

Instituto Superior de Ciências do Trabalho e da Empresa



CUSTOMER LIFETIME VALUE NA BANCA

Ana Margarida Gomes Alexandre

Relatório de Projecto
Mestrado em Prospeção e Análise de Dados

Orientadora:
Prof. Dra. Anabela Costa, Professora Auxiliar, ISCTE Business School, Departamento de
Métodos Quantitativos

Co-orientador:
Mestre Elson Filho, Customer Intelligence Practice Manager, SAS Portugal

Junho 2009

Resumo

Ao longo do tempo, as instituições financeiras, muito em particular os bancos, têm cada vez mais apostado na implementação, na gestão de processos e em métricas que possibilitem a gestão do relacionamento com os clientes, e o próprio conhecimento inteligente sobre o valor dos mesmos. Uma das possíveis métricas a utilizar é o *Customer Lifetime Value* (CLTV), que tem vindo a ganhar cada vez maior importância.

Este projecto consistiu no desenvolvimento, e implementação, do modelo de CLTV probabilístico num banco, evidenciando algumas das limitações das métricas “tradicionais” utilizadas para o cálculo do valor do cliente. A metodologia utilizada no desenvolvimento do projecto foi a metodologia “implementação plataformas inteligentes” (IPI), a qual é utilizada pelo SAS para o desenvolvimento de projectos de *Data Mining*. Para a obtenção do CLTV foi necessário efectuar o cálculo e análise da rentabilidade e potencialidade futura associada a cada cliente ao longo do período de tempo em análise. A partir do desenvolvimento do modelo de CLTV foi possível identificar quais os clientes com maior valor actual (rentabilidade) e maior valor futuro (potencialidade), com o intuito de futuramente criar estratégias específicas que permitam maximizar o ciclo de vida dos clientes no banco.

Classificação JEL: C13, C45

Palavras-chave: Valor de Cliente, Instituições Financeiras, Mineralização de Dados, Gestão do Relacionamento dos Clientes.

Abstract

Over time, financial institutions, especially banks, have increasingly focused on implementation, processes management and metrics that allow the relationship management with customers, and the intelligent understanding of the value of them. One of the metric used is the Customer Lifetime Value (CLTV), which has gained increasing importance over time.

This project was based on developing and implementing the probabilistic model CLTV of a bank, highlighting some of the limitations of the "traditional" metrics used to calculate the customer value. The methodology used for developing the project was "implementing intelligent platforms" (IPI), which is used by SAS for the development of Data Mining projects. To obtain the CLTV model was necessary to calculate and analyze the profitability and future potential associated with each client over the time period under review. From the development of the CLTV model we could identify customers with higher current value (profitability) and greater future value (potential), in order to further establish specific strategies that can maximize the life cycle of customers in the bank.

JEL Classification: C13, C45

Keywords: Customer LifeTime Value, Financial Institutions, Data Mining, Customer Relationship Management

Agradecimentos

Ao ISCTE-IUL pelas facilidades concedidas.

À minha Orientadora, Professora Doutora Anabela Costa, por ter acreditado neste projecto, e me disponibilizar conhecimentos que me permitiram desenvolver este tema, considerado até então, como sendo algo recente. Obrigada pela confiança e dedicação.

Ao Mestre Elson Filho e ao SAS Portugal por me disponibilizar informações e conhecimentos úteis para a concretização deste trabalho

Aos meus pais, ao meu irmão, e à Alice, por toda a força, apoio e confiança, que me deram na elaboração deste trabalho. Sem eles não seria possível. O meu obrigado por estarem sempre comigo e me apoiarem.

Um agradecimento muito especial ao Rui Gonçalves por todo o apoio, dedicação, conhecimento e confiança que me transmitiu. Obrigado por acreditar em mim.

Quero agradecer, por último, às minhas grandes amigas Daniela Germano, Diana Areia, Paula Pereira (“porque valem muito mais do que o sol”), à Sara Fonseca, e a todos os meus outros amigos por toda a energia, confiança e questões que me foram colocando. Sem todos vocês tudo seria mais difícil. Espero não vos desiludir.

