando minimizar o erro global da camada de saída. Deste modo, a atualização do peso (Δw_{ij}) é proporcional ao negativo da derivada parcial do erro com relação ao próprio peso:

$$\Delta\omega_{ij} = -\eta \frac{\partial E}{\partial \omega_{ij}} \qquad \text{eq. 3}$$

onde η é a taxa de aprendizado e E é a função erro definida como:

$$E = \frac{1}{2}\sum_{i=1}^{N_0}\left(t_j - s_j\right)^2 \qquad \text{eq. 4}$$

onde N_o é o número de processadores da camada de saída e t_j é o valor esperado na saída do processador j.

Derivando a eq. 3, chega-se à seguinte fórmula:

$$\Delta\omega_{ij} = \eta \cdot s_i \cdot e_j$$

$$e_j = \begin{cases} \left(t_j - s_j\right) f'\left(net_j\right) & \text{se } j \in \text{camada de saída} \\ f'\left(net_j\right)\sum_{k=1}^{N}\omega_{jk} e_k & \text{se } j \in \text{camada escondida} \end{cases} \qquad \text{eq. 5}$$

onde η é a taxa de aprendizado; s_i é a entrada associada ao peso w_{ij}; e_j é o erro do j-ésimo processador; t_j é o valor desejado de saída do processador j; s_j é o seu estado de ativação; net_j é o seu potencial interno; f' é a derivada da função de ativação; e N é o número de processadores na camada seguinte à camada do processador j.

Como pode-se verificar da equação acima, o algoritmo de aprendizado do Back Propagation tem duas fases, para cada padrão apresentado: *Feed-Forward* e *Feed-Backward*. Na primeira etapa as entradas se propagam pela rede, da camada de entrada até a camada de saída, gerando a saída da rede em resposta ao padrão apresentado. Na segunda etapa, os erros se propagam na direção contrária ao fluxo de dados, indo da camada de saída até a primeira camada escondida, atualizando os pesos sinápticos. Este procedimento de aprendizado é repetido diversas vezes, até que, para todos os processadores da camada de saída e para todos os padrões de treinamento, o erro seja menor do que o especificado.

Foi demonstrado que o algoritmo Back Propagation é um aproximador universal [8], sendo capaz de aprender qualquer mapeamento de entrada-saída. Entretanto, apesar do grande sucesso do Back Propagation nas mais diferentes aplicações, existem alguns problemas básicos: a definição do tamanho da rede, o longo processo de treinamento, e fenômenos como paralisia da rede (contornado diminuindo o valor de η) e mínimo local (que pode ser solucionado utilizando-se métodos estatísticos).

A definição do tamanho da rede, isto é, o número de camadas escondidas e número de processadores em cada uma dessas camadas, é um compromisso entre convergência e generalização. Convergência é a capacidade da Rede Neural de aprender todos os padrões do conjunto de treinamento. Generalização é a capacidade de responder corretamente aos padrões nunca vistos (conjunto de teste). O objetivo é utilizar a menor rede possível, de forma a se obter uma boa generalização, que seja capaz de aprender todos os padrões.

A taxa de aprendizado η é um parâmetro importante a ser definido no aprendizado. Esta não deve ser nem muito pequena, causando um treinamento muito lento, nem muito grande, gerando oscilações. Quando a taxa de aprendizado é pequena, e dependendo da inicialização dos pesos (feita de forma aleatória), a Rede Neural pode ficar presa em um mínimo local. Quando a taxa de aprendizado é grande, a Rede Neural pode nunca conseguir chegar ao mínimo global, pois os valores dos pesos são grandes. A solução para este problema é utilizar uma taxa de aprendizado adaptativa. Além deste parâmetro, pode-se também utilizar um termo α de momento [7], proporcional à variação no valor do peso sináptico no passo anterior. Deste modo, a equação de atualização do peso sináptico w_{ij} é modificada da seguinte forma:

$$\Delta\omega_{ij}\left(t+1\right) = \eta \cdot s_i \cdot e_j + \alpha\Delta\omega_{ij}\left(t\right) \qquad \text{eq. 6}$$

A utilização do termo de momento tem a função de acelerar a convergência da rede, sem causar oscilações.

