



Jorge Brantes Ferreira

**Mineração de Dados na Retenção de Clientes
em Telefonia Celular**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio.

Orientadores: Marley Maria B. R. Vellasco
Marco Aurélio C. Pacheco

Rio de Janeiro, março de 2005



Jorge Brantes Ferreira

**Mineração de Dados na Retenção de Clientes
em Telefonia Celular**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Marley Maria B. R. Vellasco
Orientador
PUC-Rio

Marco Aurélio C. Pacheco
Orientador
PUC-Rio

Emmanuel Piseces Lopes Passos
PUC-Rio

Carlos Hall Barbosa
PUC-Rio

Ricardo Tanscheit
PUC-Rio

Prof. José Eugenio Leal
Coordenador(a) Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, março de 2005

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Jorge Brantes Ferreira

Graduação em Engenharia Elétrica com dupla ênfase em Sistemas de Apoio à Decisão e Telecomunicações pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 2002. MBA IAG Master em Finanças Corporativas concluído em meados de 2003. Início das atividades no mestrado em Engenharia Elétrica com ênfase em Sistemas de Apoio à Decisão em 2003.

Ficha Catalográfica

Ferreira, Jorge Brantes

Mineração de dados na retenção de clientes em telefonia celular / Jorge Brantes Ferreira ; orientadores: Marley Maria B. R. Vellasco, Marco Aurélio C. Pacheco. – Rio de Janeiro : PUC-Rio, Departamento de Engenharia Elétrica, 2005.

93 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica .

Inclui referências bibliográficas

1. Engenharia Elétrica – Teses. 2. Mineração de dados. 3. Telefonia celular. 4. Churn. 5. Customer Relationship Manager (CRM). I. Vellasco, Marley Maria B. R. II. Pacheco Marco Aurélio C. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Ao meu pai pelo exemplo,
À minha mãe pelo carinho,
Ao meu irmão pela amizade.

Agradecimentos

Á minha namorada, Cristiane Giovannini, pelo amor e apoio.

Aos meus amigos, pela presença e companheirismo em infinitas jornadas.

Aos meus orientadores, Dra. Marley B. R. Vellasco e Dr. Marco Aurélio C. Pacheco, pelos valiosos ensinamentos e o essencial apoio que tornou esse trabalho em uma realidade.

Ao CNPq, pelo apoio financeiro.

Ao Departamento de Engenharia Elétrica da PUC-Rio, seus professores e alunos, pela maravilhosa formação e estrutura que me forneceram ao longo de tantos anos de convívio.

Resumo

Jorge Brantes Ferreira. **Mineração de Dados na Retenção de Clientes em Telefonia Celular**. Rio de Janeiro, 2005. 93p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O objetivo desta dissertação é propor um sistema de mineração de dados completo para a solução de problemas de retenção de clientes, presentes nas mais variadas indústrias. Tal solução reside na correta identificação, em meio a gigantescas bases de dados, dos clientes cujos perfis e históricos de comportamento denotam que sua saída da empresa é iminente. Agindo então sobre a inteligência gerada a partir desta classificação de clientes, incentivos e ações de retenção devem ser postos em prática para evitar e/ou minimizar a perda para algum concorrente de clientes valiosos. Ao longo do processo de mineração de dados, deu-se atenção ao processo de preparação e representação dos dados e métodos de seleção de variáveis, na tentativa de melhorar e otimizar o desempenho dos modelos a serem estudados. Vários modelos diferentes foram testados, otimizados e comparados na tarefa de classificação de clientes como aqueles que permanecerão na empresa ou aqueles que apresentam riscos de abandono. Entre os modelos estudados estão: redes neurais, sistemas neuro-fuzzy hierárquicos, algoritmos genéticos, árvores de decisão e máquinas de vetor de suporte. Em particular, avaliou-se a questão do abandono de clientes (*churn*) na indústria de telecomunicações móvel brasileira, devido à disponibilidade de dados reais para a análise. Foi feito um estudo abrangente do problema do *churn*, identificando suas causas, conseqüências e detalhes. Conclui-se com uma análise do impacto da implementação da metodologia proposta em ações de retenção de clientes, sob o prisma da lucratividade ou corte de despesas em que tal utilização implicaria.