Índice

Subcapa

Resumo

Abstract

Agradecimentos

Índice.....I

Índice de Figuras.....III

Índice de Tabelas.....IV

1. Introdução..... 1

1.1 Relevância do modelo CLTV na actualidade..... 2

1.2 Objectivos da tese..... 3

2. Customer Lifetime Value na Banca..... 4

2.1 Customer Relationship Management (CRM): Importância para a estratégia do negócio e vantagem competitiva..... 5

2.2 O Modelo de CLTV como elemento fundamental em todo o processo de CRM na Banca.....7

2.3 Métricas tradicionais utilizadas no processo de identificação dos melhores Clientes 9

2.3.1 *Recency-Frequency-Monetary Value* (RFM)..... 10

2.3.2 *Past Customer Value* 11

2.3.3 *Share-of-Wallet* 11

2.3.4 Métricas tradicionais vs. CLTV 12

2.4 Abordagens utilizadas no processo de obtenção do CLTV 15

2.4.1 Abordagem Agregada..... 15

2.4.2 Abordagem Individual..... 16

3. Caso Prático	21
3.1 Introdução.....	21
3.2 Metodologia.....	21
3.3 Dados.....	26
3.4 Produtos utilizados no desenvolvimento do modelo de CLTV	28
3.5 Elementos envolvidos no processo de obtenção do CLTV	28
3.5.1 Rentabilidade associada aos Clientes	29
3.5.2 Potencialidade associada aos Clientes.....	30
3.5.2.1 Valor dos clientes num futuro mais recente.....	30
3.5.2.1.1 Modelos analíticos desenvolvidos.....	32
3.5.2.1.2 Parametrização utilizada no desenvolvimento dos modelos e estratégia de selecção.....	38
3.5.2.2 Valor dos clientes num futuro mais longínquo.....	41
3.5.2.2.1 Modelos analíticos de sobrevivência desenvolvidos.....	43
3.6 Obtenção do CLTV por cada um dos clientes.....	46
3.7 Resultados Finais.....	48
4. Conclusões e Investigação Futura.....	52
4.1 Conclusões.....	52
4.2 Investigação Futura	54
5. Referências Bibliográficas	57
Anexo 1. Caracterização dos Dados.....	59
Anexo 2. Obtenção do Valor de Cliente num futuro considerado mais recente (RFCV).....	60
Anexo 3. Obtenção do Valor de Cliente num futuro considerado mais longínquo (FCV).....	69

Índice de Figuras

Figura 1. Retorno que o Banco tem perante o ciclo de vida dos Clientes.....	9
Figura 2. Metodologia adoptada no desenvolvimento do Trabalho de Projecto.....	22
Figura 3. Horizontes Temporais de cada componente que engloba o modelo de CLTV.....	23
Figura 4. Modelos Analíticos desenvolvidos para a modelação do CLTV.....	25
Figura 5. Processo de obtenção do valor de cada cliente em Setembro de 2009.....	47
Figura 6. <i>Score</i> Médio dos Modelos Preditivos desenvolvidos para a obtenção do RFCV.....	68
Figura 7. Sintaxe utilizada no desenvolvimento dos modelos de Análise de Sobrevivência.....	69

Índice de Tabelas

Tabela 1. Desvantagens associadas às métricas tradicionais utilizadas na medição da lealdade e valor dos clientes	13
Tabela 2. Vantagens associadas à métrica CLTV.....	14
Tabela 3. Análise Descritiva das Características dos Clientes.....	59
Tabela 4. Técnicas utilizadas nos modelos vencedores de <i>Cross Sell</i> , <i>Up Sell</i> e Abandono.....	60
Tabela 5. Parametrização utilizada na modelação dos modelos vencedores de <i>Cross Sell</i> , <i>Up Sell</i> e Abandono.....	61
Tabela 6. Estatísticas de avaliação da qualidade dos modelos preditivos de <i>Cross Sell</i> , <i>Up Sell</i> e Abandono desenvolvidos.....	66
Tabela 7. Análise Descritiva dos modelos de <i>Cross Sell</i> , <i>Up Sell</i> e Abandono desenvolvidos e considerados como vencedores.....	67
Tabela 8. Análise Descritiva das componentes utilizadas na obtenção do modelo de CLTV.....	48
Tabela 9. Proporção de Clientes tendo em conta a rentabilidade, potencialidade e valor dos mesmos.....	51

1. Introdução

Existe actualmente uma grande agitação económica, empresarial e mesmo social, o que leva a que seja dada uma enorme importância à obtenção de informação atempada e verdadeira. Toda esta agitação pode tornar as organizações, em particular as instituições financeiras, cada vez mais vulneráveis, mas também pode proporcionar boas oportunidades caso estas instituições saibam lidar com o grande volume de dados existentes.