4. Estudo de Caso

Esta seção apresenta os teste e resultados obtidos com a metodologia proposta.

Do histórico dos dados fornecidos pela Light foram utilizadas as faturas originadas pelo sistema FAT, isto porque durante a fase de análise mostraram-se mais representativas na base de dados, além de possuir maior número de clientes/faturas.

Dado o volume de clientes/faturas existente na base de dados, foi necessário agrupá-los. O critério usado para agrupar foi a similaridade entre as atividades econômicas de cada cliente definidas no sistema da Light.

Dessa forma, identificou-se 12 grupos de clientes/faturas representativos na base de dados.

Os 6 primeiros grupos foram definidos a partir da análise estatística da base de dados, por apresentar uma quantidade maior de clientes/faturas no período de janeiro de 2003 a setembro de 2005. Sendo que os

clientes/faturas de cada uma destas atividades apresentam entre si uma correlação significativa. Os testes comprovaram que cada uma delas constitui-se naturalmente um grupo. A Tabela 1 lista as 6 primeiros grupos de atividade.

Código da Atividade	Nome da Atividade
5211	Hotéis e Motéis
5221	Restaurantes e Lanchonetes
5900	Bancos Comerciais e Caixas Econômicas
6109	Com. Varejo de Combustíveis e Lubrificantes
6115	Supermercados
6800	Comercial - Serviços

Tabela 3 – Primeiros 6 grupos de atividades

Os 5 grupos seguintes foram obtidos dentre 30 atividades com correlações significativas e que apresentaram um número de clientes/faturas por atividades considerável.

Este agrupamento foi realizado considerando a influencia dos índices econômicos com a correlação ao percentual de inadimplência de cada atividade usando uma rede neural auto-organizável do tipo mapa de Kohonen.

O aprendizado da rede neural foi dividido em duas fases:

- Fase de ajuste: nesta fase a taxa de aprendizado deve variar entre 0.9 e 0.02, decrescendo exponencialmente. A vizinhança deve decrescer linearmente com o treinamento até que somente um neurônio vencedor seja atualizado. Tipicamente, esta fase deve ter milhares de iterações (50000) e tem como objetivo distribuir os vetores de pesos dos neurônios de acordo com a densidade dos padrões de entrada [2].
- Fase de convergência: Nesta fase a taxa de aprendizado deve ser mantida em 0.02 por mais alguns milhares de iterações (50000) e a vizinhança deve permanecer fixa em 0 [2].

Como resultado deste agrupamento foram obtidos os 5 grupos que são listados na Tabela 4.

O último grupo é formado pelo clientes/faturas dos restantes das atividades da base de dados definidas no sistema da Light, por terem pouca representatividade em relação ao número de clientes/faturas.

Para cada um destes grupos foi modelada uma rede neural que tem por função classificar cada cliente em adimplente ou inadimplente.

Para avaliar os resultados da classificação, a Light forneceu para cada mês um conjunto de clientes inadimplentes.

	Código da atividade	Nome da Atividade
Grupo 1	1299	Fabri. de outras maquinas e equipamentos
	2099	Fabri. de outros produtos químicos
	5031	Transporte rodoviários passageiros
	5032	Transporte rodoviários carga
	5433	Assistência medica odontológica e veterinária
	5599	Outros serviços comerciais não especificados.
	5910	Bancos empresas credito, financiamento,investimento.
	6104	Com.varej.de veículos e acessórios
	6199	Com.varej.de produtos não especificados
	6999	Atividades não especificadas ou não classificadas.
Grupo 2	6106	Com.varej.de moveis, art.para habitação
	6116	Magazines
	8010	Entidades religiosas
	8021	Associações beneficentes, relig.assist.
	8029	Outras associações
Grupo 3	1199	Fabri. outros artigos de metal não especificadas.
	2510	Confecções de roupas e agasalhos
	5462	Estabelecimentos particulares ensino do 2 grau
	6103	Com.varej.de veículos
	8023	Associações esportivas e recreativas
Grupo 4	6112	Com.varej.de artigos de vestuário
	2910	Impressão e edição de jornais e livros
	8023	Associações esportivas e recreativas
	3310	Construção civil
Grupo 5	2110	Fabri. de produtos farmacêuticos e veterinários
	2350	Fabri. artículos, materiais plásticos para embalagem
	2399	Prod. de outros art.materiais plásticos
	2670	Fabricação. produtos de padaria, confeitaria
	5463	Estabelecimento particulares de ensino superior
	5699	Outros serviços, diversões não espec.n.classif.
	6114	Mercearias, armazéns e padarias

