

## Modelos de Gestão do Risco de Inadimplência – Uma aplicação ao Segmento Educacional

**Autoria:** Marcelo José Vieira de Melo Sobrinho, Charles Ulises De Montreuil Carmona

**Resumo:** O crescimento da inadimplência escolar verificado em anos recentes tem aumentado a preocupação dos gestores quanto ao equilíbrio financeiro de suas instituições. É diante deste contexto que a gestão do risco creditício tem se tornado uma questão de suma importância para as instituições de ensino, e o *credit scoring* apresenta-se como uma possível alternativa para superar as limitações inerentes aos participantes deste mercado, ao passo que possibilita o gerenciamento ativo do risco de crédito. A utilização deste tipo de modelagem na previsão da inadimplência dos serviços educacionais foi testada através das técnicas de Redes Neurais Artificiais e da Regressão Logística. Os resultados obtidos demonstraram que o *credit scoring* é um método viável, pois supera o índice de acerto pelo critério de chances, e é capaz de realizar previsões com índices de acurácia de até 88,52%. Os resultados obtidos pelos diferentes modelos demonstram que não há uma supremacia de uma técnica sobre a outra, pois os desempenhos obtidos pelas Redes Neurais e pela Regressão Logística são equivalentes.

### 1 Introdução

Diante das incertezas presentes no ambiente empresarial os modelos estatísticos surgem como uma ferramenta de fundamental importância para o gerenciamento eficaz do risco.

O presente trabalho é direcionado especificamente para o risco de crédito. Este segmento do conhecimento científico tem tomado crescente significância e atraído cada vez mais a atenção da comunidade acadêmica, e dos participantes do mercado, pela sua aplicabilidade na proteção dos agentes econômicos. No Brasil este tema começou a receber maior atenção a medida que o Plano Real, implantado em 1994, possibilitou o deslocamento dos recursos disponíveis no país, até então direcionados para a especulação financeira e financiamento da dívida interna devido ao processo inflacionário descontrolado, para atividades produtivas através da concessão de crédito (PRADO; BASTOS; DUARTE JR., 2000).

Entre as instituições que predominantemente são caracterizadas pela atividade financeira, o gerenciamento do risco de crédito desempenha um papel primordial, pois está associado à atividade principal da instituição. Problemas na carteira de crédito de empresas dessa natureza possuem uma relação negativa direta com seu desempenho financeiro. Deste modo existe uma preocupação e uma cultura orientada para o gerenciamento do risco. Por outro lado, para instituições que não tem na concessão de crédito um dos objetivos primordiais da organização, o risco de crédito apresenta-se como um desafio adicional, pois não há uma competência estabelecida que possa viabilizar uma análise de crédito de maneira confiável e precisa, nos termos da necessidade de cada usuário.

Um segmento que historicamente não tem direcionado esforços suficientes para o gerenciamento do risco de crédito é o educacional, fato este que pode ser percebido pelo crescente nível de inadimplência que o setor tem enfrentado nos últimos anos, segundo o Sindicato das Escolas Particulares de Pernambuco (2005).

As instituições educacionais atuantes na rede privada têm sido severamente penalizadas com o nível de inadimplência verificados em anos recentes. A falta de uma cultura orientada para a gestão do crédito tem tornado o equilíbrio financeiro das escolas particulares vulneráveis a uma série de fatores, dentre os quais se destacam as forças

competitivas do mercado educacional, a perda do poder econômico de seu público alvo, além de uma legislação desfavorável.

As empresas que estão direcionadas para a área educacional trabalham com a prestação de serviços, o que difere bastante dos empreendimentos que recorrem a oferta de recursos monetários ou fornecimento de bens físicos e que são objeto de estudo para pesquisas de inadimplência com maior frequência, como os trabalhos de Thomas (2000), Lee e Jung (2000), Cia (2003) e Almeida e Passari (2005). No entanto, o fornecimento de um serviço educacional e sua contraprestação pecuniária também pode ser percebido como um crédito, onde a instituição arca com os custos de manutenção da estrutura, remuneração de funcionários, entre outras despesas, durante o mês e espera ser remunerada ao final deste período.

Recentemente este setor tem sido severamente penalizado pela inadimplência de seus clientes. Em 23 de novembro 1999 foi sancionada a lei 9870 limitando os meios que estavam disponíveis as estas instituições para exigir o pagamento pela contraprestação de seus serviços. Esta lei vedou a suspensão da prestação dos serviços pela inadimplência por parte do cliente, e ao mesmo tempo não permite que a documentação do inadimplente fique retida. Estas medidas funcionam como um mecanismo que não prima pelo incentivo ao cumprimento do pagamento das mensalidades, e colocam como única medida cabível recorrer a cobrança no âmbito judicial.

O gerenciamento do risco de crédito não é uma área relativamente nova, por outro lado não existem muitos estudos que se propõe a trabalhar com o desenvolvimento de modelos direcionados para empresas não financeiras. Por isso, o presente trabalho destina-se a verificar a possibilidade de implantação de modelos de risco de crédito em uma instituição de ensino, comparando o desempenho dos modelos entre diferentes metodologias de estimação.

O desenvolvimento deste trabalho teve como função principal o desenvolvimento de modelos estatísticos utilizados como um mecanismo de auxílio às instituições educacionais que historicamente não gerenciam seu risco de inadimplência de maneira ativa, de modo a muni-las dos preceitos fundamentais para uma administração financeira mais sadia.

## **2 Porque pesquisar a inadimplência em instituições de ensino?**

A atenção despendida à administração do risco de crédito tem tomado volume para corresponder às expectativas da crescente demanda por recursos.

O foco desta preocupação tem sido primordialmente direcionado, tanto pelo meio acadêmico como pelos agentes de mercado, para o desenvolvimento de modelos de risco de crédito relacionados para a concessão de recursos na forma de empréstimos, para pessoas físicas e jurídicas, ou para a previsão de concordatas que se configura como um caso específico de risco de crédito.

Como comenta Amorin Neto (2000) a concessão de crédito extrapolou os limites bancários e permeia diversas atividades empresariais como: cooperativas de crédito, empresas de cartões de crédito e o comércio em geral. Para instituições financeiras a análise minuciosa dos riscos inerentes a concessão de empréstimos faz parte do corpo de suas atividades principais, por isso existe uma maior concentração no número de trabalhos para este segmento. No entanto outras atividades empresariais têm percebido a necessidade de gerenciar melhor sua carteira de crédito.

