

Universidade do Minho

Escola de Engenharia

Fernando Jorge Machado Veloso

**Um Modelo para Previsão de Churn na
Área do Retalho**



Universidade do Minho

Escola de Engenharia

Fernando Jorge Machado Veloso

Um Modelo para Previsão de Churn na Área do Retalho

Dissertação de Mestrado

Área de Especialização Engenharia Informática

Trabalho efectuado sob a orientação do

Professor Doutor Orlando Manuel de Oliveira Belo

Novembro de 2012

Agradecimentos

No longo caminho que foi o processo de desenvolvimento desta dissertação, com bons e também alguns menos bons momentos, gostaria de agradecer a todos aqueles que direta ou indiretamente permitiram que eu aqui chegasse, à escrita destes agradecimentos.

Em primeiro lugar gostaria de agradecer ao meu orientador, o professor Orlando Belo, não só pela ajuda e orientação deste trabalho, mas também pela excelente forma de estar e de ensinar durante as atividades letivas.

Gostaria também de agradecer à minha família por permitir que eu aqui chegasse e por toda a paciência demonstrada desde o início deste caminho universitário.

Por último gostava de agradecer a todos os meus amigos pelo incentivo e força que me transmitiram e em particular ao Eduardo Costa que me acompanhou e ajudou durante grande parte do processo de desenvolvimento desta dissertação, permitindo que não tivesse sido um percurso solitário.

Resumo

Um Modelo para Previsão de *Churn* na Área do Retalho

O ambiente de grande competitividade característico do sector do retalho e crescente dificuldade na captação de novos clientes leva as empresas a apostar na implementação de estratégias adequadas para promover a satisfação dos clientes adquiridos para motivar a sua lealdade. É neste contexto que se começa a reconhecer a importância de combater o fenómeno de *churn*, ou seja, a perda de clientes. É necessário identificar os clientes que estão em risco de *churn* e, para isso, é necessário criar um método que o permita fazer com antecedência para que possam recair sobre eles as campanhas de retenção proactivas. Quanto mais eficaz for o método a identificar os clientes em riscos, maior será o retorno da aplicação da campanha. Muitos trabalhos têm sido desenvolvidos na área de previsão de *churn* nos mais diversos sectores. Contudo, na área do retalho a pesquisa têm sido muito limitada. Assim, com este trabalho de dissertação pretendeu-se estudar o fenómeno da perda de clientes com o objectivo de definir e implementar um modelo de *churning* para o sector do retalho recorrendo a técnicas de mineração de dados. Pretendeu-se fazer um levantamento das principais questões envolvidas na previsão de *churn* no retalho, na construção do conjunto de dados (assinaturas dos clientes) e na aplicação de técnicas de mineração de dados no processo de previsão. Nesse sentido, foram construídos alguns modelos para fazer a previsão de casos de *churn* baseados em cinco das técnicas de classificação mais utilizadas em trabalhos de previsão de *churn*: Árvores de Decisão, Regressão Logística, Redes Neurais, *Random Forests* e SVM. A avaliação e comparação da performance dos modelos elaborados foi feita de acordo com várias medidas como *accuracy*, *precision*, *sensitivity*, *specificity*,

f-measure e AUC e, para além disso, foi testado o impacto, na precisão do modelo, da alteração da densidade de eventos de *churn* no conjunto de treino.

Palavras-chave: *Data Mining*, Árvores de Decisão, Regressão Logística, Redes Neurais, *Random Forests*, *Support Vector Machines*, Previsão de Churn, Sistemas de Retalho.

Abstract

A Churn Prediction Model for Retail

The great competitive environment characteristic of the retail sector and increasing difficulty in attracting new customers leads firms to invest in the implementation of appropriate strategies to promote customer satisfaction to motivate their loyalty. It is in this context that we begin to recognize the importance of combating the phenomenon of churn, i.e., the loss of clients. It is necessary to identify customers who are at risk of churn and, therefore, it is necessary to create a method that allows to do it in advance so that they can be covered by the proactive retention campaigns. The more effective the method to identify customers at risk, the higher the return of applying the campaign. Many studies have been developed in the area of churn prediction in various sectors. However, in the area of retail the research has been very limited. So with this dissertation work was intended to study the phenomenon of loss of customers to define and implement a model of churning to the retail sector using data mining techniques. The intention was to make a survey of the main issues involved in the prediction of churn in retail, construction of the dataset (customer signatures) and applying data mining techniques in the forecasting process. Accordingly, some models were constructed to forecast cases of churn based on five of the most commonly used classification techniques in churn prediction: Decision Trees, Logistic Regression, Neural Networks, Random Forests and SVM. The evaluation and comparison of the performance of models developed has been made according to several measures as accuracy, precision, sensitivity, specificity, f-measure and AUC and, furthermore, has been tested the impact of the change in the density of churn events in the training set.

Keywords: *Data Mining*, Decision Trees, Logistic Regression, Neural Networks, *Random Forests*, *Support Vector Machines*, Churn Prediction, Retail Systems.

Índice

1	Introdução.....	1
1.1	Situações de <i>Churn</i> em Sistemas de Retalho	1
1.1.1	Gestão de <i>Churn</i> no Retalho	3
1.1.2	Caracterização do Sector do Retalho Português	5
1.2	Motivação, Objectivos e Metodologia	8
1.3	Organização da Dissertação	10
2	Análise do Fenómeno de <i>Churn</i>.....	11
2.1	Uma Cultura Centrada no Cliente.....	11
2.1.1	Organização Centrada no Cliente.....	13
2.1.2	Ciclo de Vida do Cliente	13
2.1.3	Processos de Negócio no Ciclo de Vida de um Cliente	14
2.2	Retenção de Clientes	16
2.2.1	Impacto nos Lucros.....	17
2.2.2	Análise de Números de Retenção	19
2.2.3	Em Prol da Retenção dos Bons Clientes	20
2.3	Situações de <i>Churn</i>	21
2.3.1	Definição e Taxa de <i>Churn</i>	21
2.3.2	Factores de <i>Churn</i>	22
2.3.3	Os Diferentes Tipos de <i>Churn</i>	23
2.4	A Gestão de <i>Churn</i>	24
2.4.1	Controlando o <i>Churn</i> Proactivamente.....	25

Diferentes Ambientes na Previsão de <i>Churn</i>	26
Definição do "Momento de <i>Churn</i> "	27
2.4.2 Modelação Preditiva de <i>Churn</i>	27
Diferentes Modelos de <i>Churn</i>	27
A Importância da Precisão e Compressibilidade.....	28
3 Previsão de Situações de <i>Churn</i>	31
3.1 Modelos de Mineração de Dados	31
3.1.1 Problema de Classificação	35
3.1.2 Metodologia CRISP-DM.....	36
3.2 Desenho da Assinatura de um Cliente	37
3.2.1 Definição da Janela Temporal do Modelo	38
3.2.2 Compreensão dos Dados	44
3.2.3 Conceptualização da Variável Dependente	49
3.2.4 Problema de Balanceamento das Classes	50
3.3 Trabalho Relacionado	52
3.4 Revisão das Técnicas de Classificação	52
3.4.1 Árvores de Decisão.....	53
3.4.2 Regressão Logística.....	56
3.4.3 Redes Neurais.....	57
3.4.4 <i>Random Forests</i>	62
3.4.5 <i>Support Vector Machines</i>	63
3.5 Métodos para a Avaliação de Modelos.....	65
3.5.1 Métricas de Avaliação de Modelos de Classificação.....	66
3.5.2 Métodos de Avaliação.....	69
4 Aplicação e Validação de Modelos de <i>Churn</i>.....	71
4.1 Conhecimento das Regras e Objectivos do Negócio.....	71
4.1.1 Enquadramento Experimental	71
4.2 Modelação	76
4.2.1 Metodologia de Testes.....	78
4.2.2 Construção dos Modelos	78
4.3 Avaliação dos Resultados	92
4.3.1 Resultados da Árvore de Decisão.....	92

4.3.2	Resultados da Regressão Logística	95
4.3.3	Resultados da Rede Neuronal.....	97
4.3.4	Resultados da <i>Random Forest</i>	99
4.4	Resultados da SVM.....	101
4.4.1	Síntese dos Resultados	103
4.4.2	Implicações financeiras.....	105
5	Conclusões e Trabalho Futuro	111
5.1	Comentários e Conclusões	111
5.2	Limitações e Trabalho Futuro	115
6	Bibliografia	117
7	Referências WWW	124

Índice de Figuras

Figura 1: Custo de novos clientes com saturação do mercado - adaptado de (Berry e Linoff, 2004)	2
Figura 2: <i>Churn</i> anual nos operadores de Telecom - adaptado de (Hung et al., 2006)	3
Figura 3: Volume de negócios empresas APED (APED, 2010)	5
Figura 4: Repartição por sector no ano 2010 (APED, 2010)	6
Figura 5: Evolução do parque de lojas dos associados APED (APED, 2010)	7
Figura 6: Ciclo de vida de um cliente - adaptado de (Berry e Linoff, 2004)	14
Figura 7: Processos de negócio - adaptado de (Berry e Linoff, 2004)	15
Figura 8: Aumento no retorno de um cliente - adaptado de (Reichheld e Teal, 1996)	17
Figura 9: Impacto do aumento da retenção - adaptado de (Van den Poel e Larivière, 2004)	18
Figura 10: Impacto da subida de 5% da taxa de retenção-adaptado de (Reichheld e Teal,1996)	19
Figura 11: Distribuição dos artigos revistos pelas categorias CRM (Ngai et al., 2009)	20
Figura 12: Percepção sobre os factores de <i>churn</i> - adaptado de (Thompson, 2005)	23
Figura 13: Número de artigos publicados por ano - adaptado de (KhakAbi et al., 2010)	26
Figura 14: Etapas do processo KDD - adaptado de (Fayyad et al., 1996)	33
Figura 15: Evolução na procura de informação - adaptado de (Lejeune, 2001)	34
Figura 16: Fases da Metodologia CRISP-DM - adaptado de (Chapman et al., 2000)	37
Figura 17: Períodos existentes nos dados para previsão de <i>churn</i>	40
Figura 18: Conjuntos de dados envolvidos na previsão de <i>churn</i>	41
Figura 19: Extração dos dados de treino - adaptado de (Yan et al., 2001)	43
Figura 20: <i>Data Mart</i> de vendas típico - baseado em (Kimball e Ross, 2002)	47
Figura 21: Tendências na previsão de <i>churn</i> - baseado em (KhakAbi et al., 2010)	53

Figura 22: Árvore de decisão - adaptado de (Han e Kamber, 2006)	54
Figura 23: Sentido de propagação dos dados - adaptado de (Cortez, 2002).....	59
Figura 24: Processamento num neurónio artificial - adaptado de (Han e Kamber, 2006).....	60
Figura 25: Retro propagação do erro - adaptado de (Cortez, 2002).....	61
Figura 26: Separação das classes pelo HLO - adaptado de (Han e Kamber, 2006).....	64
Figura 27: Mapeamento dos dados - adaptado de (Coussement e Van den Poel, 2008a)	65
Figura 29: Método de estimação <i>cross-validation</i>	70
Figura 30: Distribuição das classes dos clientes	76
Figura 31: Um fluxo de trabalho em RapidMiner	77
Figura 32: Resultados da Árvore de Decisão com a variação da densidade de <i>churners</i>	93
Figura 33: <i>Output</i> gerado pela técnica de árvores de decisão	94
Figura 34: Resultados da regressão logística com a variação da densidade de <i>churners</i>	96
Figura 35: <i>Output</i> gerado pela técnica Regressão Logística	97
Figura 36: Resultados da Rede Neuronal com a variação da densidade de <i>churners</i>	98
Figura 37: <i>Output</i> gerado pela técnica Rede Neuronal	99
Figura 38: Resultados da <i>random forest</i> com a variação da densidade de <i>churners</i>	100
Figura 39: <i>Output</i> gerado pela <i>random forest</i>	101
Figura 40: Resultados das SVM variando a densidade de <i>churners</i>	102
Figura 41: <i>Output</i> gerado pelo RapidMiner aquando da utilização das SVM.....	103
Figura 42: Impactos financeiros com o aumento da densidade de <i>churners</i> para a RL.....	108

Índice de Tabelas

Tabela 1: <i>Ranking</i> das empresas segundo volume de negócios (milhões €) (APED, 2010).	6
Tabela 2: Taxa de aderência dos consumidores ao cartão de fidelidade (TNSPortugal, 2008a).....	7
Tabela 3: Exemplo de retenção - adaptado de (Van den Poel e Larivière, 2004)	17
Tabela 4: Informação registada no POS - adaptado de (Tsiptsis e Chorianopoulos, 2010).....	45
Tabela 5: Tabela com a informação do cliente - adaptado de (Tsiptsis e Chorianopoulos, 2010) ..	46
Tabela 6: Matriz de confusão	66
Tabela 7: Variáveis independentes usadas neste estudo.....	73
Tabela 9: Avaliações da árvore de decisão com diferentes quantidades de atributos	80
Tabela 10: Resultados para diferentes valores de C e M.....	80
Tabela 11: Avaliações dos métodos de filtragem de atributos para a regressão logística	82
Tabela 12: Avaliações da regressão logística com diferentes quantidades de atributos	82
Tabela 13: Avaliações da técnica Regressão Logística variando o parâmetros I.....	83
Tabela 14: Avaliações dos métodos de filtragem de atributos para a rede neuronal.....	84
Tabela 15: Avaliações da rede neuronal com diferentes quantidades de atributos	84
Tabela 16: Avaliações medidas com a alteração do número de unidades escondidas (HU).....	85
Tabela 18: Avaliações dos métodos de filtragem de atributos para a <i>random forest</i>	86
Tabela 20: Avaliações da técnica <i>Random Forest</i> variando o parâmetros K.....	88
Tabela 21: Avaliações da técnica <i>Random Forest</i> variando os parâmetros K e I	88
Tabela 22: Avaliações da técnica SVM com diferentes funções <i>kernel</i>	89
Tabela 23: Avaliações dos métodos de filtragem de atributos para a SVM	90
Tabela 24: Avaliações da SVM com diferentes quantidades de atributos	90
Tabela 25: Resultados com a variação dos parâmetros γ e d	91
Tabela 26: Matriz de confusão da Árvore de Decisão	92

Tabela 27: Avaliação da técnica Árvore de Decisão	92
Tabela 28: Matriz de confusão da Regressão Logística	95
Tabela 29: Avaliação da técnica Regressão Logística.....	95
Tabela 30: Matriz de confusão da Rede Neuronal	97
Tabela 31: Avaliação da técnica Rede Neuronal	98
Tabela 32: Matriz de confusão da <i>Random Forest</i>	99
Tabela 33: Avaliação da técnica <i>Random Forest</i>	100
Tabela 34: Matriz de confusão da SVM	101
Tabela 35: Avaliação da técnica SVM	102
Tabela 37: Rentabilidades e Investimentos para as diferentes técnicas	106
Tabela 39: Rentabilidades para diferentes taxas de sucesso e densidades de <i>churners</i>	109

Siglas e Acrónimos

AUC - *Area Under Curve*

BI Business Intelligence

CLV – Customer Lifetime Value

CRM - *Customer Relationship Management*

CRISP-DM - *CRoss-Industry Standard Process for Data Mining*

DCBD - Descobrimto de Conhecimento em Base de Dados

KDD - *Knowledge Discovery in Databases*

LTV - *LifeTime Value*

OLAP - *Online Analytical Processing*

PCC - Percentagem de casos Corretamente Classificados

POS – Posto de Vendas

RFM – Recency, Frequency, Monetary

ROI - *Return On Investment*

SVM – *Support Vector Machine*

Capítulo 1

Introdução

1.1 Situações de *Churn* em Sistemas de Retalho

Atualmente vive-se num ambiente extremamente competitivo no mercado de venda de produtos ou de serviços. A população em geral recebe, frequentemente, propostas de produtos de um enorme número de empresas, com ofertas iguais ou similares. Além disso, existe ainda uma grande facilidade de acesso à informação sobre tais ofertas e respetivos fornecedores, o que leva a que os clientes sejam cada vez mais exigentes na sua escolha e, obviamente, mais propensos à mudança. É, pois, cada vez mais complicado às empresas manter os seus clientes ou mesmo arranjar novos clientes. Usualmente, estes não se vêm obrigados a manter uma ligação a uma companhia em particular, vão sistematicamente à procura de melhor qualidade ou de melhores preços. Este tipo de situação afecta principalmente empresas nas quais não há qualquer tipo de contrato ou assinatura, pois não existe qualquer restrição na altura do cliente optar por outra alternativa.

A perda de clientes, também designada por *churn* ou *customer attrition*, é, assim, uma preocupação em crescendo das empresas, uma vez que, para manter o nível de receitas, os clientes perdidos têm que ser substituídos por novos clientes. Estes, por sua vez, ficam bastante caros às empresas, dados os custos da sua angariação e, geralmente, geram menos receitas num futuro próximo que os clientes habituais. Isto é especialmente verdade nas indústrias onde o

mercado está bastante saturado. A Figura 1 ilustra o aumento dos custos na aquisição de novos clientes, devido à saturação do mercado e conseqüente diminuição na taxa de resposta da campanha de aquisições (Berry e Linoff, 2004).

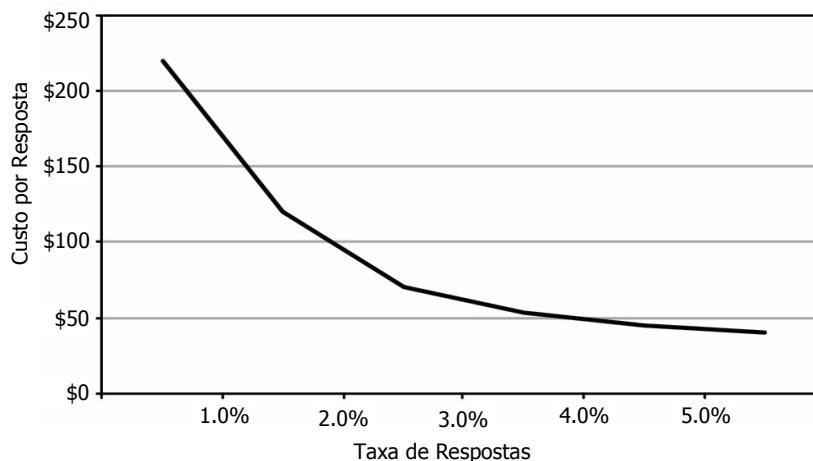


Figura 1: Custo de novos clientes com saturação do mercado - adaptado de (Berry e Linoff, 2004)

O fenômeno do *churn* requer que as empresas comecem a apostar em estratégias para o tentar evitar (ou pelo menos diminuir). No entanto, não é razoável, tentar reter todos os seus clientes, pois provavelmente estarão a investir naqueles que não dão lucro, estando a destruir algum valor da empresa, já que a retenção de clientes que não geram lucro é frequentemente lesiva para a empresa. Por outro lado, o dinheiro investido para os reter não é gasto a tentar reter os clientes que geram lucro.

A estratégia para evitar o *churn* deverá estar diretamente ligada ao valor dos clientes (Menezes e Portela, 2009). Como os clientes de longo prazo são menos custosos de servir e estão mais propensos a gerar valor para a empresa, deve ser sobre estes que o investimento deve recair, e não apenas, na atração de novos clientes (Zorn et al., 2008). Por outro lado, os clientes antigos promovem o fenômeno "*word-of-mouth*" ("de boca em boca") que tanto pode ser positivo, quando estes sugerem os serviços ou produtos da empresa a outras pessoas, como negativo, no caso dos clientes que abandonam a empresa insatisfeitos e passam uma má opinião da empresa, tendo neste caso um efeito lesivo para a empresa. Uma das chaves da análise do *churn* para além de identificar os clientes em risco é identificar as características que estão ligadas ao *churn* (Zorn et al., 2008).

A previsão de *churn* pode ser feita em diversos sectores da indústria e pode ter um grande impacto, sobretudo em sectores em que o mercado começa a ficar, ou já está, saturado e nos quais a aposta em manter os clientes é decisiva para o sucesso da empresa. Entre os diversos sectores de atividade, em que é efectuada a análise de *churn*, destacam-se as telecomunicações, a banca, as seguradoras, os operadores de televisão paga e o retalho. O *churn* significa o mesmo em todas as áreas, mas a maneira de reconhecê-lo diverge para cada uma delas.

O mundo das telecomunicações é uma das áreas onde o *churn* é relativamente simples de identificar, visto a relação empresa-cliente estar tipicamente oficializada por um contrato. Desse modo, as empresas sabem exatamente quando os clientes as abandonam. É também uma área muito sensível, na qual existe muita concorrência e os clientes podem mudar livremente de operador (Figura 2). Há uma relação significativa entre a lealdade do cliente, a sua satisfação, confiança e custo de mudança de operador no mercado das telecomunicações (Jahanzeb e Jabeen, 2007). A análise do *churn* tornou-se fundamental para as companhias de telecomunicações nos finais dos anos 90 (Jahanzeb e Jabeen, 2007).

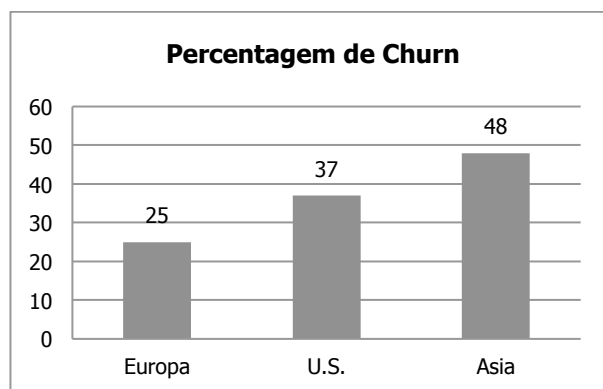


Figura 2: *Churn* anual nos operadores de Telecom - adaptado de (Hung et al., 2006)

1.1.1 Gestão de *Churn* no Retalho

Quando se fala de gestão de *churn*, não há nenhuma indústria que compreenda a inconstância dos clientes como a indústria do retalho. A informalidade da relação, devido à inexistência de um contrato, dá origem a que os clientes possam alterar os seus hábitos de forma repentina e

constante. Isto tudo porque não necessitam de informar a empresa que vão cessar a relação e, como tal, no mesmo dia podem estar a fazer as suas compras numa empresa da concorrência. No estudo *Wordpanel*, realizado pela TNSPortugal (2008b), verifica-se que, em média, cada lar visita 4 insígnias na distribuição moderna. Os retalhistas vêem-se assim obrigados a fazer um esforço permanente para convencer os clientes a permanecerem fiéis à sua empresa. Há dois aspectos que marcam esta batalha para manter clientes no retalho (Mattison, 2005):

1. Na sua maioria, os retalhistas têm acesso aos mesmos, ou similares, fornecedores. Este facto faz com que os clientes possam comprar os mesmos produtos nos diferentes retalhistas. Para além disso, a margem de manobra, sobre os preços dos produtos, está condicionada de forma semelhante.
2. Os retalhistas são obrigados a competir pela atenção dos clientes numa base diária, já que os clientes podem efetuar as compras em qualquer sítio, sem qualquer compromisso.

Num ambiente em que a competição entre as várias empresas é feroz (Buckinx e Van den Poel, 2005), as estratégias habituais, para manter os clientes fiéis, passam pela criação da "experiência da loja" (bom atendimento por parte dos funcionários, o ambiente da loja, a sua localização, etc.) e o recurso a campanhas promocionais agressivas para dirigir os clientes à sua loja (Mattison, 2005).

As novas estratégias adoptadas pelos retalhistas incluem a aposta nos produtos com marca do distribuidor e no aperfeiçoamento da aposta no consumidor, através do lançamento de páginas de comércio electrónico, cartões de fidelização e promoções cruzadas, para ir de encontro às necessidades e exigências do consumidor (Dias, 2009). Os retalhistas, reconhecem a necessidade de apostar na maximização das relações com os clientes e do consequente uso dos dados armazenados para apoiar essa estratégia. Com esta mudança para o foco no cliente e uso dos dados armazenados na tomada de decisões, surgiu a aposta no *Customer Relationship Management* (CRM). O uso de técnicas de *data mining* permite a utilização da enorme quantidade de dados, existente atualmente nos retalhistas, para apoiar os processos de CRM. No entanto, a pesquisa sobre a aplicação de técnicas de *data mining* no sector do retalho é limitada e principalmente focada em problemas diretamente relacionados com o *e-commerce* (Anderson et al., 2007).

As estratégias proactivas de prevenção de *churn* estão ainda pouco exploradas na área do retalho

e vão ganhando espaço na gestão de *churn* de clientes (Buckinx e Van den Poel, 2005). Contudo, também na aplicação deste tipo de estratégia surgem problemas relativos à inexistência de um contrato. Não tendo acesso aos dados do cliente que são disponibilizados quando é oficializado um contrato, as empresas não podem associar as transações aos clientes que as realizaram, ou seja, têm ao seu dispor uma enorme quantidade de dados relativos a transações de clientes efetuadas, mas não conseguem usá-los em campanhas de *marketing* direto. Para contornar este facto, cada vez mais empresas aderem à estratégia do cartão de fidelidade. Em troca de incentivos, que são, de modo geral, disponibilizados através de descontos em produtos da loja, os clientes disponibilizam os seus dados e ao apresentar o cartão no momento da compra é possível estabelecer a ligação entre o cliente e os dados relativos à transação por ele efectuada. A desvantagem desta solução é que nem todos os clientes aderem a esse mesmo cartão.

1.1.2 Caracterização do Sector do Retalho Português

O Mercado da Distribuição em Portugal - Ano 2010

No ano 2010, o conjunto das empresas que integram a *Associação Portuguesa de Empresas de Distribuição* (APED) registou um volume de negócios de 15.708 milhões de Euros, representando 9,1% do PIB nacional. O que representou um aumento de 6% face a 2009 (Figura 3) (APED, 2010).

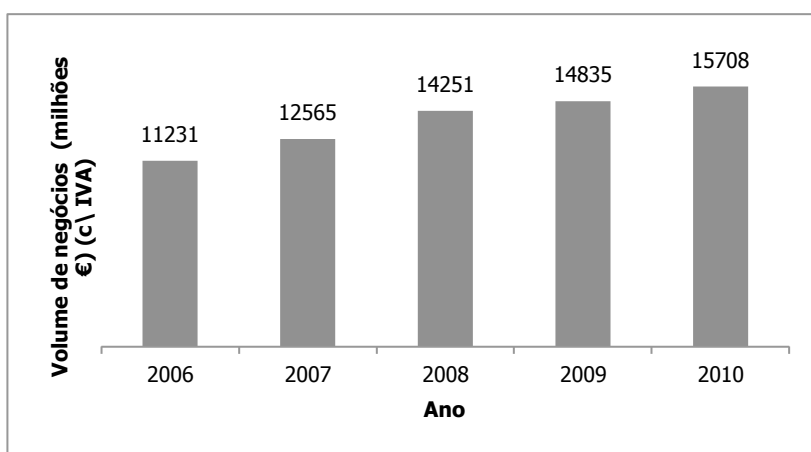


Figura 3: Volume de negócios empresas APED (APED, 2010)

O sector do calçado é o mais representado em número de empresas (33% dos associados APED). Todavia, o sector alimentar é, claramente, o sector dominante, representando 71% do volume de negócios do total das empresas associadas APED (Figura 4).

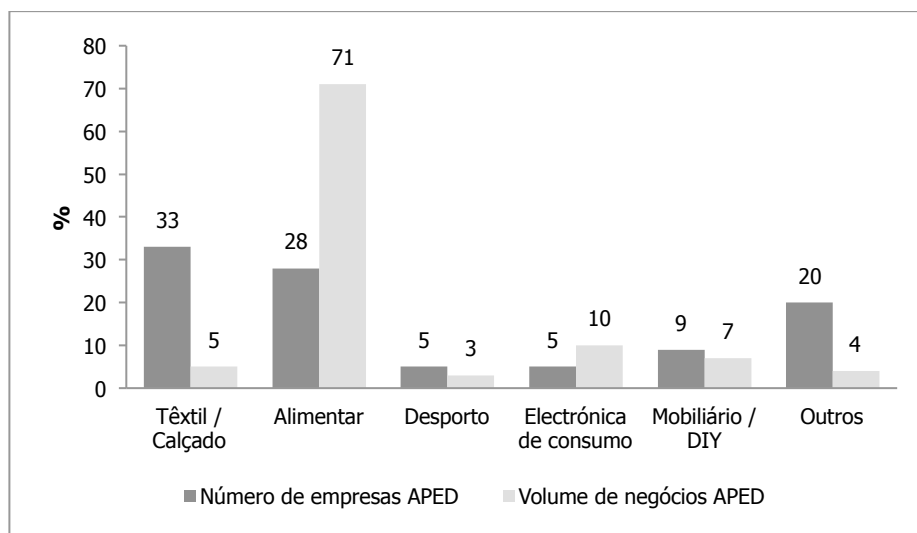


Figura 4: Repartição por sector no ano 2010 (APED, 2010).

As cinco principais empresas associadas APED, em relação ao volume de negócios, são o grupo Sonae Distribuição na liderança, a Jerónimo Martins, as internacionais Auchan e Lidl e finalmente em 5º lugar a cadeia Minipreço. É importante referir que a empresa os Mosqueteiros, uma das maiores empresas do sector, não está integrada na APED.

Tabela 1: *Ranking* das empresas segundo volume de negócios (milhões €) (APED, 2010).

Empresa	2010	2009	2010 vs 2009 (%)
Sonae	5203	5006	4
Pingo doce	3453	3112	11
Auchan	1601	1501	7
Lidl	1199	1211	-1
Minipreço	903	897	1

Os números relativos ao número de lojas dos associados APED comprovam o contínuo crescimento do sector e a grande competitividade existente (Figura 5). Nos últimos anos tem-se registado uma

desaceleração no crescimento do parque de lojas das empresas associadas APED, relativamente a anos anteriores, o que pode ser um indício da crescente saturação do sector.

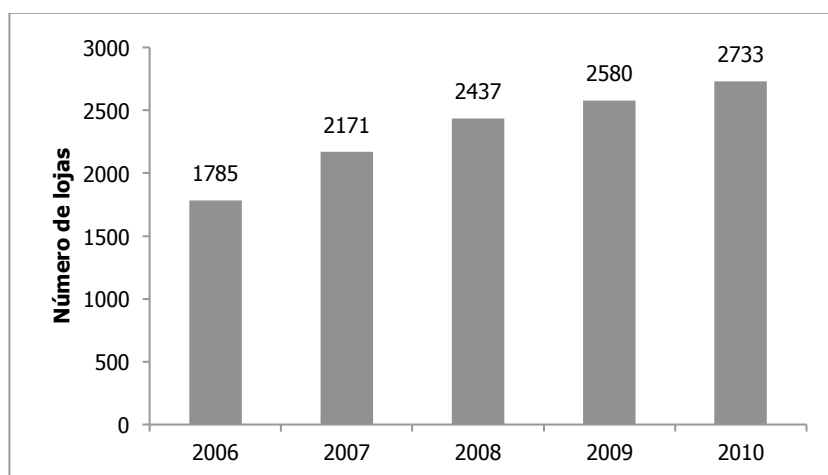


Figura 5: Evolução do parque de lojas dos associados APED (APED, 2010)

Adesão à Atividade Promocional

De acordo com o estudo *Retail Tracking* de 2008 da TNSPortugal (2008a), no acumulado até Julho de 2008, mais de metade dos entrevistados (65,3%) afirma ter pelo menos um cartão de fidelidade de um hipermercado ou supermercado, enquanto que os restantes 34,7% afirmam não ter qualquer tipo de cartão. Apesar de ter ocorrido uma diminuição ligeira face ao ano anterior, já que em 2007, 68,9% afirmava ter pelo menos um cartão deste tipo, houve um aumento incrível deste fenómeno, já que em 2006 apenas 5% dos consumidores possuíam um cartão de fidelidade.

Tabela 2: Taxa de aderência dos consumidores ao cartão de fidelidade (TNSPortugal, 2008a)

Cartão	Acumulado 08 (%)	Acumulado 07 (%)
Continente	43,4	23,9
Minipreço	39,9	32,4
Modelo	34,5	20,5
Os Mosqueteiros	25,4	17,8
Jumbo	13,7	9,8
E. Leclerc	7,3	3,5
Corte Inglés / SuperCor	2,3	1,6

Apesar da diminuição ligeira dos consumidores com pelo menos um cartão, todas as insígnias incluídas no estudo aumentaram o número de aderentes aos seus cartões, do ano 2007 para o ano 2008, destacando-se claramente o Continente no aumento de aderentes ao seu cartão (Tabela 2).

1.2 Motivação, Objectivos e Metodologia

Sendo o retalho um mercado maduro e estando já bastante saturado, a estratégia mais comum de aplicar o investimento na angariação de novos clientes perde algum interesse, tornando-se cada vez mais cara e com resultados cada vez mais pobres. As companhias começam a alterar a sua visão, sobre onde devem incidir os seus esforços, para procurar o aumento (ou mesmo para garantir a conservação) das receitas. Com este cenário, as empresas começam a mudar de atitude em relação à perda de clientes, dando cada vez mais importância a este fenómeno. Com a consciencialização do problema de *churning*, e sendo este, um fenómeno do qual as empresas podem retirar grande proveito no seu combate, pretende-se com este trabalho fazer um estudo mais aprofundado deste problema, assim como, o desenvolvimento de uma proposta de solução robusta para o combater ou tentar diminuir.

A previsão de *churn* na área do retalho torna-se apetecível, visto ser uma técnica ainda pouco explorada. Procura-se descobrir com a devida antecedência clientes que estejam em risco de se perder. Este estudo está direccionado e orientado ao sector do retalho, pois é um mercado, no qual, existe um défice de estudos em relação ao *churn* e ao seu combate. Deste modo, a motivação principal deste projeto esteve relacionada com a aquisição de conhecimento sobre as metodologias, práticas e aplicação de técnicas de *data mining* na criação de soluções para combater a perda de clientes, no contexto do sector do retalho. Foi portanto, o objectivo principal deste projeto a implementação de um modelo de *churn*, para servir como meio de combate ao fenómeno. Para isso, tornou-se necessário:

- Compreender o modelo de negócios do retalho. A estratégia de *churn* a aplicar está dependente do modelo de negócios. Por exemplo, a previsão do *churn* aplicada numa empresa de telecomunicações não pode ser aplicada numa empresa de retalho. Cada

modelo de negócio tem as suas próprias características, não se podendo extrapolar as conclusões e características de outros modelos já implementados.