Tabela 4 – Agrupamento de atividades realizado por RN de Kohonen.

Na avaliação dos resultados de classificação do sistema compara-se a lista de clientes inadimplentes real com a lista de clientes fornecida da pelo sistema após a classificação. Desta forma, é possível calcular os erros de classificação para cada grupo de atividades.

Para os 6 primeiros grupos as redes neurais back propagation usadas, foram modeladas considerando as seguintes entradas:

- Valor da Fatura,
- Tempo de Inadimplência da Fatura Anterior,
- Quantidade de vezes inadimplentes até o momento da fatura em questão,
- Código da Localidade

Como saída destas redes foi:

- Inadimplente ou não nesta fatura.

O código de localidade nos testes foi representado sob a forma binária.

Como o número de clientes/faturas inadimplentes em todos os grupos é menor que o número de clientes/faturas adimplentes, se faz necessária a aplicação de técnicas de balanceamento dos dados

para melhorar o desempenho da rede neural, isto é, ter o mesmo número de amostras de adimplentes e inadimplentes sem perda de informação.

A técnica de balanceamento dos dados usada foi a de reamostragem (ou réplica) dos dados de inadimplentes, aumentando assim a quantidade de amostras de inadimplentes até atingir o mesmo número de amostras dos clientes/faturas adimplentes.

Foram realizados diversos testes com redes neurais variando-se o número de neurônios da camada escondida para se identificar a melhor topologia. Também foram feitos testes com estas diferentes topologias para vários conjuntos de amostras de dados.

A seguir na Tabela 5 mostram-se os resultados obtidos para os 6 grupos utilizando a técnica de replicação para o balanceamento dos dados. Indicando as melhores topologias das redes neurais (número de neurônios), com os erros percentuais da classificação no treinamento e teste .

Atividade	Representação Tipo de Localidade	N° Neuronios na Camada Escondida	Erro de Treinamento (%)	Erro de Teste (%)
5211	binário	19	13.38	13
5221	binário	15	19.94	17.94
5900	binário	23	22.61	20.64
6109	binário	26	11.85	10.84
6115	binário	25	15.61	13.96
6800	binário	21	19.31	18.13

Tabela 5 – Resultados das Redes Neurais para os 6 grupos utilizando a técnica de replicação para o balanceamento.

Os clientes dos 5 grupos seguintes obtidos dentre as 30 atividades com correlações significativas e que apresentaram um número de clientes/faturas por atividades considerável. Agrupados usando uma rede neural auto-organizável do tipo mapa de Kohonen, foram classificados usando redes neurais back propagation, estas redes foram modeladas considerando as mesmas entradas que no caso anterior.

A Tabela 6 mostra os resultados da classificação obtidos para estes 5 *clusters* de clientes.

Atividade	N° Neuronios na Camada Escondida	Erro de Treinamento	Erro de Teste
Grupo 1	27	15.8165	14.6119
Grupo 2	32	21.6782	20.7477
Grupo 3	27	9.6307	8.5749
Grupo 4	29	16.0111	12.7430
Grupo 5	19	14.3016	11.2840

Tabela 6 – Resultados da classificação por Redes Neurais para os 5 grupos.

Para o último grupo, formado pelo clientes/faturas do restantes das atividades da base de dados da Light, foi modelada uma rede neural back propagation que usa como entradas além das entradas indicadas anteriormente, os índices econômicos mensais obtidos da Macrodados. Os índices utilizados estão listados na Tabela 7.

