



Revista ADM.MADE

Revista do Mestrado em Administração e
Desenvolvimento Empresarial - Universidade
Estácio de Sá

Revista ADM.MADE, ano 9, v.13, n.1, p. 38-56, janeiro/abril, 2009

Revista do Mestrado em Administração e Desenvolvimento Empresarial da Universidade Estácio de Sá – Rio de Janeiro (MADE/UNESA). ISSN: 1518-9929
Editora responsável: Isabel de Sá Affonso da Costa

Criação de Inteligência Competitiva a Partir de Regressão Logística: Estudo no Setor de Telecomunicações

Aline Gonçalves Lins¹

Wesley Vieira da Silva²

Carlos Olavo Quandt³

Artigo recebido em 23/08/08. Aceito em 30/03/09. Artigo submetido a avaliação *double-blind*.

¹ Mestranda pelo Programa de Pós-Graduação em Administração da Pontifícia Universidade Católica do Paraná (PPAD/PUC-PR). Endereço: Pontifícia Universidade Católica do Paraná – PPAD - Rua Imaculada Conceição, 1155 - Bloco Acadêmico, sala 217 - Prado Velho - Curitiba - PR - CEP: 80215-901. Email: ALINE.LINS@gvt.com.br.

² Professor Doutor do Programa de Pós-Graduação *Stricto Sensu* em Administração da Pontifícia Universidade Católica do Paraná (PPAD/PUC-PR). Email: wesley.vieira@pucpr.br.

³ Professor Doutor do Programa de Pós-Graduação *Stricto Sensu* em Administração da Pontifícia Universidade Católica do Paraná (PPAD/PUC-PR). E-mail: carlos.quandt@pucpr.br.

Criação de Inteligência Competitiva a Partir de Regressão Logística: Estudo no Setor de Telecomunicações

A crescente complexidade do ambiente empresarial na Nova Economia propicia o surgimento de diferentes configurações de oportunidades e de ameaças e, conseqüentemente, de novos desafios às empresas. Desta forma, a importância estratégica da transformação de dados e de informações em inteligência organizacional passou a determinar o nível de competitividade das empresas. Nesse contexto, o processo de criação de uma sistemática de inteligência competitiva dentro da organização assume importância fundamental. O objetivo deste trabalho é analisar a utilização da técnica multivariada de regressão logística como auxílio ao desenvolvimento de conhecimentos diferenciais, a partir do estudo de caso na prevenção de fraudes em uma empresa de telecomunicações brasileira. Os resultados demonstram que a ferramenta permite uma melhoria de eficiência nos processos de auditoria de fraude, desenvolvendo, assim, diferencial importante para a empresa dentro do seu setor de atuação.

Palavras-chave: inteligência competitiva; telecomunicações; regressão logística; fraudes.

Keywords: competitive intelligence; telecommunications; logistic regression; fraud.

Competitive Intelligence Creation Based on Logistic Regression: A Case Study in the Telecommunications Sector in Brazil

Under the increasing complexity of the business environment in the New Economy, different configurations of opportunities and threats have emerged and, consequently, new challenges to companies have arisen. Hence the strategic importance of the transformation of data in organizational intelligence has become crucial to determine the level of competitiveness of companies. In this context, the creation of a systemic competitive intelligence process within the organization assumes fundamental importance. This paper examines the use of multivariate logistic regression technique as a tool to support the development of specific knowledge, through a case study on prevention of fraud in a Brazilian telecommunications company. The results show that the tool improves efficiency in the processes of auditing potential fraudulent cases, thus providing an important advantage for the company within its sector.

1. Introdução

A nova economia que surgiu em escala global no último quarto do século XX possui características fundamentais e diferenciais, segundo Castells (1999), por ser uma economia informacional, globalizada e em rede. Este trabalho aborda um aspecto fundamental da característica informacional desta nova economia, que é a dependência entre a produtividade e a competitividade das organizações, e sua capacidade de gerar, de processar e de aplicar a informação de forma eficiente, e transformá-la em conhecimento.

A crescente complexidade identificada nos ambientes organizacionais interno e externo amplia a dificuldade na elaboração da estratégia empresarial. Dessa forma, a compreensão e a habilidade no manuseio de informações se tornaram cruciais na vida das empresas (BEUREN, 2000). A necessidade latente que as empresas têm de se manter competitivas frente ao mercado consumidor dá extrema importância estratégica ao

processo de inteligência competitiva (VALENTIN et al., 2003). Sendo assim, uma organização mais sofisticada e integrada será a resposta para a velocidade e a complexidade da Nova Economia. Um alto nível de coordenação é necessário para agregar valor e para criar riqueza a partir das atividades da organização (SAPIRO, 1993).

Com o aumento da importância da informação no mundo científico e comercial, o conhecimento tornou-se um grande diferencial, auxiliando desde a tomada de decisão até a descoberta de fraudes ou o mapeamento de perfis de consumidores. No entanto, para que estas informações sejam extraídas corretamente, é necessária a utilização de técnicas e de ferramentas que propiciem a descoberta ou mineração de padrões (ALMEIDA et al., 2003). Nesse contexto, as empresas estão usando a informação e os sistemas de informação como ferramentas para obter vantagens competitivas em relação a seus competidores diretos.

Este trabalho aborda especificamente a questão do processo de criação de inteligência organizacional utilizando-se de ferramenta de análise multivariada - a técnica de regressão logística - que apresenta êxito considerável na transformação de dados em conhecimento. Para tanto, utilizou-se estudo de caso no setor de telefonia fixa brasileira, onde a ferramenta aparece como aliada na detecção do perfil do usuário fraudulento.

O trabalho encontra-se estruturado em cinco seções: a primeira trata da parte introdutória; a segunda apresenta as principais bases conceituais do trabalho acerca da inteligência competitiva e do seu processo de formação, e uma contextualização sobre o setor de telefonia fixa no Brasil; a terceira refere-se à metodologia empregada no trabalho; a quarta refere-se à apresentação e à análise dos dados; e a quinta trata das considerações finais e recomendações.

2. Referencial Teórico-Empírico

Nesta seção, tem-se a apresentação das principais bases conceituais no trabalho acerca da inteligência organizacional, seu processo de criação, a aplicação da técnica de regressão logística visando à transformação de dados em conhecimento, e um breve panorama do setor de telecomunicações.

2.1. Inteligência organizacional

De acordo com Almeida e Motta (1997), com a consolidação das tendências do desenvolvimento da sociedade da informação e da globalização da economia nos últimos anos, configuram-se novas oportunidades e novas ameaças para as organizações. A capacidade de oferecer produtos e serviços de alta qualidade e de baixo custo, inovadores e adaptados ao meio ambiente, tomaram papel fundamental na vida das empresas. Com isso, tornou-se essencial a adequação das organizações ao novo ambiente de negócios, como forma de garantir suas vantagens competitivas e, mesmo, sua sobrevivência.