Palavras-chave

Mineração de Dados; Telefonia Celular; Churn; Customer Relationship Management (CRM)

Abstract

Jorge Brantes Ferreira. **Data Mining applied to Customer Retention in Wireless Telecommunications**. Rio de Janeiro, 2005. 93p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The goal of this work is to propose a complete data mining system for the solution of customer retention problems, commonly found in many industries. Such a solution encompasses the accurate identification among huge amounts of data of those consumers who would most likely end their relationship with the firm, based on their historical behavior and individual profile. Acting upon the intelligence provided by a precise customer classification, incentives and retention actions should be put into practice to prevent or minimize the losses of valuable clients to competitors. Throughout the data mining process designed here, great care was given to the preparation and representation of the data and to input selection methods, in an effort to optimize the performance of the classification models. Various different classification techniques have been tested, with the objective of finding the one best suited for the task at hand: to pinpoint those customers who present clear risks of abandoning the analyzed company. Among the studied models were neural networks, decision trees, genetic algorithms, neuro-fuzzy systems and SVMs (“Support Vector Machines”). As a case study, the issue of “churn” (loss of customer to a competitor) in the Brazilian wireless telecommunications was tackled, due to the availability of data. A detailed study was made, identifying the causes, consequences and details of the business problem. As a conclusion, the great impact of the implementation of the proposed system in retention strategies of wireless carriers is evaluated, under the view of the profitability that would be generated by its use.

Keywords

Data Mining; Churn; Customer Relationship Management (CRM); Wireless Telecommunications

Sumário

1 Introdução	13
1.1. Motivação	13
1.2. Objetivos	15
1.3. Descrição do Trabalho	16
1.4. Organização da Dissertação	17
2 Descoberta de Conhecimento em Bases de Dados	18
2.1. Introdução	18
2.2. Dados	19
2.2.1. Data Warehouses	19
2.2.2. OLAP, Data Marts e Bases de Dados Multidimensionais	20
2.3. Preparação dos dados	21
2.4. Mineração de dados	22
2.5. Interpretação e utilização do conhecimento gerado	23
2.6. Aplicações em CRM	23
2.6.1. O que se entende por CRM	24
2.6.2. Entendimento do cliente	24
2.6.3. Geração de inteligência empresarial	25
2.7. Resumo	25
3 Técnicas e Modelos para Análise de Dados	26
3.1. Introdução	26
3.2. Preparação dos Dados	26
3.2.1. Tratamento e Limpeza	27
3.2.2. Transformações	28
3.2.3. Oversampling	29
3.3. Métodos de seleção de variáveis	30
3.3.1. Método do Estimador de Mínimo Quadrado (LSE)	31
3.3.2. Método da Efetividade de uma Entrada Singular (SIE)	31
3.3.3. Método baseado no modelo ANFIS	32
3.4. Métodos de Classificação de Padrões	33
3.4.1. Métodos Estatísticos	34

3.4.2. Redes Neurais	35
3.4.3. Árvores de Decisão	38
3.4.4. Sistemas Neuro-Fuzzy	41
3.4.5. Máquinas de Vetor de Suporte	42
3.4.6. Algoritmos Genéticos	43
3.5. Resumo	44
4 Retenção de Clientes em Telefonia Celular	45
4.1. Introdução	45
4.2. A indústria de telefonia móvel no Brasil	45
4.3. O problema do <i>churn</i>	47
4.3.1. O significado do <i>churn</i>	48
4.3.2. Possíveis causas	48
4.3.3. Consequências e custos	50
4.3.4. Como combater o <i>churn</i>	51
4.4. Resumo	51
5 Sistema de Auxílio à Retenção de Clientes por Mineração de Dados	52
5.1. Introdução	52
5.2. O sistema de auxílio à retenção de clientes por mineração de dados	53
5.3. Coleta e estudo dos dados	55
5.4. Preparação dos dados	57
5.4.1. Definição do alvo	57
5.4.2. Adicionar variáveis derivadas	58
5.4.3. Transformação dos dados	58
5.4.4. Seleção de variáveis	59
5.4.5. Oversampling	59
5.5. Modelagem do churn	60
5.6. Análise de lucratividade	61
5.7. Operacionalização dos resultados	64
5.8. Resumo	64
6 Estudo de Caso	65
6.1. Descrição	65
6.2. Dados disponíveis e suas limitações	65
6.3. Apresentação e discussão dos resultados	67
6.3.1. Validação, exploração e limpeza dos dados	68