O segmento educacional é uma destas áreas que não tem na concessão de crédito sua atividade principal e que tem percebido a duras penas a necessidade de desenvolver métodos de se precaver contra as altas taxas de inadimplência que tem enfrentado.

A questão da inadimplência para as instituições do segmento educacional atuantes no estado de Pernambuco tem se tornado cada vez mais crítica, e alguns fatores tem se tornado

preponderantes para esta situação. A redução do poder aquisitivo das famílias que habitualmente mantêm seus filhos em escolas particulares tem sido um deles. Como um reflexo desta situação, segundo a Secretaria Estadual de Educação (2006), em 1996 as escolas particulares correspondiam a 14,4% dos alunos matriculados no ensino médio em todo o estado, já em 2005 este percentual foi reduzido a 12,3%. Associado a este fato está o aumento do número de escolas funcionando regularmente na região; dados da Secretaria de Educação (2006) indicam que nos dois últimos anos 392 novas escolas iniciaram suas atividades atingindo o número total de 2.549, aumentando assim o número de vagas disponíveis ao mesmo tempo em que o número de alunos foi reduzido.

Outro fator que tem contribuído com a inadimplência escolar, como mencionada anteriormente, foi a sanção da lei 9.870 que tornou ainda mais crítica a situação financeira das escolas particulares, limitando os meios disponíveis para que estas cobrassem as dívidas que lhes são devidas. A consequência desta lei, que visa preservar o direito a educação dos alunos, é que não há um incentivo para que os responsáveis pelo pagamento dos serviços educacionais mantenham as mensalidades quitadas, pois as sanções disponíveis, a partir de então, recaem apenas na esfera jurídica.

A soma de todos estes fatores desencadearam um fluxo de inadimplência no setor que tem atingido a marca de 35%, segundo dados do Sindicato dos Estabelecimentos de Ensino do Estado de Pernambuco (2006).

Por não fazer parte de seu objetivo principal as instituições de ensino historicamente tornaram a administração do risco de crédito como uma atividade secundária. As poucas instituições que realizam alguma atividade neste sentido limitam-se apenas a consultas ao Serviço de Proteção ao Crédito (SPC) e ao Serasa, sem que, no entanto, haja uma real intenção de limitar o acesso aos serviços da empresa.

É diante deste contexto que a inadimplência tem se tornado uma questão de suma importância para as instituições de ensino, e a modelagem estatística, mais especificamente o *credit scoring*, apresenta-se como uma possível alternativa para superar as limitações legais, ao passo que possibilita o gerenciamento ativo do risco de crédito.

Sendo assim, o objetivo central deste trabalho foi avaliar se a utilização de modelos de *credit scoring* constitui-se em uma alternativa viável para identificar o risco de inadimplência na prestação de serviços educacionais. Para tal foram pesquisados os clientes de uma instituição de ensino, atuante no ensino privado de nível médio e fundamental, localizada na cidade do Recife.

### **3 *Credit Scoring* e o segmento educacional**

O *credit scoring* consiste no desenvolvimento de um modelo capaz de fornecer uma medida quantitativa que reflete a credibilidade de um indivíduo, quanto ao pagamento de um produto ou serviço, baseando-se em um conjunto de informações preditoras (LEE; JUNG, 2000). O score de crédito pode ser percebido também como um mecanismo que permite o reconhecimento do status de diferentes grupos, como adimplentes e inadimplentes, em uma população ao se realizar a previsão de um evento, como a inadimplência. (TOMAS, 2000).

A possibilidade de realizar uma avaliação da credibilidade de um indivíduo tomando-se por base mecanismos estatísticos, concede ao *credit scoring* uma vantagem em comparação aos sistemas de julgamento subjetivo quanto a rapidez e a confiabilidade com que as decisões são tomadas, além de possibilitar uma redução de custos de análise e no monitoramento de créditos existentes.

Um dos principais benefícios da utilização do *credit scoring* é a possibilidade de otimizar a decisão de quais créditos devem ser concedidos, aumentando a lucratividade, e quais devem ser negados, reduzindo o percentual de perdas (WEST, 2000).

Deste modo a construção do *credit scoring* dentro de uma instituição educacional pode ser utilizado como um critério auxiliar na seleção de novos alunos, onde a efetivação da matrícula ficará condicionada ao resultado positivos da aplicação do modelo.

A utilização do *credit scoring* para a tomada de decisão de aceitação ou rejeição de um potencial aluno é crucial do ponto de vista de leis contra a discriminação, pois a tomada de decisão é feita de maneira objetiva, respaldada pela modelagem estatística, eliminando o potencial viés da subjetividade humana (ROSEMBERG e GLEIT, 1994).

Outro benefício esperado com a utilização do *credit scoring* é que os modelos forneçam informações importantes para um gerenciamento financeiro eficaz. Ao identificar previamente os alunos que tem alta probabilidade de se tornarem inadimplentes em algum momento dentro do ano letivo, a instituição pode ter uma medida do volume de recursos que serão recebidos ao longo dos meses. De posse destas informações a instituição educacional disporia de mais subsídios para equacionar o montante de recursos que se espera receber, com as despesas incorridas no desempenho da atividade educacional.

Por outro lado é necessário perceber que a utilização da modelagem de *credit scoring* também traz alguns desafios e algumas limitações. Este tipo de modelagem é estática, considerando as condições econômicas e as variáveis que caracterizam os indivíduos pesquisados como estáveis, além de se basear no fato de que as relações passadas sobre a inadimplência são fortes preditoras do comportamento futuro dos clientes.

Ou seja, o modelo é desenvolvido partindo do pressuposto que a relação histórica, obtida entre os dados sobre as características dos clientes e a inadimplência, permanecerá vigente em períodos subseqüentes para que o modelo possa ser aplicado. Sendo assim, o modelo parte do pressuposto que as características que no passado apresentaram forte poder preditivo na identificação da inadimplência podem ser replicadas futuramente a um novo conjunto de clientes.

No entanto a base de clientes muda lentamente ao longo do tempo, o que possivelmente pode alterar o perfil das características que melhor predizem a inadimplência.