Sapiro (1993) explica que as organizações modernas são fundamentadas na informação e no conhecimento, e que estão buscando desesperadamente adaptarem-se aos novos tempos. Neste contexto, as empresas não terão escolha a não ser lastrear suas atividades na informação.

Porter (2000) afirma que a tecnologia de sistemas de informação é particularmente penetrante na cadeia de valores, visto que cada atividade de valor cria e utiliza

informação. Segundo Beuren (2000), a decisão do posicionamento estratégico da empresa, que é a definição de como enfrentar os concorrentes, depende de informações sobre os ambientes interno e externo à organização, pois são elas que propiciarão identificar as ameaças e oportunidades, criando o cenário para uma resposta eficaz e competitiva.

Sapiro (1993) mostra, com base em vários estudos empresariais, que a falta de habilidade de monitoramento das informações por parte de organização resulta em desempenho fraco e em atraso competitivo. Segundo o autor, a formulação estratégica é derivada das informações disponíveis e, conseqüentemente, o nível de aprofundamento dessas está diretamente ligado a estratégias de sucesso.

Hall (2004) também enfatiza que a tomada de decisões é o processo pelo qual as empresas são estruturadas e reestruturadas. Para o autor, o fator obviamente crítico da tomada de decisões é a informação: a quantidade e o tipo de informações disponíveis é que determinam o grau de acerto e a dificuldade no momento em que elas são tomadas.

Castells (1999) reforça que, sem dúvida, a informação e o conhecimento sempre foram cruciais no crescimento da economia. Porém, na Nova Economia a geração de conhecimentos e a capacidade tecnológica passam a ser ferramentas fundamentais para a concorrência entre organizações de todos os tipos.

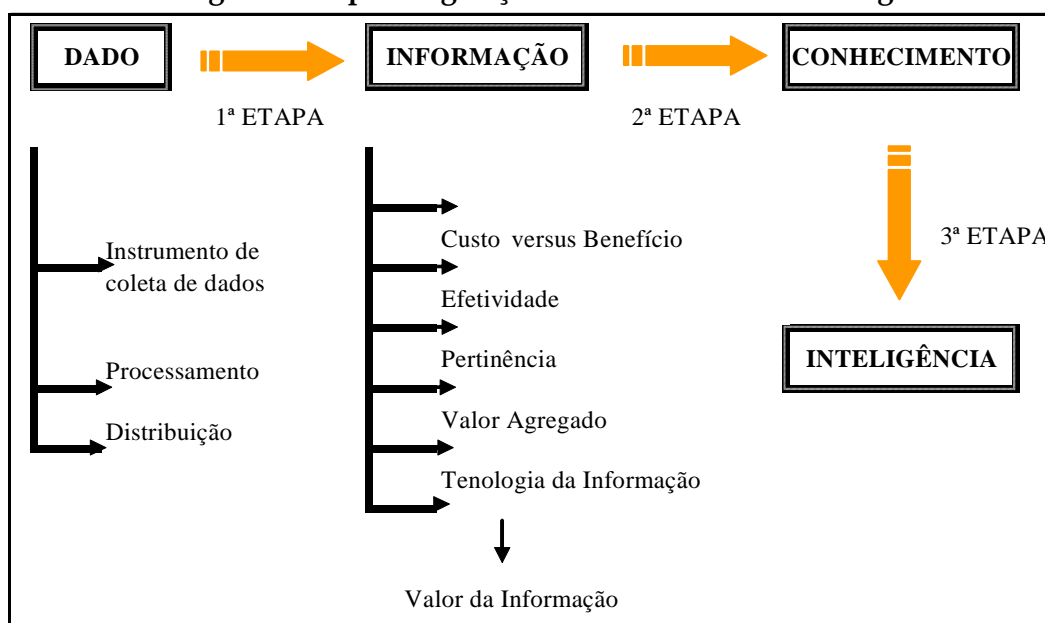
A inteligência competitiva, segundo Valentin et al. (2003), pode ser entendida como um processo dinâmico, composto pela gestão da informação e pela gestão do conhecimento. É o processo que investiga o ambiente onde a empresa está inserida, com o propósito de descobrir oportunidades, de reduzir os riscos e de diagnosticar o ambiente interno, visando ao estabelecimento de estratégias de ação a curto, médio e longo prazos.

O termo “inteligência organizacional” remete à capacidade das organizações de monitorar informações ambientais para responder satisfatoriamente aos desafios e às oportunidades que se apresentam. Nesse sentido, é possível afirmar que a inteligência visa principalmente, a imprimir comportamento adaptativo à organização, permitindo que essa realize mudanças e adapte seus objetivos, seus produtos e seus serviços em resposta a novas demandas do mercado e a mudanças no ambiente (TARAPANOFF et al., 2000). Para esses autores, o ato de possuir inteligência organizacional está associado à busca sistemática, efetiva e pró-ativa de posturas ligadas à estratégia, à relação entre organização e ambiência externa. Sendo assim, o processo de geração de inteligência nas organizações está no centro de muitos estudos do campo de administração.

A Figura 1 esquematiza o processo de geração de conhecimento e de inteligência organizacional, a partir do prisma da informação. Nela, é possível verificar que a inteligência organizacional é fruto de um processo de transformação dos dados disponíveis.

Davenport e Prusak (1998) conceituam os elementos desse processo: **dados** são simples observações sobre o estado do mundo, facilmente estruturados, obtidos por máquinas, freqüentemente quantificados e facilmente transferidos; **informação** são dados dotados de relevância e de propósito, requerem unidade de análise, exigem consenso em relação ao significado e necessariamente exigem a mediação humana; **conhecimento** é a informação valiosa da mente humana, inclui reflexão, síntese e contexto, é de difícil estruturação, transferência e captura em máquinas, bem como é freqüentemente tácito. O desafio da criação de uma inteligência competitiva passa, então, a ser o de transformar dados em conhecimento e obter sua aplicação prática.

Figura 1: Etapas da geração de conhecimento e inteligência



Fonte: Adaptado de Tarapanoff et al. (2000, p. 91).

Os trabalhadores têm autoridade para, por meio de qualquer informação, produzir um novo conhecimento que poderá ser utilizado por eles mesmos ou por outras áreas da empresa. E, uma vez que o conhecimento é criado, há uma provisão inesgotável disso e ele não pode ser consumido (TJADEN, 1996).

Gilad (1991) sugere que as conseqüências positivas da adoção de um sistema de inteligência competitiva podem ser facilmente enumeradas: a) identificação de novas oportunidades de negócios; b) ideais compartilhados; c) crescente habilidade para antecipação de surpresas; d) aumento das habilidades gerenciais; e) integração de vários pontos de vista.