6.3.2. Definição do alvo	70
6.3.3. Adição de Variáveis Derivadas e Transformações	71
6.3.4. Seleção de variáveis	71
6.3.5. Oversampling	73
6.3.6. Criação das bases de dados	74
6.3.7. Modelagem	75
6.3.8. Análise da lucratividade	79
6.4. Resumo	82
7 Conclusões e Trabalhos Futuros	84
7.1. Conclusões	84
7.2. Trabalhos Futuros	85
8 Artigos Publicados	86
9 Referências Bibliográficas	87

Lista de figuras

Figura 2.1 – Ciclo da Descoberta do Conhecimento em Bases de Dados	19
Figura 3.1 – Sistema ANFIS para Seleção de Variáveis	32
Figura 3.2 – Particionamento do sistema ANFIS utilizado	33
Figura 3.3 – Estrutura em camadas de uma Rede Neural	35
Figura 3.4 – Neurônio Artificial	36
Figura 3.5 – Arquitetura de uma Rede Neural com bases estatísticas	38
Figura 3.6 – Estrutura de uma árvore de decisão	39
Figura 3.5 – Caso Linearmente Separável	42
Figura 3.6 – Rule Evolver: Representação do Cromossoma	43
Figura 5.1 – Sistema de Retenção de Clientes por Mineração de Dados	54
Figura 6.1 – Histograma da Distribuição da variável Fatura Média, a linha que corta o valor de R\$480,00 no eixo x demonstra onde foi realizado o filtro de <i>outliers</i> .	70
Figura 6.2 – Faixas de lucratividade para uma taxa de retenção de 80%.	80
Figura 6.3 – Faixas de lucratividade para uma taxa de retenção de 60%.	81
Figura 6.4 – Faixas de lucratividade para uma taxa de retenção de 40%.	81
Figura 6.5 – Faixas de lucratividade para uma taxa de retenção de 20%.	82

Lista de tabelas

Tabela 4.1 – Situação do mercado de telefonia móvel no Brasil ao final de 2004, dados fornecidos pelas operadoras, em milhares. O número total de celulares no país é de 66,605 milhões, segundo a Agência Nacional de Telecomunicações, ANATEL.	47
Tabela 4.2 – <i>Churn</i> mensal para as quatro principais operadoras atuantes no Brasil, ao final de 2004. Dados do site www.teleco.com.br .	48
Tabela 4.3 – Causas do <i>churn</i> .	49
Tabela 6.1 – Estrutura da base de dados real para o estudo de caso.	67
Tabela 6.2 – Estatísticas descritivas básicas das variáveis base de dados original.	69
Tabela 6.3 – Resultados da aplicação do método de seleção de variáveis LSE.	72
Tabela 6.4 – Resultados da aplicação do método de seleção de variáveis SIE.	72
Tabela 6.5 – Resultados da aplicação do método de seleção de variáveis ANFIS.	72
Tabela 6.6 – Variáveis restantes para a modelagem após a seleção de variáveis	74
Tabela 6.7 – Resultados obtidos na literatura para modelagem do <i>churn</i> através de métodos de mineração de dados.	76
Tabela 6.8 – Matrizes de confusão para os melhores modelos de cada família de métodos avaliados.	76
Tabela 6.9 – Lucratividade para os modelos segundo os resultados do conjunto de testes.	79