Muito embora este seja um contratempo a ser resolvido a médio prazo, uma vez que levaria pelo menos treze anos para alterar toda a base de alunos da escola, considerando que ao final de cada ano apenas uma turma deixa a instituição. Outro ponto que pode promover a alteração desta relação é a mudança no cenário econômico. O perfil do cliente inadimplente possivelmente é influenciado diferentemente em ciclos de expansão e retração econômica.

Todas estas alterações exigem que os modelos de crédito sejam constantemente reavaliados, e refeitos, o que potencialmente seria uma fonte de problemas para a instituição pesquisada. O desenvolvimento de um *credit scoring* exige o conhecimento das técnicas estatísticas utilizadas e atualmente nenhum dos colaboradores da instituição dispõe de uma formação que lhe permita ter a compreensão deste tipo de informação.

Outro fator que potencialmente pode apresentar-se como uma limitação para a utilização do *credit scoring* em uma instituição de educação é o impacto que será causado na opinião dos clientes da instituição. A implantação deste sistema deve ser feita de maneira transparente, onde os pais e responsáveis devem ser informados sobre os objetivos e a imparcialidade do modelo, caso contrário pode causar constrangimento entre os clientes e afastá-los da instituição.

Diante dos fatos expostos é possível perceber os benefícios que potencialmente podem ser auferidos através do modelo de *credit scoring*. No entanto, sua utilização prescinde de uma cautelosa avaliação dos efeitos diretos e indiretos que podem repercutir de maneira desfavorável a instituição.

## 4 Metodologia

### 4.1 População e amostra

A instituição de ensino escolhida para este estudo está localizada na cidade de Recife e presta serviços educacionais para os níveis fundamental e médio para uma população de 303 alunos. A identidade da instituição e de seus clientes é omitida durante toda a pesquisa por solicitação da direção da escola uma vez que aborda uma questão sigilosa e delicada, relativa a inadimplência.

Do total de 303 alunos que constam no registro da ficha cadastral da escola, 58 indivíduos foram descartados. Dentre eles 25 são contemplados com bolsas de estudo, 12 alunos que possuíam ficha cadastral incompleta, além daqueles alunos que ingressaram na instituição após o mês de janeiro ou que saíram antes agosto, período estabelecido para a coleta dos dados, que totalizam mais 21 alunos. Deste modo, a amostra final utilizada na composição dos modelos é de 245 alunos.

Sendo assim, devido as características da modelagem utilizada no estudo, do número limitado de indivíduos da população e da disponibilidade da instituição de ensino em ceder o banco de dados completo de seus clientes, optou-se pela utilização de todos os alunos para composição da amostra, sendo descartadas apenas alguns indivíduos conforme os critérios citados anteriormente, totalizando 245 observações.

As observações que compuseram os modelos são referentes ao período de janeiro a agosto do ano de 2006. Por sua vez os dados sobre a situação de inadimplência envolvem também os anos de 2005 e 2004 e foram utilizados como variáveis defasadas nos modelos de escoragem comportamental.

Ao agrupar cada uma das observações em uma base de dados única, procurou-se considerar o maior número possível de informações que pudessem ser quantificadas, uma vez que devido a natureza do estudo não existem muitos indícios de quais variáveis são mais relevantes, reservando a escolha de quais deveriam realmente compor o modelo para os testes de significância estatística.

### 4.2 Definição de inadimplência

A inadimplência pode ser percebida de maneira ampla como a incapacidade de pagar obrigações financeiras na data de vencimento (BEAVER, 1967). No entanto, existem peculiaridades ligadas a certos ramos de atividade que exigem uma compreensão mais elaborada a respeito deste conceito.

Ao se avaliar a qualidade de crédito no contexto da prestação de serviços educacionais faz-se necessário uma análise prévia do histórico dos atrasos, no intuito de identificar o perfil da demora no pagamento das mensalidades. Deste modo é possível estabelecer os limites que caracterizam os atrasos na quitação dos débitos como uma simples demora ou como caso de real incapacidade de arcar com as mensalidades pendentes.

Ao avaliar o perfil dos atrasos no pagamento das mensalidades foi possível perceber que a maior parte dos atrasos, 64,34% dos casos, em média, ocorrem de 1 a 30 dias, com uma predominância dos casos onde a pendência na quitação leva de 1 a 10 dias. Este fato demonstra que estes pequenos atrasos são frequentes, mas que rapidamente são solucionados não devendo, portanto, representar uma grande fonte de preocupação para a instituição.

Superando 30 dias entre a data de vencimento e o não pagamento da dívida uma nova mensalidade entra em vigor, praticamente duplicando o valor a ser pago o que reduz a capacidade de pagamento que já se apresentava limitada. Suscita-se então a possibilidade de utilizar atrasos acima de 30 dias para indicar a situação de crédito como inadimplente. No

entanto, os dados revelam que 86,53% dos alunos apresentam em algum momento um ou mais atrasos superiores a 30 dias, demonstrando ser uma situação que ocorre ao menos eventualmente com a grande maioria dos indivíduos.

Deste modo, diante do padrão natural de atraso encontrado ao se observar os dados, faz-se necessário flexibilizar o conceito de inadimplência, dilatando o prazo de tolerância entre a data de vencimento da mensalidade e a data de sua quitação. De modo que seja possível separar os indivíduos que apresentam uma real dificuldade em arcar com os custos da mensalidade, daqueles que eventualmente apresentaram alguma dificuldade de pagamento durante o período de análise.

Sendo assim os indivíduos considerados como inadimplentes na composição do modelo são aqueles que apresentam mais de 60 dias de atraso no pagamento da mensalidade pelos serviços educacionais, observando como referência a data para quitação estabelecida em contrato. Ao utilizar este critério foram identificados 62 indivíduos indicando que 25,31% dos alunos ficaram devendo mais de 2 meses a mensalidade, pelo menos uma vez dentro do período de janeiro a outubro de 2006. Dentre estes alunos que ultrapassavam o prazo de 60 dias a média de atraso é de aproximadamente 5 meses.

### **4.3 Variáveis utilizadas no estudo**

O procedimento seguinte à seleção da qualidade de crédito foi a identificação das variáveis independentes, ou explicativas, que são relacionadas às características dos indivíduos que consomem os serviços educacionais da instituição em análise.