2.2. A criação de inteligência organizacional

No cenário organizacional atual, a crescente competitividade e o incremento exponencial dos dados torna indispensável a utilização de ferramentas estratégicas que possam extrair informações que auxiliem na tomada de decisão com rapidez e eficiência, a partir de grandes bases de dados (DALFOVO et al., 2003). Por outro lado, Tjaden (1996) destaca que, na era da inteligência, a seleção de um correto conjunto de métricas de negócio é o mais crucial e fundamental passo no desenvolvimento de uma ciência dentro da administração de um negócio.

As tecnologias de informação são muito úteis para apoiar o processo de inteligência competitiva, desde a fase de identificação das necessidades de informação, passando pela coleta, análise e disseminação, até a avaliação dos produtos entregues (VALENTIN et al., 2003). Para o processo decisório, as empresas precisam de informações

históricas, que podem ser mais bem visualizadas a partir de algumas técnicas computacionais.

Dalfovo et al. (2003) discutem que as empresas, no geral, possuem informações que proporcionam sustentação para as suas decisões. O que as diferencia é que apenas algumas possuem um sistema estruturado de informações gerenciais, voltadas à gestão ambiental. As companhias que estão neste grupo seguramente possuem vantagem competitiva.

Estudos para o desenvolvimento de meios e de métodos que permitam a eficaz gestão do conhecimento constituem uma tendência no atual mundo competitivo e globalizado. No entanto, o diferencial competitivo é justamente a aplicação da tecnologia certa para cada tipo de demanda. Ou seja, o profissional envolvido no processo de transformação de inteligência em conhecimento ainda representa um importante fator crítico de sucesso para que a adequada tecnologia seja empregada, o que representará investimento correto de tempo, de recursos materiais e financeiros (TARAPANOFF et al., 2000).

2.3. A técnica de regressão logística

A geração de um volume cada vez maior de informações leva à necessidade de maximizar o aproveitamento deste investimento. Provavelmente, a forma mais nobre de se utilizarem esses vastos repositórios seja tentar descobrir se há algum conhecimento escondido neles. Contudo, embora essa idéia básica seja facilmente compreensível, fica a dúvida sobre como um sistema é capaz de obter esse tipo de relação.

A regressão logística binária - ou simplesmente *regressão logística* - é definida como uma técnica estatística de análise multivariada comumente empregada para desenvolvimento de modelos que visem a entender ou a prever a relação existente entre uma variável categórica, que assume um entre dois valores possíveis (ex. **zero** e **um**, **sim** e **não**, etc.), e um conjunto de variáveis explicativas (HOSMER; LEMESHOW, 2000).

A desproporcionalidade entre as observações é, em geral, a responsável pela não normalidade dos dados analisados, o que torna a técnica de regressão logística a mais adequada para este tipo de análise, uma vez que ela não pressupõe a normalidade.

Ao contrário da regressão logística binária, a regressão multinomial pode ser aplicada quando a variável dependente categórica não é necessariamente dicotômica. Para análise das informações, a regressão logística multinomial internamente agrega conjunto de dados para formar subpopulações com padrões de covariância idênticos, facilitando os testes preditivos, o cálculo de residuais, e o teste de ajustamento do modelo *goodness-of-fit* que mede se os valores preditos pelo modelo representam adequadamente os valores observados (SPSS, 2004; HOSMER; LEMESHOW, 2000).

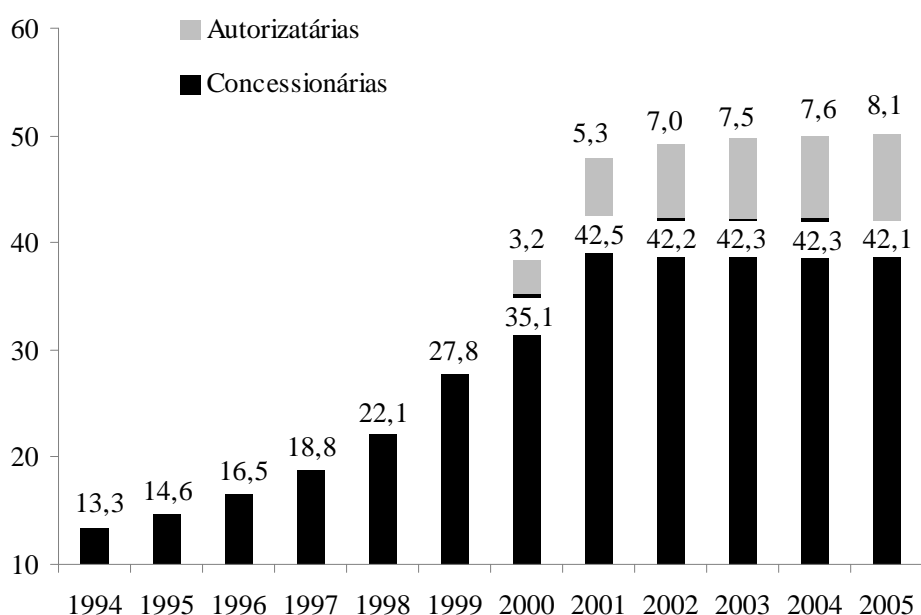
2.4. O setor de telecomunicações

O setor de telecomunicações mundial passou, a partir dos anos de 1990, por fase de rápida evolução devido ao desenvolvimento tecnológico e à disseminação da telefonia celular, da TV a cabo e da Internet. Estes serviços, além de agregarem valor ao consumidor, também tiveram um importante efeito social, na medida em que transformaram muitos hábitos da população (LINS, 2001).

No Brasil, as telecomunicações foram profundamente afetadas pelo processo de privatização do setor. A partir de 1994, foi iniciado esforço no sentido de abrir o setor à iniciativa privada, o que propiciou a entrada de novos investimentos e culminou com a privatização do sistema Telebrás em 1998 e com as licitações para as *autorizatórias* (empresas espelhos) em 1999 (LINS, 2001). Esse cenário proporcionou a rápida expansão da oferta de serviços e o surgimento de novos conceitos na sua comercialização, tais como o serviço pré-pago na telefonia celular, a seleção do provedor pelo assinante através do código na telefonia de longa distância, e o serviço gratuito no acesso à Internet (LINS, 2001). Aliado a este fator, é preciso avaliar que o setor de telecomunicações é aquele, dentre os de infra-estrutura, que apresenta atualmente o maior dinamismo tecnológico. As utilizações potenciais da rede são cada vez mais numerosas (FIUZA; NERI, 1998).

A telefonia fixa, destinada ao uso do público em geral, é prestada no Brasil por detentoras de concessão ou autorizadas para o Serviço Telefônico Fixo Comutado (STFC). Em 2005 existiam em operação, no Brasil, 27 operadoras de STFC, sendo 6 concessionárias - Telemar, BRT, Telefônica, CTBC, Sercomtel e Embratel - e 21 *autorizatórias* - sendo as maiores a GVT e a Intelig (TELECO, 2006).