- Definir a estratégia de prevenção de *churn*. Foi necessário definir uma série de características que determinaram qual a resposta a que o modelo implementado deve responder. Teve que ser definido o modo de atuação do modelo de *churning*.
- Estudar e compreender os dados disponíveis. Como existe uma grande variedade de dados, que podem ser usados no problema de previsão de *churn*, tornou-se necessário compreender perfeitamente o seu significado. Uma análise cuidadosa e prévia dos dados é fundamental para evitar eventuais ambiguidades e erros de interpretação.
- Escolher um subconjunto desses dados e características com maior relevância para o estudo, para aplicar no problema. Diferentes conjuntos de dados adequam-se de forma diferente a diferentes problemas e aos vários sectores do serviço. A identificação dos melhores dados numa fase inicial é um factor determinante para a eficiência e precisão do modelo (Hadden et al., 2007). Foi também necessário escolher as variáveis a introduzir no modelo, aquelas que eram melhores para o processo de previsão. É uma etapa importante porque auxilia a limpeza e redução dos dados, seleccionando as variáveis importantes e excluindo as redundantes, as que apenas são fonte de ruído e as que tem pouco poder informativo (Hadden et al., 2007).
- Determinar o(s) modelo(s) a aplicar no problema. Havendo uma vasta variedade de modelos de previsão foi necessário escolher aqueles que melhor se adequam ao problema e os que aumentam a qualidade e validade dos resultados.

Para pôr em prática e concretizar os objectivos que foram propostos, foi necessário numa primeira fase, fazer um levantamento de requisitos e uma recolha de informação mais exhaustiva, para obter um plano devidamente sustentado e com os objectivos perfeitamente delineados. Posteriormente, já com o devido estudo e compreensão do problema foram preparados e escolhidos os dados mais apropriados para o problema e foi criado o conjunto de dados sobre o qual se aplicaram os modelos. Só então, foram efectuados todos os testes, escolhas de modelos e implementações necessárias para por em prática uma solução para o problema de *churn*. Por fim, foram avaliados e discutidos os resultados obtidos e foi definida a estratégia a aplicar.

1.3 Organização da Dissertação

Para além do presente capítulo, esta dissertação está organizada em mais quatro capítulos, nomeadamente :

- **Capítulo 2 - Análise do Fenómeno de Churn.** Este capítulo fornece uma visão global da relação empresa-cliente, dos estágios pelos quais o cliente passa nessa relação e dos processos aplicados pelas empresas ao longo desses estágios. É, também, abordado o processo de retenção com maior pormenor e são enumeradas algumas razões para apostar neste tipo de estratégia. É discutido o fenómeno de *churn*, as suas vertentes e os factores que o rodeiam e por fim este capítulo foca-se na gestão do *churn* e na modelação para previsão de *churn*.
- **Capítulo 3 - Previsão de Situações de Churn.** Aqui pretendeu-se discutir os principais assuntos envolvidos na previsão de *churn*. Este capítulo é iniciado com uma descrição superficial do processo de descoberta de conhecimento em base de dados e a sua fase de maior relevo, a fase de *data mining*. De seguida, faz-se um levantamento das principais questões e problemas relacionados com a construção das assinaturas dos clientes. São dados alguns exemplos de trabalhos relacionados e por fim é feita uma pequena revisão de algumas técnicas de classificação e de alguns métodos de avaliação de modelos de *data mining*.
- **Capítulo 4 - Aplicação e Validação de Modelos de Churn.** O quarto capítulo começa com a apresentação de um cenário sobre o qual se irá desenvolver o estudo. Apresenta a metodologia adoptada no projeto bem como descreve as experiências feitas na produção do conjunto de dados e na fase de modelação. Por fim, dedica-se à fase de avaliação dos modelos sendo feita a descrição dos resultados obtidos na tarefa de modelação.
- **Capítulo 5 - Conclusões e Trabalho Futuro.** Por fim chega-se ao capítulo da conclusão em que são discutidas as limitações e possibilidades de extensão do trabalho no futuro, e obviamente, são tecidas algumas conclusões sobre o estudo desenvolvido.

Capítulo 2

Análise do Fenómeno de *Churn*

2.1 Uma Cultura Centrada no Cliente

Com a revolução industrial surgiu a produção em massa, as empresas começaram a apostar em estratégias de marketing não direcionadas, orientadas ao produto, onde o objectivo era chegar ao maior número de clientes possível (Rygielski et al., 2002). Não obstante, com o decorrer dos anos, o contexto económico sofreu alterações profundas, essas estratégias não tinham capacidade para lidar com os problemas que entretanto surgiram devido ao aumento da competitividade e da saturação em diversas áreas de negócios. Assim, foi necessário às empresas a procurarem melhores soluções para os seus problemas, ou mesmo em alguns casos, mudar completamente a sua estratégia. Como consequência, tem-se assistido nos últimos anos a uma mudança de cultura no mundo dos negócios. As empresas percebem agora a importância da relação com os clientes e da sua gestão apropriada, estando estas a substituir o modelo estratégico centrado no produto pelo modelo centrado no cliente (Rygielski et al., 2002). Este último assenta na utilização de técnicas e ferramentas que permitem obter um conhecimento mais adequado do cliente, para que possam ser aplicadas técnicas de marketing direcionadas, ou seja, técnicas adaptadas a cada cliente em particular.

Esta mudança de paradigma necessitou do impulso da evolução tecnológica e consequente criação de ferramentas para a suportar (Parvatiyar e Sheth, 2001; Berry e Linoff, 2004). Num ambiente em que as empresas têm milhares de clientes, só é possível um conhecimento adequado dos clientes com o auxílio de infraestruturas e tecnologias de sistemas de informação adequadas. É necessário a captura e armazenamento de dados sobre os clientes para análise posterior. O aparecimento de tecnologias de base de dados, sistemas de *data warehousing*, sistemas *Online Analytical Processing* (OLAP), técnicas de *data mining* e ferramentas de *Customer Relationship Management* (CRM) proporcionou uma relação individualizada com os clientes e a aplicação de técnicas de marketing direcionado (Rygielski et al., 2002).

Outro factor para a mudança de cultura nas empresas foi a facilidade da obtenção de informação por parte dos consumidores (Rygielski et al., 2002). Num mundo extremamente competitivo, em que a existência da internet e dos mais diversos canais de informação facilita o acesso a informação sobre produtos, serviços e alternativas existentes no mercado, os consumidores tornaram-se mais informados, procurando de forma já muito sistemática o melhor produto ou aquele que mais se lhe adequa às suas necessidade ou possibilidades (Rygielski et al., 2002; Lejeune, 2001). Para fazer face a estas contrariedades, hoje as empresas necessitam de arranjar meios para se distinguirem nos mercados nos quais competem, sendo que uma interação personalizada com o cliente é uma maneira eficaz de o fazer. Essa interação direcionada é possível com a recolha de dados demográficos e comportamentais do cliente.

Este tipo de contextos leva as empresas a investir no sentido de aumentar o valor dos clientes, analisando as diferentes fases do seu ciclo de vida (Rygielski et al., 2002). É nesta análise das fases do ciclo de vida de um cliente, que entra o CRM. Com a percepção do elevado custo da atracção de novos clientes e da maior rentabilidade dos clientes antigos, começa a fazer sentido que as empresas mudem o seu foco de ação e apostem na manutenção de relações de longa duração com os clientes. O desenvolvimento de relações e uma gestão orientada para o cliente são factores chaves no sucesso de uma empresa (Lejeune, 2001).

2.1.1 Organização Centrada no Cliente

Para gerir a relação que tem com os clientes, uma empresa necessita de uma grande quantidade de informação sobre eles. É necessária a criação de procedimentos para tirar proveito dessa informação (Lejeune, 2001). As empresas registam transações, comunicações e interações com os clientes e relacionam essa informação com os seus dados identificativos e demográficos. A exploração dos dados guardados nas bases de dados das empresas permitirá posteriormente prever o comportamento dos clientes, o seu valor, nível de satisfação e ainda o seu nível de lealdade. Contudo, a exploração e uso dessa enorme quantidade de dados, para melhor servir e ganhar valor dos clientes, pode ser um processo bastante complexo. Exige que as empresas comprem *software* comercial especializado ou implementem soluções à medida, baseadas nas muitas técnicas de *data mining* conhecidas e disponíveis no mercado. Deste modo, uma empresa centrada no cliente é aquela que tira partido dos dados adquiridos sobre os clientes, aprende a responder melhor às suas necessidades, ao longo do seu ciclo de vida, ganhando com isso vantagem sobre a concorrência, o que é fundamental em mercados considerados saturados (Berry e Linoff, 2004; Hadden et al., 2007).

2.1.2 Ciclo de Vida do Cliente

A compreensão dos diferentes estágios de uma relação cliente-empresa, também conhecidos no seu conjunto como ciclo de vida de um cliente (Figura 6), é bastante relevante para potencializar a rentabilidade do cliente (Rygielski et al., 2002). Estando as empresas orientadas para uma cultura centrada no cliente, a visão que têm dele tem que refletir isso mesmo. A percepção dos clientes como sendo agentes estáticos está longe de estar correta. Estes, são entidades dinâmicas que ao longo da relação com a empresa passam por várias fases, podendo o seu comportamento variar entre cada uma delas (Berry e Linoff, 2004).

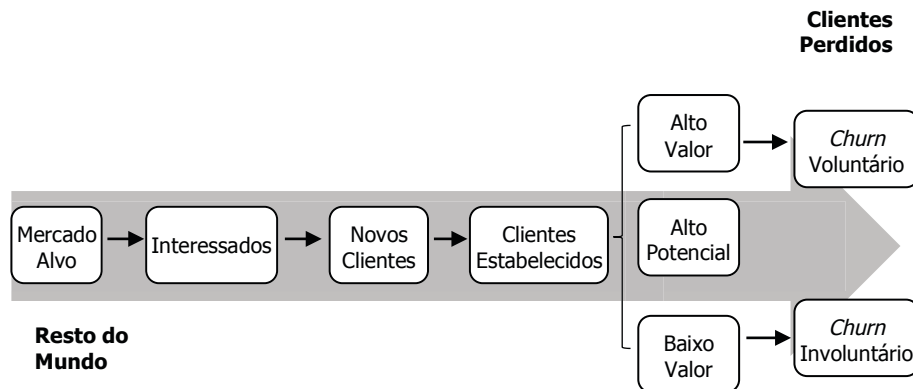


Figura 6: Ciclo de vida de um cliente - adaptado de (Berry e Linoff, 2004)

Em geral podem-se identificar quatro estágios fundamentais no ciclo de vida de um cliente (Berry e Linoff, 2004; Rygielski et al., 2002):

- Clientes Potenciais, que estão no mercado onde a empresa atua, mas não são clientes.
- Clientes Interessados, que estão no mercado e demonstram interesse na empresa como, por exemplo, os que se registam no *website* da empresa.
- Clientes Ativos, que estão atualmente a adquirir serviços ou produtos. Estes podem ser divididos em dois subgrupos:
 - o Novos Clientes, todos aqueles que recentemente adquiriram pela primeira vez um produto ou serviço.
 - o Clientes Estabelecidos, que são os que regressam e voltam a adquirir produtos ou serviços. Destes é esperado que a sua relação com a empresa se vá alargando e aprofundando.
- Clientes Perdidos, que são aqueles que já foram clientes, mas que por algum motivo deixaram de o ser.

2.1.3 Processos de Negócio no Ciclo de Vida de um Cliente

Atualmente, verifica-se a necessidade das empresas adoptarem diferentes estratégias para as diferentes fases da relação e para os diferentes tipos de clientes. A relação que a empresa tem com um cliente novo deverá ser bastante distinta da relação com um cliente antigo. As empresas

querem tornar os clientes mais valiosos com o desenrolar da relação, usando diversos processos de negócio para chegar a esse objectivo (Figura 7).

Os principais processos de negócio utilizados para acompanhar o ciclo de vida do cliente são (Berry e Linoff, 2004):

- Aquisição: é o processo de captar clientes potenciais para a empresa. É feito através de campanhas de publicidade, pelo processo "word-of-mouth" ou campanhas de marketing direcionado.
- Ativação: é o processo de oficializar como clientes as pessoas que demonstraram interesse em adquirir produtos ou serviços. Existe apenas em empresas que mantêm uma relação contratual com o cliente.
- Gestão (Desenvolvimento) da relação: é o processo responsável pelo desafio de potenciar o valor dos clientes para a empresa.
- Retenção: é o processo que trata da manutenção dos clientes estabelecidos. É um processo com grande relevância para as empresas e um dos que é aplicado mais frequentemente.
- Recuperação: é a tentativa de recuperar os bons clientes que abandonaram recentemente a empresa para a concorrência, através da oferta de incentivos.

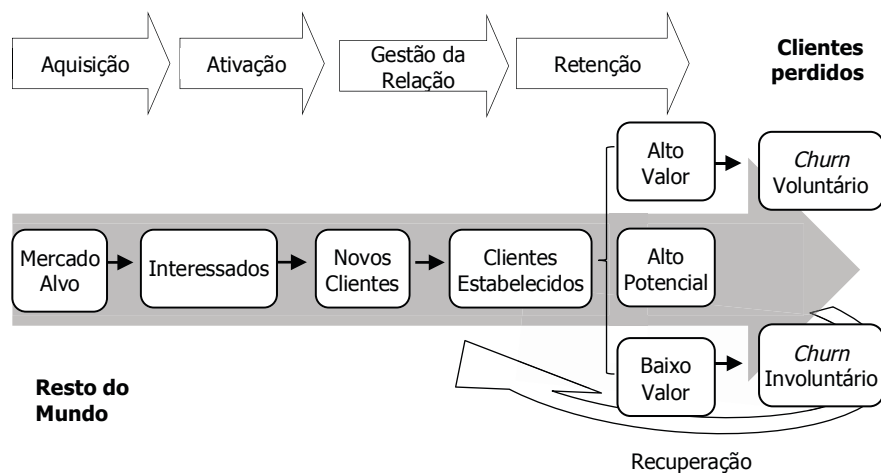


Figura 7: Processos de negócio - adaptado de (Berry e Linoff, 2004))

2.2 Retenção de Clientes

A retenção de clientes é um dos processos mais importantes para as empresas (Burez e Van den Poel, 2009; Ngai et al., 2009) e onde a modelação preditiva é aplicada mais frequentemente (Berry e Linoff, 2004). Para além disso, é defendido que, em algumas situações, o custo de adquirir um novo cliente é no mínimo 5 vezes mais caro do que a retenção de um cliente existente (Chu et al., 2007; Menezes e Portela, 2009).

Reichheld e Teal (1996) defendem que os lucros gerados por um cliente aumentam com o tempo que este permanece na empresa (Figura 8) e apontam vários factores pelos quais é importante investir na retenção de clientes:

- Custo de aquisição. A poupança em custos associados a campanhas de publicidade, custos de instalação, comissões nas vendas a novos clientes, entre outros. Naturalmente, este ponto é aqui referido devido à sua inexistência. Exemplificando: uma empresa de cartões de crédito que faça uma campanha de correspondência para adquirir clientes que tenha uma taxa de resposta de 2 ou 3 por cento, terá que enviar 30 mil ou 50 mil solicitações para adquirir umas simples mil aplicações. Adicionando os custos relacionados a adaptar o cliente ao sistema e os preços de aquisição por cada novo cliente irão para valores entre os 50\$ e 100\$.
- Lucro base. Este é o lucro natural que advém da venda dos produtos ou serviços que os clientes adquirem. Normalmente, os preços pagos pelos clientes são superiores aos custos que a empresa tem. É o lucro básico não influenciado por factores como tempo da relação, lealdade ou eficiência. Quanto mais tempo se mantiver um cliente, mais tempo se lucrará com o seu lucro base, verificando-se a boa aposta no investimento de aquisição.
- Crescimento da receita. Na maioria dos negócios, os gastos praticados pelos clientes tendem a acelerar com o tempo. Os clientes tornam-se cada vez mais familiarizados com a linha de produtos ou serviços oferecidos pela empresa, o que os leva a comprar outros produtos para além daqueles que compravam inicialmente.
- Corte nos custos. Quanto mais os clientes conhecem o negócio mais eficientes se tornam na relação. Sabem o que lhes é oferecido e o que não é. Tornam-se mais autónomos, poupando desta forma, tempo aos empregados.

- Referências. Clientes satisfeitos com um negócio recomendam-no a outras pessoas. É também conhecido como fenómeno "word-of-mouth". Por outro lado, os clientes que abandonam o negócio insatisfeitos, podem promover o mesmo fenómeno, mas de forma negativa, ou seja, influenciando outras pessoas contra o negócio (Ng e Liu, 2000).
- Tolerância a preços. Clientes de longa data são mais tolerantes a preços elevados.

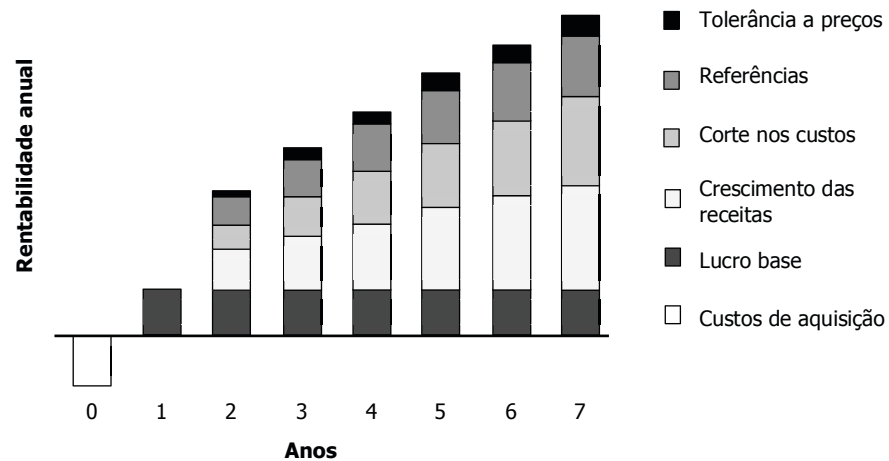


Figura 8: Aumento no retorno de um cliente - adaptado de (Reichheld e Teal, 1996)

2.2.1 Impacto nos Lucros

Tabela 3: Exemplo de retenção - adaptado de (Van den Poel e Larivière, 2004)

Taxa de Retenção (TR)	Número de clientes (milhares)					CM ¹ por ano (Euro)	CM ¹ total depois de 25 anos (milhares de euros)	CM ¹ adicional sobre a situação de 93% de retenção (milhares de euros)	CM ¹ adicional sobre a situação de 75% de retenção (milhares de euros)
	Início	2º Ano	3º Ano	...	25º Ano				
75%	1000	750	563	...	1	50	174 688	/	0
93%	1000	930	865	...	175	50	392 213	0	217 525
94%	1000	940	884	...	227	50	419 757	27 544	245 069
100%	1000	1000	1000	...	1000	50	682 518	290 305	507 830

¹ Contribuição Média

Utilizando um exemplo real de uma empresa da área financeira, Van den Poel e Larivière (2004) mostram que, um simples aumento de 1% na taxa de retenção pode gerar lucros significativos para a empresa. No exemplo, supõe-se que a empresa tem 1 milhão de clientes, nos quais a média da taxa de *churn* é de 93% e a contribuição média anual dos clientes é de 50€, o que corresponde à situação real da empresa. A Tabela 3 compara os dados económicos de várias situações alternativas. Foram incluídos os resultados para uma taxa de retenção de 75% por ser a situação de um grande banco do Reino Unido. É ainda incluída a situação onde a aplicação de técnicas de gestão de *churning* proporciona o ganho de um ponto percentual na taxa de retenção, isto é, permite alcançar uma taxa de retenção de 94%. É demonstrado na tabela que este ganho residual proporciona um aumento substancial nas receitas. As receitas aumentaram de 392.2 milhões de Euro (a uma taxa de 93% de retenção) para 419.7 milhões de Euro (a uma taxa de 94% de retenção), ou seja, um aumento de 27.5 milhões de Euro, considerando um período de 25 anos e uma taxa de desconto de 6% (Tabela 3). O impacto deste aumento na retenção no lucro pode ser visualizado na Figura 9.

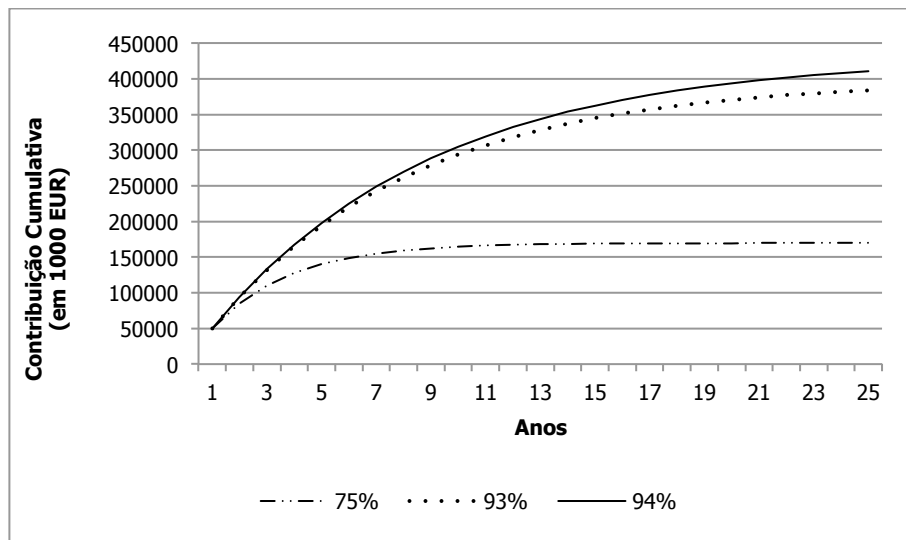


Figura 9: Impacto do aumento da retenção - adaptado de (Van den Poel e Larivière, 2004)

A Figura 10, mostra o impacto da subida de 5% da taxa de retenção no *Net Present Value* (NPV), médio por cliente em várias indústrias. Por exemplo, se uma empresa de cartões de crédito aumentar a taxa de retenção em 5% (aumentando, por exemplo, de 90% para 95%) então os

lucros totais do ciclo de vida de um cliente irão aumentar, em média, cerca de 75% (Reichheld e Teal, 1996).

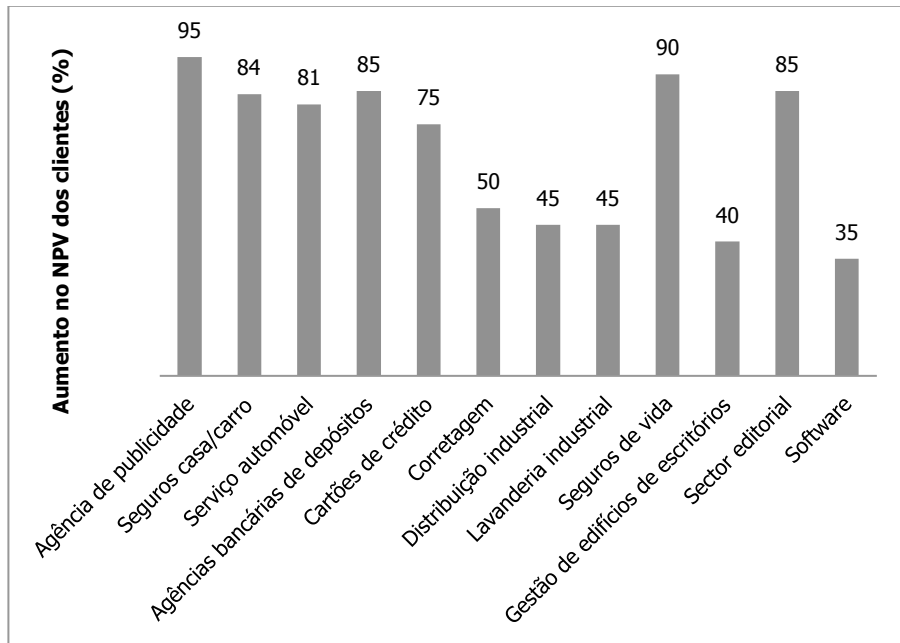


Figura 10: Impacto da subida de 5% da taxa de retenção - adaptado de (Reichheld e Teal, 1996)

2.2.2 Análise de Números de Retenção

Ngai et al. (2009) fizeram uma revisão de literatura académica sobre a aplicação de técnicas de *data mining* em CRM na qual expõem uma base de dados de artigos proveniente de 24 periódicos científicos, referentes ao período entre 2000 e 2006. Nesse trabalho foram revistos 900 artigos, dos quais 87 foram selecionados, revistos e classificados. Os artigos foram categorizados segundo 4 dimensões de CRM: identificação, atração, retenção e desenvolvimento de clientes. Estas dimensões não correspondem exatamente aos processos identificados anteriormente neste trabalho (secção 2.1.3), mas permitem perceber a importância da retenção (esta dimensão corresponde exatamente ao processo identificado com o mesmo nome anteriormente) no seio da comunidade académica (Figura 11).

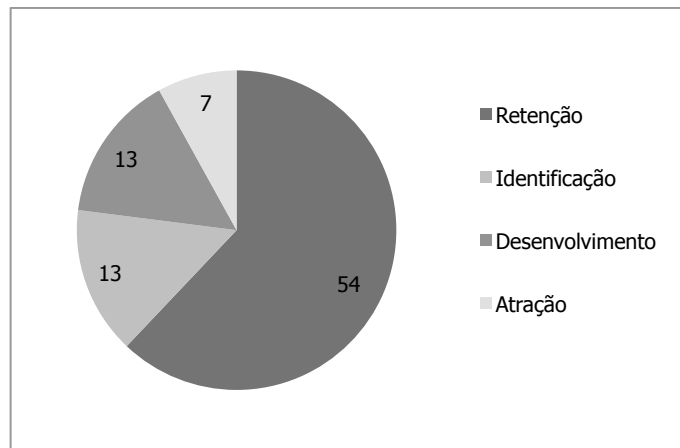


Figura 11: Distribuição dos artigos revistos pelas categorias CRM (Ngai et al., 2009)

2.2.3 Em Prol da Retenção dos Bons Clientes

Os clientes não são todos iguais. Reichheld e Teal (1996) salientam 3 regras que as empresas devem ter em consideração na análise dos seus clientes:

1. "Alguns clientes são inerentemente previsíveis e leais, não interessando qual a companhia com quem estão a fazer o negócio, simplesmente preferem relações estáveis e duradouras."
2. "Alguns clientes são mais lucrativos que outros, gastam mais dinheiro, pagam as contas prontamente e requerem menos serviço das empresas."
3. "Alguns clientes irão ver os produtos ou serviços da empresa mais valiosos do que os da concorrência. Nenhuma empresa pode ser "tudo" para toda a gente. As mais-valias da empresa irão simplesmente servir melhor as necessidades e oportunidades de certos clientes."

A lealdade dos clientes pode ser vista em um, ou mais, destes 3 pontos. As empresas podem dirigir as suas estratégias de retenção a todos os clientes em risco, ou então, no seio desses clientes, focar ou priorizar essas estratégias apenas nos clientes mais lucrativos (Hung et al., 2006).

O Valor de um Cliente

O cálculo do valor de um cliente pode ser bastante complicado. Numa reflexão inicial, pode-se pensar que é uma simples subtração dos custos, associados à manutenção do cliente, às receitas geradas por ele. No entanto, a dificuldade reside em calcular as parcelas do cálculo. Por exemplo, terá de se definir a janela temporal para o cálculo das receitas de um cliente (último mês, último ano, etc.) e se estas deverão incluir previsões das suas receitas futuras. Por outro lado, os gastos de um cliente também podem ser complicados de calcular, i.e., dos montantes aplicados em campanhas de publicidade e de retenção não direcionadas, qual será o montante a atribuir ao cliente em questão? Só se poderá efetuar o cálculo do valor do cliente quando todas estas (e outras) questões estiverem respondidas e quando estiver definida uma fórmula para o cálculo desse valor (Berry e Linoff, 2004).

2.3 Situações de *Churn*

2.3.1 Definição e Taxa de *Churn*

Para pôr em prática uma estratégia de retenção o primeiro passo será detectar com antecedência os clientes que poderão abandonar a empresa no futuro. Ou seja, prever o *churn*. De forma semelhante a Chu et al. (2007) e Jinbo et al. (2007) o termo *churn* é definido neste trabalho como “a quebra da relação entre uma empresa e um cliente”. Do mesmo modo, se denomina *churner* o cliente para o qual ocorreu o evento de *churn* e *churning* o ato de um cliente incorrer em *churn*. Para além da definição do significado gramatical da palavra *churn* é também necessário definir conceptualmente o evento (secção 3.2.3), ou seja, definir, na prática, quando é considerado que houve *churn* de um cliente.

A taxa de *churn* é um dado estatístico sobre a quantidade de *churn* num determinado período de tempo (Page et al., 1996). Por exemplo, se não se levar em consideração os clientes adquiridos, numa empresa que tem 100 clientes, dos quais, num período de 1 ano, 10 abandonam a empresa, a taxa de *churn* anual é na ordem dos 10%. Levando os clientes adquiridos em conta a taxa de *churn* mensal pode ser calculada como (Jahanzeb e Jabeen, 2007):

$$\textit{Churn Mensal} = \frac{C0 + A1 - C1}{C0}$$

em que C0 é número de clientes no início do mês, C1 o número de clientes no final do mês e A1 o número bruto de novos clientes durante o mês

2.3.2 Factores de *Churn*

Os factores que levam ao *churn* podem variar muito de indústria para indústria. Ghorbani e Taghiyareh (2009) defendem que a investigação das causas do *churn* é o primeiro pré-requisito no desenvolvimento de um sistema para combater o fenómeno. As causas que levam os clientes a partir podem ter diversas origens, tais como, o serviço prestado ao cliente, qualidade, preço, funcionalidade ou conveniência.

Uma pesquisa da CMRGuru (Thompson, 2005) revelou que os gestores das empresas podem não ter a noção correta das razões que levam ao abandono dos clientes. Não estão em sintonia com o sentimento que os clientes revelam. A Figura 12 mostra a diferença na percepção dos factores que levam o *churn* entre os clientes e os gestores. Enquanto o serviço ao cliente e a qualidade são os factores mais relevantes nas respostas dos clientes, para as empresas os factores mais relevantes são o preço e a mudança das necessidades dos clientes. Como todos os negócios têm as suas particularidades, os gestores devem refletir sobre os seus pressupostos e investir no estudo das razões que levam os clientes a partir. É próprio da natureza humana procurar a culpa noutra lugar (Thompson, 2005).

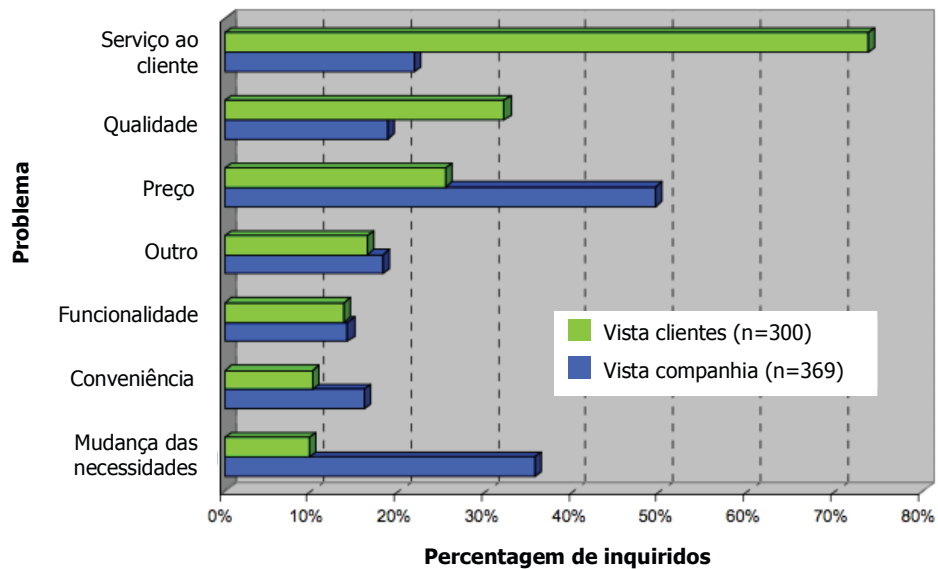


Figura 12: Percepção sobre os factores de *churn* - adaptado de (Thompson, 2005)

2.3.3 Os Diferentes Tipos de *Churn*

O *churn* tem diversas origens, que podem ser bastante distintas. Como tal, é importante saber distingui-las. É perigoso tratar da mesma forma clientes que a empresa quer manter e clientes que não são importantes para a empresa. Por exemplo, a oferta de incentivos a um cliente que foi classificado como estando em risco de *churn*, quando na verdade este era um mau cliente que não trazia retorno à empresa. Por outro lado, um bom cliente pode abandonar a empresa, em consequência de esta chegar à percepção errada de que este é um mau cliente e o tratar em conformidade com essa situação.

***Churn* Involuntário**

Verifica-se quando é a empresa a terminar a relação e não o cliente. Normalmente ocorre pelo não-pagamento das contas, o não-cumprimento do contrato estabelecido ou abuso do serviço. Consequentemente, no contexto da retenção de clientes, as empresas não têm interesse na previsão deste tipo de *churn* (Hadden et al., 2007; Berry e Linoff, 2004).

Churn Voluntário

O *churn* voluntário é mais difícil de detectar, pois a decisão de quebrar a relação parte do cliente (Berry e Linoff, 2004). Este tipo de *churn* pode ter causas circunstanciais como, por exemplo, alterações na situação financeira do cliente que o levam a deixar de ter a possibilidade de pagar o produto, ou, então, quando o cliente muda de área de residência e o serviço não está disponível na nova área de habitação. Por outro lado o *churn* voluntário pode ser deliberado. É este tipo de *churn* a principal preocupação das empresas. Ocorre quando o cliente decide procurar o mesmo produto num competidor da empresa. Entre várias possíveis razões, este tipo de *churn* pode ter origens financeiras, quando, por exemplo, a empresa não consegue oferecer preços tão bons como os seus competidores, pode estar relacionada com factores de qualidade como, por exemplo, o mau atendimento por parte de um funcionário ou quando um serviço não corresponde à sua descrição (Hadden et al., 2007).

Churn Parcial

Quando se fala de *churn* parcial, refere-se à situação em que o cliente transfere alguns dos seus gastos que previamente eram efectuados na empresa para uma outra empresa da concorrência. O cliente diminui a sua atividade na empresa, todavia, não a termina completamente. Este tipo de *churn* pode conduzir ao *churn* total (Buckinx e Van den Poel, 2005). Este tipo de *churn* pode beneficiar a detecção do *churn* considerado como total, já que, pode ser visto como um sinal de aviso dessa situação.