O processo de identificação das variáveis explicativas que estão fortemente ligadas ao fenômeno da inadimplência deve seguir uma composição de variáveis relatadas na literatura e em estudos anteriores, assim como na experiência adquirida pelo pesquisador ao se envolver com o tema pesquisado (HAIR et al, 2002).

Estudos prévios sobre a inadimplência indicam que não existe um consenso na literatura sobre concessão de crédito, quanto as variáveis que melhor identificam o fenômeno em questão. Muitos estudos apontam para características explicativas específicas ao contexto de cada pesquisa.

Uma vez que as pesquisas sobre inadimplência no segmento de serviços são escassas, e não conclusivas quanto à determinação das variáveis independentes que melhor identificam indivíduos inadimplentes, procurou-se levar em consideração não só as características encontradas nos trabalhos pesquisados, como também o maior número possível de informações disponíveis nos registros da instituição de ensino pesquisada. Reservando, assim, a função de identificar quais variáveis são efetivamente preditoras da inadimplência para os testes estatísticos realizados no desenvolvimento dos modelos.

No entanto como é possível observar, algumas destas características são de natureza qualitativa, não apresentando, a princípio, uma escala de medida numérica. Deste modo precisaram receber um tratamento para que pudessem ser enquadradas dentro de categorias e assim inseridas nos modelos estatísticos.

**Quadro 1 : Variáveis consideradas no estudo.**

Variáveis	
1. Reside em casa ou apartamento.	11. Empresa que a mãe trabalha.
2. Bairro	12. Natureza da empresa empregadora da mãe.
3. Telefone residencial.	13. Telefone celular da mãe.
4. Instrução do pai	14. Possui contrato assinado com a escola.
5. Função exercida pelo pai.	15. Possui irmão na instituição.
6. Empresa que o pai trabalha	16. Plano de saúde.
7. Natureza da empresa empregadora do pai.	17. Data de vencimento da mensalidade.
8. Telefone celular do pai.	18. Valor pago pela mensalidade.
9. Instrução da mãe.	19. Débitos no ano de 2004.
10. Função exercida pela mãe.	20. Débitos no ano de 2005.

## 5 Resultados

### 5.1 Desenvolvimento dos modelos: Regressão logística

Dentre os 245 casos obtidos junto a instituição pesquisada, que compuseram a amostra final utilizada no estudo, aproximadamente 25% dos dados, 61 observações, foram reservadas como amostra de teste no intuito de realizar uma validação cruzada. Neste procedimento os 184 casos restantes foram utilizados no desenvolvimento dos modelos e sua capacidade avaliada através do desempenho em prever corretamente os dados contidos na amostra de teste.

A literatura financeira direcionada ao risco de crédito não especifica o percentual de dados que devem ser reservados para as etapas de modelagem e teste. A escolha que foi utilizada neste estudo está baseada no fato de que a quantidade de casos utilizados na modelagem deve ser em número suficiente para incluir operações de crédito representativas de todas as operações, ou seja, preferencialmente que o maior número possível de informações sejam abarcadas durante a modelagem (VASCONCELOS, 2002).

Operacionalizando a regressão logística foi utilizada a opção *stepwise* como mecanismo de seleção das variáveis que compuseram a equação final. Dentre as diversas combinações de variáveis independentes elaboradas pelo *stepwise* as seguintes observações foram selecionadas como as melhores preditoras da inadimplência, baseando-se no poder discriminatório.

**Tabela 1 : Resultado do modelo de aprovação de crédito - regressão logística**

Variáveis	Coefficiente	Significância.
INSTPAI – Nível de instrução do pai	-0,3041	0,0198
INSTMAE – Nível de instrução da mãe	-0,4175	0,0011
FUNCM – Função que a mãe exerce	0,8460	0,0026
CONTRATO – Contrato de prestação de serviço com a escola	-1,1645	0,0081
CONSTANTE	2,5446	0,0024

O resultado da regressão indica que o coeficiente estimado para o nível de instrução do pai sugere que para o aumento em uma unidade no nível de instrução indica que a sua pontuação de crédito, ou seja, a chance de se tornar inadimplente, vai sendo reduzida em - 0,3041. Comportamento análogo foi observado na variável nível de instrução da mãe.

A variável relacionada a função que a mãe desempenha no emprego apresenta um sinal positivo indicando que o a probabilidade de inadimplência aumenta em 0,8460 a medida que aumenta em uma unidade o valor atribuído a categoria da função da mãe. Vale ressaltar

que as funções de nível estratégico fazem parte da categoria 1, as gerenciais compreendem a categoria 2 e as operacionais a terceira categoria, ou seja, quanto mais operacional a função exercida maiores as chances de se tornar inadimplente.

Por fim, o coeficiente apresentado pela variável contrato indica que sua existência tem efeito negativo sobre a inadimplência.

Vale ressaltar que o contrato é entregue a todos os pais no ato da matrícula, onde alguns são assinados e devolvidos a instituição de imediato, e outros preferem não assinar sob o pretexto da necessidade de realizar uma análise posterior, e não devolvem. A existência ou não do contrato provavelmente não influencia na capacidade de pagamento das mensalidades, mas a julgar pela significância do coeficiente da regressão, a existência do contrato funciona como um motivador para a quitação dos débitos, ou pelo menos demonstra, a princípio, a intenção de cumprir com os pagamentos assumidos.

No desenvolvimento da regressão logística para a construção do modelo *behavioral scoring* foi utilizada a mesma partição dos dados, em amostras de modelagem e teste, que foi apresentada anteriormente. O que difere este modelo do anterior resume-se a inclusão das variáveis referentes ao histórico de inadimplência dos anos de 2004 e 2005.

Utilizando mais uma vez a opção *stepwise* para a seleção das variáveis mais significantes foi obtido o seguinte resultado:

**Tabela 2 : Resultado do modelo de behavioral scoring - regressão logística**

	Coeficiente	Significância
INSTPAI – Nível de instrução do pai	-0,3176	0,0385
INSTMAE – Nível de instrução da mãe	-0,3806	0,0091
FUNCM – Função que a mãe exerce	0,9875	0,0029
CONTRATO – Contrato de prestação de serviço com a escola	-1,4068	0,0070
DUMMY05 – Presença de débitos no ano de 2005	2,7873	0,0000
CONSTANT	1,7070	0,0844

Dentre as variáveis que compuseram o modelo, a instrução do pai e da mãe, a função da mãe e a presença de contrato se mostraram novamente significantes e apresentaram os mesmos sinais.