Figura 2: Evolução do número de acessos fixos instalados no Brasil



Fonte: Teleco (2006)

A Figura 2 apresenta a evolução do número de acessos fixos instalados no Brasil. É possível observar um grande salto ocorrido entre 2000 e 2002, quando as operadoras fizeram grande esforço para antecipar o cumprimento das metas de universalização integrantes dos contratos de concessão.

Esse aumento na prestação do serviço de telefonia, por sua vez, não deixou de trazer consigo a possibilidade de fraude, subproduto dos preços de mercado que

passaram a imperar no novo quadro de oferta destinada a satisfazer a demanda até então insatisfeita devido à incapacidade de investimento estatal. Segundo dados da empresa de tecnologia preditiva SPSS (2007), a maioria das pessoas não aluga apartamento com a intenção de instalar um quadro de distribuição para fazer chamadas telefônicas gratuitas. Contudo, infelizmente, algumas pessoas inescrupulosas fazem isso, causando um prejuízo de bilhões de dólares a essas empresas.

A Neural Technologies, consultoria especializada modelagem preditiva, com base nas informações prestadas pelos seus clientes em todo mundo, estima que entre 3% e 8% do faturamento das empresas do setor de telefonia é perdido em escoamento de receita; em companhias jovens, este valor pode chegar a 10% do faturamento. A fraude foi identificada como a principal causa de perdas de receita no setor de telecomunicações, com prejuízos estimados em 5% do faturamento anual do setor.

Para Jonsson, Lundin e Kvarnström (2000), a fraude pode ser caracterizada como uma decepção ou má-representação em que um indivíduo assume ser falso, sabendo que a decepção pode resultar em benefício não autorizado para si próprio ou para outras pessoas.

É neste cenário competitivo que o presente trabalho procura mostrar como a criação de uma inteligência organizacional pode contribuir para a melhoria dos resultados de uma companhia. Os casos de fraude representam problemas graves, no sentido de que provocam evasão de receitas, gastos adicionais com infra-estrutura e até despesas judiciais. A detecção ágil de um caso fraudador melhora o desempenho da empresa e libera os recursos para investir esforços em posições estratégicas que visem a melhorar o posicionamento competitivo dentro do setor.

3. Metodologia da Pesquisa

Esta seção descreve os aspectos metodológicos utilizados na realização do estudo. Apresenta-se inicialmente a caracterização da pesquisa, seguida pela descrição das variáveis utilizadas no estudo, pela identificação da população e da amostra, e pela explanação dos métodos estatísticos aplicados.

3.1. Caracterização da pesquisa

O objetivo deste estudo é encontrar um modelo econométrico que, por meio das características do cliente, possa identificar eventuais fraudadores em uma empresa de telecomunicações. O método adotado nesta pesquisa é o hipotético dedutivo, a partir do qual se formulam hipóteses que são testadas pelo processo de inferência-dedutiva. A pesquisa proposta é caracterizada como um estudo descritivo, uma vez que procura determinar qual influência as variáveis independentes ou explicativas exercem sobre a variável dependente ou explicada (HAIR JR. et al., 2005).

3.2. Descrição das variáveis

A variável resposta, ou dependente, refere-se ao fato de o cliente ser identificado e comprovado como fraudulento ou não. Desta maneira, a variável é caracterizada pela *dummy TIPO*, onde 0 representa o cliente BOM (sem fraude comprovada) e 1 representa o MAU (com fraude comprovada).

As características integrantes do estudo e classificadas como variáveis independentes ou explicativas são de cunho sócio-demográfico, ou seja, permitem exibir um perfil sócio-demográfico do cliente e de perfil de compra na empresa. São elas:

- Número de linhas (**NLINH**): número de linhas telefônicas adquiridas pelo cliente;
- Tecnologia (**TECNO**): tecnologia utilizada na linha telefônica adquirida pelo cliente. Pode ser **cabo** que é a mais conhecida e utilizada no mercado, ou **wll**, que utiliza ondas de rádio na transmissão da voz;
- Sexo (**SEXO**): gênero informado pelo cliente;
- Tipo do cliente (**TPCLI**): tipo de utilização do serviço de telefonia. Pode ser **residencial** (doméstico) ou **soho** (comercial);
- Possui email cadastrado? (**EMAIL**);
- Possui computador? (**COMPU**);
- Forma de pagamento (**FPGTO**): forma de pagamento solicitada pelo cliente. Pode ser **débito automático** ou **recebe a fatura** para pagamento;
- Data de vencimento (**DVCTO**): data de vencimento da fatura. É solicitada pelo cliente e pode ser nos dias 2, 5, 10, 15, 20 e 28 de cada mês;
- Quantidade de pessoas (**QPESS**): quantidade de pessoas, informado pelo cliente, que irão utilizar a linha telefônica;
- Cidade para envio da fatura (**CDBIL**): cidade onde se situa o endereço de correspondência.
- Profissão informada pelo cliente (**PROF**);
- CEP 2 dígitos da instalação (**CEPSV**): os dois primeiros dígitos do Cep do endereço (sub-região) onde as linhas foram instaladas.

Vale salientar que, ao elevar o nível de assertividade e tratar a qualidade das informações, todos os atributos “Não Informado” e “Outros” foram ignorados na análise estatística.

3.3. População e amostra

O estudo de caso foi realizado em uma empresa de telefonia fixa brasileira. Visando a preservar as informações da companhia e os dados dos clientes, não será divulgado o nome e a região onde ela atua. O universo desta pesquisa refere-se a todos os clientes pessoa física da empresa que tiveram a primeira fatura emitida no período de 01/01/2004 a 31/12/2006. Os dados foram coletados por meio do *Data Warehouse* da empresa, e nos bancos de dados de informações cadastrais e de comportamento de pagamentos. O perfil de interesse para o estudo é o do cliente com FRAUDE COMPROVADA, que representa 0,76% da população.

Seguindo esta proposta, realizou-se a separação dos clientes. Para que a seleção das variáveis e atributos que diferenciam o perfil do cliente fraudador (MAU) e do não fraudador (BOM) não fosse influenciada pela pequena quantidade dentro do total de clientes analisados, foi realizada uma amostragem aleatória dos clientes BONS, de modo

que a proporção entre bons e maus foi forçada à relação 50% / 50%, conforme ilustra a Tabela 1.

Tabela 1: Proporção de clientes na população e na amostra

Tipo de Cliente	População		Amostra	
	Qtde.	%	Qtde.	%
BONS	247.811	99,24%	1.894	50,00%
MAUS	1.894	0,76%	1.894	50,00%
Total	249.705	100%	3.788	100%

Passini e Toledo (2002) sugerem que, para a construção dos modelos, os consumidores que já tenham alguma fraude identificada sejam separados dos que nunca tiveram fraudes identificadas. Posteriormente, são geradas duas tabelas, contendo 100% dos fraudadores e 100% dos não fraudadores.