2.4 A Gestão de *Churn*

"A gestão do *churn* consiste no desenvolvimento de técnicas que permitam às empresas manter os seus clientes lucrativos e procura aumentar a lealdade dos clientes" (Lejeune, 2001). Pode ser visto como um processo de *Business Intelligence* (BI) que tem como objectivos prever os clientes que vão abandonar a empresa e avaliar a maneira mais eficaz de atuar para evitar esse fenómeno (Hung et al., 2006).

As campanhas de retenção aplicadas à totalidade dos clientes podem trazer excelentes resultados relativos aos números de retenção, mas podem também tornar-se bastante dispendiosas (Berry e Linoff, 2004). Ao tentarem reter todos os clientes, as empresas desperdiçam recursos, e as campanhas de retenção tornam-se ineficazes e não lucrativas (Hadden et al., 2007; Hu, 2005). O problema reside no facto de que os incentivos ou programas para os clientes permanecerem na empresa serão aceites pela generalidade dos clientes, incluindo aqueles que não estariam em risco de ser perdidos. Para contornar esse desperdício de recursos, as campanhas de retenção devem ser direccionadas apenas aos clientes que estão em risco. Para além disso, as empresas poderão controlar ainda mais os custos e, de entre os clientes que estão em risco, escolher apenas os que geram maior valor para empresa. Isto significa que, não vale a pena reter clientes dos quais o seu retorno não compensa os custos de manutenção e retenção (Hadden et al., 2007). Existem, então, dois tipos de campanhas direccionadas:

- Reativas. As empresas só agem quando o cliente pede para terminar o contrato ou, no caso de relações não contratuais, quando percebem que o cliente partiu para a concorrência. Só nesse momento é que as empresas oferecem incentivos para tentar contrariar essa situação (Burez e Van den Poel, 2007).
- Proactivas. As empresas tentam identificar os clientes que estão em risco e agem com alguma antecedência, oferecendo incentivos especiais, para prevenir o *churn*. Este tipo de estratégia tem a vantagem, potencial, de ter custos mais reduzidos (Burez e Van den Poel, 2007; KhakAbi et al., 2010).

2.4.1 Controlando o *Churn* Proactivamente

A prevenção do *churn* de clientes, como parte de uma *framework* de CRM e em particular de uma estratégia de retenção, tem gerado um interesse crescente. Para pudermos detectar com antecedência os clientes que estão em risco de *churn*, é do interesse das empresas, implementar modelos de previsão de *churn* baseados em técnicas de *data mining* (Burez e Van den Poel, 2009). Esse crescimento do interesse na gestão de *churn* é demonstrado numa revisão de literatura, proposta por KhakAbi et al. (2010), onde foram examinados 32 artigos, publicados entre 2003 e 2009, sobre o assunto (Figura 13).

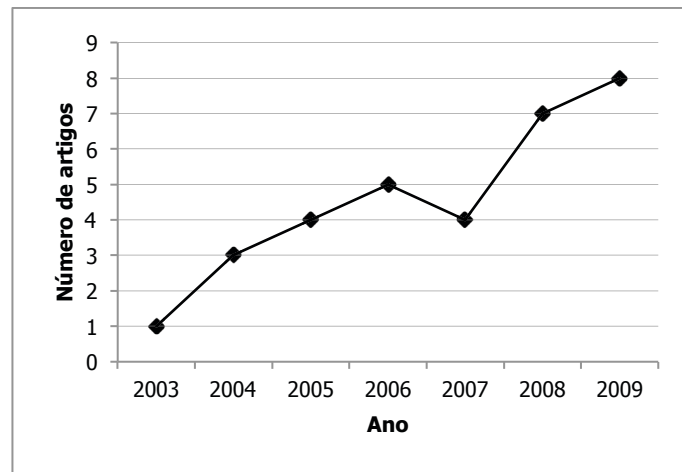


Figura 13: Número de artigos publicados por ano - adaptado de (KhakAbi et al., 2010)

Diferentes Ambientes na Previsão de *Churn*

Quando se fala de detectar que um cliente abandona uma empresa têm que ser referidos dois cenários distintos. A dificuldade e a forma de detectar *churn* varia largamente para os ambientes em que estão em causa os seguintes tipos de relacionamentos:

- **Relações Contratuais:** Estas são caracterizadas pela existência de um contrato que as formaliza. É assinado um contrato quando se estabelece a relação e, geralmente, quando se quer quebrar o contrato, comunica-se essa intenção à outra parte envolvida no negócio. Neste caso o *churn* é detectado facilmente, pois tem que existir uma comunicação entre as partes para a quebra do contrato. São exemplo de empresas que praticam este tipo de relação as seguradoras, bancos e determinadas empresas de telecomunicações.
- **Relações não Contratuais:** Aqui como não existe uma ligação oficial, a detecção de *churn* pode tornar-se muito complicada, pois os clientes podem mudar de empresa facilmente sem ter que informar previamente (Berry e Linoff, 2004; Buckinx e Van den Poel, 2005). Em grande parte das empresas da área de retalho é praticado este tipo de relação.

Definição do “Momento de *Churn*”

A forma de reconhecer o *churn* de clientes depende da indústria em estudo e varia de caso para caso. Normalmente, na análise de *churn*, considera-se que ocorreu *churn* de um cliente quando este termina totalmente a relação com a empresa, isto porque, existe um contrato a formalizar a relação. No entanto, em ambientes não contractuais, não é claro o momento em que os clientes quebram a relação. É necessário definir claramente o conceito de *churn* no contexto da empresa, ou seja, definir em concreto o momento em que se considera que ocorreu *churn* de um cliente (secção 3.2.3).

2.4.2 Modelação Preditiva de *Churn*

As campanhas de prevenção de *churn* proactivas fornecem um aumento expectável da margem de manobra para evitar o *churn* de clientes e, como o objectivo é antecipar comportamentos de *churn* por parte dos clientes, estas campanhas deverão ser suportadas por modelos de previsão. Estes são modelos que baseados em factos históricos propõem prever eventos similares no futuro. Contextualizando, no caso de estudo sobre o fenómeno de *churn*, a previsão de futuros comportamentos de *churn* numa empresa será baseada no comportamento demonstrado pelos clientes do passado, ou seja, serão usados os dados de diversos clientes que foram sendo armazenados e que descrevem o seu comportamento ao longo de um determinado período de tempo passado na tentativa de encontrar semelhanças com os dados e os comportamentos dos clientes que se pretende prever o futuro.

Diferentes Modelos de *Churn*

Em termos gerais, a previsão de *churn* pode ser feita através de dois tipos de modelação. O primeiro tipo dá como resposta um resultado do tipo binário, ou seja, prevê quais os clientes que irão terminar a sua ligação com a companhia e os que irão continuar. Vulgarmente, estas previsões são feitas para um horizonte relativamente curto. Por outro lado, pode ser usado um segundo tipo de modelação que tenta estimar o tempo de vida restante do cliente na empresa, ou seja, quanto mais tempo o cliente vai permanecer na empresa - são modelos suportados por técnicas de *data*

mining (Berry e Linoff, 2004). Resumidamente os dois tipos de modelação normalmente aplicados na previsão de *churn* são:

- Modelos de *churn* binários. Este é o tipo mais utilizado na previsão de *churn* (Jinbo et al., 2007). O objectivo destes modelos é a previsão dos clientes que estão em risco de *churn* num horizonte temporal próximo. Posto isto, é necessário definir esse período de tempo, que não poderá ser muito curto para possibilitar a intervenção sobre estes clientes (Berry e Linoff, 2004). Os dados históricos do cliente são combinados com uma *flag* que marca o cliente como ativo ou não. Depois, serão então aplicados algoritmos de classificação como a regressão logística, árvores de decisão ou redes neuronais. O resultado será um modelo que aplicado aos dados gera uma probabilidade de *churn* dos clientes em causa (Berry e Linoff, 2004). Deste modo, os clientes que tiverem uma probabilidade de *churn* superior ao limite definido serão inseridos no programa de retenção.
- Modelos do tempo de vida restante. Estes modelos poderão não ser tão utilizados como os modelos binários, mas também proporcionam resultados interessantes. Propõem prever o tempo restante dos clientes na empresa, proporcionando mais informação do que simplesmente "quem vai abandonar a empresa nos próximos 60 dias", na medida em que, a estimativa do "tempo de vida" (tempo que o cliente permanecerá na empresa) restante de um cliente pode ser usada em modelos *LifeTime Value* (LTV) para calcular o valor do cliente ou como base para a definição de um *score* de lealdade de um cliente. Neste último caso, a lealdade de um cliente é baseada no tempo futuro que o cliente vai permanecer na empresa e não no tempo que o cliente permaneceu até ao presente.

A Importância da Precisão e Compressibilidade

Dois aspectos fundamentais na implementação de um modelo de *churning* são a compreensibilidade e precisão obtidas (Verbeke et al., 2011). Um modelo preciso permite a rentabilização das campanhas de retenção focando-as nos clientes certos. Por seu lado, a compreensibilidade é a propriedade de descobrir informação sobre as causas dos clientes partirem que pode ser usada para melhorar as campanhas de retenção. Resumidamente, as razões da importância destas propriedades são:

- Precisão (Poder Preditivo). Uma boa precisão significa um erro estimado baixo (Breiman, 2001b). Na modelação de *churn*, uma característica fundamental, é que o número de clientes no estado *churn* é extremamente baixo, o que leva a que a precisão seja uma propriedade fundamental para o seu sucesso (Zhao et al., 2005). Para além disso, com previsões de *churn* imprecisas, as empresas poderão estar a desperdiçar algum dinheiro em clientes que não estavam em risco de *churn* (Burez e Van den Poel, 2007). É essencial, pois, que a previsão de casos de *churn* seja o mais precisa possível.
- Compreensibilidade dos modelos. Kim e Yoon (2004) defendem que é essencial entender as causas que levam ao *churn* de clientes, de forma a minimizar o fenómeno e aumentar a lealdade dos clientes. Quando se quer agir em campanhas de retenção, terá que se interagir com os clientes, sendo necessário compreender os problemas existentes. Portanto, torna-se necessário perceber e estudar as razões de *churn* (Ahn et al., 2006; Verbeke et al., 2011; Wang et al., 2009). Essa informação permite melhorar as campanhas de retenção (Verbeke et al., 2011) e permite ainda reavaliar e melhorar os modelos de *churn* (Wang et al., 2009). Contudo, a compreensão das causas do *churn* e a compreensibilidade dos modelos estão a ser, de certa forma, subvalorizados na literatura sobre o tema (Ahn et al., 2006; Verbeke et al., 2011).

Infelizmente, na previsão, a precisão e a compreensibilidade dos modelos estão em conflito (Breiman, 2001b).

Avaliar a Rentabilidade do Projeto

Para implementar uma campanha de prevenção de *churn* é necessário estabelecer métodos para medir a sua eficácia. Para além de avaliar o modelo de previsão usado, é necessário estimar, no início, e calcular, no final, o retorno real que advém de implementar a campanha de prevenção de forma a justificar o investimento. Neslin et al. (2006) apresentaram uma forma simples de calcular o retorno a partir das seguintes variáveis:

- N , número total de clientes.
- α , fracção dos clientes abrangidos pelo programa de gestão de *churn*.
- β , fracção dos clientes abrangidos que eram de facto *churners*.
- δ , custo do incentivo usado para manter o cliente.
- γ , fracção dos clientes abrangidos pelo programa, que eram de facto *churners*, que

decidiram permanecer na empresa devido ao incentivo (i.e. taxa de sucesso do incentivo).

- c , custo de contactar o cliente para lhe oferecer o incentivo.
- CLV , *Customer Lifetime Value* (i.e. o valor para a firma se o cliente for mantido).
- A , custos administrativos fixos do programa de gestão de *churn*.

A partir destas variáveis, o lucro de uma campanha de gestão de *churn* pode ser calculado da seguinte forma:

$$\Pi = N\alpha[\beta\gamma(CLV - c - \delta) + \beta(1 - \gamma)(-c) + (1 - \beta)(-c - \delta)] - A$$

Em que a parcela $\beta\gamma(CLV - c - \delta)$ diz respeito à fracção dos clientes que decidiu permanecer na empresa devido ao programa, a parcela $\beta(1 - \gamma)(-c)$ diz respeito à fracção de clientes no qual o programa não obteve efeito e $(1 - \beta)(-c - \delta)$ reflete os custos dos clientes que aceitaram a oferta mas que não iam abandonar a empresa

O termo β reflete a eficiência do modelo e pode ser expresso como:

$$\beta = \lambda\beta_0$$

em que β_0 é a taxa de *churners* do total de clientes da empresa e λ é o "*lift*" (diz-nos o quanto mais provável que aconteça *churn* a um cliente do grupo abrangido pelo programa relativamente ao total de clientes da empresa)

Capítulo 3

Previsão de Situações de *Churn*

3.1 Modelos de Mineração de Dados

Historicamente, a noção de encontrar padrões úteis nos dados tem sido batizada com uma grande variedade de nomes como *data mining*, extração de conhecimento ou descoberta de informação. O termo *data mining* foi popularizado pelos estatísticos, analistas de dados e comunidades dos sistemas de informação de gestão (Fayyad et al., 1996). Contudo, em 1989 foi criada a frase *Knowledge Discovery in Databases* (KDD) para designar o processo geral de descobrir conhecimento útil em bases de dados, no qual, *data mining* diz respeito apenas a uma etapa de todo o processo (Fayyad et al., 1996). Como, na indústria, se tem adoptado uma definição mais abrangente do termo *data mining* (Han e Kamber, 2006) e estando este estudo ligado à indústria, aceita-se, assim, a interdisciplinaridade normalmente atribuída ao conceito neste contexto.

O conceito de Descobrimto de Conhecimento em Base de Dados (DCBD) do inglês *KDD* foi definido em Fayyad et al. (1996) como "*o processo não trivial de identificação de padrões válidos e potencialmente úteis, perceptíveis através dos dados*". É um processo iterativo que consiste numa sequência iterativa de etapas. É interativo porque requer o envolvimento do utilizador na tomada

de decisões e iterativo já que pode haver retrocesso a etapas anteriores. A Figura 14 ilustra de forma global o processo KDD, que inclui as seguintes etapas (Fayyad et al., 1996):

- Desenvolver uma compreensão do domínio da aplicação e do conhecimento prévio relevante, assim como identificar o objectivo a atingir pelo processo.
- Criação do conjunto de dados (*dataset*), que é fazer a seleção dos dados mais apropriados de onde se extrairá o conhecimento; esta pode incluir processos de seleção de variáveis, bem como, de amostragem dos dados.
- Limpeza e pré-processamento dos dados, que inclui algumas operações básicas como: remover ruído dos dados ou definir estratégias para tratamento de valores em falta.
- Redução e projecção dos dados, que se desenvolve através da escolha dos atributos mais úteis para a execução da tarefa.
- Definição da tarefa de *data mining*, por exemplo classificação (3.1.1).
- Seleção do algoritmo de *data mining* a ser utilizado e do(s) método(s) para fazer a procura de padrões.
- Mineração de dados, ou seja, fazer a procura de padrões que sejam interessantes.
- Interpretação dos padrões encontrados, a visualização do conhecimento extraído, remoção de padrões irrelevantes ou redundantes e possivelmente o regresso a um das etapas anteriores.
- Ação sobre o conhecimento obtido, não é mais do que usar o conhecimento diretamente no processo de tomada de decisão.

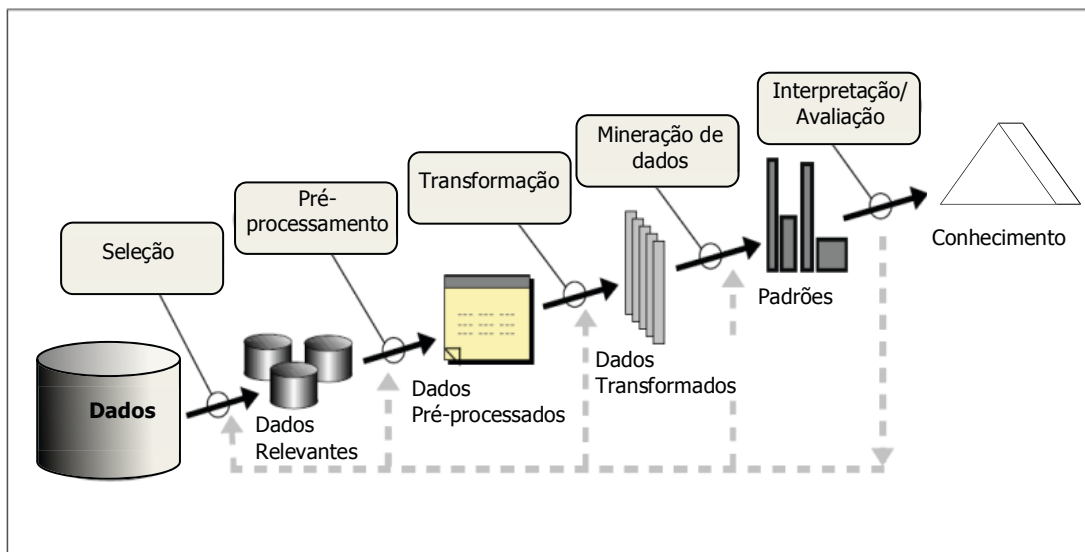


Figura 14: Etapas do processo KDD - adaptado de (Fayyad et al., 1996)

Data Mining

Devido ao constante aumento de dados, operáveis de forma eficiente via computadores de alta capacidade computacional e canais de comunicação de grande velocidade, tem havido um aumento constante da dependência de métodos estatísticos para extrair informação útil dos conjuntos de dados abundantes. Os métodos estatísticos fornecem uma maneira para estruturar e organizar dados que se encontram num estado bruto. Na Figura 15 podemos observar as diferentes fases da procura crescente de informação refinada. É nesta perspectiva que os métodos de *data mining* aparentam ser proeminentes (Lejeune, 2001). Estes complementam a análise de dados feita através de métodos como a estatística, as folhas de cálculo ou o processamento analítico de dados.

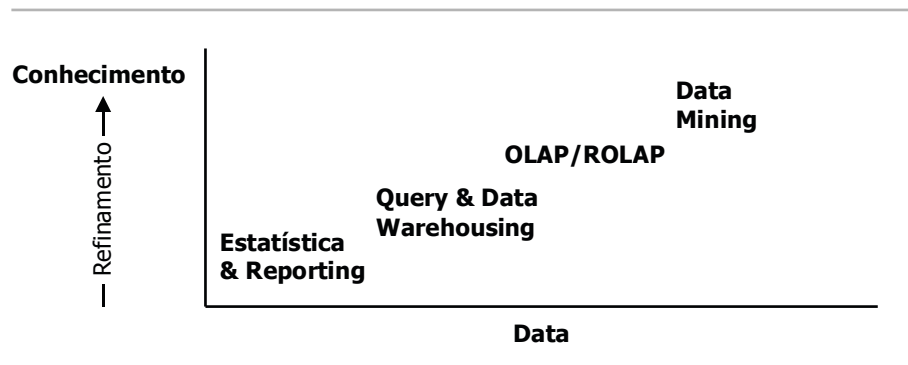


Figura 15: Evolução na procura de informação - adaptado de (Lejeune, 2001)

Data mining é uma área multidisciplinar que evoluiu a partir da "intercepção" de vários campos de investigação como a aprendizagem automática (*machine learning*), a inteligência artificial, o reconhecimento de padrões, as bases de dados, a estatística e a visualização de dados. O objectivo unificador é extrair conhecimento de alto nível a partir de dados em bruto no contexto de grandes conjuntos de dados (Fayyad et al., 1996). Berry e Linoff (2004) definem *data mining* como "a exploração e análise de grandes quantidades de dados com o objectivo de descobrir padrões e regras significativas".

As técnicas de *data mining* constroem, a partir dos dados, uma representação da realidade na forma de um modelo. O modelo descreve os padrões e relacionamentos presentes nos dados (Rygielski et al., 2002). Genericamente, as técnicas de *data mining* podem ser aplicadas tendo em vista dois tipos de objectivos (Fayyad et al., 1996; Lejeune, 2001):

- Descrição, com vista a aumentar a compreensão dos dados e seu conteúdo; a segmentação (*clustering*) é o exemplo de uma tarefa deste tipo.
- Previsão, para descobrir padrões escondidos nos dados e usa-los para prever informação ou eventos futuros; a classificação e a regressão são exemplos de tarefas deste tipo.

Algumas das tarefas de *data mining* que podem ser realizadas para alcançar estes objectivos são a classificação, a regressão, a segmentação, a associação e a sumarização.

3.1.1 Problema de Classificação

É um dos problemas mais comuns de *data mining* e está particularmente orientado para a aplicações de CRM e detecção de casos de *churn* (Berry e Linoff, 2004; KhakAbi et al., 2010; Ngai et al., 2009). Num problema de classificação é assumido que existe um conjunto de objetos (tuplos, instâncias ou exemplos), - e.g. clientes - caracterizados por vários atributos (variáveis independentes ou *predictors*), que pertencem a diferentes classes (variável dependente, *target* ou *label*). O atributo classe é um valor discreto e é conhecido para cada objecto na fase de treino. O objectivo é construir um modelo de classificação, também conhecido como classificador, que atribui o valor correto da classe, a novos objetos que lhe são submetidos. As técnicas de classificação são sobretudo usadas em problemas de previsão. Como, na fase de treino, o atributo classe, de cada tuplo, participa no processo, a classificação é um tipo de aprendizagem supervisionada (Han e Kamber, 2006).

Um processo de classificação envolve dois momentos:

1. o treino, no qual é construído um classificador que descreve os padrões associados às classes; é a fase de aprendizagem (ou fase de treino), na qual um algoritmo de classificação constrói o classificador "aprendendo" do conjunto de dados de treino, ou seja, por tuplos da base de dados que contêm associado às variáveis independentes um atributo classe;
2. a previsão, na qual são usados os novos dados que se querem classificar; estes são descritos pelo mesmo conjunto de atributos dos dados de treino com a exceção do atributo classe, são processados pelo classificador, saindo com um atributo classe atribuído pelo classificador.

O problema de gestão proactiva de *churn* é tipicamente conceptualizado como um problema de classificação, mais concretamente, como um problema de classificação binário em que existem duas classes "*churner*" e "*não churner*" (e.g. (Coussement e Van den Poel, 2008a; Ju e Guo, 2008; Ghorbani e Taghiyareh, 2009)). O modelo é gerado através da informação existente sobre os clientes e o respectivo comportamento de *churn* registado.

3.1.2 Metodologia CRISP-DM

Uma realidade nos projetos de *data mining* é que, mesmo que o problema e o objectivo estejam claramente definidos, o processo é complicado, pois envolve a realização de muitas tarefas ao longo de várias fases (Ng e Liu, 2000). Torna-se, portanto, mais fácil de compreender, implementar e desenvolver se for enquadrado no contexto de uma metodologia. Duas das metodologias mais conhecidas são a CRISP-DM (*Cross-Industry Standard Process for Data Mining*) e a SEMMA (*Sample, Explore, Modify, Assessment*) (Santos e Azevedo, 2005).

A metodologia CRISP-DM foi concebida em finais de 1996, pelo consórcio formado pelas empresas DaimlerChrysler, SPSS e NCR. O seu desenvolvimento foi motivado pelo interesse crescente e generalizado no mercado de *data mining* e pela consequente necessidade, da indústria, de adoptar um processo padronizado (Chapman et al., 2000). É uma metodologia bem-sucedida porque se baseia na experiência prática, do mundo real, de como as pessoas desenvolvem os projetos de *data mining*. Está bem documentada, encoraja o uso das melhores práticas e oferece às organizações uma base para realizar os seus projetos da melhor maneira de forma obter os melhores resultados (Shearer, 2000).

A metodologia CRISP-DM é descrita em termos de um processo hierárquico, com um ciclo de vida que se desenvolve em 6 fases (Figura 16), que ajudam as organizações a perceber o processo de descoberta de conhecimento e fornecem um "guião" a seguir no planeamento e execução de um projeto de *data mining* (Chapman et al., 2000; Shearer, 2000):

- Estudo do negócio, que visa converter os objectivos e requisitos de negócio num problema de mineração de dados.
- Estudo dos dados, que tem como objectivo ganhar familiaridade com os dados.
- Preparação dos dados, que cobre todas as actividades envolvidas na construção do conjunto de dados final que vai ser usado pelos modelos de mineração de dados.
- Modelação, que integra as várias técnicas de modelação são aplicadas e são calibrados os seus parâmetros.
- Avaliação, em que se avalia o modelo e revê-se todo o processo para a certificação de que são atingidos os objectivos de negócio.
- Implementação, em que se faz a aplicação do modelo no processo de tomada de decisão.

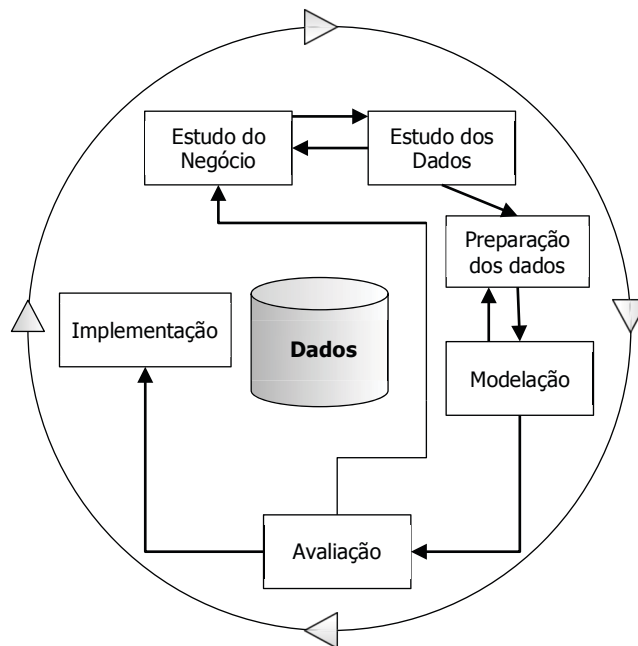


Figura 16: Fases da Metodologia CRISP-DM - adaptado de (Chapman et al., 2000)

3.2 Desenho da Assinatura de um Cliente

Nas empresas, normalmente não existe um conjunto de dados preparado para ser usado diretamente no processo de descoberta de conhecimento. Usualmente, as bases de dados possuem inúmeras tabelas, centenas de atributos e são guardados milhões de registos transacionais todos os dias, o que leva a que seja impraticável a aplicação de tarefas de *data mining* sobre esses dados não tratados (Ng e Liu, 2000). É necessário fazer o seu tratamento, de modo que, a informação mais útil, para caracterizar cada item e para a tarefa de modelação, esteja agrupada numa única linha. Consequentemente, o conjunto destas linhas dá origem a uma tabela, na qual cada coluna representa um atributo de descrição das linhas. Quando estas linhas descrevem clientes são habitualmente conhecidas como assinaturas de clientes. Estas têm que ter a capacidade de capturar os atributos correntes do cliente, assim como as mudanças de comportamento ao longo do tempo. É pois, um sumário ao nível do cliente, de tudo o que pode ser extraído dos dados para caracteriza-lo (Berry e Linoff, 2004). Os modelos de *data mining*, embora sejam poderosos, apenas são efetivos com a introdução de conhecimento do negócio (Tsiptsis e Chorianopoulos, 2010).

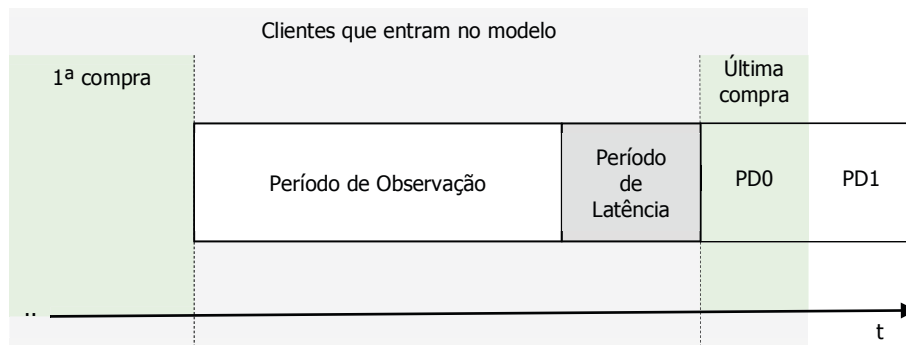
A etapa inicial envolve entender o problema de negócio e o ambiente em que o sistema de *data mining* se insere. O novo sistema deve estender os processos existentes em vez de tentar modificar os processos existentes (Masand et al., 1999). Deve haver, também, uma colaboração entre os especialistas de *data mining* e os de negócio. Uma parte significativa do esforço do processo de modelação envolve essa colaboração, no sentido de compreender o significado e contexto dos dados disponíveis, bem como dissecar as regras do negócio e integrar conhecimento sobre o negócio (Mozer et al., 2000). Um dos principais meios de introdução deste conhecimento, no processo de modelação é na especificação das assinaturas dos clientes e representação dos dados (Yan et al., 2004). Naturalmente, os dados que entram na assinatura têm que ser escolhidos e tratados. Têm que ser feitos compromissos entre os diferentes requisitos, como por exemplo, a quantidade de dados e a escalabilidade dos algoritmos. Com o objectivo de aumentar a probabilidade de sucesso do projeto, o processo de construção da assinatura deverá seguir a metodologia CRISP-DM, nomeadamente as fases de compreensão e preparação dos dados.

De seguida serão apresentadas algumas das questões essenciais da construção do conjunto de dados que irá ser usado para a previsão do *churn* de clientes. Também será fornecida uma perspectiva conjunta da construção do conjunto de dados para o modelo de *churn* e das várias regras relativas ao negócio que têm que ser levadas em conta.

3.2.1 Definição da Janela Temporal do Modelo

Normalmente, os clientes não abandonam uma empresa subitamente. Antes de isso acontecer, eles vão transferindo parte dos seus gastos para outras lojas (Buckinx e Van den Poel, 2005). Esta perturbação no comportamento do cliente, juntamente com outras informações que descrevem o seu perfil, pode ser usada para prever o evento futuro de *churn*. Uma das primeiras etapas na construção das assinaturas dos clientes é a identificação da janela temporal do modelo e dos vários períodos que a constituem (Berry e Linoff, 2004). Deste modo, tal como ilustrado na Figura 17, um *dataset* para treino de um modelo de previsão de *churn*, cobre habitualmente os seguintes períodos (Wei e Chiu, 2002; Burez e Poel, 2008; Ju e Guo, 2008; Tsiptsis e Chorianopoulos, 2010):

- Período de Observação (PO). É deste período que é extraída a "vista" que caracteriza o cliente e o seu comportamento anterior à ocorrência do evento ("*churn*" ou "não *churn*"). O tamanho deste período afecta vários factores. Se, por um lado, o aumento da quantidade de dados influencia positivamente a capacidade preditiva do modelo, por outro, tem um impacto negativo na eficiência do processo de modelação. Deste modo, baseando nos requisitos previamente definidos para esses factores e tal como em (Datta et al., 2000) deverão ser testados diferentes tamanhos para o período de observação para chegar ao melhor compromisso.
- Período de Latência (PL). Um modelo de previsão de *churn* de clientes insere-se num projeto de gestão de *churn*, no qual, após se identificarem os clientes em risco, terá que se agir sobre eles para tentar evitar a situação de *churn*. A empresa necessita de tempo para abordar o cliente com as ações de retenção e é neste contexto que nasce a necessidade de especificar um período de latência (Wei e Chiu, 2002; Neslin et al., 2006; Ju e Guo, 2008; Ghorbani e Taghiyareh, 2009), para que o período de previsão não seja imediatamente após o período de observação. Para além disso, é necessário tempo para colecionar e preparar os dados e para o treino e aplicação do modelo de mineração, operações de computação, frequentemente, complexas que requerem bastante tempo. A duração deste período é habitualmente de um mês, mas, essencialmente, um tamanho apropriado é aquele que permite executar eficientemente os planos de retenção e as tarefas relativas ao modelo de mineração de dados.
- Período de Definição da Variável Dependente (PD). Este período segue o período de latência. É durante este período que é definido o estado do cliente, ou seja, a variável dependente do modelo supervisionado ("*churn*", "não *churn*"). Na sua definição é usado um critério previamente estabelecido para definir a variável dependente (3.2.3). É neste período que ocorre a última compra do cliente (período PD0 Figura 17). Como no retalho (o nosso caso de estudo) não há uma comunicação da intenção de quebrar a relação por parte do cliente, sendo assim necessário acrescentar ao período de definição um período de tempo extra (período PD1 Figura 17) para verificar a ocorrência (ou não) do evento de *churn*. Por exemplo, se for considerado *churner* um cliente que não efetua qualquer compra durante 2 meses, o período PD1 será de 2 meses.

Figura 17: Períodos existentes nos dados para previsão de *churn*

A estratégia de treino do modelo passa por associar os padrões existentes nos dados do período de observação ao evento extraído no período de definição da variável dependente, para que na fase de aplicação do modelo sejam feitas as previsões de modo análogo, mas numa janela temporal diferente.

Tempo de Vida do Modelo

Existem dois conjuntos de dados num processo de modelação preditiva de *churn*: treino e previsão. Como se pode ver através da Figura 18, os dados dos conjuntos pertencem a janelas temporais diferentes.