Sendo os clientes um dos principais pilares para o sucesso, competitividade, e rentabilidade das instituições financeiras, torna-se necessário avaliar novas formas de compreender, e analisar, as necessidades e expectativas dos clientes. É neste contexto que a satisfação dos clientes assume especial importância de forma a contribuir para a sobrevivência das organizações e seu desenvolvimento - a construção de fortes relações entre as instituições e os seus clientes é uma forma de ganhar vantagem competitiva (McKenna, 1991; Reichheld, 1993).

Ao longo do tempo, as instituições financeiras têm cada vez mais apostado na implementação de processos e métricas, que permitam um modelo de gestão de relacionamento com os clientes, e todo um conhecimento do próprio cliente, que potencie, ou aumente, o valor e sustentabilidade do cliente no banco a longo prazo. Este foco nos clientes tem-se verificado cada vez mais, uma vez que as instituições compreenderam que é o cliente quem dita as regras de consumo, existindo, por isso, uma maior orientação para os próprios clientes e não tanto aos produtos (ao contrário do que sucedia anteriormente). Assim, existe uma tendência para as instituições reconhecerem que o diálogo com os clientes tornou-se uma conversa de igual para igual, uma vez que existe um fácil acesso à informação, o que leva a que as instituições obtenham informações sobre os clientes e vice-versa (Prahalad e Ramaswamy, 2002).

Face a toda a concorrência e competitividade existente actualmente, as instituições financeiras enfrentam alguns desafios no que diz respeito à forma de como manter clientes fiéis, mais duradouros e rentáveis, já que a dinâmica dos critérios de fidelidade dos clientes têm sofrido alterações face às exigências com que as instituições se deparam. De modo a contrariar esta tendência, as instituições não podem somente estar interessadas em realizar a venda de produtos ou serviços a qualquer custo, mas sim em detectar quais os clientes considerados mais rentáveis, através da aquisição de novos produtos e do aumento do grau de

envolvimento dos clientes na instituição. Surge então o conceito de *Customer Lifetime Value* (CLTV), que tem vindo a ganhar cada vez mais importância.

1.1 Relevância do modelo CLTV na actualidade

De modo a que haja um crescimento das organizações, e muito em especial das instituições financeiras, é importante assegurar o retorno de investimento (ROI). Para tal, as instituições têm de se concentrar, quer em maximizar as receitas (proveitos), quer em controlar ou mesmo minimizar os custos associados (Kumar, 2006). Tendo em conta esta dinâmica, é importante a construção de estratégias que permitam gerir os clientes de uma forma eficaz e audaz, bem como fidelizar os clientes com maior valor associado. Ao pensarmos nestas problemáticas surge um conjunto de questões às quais é necessário, e importante, dar resposta. O que é o valor de um cliente? Qual o valor de cada cliente em cada momento do seu ciclo de vida na organização? Qual o valor futuro de cada cliente? Qual a melhor métrica para medir o valor futuro dos clientes nas organizações? Estas são algumas questões que as organizações enfrentam ao pretenderem identificar, e seleccionar, os clientes com maior valor futuro. Como resposta a estas questões surge o conceito de *Customer Lifetime Value* (CLTV).

O CLTV é um conceito que permite adquirir o conhecimento sobre os clientes ao nível da sua rentabilidade e valor futuro na organização, sendo por isso uma peça fundamental para o sucesso desta. Através do cálculo do CLTV para cada um dos clientes, é possível ordenar os mesmos tendo em conta a sua contribuição para o valor e rentabilidade (Kumar e Reinartz, 2006), sendo por isso uma estratégia essencial para maximizar o ciclo de vida dos clientes na organização (Kumar, 2007).

Ao longo do tempo, a utilização e importância do CLTV tem sido crescente no sector da Banca, bem como em diversos outros sectores, isto porque, permite perceber qual o valor actual (rentabilidade até ao momento presente) de cada cliente e qual o seu valor futuro (potencialidade), ou seja, qual o valor do cliente ao longo de todo o seu ciclo na organização. Como resultado do modelo de CLTV obtém-se uma identificação do valor de cada cliente, o que, como acção futura, permitirá agrupar os mesmos por segmento, tendo em conta tais valores. Para cada segmento é então possível desenvolver estratégias que permitam maximizar o valor do cliente ao longo do tempo.

1.2 Objectivos da tese

O principal objectivo deste projecto consistiu na determinação e análise do valor associado a cada cliente existente num banco, com o intuito de identificar e seleccionar os clientes considerados como sendo os mais rentáveis e de maior valor futuro. Para tal, procedeu-se ao desenvolvimento e implementação do modelo de *Customer Lifetime Value* (CLTV) probabilístico. O modelo de CLTV foi desenvolvido para clientes com posse de: (1) créditos (Crédito Habitação, Crédito Pessoal e Crédito Especializado); (2) recursos (Seguros, Fundos, Depósitos a Prazo e Poupanças) e (3) serviços (Depósitos a Ordem e Cartões de Crédito). Estes produtos são considerados pelo banco como sendo aqueles cuja rentabilidade é superior.