3.4. Métrica de análise dos dados

A métrica estatística empregada para a modelagem dos dados refere-se à Análise de Regressão Logística (RL), que se aplica à obtenção da probabilidade de que uma observação pertença a um conjunto determinado, em função do comportamento das variáveis independentes (GIMENEZ; OPAZO, 2001). Nesse caso, as variáveis independentes são as variáveis sócio-demográficas registradas para cada cliente, e a variável dependente (ter fraude comprovada) pode tomar o valor compreendido entre 0 e 1.

É importante destacar que a vantagem da utilização do modelo de regressão logística é que ele não requer os pressupostos de normalidade e independência que a regressão linear exige. A expressão algébrica da **regressão logística** (RL) pode ser visualizada por meio da equação (1).

$$y_i = \frac{1}{1 - \exp(\alpha + \beta_i)} + e_i \quad (1)$$

O modelo de RL é estimado pelo valor do *log* de verossimilhança, o qual avalia o nível de ajuste do modelo aos dados disponíveis; ou seja, tenta levantar qual o modelo de onde teriam sido gerados os dados presentes na amostra. O indicador *R² de Cox e Snell* é interpretado como o *R²* (coeficiente de determinação) na regressão linear múltipla, ressaltando que seu valor máximo não atinge um valor igual à unidade como nos casos de regressão pelo método dos mínimos quadrados ordinários. A regressão logística, diferentemente da regressão linear, prediz diretamente a probabilidade de um evento ocorrer, que pode ser qualquer valor entre 0 e 1.

Como medida auxiliar à modelagem por meio da RL foi empregada a análise do **risco relativo (RR)**, que é uma relação da probabilidade de um evento ocorrer no grupo exposto ao fator estudado (nesse caso, a presença de fraude) contra o grupo não exposto ao fator (clientes sem fraude). Caso o risco seja igual para os dois grupos, o valor do risco relativo é igual a 1.

Outra estatística auxiliar empregada é a análise de correlação a partir do coeficiente de **correlação de Spearman**, r_s , que é calculado de forma a medir a força entre duas variáveis ordinais, não necessitando da garantia de normalidade das variáveis pesquisadas. É utilizado para determinar se existe evidência estatística significativa da associação entre duas variáveis (LEVINE et al., 2005). É um índice que varia de -1 a +1, e quanto mais perto de $|1|$ maior é a correlação das variáveis.

Pelo fato de as variáveis nesta pesquisa serem fundamentalmente categóricas, foi utilizado o método das **variáveis dummy**. As variáveis *dummy* são expressões artificiais e dicotômicas. O valor 1 é atribuído para casos onde ocorre determinada situação e 0 no caso contrário (HAIR JR. et al., 2005). A expressão denotada por (2) pode ser desdobrada da seguinte forma:

$$y_i = \begin{cases} 1 & \text{se } w_i = x_i \\ 0 & \text{c.c} \end{cases} \quad \text{onde, } y_i = \text{dummy}; w_i = \text{categoria}; x_i = \text{valor} \quad (2)$$

Vale destacar que a análise univariada das variáveis, bem como as estimações econométricas da regressão logística foram efetuadas pautando-se no *software* XLSTAT versão 7.5.

4. Apresentação e Análise dos Dados

Detectar fraudes é assunto de grande interesse para empresas de telecomunicações, para agências governamentais, para companhias de cartões de crédito e de seguros, dentre outras, uma vez que as perdas geradas representam fator negativo. A detecção prévia pode prevenir futuros atos fraudulentos ou, mesmo, tentar obter restituição dos prejuízos (PASSINI; TOLEDO, 2002).

O objetivo dessa pesquisa empírica é demonstrar o ganho que a técnica de regressão logística pode representar para uma organização, a partir da detecção de fraudes. O trabalho procurou transformar os dados registrados por uma companhia de telecomunicações sobre os seus clientes em informações úteis na prevenção de fraudes.

4.1. Análise univariada das variáveis pesquisadas

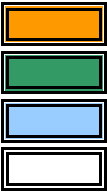
O primeiro passo da análise dos dados constituiu em avaliar se as variáveis integrantes do estudo têm influência entre si - ou seja, detectar a correlação não-paramétrica entre elas. Este passo é relevante na medida em que permite detectar se é possível reduzir o número de características utilizadas na predição.

Os resultados encontrados podem ser verificados na Figura 3, que destaca os níveis de correlação entre as variáveis com $\alpha = 0,05$. Pela análise, é possível identificar

que mesmo as correlações estatisticamente significantes não demonstraram valores elevados de associação por meio da estatística de Spearman - ou seja, as variáveis não são interdependentes. Deste modo, as 12 características levantadas foram utilizadas na modelagem.

Figura 3: Correlações de Spearman entre as variáveis pesquisadas

Variáveis	TECNO	SEXO	TPCLI	PROF	QPESS	CDBIL	CEPSV	EMAIL	COMPU	DVCTO	FPGTO
NLINH	-0,18	-0,09	0,16	0,02	0,17	0,01	-0,04	0,11	-0,13	-0,05	0,02
TECNO		-0,01	-0,05	-0,04	-0,09	-0,08	-0,01	-0,09	0,10	-0,01	0,02
SEXO			-0,02	0,09	-0,08	0,03	0,02	0,00	0,03	-0,01	0,00
TPCLI				-0,03	0,13	-0,05	-0,04	-0,02	-0,07	0,01	-0,01
PROF					-0,04	0,06	0,00	0,10	0,03	-0,07	0,02
QPESS						-0,01	0,01	0,02	-0,34	-0,03	0,00
CDBIL							0,12	0,07	-0,01	-0,05	0,01
CEPSV								-0,07	0,01	-0,03	0,00
EMAIL									-0,04	-0,05	0,05
COMPU										0,04	0,01
DVCTO											-0,01



- Correlação estísticamente significativa: $|0,2| > r$
- Correlação estísticamente significativa: $|0,1| > r \geq |0,2|$
- Correlação estísticamente significativa: $|0,0| > r \geq |0,1|$
- Não existe correlação estatisticamente significativa

O segundo passo utilizado na mineração dos dados é o cálculo do risco relativo de cada um dos atributos dentro das 12 características integrantes do estudo. Aqui o objetivo é selecionar aqueles atributos com maior associação positiva ou negativa à fraude. A partir da análise dos riscos relativos mais relevantes, foram criadas variáveis *dummy* para cada um dos atributos escolhidos.

Na Tabela 2 estão apresentados os riscos relativos, o percentual na amostra e a variável *dummy* criada para cada um dos atributos significantes das características estudadas. O critério para identificar os atributos com riscos relativos mais significativos foi preservar intervalo de confiança de 10% em torno do valor 1 (proporções iguais); ou seja, os atributos com $0,90 \leq RR \leq 1,10$ foram retirados do estudo.