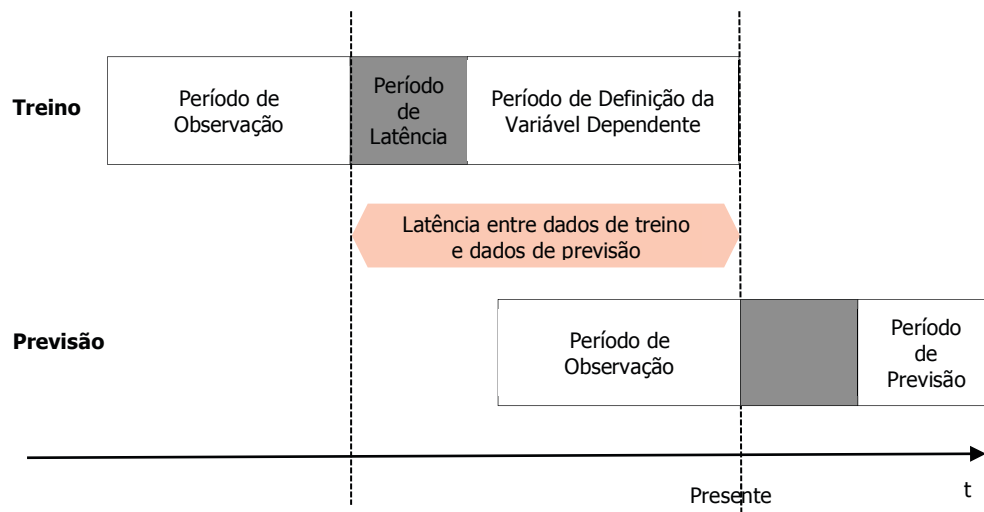


Figura 18: Conjuntos de dados envolvidos na previsão de *churn*

Um dos muitos problemas que podem surgir em aplicações do mundo real está relacionado com o dinamismo do ambiente a que pertence a empresa/sector e dos comportamentos humanos (Ng e Liu, 2000; Yan et al., 2001). Num ambiente dinâmico, que esteja em constante mudança ou que demonstre características de sazonalidade, podem existir flutuações nos dados que façam com que o modelo treinado com dados do passado e os padrões detectados se tornem inválidos. Os resultados obtidos na avaliação através de validação cruzada podem ser traiçoeiros, pois o modelo irá ser usado com dados relativos a uma janela temporal futura, que podem ter sido afectados por mudanças do ambiente, levando conseqüentemente a uma menor eficácia do modelo (Yan et al., 2001).

Através da Figura 18 podemos ver que existe um período de atraso, entre o período de origem dos dados de treino e o período de onde são extraídos os dados para previsão. O tamanho deste período, para além de depender das decisões tomadas para os períodos de latência e de previsão do evento, depende também da estratégia adoptada para a conceptualização da variável dependente. Naturalmente que, o tamanho do período de atraso entre os dados de treino e os dados de previsão deve ser o menor possível. Conseqüentemente, deverão ser testadas várias configurações da janela temporal com o objectivo de obter o melhor compromisso entre:

- o período de observação (PO);

- o período de latência (PL);
- o período de definição da variável dependente (PD0 + PD1);
- a estratégia de conceptualização da variável dependente e, conseqüente, o período de tempo necessário (PD1).

Uma outra questão importante é a estratégia de atualização do modelo, ou seja, com que frequência o modelo deve ser reconstruído. Um método para auxiliar a definição desta estratégia é calcular a deterioração das previsões ao longo do tempo. Isso faz-se avaliando um modelo construído num determinado período com conjuntos de dados extraídos de períodos sucessivamente mais avançados no tempo. A prova de que um modelo é estável reside na sua habilidade de obter bons resultados ao longo de vários meses consecutivos.

Estratégias para Combater a Instabilidade do Modelo ao Longo do Tempo

A primeira estratégia que salta à vista é o uso de um conjunto maior de dados de treino. Todavia, a quantidade de dados, geralmente, não é o problema. Outra maneira intuitiva de aumentar a estabilidade do modelo é usar dados provenientes de um período histórico maior, ou seja, de um período de observação maior. Esta estratégia pode não ser a mais adequada, se muitos clientes tenham um tempo de vida insuficiente para incluir no modelo (Yan et al., 2001).

Uma estratégia para criar modelos estáveis, ou seja, modelos que reduzem o impacto das alterações do ambiente nas previsões e que mantêm a sua eficácia ao longo de um período de tempo mais extenso, é usar vários conjuntos de dados provenientes de janelas temporais separadas (Figura 19). O modelo é treinado com o conjunto de dados resultante da união dos vários conjuntos de treino (Yan et al., 2001). A ideia é generalizar a partir do passado e não memorizar o que aconteceu num momento particular do passado (Berry e Linoff, 2004). O aspecto negativo desta estratégia é o aumento do conjunto de treino e conseqüente aumento do tempo de aprendizagem do modelo (Yan et al., 2001).

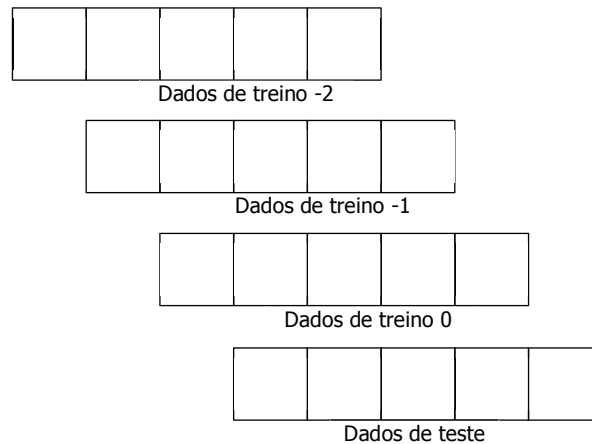


Figura 19: Extração dos dados de treino - adaptado de (Yan et al., 2001)

Alternativamente, de maneira a evitar o aumento do tempo de treino do modelo, pode-se treinar um modelo com cada conjunto de treino e, no final, usar um método de combinação de modelos para combinar as previsões individuais. Desta forma, o modelo é atualizado periodicamente com a inclusão do modelo treinado com o conjunto de dados da janela temporal mais recente. Esta estratégia requer apenas que sejam guardados os modelos individuais gerados no passado e reduz drasticamente o tempo de treino relativamente à estratégia anterior. As previsões finais são obtidas através do voto ponderado dos vários modelos (Yan et al., 2001).

Critérios de Seleção de Clientes

Existem requisitos que impedem o uso da base de dados total de clientes. Esses requisitos, podem ser impostos por objectivos de negócio ou pelo próprio processo de modelação. Por exemplo, a qualidade dos dados pode impedir a inclusão de certos clientes no modelo. Um outro exemplo de restrição do uso de determinados clientes, que surge no processo de modelação, é devido à janela temporal e a um princípio que lhe está subjacente. O princípio que o período de observação deve ser equivalente para todos os clientes, ou seja, os clientes que não estejam ativos durante o período total de observação não devem ser considerados (Wei e Chiu, 2002; Zan et al., 2007). Logo, só deverão ser incluídos os clientes que:

- estão ativos desde um ponto anterior ao início do período de observação;
- são considerados como ativos até ao período de conceptualização da variável dependente,

ou seja, fizeram a última compra nesse mesmo período.

Por outro lado, as estratégias de retenção das empresas têm um limite de recursos a utilizar. O objectivo é obter o máximo de retorno possível recorrendo aos recursos disponíveis. Um modo de otimizar a utilização dos recursos é focar os esforços de retenção nos melhores clientes, consoante um determinado critério. Deste modo, apenas são incluídos no processo de modelação os clientes que respeitam determinadas regras, como ocorre no trabalho de Buckinx e Van den Poel (2005), em que apenas são incluídos os clientes considerados "comportamentalmente leais".

3.2.2 Compreensão dos Dados

Naturalmente que, o tipo de dados guardados pelas empresas varia conforme a indústria a que pertencem. Mesmo em empresas do mesmo ramo os dados não são armazenados de forma idêntica. É neste contexto, que vários autores (Masand et al., 1999; Datta et al., 2000; Ng e Liu, 2000; Hadden et al., 2007; Hu, 2005; Hung et al., 2006) destacam a enorme importância das fases de compreensão e preparação dos dados no desenvolvimento de um modelo de previsão de *churn*.

Conjunto de Dados Disponível

O objectivo nesta fase do trabalho foi fazer um levantamento dos dados necessários (para a resolução do problema) e investigar a disponibilidade de tais dados (Tsiptsis e Chorianopoulos, 2010). É, ainda, fundamental fazer uma boa compreensão da semântica dos dados, ou seja, o seu significado na base de dados (Ghorbani e Taghiyareh, 2009). Resumidamente as principais tarefas desta fase da modelação relativamente aos dados são (Chapman et al., 2000):

- Coleção inicial: Adquirir os dados disponíveis para o projecto.
- Descrição: Examinar e descrever as propriedades dos dados adquiridos.
- Exploração: Análise dos dados de acordo com os objectivos de mineração de dados.
- Verificação da qualidade: Levantamento de eventuais problemas no conjunto de dados e possíveis resoluções para os mesmos.

Tabela 4: Informação registada no POS - adaptado de (Tsipstis e Chorianopoulos, 2010)

Nome do Campo	Descrição
Transaction_Line_ID	Identificador associado a cada linha da transação
Transaction_ID	Identificador que identifica unicamente cada transação. Cada transação é um evento único de compra compreendendo vários <i>records</i> de dados (Transaction_Line_ID) e muitos itens comprados
Transaction_Type	Tipo da transação (Loja, internet, telefone)
Customer_ID	Código do cartão de fidelidade apresentado pelo cliente
Data_Tempo	A data e hora da transação
Store_Code	Código da loja onde foi feita a transação
Check_Out_Code	Código <i>checkout</i> da transação
Item_Code	Código UPC do item comprado
Num_of_Items	Número de itens comprados
Rated_Amount	Valor do item
Discount_Amount	Valor descontado ao valor do item

Como em qualquer outra indústria, no retalho a relação da empresa com o cliente é monitorizada de forma contínua. Normalmente, isto é feito através dos postos de venda (POS). As principais fontes de dados para a tarefa de modelação são:

- Registos transacionais: Dados sobre as transações registados nos pontos de venda. A Tabela 4 representa a estrutura típica da informação registada nos pontos de vendas.
- Informação sociodemográfica do cliente: São tabelas que contêm a informação mais recente sobre os clientes atuais e antigos. São dados obtidos através de campanhas de fidelidade desenvolvidas pelas empresas e têm sido amplamente usados em estudos de atrito de clientes (Buckinx e Van den Poel, 2005). A Tabela 5 representa uma estrutura típica dessa tabela.

Tabela 5: Tabela com a informação do cliente - adaptado de (Tsipstis e Chorionopoulos, 2010)

Nome do atributo	Descrição
Customer_ID	Número do cartão de fidelidade apresentado pelo cliente
VAT_Num	Número de contribuinte
Gender_Code	Código do género
Marital_Status_Code	Código do estado civil
Children_Num	Número de filhos
Child_Birth_Date	A data de nascimento do 1º filho
Educational_Status_Code	Código das habilitações
Occupation_Code	Código da ocupação
Annual_Income_Code	Código da faixa dos rendimentos anuais
Birth_Date	Data de nascimento
Death_Date	Data de óbito ou caso contrário data de controlo
Registration_Date	Data do primeiro registo
Closure_Date	Data do encerramento do relacionamento com o cliente
First_Name	Nome próprio
Last_Name	Apelido
Middle_Name	Nome do meio
Home_Street	Nome da rua
Home_Num	Número da rua
Home_City	Nome da localidade
Home_State	Nome da região
Home_ZIP	Código Postal
Home_Country	Nome do País
Fixed_Phone	Número de telefone fixo
Mobile_Phone	Número de telemóvel
Home_Email	Email próprio
Work_Email	Email do emprego
Last_Update	Data da última atualização do registo

Todos estes dados são complementados com outros dados existentes relativos a produtos, promoções e lojas. A Figura 20 ilustra um exemplo típico, simplificado, de um diagrama de um *data mart* de vendas de um sistema de suporte à decisão. São dados que se encontram ao nível da transação, mas para poderem ser usados num modelo de previsão terão que ser adaptados ao nível do cliente, ou seja, permitir descrever o perfil do cliente e o seu comportamento durante o período de observação selecionado. Esta adaptação é feita essencialmente através de processos de agregação.

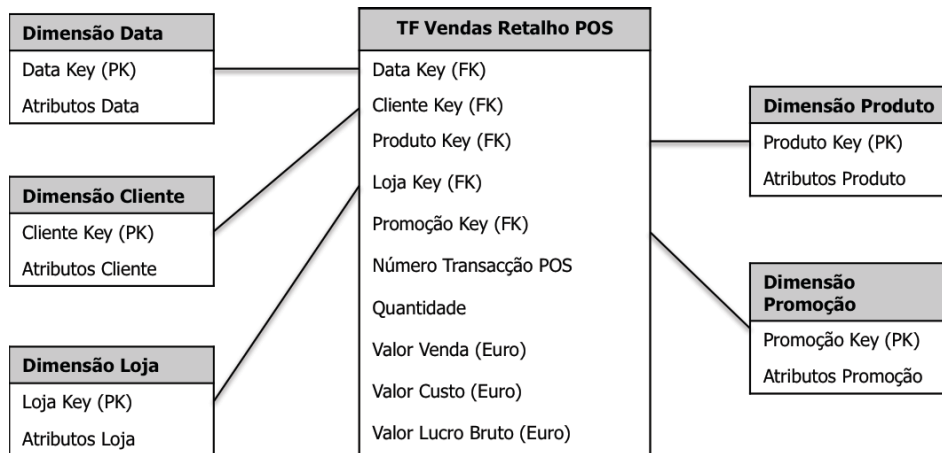


Figura 20: *Data Mart* de vendas típico - baseado em (Kimball e Ross, 2002)

Dados para a Assinatura

Os dados a serem incluídos nas assinaturas devem descrever o comportamento passado do cliente, bem como, refletir mudanças ao longo do tempo do período de observação. Alguns dos indicadores que podem ser extraídos dos dados disponíveis para complementar a informação sociodemográfica dos clientes são (Berry e Linoff, 2004; Buckinx e Van den Poel, 2005; Tsiptsis e Chorianopoulos, 2010):

- Atributos RFM, que podem estar descritos globalmente ou por categoria de produto. Mais especificamente são: recente (*Recency*), o intervalo de tempo desde a última transação; frequência (*Frequency*), o número de visitas com compras; e monetário (*Monetary*), o valor total gasto em compras pelo cliente.
- Duração da relação: é a dimensão do período entre a primeira e a última compra efetuadas.
- Loja(s)/departamento(s) preferidas: é a loja em que o cliente fez o maior número de visitas com compras.
- Horas e dias das visitas: indica a frequência das visitas relativamente ao dia da semana e à hora em que ocorrem.
- Modo de pagamento (dinheiro, cartão de crédito) e canal (loja, telefone, internet).
- Contactos de campanhas e respectivas respostas: integra informação sobre a reação do cliente a ofertas promocionais passadas.

- Promoções: dados sobre o uso de cupões de desconto, pontos do cartão, etc.
- Informação relativa a reclamações: é constituída por dados relativos a eventuais reclamações efectuadas pelo cliente.

Factores Macroeconómicos

Como o período de observação para a descrição dos clientes é igual para todos, a inserção de indicadores macroeconómicos como, por exemplo, o índice da inflação, não terá qualquer efeito na tarefa de previsão, pois teria o mesmo valor para todas as assinaturas, ou seja, não teria qualquer valor informativo. Uma possível forma de incluir essa informação é através da sua extração do comportamento do cliente, como por exemplo, calcular a variação dos preços do top de produtos mais comprados pelo cliente. Uma outra solução para incluir indicadores macroeconómicos é a extração das assinaturas de diversas janelas temporais.

Preparação dos Dados

Após a compreensão dos dados disponíveis e avaliação da informação necessária para a previsão de *churn*, é necessário adapta-los para que possam ser usados no modelo de *data mining*. Geralmente, são necessárias diversas *queries* que envolvem um elevado nível de complexidade e que necessitam de bastante processamento sobre as fontes de dados, pois são necessárias diversas operações de seleção, agregação e junção de dados de diversas tabelas. Após a consolidação dos dados é, ainda, necessário efetuar operações de limpeza e de transformação dos dados para ir ao encontro dos requisitos dos algoritmos a serem aplicados (Tsiptsis e Chorianopoulos, 2010).

Seguindo as tarefas descritas, para a fase da preparação de dados, pela metodologia CRISP-DM, algumas das coisas que mais se destacam na preparação do conjunto das assinaturas para usar na previsão de *churn* são:

- A seleção dos dados de acordo com a janela temporal e critérios de seleção de clientes.
- A transformação dos dados para o nível de abstração correto (Ng e Liu, 2000; Berry e Linoff, 2004): sumarização de variáveis para o nível cliente em vez do nível de transação (agregação).
- A minimização do efeito da desproporção entre as classes.

- A seleção dos atributos: Escolha dos atributos relevantes para o processo de modelação.
- O tratamento de nulos e dados inconsistentes.
- A definição e introdução da variável dependente.

3.2.3 Conceptualização da Variável Dependente

Uma definição apropriada da noção de *churn* é crucial no contexto da modelação preditiva deste fenómeno. É uma das tarefas mais importantes do processo de modelação e pode ser bastante complicada (Yan et al., 2004; Owczarczuk, 2010). Assim, é necessário fazer uma análise cuidadosa dos dados, das regras de negócio e do cenário geral da aplicação, assim como desenvolver uma boa cooperação entre os utilizadores de negócio e os utilizadores da parte da modelação, para chegar a uma boa definição do conceito de *churn* (Yan et al., 2004; Tsiptsis e Chorianopoulos, 2010). Uma boa definição de *churn* requer um bom compromisso entre os seguintes objectivos (Owczarczuk, 2010):

- Permitir uma verificação rápida. O período de verificação está directamente relacionado com o tempo de latência entre a janela temporal de aprendizagem e a janela temporal para previsão (Figura 18). O objectivo é esperar o menor intervalo de tempo possível, para verificar que ocorreu, de facto, o *churn* do cliente.
- Ter um elevado grau de certeza. Contemplar o menor número de casos possível de exemplos que, apesar de entrarem nos critérios da definição positiva de *churn*, continuaram, na realidade, ativos.

Desta forma, algumas das estratégias adoptadas na definição da variável dependente, *churn*, são:

- Definição estática, em que é definido um período de tempo fixo, no qual, se o cliente não efetuar qualquer compra é considerado como *churner*. É a estratégia mais simples, já tendo sido usada, inclusivamente, na modelação preditiva de *churn* no retalho (Ju e Guo, 2008), em que foi definido que se o cliente não efetuasse compras durante um período de 6 meses era considerado *churner*.
- Definição dinâmica, que é uma estratégia que leva em conta o comportamento do cliente, ou seja, a definição é distinta para cada cliente. São exemplos, desta estratégia, duas

definições usadas em Chiang et al. (2003) que fazendo a devida analogia para o domínio do retalho seriam:

- É considerado *churn* quando o cliente não efetua qualquer compra durante um intervalo de tempo superior ao seu período médio entre compras no passado (pode ser generalizado para n períodos sem fazer compras).
 - É considerado *churn* quando o cliente passa um intervalo de tempo, sem efetuar qualquer compra, superior ao seu maior período entre compras do passado.
- Definição parcial, que essencialmente diverge das duas anteriores por modelar o evento de *churn* apenas parcialmente, ou seja, é assumido o *churn* parcial. Buckinx e Van den Poel (2005) adotam esta estratégia incluindo no modelo apenas aqueles que são considerados bons clientes, e definindo como estando no estado de *churn* parcial, os clientes que desrespeitam os critérios pelos quais foram escolhidos como bons clientes, estando estes conceptualizados com as seguintes regras:
- A frequência de compras é superior à média.
 - A relação entre o desvio padrão do tempo entre compras e o tempo médio entre compras está abaixo da média.

Além disso, tem-se que considerar como momento de *churn* o indicador mais antigo do evento (Yan et al., 2004), ou seja, no caso de uma definição estática, é considerado como o momento de *churn* o início do período que o cliente esteve sem fazer compras. Quando tal não é feito, corre-se o risco de se estar a incluir falsos indicadores nos dados de treino. Um simples exemplo disto é incluir, no treino do modelo, clientes que entram na condição de *churn*, mas que na realidade o momento de *churn* ocorreu antes do período de definição de *churn*.

3.2.4 Problema de Balanceamento das Classes

Uma importante característica na previsão de *churn* é o facto de as classes serem altamente desequilibradas, estando a classe relativa aos eventos de *churn* muito pouco representada. Este facto influencia negativamente o modelo de previsão e a sua precisão. Weiss (2004) faz um levantamento dos problemas que surgem na aplicação de técnicas de *data mining* a conjuntos de dados em que a representação das diferentes classes é muito diferente:

- Uso de medidas de avaliação erradas – Uso de métricas que não valorizam adequadamente a raridade, levando a uma pobre avaliação da classe menos representada.
- Falta de dados no sentido absoluto - O número de exemplos da classe minoritária é pequeno no sentido absoluto, o que torna difícil a detecção de regularidades existentes nessa mesma classe.
- Falta de dados no sentido relativo - Os exemplos não são raros no sentido absoluto, mas são raros relativamente à classe majoritária, o que torna mais difícil para heurísticas de procura *greedy*, e métodos mais globais não são, normalmente, efetivos.
- Fragmentação dos dados - Muitos algoritmos de *data mining*, tal como as árvores de decisão, adoptam uma estratégia *divide-and-conquer*, onde o problema inicial é decomposto sucessivamente em problemas mais pequenos. Isto pode ser uma adversidade, porque as regularidades (padrões) existentes nos dados, só podem ser encontradas nas partições individuais, que irão conter uma menor quantidade de dados.
- *Bias* indutivo inapropriado - A generalização a partir de exemplos, ou indução, requer um *bias* extra evidência. Sem este *bias* os "saltos indutivos" não são possíveis e aprendizagem não pode ser feita. Deste modo, este *bias* é um factor fundamental para um sistema de *data mining*. Muitos modelos utilizam um *bias* geral para aumentar a generalização e evitar o *overfitting* (secção 3.4.1) e isso pode ter um efeito adverso na capacidade de aprendizagem da classe rara.
- Ruído dos dados - O ruído nos dados afecta o comportamento de qualquer sistema de *data mining*, mas o seu impacto é maior nas classes menos representadas.

Para Burez e Van den Poel (2009), os problemas mais significantes que surgem na previsão de *churn* em *datasets* desequilibrados, relativamente à distribuição das classes são o uso de medidas de avaliação erradas, a falta de dados no sentido relativo e o ruído nos dados. Uma das técnicas mais utilizadas para combater a pouca representatividade dos casos de *churn* é a amostragem. A ideia é minimizar a diferença de representação das classes alterando a distribuição dos exemplos de treino. Isto pode ser feito de 2 maneiras: através da duplicação de exemplos da classe minoritária (*over-sampling*) ou através da eliminação de exemplos da classe majoritária (*under-sampling*) (Yan et al., 2004; Burez e Van den Poel, 2009).

3.3 Trabalho Relacionado

Dois raros exemplos da previsão de *churn* no retalho são os trabalhos de Buckinx e Van den Poel (2005) e Ju e Guo (2008). Nestes trabalhos são adotadas abordagens distintas. No estudo de Buckinx e Van den Poel (2005) o objectivo da modelação é prever o *churn* parcial (3.2.3) dos clientes considerados fiéis, ou seja, a previsão é feita apenas sobre os clientes que cumprem um determinado critério de fidelidade (clientes que fazem compras com muita frequência e que exibem um padrão de visitas regular). Já em Ju e Guo (2008) o *churn* é modelado de forma total e estática (secção 3.2.3). São considerados *churners* os clientes que não efetuaram qualquer compra durante um período de seis meses. Também aqui, os autores limitaram o modelo aos clientes VIP. Em ambos os trabalhos são usados dados de duas categorias:

- Sociodemográficos.
- Comportamentais.

Além disso, os dois estudos divergem nas técnicas de classificação adoptadas. No trabalho de modelação de *churn* parcial os autores optaram pelas técnicas de regressão logística (secção 3.4.2), redes neuronais (secção 3.4.3) e *random forests* (secção 3.4.4) (Buckinx e Van den Poel, 2005). Foram obtidos resultados semelhantes para as diferentes técnicas, mas as *random forests* apareceram consistentemente no topo. Ju e Guo (2008) adotaram o método SVM (3.4.5).

3.4 Revisão das Técnicas de Classificação

O problema de previsão de *churn*, como tem sido descrito, é, de um modo geral, conceptualizado como um problema de classificação. Existe uma enorme variedade de técnicas de *data mining* para aplicar em tarefas de classificação. Uma das principais decisões a tomar é a escolha da técnica a usar no processo, pois esta pode implicar variações de centenas de milhares de euros no retorno das campanhas (Neslin et al., 2006). Um dos principais factores a ter em conta, na escolha da técnica a usar, é o compromisso entre a interpretação e a precisão que se pretende obter (secção 2.4.2). Por um lado, existem técnicas de fácil interpretação, como as árvores de decisão, que nos dão as razões, ou pelo menos, os sintomas do *churn*. Para além disso, permitem detetar facilmente

erros no tratamento dos dados, como a falta de determinados atributos ou valores anormais nos dados [Owczarczuk, 2010]. Por outro lado, existem os modelos do tipo "*black-box*", como as *random forests* e as SVM, que podem, por vezes, apresentar melhores resultados na tarefa de previsão. Neslin et al. (2006) apontam que, no contexto da previsão de *churn*, as técnicas de mais fácil interpretação poderão não obter uma performance tão boa na previsão como esses modelos. É então necessário chegar a um compromisso entre o tempo despendido na modelação e a performance a alcançar (Coussement e Van den Poel, 2008a).

Algumas das técnicas apresentadas podem ser utilizadas para diferentes finalidades, mas são apresentadas no contexto da tarefa de classificação. As técnicas a estudar foram escolhidas com base no estudo, de KhakAbi et al. (2010), sobre as técnicas de mineração de dados utilizadas na previsão de *churn*. Entre as muitas técnicas existentes para esse fim foram selecionadas as cinco mais utilizadas (Figura 21): árvores de decisão, regressão logística, redes neuronais, *random forests* e SVM.

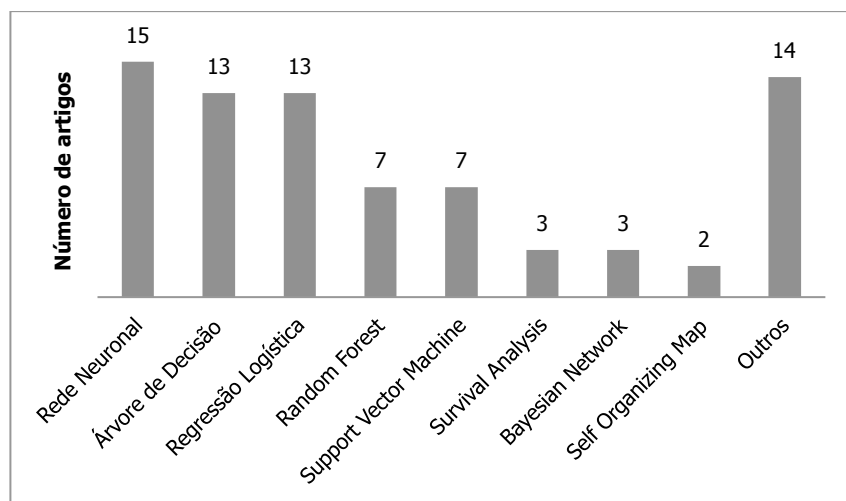


Figura 21: Tendências na previsão de *churn* - baseado em (KhakAbi et al., 2010)

3.4.1 Árvores de Decisão

Uma árvore de decisão (Figura 22) é uma estrutura, em forma de árvore, que é usada para representar a divisão de um conjunto de elementos em conjuntos sucessivamente menores e mais homogêneos através de regras de decisão (Berry e Linoff, 2004). A indução de uma árvore de

decisão é o processo de aprendizagem a partir de um conjunto de dados de treino em que o atributo que se pretende classificar (classe) está disponível. Um dos principais trunfos deste algoritmo de classificação é a sua representação do conhecimento adquirido na forma de árvore, o que o torna bastante intuitivo e de fácil compreensão (Hu, 2005; Zan et al., 2007; Breiman, 2001b). É ainda um algoritmo conceptualmente simples, tanto na fase de treino, como na fase de classificação. Na sua generalidade, os classificadores de árvores de decisão obtêm uma boa precisão e conseguem tratar problemas com dados de grande dimensionalidade (Han e Kamber, 2006).

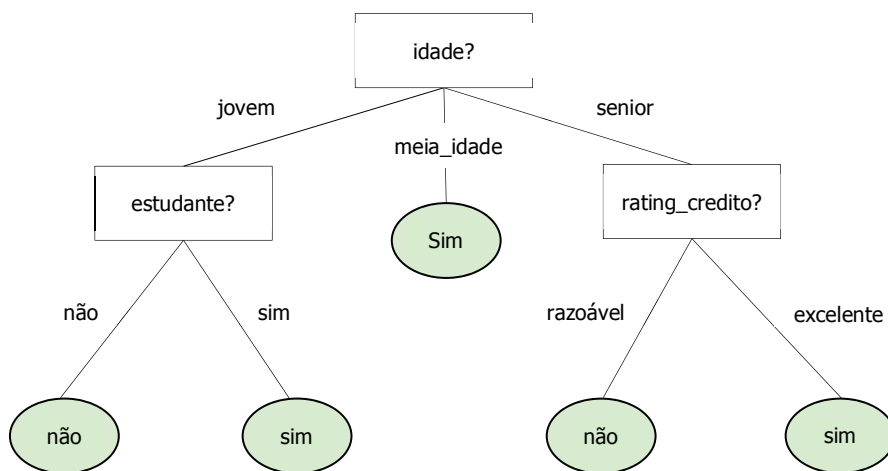


Figura 22: Árvore de decisão - adaptado de (Han e Kamber, 2006)

As árvores de decisão são constituídas por nós, que representam os sucessivos testes lógicos (aos atributos) das regras de decisão, por ramos, que contêm o valor do atributo testado e finalmente, por folhas, que representam o valor resultado (*output*) da regra. O nó do topo da estrutura é conhecido como nó raiz e representa o teste ao primeiro atributo escolhido pelo algoritmo de indução da árvore. Os restantes nós, entre o nó raiz e as folhas, são conhecidos como nós internos. A Figura 22 ilustra uma árvore de decisão do conceito "compra_computador" (classe) que indica se é provável que um cliente compre um computador. Por exemplo, se a árvore for sempre percorrida pela direita conseguimos identificar a seguinte regra de classificação:

"Se idade = senior e rating_credito = excelente então sim (compra)"

Algoritmos de Indução de Árvores de Decisão

Existem vários algoritmos de indução de árvores de decisão. Contudo, todos os algoritmos fundamentam-se no mesmo procedimento: separar repetidamente os dados em conjuntos sucessivamente menores, para que os novos subconjuntos tenham maior "pureza" que o conjunto a partir do qual foram gerados, relativamente à variável dependente (Berry e Linoff, 2004). A construção da árvore começa no nó raiz e desenvolve-se no sentido das folhas. Para cada nó é selecionado o atributo mais "informativo" através de uma medida de seleção de atributos. Os dados são particionados em 2 (ou mais) novos subconjuntos. O processo é então repetido nos novos subconjuntos criados. O processo deve terminar com uma das seguintes condições (Berry e Linoff, 2004):

- todos os exemplos pertencem à mesma classe;
- o número de exemplos é inferior a um dado limite;
- não há melhoria significativa com a divisão do conjunto de dados.

As principais diferenças existentes nos algoritmos de indução de árvores são relativas:

1. ao método escolhido para a seleção dos atributos que servem de teste nos nós;
2. aos mecanismos de poda (*prunning*) da árvore (Han e Kamber, 2006).

Medidas de Seleção de Atributos

Uma medida de seleção de atributos é usada para selecionar o atributo que melhor discrimina um conjunto de exemplos de treino nas diferentes classes (i.e. o atributo mais informativo), para um determinado nó da árvore. Os atributos são ordenados por ordem, consoante os resultados obtidos pela medida, e o que tiver o melhor resultado é escolhido para ser o atributo teste nesse nó (Han e Kamber, 2006). Três das medidas de seleção mais populares são: entropia, ganho de informação e índice Gini.

Prunning e Overfitting

Uma árvore que descreve, com demasiada precisão, os dados a partir dos quais foi construída é pouco provável que generalize corretamente para diferentes amostras da mesma população. À medida que os conjuntos de dados testados nos nós ficam mais reduzidos, as particularidades existentes no conjunto de dados começam a surgir representados pela árvore. Uma forma de pensar nisto é que os grandes nós representam padrões gerais e os pequenos nós representam padrões específicos ao conjunto de dados de treino. A este fenómeno dá-se o nome de *overfitting* e uma forma de o combater é fundir as folhas mais pequenas, um processo também conhecido como poda ou *prunning* (Berry e Linoff, 2004). Existem dois tipos de *prunning*:

- *preprunning*, no qual a poda é feita durante o processo de aprendizagem, o que torna o processo mais complexo.
- *Postprunning*, é o método mais comum, e acontece após a árvore estar totalmente construída.

3.4.2 Regressão Logística

A regressão logística é uma técnica estatística poderosa e bem estabelecida na tarefa de classificação binária (Coussement e Van den Poel, 2008a; Tsiptsis e Chorianopoulos, 2010). É uma técnica conceptualmente simples, mas que tem sido bastante utilizada em modelos de previsão de *churn* (KhakAbi et al., 2010; Neslin et al., 2006), apresentando boas performances comparativamente a outras técnicas de classificação (Neslin et al., 2006). No entanto, esta técnica é muitas vezes usada apenas como referência de comparação a outras técnicas em estudo (Buckinx e Van den Poel, 2005; Coussement e Van den Poel, 2008a; Jing e Xing-Hua, 2008).

A inclusão de todos os atributos independentes num modelo de regressão resulta regularmente em *overfitting* (secção 3.4.1) do modelo e fracas previsões, em aplicações que muitas variáveis estão pouco relacionadas com a variável dependente (Coussement e Van den Poel, 2009).

Na regressão linear a probabilidade de um resultado binário está relacionada com um conjunto de

atributos potencialmente explicativos na forma (Leung e Tran, 2000):

$$\log[p / (1 - p)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

na qual p é a probabilidade do evento a ser previsto e β_1, \dots, β_n são os coeficientes associados a cada variável independente x_1, \dots, x_n . A variável dependente é o logaritmo das *odds*¹ do evento. Exemplificando, para o caso da previsão de *churn* essa equação ficaria da seguinte forma:

$$\log\left(\frac{p(\text{churn})}{p(\text{"não churn"})}\right) = \beta_0 + \beta_1.\text{atributo1} + \beta_2.\text{atributo2} + \dots + \beta_n.\text{atributon}$$

O logaritmo das *odds* está relacionado linearmente com as variáveis independentes e, desse modo, a utilização de diferentes técnicas de seleção de atributos pode dar origem a diferentes modelos de regressão, mas que são, geralmente, similares. Os coeficientes β_1, \dots, β_n são estimados através do método *maximum-likelihood* (Leung e Tran, 2000).