Por sua vez, a variável relacionada ao histórico de débitos do ano de 2005 tornou-se significativa, sendo incorporada pelo modelo. Como esperado o coeficiente apresentou sinal positivo, indicando que a probabilidade de adimplência aumenta no caso da existência de débitos na instituição no ano de 2005. É válido ressaltar que a nova variável apresenta o maior coeficiente dentre as variáveis independentes. O que significa dizer que os indivíduos que possuem um histórico de débitos no ano de 2005 têm uma forte tendência a incorrer em atrasos no ano subsequente.

A própria estrutura do modelo de regressão logística é voltada para lidar com a variável dependente dicotômica, como no caso em que esta sendo trabalhado. Sendo assim, o próprio valor resultante da multiplicação das variáveis selecionadas pelos seus respectivos pesos discriminantes fornece uma pontuação, que ao ser aplicada a função de distribuição logística fornece a probabilidade da ocorrência do evento.

Função de distribuição logística:

$$P_i = \frac{1}{1 + e^{-Z}}, \text{ onde } Z_i \text{ indica o escore de cada indivíduo.}$$

Deste modo o valor resultante da aplicação da equação acima anuncia a probabilidade direta de que determinado indivíduo pertença ou não ao grupo dos inadimplentes.

A previsão da inadimplência, contrariamente a adimplência é resultado da opção por atribuir o valor de 1 aos casos inadimplentes, informando ao modelo a presença deste atributo e atribuindo 0 aos adimplentes informando a ausência de inadimplência.

Sendo assim valores altos da probabilidade indicam que o modelo prevê a situação de indivíduos como inadimplentes. Por outro lado, baixos valores de probabilidade indicam um distanciamento da previsão de inadimplência, que será interpretada como uma aproximação da adimplência.

A questão chave então se torna decidir o limite a partir do qual a probabilidade é considerada como inadimplente e quando será adimplente. Tomando-se como referência que a probabilidade de 0,5 indica chances iguais de obtenção de um resultado quando se trata de duas escolhas (HAIR et al, 2002), optou-se por atribuir o status de inadimplente as previsões com probabilidade superior a 0,5 e adimplentes caso contrário.

### 5.1.1 Matrizes de classificação dos modelos de Regressão Logística

Nas tabelas 3 e 4, apresentados a seguir, são demonstradas as matrizes de classificação dos modelos de aprovação de crédito e behavioral scoring baseados na regressão logística:

**Tabela 3 : Matriz de classificação aprovação de crédito - regressão logística.**

	Resultado Original	Previsão do Modelo	% Corretamente classificado
Inadimplentes	15	7	46,67%
Adimplentes	46	42	91,30%
Classificação correta total: 80,33%.			

**Tabela 4: Matriz de classificação behavioral scoring - regressão logística.**

	Resultado Original	Previsão do Modelo	%Corretamente classificado
Inadimplentes	15	11	73,33%
Adimplentes	46	43	93,47%
Classificação correta total: 88,52%.			

O modelo de aprovação de crédito com regressão logística apresentou desempenho na previsão dos dados de maneira satisfatória, onde 80,33% dos casos foram classificados de maneira correta. Este percentual pode ser classificado como suficiente, pois supera o critério de acerto por chances proporcional, o qual permite uma classificação correta de pelo menos 62,50%, superando também o critério de chance máxima de 75%, que indica o percentual de acerto que se pode obter ao classificar todas as observações como pertencentes a um único grupo.

De todo modo, a comparação com o critério de chances proporcional é mais indicado para este estudo, pois procura-se um modelo que seja capaz de identificar corretamente o maior número possível de indivíduos dentro dos dois grupos distintos, por isso o modelo pode ser considerado como válido.

Analisando o desempenho dentro dos grupos, percebe-se que o modelo de aprovação de crédito obteve um fraco desempenho na classificação dos inadimplentes onde apenas 46,67% dos indivíduos pertencentes a este grupo foram identificados corretamente. Em contra partida 91,30% dos casos de adimplentes são classificados corretamente.

Por sua vez, o modelo *behavioral scoring* apresenta um percentual de acerto global de 88,52% dos casos, superando o critério de acerto por chances máxima e proporcional.

Analisando de maneira mais específica é possível perceber que o percentual de acerto dentro dos grupos é melhor que o modelo de aprovação de crédito obtido com a regressão logística, obtendo 93,47% de acerto no grupo de adimplentes e 73,33% nos inadimplentes.

Para se certificar de que o poder discriminatório da matriz de classificação é estatisticamente superior ao critério por chances, foi empregado o teste Q de Press.

Considerando que o número de observações na amostra de teste somam 61 dados, entre os quais 49 foram classificados corretamente no modelo de aprovação de crédito e 54 no modelo comportamental, e existem apenas os grupos de adimplentes e inadimplentes, o teste Q de Press foi calculado da seguinte maneira:

Aprovação de crédito – Regressão Logística.  

$$Q \text{ de Press} = \frac{[61 - (49*2)]^2}{61*(2-1)} = 22,44$$

Behavioral - Regressão Logística.  

$$Q \text{ de Press} = \frac{[61 - (54*2)]^2}{61*(2-1)} = 36,21$$

O valor tabelado de 3,84 encontrado para um nível de confiança de 95% é bem inferior aos valores encontrados no teste. Sendo assim, é possível considerar que a matriz de classificação obtida através da regressão logística é estatisticamente superior ao nível de acertos obtidos ao acaso, em ambos os modelos.

## 5.2 Redes neurais artificiais

O desenvolvimento dos modelos de *credit scoring* baseados nas redes neurais artificiais exigiu algumas modificações em relação à seleção da amostra, assim como foram necessários estabelecer novos parâmetros inerentes às redes neurais artificiais e que serão explicitados a seguir.

A operacionalização das RNA's é desmembrada nas etapas de treinamento, validação e teste. Sendo assim foi necessário redimensionar a partição da amostra que havia sido utilizada anteriormente na regressão logística. A fim de preservar as características da amostra para efeito de comparação entre os modelos que utilizam diferentes técnicas, os dados resguardados para amostra de testes foram mantidos e os demais dados alocados para treinamento e validação. O resumo da nova partição dos dados é apresentado na tabela 5.