ÍNDICES
Indústria Geral - RJ
Indústria de Transformação - RJ
Alimentos - RJ
Bebidas - RJ
Farmacêutica - RJ
Metalurgia Básica - RJ
Vestuário e Acessórios - sem Ajuste - RJ
Nominal Vendas no Varejo (Comb.e Lubr.) - RJ
Nominal Vendas Var.(Hip,Sup.,Prod.Alim.,Beb.Fumo)-RJ
Nominal Vendas no Varejo (Hiperm. e Superm.) - RJ

Tabela 7 – Índices econômicos usados como entreda na Rede Neural.

A saída da rede neural é o percentual de inadimplência do grupo mês a mês. Este percentual de inadimplência do grupo no mês foi calculada segundo a seguinte relação:

$$\%Inadimplencia = \frac{Número\ \ Inadimplentes\ \ Mês_i}{Número\ \ Clientes\ \ Mês_i} * 100 \qquad \text{eq. 7}$$

Foram realizados diversos testes com redes neurais variando-se o número de neurônios da camada escondida para se identificar a melhor topologia.

A Tabela 8 apresenta o resultado da melhor topologia da rede neural, com os erros de treinamento e teste para a previsão da inadimplência.

N° Neurônios Camada Escondida	Erro Treinamento (%)	Erro Teste (%)
5	34.92	48.58

Tabela 8 – Resultados da Rede Neural para a previsão da inadimplência.

A Figura 2 apresenta o gráfico com o percentual de inadimplência prevista pela rede neural e o real ao longo do período de 30 meses.

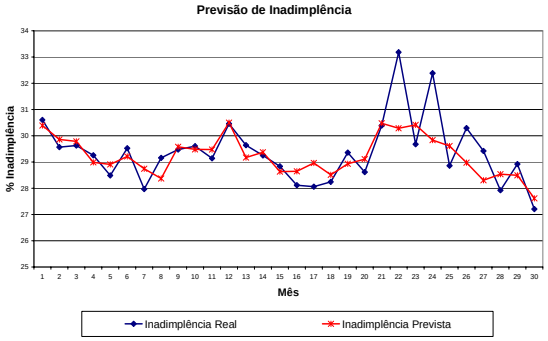

Figura 2 – Percentual de inadimplência real e prevista por redes neurais

Das tabelas 5, 6 e 8 observa-se como o sistema com a metodologia proposta conseguiu identificar os clientes inadimplentes, com erros baixos.

5. Conclusões

Este artigo apresentou o desenvolvimento de uma metodologia para detecção de inadimplência no faturamento de energia elétrica.

Os resultados obtidos com o sistema evidenciaram a capacidade do sistema em detectar clientes inadimplentes, o que foi comprovado e confirmado pelos dados históricos da Light.

Referências Bibliográficas

[1] Camargo, S.H.C.R.V., Farah, O.E., Martinelli, D.P., Estudo Sobre Inadimplência E Decisões Tomadas Pelas Empresas.

[2] Kohonen, T.: Self-Organizing Maps. 2nd Edition. Springer-Verlag Berlin Heidelberg New York, 1997.

[3] Bezdec, J.; Pal., S. K. Fuzzy Models for Pattern Recognition. New Jersey: IEEE Press, 1991.

[4] Guha, S.; Rastogi, R.; Shim K. Cure: An Efficient Clustering Algorithm for Large Databases. In: ACM SIGMOD International Conference on Management of Data, Seattle. 1998.

[5] Michael, J. A.; Berry, G. L. Data Mining Techniques For Marketing, Sales and Customer Support. New York: John Willey & Sons, 1996.

[6] Zurada, Jacek M.: Introduction to Artificial Neural System. West Publishing Company, 1992.

[7] Haykin, S.: Neural Networks. A Comprehensive Foundation. Prentice-Hall, Inc., 1994.

[8] Hornik, K., Stinchcombe, M, White, H., Multilayer Feedforward Networks are Universal Approximators, Neural Networks, Vol. 2, pp. 359-366, 1989.