3.4.3 Redes Neurais

As redes neurais foram originalmente pensadas por psicólogos e neurobiólogos, com o objectivo de tentar modelar o comportamento dos neurónios humanos por via computacional (Han e Kamber, 2006). Estas, propõem, de certa forma, com a generalização e aprendizagem a partir de dados, imitar a nossa aprendizagem através da experiência (Berry e Linoff, 2004). A aprendizagem numa rede neuronal faz-se através do ajustamento de pesos distribuídos pela rede e tal como o processo do pensamento humano nos é desconhecido, o valor dos pesos também não nos permite a compreensão do processo que levou ao resultado final.

Topologia de uma Rede *Feed-Forward*

Uma rede neuronal é constituída por uma camada de entrada, uma ou mais camadas escondidas e uma camada de saída. Cada camada é constituída por uma ou mais unidades de processamento dos dados. A ligar as unidades das diferentes camadas existem conexões que têm um peso a si

¹ Razão entre duas probabilidades: a probabilidade do evento acontecer sobre a probabilidade do evento não acontecer.

associado. As redes *feed-forward*, como o próprio significado em inglês sugere, propagam os dados desde a camada de entrada até à camada de saída não existindo qualquer mudança de sentido no interior da rede (Figura 23).

A camada de entrada de uma rede neuronal serve de interface entre a rede e os dados. O número de unidades desta camada é variável e corresponde ao número de campos (atributos) que são processados pela rede. A sua função é a transformação dos dados para uma gama de valores razoável, normalmente compreendidos entre -1 e 1 (Berry e Linoff, 2004) ou 0 e 1 (Han e Kamber, 2006). Após essa transformação, os valores são enviados para a camada adjacente seguinte.

As camadas intermédias são também designadas por camadas escondidas. Os valores recebidos provêm de cada uma das unidades da camada de entrada, e dos pesos guardados nas conexões com essas unidades. O número de camadas escondidas é definido pelo utilizador, mas normalmente uma é suficiente (Berry e Linoff, 2004; Han e Kamber, 2006). Estas são constituídas por unidades designadas por neurónios artificiais. Quantas mais unidades tiver, maior será a capacidade, da rede, de reconhecer padrões. Todavia, se for um número demasiado elevado corre-se o risco de ocorrer *overfitting* (secção 3.4.1) sobre os dados de treino (Berry e Linoff, 2004). O valor resultante da atividade de cada uma das unidades é então enviado para a próxima camada, de modo geral, a camada de saída.

Na camada de saída cada unidade recebe, tal como nas camadas escondidas, o conjunto dos valores calculados e enviados por cada uma das unidades da camada anterior e os pesos guardados nas respectivas conexões. É desta camada que sai o resultado final de cada tuplo processado pela rede. O número de unidades depende do tipo de previsão que se pretende fazer. Se o problema for de classificação numérica ou de classificação binária uma unidade é suficiente. Caso seja um problema de classificação, esta camada terá o número de unidades igual ao número de classes existentes.

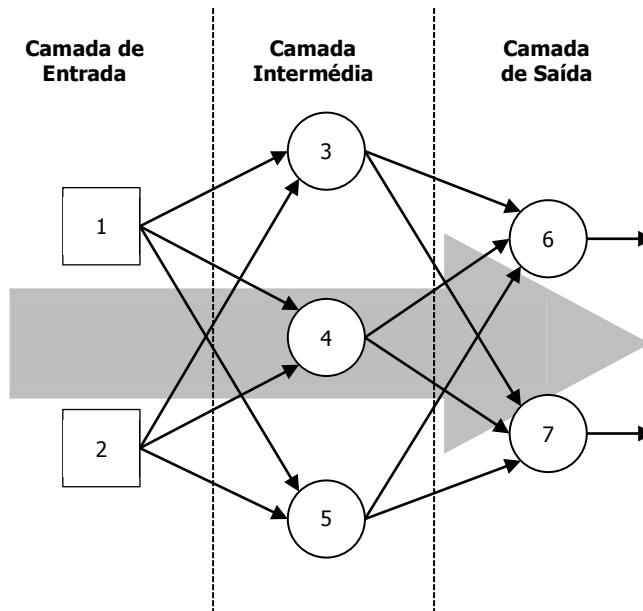


Figura 23: Sentido de propagação dos dados - adaptado de (Cortez, 2002)

Neurónio Artificial

Os neurónios artificiais são as unidades que compõem as camadas escondidas e a camada de saída. São os componentes principais no processamento dos dados, que estão interligados, através de conexões, às unidades pertencentes a outras camadas adjacentes e têm um valor a si associado conhecido por *bias* (Berry e Linoff, 2004; Han e Kamber, 2006). O cálculo efectuado nestas unidades divide-se em duas etapas (Figura 24):

1. Combinação de todos os valores recebidos num único valor. De modo geral, a função de combinação corresponde ao somatório dos valores resultantes da multiplicação de cada valor recebido com o peso associado à conexão. Ao resultado é ainda somado o valor *bias* da unidade (Berry e Linoff, 2004).
2. Aplicação da função conhecida como função de ativação ao resultado do 1º passo (Han e Kamber, 2006). As funções de ativação mais usadas são a logística, linear e tangente hiperbólica. No entanto, o resultado de aplicar a função linear é equivalente a efetuar uma regressão linear (Berry e Linoff, 2004).

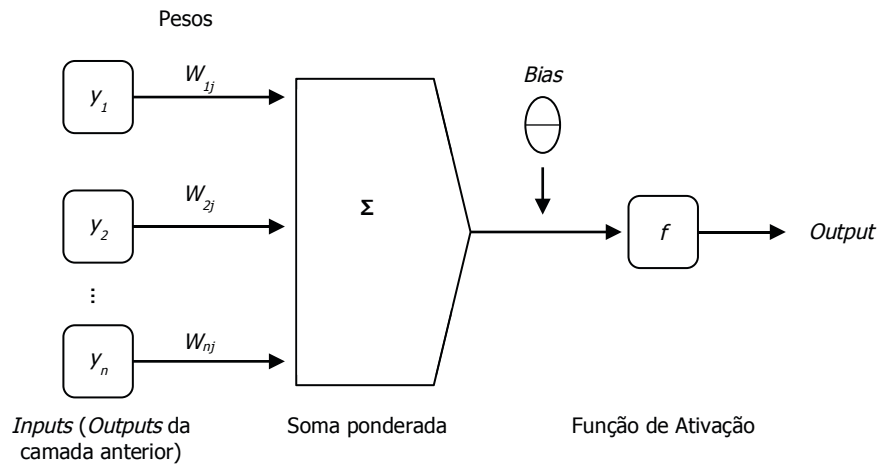


Figura 24: Processamento num neurónio artificial - adaptado de (Han e Kamber, 2006)

Estas unidades estão também associadas ao cálculo e propagação dos erros, no processo de optimização dos pesos das conexões, apresentado de forma resumida de seguida.

Algoritmo de Aprendizagem *Backpropagation*

Estando especificada a estrutura da rede, o objectivo do algoritmo de aprendizagem é determinar os pesos associados às suas conexões e unidades, de maneira a que os resultados de saída estejam o mais próximo possível dos valores reais dos dados treino. A aprendizagem *backpropagation* (BP) é feita iterativamente através do processamento do conjunto de tuplos de treino, comparando, para cada um deles, o valor calculado, pela rede, da variável classe com o seu valor real. Para cada exemplo de treino, os pesos, distribuídos pelas conexões, são modificados com o objectivo de minimizar o erro quadrado médio entre a previsão da rede e o verdadeiro valor da classe. Estas modificações são feitas a partir da camada de saída e no sentido da camada de entrada (Figura 25) (Han e Kamber, 2006). O processo do algoritmo de BP pode ser resumido em três fases (Berry e Linoff, 2004):

1. Cálculo do valor de saída de um exemplo de treino, através do uso dos pesos existentes na rede (estão fixos nesta fase), com o processamento a ser efectuado da camada de entrada para a camada de saída (*feed-forward*).
2. Cálculo da diferença entre o resultado processado pela rede e o valor real do exemplo, ou seja, o erro associado ao processamento do *tuplo*.

3. Propagação do erro da camada de saída até à camada de entrada pelos pesos existentes na rede (origem do nome *backpropagation*); neste processo, os pesos existentes nas conexões da rede são ajustados para minimizar o erro.

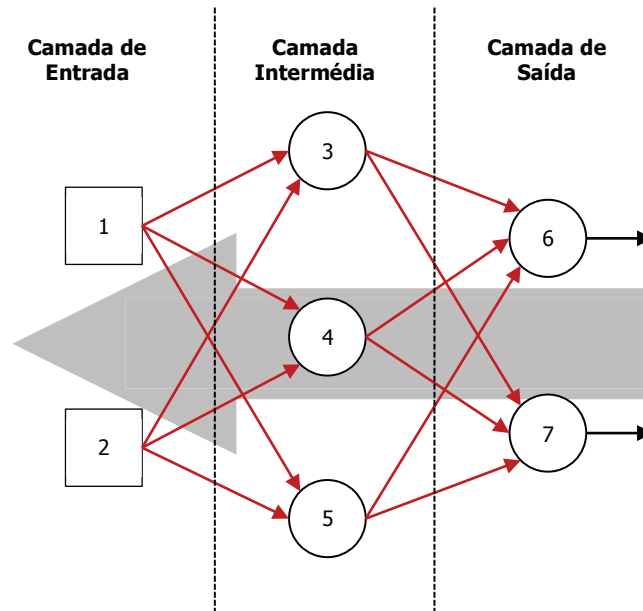


Figura 25: Retro propagação do erro - adaptado de (Cortez, 2002)

Os parâmetros típicos que um utilizador tem para definir numa rede neuronal são:

- Número de Camadas Escondidas. Normalmente, uma camada escondida é suficiente (Berry e Linoff, 2004).
- Número de unidades nas camadas escondidas. Não existe uma regra clara para definir este parâmetro. Quando mais unidades tiver a rede maior será a sua capacidade de reconhecer padrões. Não obstante, se o número de unidades for demasiado elevado corre-se o risco de acontecer *overfitting* sobre os dados de treino (Berry e Linoff, 2004).
- *Momentum*. Este parâmetro refere-se à tendência que os pesos têm para mudar a direção da sua variação, sendo utilizado para estabilizar as mudanças de direção da variação dos pesos. Uma rede com um valor de *momentum* alto responde lentamente aos novos tuplos que queiram mudar a direção de correção do peso. Por sua vez, se tiver um valor baixo, é permitida uma mais livre mudança da direção de atualização dos pesos (Berry e Linoff, 2004).

- *Learning Rate*. É um parâmetro utilizado para controlar a amplitude da variação dos pesos no processo de otimização. O seu valor está, normalmente, inserido no intervalo entre 0.0 e 1.0. Se o valor for demasiado baixo o treino será muito lento. Por outro lado, se o valor for demasiado alto pode ocorrer uma oscilação que não encontre uma boa solução (Han e Kamber, 2006). Berry e Linoff (2004) afirmam que a melhor estratégia para usar neste parâmetro é começar com um valor alto e fazê-lo decrescer lentamente durante o treino da rede.

3.4.4 Random Forests

Recentemente tem havido muito interesse por métodos de composição de modelos, ou seja, algoritmos de aprendizagem nos quais é construído um conjunto de classificadores que posteriormente classificam os novos dados através de um processo de votação (Dietterich, 2000). Dois dos métodos de composição mais conhecidos são o *boosting* (ver, e.g., (Schapire et al., 1998)) e o *bagging* (Breiman, 1996) de árvores de classificação.

O método *bagging*, o mais simples, consiste em usar um conjunto de treino independente na construção de cada classificador, obtido através de amostragem com reposição. A previsão é feita por voto maioritário, ou seja a votação é uniforme. No método de *boosting* cada exemplo de treino tem um peso a si associado. Assim, quando um exemplo é mal classificado por um classificador, o seu peso é aumentado tendo assim uma maior influência nos classificadores seguintes. Além disso, fica associado um peso a cada classificador, relativo à taxa de erro obtida, para ser usado posteriormente na tarefa de classificação. Esta é feita através do voto ponderado de todos os classificadores (Liaw e Wiener, 2002).

O método *random forests* foi proposto em Breiman (2001a). Este tem um funcionamento que se assemelha ao método *bagging*, mas acrescenta uma camada extra de aleatoriedade. Para além do que é feito no método *bagging*, este método modifica o processo de crescimento das árvores de classificação, mais particularmente, o método de divisão dos nós. Nas *random forests*, cada nó é dividido usando o melhor atributo de um subconjunto aleatoriamente escolhidos para esse nó (Liaw e Wiener, 2002).

As *random forests* já foram aplicadas em vários domínios como a bioinformática, criminologia quantitativa, geologia e medicina (Burez e Poel, 2008) e nos últimos anos têm recebido especial atenção na prevenção de *churn* de clientes (Buckinx e Van den Poel, 2005; Burez e Poel, 2008; Larivière e Van den Poel, 2005). As suas maiores vantagens são a robustez contra *overfitting* e a facilidade de uso, já que, o utilizador só necessita definir dois parâmetros para o uso deste algoritmo (Breiman, 2001a). Para além destas vantagens, foram já usadas na área da prevenção de *churn* no retalho onde obtiveram excelentes resultados e com tempos de execução razoáveis (Buckinx e Van den Poel, 2005). Os dois parâmetros que são necessários definir são (Breiman, 2001a): o número de árvores a serem construídas, que deve ser um número elevado; e o número de atributos que irão ser escolhidos aleatoriamente em cada nó.

3.4.5 Support Vector Machines

As maiores vantagens das SVM são a sua capacidade de modelação de dados complexos e altamente não lineares e a sua capacidade de evitar *overfitting* aos dados de treino. Por outro lado, é um método que exige elevados recursos de memória e de tempo de processamento, podendo ser extremamente lento (Han e Kamber, 2006; Tsitsis e Chorianopoulos, 2010). As SVM têm demonstrado um excelente performance num vasto conjunto de problemas incluindo a bioinformática, classificação de texto e detecção facial através de imagens. Têm sido também frequentemente aplicadas no contexto comercial, nomeadamente em aplicações de *marketing*, previsão de bancarotas e avaliação de crédito, assim como, em aplicações de previsão de *churn* (KhakAbi et al., 2010; Kim et al., 2005).

Dados Linearmente Separáveis

Na Figura 26 podemos verificar uma situação em que os dados são linearmente separáveis, ou seja, é possível traçar uma linha recta que discrimina perfeitamente os exemplos das duas classes. Também aí, é possível verificar que existe um número infinito de rectas que o podem fazer. O objectivo é encontrar a melhor recta, ou seja, aquela que minimiza o erro na classificação de novos exemplos. Para além disso, também se chega facilmente à conclusão que para separar exemplos com mais de duas dimensões será necessário um plano. Assim, generalizando o problema de classificação para n dimensões, o objectivo do algoritmo SVM é o de encontrar o melhor plano para

separar as diferentes classes, que é conhecido tipicamente por hiperplano linear óptimo (HLO). Este é o plano que maximiza as margens de separação entre os vectores de suporte (os exemplos mais próximos ao hiperplano) das diferentes classes (Coussement e Van den Poel, 2008a; Han e Kamber, 2006).

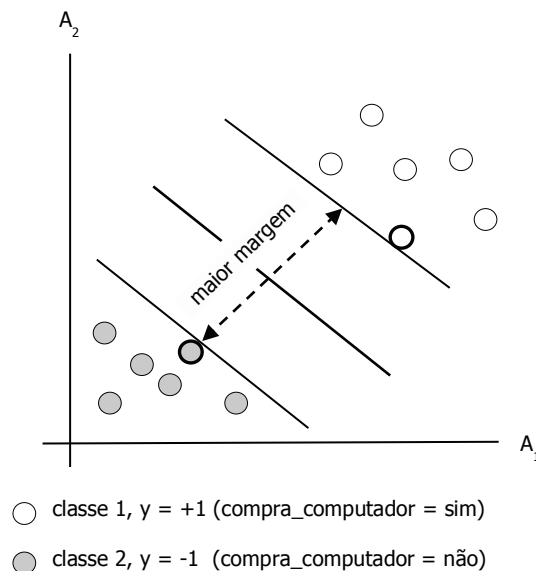


Figura 26: Separação das classes pelo HLO - adaptado de (Han e Kamber, 2006)

Dados Linearmente Inseparáveis

Contudo é provável que os dados tenham uma estrutura complexa e não linear relativamente às classes (Ju e Guo, 2008). Para que se possa fazer a classificação de dados que não sejam linearmente separáveis, os algoritmos SVM transformam os dados de treino para um espaço dimensional superior através de um mapeamento não linear, com o objectivo de os tornar separáveis através de uma função linear. Essa transformação dos dados pode ser feita recorrendo a funções não lineares conhecidas como funções *kernel*. Este processo está ilustrado de forma simples na Figura 27. Após a aplicação de uma função *kernel* o procedimento é o mesmo do que foi descrito para os dados separáveis linearmente, ou seja, é procurado o melhor hiperplano linear que separe as diferentes classes acabando, do mesmo modo, num problema de optimização quadrática (Han e Kamber, 2006).

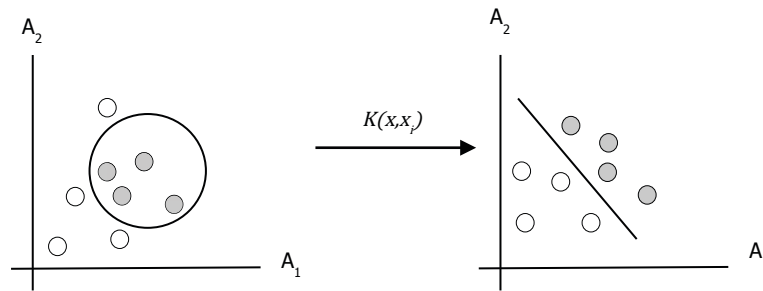


Figura 27: Mapeamento dos dados - adaptado de (Coussement e Van den Poel, 2008a)

3.5 Métodos para a Avaliação de Modelos

A avaliação dos modelos é um ponto-chave do processo de mineração de dados. Durante o processo de modelação são testados vários algoritmos e diferentes parametrizações para cada um deles, o que faz com que seja necessário ter um método de avaliar a evolução do processo e comparar os diferentes algoritmos e respectivos parâmetros.

Para além da avaliação da sua precisão, um método pode ser avaliado por diversos critérios, nomeadamente (Han e Kamber, 2006):

- Velocidade, que é o custo computacional envolvido no treino e uso do classificador.
- Robustez, que é a capacidade do classificador fazer previsões corretas com existência de ruído nos dados ou falta de dados.
- Escalabilidade, que é a capacidade de construção do modelo de forma eficiente a partir de grandes conjuntos de dados.
- Interpretabilidade, que é o nível de compreensão e discernimento sobre o fenómeno modelado que é possível retirar do classificador.

Tal como nas restantes fases do processo de modelação, na fase de avaliação, também é necessário fazer várias opções relativas ao processo. É necessário escolher o método e a(s) medida(s) para avaliação do modelo que melhor se adequem às características e requisitos do problema. As principais decisões a tomar nesta fase são:

- o método de avaliação, ou seja, a forma como são divididos os dados para os conjuntos de treino e de teste e o tipo de procedimento para estimar a precisão do modelo;
- as métricas de avaliação do modelo.

3.5.1 Métricas de Avaliação de Modelos de Classificação

Uma ferramenta muito útil e direta para analisar a precisão do modelo para as diferentes classes é a matriz de confusão. Uma matriz de confusão para um problema de classificação binária discrimina, para ambas as classes, os casos corretamente classificados e os erros de classificação (Tabela 6). As colunas dizem respeito às classes previstas e as linhas às classes verdadeiras.

Tabela 6: Matriz de confusão

		Previsto		
		<i>Churner</i>	Não <i>Churner</i>	
Real	<i>Churner</i>	Verdadeiro Positivo (<i>hit</i> , VP)	Falso Negativo (<i>miss</i> , erro Tipo II, FN)	VP + FN = P
	Não <i>Churner</i>	Falso Positivo (falso alarme, erro Tipo I, FP)	Verdadeiro Negativo (rejeição correta, VN)	FP + VN = N

A partir de uma matriz de confusão facilmente se deduzem outras medidas também bastante relevantes. A *Percentagem de casos Corretamente Classificados* (PCC), também conhecida por *accuracy*, é sem dúvida a medida de performance de classificadores mais usada. Todavia, poderá não ser a medida mais adequada para alguns problemas, pois assume, que o custo de classificação errada é igual para ambas as classes (Baesens et al., 2002). No caso da previsão de *churn*, este pressuposto é problemático, já que a principal meta deste problema é encontrar o máximo de casos de *churn*, ou seja, os falsos positivos e os falsos negativos têm impactos claramente diferentes.

$$Accuracy (PCC) = \frac{VP + VN}{VP + FP + VN + FN} = \frac{VP + VN}{P + N}$$

Uma outra característica, para a qual pode ser problemático o uso desta medida, é o desequilíbrio

da distribuição das classes no conjunto de dados. Por exemplo, para o caso de previsão de *churn*, se o conjunto de dados possui 90% de casos negativos e 10% de casos positivos, e o modelo classifica todos os casos como negativos a precisão do modelo será de 90%. À primeira vista, pode parecer que o modelo tem uma boa performance, no entanto, não consegue identificar nenhum caso verdadeiro de *churn*.

Contudo, existem outras medidas que fornecem resultados mais significativos para os casos em que a *accuracy* não é uma boa opção. De uma matriz de confusão também se extraem essas métricas. A *sensitivity* (ou, em português, sensibilidade) mede a proporção de exemplos positivos que foram corretamente classificados como positivos. A *specificity* (ou, em português, especificidade) mede a proporção dos exemplos negativos corretamente classificados como negativos (Baesens et al., 2002). A *precision* diz-nos a proporção de casos corretamente classificados no total de casos classificados como positivos. Estas métricas podem ser calculadas da seguinte forma:

$$Sensitivity = \frac{VP}{VP + FN} = \frac{VP}{P}$$

$$Specificity = \frac{VN}{VN + FP} = \frac{VN}{N}$$

$$Precision = \frac{VP}{VP + FP}$$

O principal objectivo quando se trabalha com conjuntos de dados desequilibrados, relativamente à distribuição das classes, é melhorar a *sensitivity* sem piorar a *precision*. Contudo, as duas métricas estão usualmente em conflito, pois com o aumento dos VP da classe minoritária os FP também podem aumentar, o que irá reduzir a *precision*. A métrica *F-value* (ou *F-measure*) é uma medida que combina a sensibilidade e a precisão e que estabelece o compromisso entre as duas medidas dando como resultado um número singular (Chawla, 2005):

$$f_value = \frac{(1 + \beta^2) * sensitivity * precision}{\beta^2 * sensitivity + precision}$$

em que o β corresponde à importância relativa entre *precision* e *sensitivity*, sendo normalmente definida como 1.

AUC

Uma curva ROC (*Receiver Operating Characteristic*) mostra a relação entre a taxa de verdadeiros positivos (*sensitivity*) e a taxa de falsos positivos ($1 - \textit{specificity}$), para um modelo de classificação binária. A Figura 28 mostra as curvas ROC de dois modelos de classificação. O eixo vertical representa a taxa de verdadeiros positivos e o eixo horizontal representa a taxa de falsos positivos ($\%FP = FP / (VN + FP)$). Para além disso, mostra uma linha diagonal que representa uma fronteira em que a probabilidade de encontrar um verdadeiro positivo é igual à probabilidade de encontrar um falso positivo. Quanto melhor for o modelo, mais abruptamente a curva ROC sobe a partir do zero. Por outro lado, à medida que vão sendo encontrados menos exemplos de positivos verdadeiros a curva vai se tornando mais horizontal. A *Area Under Curve* (AUC) fornece uma métrica para avaliar, em média, qual o melhor algoritmo (Han e Kamber, 2006; Chawla, 2005).

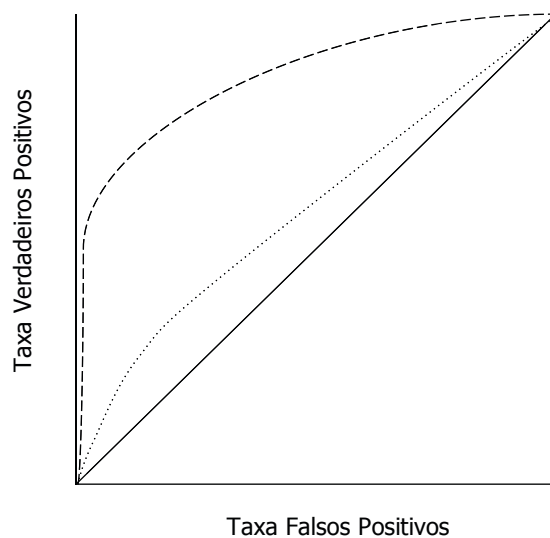


Figura 28: Curvas ROC de dois modelos - adaptado de (Han e Kamber, 2006)

Lift

Quando, na fase de validação, se ordena o conjunto de dados de validação por probabilidade decrescente de ocorrência de eventos (e.g. *churners*), se o método de classificação for eficaz, deve-se encontrar uma maior concentração de verdadeiros positivos no topo da lista. Essa maior proporção pode ser medida em termos de *lift*, que nos diz o quanto melhor é o classificador relativamente à classificação aleatória. Enquanto que, a medida *accuracy* mede as previsões

corretas para o conjunto de dados total, o *lift* mede o aumento da precisão num determinado subconjunto, e.g. no top 10% da lista ordenada por *score* de um modelo (Piatetsky-Shapiro e Masand, 1999). Este pode ser calculado através da seguinte equação:

$$Lift = \frac{Precision}{P/(P+N)} = \frac{VP/(VP+FP)}{P/(P+N)}$$

De um modo geral, o *lift* reflete o aumento da densidade de *churners* previstos pelo classificador, relativamente à densidade de *churners* existente na conjunto de dados total (Coussement e Van den Poel, 2008b).

3.5.2 Métodos de Avaliação

Uma outra questão, também importante, é a maneira como vão ser aplicadas as medidas descritas anteriormente de forma a avaliar os modelos de forma fidedigna, isto é para que os resultados obtidos, na fase de avaliação, sejam idênticos aos resultados que o modelo irá obter na aplicação do modelo na previsão de novos casos. Dois dos métodos mais utilizados são a validação por divisão da amostra (*holdout validation*) e validação cruzada (*cross-validation*).

O estimador *holdout* procede dividindo os dados em dois conjuntos: um conjunto para treinar o modelo e o outro para testar o modelo. Se por um lado é um estimador simples e rápido, por outro, reduz a quantidade de dados para o processo de aprendizagem. Geralmente são usados 2/3 dos dados para a aprendizagem do modelo e 1/3 é reservado para o teste (Santos e Azevedo, 2005).

Ao contrario do método *holdout*, o método *cross-validation*, permite a utilização do conjunto total dos dados para aprendizagem do modelo. Os dados são divididos em k conjuntos mutuamente exclusivos de tamanhos idênticos. Os modelos são treinados com k-1 desses conjuntos e testados com o conjunto restante. O processo é repetido k (tipicamente 10) vezes. No final, o valor da avaliação do estimador é dado pela média dos valores obtidos em cada iteração. A desvantagem deste método é o incremento considerável do esforço computacional. Na Figura 29 apresenta-se, de uma forma bastante simples, o processo de divisão dos dados para as referidas k iterações.

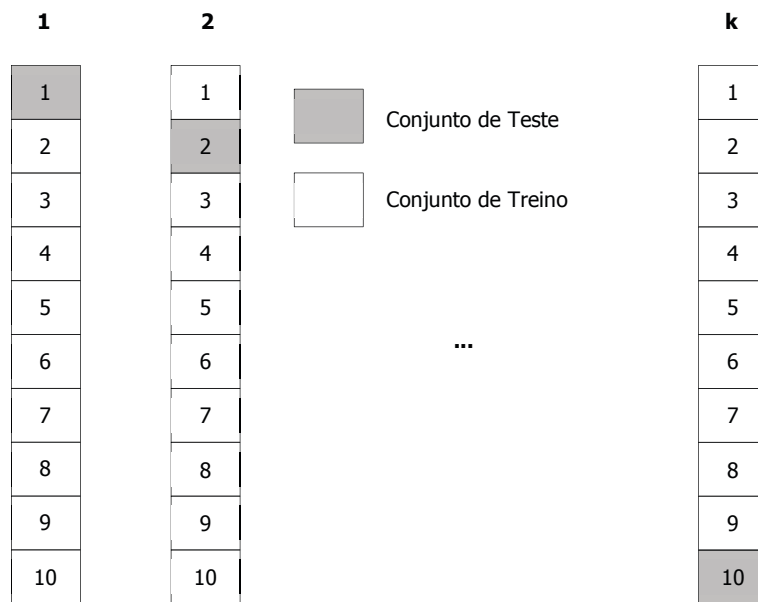


Figura 29: Método de estimação *cross-validation*

Capítulo 4

Aplicação e Validação de Modelos de *Churn*

4.1 Conhecimento das Regras e Objectivos do Negócio

4.1.1 Enquadramento Experimental

A experiência realizada neste trabalho foi desenvolvida no contexto de um retalhista² com posição de destaque no mercado, com milhares de clientes registados. O comportamento dos clientes é seguido e associado aos dados demográficos via cartão de fidelidade. Neste retalhista cerca de 60% das compras são registadas pelo cartão de fidelidade. Os dados dos clientes e transações são registados no *data warehouse* da empresa que acumula o histórico das suas vendas dos últimos 5 anos. Neste retalhista o tempo médio entre atos de compra consecutivos é de 16 dias. Baseados no trabalho de Ju e Guo (2008) e de Buckinx e Van den Poel (2005), nos quais os autores acreditam que, respectivamente, 4 e 5 meses são suficientes para o período de observação, iremos

² Cenário fictício, preparado de forma a poder estudar e aplicar os modelos e técnicas de deteção de churn.

utilizar um período de observação de 6 meses neste processo de estudo e análise de modelos para a detecção de *churn*.

Descrição dos Objectivos do Modelo de Previsão

O objetivo da empresa é detectar clientes que estão inclinados a mudar para uma empresa da concorrência, de modo a poder desenvolver ações com vista à sua manutenção, de forma a que eles não abandonem a empresa. Como é necessário um período para aplicar as estratégias de retenção, o objectivo deste trabalho prático é detectar os clientes que vão abandonar a empresa no período de 30 a 90 dias após a aplicação das estratégias referidas.

Definições da Janela Temporal e Clientes Escolhidos

Numa primeira abordagem optou-se por uma definição estática da variável dependente *churn*. Deste modo, tomaram-se algumas decisões pertinentes para este trabalho relativas aos períodos de tempo e clientes envolvidos na modelação, nomeadamente:

- O modelo é treinado apenas com clientes que estejam ativos no final do período de observação.
- O período de observação tem uma duração de 6 meses.
- Após o período de observação é reservado 1 mês para o período de latência.
- A variável *target* será definida nos 4 meses seguintes ao período de latência (período de definição).
- São considerados *churners* todos aqueles que não efetuaram qualquer compra durante 60 dias consecutivos. O início desse período terá que ocorrer durante os 2 primeiros meses do período de definição.

Coleção e Seleção dos Dados

Neste trabalho de análise foi selecionado um conjunto de variáveis que ilustram os vários aspectos da relação entre o cliente e a organização. Foram reunidos os dados a partir do *data mart* de vendas do *data warehouse* da empresa. O *data mart* armazena informação relativa às transações de venda realizadas (Figura 20) - daí foi extraído o conjunto de variáveis apresentado e descrito na

Tabela 7. As variáveis foram escolhidas e construídas com base em relatos recolhidos na literatura do domínio sobre previsão de *churn* em ambientes de retalho (Buckinx e Van den Poel, 2005; Ju e Guo, 2008).

Os dados para construção das assinaturas foram extraídos de um período temporal correspondente a 6 meses de dados. Foram selecionados os clientes que respeitam o critério referido anteriormente, ou seja, os clientes que efetuaram a primeira compra antes do período considerado e a última compra no período em questão. Desse conjunto de clientes foram selecionados os dados de um conjunto de 10000 clientes para o processo de modelação.

Tabela 7: Variáveis independentes usadas neste estudo

Tipo	Nome	Descrição
ID	CustomerKey	ID do cliente
<i>Label</i>	Churn	Variável dependente
Demográficos	Age	Idade do cliente
	maritalStatus	Estado Civil
	Gender	Sexo
	yearlyIncome	Rendimentos anuais
	totalChildren	Número de filhos
	nrChildrenAtHome	Número de filhos a viver em casa
	education	Habilitações
	occupation	Profissão - classe
	houseOwnerFlag	Indicador de posse de casa
	nrCarsOwned	Número de carros que cliente possui
	city	Cidade de habitação do cliente
Comportamentais	NrVisits	Número de visitas à loja
	frequency	Número de visitas à loja por dia
	frequencyLastMonth	Número de visitas por dia no último mês do PO
	frequencyLastMonthR	frequencyLastMonth / frequency
	frequencyVariation	Varição da frequência de compras no PO
	nrProducts	Número total de produtos comprados
	nrProductsPerVisit	Número de produtos comprados por visita
	monetary	Total gasto durante o PO
	monetaryPerVisit	Gasto médio por visita (€)
	monetaryLastMonth	Gasto médio por visitas no último mês
	monetaryVariation	Varição nos gastos do cliente durante do PO

Tabela 7: Variáveis independentes usadas neste estudo (continuação).

	nrReturnedProducts	Número de produtos devolvidos durante o PO
	daysFromLastShop	Número de dias desde a última compra
	ipt	Intervalo médio entre compras consecutivas
	maxIPT	Máximo intervalo entre compras consecutivas
	nrPromotions	Nº de visitas em que foi usada uma promoção
	lastPromotion	Número de dias desde a última promoção
	lor	Duração total da relação (dias)
	rMonetary_category_[1..12]	Porcentagem dos gastos por categoria
	relative_mop_[1..6]	Proporção de pagamentos efectuados por método de pagamento
	rNrVisits_WeekDay_[1..7]	Proporção das visitas por dia da semana
	preferentialStore	Loja onde efetuou o maior nº de transações
Macroeconómicos	inflationRateVariation	Varição média dos preços nos 10 produtos mais comprados
	topProductsIvaVariation	Varição média do IVA nos 10 produtos mais comprados pelo cliente.