**Tabela 5. Partição das amostras para RNA**

	% na amostra completa	Casos na amostra completa	Casos reservados para treinamento	Casos reservados para validação.	Casos reservados para teste.
Adimplentes	75%	183	91	46	46
Inadimplentes	25%	62	32	15	15
Total	100%	245	123	61	61

A amostra de treinamento conta com 50% das observações, pois é uma fase crítica onde os dados estão sendo apresentados pela primeira vez à rede. Os demais 25% foram destinados para a fase de validação dos dados.

Não existe uma regra que estabeleça o limite ótimo de dados que devam constar em cada etapa, sendo assim a opção por esta segregação das observações segue o mesmo padrão observado na composição da amostra completa. Deste modo, a composição de inadimplentes e adimplentes dentro de cada uma das três fases de construção do modelo segue aproximadamente a mesma proporção em cada uma das etapas, onde 25% dos dados referem-se a inadimplentes e 75% a adimplentes.

### 5.2.1 Rede Neural – Parâmetros dos modelos

O modelo desenvolvido a partir da técnica de rede neural baseou-se no mecanismo mais difundido entre os estudos de *credit scoring* voltados para a previsão da inadimplência, denominado de *perceptron* multicamadas.

O ponto de partida para a geração dos modelos foi o estabelecimento da quantidade de camadas ocultas a serem utilizadas, e o número de neurônios em cada camada. Objetivando a criação de redes neurais com o maior poder de explicação possível do fenômeno da inadimplência, mas prezando pela parcimônia do modelo, as redes foram desenvolvidas de modo incremental.

A princípio foram desenvolvidas redes com apenas uma camada de entrada, uma oculta e uma de saída, e posteriormente novos testes foram realizados com o acréscimo de mais uma camada oculta. A opção por apenas duas camadas ocultas deve-se ao fato de que estas são suficientes para aproximar qualquer função matemática (Cybenko, *apud* Braga et al, 2000). Raciocínio análogo foi utilizado para o estabelecimento do número de unidades de processamento. Iniciando com dois neurônios, novas unidades foram acrescentadas na tentativa de identificar modelos que obtivessem melhor desempenho na amostra de teste. A adição dos neurônios ocorreu até que se atingiu o número máximo de 10 nodos. A escolha deste limite baseou-se no fato de que o acréscimo excessivo de neurônios pode causar o problema de *overfitting*, onde informações imprecisas são assimiladas, prejudicando a capacidade de generalização ou o reconhecimento de padrões não vistos durante o treinamento. Além do mais foi possível notar que o aumento da complexidade da rede, decorrente do crescimento de nodos, não foi capaz de conferir desempenho superior ao obtido pelas redes mais simples.

Seguindo o procedimento sugerido por Desai et al (1996), cada rede foi elaborada 10 vezes, onde a cada uma delas foram atribuídos pesos aleatórios as conexões entre neurônios.

O objetivo deste procedimento foi atenuar os efeitos das redes neurais que não convergiram adequadamente. Sendo assim dentre as 10 vezes que cada rede foi gerada apenas aquela que obteve melhor desempenho foi considerada como válida e utilizada na análise comparativa entre modelos.

O aprendizado da rede ocorreu de modo supervisionado, onde foram apresentadas as variáveis de entrada e o respectivo resultado esperado. Deste modo o algoritmo de aprendizado de *back-propagation* apresentou-se com viável, pois foi possível calcular o erro obtido em cada ciclo e compará-lo com o desejado, permitindo que essa diferença fosse repassada na forma de insumo para as camadas iniciais.

O número de épocas utilizado nas etapas de treinamento e validação foi limitado a mil, e a taxa de aprendizado foi de 0,3, ou seja, todo o processo de entrada dos dados até a realimentação através do erro ocorreu mil vezes e os pesos foram alterados de maneira moderada. A opção por estas escolhas, mais uma vez, esteve baseada na percepção empírica, uma vez que não foram identificadas regras estabelecendo limites para a escolha ideal destes parâmetros.

Durante a fase de elaboração das primeiras redes foi possível perceber que o ponto de convergência de menor erro ocorria antes da milésima época, o que demonstrou que a utilização de um número de épocas superior a mil seria improdutivo e consumiria bastante tempo. A tabela a seguir apresenta a síntese dos melhores resultados obtidos na geração das redes.

A medida que as redes começaram a ser construídas foi possível perceber que apesar do aumento da complexidade da rede não houve progressos significativos na melhora da performance da previsão que justificasse a inclusão de mais variáveis, em uma contínua busca por novas redes que proporcionassem melhores resultados.

O melhor nível de previsão obtido no modelo de aprovação de crédito atingiu 86,89% de acerto quando a rede ainda estava sendo construída com uma camada inicial com 18 entradas, relativas as 18 variáveis utilizadas, e apenas uma camada oculta com 4 neurônios.

No modelo comportamental foram introduzidas das variáveis defasadas referentes aos débitos de 2006 e 2005. Contando agora com vinte variáveis a melhor previsão do modelo atingiu o nível de 88,52% quando a rede foi construída com uma camada de entrada, duas ocultas e uma de saída. Na primeira camada oculta foram inseridos seis neurônios, e 10 neurônios na segunda camada oculta.

Apesar da simplicidade das redes foram obtidos bons desempenhos na previsão da amostra de teste como será demonstrado posteriormente. Resultados satisfatórios como este, utilizando redes simples com apenas uma camada oculta, também foram encontrados pelos autores Lee e Jung (2000) e Castro Junior (2003), utilizando 7 e 5 neurônios respectivamente.

### 5.2.2 Matrizes de classificação do modelo de rede neural

Nas tabelas 6 e 7 demonstrados a seguir apresentam de maneira resumida os resultados obtidos com a utilização das redes neurais através das matrizes de classificação.

**Tabela 6: Matriz de classificação aprovação de crédito – redes neurais**

	Resultado Original	Previsão do Modelo	% Corretamente classificado
Inadimplentes	15	7	46,67%
Adimplentes	46	46	100%
Classificação correta total: 86,89%.			

**Tabela 7: Matriz de classificação behavioral scoring - redes neurais**

	Resultado Original	Previsão do Modelo	% Corretamente classificado
Inadimplentes	15	11	73,33%
Adimplentes	46	43	93,48%
Classificação correta total: 88,52%.			