O risco relativo deve ser entendido como a chance de ser mais ou menos fraudador. Se o RR se situar abaixo de 1, o atributo deve ser analisado como percentualmente mais propenso a ter fraude. O oposto (quando o RR fica acima de 1) mostra que o atributo é mais propenso a não ter fraude. Para exemplificar, analisando-se a variável Data de Vencimento da Fatura (DVCTO), quando a data é igual a 2, a chance de ocorrer fraude é 29% ($1 - 0,71$) maior que nas outras datas. Quando a data é 5, a chance é 23% maior de não ocorrer fraude.

Tabela 2: Variáveis *dummy* criadas para a pesquisa

CARACTERÍSTICA	ATRIBUTO	RR	%TOTAL	DUMMY
NLINH	2	1,18	34%	Dummy 1
	3	0,75	9%	Dummy 2
	4	0,45	4%	Dummy 3
	Mais que 5	0,26	4%	Dummy 4
TECNO	WLL	0,67	28%	Dummy 5
SEXO	M	0,87	56%	Dummy 6
TPCLI	SOHO	0,51	4%	Dummy 7
EMAIL	Não	0,79	76%	Dummy 8
COMPU	Sim	1,21	33%	Dummy 9
FPGTO	Débito Automático	21,00	1%	Dummy 10
DVCTO	2	0,71	5%	Dummy 11
	5	1,23	10%	Dummy 12
	10	1,14	34%	Dummy 13
	15	1,28	20%	Dummy 14
	28	0,51	17%	Dummy 15
QPESS	1	1,18	9%	Dummy 16
CDBIL	GRUPO 1	0,39	8%	Dummy 17
	GRUPO 2	0,66	12%	Dummy 18
	GRUPO 3	0,78	8%	Dummy 19
	GRUPO 5	1,20	14%	Dummy 20
	GRUPO 6	1,49	9%	Dummy 21
	GRUPO 7	1,84	6%	Dummy 22
	GRUPO 8	2,26	6%	Dummy 23
	GRUPO 1	0,25	9%	Dummy 24
PROF	GRUPO 2	0,48	15%	Dummy 25
	GRUPO 3	0,72	7%	Dummy 26
	GRUPO 5	1,18	10%	Dummy 27
	GRUPO 6	1,50	9%	Dummy 28
	GRUPO 7	2,69	13%	Dummy 29
CEPSV	70	1,27	4%	Dummy 30
	73	1,67	1%	Dummy 31
	74	1,26	6%	Dummy 32
	79	0,67	4%	Dummy 33
	80	0,81	11%	Dummy 34
	85	1,22	5%	Dummy 35
	86	0,44	7%	Dummy 36
	87	1,53	4%	Dummy 37
	88	1,31	4%	Dummy 38
	89	1,11	5%	Dummy 39
	90	1,32	7%	Dummy 40
91	1,03	4%	Dummy 41	

As variáveis CDBIL (cidade de faturamento) e PROF (profissão) possibilitam agrupamentos entre os atributos, por meio da semelhança entre seus riscos relativos. Estes atributos estavam significantemente diferentes de 1; no entanto, o número de clientes era baixo dentro de cada um deles. Em função disso, a junção desses atributos formando novos colaborou com a sua inclusão na modelagem dos dados. Sendo assim, os grupos de

idades e de profissões criados passaram a fazer parte do cálculo como se fossem um único atributo.

4.2. Resultados da análise de regressão logística

A análise multivariada das variáveis foi realizada por meio da regressão logística. A partir dela, é possível construir um modelo que melhor descreve a relação entre as características sócio-demográficas e a propensão a ser um cliente fraudador. A Tabela 3 mostra os atributos (em forma de *dummy*) que entraram no modelo. A regressão se mostrou significativa, por meio dos testes de Log Verossimilhança, *Score* e *Wald*.

A forma funcional do modelo de propensão a fraude encontrado é descrita por meio da expressão algébrica (3).

$$Tipo = \frac{1}{1 + e^{(-(-1,583+0,626*dummy2+1,114*dummy3+1,608*dummy4+0,539*dummy5+0,218*dummy6+0,916*dummy8-0,159*dummy9-3,097*dummy10+0,556*dummy11+0,0758*dummy15+0,839*dummy25+0,55*dummy26-0,733*dummy29+0,598*dummy34))}} \quad (3)$$

É importante salientar que a influência dos coeficientes na variável resposta se dá de forma inversamente proporcional ao sinal. Ou seja, a variável com um β positivo influi negativamente na probabilidade de ser um fraudador.

Tabela 3: Estimação dos parâmetros do modelo de propensão a fraude

ATRIBUTO		β	Erro Padrão	Pr > Qui ²
α	Intercepto	-1,734	0,113	< 0,0001
β_1	<i>Dummy 2</i>	0,317	0,129	< 0,0001
β_2	<i>Dummy 3</i>	0,785	0,210	< 0,0001
β_3	<i>Dummy 4</i>	1,287	0,229	< 0,0001
β_4	<i>Dummy 5</i>	1,775	0,085	< 0,0001
β_5	<i>Dummy 6</i>	0,619	0,074	0,003
β_6	<i>Dummy 8</i>	0,211	0,090	< 0,0001
β_7	<i>Dummy 9</i>	0,945	0,080	0,047
β_8	<i>Dummy 10</i>	-0,215	1,049	0,003
β_9	<i>Dummy 11</i>	-3,168	0,171	0,001
β_{10}	<i>Dummy 15</i>	0,556	0,102	< 0,0001
β_{11}	<i>Dummy 17</i>	0,758	0,149	< 0,0001
β_{12}	<i>Dummy 18</i>	0,829	0,118	< 0,0001
β_{13}	<i>Dummy 19</i>	0,587	0,138	< 0,0001
β_{14}	<i>Dummy 22</i>	0,563	0,160	0,010
β_{15}	<i>Dummy 23</i>	-0,410	0,159	0,009
β_{16}	<i>Dummy 24</i>	-0,422	0,148	< 0,0001

Cont....

Tabela 3 (cont.): Estimação dos parâmetros do modelo de propensão a fraude

β_{17}	Dummy 25	1,486	0,106	< 0,0001
β_{18}	Dummy 26	0,680	0,145	0,000
β_{19}	Dummy 29	0,552	0,119	< 0,0001
β_{20}	Dummy 34	-0,759	0,125	< 0,0001
N = 3.788		R ² (Cox e Snell) = 0,197		
		Estatística	Pr > Qui ²	
		-2 Log(Veros)	< 0,0001	
		Score	< 0,0001	
		Wald	< 0,0001	

Cabe destacar que o coeficiente de determinação R² situa-se em torno de 0,197, mostrando que a variabilidade nas variáveis independentes explica 19,7% da variabilidade na variável dependente.