Descrição dos Cenários e Experiências Desenvolvidas

O conjunto de dados usado para a modelação foi criado artificialmente de forma a recriar, com as características habituais do mercado em questão, um *dataset* suficientemente capaz de suportar as experiências de previsão de *churn* que pretendíamos realizar. O atributo classe, por sua vez, foi definido ("*churn*" ou "*não churn*") usando cenários fictícios baseados num conjunto limitado de atributos que foram utilizados em outras pesquisas de modelos de previsão de *churn* no retalho. Refira-se que, no estudo sobre *churn* parcial no retalho que Buckinx e Van den Poel (2005) realizaram, estes concluem que as variáveis comportamentais são bastante mais importantes, têm mais influência, que as variáveis demográficas, na separação dos bons clientes daqueles que têm tendência de *churn* (parcial). No entanto, as últimas não podem ser descartadas. De entre as variáveis comportamentais destacavam-se as variáveis RFM no poder informativo. Baseados nesses critérios, os cenários de *churn* descritos neste trabalho foram baseados no trabalho desses dois autores.

Buckinx e Van den Poel (2005) defendem que quanto menor é o tempo desde que o cliente efetuou a última compra (*recency*) maior é a probabilidade de o cliente permanecer leal. Este

pode ser um importante indicador do estado ativo ou não do cliente em ambientes não contratuais (Reinartz e Kumar, 2002). A frequência de compras dos clientes pode ser preditiva do seu comportamento futuro (Schmittlein e Peterson, 1994) porque está positivamente relacionada com o uso futuro esperado do cliente (Lemon et al., 2002). Por seu lado Bhattacharya (1998) e Anderson e Weitz (1989) indicam que a duração da relação está positivamente relacionada com a estabilidade futura da relação. Também, Mittal e Kamakura (2001) mostram que, entre outros, o género, o número de crianças em casa, bem como a área de residência, são indicativos do comportamento dos clientes. Além disso, Vakratsas (1998) confirmou que o tamanho da família tem influência no comportamento dos clientes: pequenas famílias são mais propensas ao negócio. Então, é esperado que este tipo de famílias seja menos fiel para com o retalhista (Buckinx e Van den Poel, 2005).

A partir destas premissas, neste trabalho de dissertação foram criados quatro cenários de *churn* para clientes da empresa considerada. Assim, os registos de compra de clientes que revelem um dos seguintes padrões fazem com que esse cliente seja apontado como *churner*:

- Cenário 1: Frequência último mês / Frequência < 0,5.
- Cenário 2: Dias desde a última compra >20.
- Cenário 3: Número de visitas < 10 && Montante gasto por visita < 15.
- Cenário 4: Duração da relação < 1500 && Número de crianças em casa < 2 && Frequência último mês / Frequência < 0,7 && Género = "M".

O ruído nos dados afecta o comportamento dos sistemas de mineração de dados, tendo um maior impacto na classe mais rara (Weiss, 2004). Como é frequente a existência de ruído em bases de dados ligadas a negócios e a clientes (Xu et al., 2006; Bose e Mahapatra, 2001), após a criação do conjunto de dados, adicionámos 5% de ruído à variável classe, ou seja, foram alterados aleatoriamente 5% dos exemplos dessa mesma variável.

4.2 Modelação

Após o processamento dos dados seguiu-se a fase de modelação. É nesta fase que deve ser seleccionada a técnica de modelação de acordo com os requisitos levantados em fases anteriores. Aqui, são, também, testados alguns algoritmos de modelação alternativos relativos às diferentes técnicas de classificação e ainda algoritmos concorrentes para cada uma das técnicas consideradas, como por exemplo, no caso das árvores de decisão, onde estão disponíveis diversos algoritmos: uns nativos do Rapidminer e outros importados do software Weka. O processo de teste dos diferentes algoritmos é acompanhado pelo respectivo processo de optimização dos seus parâmetros. A complexidade do processo de optimização dos algoritmos varia significativamente para os diferentes algoritmos. O número de parâmetros e a influência de cada um deles no resultado final leva a grandes variações na dificuldade de optimização dos algoritmos. Alguns dos algoritmos requerem um processamento extra de parte dos dados, de modo a serem compatíveis com o seu procedimento. Também, é necessário definir uma metodologia de teste (secção 4.2.1) para avaliar a qualidade e validade dos modelos.

Densidade do Conjunto de Dados

Nesta aplicação seleccionámos uma amostra de 10000 exemplos, em que 8707 deles eram da classe negativa (não *churners*) e 1293 eram da classe positiva (*churners*) (Figura 30). A amostra de dados seleccionada foi relativamente pequena para facilitar a exploração e modelação do problema através dos vários algoritmos escolhidos para teste, já que alguns deles requeriam um elevado tempo de treino e de um processo de optimização relativamente complexo.

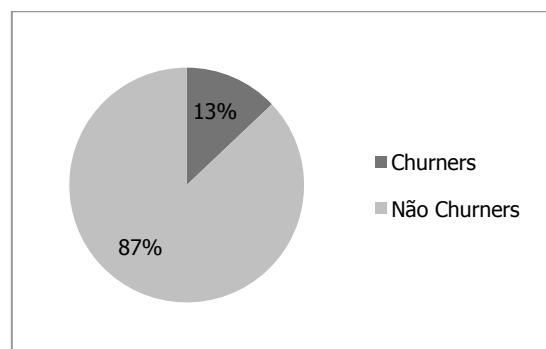


Figura 30: Distribuição das classes dos clientes

A Ferramenta Utilizada

A ferramenta usada para a realização da mineração de dados foi o RapidMiner 5.1³, simplesmente por ser uma das ferramentas mais utilizadas na área da mineração de dados (KDnuggets, 2011), *open source*, e por ser compatível com as principais plataformas e sistemas operativos existentes no mercado [1]. Além desses aspetos, o RapidMiner disponibiliza uma quantidade enorme de ferramentas e operadores especialmente orientados para o processamento e mineração de dados. Refira-se, também, um dos seus grandes trunfos: a facilidade de uso devido à sua interface gráfica intuitiva, que assenta no conceito da construção de um *workflow* baseado em blocos de processamento correspondentes às diversas tarefas do processo de tratamento de dados e de mineração (Figura 31). Este último aspecto facilita muito a compreensão do processo, bem como a sua parametrização ou configuração. Para além disso, existem operadores que podem agregar conjuntos de operadores internamente, como os dois operadores visíveis na Figura 31, à direita, o que permite um ambiente de trabalho simples e limpo.

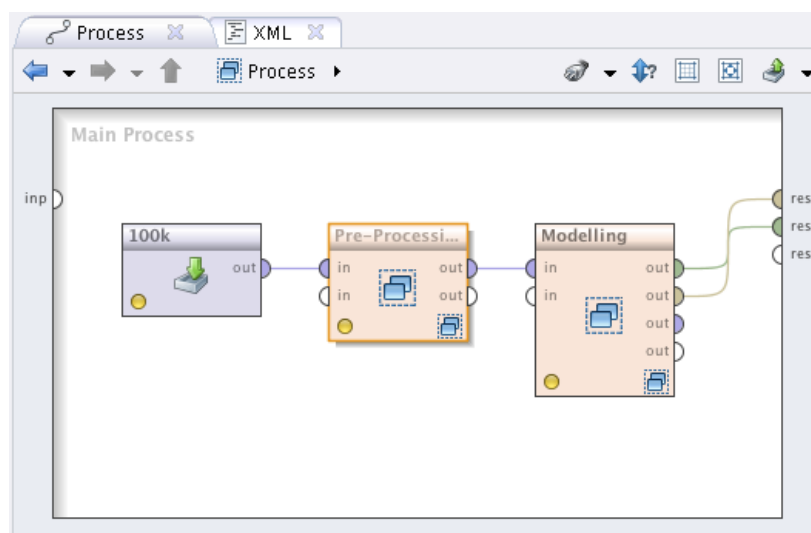


Figura 31: Um fluxo de trabalho em RapidMiner

³ <http://rapid-i.com/content/view/181/196/>

4.2.1 Metodologia de Testes

Nesta fase pretendeu-se estudar o comportamento dos algoritmos de classificação mais usados na previsão de *churn*. Foi feita a otimização dos parâmetros com base num conjunto de medidas de performance. Deste modo, todos os algoritmos são avaliados pelas medidas *accuracy*, *AUC*, *sensitivity*, *specificity*, *precision*, *f-measure* e *lift*, medidas estas já apresentadas anteriormente no capítulo 3. Contudo, como a classe positiva tem um peso maior devido à natureza da tarefa, deu-se importância acrescida às medidas *sensitivity*, *precision* e *f-measure*. O método escolhido para estimar as medidas de comparação dos diferentes modelos foi o *cross-validation*. É um método de estimação escolhido frequentemente na previsão de *churn* (Wei e Chiu, 2002; Coussement e Van den Poel, 2008a; Burez e Van den Poel, 2009). Mais precisamente foi usado o método *10-fold cross-validation*.

Para além da otimização dos algoritmos, pretendeu-se avaliar também a influência no comportamento dos algoritmos nos seguintes aspectos:

- método de filtragem de atributos (são escolhidos os n melhores atributos);
- número de atributos - 10, 20, 30, 40, 50, todos;
- densidade do conjunto de dados de treino (*undersampling* classe negativa) (classe positiva: 13%, 15%, 20%, 30%, 40%, 50%).

De acordo com Masand et al. (1999), uma metodologia possível para efetuar os testes é fazer experiências isoladas sobre cada um destes aspectos, escolhendo um em cada momento, como por exemplo o efeito da distribuição das classes no modelo, e efetuar vários testes comparando diferentes alternativas de forma a obter a melhor. As várias alternativas foram, depois, comparadas segundo as medidas escolhidas.

4.2.2 Construção dos Modelos

Entre as muitas técnicas de classificação existentes foram escolhidas e analisadas as árvores de decisão (secção 3.4.1), regressão logística (secção 3.4.2), redes neuronais (secção 3.4.3), *random forests* (secção 3.4.4) e SVM (secção 3.4.5). Foram testados também vários métodos de filtragem

de atributos para que se pudesse escolher o melhor subconjunto a usar no processo de previsão, como a estatística qui-quadrado (X^2), a correlação, o ganho de informação (GI) e a razão do ganho de informação (RGI) (do inglês *information gain ratio*). Usualmente, os melhores resultados dividem-se pelos métodos correlação e estatística qui-quadrado.

Árvore de Decisão

A primeira técnica analisada foi a das árvores de decisão. Apesar de existir uma implementação do algoritmo de indução própria no *Rapidminer*, numa fase de testes preliminares decidiu-se usar o algoritmo *J48* importado do *software Weka*, pois possui um reduzido número de parâmetros. Isto permite um processo de optimização mais simples, rápido e com bons resultados à partida. O algoritmo *J48* usado corresponde à implementação do algoritmo de construção de árvores de decisão C4.5 existente no *software Weka* (Quinlan, 1993).

O primeiro teste realizado foi relativo ao método de filtragem de atributos. Dos quatro métodos testados, a correlação e a estatística qui-quadrado, com resultados iguais, obtiveram os melhores resultados para todas as medidas de avaliação com a exceção da *sensitivity* (Tabela 8).

Tabela 8: Avaliações dos métodos de filtragem de atributos para a árvore de decisão

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X^2	95,29	0,834	96,08	743,05	78,44	66,28	99,60
Correlação	95,29	0,834	96,08	743,05	78,44	66,28	99,60
GI	93,80	0,827	82,02	634,32	73,55	66,67	97,83
RGI	92,95	0,821	93,49	723,06	64,20	48,88	99,49

Apesar do melhor resultado de *sensitivity* ter sido obtido com o ganho de informação, a correlação e o método qui-quadrado demonstraram resultados muito próximos. Para além disso, como nas restantes medidas de avaliação se obtiveram resultados superiores, a escolha limitou-se a estes dois métodos. A decisão sobre o método de filtragem a usar, entre a correlação e a estatística qui-quadrado, foi adiada para uma das seguintes etapas.

O teste que se seguiu foi relativo ao número de atributos de entrada usados para treino do modelo (N) (Tabela 9). Foram testados diferentes N para os dois métodos de filtragem de atributos escolhidos na fase anterior.

Tabela 9: Avaliações da árvore de decisão com diferentes quantidades de atributos

	N	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	10	95,29	0,834	96,08	743,05	78,44	66,28	99,60
	20	95,33	0,832	95,89	741,60	78,70	66,74	99,58
	30	95,11	0,825	95,48	738,40	77,54	65,27	99,54
	40	95,14	0,826	95,59	739,31	77,69	65,43	99,55
	50	95,15	0,829	95,50	738,56	77,76	65,58	99,54
	57	95,12	0,827	95,17	736,07	77,66	65,58	99,51
Correlação	10	95,29	0,834	96,08	743,05	78,44	66,28	99,60
	20	95,20	0,830	95,52	738,75	78,04	65,97	99,54
	30	95,12	0,827	95,28	736,86	77,64	65,51	99,52
	40	95,14	0,829	95,08	735,37	77,79	65,82	99,49
	50	95,05	0,824	94,53	731,10	77,39	65,51	99,44
	57	95,12	0,827	95,17	736,07	77,66	65,58	99,51

Com os melhores resultados distribuído pelas configurações de 10 e 20 atributos de entrada com o método de filtragem qui-quadrado e 10 atributos de entrada com o método de filtragem correlação, a próxima análise foi efectuada sobre estas 3 configurações base. Sobre estas, foram experimentadas diversas combinações de parâmetros, nomeadamente:

- C - Nível de confiança para *prunning* (0,05, 0,10, 0,20, 0,25).
- M - Número mínimo de instâncias por folha (2, 5).

Como se pode verificar na Tabela 10, os resultados usando 10 atributos de entrada foram iguais, independentemente do método de filtragem usado (X² ou Correlação).

Tabela 10: Resultados para diferentes valores de C e M

	C	M	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ² /Correlação (10 atributos)	0,05	2	95,28	0,836	96,18	743,82	78,37	66,13	99,61
	0,10	2	95,28	0,836	96,18	743,82	78,37	66,13	99,61
	0,20	2	95,31	0,836	96,09	743,12	78,56	66,43	99,60
	0,25	2	95,29	0,834	96,08	743,05	78,44	66,28	99,60
	0,05	5	95,28	0,836	96,18	743,82	78,37	66,13	99,61
	0,10	5	95,28	0,836	96,18	743,82	78,37	66,13	99,61
	0,20	5	95,28	0,836	96,18	743,82	78,37	66,13	99,61
	0,25	5	95,28	0,836	96,18	743,82	78,37	66,13	99,61

Tabela 10: Resultados para diferentes valores de C e M (continuação)

X ² (20 atributos)	0,05	2	95,32	0,838	96,09	743,15	78,61	66,51	99,60
	0,10	2	95,36	0,838	96,11	743,29	78,83	66,82	99,60
	0,20	2	95,34	0,834	96,00	742,42	78,74	66,74	99,59
	0,25	2	95,33	0,832	95,89	741,60	78,70	66,74	99,58
	0,05	5	95,33	0,838	96,20	743,98	78,65	66,51	99,61
	0,10	5	95,33	0,838	96,20	743,98	78,65	66,51	99,61
	0,20	5	95,35	0,835	96,00	742,46	78,80	66,82	99,59
	0,25	5	95,34	0,835	95,89	741,64	78,76	66,82	99,58

Com 10 atributos de entrada, apenas foram obtidas algumas configurações com o melhor resultado de *specificity*. E mesmo estas têm que dividir o melhor resultado com algumas configurações com 20 atributos de entrada. Os resultados mais importantes foram obtidos com a configuração que usa o método de filtragem qui-quadrado e com 20 atributos de entrada para treino do modelo. Com os melhores valores de *sensitivity* e *precision* a serem obtidos por diferentes configurações, optou-se por escolher como configuração final aquela com o melhor compromisso entre as 2 medidas, ou seja, a melhor *f-measure*. Decidiu-se, então, pela seguinte configuração final:

- Método de filtragem de atributos: Qui-quadrado.
- Número de atributos de entrada: 20.
- Número mínimo de instâncias por folha: 2.
- Nível de confiança para *prunning*: 0,10.

Regressão Logística

Após as árvores de decisão, o processo de análise evoluiu para a aplicação da regressão logística. Seguindo a metodologia definida no princípio, o primeiro teste a ser realizado foi relativo ao método de filtragem de atributos. Em todas as medidas de avaliação, com a exceção da medida AUC, a correlação obteve melhores resultados que os restantes métodos (Tabela 11). Devido a isto, os testes seguintes foram realizados utilizando a correlação como método de filtragem de atributos.

Tabela 11: Avaliações dos métodos de filtragem de atributos para a regressão logística

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	88,69	0,759	80,45	622,21	27,45	16,55	99,40
Correlação	88,82	0,759	83,02	642,06	28,24	17,01	99,48
GI	88,62	0,762	79,69	616,35	26,77	16,09	99,39
RGI	88,41	0,755	76,59	592,32	24,98	14,93	99,32

Seguidamente foi testada a influência do número de atributos de entrada na performance da regressão logística (Tabela 12). A técnica não se mostrou muito sensível com as diferentes quantidades sofrendo alterações pouco significativas em todas as medidas de avaliação. Com 10 atributos foram obtidos os melhores resultados nas medidas de avaliação *accuracy*, *precision*, *lift*, *f-measure* e *specificity*. O valor mais elevado da medida AUC foi obtido com 20 atributos e para a *sensitivity* o melhor resultado foi obtido com 40 atributos.

Tabela 12: Avaliações da regressão logística com diferentes quantidades de atributos

N	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
10	88,82	0,759	83,02	642,06	28,24	17,01	99,48
20	88,77	0,763	81,95	633,84	27,97	16,86	99,45
30	88,64	0,762	81,27	628,58	26,42	15,78	99,46
40	88,72	0,762	79,78	617,04	28,15	17,09	99,36
50	88,69	0,761	80,22	620,45	27,55	16,63	99,39
57	88,72	0,760	81,37	629,30	27,51	16,55	99,44

Embora o valor de *sensitivity* para 40 atributos seja superior, foi decidido escolher 10 atributos de entrada para os seguintes testes, pois este tem um valor superior na maioria das medidas de avaliação e principalmente um melhor compromisso entre *sensitivity* e *precision* (*f-measure*). O parâmetro de optimização testado relativo ao método foi o número de iterações para optimização da função de regressão (I) (Tabela 13). A partir do valor de iterações igual a 80 os resultados para as diferentes medidas de avaliação não mais se alteraram.

Tabela 13: Avaliações da técnica Regressão Logística variando o parâmetros I

I	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
0	88,82	0,759	83,02	642,06	28,24	17,01	99,48
10	88,66	0,757	85,02	657,56	25,39	14,93	99,61
20	88,74	0,758	82,49	637,98	27,35	16,40	99,48
30	88,83	0,759	82,84	640,65	28,44	17,17	99,47
40	88,81	0,759	81,29	628,73	28,77	17,48	99,40
50	88,82	0,760	81,14	627,52	28,97	17,63	99,39
60	88,80	0,760	80,78	624,77	28,84	17,56	99,38
70	88,80	0,760	80,78	624,77	28,84	17,56	99,38
80	88,79	0,760	80,71	624,24	28,73	17,48	99,38
90	88,79	0,760	80,71	624,24	28,73	17,48	99,38
100	88,79	0,760	80,71	624,24	28,73	17,48	99,38
200	88,79	0,760	80,71	624,24	28,73	17,48	99,38

Apesar do melhor resultado de *precision* ser para $I = 10$, optou-se por escolher para configuração final de $I = 50$ por esta exibir os melhores valores de *f-measure* e *sensitivity*. Deste modo, no final da otimização da técnica chegou-se à seguinte configuração:

- Método de filtragem de atributos: Correlação.
- Número de atributos de entrada: 10.
- Iterações para otimização da função de regressão : 50.

Rede Neuronal

As redes neuronais foram a terceira técnica utilizada. O operador utilizado constrói o modelo por meio de uma rede neuronal *feed-forward* treinada pelo algoritmo de *backpropagation*. A função de ativação usada neste operador é a função *sigmoid*. Assim, seguindo a metodologia que tem vindo a ser aplicada, o primeiro aspecto a ser testado foi o método de filtragem de atributos. Cada um dos métodos obteve o melhor resultado em pelo menos uma das medidas da avaliação. Para a *precision lift* os resultados máximos foram obtidos usando o método da correlação para a *accuracy*, *f-measure* e *sensitivity* os melhores resultados foram obtidos pelo método do qui-quadrado. Usando o método RGI obteve-se o melhor resultado de AUC e finalmente a *specificity* máxima foi obtida pelo método de filtragem GI (Tabela 14).

Tabela 14: Avaliações dos métodos de filtragem de atributos para a rede neuronal

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
χ^2	94,10	0,829	90,64	700,97	72,66	60,63	99,07
Correlação	94,03	0,827	90,85	702,59	72,17	59,86	99,10
GI	92,88	0,830	81,96	633,86	67,67	57,62	98,12
RGI	92,28	0,821	89,17	689,66	60,57	45,86	99,17

Como os melhores resultados de *precision*, *f-measure* e *sensitivity* são muito aproximados e dividiam-se pelos métodos de correlação e qui-quadrado, a decisão sobre a escolha do método de filtragem foi remetida para os seguintes testes.

Um dos principais factores para obtenção de bons resultados com a rede neuronal é a sua estrutura, ou seja, os atributos escolhidos para alimentar a rede e o número de unidades escondidas usado. O factor analisado de seguida foi a influência do número de atributos de entrada (Tabela 15).

Tabela 15: Avaliações da rede neuronal com diferentes quantidades de atributos

	N	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
χ^2	10	94,10	0,829	90,64	700,97	72,66	60,63	99,07
	20	93,29	0,816	84,71	655,14	69,35	58,70	98,43
	30	92,09	0,802	77,70	600,96	64,03	54,45	97,68
	40	90,33	0,793	67,60	522,84	56,42	48,41	96,55
	50	90,36	0,793	67,48	521,90	56,85	49,11	96,49
	57	89,44	0,791	62,38	482,46	53,07	46,17	95,87
Correlação	10	94,03	0,827	90,85	702,59	72,17	59,86	99,10
	20	93,23	0,822	84,38	652,55	69,07	58,47	98,39
	30	91,77	0,797	77,71	601,02	61,56	50,97	97,83
	40	90,71	0,799	69,08	534,24	58,66	50,97	96,61
	50	90,02	0,791	65,71	508,18	55,29	47,72	96,30
	57	89,60	0,790	63,03	487,45	54,06	47,33	95,88

Os melhores resultados foram obtidos recorrendo a apenas 10 atributos de entrada, no entanto, dividem-se entre a configuração que usa o método qui-quadrado, para filtragem de atributos, e a configuração que usa a correlação. Com isto, prosseguiu-se usando duas configurações, ambas com 10 atributos de entrada e cada uma com o respectivo método de filtragem. A próxima etapa

visa avaliar a influência do número de unidades escondidas (HU) na performance da técnica (Tabela 16).

Tabela 16: Avaliações medidas com a alteração do número de unidades escondidas (HU)

	HU	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	1	90,15	0,680	80,68	623,95	45,13	31,32	98,89
	2	93,05	0,784	88,73	686,24	66,34	52,98	99,00
	3	93,50	0,830	86,74	670,87	70,02	58,70	98,67
	4	93,65	0,818	88,17	681,88	70,53	58,78	98,83
	5	93,85	0,817	90,07	696,60	71,25	58,93	99,04
	6	94,12	0,824	91,42	707,05	72,57	60,17	99,16
	7	94,10	0,829	90,64	700,97	72,66	60,63	99,07
	8	94,34	0,827	91,64	708,72	73,87	61,87	99,16
	9	94,30	0,819	91,22	705,49	73,73	61,87	99,12
	10	94,16	0,819	90,61	700,75	73,04	61,18	99,06
Correlação	1	90,25	0,677	81,80	632,64	45,62	31,63	98,95
	2	93,14	0,783	88,86	687,24	66,92	53,67	99,00
	3	93,51	0,830	86,67	670,33	70,11	58,86	98,66
	4	93,72	0,829	87,83	679,25	71,09	59,71	98,77
	5	93,94	0,826	89,26	690,31	72,05	60,40	98,92
	6	93,96	0,822	88,75	686,40	72,32	61,02	98,85
	7	94,03	0,827	90,85	702,59	72,17	59,86	99,10
	8	94,16	0,820	90,15	697,19	73,16	61,56	99,00
	9	94,00	0,827	90,62	700,86	72,04	59,78	99,08
	10	94,15	0,829	90,97	703,57	72,88	60,79	99,10

Com a exceção da medida AUC, os melhores resultados foram obtidos quando a estrutura da rede era constituída por 8 unidades escondidas, usando como método de filtragem a estatística qui-quadrado. Deste modo, o processo continuou a usar estas configurações com que foram obtidos os melhores resultados.

Após ser definida a estrutura da rede, a ultima etapa do processo de otimização da rede neuronal visou analisar a influência dos parâmetros *momentum* (M) e *learning rate* (LR) nos seus resultados, e para isso foram testadas várias combinações desses parâmetros (Tabela 17).

Tabela 17: Avaliações da técnica Rede Neuronal variando os parâmetros M e LR

M	LR	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
0,2	0,3	94,34	0,827	91,64	708,72	73,87	61,87	99,16
0,2	0,2	94,23	0,829	90,96	703,49	73,37	61,48	99,09
0,2	0,1	94,24	0,827	91,16	705,02	73,38	61,41	99,12
0,2	0,05	94,29	0,827	91,88	710,59	73,50	61,25	99,20
0,1	0,3	94,16	0,824	91,27	705,87	72,86	60,63	99,14
0,1	0,2	94,31	0,827	91,23	705,57	73,79	61,95	99,12
0,1	0,1	94,23	0,832	91,15	704,95	73,32	61,33	99,12
0,1	0,05	94,34	0,828	92,02	711,70	73,77	61,56	99,21

No final, como houve diferentes configurações com os melhores resultados para as medias *sensitivity*, *precision* e *f-measure*, optou-se pela configuração com o melhor resultado de *f_measure*. A configuração final da rede neuronal:

- Método de filtragem de atributos: X^2 .
- Número de atributos de entrada: 10.
- *hidden units* (unidades da camada escondidas): 8.
- *learning rate*: 0.3.
- *momentum*: 0.2.

Random Forest

A quarta técnica experimentada foi a *random forest*. A otimização desta técnica foi realizada tendo em consideração quatro factores: o método de filtragem de atributos, os atributos escolhidos para treino (N), o número de atributos escolhidos aleatoriamente em cada nó (K) e o número de árvores (I) a construir. No primeiro teste, os melhores resultados dividiram-se entre os métodos de filtragem correlação, qui-quadrado e ganho de informação (Tabela 18).

Tabela 18: Avaliações dos métodos de filtragem de atributos para a *random forest*

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X^2	95,06	0,828	95,45	738,20	77,26	64,89	99,54
Correlação	95,07	0,829	95,15	735,86	77,37	65,20	99,51
GI	93,38	0,830	85,97	664,93	69,49	58,31	98,59
RGI	92,52	0,808	85,62	662,19	63,65	50,66	98,74

Apesar de ter obtido o melhor resultado para a medida AUC, o método de ganho de informação obteve resultados significativamente inferiores nas medidas *sensitivity* e *precision* e, deste modo, não entrou na consideração para prosseguir com a análise. Como tem sido o método aplicado anteriormente, prossegue-se, adiando a decisão entre os métodos correlação e qui-quadrado para as seguintes etapas, ou seja, o teste que se realizou de seguida consistiu em verificar a influência do número de atributos de entrada no treino do modelo para ambos os métodos de filtragem (Tabela 19).

Tabela 19: Avaliações da *random forest* com diferentes quantidades de atributos

	N	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	10	95,06	0,828	95,45	738,20	77,26	64,89	99,54
	20	95,05	0,828	96,07	743,03	77,07	64,35	99,61
	30	94,61	0,828	95,31	737,14	74,64	61,33	99,55
	40	93,75	0,831	95,63	739,59	69,14	54,14	99,63
	50	92,87	0,828	95,45	738,24	63,08	47,1	99,67
	57	92,49	0,819	94,28	729,17	60,58	44,62	99,60
Correlação	10	95,07	0,829	95,15	735,86	77,37	65,2	99,51
	20	95,07	0,835	95,87	741,47	77,23	64,66	99,59
	30	94,6	0,826	95,53	738,79	74,53	61,1	99,58
	40	94,05	0,836	95,32	737,24	71,16	56,77	99,59
	50	93,65	0,83	94,82	733,36	68,67	53,83	99,56
	57	92,49	0,819	94,28	729,17	60,58	44,62	99,60

Na etapa de teste do número de atributos de entrada os melhores resultados ficaram distribuídos por diversas configurações. O melhor valor de *accuracy* foi obtido usando o método de filtragem correlação com 10 e 20 atributos. O resultado mais elevado de AUC foi obtido com 40 atributos filtrados com o método de correlação. Com a *precision* e *lift* mais altos veio o método de filtragem qui-quadrado com 20 atributos. Os melhores resultados de *f-measure* e *sensitivity* foram obtidos com 10 atributos de entrada e correlação. Finalmente, com a melhor *specificity*, tivemos a configuração com 50 atributos de entrada e o método de filtragem qui-quadrado. Para continuar com a optimização foram selecionadas as configurações com os melhores valores de *accuracy*, *precision*, *lift*, *f-measure* e *sensitivity*:

- 20 atributos de entrada com o método de filtragem qui-quadrado.
- 10 atributos de entrada com o método correlação.

- 20 atributos de entrada com o método correlação.

Para cada uma das configurações descritas em cima, foi testado o número de atributos escolhidos aleatoriamente em cada nó (K), de 1 até N (número de atributos de entrada). Os melhores resultados foram obtidos usando 20 atributos de entrada, mas divididos entre os dois métodos de filtragem (Tabela 20).

Tabela 20: Avaliações da técnica *Random Forest* variando o parâmetros K

	K	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	15	95,28	0,837	96,07	743,01	78,39	66,20	99,60
	16	95,22	0,842	95,84	741,21	78,09	65,89	99,58
Correlação	1	90,95	0,819	93,69	724,62	47,90	32,17	99,68
	12	95,20	0,833	96,25	744,33	77,90	65,43	99,62

Seguindo a recomendação de Breiman (2001a), em escolher um número elevado para o número de árvores a construir (I), foram testados diferentes valores para este parâmetro, para cada uma das melhores configurações da etapa anterior, com a exceção de K = 1 que teve avaliações significativamente inferiores (Tabela 21).

Tabela 21: Avaliações da técnica *Random Forest* variando os parâmetros K e I

	K	I	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	15	10	95,28	0,837	96,07	743,01	78,39	66,20	99,60
	15	50	95,36	0,833	96,21	744,08	78,81	66,74	99,61
	15	100	95,37	0,830	96,32	744,91	78,85	66,74	99,62
	15	500	95,37	0,830	96,32	744,91	78,85	66,74	99,62
	16	10	95,22	0,842	95,84	741,21	78,09	65,89	99,58
	16	50	95,35	0,834	96,21	744,05	78,76	66,67	99,61
	16	100	95,35	0,832	96,31	744,85	78,74	66,59	99,62
	16	500	95,36	0,829	96,31	744,88	78,79	66,67	99,62
Correlação	12	10	95,20	0,833	96,25	744,33	77,90	65,43	99,62
	12	50	95,29	0,832	95,97	742,25	78,46	66,36	99,59
	12	100	95,30	0,829	95,98	742,29	78,52	66,51	99,59
	12	500	95,31	0,828	95,98	742,32	78,57	66,51	99,59

Com a exceção da medida AUC, os melhores resultados foram obtidos para $K = 15$. Relativamente ao número de árvores a construir, os melhores resultados foram obtidos para 100 e 500 árvores. Como os valores das avaliações de 500 não superavam os de 100 e apenas aumentavam o tempo de treino do modelo, decidiu-se usar $I = 100$. Assim, a configuração final foi definida com os seguintes factores:

- Método de Filtragem: X^2 .
- Número de atributos de entrada: 20.
- Número de atributos escolhidos aleatoriamente em cada nó: 15.
- Número de árvores a construir: 100.

SVM

A última técnica utilizada foi a SVM. Ao contrário das restantes técnicas, o primeiro teste recaiu sobre a função *kernel* a utilizar, já que com a configuração original os casos eram todos classificados como negativos (não *churners*). É possível configurar o operador com diferentes funções *kernel* e, por essa razão, foi testado com as funções presentes na Tabela 22. O procedimento foi realizado utilizando 10 atributos de entrada e com o método qui-quadrado como método de filtragem.

Tabela 22: Avaliações da técnica SVM com diferentes funções *kernel*

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
Radial	82,29	0,811	88,54	684,78	32,26	19,72	99,62
Anova	93,00	0,832	91,94	711,04	65,00	50,27	99,35
Polynomial	89,98	0,778	95,47	738,39	30,21	17,94	99,87
Neural	75,03	0,486	14,50	112,18	16,46	19,03	83,35
Epachnenikov	87,08	0,686	57,14	441,94	0,62	0,31	99,97

Com a função *anova* foram obtidos os melhores resultados relativos às medidas *accuracy*, *AUC*, *f-measure* e *sensitivity*. Para as medidas *precision* e *lift* os melhores resultados foram obtidos com a função *polynomial*. Finalmente, a função *epachnenikov* obteve a melhor *specificity*. Como os valores de *f-measure* e *sensitivity* obtidos com a função *kernel anova* são muito superior às restantes, e a *precision* está próxima do valor mais elevado obtido, a escolha para prosseguir com os restantes testes recaiu sobre esta.

O teste que se seguiu foi realizado com o método de filtragem de atributos. O melhor resultado de *accuracy* foi obtido pelos métodos qui-quadrado e correlação. Para as métricas *AUC*, *f-measure* e *sensitivity* o método de correlação levou vantagem e para a *precision*, *lift* e *specificity* os melhores resultados foram obtidos pelo método de filtragem RGI (Tabela 23).

Tabela 23: Avaliações dos métodos de filtragem de atributos para a SVM

	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
X ²	93,00	0,832	91,94	711,04	65,00	50,27	99,35
Correlação	93,00	0,835	91,35	706,52	65,17	50,66	99,29
GI	92,54	0,831	91,63	708,65	61,74	46,56	99,37
RGI	92,34	0,779	94,59	731,52	59,34	43,23	99,63

Apesar do método RGI ter o melhor resultado de *precision*, os valores de *sensitivity* e *f-measure* ficaram bastante aquém dos valores obtidos com o método correlação e por esta razão a análise prossegue usando a correlação como método de filtragem de atributos.

O próximo factor a ser testado foi o número de atributos que contribuem para o treino do modelo (Tabela 24). O aumento do número de atributos de entrada influenciava negativamente os resultados obtidos na *accuracy*, *f-measure* e *sensitivity*. Para as outras medidas o aumento do número de atributos não resultou numa variação muito significativa. Aliando, este facto, às implicações no tempo de treino do modelo (aproximadamente 4,5 horas para N = 40), não foram testados as configurações com o número de atributos superior. Seguiu-se então com a configuração com 10 atributos.