Como é possível verificar o modelo que abrange variáveis comportamentais obteve desempenho global superior ao apresentado no modelo de aprovação de crédito utilizando a mesma técnica, onde o primeiro atingiu 88,52% contra 86,89% do segundo.

O desempenho superior do modelo comportamental também foi observado na previsão de cliente inadimplentes, atingindo 73,33% contra 46,67% do modelo de aprovação. O modelo de aprovação de crédito mostrou-se superior apenas quanto a classificação dos clientes adimplentes, alcançando 100% de acerto em contra partida o modelo comportamental acertou 93,48%.

Considerando que o custo envolvendo uma má classificação dos clientes inadimplentes é superior a uma má classificação dos adimplentes o modelo *behavioral scoring* pode seguramente ser considerada como detentor de melhor desempenho, pois apresenta melhor poder de previsão global e melhor previsão do grupo inadimplente.

No intuito de verificar se as matrizes de classificação dos modelos de RNA são significativamente maiores que o critério aleatório de chances foi utilizado o teste Q de Press, resultando nos seguintes valores:

$$\begin{aligned} &\text{Aprovação de crédito - RNA} \\ \text{Q de Press} &= \frac{[61 - (53*2)]^2}{61*(2-1)} = 33,20 \end{aligned}$$

Behavioral Scoring - RNA

$$Q \text{ de Press} = \frac{[61 - (54*2)]^2}{61*(2-1)} = 36,21$$

Avaliando a significância dos dados a um nível de 95% confiança é possível perceber que o valor obtido no teste é bastante superior ao valor tabelado de 3,84. Deste modo é possível afirmar que ambos os modelos possuem poder de previsão significativamente superior ao critério de chances.

## 6 Conclusões

A pontuação de crédito não é uma modelagem recente, os primeiros estudos datam de 1930, no entanto novas técnicas têm sido desenvolvidas e apresentado bons resultados como é o caso das redes neurais artificiais, permitindo, conjuntamente com o desenvolvimento das técnicas computacionais, a disseminação de sua utilização para segmentos econômicos ainda pouco explorados.

Neste estudo a viabilidade do *credit scoring* foi testada através de duas técnicas, são elas a regressão logística e redes neurais artificiais.

Os dados utilizados no estudo foram retirados da base de dados concedida pela instituição pesquisada, contando com a uma amostra final de 245 indivíduos. Dentre os dados fornecidos pela instituição educacional foram identificadas 20 variáveis que potencialmente poderiam contribuir na previsão do fenômeno em estudo, a inadimplência.

Após uma revisão das principais obras de referência sobre a inadimplência foi possível constatar que não existe um consenso na literatura quanto as variáveis que melhor contribuem para a previsão dos modelos de *credit scoring*. Deste modo todas as variáveis disponíveis foram utilizadas no desenvolvimento do estudo, sendo a opção por sua utilização, ou não, no modelo final condicionada a significância em relação ao poder de previsão.

Para verificar o poder de previsão, assim como a capacidade de generalização dos modelos a amostra foi particionada em dados de validação e teste. Os primeiros foram utilizados no desenvolvimento dos modelos enquanto que os últimos utilizados para averiguar o desempenho ao realizar as previsões.

Durante o desenvolvimento da pesquisa a utilização da regressão logística apresentou-se como um método bem ajustado ao modelo, pois é uma técnica direcionada para a utilização de dados dicotômicos como variável dependente, além de não estabelecer pressupostos sobre as variáveis explicativas.

O *credit scoring* desenvolvido com a regressão logística demonstrou que as observações corretamente classificadas alcançaram 80,33% no modelo de aprovação de crédito e 88,52% no *behavioral scoring*

A estrutura flexível subjacente ao modelo de redes neurais também se apresentou como uma técnica com alta capacidade preditiva no desenvolvimento do *credit scoring*, onde o modelo de aprovação de crédito alcançou 86,89% de acerto em sua previsão geral, enquanto que o *behavioral scoring* atingiu o patamar de 88,52% de casos identificados corretamente.

Baseando-se nos resultados obtidos é possível afirmar que todos os modelos elaborados apresentaram-se como instrumentos viáveis na previsão de inadimplência, uma vez que seu poder de previsão, quanto ao status de pagamento de clientes, é superior ao critério de identificação baseado em chances proporcionais, o qual permite um acerto mínimo de 62,50%.

Ficou evidente também que em todos os modelos que incorporaram a variável comportamental houve um aumento na capacidade preditiva, demonstrando a superioridade do *behavioral scoring* sobre os modelos de aprovação de crédito.

Analisando separadamente o poder de previsão do status de pagamento dentro dos grupos, é possível perceber que os modelos obtiveram melhor desempenho ao identificar clientes considerados inadimplentes em comparação aos adimplentes.

As redes neurais artificiais e a regressão logística apresentaram a mesma capacidade preditiva geral no modelo comportamental, além de obter os mesmos percentuais de previsão dentro dos grupos, com 73,33% de acerto entre inadimplentes e 93,48% entre os adimplentes.

Trazendo a discussão do desempenho dos modelos para a aplicação no contexto da instituição educacional pesquisada, é válido perceber que a possível utilização de cada uma das técnicas deve considerar as limitações presentes na instituição.

O fato das redes neurais não apresentarem nenhum tipo de restrição ao formato dos dados, associados a sua capacidade de identificar relações não lineares entre as variáveis, é um fator que torna sua utilização atrativa. Porém, deve-se ponderar que seu desenvolvimento e utilização por parte da instituição de educação pesquisada exigem um maior nível de complexidade, tornando mais restrita a possibilidade da adoção desta técnica pela instituição.

Por outro lado, a regressão logística possui um mecanismo de elaboração mais difundido e disponível em maior número de pacotes estatísticos. Associado a este fator está a maneira intuitiva com que os resultados podem ser interpretados, onde o logit oferece o percentual direto de probabilidade de inadimplência, o que torna esta técnica mais atraentes. Aliado a este fato está a possibilidade de obtenção de um desempenho preditivo tão bom quanto o modelo de rede neural artificial.

De uma maneira geral é possível perceber, através dos resultados da pesquisa, que diante da existência de uma legislação que protege os que se tornam inadimplentes em relação a contraprestação dos serviços educacionais, e que limita os esforços das instituições de educação em reaver os pagamentos pendentes, os modelos de *credit scoring* apresentam-se como uma alternativa viável no objetivo de tentar amenizar os efeitos do risco de crédito, inerente ao desempenho da prestação dos serviços educacionais.