Após o desenvolvimento do modelo de CLTV é possível perceber quais os clientes com créditos, recursos, e serviços associados que tenham maior valor. Como resultado do desenvolvimento e implementação deste modelo o banco terá diversos benefícios associados: (1) investimentos em marketing e pós-venda mais eficientes (existindo assim uma gestão mais orientada para o perfil de cada valor de cliente e a priorização das acções de CRM); (2) aumento da rentabilidade do banco (já que o banco estabelecerá acções sobre esses clientes, adaptadas a cada um dos perfis de valor de cliente); (3) maior duração do ciclo de vida do cliente dentro do próprio banco, e (3) redução de custos de marketing. Assim, a partir da identificação do valor dos clientes é possível perceber quais os clientes mais rentáveis e de maior potencial, o que permite uma atribuição de níveis de prioridade distintos a cada um dos clientes e uma maior duração do ciclo de vida dos mesmos dentro do banco.

Com vista ao cálculo do CLTV no banco foi necessário efectuar o cálculo e a análise de dois aspectos fundamentais: (1) rentabilidade dos clientes, que consiste na obtenção do valor actual (*Past Customer Value*) de negócio para cada um dos clientes e (2) potencialidade futura associada a cada cliente (*Future Customer Value*). O cálculo da rentabilidade dos clientes será dado pela diferença entre a receita e os custos que podem ser associados a cada um dos clientes e produtos (margens), considerando a taxa de juro existente no período, de forma a medir o valor real de negócio associado a cada cliente. Já a potencialidade de cada um dos clientes terá como *input* a conjugação: (1) das propensões (*scores*) associadas a diversos modelos preditivos (tais como modelos de *cross/up sell*, modelos de abandono de cliente e modelos de análise de sobrevivência); (2) das margens previstas associadas aos produtos e/ou serviços que cada cliente possui e (3) das taxa de juro prevista. Cada um dos elementos constituintes do modelo de CLTV, bem como a forma de obtenção dos mesmos, será posteriormente definido e analisado em pormenor.

2. Customer Lifetime Value na Banca

A missão do marketing é cada vez mais complexa, já que lida com todo um conjunto de questões às quais é preciso dar resposta. Kumar, Ramani e Bohling (2004) sugerem algumas destas questões: (1) como satisfazer as necessidades e desejos dos clientes?; (2) como captar, reter e fidelizar clientes?; (3) como ganhar vantagem competitiva face à concorrência? e (4) como obter rentabilidade ao longo do ciclo de vida dos clientes?. Todas estas questões são de difícil resposta, já que os clientes são cada vez mais dinâmicos ao longo do seu ciclo de vida, uma vez que um cliente tem diferentes expectativas, vontades e gostos.

Tendo em conta todas as agitações que existem actualmente na Banca, o comportamento dinâmico dos clientes, e as próprias características do mercado, é necessário a obtenção de informações atempadas e verdadeiras que permitam encontrar padrões, conhecimento, tendências, relações, e associações, que sejam úteis para a construção de estratégias (a partir de dados e informação) que permitam uma rápida alteração e evolução do negócio.

Desde há muitos anos que a Banca dispõe de sistemas operativos que permitem a captação e gestão de dados transaccionais dos clientes. Posteriormente, surgiram os chamados *data warehouses* (DW) corporativos, que visam o armazenamento de dados relativos às actividades do banco, e disponibiliza-los para todo o banco. Face à necessidade de existirem sub-conjuntos de dados de um DW por áreas específicas (por exemplo, de Marketing) surgiu mais recentemente os chamados *Data Marts* (DM), que disponibilizam a informação comportamental dos clientes ao longo do tempo.

A estratégia não é mais do que a “espinha dorsal” que orienta toda a acção de uma empresa, muito em especial dos bancos. Cada estratégia é composta, e originada, por tácticas que, apesar de poderem ser independentes do contexto e do tempo, estão alinhadas com a direcção estratégica. Outro aspecto bastante importante de referir é que as estratégias a adoptar, por exemplo na Banca, devem ser baseadas num conhecimento profundo dos clientes. Surge assim os conceitos de *Customer Intelligence* (CI) e *Customer Relationship Management* (CRM), que dão suporte a uma estratégia abrangente de relacionamento com os clientes.