Tabela 24: Avaliações da SVM com diferentes quantidades de atributos

N	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
10	93,00	0,835	91,35	706,52	65,17	50,66	99,29
20	92,26	0,834	91,65	708,84	59,60	44,16	99,40
30	91,97	0,823	92,39	714,52	57,08	41,30	99,49
40	91,77	0,825	91,67	708,95	55,68	39,98	99,46

A SVM foi a técnica em que a optimização dos parâmetros foi mais demorada. Este facto deve-se ao tempo de treino muito elevado. A função *anova* possui dois parâmetros - γ e d - que controlam

a sua forma (Ivanciuc, 2002). Devido a isto a optimização dos parâmetros γ e d foi efectuada utilizando o método de validação cruzada com 5 iterações, ao contrário das 10 iterações que têm sido usadas (Tabela 25).

Tabela 25: Resultados com a variação dos parâmetros γ e d

γ	Degree	Accuracy	AUC	Precision	Lift	F-measure	Sensitivity	Specificity
8	2	94,45	0,831	93,93	726,44	73,98	61,02	99,41
8	1	91,87	0,739	94,44	730,43	55,65	39,44	99,66
8	3	94,29	0,832	94,35	729,69	72,9	59,4	99,57
4	1	91,89	0,741	94,8	733,14	55,71	39,44	99,68
4	2	94,38	0,831	93,77	725,23	73,59	60,56	99,4
4	3	94,52	0,833	94,93	734,21	74,18	60,87	99,52
2	1	91,89	0,717	94,8	733,14	55,71	39,44	99,68
2	2	93,89	0,831	93,49	723,09	70,58	56,69	99,41
2	3	94,13	0,833	94,12	727,96	71,95	58,24	99,46
1	1	91,81	0,746	94,05	727,39	55,27	39,13	99,63
1	2	92,85	0,832	91,05	704,19	64,2	49,57	99,28
1	3	93,55	0,835	92,41	714,68	68,64	54,6	99,33
0,5	1	91,4	0,794	93,74	724,96	51,9	35,89	99,64
0,5	2	92,12	0,831	90,92	703,2	58,74	43,39	99,36
0,5	3	92,75	0,833	91,04	704,1	63,48	48,72	99,29

De modo geral, os melhores resultados foram obtidos para $\gamma = 4$ e $d = 3$. Esta configuração não obteve o melhor resultado relativamente à métrica *sensitivity* mas, no entanto, esteve muito próxima desse valor. Com os melhores valores para a *precision* e *f-measure* a escolha recaiu sobre este conjunto de parâmetros. Desta forma a configuração final foi a seguinte:

- Função *kernel*: *anova*.
- Método de filtragem de atributos: correlação.
- Número de atributos de entrada: 10.
- γ : 4.
- d : 3

4.3 Avaliação dos Resultados

4.3.1 Resultados da Árvore de Decisão

A Tabela 26 diz respeito à matriz de confusão obtida com a aplicação do modelo de árvore de decisão, em que são discriminados os valores de acertos e erros para ambas as classes. Exemplificando, da tabela consegue-se extrair que 864 *churners* foram corretamente classificados e que apenas 35 casos foram classificados como *churners*, quando na realidade eram não *churners* (taxa de acerto de 96,11%). Relativamente à previsão de eventos negativos (não *churners*), 8672 foram corretamente classificados e 429 eram, na realidade, eventos positivos de *churn* (taxa de acerto de 95,29%).

Tabela 26: Matriz de confusão da Árvore de Decisão

		Previsto	
		<i>Churner</i>	Não <i>Churner</i>
Real	<i>Churner</i>	864	429
	Não <i>Churner</i>	35	8672

Como se pode verificar a partir da Tabela 27, os resultados da árvore de decisão são muito interessantes. Conseguiu-se obter um resultado excelente na precisão da classificação dos eventos de *churn*. Do total de eventos previstos como positivos, 96,17% foram corretamente classificados, o que leva a um nível reduzido de desperdício de fundos relativos à campanha de retenção de clientes que eventualmente não estariam em risco de ser perdidos. Relativamente à sensibilidade do algoritmo, este consegue detectar 66,82% do total de eventos de *churn*.

Tabela 27: Avaliação da técnica Árvore de Decisão

Medida	Resultado
AUC	0,838 +/- 0,024
Accuracy	95,36% +/- 0,49%
Sensitivity	66,82% +/- 3,94%
Specificity	99,60% +/- 0,21%
F-measure	78,77% +/- 2,71%
Precision	96,17% +/- 1,90%
Lift	741,70% +/- 16,18%

Devido à natureza do conjunto de dados, que é caracterizado pelo desbalanceamento entre as duas classes, testou-se o efeito da alteração da densidade dos casos de *churn*, do conjunto de dados de treino, no resultado final. A Figura 32 mostra o resultado para as diferentes densidades de casos de *churn* segundo as diversas medidas. É bem visível que o aumento de casos de *churn* no conjunto de dados apenas tem um impacto positivo na sensibilidade do modelo. Em todas as outras medidas, o aumento de casos de *churn* revelou-se prejudicial para o resultado final, resultando mesmo numa elevada degradação dos resultados a partir de uma densidade de 30% de casos de *churn*. Concluindo, o aumento da densidade de casos de *churn* resulta na identificação de um maior número de casos efetivos de *churn*, mas também num aumento de falsos positivos.

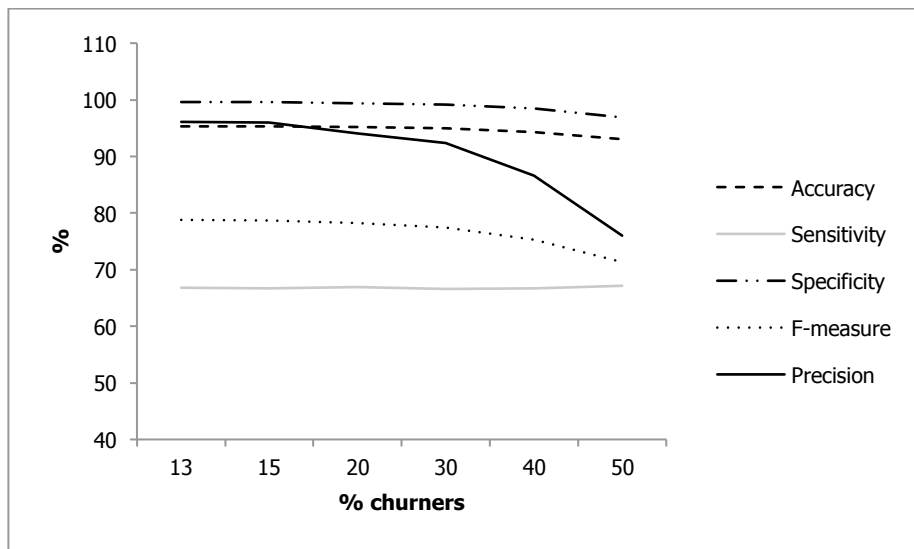


Figura 32: Resultados da Árvore de Decisão com a variação da densidade de *churners*

Relativamente ao nível de facilidade de interpretação são perfeitamente visíveis os resultados superiores demonstrados pela árvore de decisão (Figura 33). Das regras definidas na definição dos casos de *churn*, podem-se extrair facilmente grande parte delas da estrutura da árvore. Como se pode verificar, as regras apresentam uma estrutura hierárquica (da esquerda para a direita) sendo as regras de topo (esquerda) as mais importantes. Para auxiliar a comparação dos resultados com os cenários, traduz-se os mesmos para a linguagem dos atributos:

- Cenário 1: $\text{frequencyLastMonthR} < 0,5$.
- Cenário 2: $\text{daysFromLastShop} > 20$.

- Cenário 3: $NrVisits < 10 \ \&\& \ monetaryPerVisit < 15$.
- Cenário 4: $lor < 1500 \ \&\& \ nrChildrenAtHome < 2 \ \&\& \ frequencyLastMonthR < 0,7 \ \&\& \ gender = "M"$.

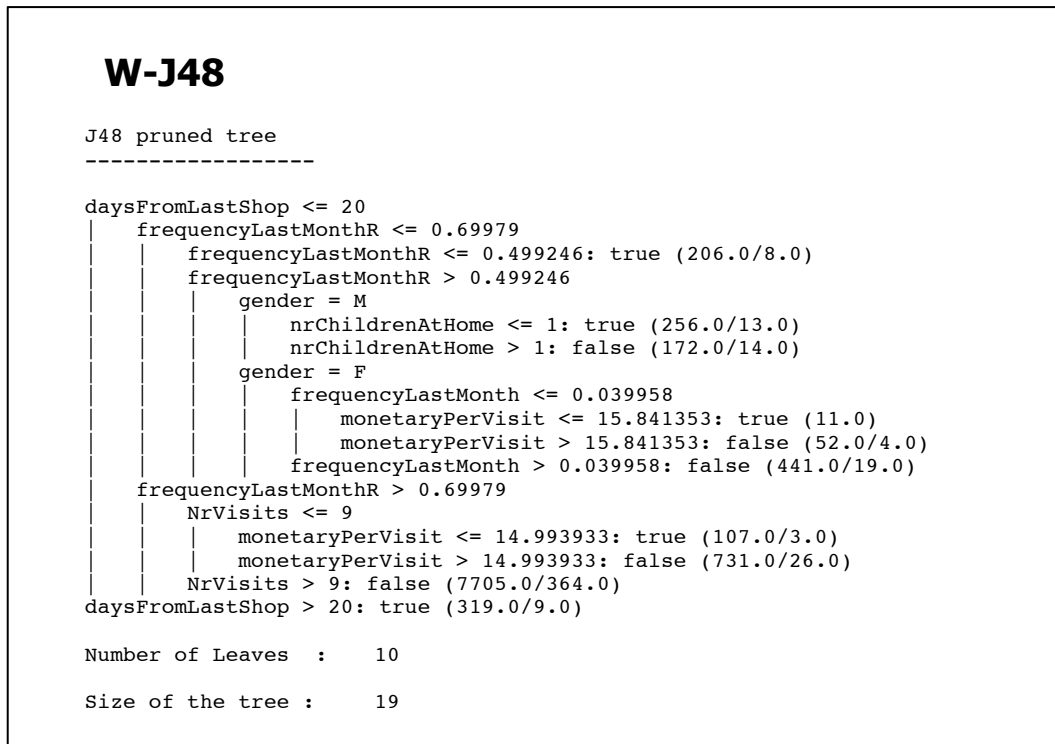


Figura 33: *Output* gerado pela técnica de árvores de decisão

Para além das regras, que se podem extrair do modelo, também se consegue verificar a sua taxa de acerto. Por exemplo, para a regra "daysFromLastShop > 20: true" é possível verificar que 319 casos foram corretamente classificados e apenas 9 deles estavam errados, o que nos dá o nível de confiança para cada uma das regras. Exemplificando com uma regra mais complexa teríamos "daysFromLastShop <= 20 && frequencyLastMonthR > 0.69979 && NrVisits <= 9 && monetaryPerVisit <= 14.993933: true" na qual se pode verificar que 107 casos foram corretamente classificados e apenas 3 estavam errados.

4.3.2 Resultados da Regressão Logística

A matriz de confusão da regressão logística (Tabela 28) mostra que 228 previsões de eventos de *churn* estavam corretas e que 53 delas estavam incorretas (taxa de acerto de 81,14%). Dos casos classificados como não *churners*, 8655 foram corretamente classificados e 1065 estavam errados (taxa de acerto de 89,04%).

Tabela 28: Matriz de confusão da Regressão Logística

		Previsto	
		<i>Churner</i>	Não <i>Churner</i>
Real	<i>Churner</i>	228	1065
	Não <i>Churner</i>	53	8654

A regressão logística revelou-se o método com os piores resultados. Em termos de *sensitivity* apenas foram previstos 17,64% do total de eventos de *churn*. Já o nível de *precision* foi de 80,98%, resultado bem inferior aos restantes métodos. Com a exceção da *specificity* o algoritmo exibiu resultados significativamente mais pobres que os restantes métodos (Tabela 29).

Tabela 29: Avaliação da técnica Regressão Logística

Medida	Resultado
AUC	0,760 +/- 0,019
Accuracy	88,82% +/- 0,51%
Sensitivity	17,64% +/- 3,91%
Specificity	99,39% +/- 0,23%
F-measure	28,97% +/- 17,63%
Precision	80,98% +/- 6,16%
Lift	626,35% +/- 48,06%

A variação da densidade dos eventos de *churn* (Figura 34) revela a mesma tendência que a árvore de decisão, ou seja, a sensibilidade a aumentar e a precisão a diminuir. No entanto, até a uma densidade de 30% de *churners* esta tendência é mais acentuada em ambos os sentidos. E para além disso, é importante referir que neste caso a *f-measure* revela uma tendência de subida até à taxa de 30% de *churners*.

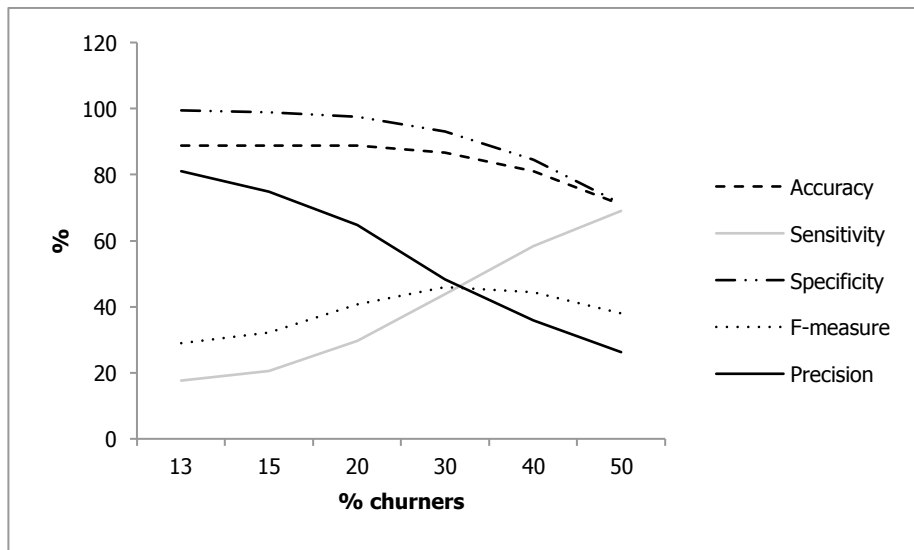


Figura 34: Resultados da regressão logística com a variação da densidade de *churners*

Relativamente à compreensibilidade do modelo de regressão logística, o resultado do algoritmo é dado sob a forma de uma equação matemática (Figura 35). Porém, não são facilmente perceptíveis padrões que expliquem o fenómeno. Os atributos apresentados são aqueles escolhidos pelo método de filtragem. Não se aproxima, portanto, do tipo de interpretabilidade disponibilizado pela árvore de decisão.

```

W-SimpleLogistic

SimpleLogistic:

Class 0 :
-0.13 +
[gender] * 0.24 +
[nrChildrenAtHome] * 0.09 +
[frequencyLastMonth] * -0.49 +
[frequencyLastMonthR] * 1.78 +
[monetaryPerVisit] * 0 +
[daysFromLastShop] * -0.07 +
[ipt] * -0.01 +
[maxIPT] * -0.01 +
[relative_mop1] * 0.51 +
[relative_mop6] * 0.36

Class 1 :
0.13 +
[gender] * -0.24 +
[nrChildrenAtHome] * -0.09 +
[frequencyLastMonth] * 0.49 +
[frequencyLastMonthR] * -1.78 +
[monetaryPerVisit] * 0 +
[daysFromLastShop] * 0.07 +
[ipt] * 0.01 +
[maxIPT] * 0.01 +
[relative_mop1] * -0.51 +
[relative_mop6] * -0.36
    
```

Figura 35: Output gerado pela técnica Regressão Logística

4.3.3 Resultados da Rede Neuronal

Com a rede neuronal foram obtidas uma taxa de 91,53% de acerto para a classe positiva (800 casos corretamente classificados e 74 erros) e uma taxa de 94,60% de acerto para a classe negativa - 8633 previsões corretas e 493 previsões incorretas (Tabela 30).

Tabela 30: Matriz de confusão da Rede Neuronal

		Previsto	
		Churner	Não Churner
Real	Churner	800	493
	Não Churner	74	8633

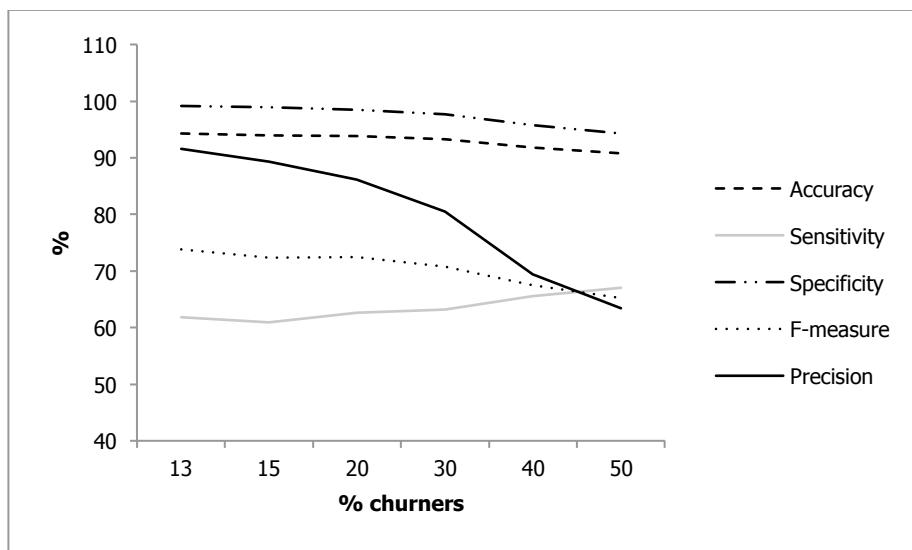
Os resultados da rede neuronal são bastante interessantes. No entanto, não estão ao nível dos

melhores resultados obtidos pela árvore de decisão e pela *random forest*, situando-se numa segunda linha a par com os resultados da técnica SVM (Tabela 31).

Tabela 31: Avaliação da técnica Rede Neuronal

Medida	Resultado
AUC	0,823 +/- 0,030
Accuracy	94,33% +/- 0,55%
Sensitivity	61,87% +/- 3,66%
Specificity	99,15% +/- 0,27%
F-measure	73,78% +/- 2,98%
Precision	91,56% +/- 2,49%
Lift	708,08% +/- 18,72%

A alteração da distribuição das classes revela a mesma tendência verificada com as restantes técnicas, ou seja, a acompanhar o aumento da densidade de *churners* regista-se também o aumento da *sensitivity* e a diminuição da *precision* (Figura 36).

Figura 36: Resultados da Rede Neuronal com a variação da densidade de *churners*

O *output* que resulta do treino da rede neuronal (Figura 37), exhibe a estrutura final da rede, permitindo consultar os pesos associados às ligações. Para além disso, é possível reconhecer

visualmente os pesos com maiores valores, já que estes estão representados por uma ligação com uma cor mais forte.

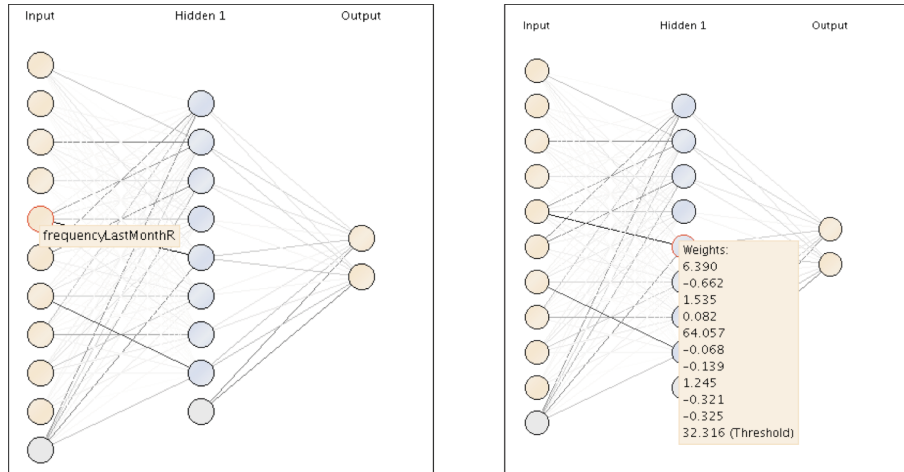


Figura 37: *Output* gerado pela técnica Rede Neuronal

4.3.4 Resultados da *Random Forest*

A matriz de confusão da *random forest* (Tabela 32) revela que dos 896 dos casos previstos como eventos de *churn* apenas 33 estavam errados. Para a classe negativa, foram classificados acertadamente 8674 casos e apenas 430 foram mal classificados.

Tabela 32: Matriz de confusão da *Random Forest*

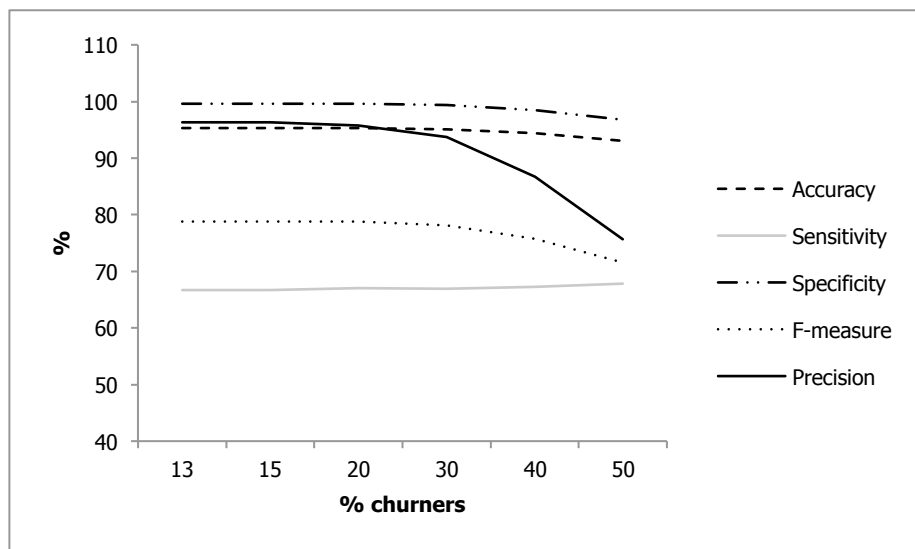
		Previsto	
		<i>Churner</i>	Não <i>Churner</i>
Real	<i>Churner</i>	863	430
	Não <i>Churner</i>	33	8674

A *random forest* acompanha a árvore de decisão na obtenção dos melhores resultados (Tabela 33). Aliado a isso, demonstrou ser uma técnica muito robusta pois relevou sempre resultados muito elevados, durante todo o processo de otimização e de experimentação.

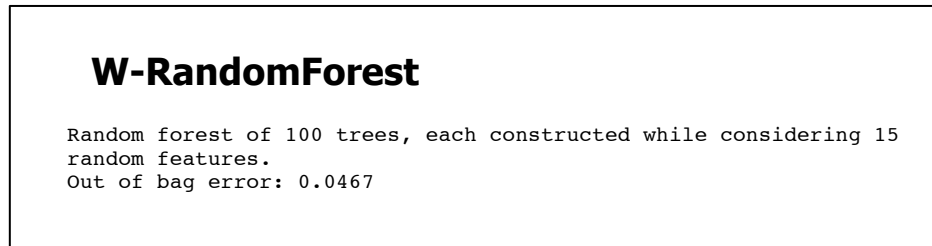
Tabela 33: Avaliação da técnica *Random Forest*

Medida	Resultado
AUC	0,830 +/- 0,023
Accuracy	95,37% +/- 0,48%
Sensitivity	66,75% +/- 3,89%
Specificity	99,62% +/- 0,18%
F_measure	78,79% +/- 2,69%
Precision	96,37% +/- 1,63%
Lift	745,31% +/- 13,30%

O *undersampling* da classe respectiva aos eventos de *churn* negativos indica a robustez deste algoritmo, ou seja, a baixa variação dos resultados relativamente a conjuntos de dados diferentes (Figura 38). Até uma taxa de eventos de *churn* de 30% a variação dos resultados é pouco significativa. A partir dos 30% a precisão sofre já um impacto negativo acrescido, o que também é proporcionado pela crescente escassez dos dados.

Figura 38: Resultados da *random forest* com a variação da densidade de *churners*

Apesar dos bons resultados na tarefa de previsão, o algoritmo *random forest* também revela um aspecto negativo, a sua compreensibilidade. Os resultados da *random forest* não permitem extrair qualquer informação relativa às causas, ou atributos, que revelam as tendências de *churning* (Figura 39).

Figura 39: *Output* gerado pela *random forest*

4.4 Resultados da SVM

A técnica SVM determinou 833 casos de *churn* de entre os quais 795 foram corretamente classificados e os restantes 38 eram na realidade não *churners*. A taxa de acerto para a classe positiva é portanto de 95,44%. Para a classe negativa foi obtida uma taxa de acerto de 94,57% resultante dos 8669 corretamente classificados e dos 498 erros de classificação como “não *churner*” (Tabela 34).

Tabela 34: Matriz de confusão da SVM

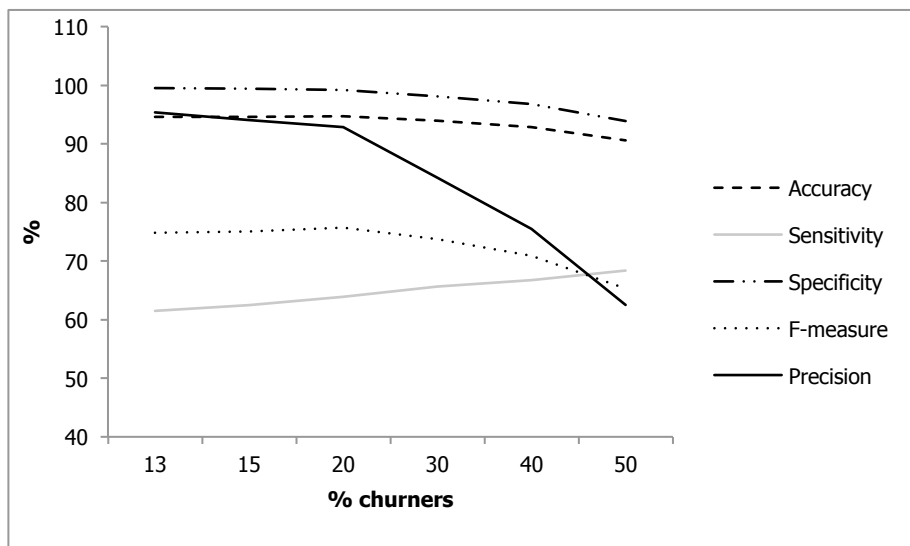
		Previsto	
		<i>Churner</i>	Não <i>Churner</i>
Real	<i>Churner</i>	795	498
	Não <i>Churner</i>	38	8669

Os resultados obtidos para este algoritmo, revelam a importância da escolha da medida adequada para avaliar os modelos. Comparando com a rede neuronal obtém melhor resultado de *precision* e *f-measure* mas um resultado de *sensitivity* ligeiramente inferior (Tabela 35).

Tabela 35: Avaliação da técnica SVM

Medida	Resultado
AUC	0,831 +/- 0,029
Accuracy	94,64% +/- 0,46%
Sensitivity	61,49% +/- 3,78%
Specificity	99,56% +/- 0,21%
F-measure	74,72% +/- 2,76%
Precision	95,52% +/- 2,07%
Lift	738,75% +/- 16,42%

O aumento da taxa de *churners* no conjunto de dados de treino segue a tendência já revelada em outros métodos, ou seja, a melhoria na sensibilidade e o depreciação de todos os outros indicadores com o aumento da densidade de *churners* (Figura 40).

Figura 40: Resultados das SVM variando a densidade de *churners*

Como se pode verificar pela Figura 41, que corresponde ao output do algoritmo SVM, a extração de informação útil que permita uma compreensão das causas e factores por de trás do fenómeno de *churn* não é o ponto forte deste algoritmo. Os atributos que aparecem representados são simplesmente aqueles escolhidos na pré-seleção efectuada antes do processamento do próprio algoritmo.

```
Kernel Model  
  
Total number of Support Vectors: 10000  
Bias (offset): -0.257  
  
w[gender] = -0.002  
w[nrChildrenAtHome] = -0.017  
w[frequencyLastMonth] = -0.054  
w[frequencyLastMonthR] = -0.166  
w[monetaryPerVisit] = -0.035  
w[daysFromLastShop] = 0.300  
w[ipt] = -0.013  
w[maxIPT] = 0.002  
w[relative_mop1] = -0.005  
w[relative_mop6] = -0.003
```

Figura 41: *Output* gerado pelo RapidMiner aquando da utilização das SVM

4.4.1 Síntese dos Resultados

Na Tabela 36 resumimos os resultados obtidos com a aplicação dos vários métodos que aplicámos sobre o nosso caso de estudo. Através de uma rápida observação, facilmente constatamos a superioridade das árvores de decisão e das *random forests* relativamente à capacidade de previsão, o que está reflectido nas diversas medidas de avaliação. Entre estas duas técnicas a diferença é pouco significativa e os melhores resultados dividem-se entre si. A árvore de decisão, porém, apresenta os melhores resultados de AUC e *sensitivity*, enquanto a *random forest* revela os valores máximos de *accuracy*, *precision* e *sensitivity*. Estas apresentam resultados muito semelhantes em todas as medidas de avaliação. Com estas duas técnicas foram detectados, aproximadamente, 67% de eventos verdadeiros de *churn*. O nível de desperdício que revelam é pouco significativo, já que, são menos de 4% os casos classificados erradamente como *churners*.

A um nível intermédio, com resultados satisfatórios, encontram-se a rede neuronal e a SVM. Enquanto a rede neuronal leva ligeira vantagem na proporção de verdadeiros eventos detectados (*sensitivity*), o desperdício com a SVM é menor (*precision* maior). Isto reflete-se nos valores da *f-measure* em que a SVM tem ligeiramente melhor resultado.

Com os piores resultados está a regressão logística que apresenta valores muito inferiores a todas as outras técnicas em todos os factores de avaliação. A capacidade de detecção de verdadeiros eventos de *churn* (*sensitivity*) não chega a um terço dos valores apresentados pelas outras técnicas. Já os erros de classificação, na classe positiva, acontecem aproximadamente em 20% dos casos, uma proporção significativamente superior às restantes técnicas testadas.

Tabela 36: Síntese dos resultados obtidos.

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)	Precision (%)	Lift (%)
Árvore de Decisão	0,838 +/- 0,024	95,36% +/- 0,49%	66,82% +/- 3,94%	99,60% +/- 0,21%	78,77% +/- 2,71%	96,17% +/- 1,90%	741,70% +/- 16,18%
Regressão Logística	0,760 +/- 0,019	88,82% +/- 0,51%	17,64% +/- 3,91%	99,39% +/- 0,23%	28,97% +/- 17,63%	80,98% +/- 6,16%	626,35% +/- 48,06%
Rede Neuronal	0,823 +/- 0,030	94,33% +/- 0,55%	61,87% +/- 3,66%	99,15% +/- 0,27%	73,78% +/- 2,98%	91,56% +/- 2,49%	708,08% +/- 18,72%
Random Forest	0,830 +/- 0,023	95,37% +/- 0,48%	66,75% +/- 3,89%	99,62% +/- 0,18%	78,79% +/- 2,69%	96,37% +/- 1,63%	745,31% +/- 13,30%
SVM	0,831 +/- 0,029	94,64% +/- 0,46%	61,49% +/- 3,78%	99,56% +/- 0,21%	74,72% +/- 2,76%	95,52% +/- 2,07%	738,75% +/- 16,42%

Relativamente à facilidade de interpretação dos modelos, como foi anteriormente referido, o destaque vai para a árvore de decisão. Com a geração de um *output* estruturado em forma de árvore, supera largamente os restantes métodos na extração de informação, característica muito relevante para a explicação do fenómeno de *churn*. Para além disso, a árvore de decisão, fornece um bom compromisso entre a interpretabilidade e o poder de previsão, o que pesa a seu favor numa escolha final do método a utilizar. Com a exceção da *random forest*, que não fornece qualquer *output* relevante, as restantes técnicas apenas disponibilizam pesos relacionados com as variáveis de entrada.

Quanto à dificuldade de otimização, podemos separar os métodos em dois grupos. A árvore de decisão, regressão logística e *random forest* fazem parte do grupo de mais simples e rápida otimização. Isto deve-se ao facto de possuírem um conjunto reduzido de parâmetros bem como um tempo de treino relativamente pequeno. No outro grupo encontram-se a rede neuronal e a SVM. Para isto contribuiu o elevado tempo de treino destas técnicas, que varia com os valores dos parâmetros. Como se verificou, com a técnica SVM o tempo de treino para um conjunto de dados

com dez mil exemplos e 40 atributos de entrada demorou aproximadamente quatro horas e meia. Esta condição pode ser problemática para conjuntos de dados de treino maiores.

O aumento da densidade de *churners* através da redução de exemplos da classe negativa não revelou resultados muito animadores. Isto pode dever-se, entre outras coisas, à redução do tamanho do *dataset* de treino. Um conjunto de dados de treino maior poderia sofrer menos com a redução da densidade da classe negativa. O que se verifica é o aumento do número de eventos de *churn* com o aumento da densidade de *churners* no conjunto de dados de treino, mas, por outro lado, também aumentam (na maioria das vezes em maior proporção) os casos classificados erradamente como positivos. Apenas na regressão logística os resultados desta experiência foram razoáveis já que se verifica uma melhoria da *f-measure*, o que significa que o custo do aumento da *sensitivity* na *precision* do modelo é razoável.

4.4.2 Implicações financeiras

A principal razão para a aplicação de técnicas de análise e mineração de dados numa campanha de prevenção de *churn* é otimizar a aplicação dos recursos monetários disponíveis. Com fundos limitados, numa campanha de prevenção em massa, os valores para investir sobre cada cliente são bastante limitados. Para além disso, está a ser desperdiçado muito dinheiro em clientes que não estão, de facto, em risco de serem perdidos. A aplicação de técnicas de mineração de dados permite diminuir o desperdício e aumentar o valor a investir sobre os clientes alvo. Mesmo com a aplicação de técnicas de mineração de dados, é necessário um grande investimento para efetuar uma campanha de prevenção de *churn*. É importante que se faça uma análise financeira prévia para avaliar se os proveitos da campanha justificam o investimento. Deste modo, para avaliar a campanha foi definida uma simples campanha de retenção.