4.3. Validação do modelo estimado

A partir da construção do modelo, a etapa posterior é a validação do modelo encontrado. Neste caso, utiliza-se a própria amostra de desenvolvimento como padrão para o cálculo dos índices de bondade de ajuste. A Tabela 4 mostra a matriz de confusão que é a comparação entre a classificação, *a priori* (classificação dos clientes na amostra) e *a posteriori* (obtidos pela estimativa do modelo de regressão logística). Por meio dela, é possível visualizar que o percentual de acerto situa-se em torno de 68,88%.

Tabela 4: Validação do modelo estimado

REAL	BOM	MAU	Total	% correto	
BOM	1.324	570	1.894	69,90%	% especificidade
MAU	609	1.285	1.894	67,85%	% sensibilidade
Total	1.933	1.855	3.788	68,88%	% ACERTO

Os índices de especificidade (clientes bons reclassificadas como bons), e sensibilidade (clientes maus reclassificadas como maus) ficaram em 69,90% e 67,85%, respectivamente. Estas medidas são utilizadas para descrever a precisão do ajuste, e demonstram que o modelo está conseguiu identificar características relevantes que diferenciam o perfil do cliente fraudador do não fraudador.

4.4. Aplicação do Modelo Estimado

A partir da construção do modelo, é necessário criar um padrão para implementação e utilização por parte da companhia. Esta etapa é muito importante e diferencia o processo de mineração de dados aplicado a uma empresa do processo de modelagem estatística com fins puramente acadêmicos.

Sendo assim, são desenvolvidas análises visando a entender os ganhos real e estratégico que o presente estudo pode representar. Com este objetivo foram criadas

classes de score, que são as probabilidades de ser fraudador encontradas a partir da aplicação do modelo na amostra de desenvolvimento.

Na Tabela 5 está apresentada a distribuição por *classe score* dos bons e maus clientes da amostra de desenvolvimento. Esta tabela fornece o % **Total** na classe, que serve como o percentual esperado de clientes em cada classe de *score* (%ECS). A partir da obtenção do % ECS, é feita a distribuição dos clientes bons da população. Este valor foi estimado a partir da multiplicação do total de clientes bons não amostrados pelo (%ECS) + clientes bons da amostra. Desta forma, foi possível retornar ao % de Maus Total existente na população total (0,76% de fraudes comprovadas).

Tabela 5: Distribuição por classe de *score* na amostra de desenvolvimento

CLASSE DE SCORE	BOM	MAU	TOTAL	% TOTAL	$\Rightarrow \frac{\#total_{classe}}{\#total}$
< 0,1	51	6	57	1,5%	
[0,1 a 0,2)	265	56	321	8,5%	
[0,2 a 0,3)	340	98	438	11,6%	
[0,3 a 0,4)	373	207	580	15,3%	
[0,4 a 0,5)	295	242	537	14,2%	
[0,5 a 0,6)	250	284	534	14,1%	
[0,6 a 0,7)	172	337	509	13,4%	
[0,7 a 0,8)	92	281	373	9,8%	
[0,8 a 0,9)	42	266	308	8,1%	
≥ 0,9	14	117	131	3,5%	
TOTAL	1.894	1.894	3.788		

A Tabela 6 permite o cálculo do percentual esperado de clientes fraudadores por *classe de score* (%MAU), a partir da distribuição do % de Maus Total pelas diferentes *classes de score*. É interessante observar que existe aderência entre o modelo e o cenário real, uma vez que o %MAU cresce junto com a probabilidades de ser fraudador encontrada pelo modelo.

Após estimar o %MAU é possível calcular a quantidade esperada de clientes que são fraudadores em cada classe. Este cálculo torna-se útil para que a empresa possa estimar a quantidade de casos de fraude que irão aparecer em cada período.

Com o intuito de demonstrar o ganho possível representado pelo modelo, é feita a escoragem de todos os clientes que tiveram a 1ª fatura emitida no período considerado, e que não tiveram fraude comprovada até o momento.

Tabela 6: Percentual esperado de clientes fraudadores por classe de score (%CFCS)

CLASSE DE SCORE	BOM - POPULAÇÃO	MAU	TOTAL	%MAU
< 0,1	3.751	6	3.757	0,16%
[0,1 a 0,2)	21.104	56	21.160	0,26%
[0,2 a 0,3)	28.775	98	28.873	0,34%
[0,3 a 0,4)	38.027	207	38.234	0,54%
[0,4 a 0,5)	35.157	242	35.399	0,68%
[0,5 a 0,6)	34.917	284	35.201	0,81%
[0,6 a 0,7)	33.216	337	33.553	1,00%
[0,7 a 0,8)	24.307	281	24.588	1,14%
[0,8 a 0,9)	20.037	266	20.303	1,31%
>= 0,9	8.519	117	8.636	1,35%
TOTAL	247.811	1.894	249.705	0,76%

% Maus cresce à medida que a probabilidade de ser fraude cresce. Demonstra aderência do modelo ao cenário real.

A Tabela 7 apresenta a quantidade de fraudes esperadas dentro deste grupo, tornando possível identificar o ganho representado pelo modelo de propensão à fraude. Em um universo de 57.444 clientes, escolhendo-se os que têm probabilidade superior a 40% de serem fraudadores, o universo de trabalho fica resumido em 39% dos clientes originais (11.297 clientes), o tratamento deste novo conjunto possibilita detectar 54% de todas as fraudes. Isto representa um ganho de efetividade de 21% nas auditorias de fraude que são feitas pela companhia.

Tabela 7: Distribuição da probabilidade nos clientes com 1ª fatura

CLASSE DE SCORE	TOTAL	% TOTAL ACUM	FRAUDE ESPERADA	% FRAUDE ACUM.
>= 0,9	18	0,03%	0	0,07%
[0,8 a 0,9)	190	0,36%	2	0,83%
[0,7 a 0,8)	694	1,57%	8	3,24%
[0,6 a 0,7)	2.564	6,03%	26	11,05%
[0,5 a 0,6)	7.818	19,64%	63	30,18%
[0,4 a 0,5)	11.297	39,31%	77	53,61%
[0,3 a 0,4)	19.967	74,07%	108	86,40%
[0,2 a 0,3)	7.265	86,72%	25	93,88%
[0,1 a 0,2)	7.607	99,96%	20	99,99%
< 0,1	24	100,00%	0	100,00%
TOTAL	57.444			

Assim, percebe-se que a pesquisa propicia a redução efetiva no número de análises que deverão ser realizadas dentro da companhia, poupando recursos para atividades que efetivamente geram receita. É possível verificar que o modelo desenvolvido de propensão à fraude criou um tipo de inteligência organizacional, de maneira que possibilita, à organização, antever casos de fraude e, com isso, reduzir o número de auditorias necessárias à prevenção.