Deste modo, os modelos de *credit scoring* demonstraram ser de grande valia a medida que se apresentaram como instrumento de previsão viável, permitindo sua utilização como instrumento auxiliar na seleção de potenciais clientes ou ainda como uma ferramenta de apoio ao planejamento financeiro.

### Referências Bibliográficas

- ALMEIDA, F. C.; SOUZA, PASSARI, F. A. Aplicação de Redes Neurais na previsão de vendas no varejo. In: XXXIV Encontro Anual Da Associação Nacional De Pós-Graduação Em Administração. **Anais...**, Brasília: ANPAD, 2005.
- ALMEIDA, F. C.; SOUZA, PASSARI, F. A. Aplicação de Redes Neurais na previsão de vendas no varejo. In: XXXIV Encontro Anual Da Associação Nacional De Pós-Graduação Em Administração. **Anais...**, Brasília: ANPAD, 2005.
- ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate failure. **Journal of Finance**, n. 23, p. 589-609, 1968.
- \_\_\_\_\_; HALDEMAN, Robert G.; NARAYANAN, Paul. ZETA Analysis: a new model to identify bankruptcy risk of corporations. **Journal of Banking and Finance**, n. 1, p. 29-54, 1977.
- AMORIN NETO, A. **Modelagem do Risco de Crédito: Um Estudo do Segmento de Pessoas Físicas de um Banco de Varejo**. 2002. 103f. Dissertação (Mestrado em Administração) – Programa de Pós-Graduação em Administração, Universidade Federal de Pernambuco, Recife, 2002.
- ASSAF NETO, A. **Mercado Financeiro**. 5 ed. São Paulo: Atlas, 2003.

- BEAVER, W. H. Financial Ratios as Predictors of Failure. **Empirical research in accounting: selected studies**, suplemento do **Journal of Accounting Research**, autumn, 1966.
- BRAGA, A. P.; LUDEMIR, T. B.; CARVALHO, A. C. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 2000.
- CAOUCETTE, J. ALTMAN, E; NARAYANAM, P. **Gestão do Risco de Crédito: o Próximo Grande Desafio Financeiro**. Rio de Janeiro: Qualitymark, 1999.
- CASTRO JR, F. H. F. **Previsão de Insolvência de Empresas Brasileiras Usando Análise Discriminante, Regressão Logística e Redes Neurais**. 2003. 169f. Dissertação (Mestrado em Administração) - Programa de Pós Graduação em Administração, Faculdade de Economia, Administração e Contabilidade, São Paulo, 2003.
- CIA, J.C. Propostas de Medidas de Inadimplência para o Mercado Brasileiro. In: XXXII Encontro Anual da Associação Nacional de Pós-Graduação em Administração, **Anais...**, Atibaia: ANPAD, 2003.
- DESSAI, V.S.; CONWAY, J.N.; OVERSTREET, G.A. A comparison of neural network and linear scoring models in the credit union environment. **European Journal of Operational Research**. v. 95, p. 24-37, 1996.
- GIL, A. C. **Métodos e Técnicas de Pesquisa Social**. 4. ed. São Paulo: Atlas, 1994.
- HAIR JR., J. F.; ANDERSON, R. E.; TATHAM, R. L.; BLACK, W. C. **Análise Multivariada de Dados**. 5 ed. Porto Alegre: Bookman, 2002.
- JORION, P. **A Nova Fonte de Referência para o Controle do Risco Financeiro: Value at Risk**. São Paulo: BM&F, 2003.
- LEE, T. H.; JUNG, S. Forecasting creditworthiness: logistic vs. artificial neural net. **The Journal of Business Forecasting Methods & Systems**, 1999/2000, winter, v.18, n. 5, p. 28-30.
- MARTINS, G. A.; LINTZ, A. **Guia para Elaboração de Monografias e Trabalho de Conclusão de Curso**. São Paulo: Atlas, 2000.
- OHLSON, James A. Financial ratios and the probabilistic prediction of bankruptcy. **Journal of Accounting Research**, v. 18, n. 1, p. 109-131, spring, 1980.
- PRADO, R.G. A; BASTOS, N.T; DUARTE JÚNIOR, A. M; Gerenciamento de Risco de Crédito em Bancos de Varejo no Brasil. São Paulo, **Tecnologia de Crédito**, n.43, p.7-30, Julho, 2000.
- ROSEMBERG, E.; GLEIT, A. Quantitative Methods in Credit Management: a Survey. **Operational Research**. v. 42, n. 4, p. 589-613, Jul/Aug, 1994.
- SANTOS, J. O. dos. **Análise de Crédito: Pessoas Físicas e Jurídicas**. 2 ed. São Paulo: Atlas, 2003.
- SECRETARIA ESTADUAL DE EDUCAÇÃO. Disponível em <<http://educação.pe.gov.br>>. Acesso em: 18 fev. 2006.
- SECURATO, J.R. (Coord.) **Crédito: Análise e Avaliação do Risco: Pessoas Físicas e Jurídicas**. São Paulo: Saint Paul, 2002.
- SILVA, J. P. **Gestão e Análise do Risco de Crédito**. São Paulo: Atlas, 1998.
- SINDICATO DAS ESCOLAS PARTICULARES DE PERNAMBUCO. Disponível em <<http://www.sinepe.com.br>>. Acesso em: 11 fev, 2006.
- SINDICATO DOS ESTABELECIMENTOS DE ENSINO DO ESTADO DE PERNAMBUCO. Disponível em <<http://www.sinepe.com.br>>. Acesso em: 17 fev. 2006.
- THOMAS, Lyn C. A survey of credit and behavioral scoring: forecasting financial lending to consumer. **International Journal of Forecasting**. n. 16, p. 149-172. 2000.
- VASCONCELLOS, M.S. **Proposta de método para análise de concessão de crédito a pessoas físicas**. 2002. 142f. Dissertação (Mestrado Economia) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2002.

VERGARA, Sylvia C. **Projetos e Relatórios de Pesquisa em Administração**. São Paulo: Atlas, 1997.

WEST, D. Neural Network Credit Scoring Model. **Computers & Operations Research**, v.27, p. 1131-1152, 2000.