Com um retorno médio por cliente de aproximadamente 924€, durante um período de relação que se estende em média por 779 dias (2,134 anos), admite-se que a campanha de prevenção estenderá a relação por um ano, o que dá origem a um retorno médio de 433€ por cliente. A campanha de prevenção consiste em oferecer incentivos (cupões de desconto, pontos no cartão de fidelidade, etc.) ao longo do período de extensão da relação esperado. Pretende-se investir um valor inferior a um décimo do retorno previsto por cliente, logo, foi definido que o valor a investir

na retenção dos clientes seria de 40 euros por cliente, que se traduziriam em ofertas promocionais diretas. Admitindo uma taxa de sucesso de 30% da campanha de prevenção (clientes que permanecem fieis à empresa devido ao incentivo que receberam), pode-se verificar na Tabela 37 o impacto que a escolha da técnica a usar tem na rentabilidade do projeto. Para além disso, é associado o valor de rentabilidade no caso de um investimento equivalente ser efectuado de forma aleatória, ou seja, se a aplicação dos incentivos fosse feita sobre um número de clientes igual à quantidade dos clientes classificados como *churners* pela técnica usada, mas escolhendo-os de forma aleatória.

Tabela 37: Rentabilidades e Investimentos para as diferentes técnicas

	Rentabilidade	Investimento	ROI	Rentabilidade (seleção aleatória)
Árvore de Decisão	76273,6	35960	3,12	-20761,7
Regressão Logística	18377,2	11240	2,63	-6433,7
Rede Neuronal	69000	34920	2,98	-20241,3
<i>Random Forest</i>	76263,7	35840	3,13	-20771,6
SVM	69071,3	33160	3,08	-19130,8

É possível verificar também em termos financeiros a superioridade das técnicas árvore de decisão e *random forest*. Os resultados destas técnicas são muito semelhantes. Apesar da árvore de decisão ter ligeiramente melhor rentabilidade, o retorno sobre o investimento (do inglês, *return on investment* - ROI) é superior na *random forest*. Para as restantes técnicas, os resultados financeiros seguem, da mesma forma, a tendência registada com as medidas de avaliação, ou seja, com resultados intermédios a rede neuronal e a SVM, e com os piores resultados a regressão logística. Para além disso, facilmente se constata a vantagem do uso de técnicas de mineração de dados na campanha de prevenção de *churn*.

A Tabela 38 mostra a variação de rentabilidade, investimento, retorno e ROI que ocorre com o aumento da densidade de *churners* no conjunto de dados de treino. Para a árvore de decisão e para a rede neuronal o aumento da densidade de *churners* apenas prejudica os resultados obtidos com o conjunto de dados original. Já para os restantes métodos o aumento da percentagem de *churners* proporciona ganhos na rentabilidade embora suportados pelo aumento no investimento. No entanto, para estes últimos, o retorno sobre o investimento diminui à medida que se aumenta

o número de *churners*. Para a regressão logística a melhor rentabilidade é obtida com 30% *churners*. Já a *random forest* e a SVM obtêm a melhor rentabilidade com 20% de *churners*.

Tabela 38: Impactos financeiros com o aumento da densidade de *churners*

	% <i>Churners</i>	Rentabilidade	Investimento	Retorno	ROI
Árvore de Decisão	13	76273,6	35960	112233,6	3,121
	15	76143,7	35960	112103,7	3,117
	20	75603,5	36760	112363,5	3,057
	30	74563,9	37280	111843,9	3,000
	40	72133,8	39840	111973,8	2,811
	50	67073,2	45680	112753,2	2,468
Regressão Logística	13	18377,2	11240	29617,2	2,635
	15	20263,5	14160	34423,5	2,431
	20	26071,7	23680	49751,7	2,101
	30	26613,3	47040	73653,3	1,566
	40	13744,6	84200	97944,6	1,163
	50	-20009,2	135880	115870,8	0,853
Rede Neuronal	13	68960	34960	103920	2,973
	15	66991,3	35240	102231,3	2,901
	20	67529,1	37560	105089,1	2,798
	30	65528,3	40600	106128,3	2,614
	40	61315,2	48840	110155,2	2,255
	50	57943,3	54680	112623,3	2,060
<i>Random Forest</i>	13	76263,7	35840	112103,7	3,128
	15	76173,8	35800	111973,8	3,128
	20	76333,4	36160	112493,4	3,111
	30	75443,5	36920	112363,5	3,043
	40	72893	40120	113013	2,817
	50	67562,3	46360	113922,3	2,457
SVM	13	69950,5	33320	103270,5	3,099
	15	70599,2	34360	104959,2	3,055
	20	71787,3	35640	107427,3	3,014
	30	69965,1	40320	110285,1	2,735
	40	66383,7	45720	112103,7	2,452
	50	58271,6	56560	114831,6	2,030

Na Figura 42 é possível visualizar as evoluções da rentabilidade e retornos obtidos em comparação com a evolução dos investimentos necessários com o aumento da densidade de *churners*. São os valores obtidos pela regressão logística, onde se pode verificar a rentabilidade máxima para 30% de *churners*, mas também um investimento necessário substancialmente superior ao inicial.

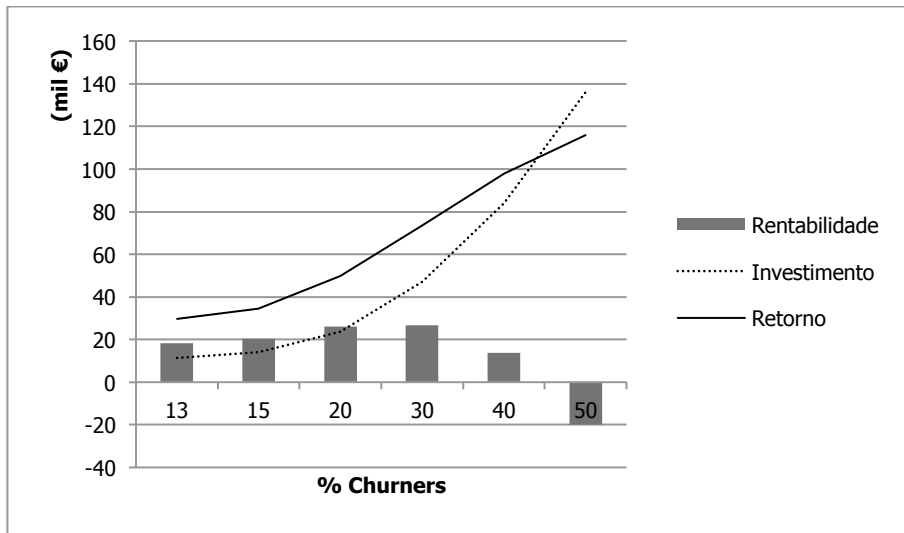


Figura 42: Impactos financeiros com o aumento da densidade de *churners* para a RL

Um factor importante a ter em conta é a escolha dos valores admitidos como parâmetros da fórmula de rentabilidade. A escolha da técnica, e respectivos parâmetros, poderá ser diferente conforme os valores admitidos como taxa de sucesso da campanha ou valor do cliente. A Tabela 39 descreve as rentabilidades obtidas com a variação da taxa de sucesso para diferentes densidades de *churners*. Pegando no exemplo da *random forest* verifica-se que o valor de densidade de *churners* que se escolheria seria diferente consoante a taxa de sucesso admitida.

Tabela 39: Rentabilidades para diferentes taxas de sucesso e densidades de *churners*

	% <i>Churners</i>	Taxa de sucesso		
		10%	30%	50%
Árvore de Decisão	13	1451,2	76273,6	151096
	15	1407,9	76143,7	150879,5
	20	694,5	75603,5	150512,5
	30	1,3	74563,9	149126,5
	40	-2515,4	72133,8	146783
	50	-8095,6	67073,2	142242
Regressão Logística	13	-1367,6	18377,2	38122
	15	-2685,5	20263,5	43212,5
	20	-7096,1	26071,7	59239,5
	30	-22488,9	26613,3	75715,5
	40	-51551,8	13744,6	79041
	50	-97256,4	-20009,2	57238
Rede Neuronal	13	-320	68960	138240
	15	-1162,9	66991,3	135145,5
	20	-2530,3	67529,1	137588,5
	30	-5223,9	65528,3	136280,5
	40	-12121,6	61315,2	134752
	50	-17138,9	57943,3	133025,5
<i>Random Forest</i>	13	1527,9	76263,7	150999,5
	15	1524,6	76173,8	150823
	20	1337,8	76333,4	151329
	30	534,5	75443,5	150352,5
	40	-2449	72893	148235
	50	-8385,9	67562,3	143510,5
SVM	13	1103,5	69950,5	138797,5
	15	626,4	70599,2	140572
	20	169,1	71787,3	143405,5
	30	-3558,3	69965,1	143488,5
	40	-8352,1	66383,7	141119,5
	50	-18282,8	58271,6	134826

A valor dos parâmetros admitidos também tem influência na importância dos valores das medidas de avaliação das técnicas. O aumento do valor admitido de cada cliente retido valoriza a medida *sensitivity* e, por outro lado, o aumento dos custos médios por cliente aumenta a importância da *precision*.

Capítulo 5

Conclusões e Trabalho Futuro

5.1 Comentários e Conclusões

A mudança para uma cultura centrada no cliente pode trazer grandes benefícios às empresas. A percepção de que a relação com os clientes passa por várias fases, bem como a aplicação de diferentes processos a acompanhá-las, são essenciais no sentido de retirar o maior proveito possível destas ligações. Com o aumento da competitividade entre as empresas, a retenção de clientes passa a ser uma atividade prioritária. Para além disso, como sabemos, as campanhas de retenção revelam-se menos dispendiosas do que a aquisição de novos clientes e incidem sobre os clientes mais fieis que eventualmente serão os mais rentáveis.

A consciencialização da necessidade de combater o *churn* de clientes leva as empresas a procurar conhecer os clientes e a forma de obter esse conhecimento é através do uso de uma quantidade de dados abundante obtida através da interação com os clientes no dia a dia da empresa. A informação relevante sobre os clientes pode, depois, ser extraída e tratada recorrendo a ferramentas estatísticas de CRM ou de mineração de dados. Uma das principais estratégias de retenção é a prevenção proactiva do *churn* do clientes, que visa, na sua essência, prever quais os

clientes que estão em risco de abandonar a empresa voluntariamente, com o objectivo de se agir de forma direccionada e antecipada a fim de evitar essa situação.

O trabalho de pesquisa apresentado em (Ngai et al., 2009) confirma a aposta na retenção de clientes verificada nos últimos anos. No entanto, a pesquisa de artigos sobre previsão de *churn*, especificamente no sector do retalho, revelou-se pouco frutífera, ajudando apenas a sustentar a afirmação de (Buckinx e Van den Poel, 2005) de que o problema de *churn* não tem sido suficientemente estudado nesta área em particular. Concretamente, foram encontrados apenas dois artigos relativos à previsão de *churn* no retalho, nomeadamente (Buckinx e Van den Poel, 2005; Ju e Guo, 2008).

A previsão de *churn* de clientes é um processo complexo, surgindo, sobretudo, devido às particularidades das diferentes situações e contextos em que as empresas estão inseridas. Devido a esse facto as soluções CRM comerciais fornecem pouco suporte à análise de *churn*, concentrando-se essencialmente nas áreas de *cross-selling* e *up-selling* (Hadden, 2008). Por isso, é geralmente necessário construir soluções à medida para abordar o problema. O uso de técnicas de mineração de dados é uma abordagem natural na abordagem ao problema da previsão de *churn*. A sua modelação preditiva é geralmente formulada como um problema de classificação binária ("*churn*" ou "não *churn*"). O uso de uma metodologia bem documentada e padronizada, como a CRISP-DM, facilita e guia o processo de mineração de dados, servindo como um guião ao analista de dados. O facto das ligações entre os clientes e a empresa não estarem contratualizadas acrescenta mais uma camada de complexidade ao processo de previsão. A razão disto assim acontecer é que não há uma comunicação da potencial quebra da relação entre as partes e, desta forma, terá que ser conceptualizada uma definição do momento de *churn* dos clientes.

A construção da assinatura que caracteriza um cliente é uma das principais etapas no processo de mineração. É uma fase de extrema importância para o sucesso da aplicação. É no desenho desta assinatura que se introduz o conhecimento de negócio e se escolhem as principais características relativas ao cliente e aos períodos envolvidos no processo de previsão. Esta fase inclui a conceptualização do conceito de *churn* (i.e. se não estiver já definido no contexto da empresa), uma etapa de enorme importância no processo de modelação.

Durante o processo de modelação terá de se escolher uma janela temporal durante a qual se extraem os dados e se efetua a própria previsão. São várias as escolhas que se têm que fazer, pois

existem vários períodos com objectivos distintos para os quais deve haver uma boa ponderação. Os principais períodos reconhecidos em processos de previsão de *churn* são: o período de observação, o período de latência e o período de definição da variável dependente. No primeiro efetua-se a extração dos dados históricos que vão ser usados na tarefa de previsão, o período de latência é usado essencialmente por questões logísticas e o ultimo refere-se à janela onde é definida a variável dependente, ou seja, são definidos os clientes como "*churners*" e os "*não churners*".

A conceptualização da variável dependente é também uma definição muito importante no processo e que beneficia muito de um bom conhecimento do negócio. Uma boa definição poderá ter um impacto muito positivo na detecção dos verdadeiros *churners*. Uma má definição levará a classificar erradamente como *churners*, clientes que na realidade não o eram, assim como, não abranger verdadeiros *churners* classificando-os como não *churners*. Resumidamente, uma boa definição (do conceito de *churner*) deverá ter origem num bom compromisso entre as seguintes premissas: permitir uma verificação rápida e ter um elevado grau de certeza.

Uma propriedade natural dos problemas de previsão de *churn* é a enorme diferença existente no número de casos de ambas as classes ("*churners*" e "*não churners*"). A existência de uma proporção muito superior dos casos de "*não churners*" leva a uma dificuldade acrescida na classe positiva, sendo esta a classe com maior relevância no processo. A detecção dos casos positivos de *churn* é bastante mais valiosa que a previsão de casos de "*não churners*". Um forma de combater esta contrariedade é alterando as densidades das classes. Por um lado pode-se aumentar (através de duplicação de casos) a densidade dos casos positivos, e por outro lado, pode-se diminuir a densidade dos casos da classe negativa (eliminando casos).

A compreensão dos dados é fundamental para uma boa definição dos clientes. É necessário ter um bom conhecimento dos dados para escolher os mais relevantes, para o processo de previsão, e os que melhor caracterizam os clientes. Os dados usados para construir as assinaturas são provenientes, fundamentalmente, de dois tipos de dados: dados transacionais registados no momento da compra e dados sociodemográficos, relativos ao cliente e associados ao seu cartão de cliente.

A escolha da técnica a aplicar no processo de previsão é uma das etapas principais da fase de modelação. O principal aspecto a ter em conta é o compromisso pretendido entre a capacidade de

previsão e a compreensibilidade dos resultados do modelo. Além disso, é necessário ter em conta os recursos e limites temporais disponíveis para implementar a campanha. O teste e a otimização de vários métodos em simultâneo pode se tornar muito dispendioso, sobretudo quando se utilizam algoritmos mais complexos como as redes neuronais ou as SVM.

Como vimos, existem variadíssimas técnicas para processos de classificação. No entanto, para este trabalho foram selecionadas 5 das técnicas mais utilizadas no contexto de previsão de *churn* (KhakAbi et al., 2010): árvores de decisão, regressão logística, redes neuronais, *random forest* e SVM. A superioridade apresentada pelos modelos de árvore de decisão, rede neuronal, *random forest* e SVM relativamente à regressão logística, é um possível indicador do poder das técnicas de mineração de dados aplicadas na previsão de eventos de *churn*.

Do ponto de vista da compreensibilidade as árvores de decisão proporcionam o melhor resultado com o seu *output* estruturado em forma de árvore e, para além disso, podem fornecer excelentes resultados de previsão como foi demonstrado. Para além da árvore de decisão, também a *random forest* demonstrou ser uma técnica robusta e com excelente poder preditivo. Contudo, aceita-se que os resultados possam variar com diferentes conjuntos e tipos de dados. Uma boa decisão deverá resultar de uma boa comunicação entre os utilizadores da parte técnica e os utilizadores da parte de negocio. Relativamente à dificuldade de treino e optimização destacam-se pela negativa as redes neuronais e, sobretudo, as SVM. Isto deve-se sobretudo ao elevado tempo de treino destes métodos e à influência que alguns dos seus parâmetros têm neste aspecto.

Um outro aspecto que se revela de enorme importância é o conhecimento das particularidades das diferentes medidas de avaliação dos modelos e das suas características. A escolha da medida correta pode resultar em ganhos adicionais na campanha. Como se verificou com as experiências de aumento da densidade de *churners* a melhoria nos valores registados para uma dada medida são, por vezes, contrapostos pela deterioração dos valores obtidos para outras medidas. Concretamente, as melhorias registadas nos valores de *sensitivity* eram normalmente acompanhadas pelo empobrecimento da *precision* do modelo. A escolha da melhor medida pode ser fundamentada através de cálculos de avaliação da rentabilidade como exemplificado anteriormente (secção 2.4.2) em que o peso dos custos para manter os clientes, dos recursos disponíveis e do retorno que os clientes proporcionam são factores chave. Se por um lado os recursos disponíveis para a companhia são relativamente escassos a precisão pode ser uma boa medida. Se por outro lado os recursos são amplos e o retorno proporcionado pela manutenção dos

clientes é consideravelmente elevado relativamente ao custos de manutenção a sensibilidade também é uma opção muito útil. A medida *f-measure* representa um compromisso entre a sensibilidade e a precisão.

Por fim, uma palavra relativamente ao caso de estudo utilizado. A indisponibilidade de um conjunto de dados real criou algumas limitações no desenvolvimento do trabalho. Mas, por seu lado, um conjunto de dados real levanta uma série de problemas que um conjunto de dados gerado artificialmente não. Normalmente, grande parte do esforço do processo de descoberta de conhecimento é dedicado ao processamento e tratamento de dados. Desta forma, não foi possível experimentar e testar esse conjunto de tarefas que normalmente seriam de esperar. Uma dessas tarefas, a seleção dos melhores dados para definir as assinaturas, é uma das mais importantes de todo o processo de mineração. Como já referido, os dados usados para mineração, estão sujeitos ao efeito do ambiente circundante e às alterações que este regista com o decorrer do tempo. A ausência de um conjunto de dados real impossibilita a realização de uma série testes sobre o impacto que o tempo tem nas previsões efectuadas. Não permite verificar se os dados adquiridos numa determinada estação do ano podem ser aplicados na previsão de eventos numa outra estação do ano; não permite estudar a validade do modelo criado, ou seja, durante quantos meses as previsões vão ser fidedignas e, por outro lado, ao fim de quanto tempo o modelo terá de ser atualizado. Também não foi possível testar algumas soluções com vista a aumentar a estabilidade e durabilidade do modelo, como o uso de dados retirados de diferentes janelas temporais.

5.2 Limitações e Trabalho Futuro

O trabalho futuro assentará, sobretudo, na realização de algumas das tarefas que a inexistência de um conjunto de dados real impossibilitou de serem realizadas. Todos os testes e processos que deveriam ter sido realizados sobre os dados como foi anteriormente referido. Para além disso, uma previsão eficaz dos eventos de *churn* apenas serve para as empresas adquirirem o conhecimento dos clientes que estão em risco. Logo, para ter sucesso, uma campanha de gestão de *churn* deverá incluir também medidas práticas de retenção (Hung et al., 2006). Para isso, deverão ser conhecidos os diversos indicadores necessários ao cálculo efetivo dos custos e rentabilidades provenientes da aplicação da campanha, como o valor dos clientes, o custo de retenção e os recursos disponíveis (monetários e humanos).

Para uma gestão e controlo eficientes da previsão e prevenção de eventos de *churn*, o módulo de previsão deverá ser integrado numa plataforma de CRM que permita definir as diferentes rotinas associadas ao processo de prevenção de *churn* como atualização das assinaturas e dos modelos, tratamento dos dados e alarmes de aviso de clientes em risco.

A análise de técnicas de mineração de dados que não pertençam ao domínio das técnicas de classificação, como por exemplo *surviver analysis* e regras de associação, com vista a estudar o fenómeno por diferentes perspectivas e também analisar os factores que conduzem os clientes a abandonar a empresa. O uso de técnicas de segmentação para dividir os clientes em diferentes grupos, segundo os principais atributos, também poderá ter efeitos positivos no processo de classificação.

Experimentar novos métodos para reduzir o impacto da diferença de densidade das classes como por exemplo o *oversampling* da classe positiva ou uso de técnicas de mineração de dados sensíveis a custos, como o que foi sugerido por Weiss (2004). No caso concreto de *churn* isto é feito associando um custo maior aos falsos negativos do que aos falsos positivos (Burez e Van den Poel, 2009).

Bibliografia

- AHN, J.-H., HAN, S.-P. e LEE, Y.-S. 2006. Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, 30, 552-568.
- ANDERSON, E. e WEITZ, B. 1989. Determinants of continuity in conventional industrial channel dyads. *Marketing Science*, 310-323.
- ANDERSON, J. L., JOLLY, L. D. e FAIRHURST, A. E. 2007. Customer relationship management in retailing: A content analysis of retail trade journals. *Journal of Retailing and Consumer Services*, 14, 394-399.
- APED. 2010. *Ranking APED* [Online]. Available: http://www.aped.pt/Media/content/313_1_G.pdf [Accessed 14 February 2012].
- BAESENS, B., VIAENE, S., VAN DEN POEL, D., VANTHIENEN, J. e DEDENE, G. 2002. Bayesian neural network learning for repeat purchase modelling in direct marketing. *European Journal of Operational Research*, 138, 191-211.
- BERRY, M. e LINOFF, G. 2004. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, *Wiley Computer Publishing.
- BHATTACHARYA, C. 1998. When customers are members: Customer retention in paid membership contexts. *Journal of the Academy of Marketing Science*, 26, 31-44.
- BOSE, I. e MAHAPATRA, R. K. 2001. Business data mining—a machine learning perspective. *Information & Management*, 39, 211-225.
- BREIMAN, L. 1996. Bagging predictors. *Machine Learning*, 24, 123-140.
- BREIMAN, L. 2001a. Random Forests. *Machine Learning*, 45, 5-32.
- BREIMAN, L. 2001b. Statistical modeling: The two cultures. *Statistical Science*, 16, 199-231.

-
- BUCKINX, W. e VAN DEN POEL, D. 2005. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164, 252-268.
- BUREZ, J. e POEL, D. V. D. 2008. Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications*, 35, 497-514.
- BUREZ, J. e VAN DEN POEL, D. 2007. CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32, 277-288.
- BUREZ, J. e VAN DEN POEL, D. 2009. Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36, 4626-4636.
- CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C. e WIRTH, R. 2000. CRISP-DM 1.0 Step-by-step data mining guide.
- CHAWLA, N. 2005. Data Mining for Imbalanced Datasets: An Overview
Data Mining and Knowledge Discovery Handbook. In: MAIMON, O. e ROKACH, L. (eds.). Springer US.
- CHIANG, D.-A., WANG, Y.-F., LEE, S.-L. e LIN, C.-J. 2003. Goal-oriented sequential pattern for network banking churn analysis. *Expert Systems with Applications*, 25, 293-302.
- CHU, B.-H., TSAI, M.-S. e HO, C.-S. 2007. Toward a hybrid data mining model for customer retention. *Know.-Based Syst.*, 20, 703-718.
- CORTEZ, P. 2002. *Modelos inspirados na natureza para a previsão de séries temporais*. PhD, Universidade do Minho.
- COUSSEMENT, K. e VAN DEN POEL, D. 2008a. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34, 313-327.
- COUSSEMENT, K. e VAN DEN POEL, D. 2008b. Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, 45, 164-174.
- COUSSEMENT, K. e VAN DEN POEL, D. 2009. Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*, 36, 6127-6134.
- DATTA, P., MASAND, B., MANI, D. R. e LI, B. 2000. Automated Cellular Modeling and Prediction on a Large Scale. *Artificial Intelligence Review*, 14, 485-502.

-
- DIAS, P. A. G. D. S. 2009. *Reposicionamento estratégico no retalho: Aplicação a um caso de estudo - Pingo Doce*. MSc, ISCTE Business School.
- DIETTERICH, T. G. 2000. Ensemble Methods in Machine Learning. *Proceedings of the First International Workshop on Multiple Classifier Systems*. Springer-Verlag.
- FAYYAD, U., PIATETSKY-SHAPIO, G. e SMYTH, P. 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17, 37-54.
- GHORBANI, A. e TAGHIYAREH, F. CMF: A framework to improve the management of customer churn. Services Computing Conference, 2009. APSCC 2009. IEEE Asia-Pacific, 7-11 Dec. 2009 2009. 457-462.
- HADDEN, J. 2008. *A Customer Profiling Methodology for Churn Prediction*. PhD, Cranfield University.
- HADDEN, J., TIWARI, A., ROY, R. e RUTA, D. 2007. Computer assisted customer churn management: State-of-the-art and future trends. *Comput. Oper. Res.*, 34, 2902-2917.
- HAN, J. e KAMBER, M. 2006. *Data Mining: Concepts and Techniques*, Morgan Kaufmann.
- HU, X. 2005. A Data Mining Approach for Retailing Bank Customer Attrition Analysis. *Applied Intelligence*, 22, 47-60.
- HUNG, S.-Y., YEN, D. C. e WANG, H.-Y. 2006. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31, 515-524.
- IVANCIUC, O. 2002. Structure-odor relationships for pyrazines with support vector machines. *Internet Electron. J. Mol. Des*, 1, 269-284.
- JAHANZEB, S. e JABEEN, S. 2007. Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. *J Database Mark Cust Strategy Manag*, 14, 120-129.
- JINBO, S., XIU, L. e WENHUANG, L. The Application of AdaBoost in Customer Churn Prediction. Service Systems and Service Management, 2007 International Conference on, 2007. Ieee, 1-6.
- JING, Z. e XING-HUA, D. Bank Customer Churn Prediction Based on Support Vector Machine: Taking a Commercial Bank's VIP Customer Churn as the Example. Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on, 12-14 Oct. 2008 2008. 1-4.
- JU, C. e GUO, F. 2008. Research and Application of Customer Churn Analysis in Chain Retail Industry. *Proceedings of the 2008 International Symposium on Electronic Commerce and Security*. IEEE Computer Society.

-
- KDNUGGETS. 2011. *Data Mining/Analytic Tools Used* [Online]. Available: <http://www.kdnuggets.com/polls/2011/tools-analytics-data-mining.html> [Accessed 14 February 2012].
- KHAKABI, S., GHOLAMIAN, M. R. e NAMVAR, M. Data Mining Applications in Customer Churn Management. Intelligent Systems, Modelling and Simulation (ISMS), 2010 International Conference on, 27-29 Jan. 2010. 220-225.
- KIM, H.-S. e YOON, C.-H. 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28, 751-765.
- KIM, S., SHIN, K.-S. e PARK, K. 2005. An Application of Support Vector Machines for Customer Churn Analysis: Credit Card Case
Advances in Natural Computation. In: WANG, L., CHEN, K. e ONG, Y. (eds.). Springer Berlin / Heidelberg.
- KIMBALL, R. e ROSS, M. 2002. *The data warehouse toolkit: the complete guide to dimensional modeling*, Wiley.
- LARIVIÈRE, B. e VAN DEN POEL, D. 2005. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29, 472-484.
- LEJEUNE, M. A. P. M. 2001. Measuring the impact of data mining on churn management. *Internet Research*, 11, 375-387.
- LEMON, K. N., WHITE, T. B. e WINER, R. S. 2002. Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *The Journal of Marketing*, 1-14.
- LEUNG, P. e TRAN, L. T. 2000. Predicting shrimp disease occurrence: artificial neural networks vs. logistic regression. *Aquaculture*, 187, 35-49.
- LIAW, A. e WIENER, M. 2002. Classification and Regression by randomForest. *R News*, 2, 18-22.
- MASAND, B., DATTA, P., MANI, D. R. e LI, B. 1999. CHAMP: A Prototype for Automated Cellular Churn Prediction. *Data Mining and Knowledge Discovery*, 3, 219-225.
- MATTISON, R. 2005. *The Telco Churn Management Handbook*, Lulu.com.
- MENEZES, R. e PORTELA, S. 2009. A Logit Model of Customer Churn as a Way to Improve the Customer Retention Strategy in the Fixed Telecommunications Industry in Portugal. *Australian and New Zealand Marketing Academy Conference*. Melbourne, Australia.

-
- MITTAL, V. e KAMAKURA, W. A. 2001. Satisfaction, repurchase intent, and repurchase behavior: Investigating the moderating effect of customer characteristics. *Journal of Marketing Research*, 131-142.
- MOZER, M. C., WOLNIEWICZ, R., GRIMES, D. B., JOHNSON, E. e KAUSHANSKY, H. 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *Neural Networks, IEEE Transactions on*, 11, 690-696.
- NESLIN, S., GUPTA, S., KAMAKURA, W., LU, J. e MASON, C. 2006. Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. *Journal of Marketing Research*, 43, 204-211.
- NG, K. e LIU, H. 2000. Customer Retention via Data Mining. *Artif. Intell. Rev.*, 14, 569-590.
- NGAI, E. W. T., XIU, L. e CHAU, D. C. K. 2009. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36, 2592-2602.
- OWCZARCZUK, M. 2010. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, 37, 4710-4712.
- PARVATIYAR, A. e SHETH, J. N. 2001. Customer relationship management: emerging practice, process, and discipline. *Journal of Economic and Social Research*, 3, 1-34.
- PIATETSKY-SHAPIRO, G. e MASAND, B. 1999. Estimating campaign benefits and modeling lift. *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. San Diego, California, United States: ACM.
- QUINLAN, J. R. 1993. *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*, Morgan Kaufmann.
- REICHHELD, F. F. e TEAL, T. 1996. *The loyalty effect: the hidden force behind growth, profits, and lasting value*, Harvard Business School Press.
- REINARTZ, W. e KUMAR, V. 2002. The mismanagement of customer loyalty. *Harvard business review*, 80, 86-95.
- RYGIELSKI, C., WANG, J.-C. e YEN, D. C. 2002. Data mining techniques for customer relationship management. *Technology in Society*, 24, 483-502.
- SANTOS, M. F. e AZEVEDO, C. S. 2005. *Data Mining - Descoberta de Conhecimento em Bases de Dados*.
- SCHAPIRE, R., FREUND, Y., BARTLETT, P. e LEE, W. 1998. Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *The Annals of Statistics*, 26, 1651-1686.

-
- SCHMITTLEIN, D. C. e PETERSON, R. A. 1994. Customer base analysis: An industrial purchase process application. *Marketing Science*, 41-67.
- SHEARER, C. 2000. The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 5.
- THOMPSON, B. 2005. *The loyalty connection: Secrets to customer retention and increased profits* [Online]. CMRGuru. Available: [http://www.rightnow.com/briefcase-files/PDFs/The Loyalty Connection Secrets to Customer Retention and Increased Profits.pdf](http://www.rightnow.com/briefcase-files/PDFs/The_Loyalty_Connection_Secrets_to_Customer_Retention_and_Increased_Profits.pdf) [Accessed 12 February 2012].
- TNSPORTUGAL. 2008a. *Retail Tracking* [Online]. Available: <http://www.tns.pt/?lop=conteudo&op=c9f0f895fb98ab9159f51fd0297e236d&id=45fbc6d3e05ebd93369ce542e8f2322d> [Accessed 14 February 2012].
- TNSPORTUGAL. 2008b. *Wordpanel* [Online]. Available: http://www.rousseau.com.pt/estudos/tns_kantar/opportunidade_de_crescimento [Accessed 14 February 2012].
- TSIPTISIS, K. e CHORIANOPOULOS, A. 2010. *Data Mining Techniques in CRM: Inside Customer Segmentation*, Wiley.
- VAKRATSAS, D. 1998. Household cost effects on purchase timing decisions: do demographics matter? *Journal of Consumer Marketing*, 15, 6-22.
- VAN DEN POEL, D. e LARIVIÈRE, B. 2004. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157, 196-217.
- VERBEKE, W., MARTENS, D., MUES, C. e BAESENS, B. 2011. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38, 2354-2364.
- WANG, Y.-F., CHIANG, D.-A., HSU, M.-H., LIN, C.-J. e LIN, I. L. 2009. A recommender system to avoid customer churn: A case study. *Expert Systems with Applications*, 36, 8071-8075.
- WEI, C.-P. e CHIU, I. T. 2002. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, 23, 103-112.
- WEISS, G. M. 2004. Mining with rarity: a unifying framework. *SIGKDD Explor. Newsl.*, 6, 7-19.
- XU, E., SHAO, L., GAO, X. e ZHAI, B. An algorithm for predicting customer churn via BP neural network based on rough set. 2006. IEEE, 47-50.
- YAN, L., MILLER, D. J., MOZER, M. C. e WOLNIEWICZ, R. Improving prediction of customer behavior in nonstationary environments. *Neural Networks*, 2001. Proceedings. IJCNN '01. International Joint Conference on, 2001 2001. 2258-2263 vol.3.

- YAN, L., WOLNIEWICZ, R. H. e DODIER, R. 2004. Predicting customer behavior in telecommunications. *Intelligent Systems, IEEE*, 19, 50-58.
- ZAN, M., SHAN, Z., LI, L. e AI-JUN, L. A Predictive Model of Churn in Telecommunications Based on Data Mining. *Control and Automation, 2007. ICCA 2007. IEEE International Conference on, 2007. IEEE*, 809-813.
- ZHAO, Y., LI, B., LI, X., LIU, W. e REN, S. 2005. Customer Churn Prediction Using Improved One-Class Support Vector Machine. *In: LI, X., WANG, S. e DONG, Z. (eds.) Advanced Data Mining and Applications. Springer Berlin / Heidelberg*.
- ZORN, S., BELLMAN, S. e JARVIS, W. 2008. Identifying Customers Likely to Churn. *Australian and New Zealand Marketing Academy Conference. University of Western Sydney, Sydney*.

Referências WWW

- [1] <http://rapid-i.com/content/view/181/190/>
- [2] <http://www.cs.waikato.ac.nz/ml/weka/>