5. Considerações Finais e Recomendações

Este trabalho desenvolveu uma pesquisada aplicada com base no conceito de inteligência organizacional e os seus métodos de criação. Foi elaborado estudo empírico que procurou demonstrar como a ferramenta de regressão logística aliada a técnicas estatísticas descritivas de associação pode possibilitar o desenvolvimento de inteligência dentro de uma empresa de telecomunicações brasileira.

O resultado encontrado, a partir de um modelo de propensão à fraude, foi satisfatório, no sentido de que permite melhoria de eficiência nos processos de auditoria de fraude. É importante salientar que as predições devem ser utilizadas com extremo cuidado e parcimônia, pois os clientes devem ter os seus direitos respeitados, e que uma probabilidade alta de fraude não indica que o consumidor é de fato um fraudador, mas sim que possuiu um perfil a ser analisado.

Cabe salientar que, após a entrada da *Lei Sarbanes & Oxley*, nos Estados Unidos, as empresas desse setor em específico no Brasil vêm sendo levadas a implantar estruturas de gestão de riscos operacionais - por exemplo, com o intuito de mostrar, ao mercado investidor, uma maior transparência acerca da organização e visando também a uma maior garantia dos controles gerenciais na condução dos negócios.

Notadamente, o uso da ferramenta descrita anteriormente depende da implantação de uma cultura de gestão de riscos na organização estudada. Também depende do apoio e do patrocínio da alta administração da empresa avaliada nesse processo, além do comprometimento de todos os funcionários e gestores, visando a criar estrutura de coordenação e de monitoramento de riscos na organização, ao incluir o risco operacional vinculado às fraudes como prioridade na estratégia de negócio da organização.

Finalmente, como sugestão para futuros trabalhos, é possível continuar investigando fraudes através de ferramental de mineração de dados, utilizando outras técnicas como redes neurais ou árvores de decisão. Outra possibilidade de análise envolveria dados de consumo ao invés de dados sócio-demográficos. Esse tipo de aplicação poderá também ser testado em um setor diferente da economia.

Referências

- ALMEIDA, L. M. et al. Uma ferramenta para extração de padrões. **Revista Eletrônica de Iniciação Científica**, v.3, 2003. Disponível em: <<http://www.sbc.org.br/reic/edicoes/2003e4/>>. Acesso em: 08/01/ 2007.
- ALMEIDA, A.M.P.; MOTTA, F.M.V. O desenvolvimento de competências gerenciais como fator de vantagem competitiva: um estudo de caso em uma empresa estatal de telecomunicações. In: VIII ENANGRAD, 1997. **Anais Eletrônicos...** Rio de Janeiro - RJ, 1997.

- BEUREN, I. M. **Gerenciamento da informação**. 2. ed. São Paulo: Atlas, 2000.
- CASTELLS, M. **A sociedade em rede**. 6. ed. São Paulo: Paz e Terra, 1999.
- DALFOVO, O.; MOMO, M.R.; MAIA, L.F.J. Sistema de informação baseado em *data mining* aplicado à gestão ambiental como recurso estratégico. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, 3., 2003. **Anais...** Itajaí - SC, 2003. v. 1.
- DAVENPORT, T.; PRUSAK, L. **Ecologia da informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.
- FERREIRA, A.; ALVES, L. **Inteligência artificial e data mining**. Departamento de Ciência da Computação. Universidade Federal da Bahia (UFBA), 2005.
- FIUZA, E.P.S.; NERI, M.C. Reflexões sobre os mecanismos de universalização do acesso disponíveis para o setor de telecomunicações no Brasil. **Texto para Discussão** nº 573. Rio de Janeiro: IPEA, 1998.
- GILAD, B. U. S. Intelligence System: model for corporate chiefs?. **The Journal of Business Strategy**, Boston, v. 12, n. 3, p. 20-25, mai./jun., 1991.
- GIMENES, R.M.T.; OPAZO, M.A.U. Previsão de insolvência de cooperativas agropecuárias por meio de modelos multivariados. **Revista FAE**, Curitiba, v. 4, n.3, p.65-78, set./dez. 2001.
- HAIR, J.F. et al. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005.
- HALL, R.H. **Organizações**: estruturas, processos e resultados. 8. ed. São Paulo: Prentice Hall, 2004.
- HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. 2nd ed. New York: John Wiley & Sons, 2000.
- JONSSON, E.; LUNDIN, E.; KVARNSTRÖM, H. Combining fraud and intrusion detection – meeting new requirement. In: NORDIC WORKSHOP ON SECURE IT SYSTEMS - NORDSEC, 5., 2000. **Proceedings...**Iceland, out, 2000.
- LEVINE, D. M.; STEPHAN, D.; KREHBIEL, T. C.; BERENSON, M. L. **Estatística** - teoria e aplicações usando o Microsoft Excel em português. Rio de Janeiro: LTC, 2005.
- LINS, B. F. E. **Avanços recentes no setor de telecomunicações e suas implicações nos campos educacional e social**. Consultoria Legislativa da Câmara dos Deputados. Brasília, 2001.
- NAVEGA, S. Princípios essenciais do data mining. **Anais Eletrônicos...** Infoimagem 2002, Cenadem, Novembro. Disponível em < <http://www.intelliwise.com/reports/i2002.htm>>. Acesso em 08 jan. 2007.
- PASSINI, S.R.R.; TOLEDO, C.M.T. Mineração de dados para detecção de fraudes em ligações de água. In: XI SEMINCO - Seminário de computação, 2002. **Anais Eletrônicos...** Blumenau - SC, 2002. Disponível em: <<http://www.inf.furb.br/seminco/2002/artigos/Passini-seminco2002-6.pdf>>. Acesso em: 08 jan. 2007.
- PORTER, M.E. **Vantagem competitiva**: criando e sustentando um desempenho superior. 15. ed. Rio de Janeiro: Campus, 2000.
- SAPIRO, A. Inteligência empresarial: a revolução informacional da ação competitiva. **Revista de Administração de Empresas**. São Paulo, v. 33, n. 3, p. 106-124, mai./jun. 1993.
- SPSS. **Statistical Package for the Social Sciences**. 2007. <http://www.spss.com.br/>. Acesso em: 09 de janeiro de 2007.
- TARAPANOFF, K.; ARAÚJO JÚNIOR, R.H.; CORMIER, P.M.J. Sociedade da informação e inteligência em unidades de informação. **Ci. Inf.**, Brasília, v. 29, n. 3, p. 91-100, set./dez., 2000.
- TJADEN, G.S. Measuring the information age business. **Technology Analysis & Strategic Management**, v. 8, n. 3, p. 233-246, 1996.
- TELECO **Relatórios Teleco 2006** - Telefonia Fixa no Brasil: Dados 2000 a 2005. Disponível em: < <http://www.teleco.com.br/> >. Acesso em 08 jan. 2007.
- VALENTIN, M.L.P. et al. O processo de inteligência competitiva em organizações. **DataGramZero - Revista de Ciência da Informação**, v.4, n.3, jun. 2003.