O CI é um conceito que permite captar, armazenar, processar, aceder, organizar, e analisar/modelar, os dados sobre os clientes, ou seja, permite compreender, entender e prever os comportamentos dos clientes, com vista a desenvolver uma estratégia eficiente e eficaz de relacionamento com os clientes em todas as áreas funcionais. Desta forma será possível: (1) identificar os melhores clientes do banco; (2) reter e atrair os clientes mais rentáveis com

produtos e serviços altamente personalizados; (3) criar perfis de cada cliente; (4) descobrir, e planejar, como comunicar com os clientes, ou seja, escolher qual o tipo de canal mais adequado para a comunicação; (5) antecipar e dirigir as necessidades dos clientes, e (6) otimizar campanhas de marketing através de vários canais de comunicação.

2.1 Customer Relationship Management (CRM): Importância para a estratégia do negócio e vantagem competitiva

Desde muito cedo que se verificou que o papel dos clientes é essencial para a performance das organizações, de tal modo que, quando as relações destas com os seus clientes se fortalecem, então os lucros das organizações aumentam (Reichheld, 1990), sendo esta uma forma de ganhar vantagem competitiva sobre os concorrentes (Mc Kenna, 1991; Reichheld, 1993).

Berry (1933) introduziu o conceito de CRM na literatura, tendo-o definido como sendo a “forma de atrair, manter e melhorar o relacionamento com os clientes, tendo em conta os vários serviços existentes numa organização”. Assim, a origem do CRM começou com a necessidade das empresas de tratarem os clientes da melhor forma possível, e compreenderem as suas expectativas e necessidades, com vista a alcançar uma vantagem competitiva e sustentável. Isto porque clientes leais originam menos custos (Jarvis e Wilcox, 1977), visto que o custo de reter clientes frequentes é muito menor do que o custo de adquirir novos clientes (Sheth e Parvatiyar, 1995).

Ao longo do tempo têm surgido variações da definição de CRM, já que existem várias formas e perspectivas de definir tal conceito. Para Hawkes (2000) a grande essência do CRM consiste no entendimento da rentabilidade associada a cada cliente e na retenção dos clientes com maior valor. Kotler (2000) defende que as empresas mais eficientes e “inteligentes” são aquelas que acabam por desenvolver uma maior confiança no que diz respeito ao relacionamento a longo prazo com os consumidores, distribuidores, revendedores e fornecedores. Já Limeira (2003) define CRM como sendo uma estratégia de marketing cujo objectivo consiste em construir uma relação de longo-prazo baseada em confiança, parceria, colaboração, investimentos e benefícios mútuos entre o cliente e a organização.

Genericamente, e tendo em mente todas as alterações que decorreram ao longo do tempo, podemos definir CRM como sendo todo um conjunto abrangente de tecnologias e processos utilizados na gestão das relações com os clientes actuais e futuros, independentemente do canal de comunicação, de forma a reter informações úteis sobre os mesmos, com vista à

compreensão das suas necessidades e desejos. Como tal, é um sistema de suporte à decisão cujo objectivo principal consiste em otimizar a satisfação dos clientes, e otimizar a eficiência e os ganhos existentes numa organização. Através da implementação do CRM, tal como Jutla, Craig e Bodorik (2001) afirmam, verifica-se uma maior lealdade e confiança dos clientes, e consequentemente são obtidas maiores receitas e menores custos - já que todos os esforços são direccionados segundo as características, necessidades e desejos dos clientes.

O sucesso do CRM foca-se no entendimento das necessidades, e desejos dos clientes, através da integração, e interacção, de estratégias de negócio, pessoas, tecnologias de informação e processos (Fox e Stead, 2001).

Apesar de existirem várias formas de definir o conceito de CRM, Ferrão (2003) sugere a existência três tipos de CRM: (1) CRM Operacional; (2) CRM Analítico e (3) CRM Colaborativo. O CRM Operacional diz respeito a todo um conjunto de aplicações/canais de contacto que permitam todo um processo eficiente de relacionamento e de interacção dos clientes com a empresa, tais como, a automatização das tarefas de marketing¹, automatização da força de vendas², e automatização de atendimentos aos clientes. Segundo Bayer (2002) o CRM Analítico é constituído por um conjunto de aplicações, processos e técnicas que permitem a construção e obtenção de modelos³, modelos estes que permitem analisar os dados dos clientes, com o intuito de conhecer, otimizar e prever o comportamento dos clientes, o que possibilita todo um ajuste às estratégias de diferenciação dos clientes. A este tipo de CRM está associado o *Data Mining* (conceito apresentado posteriormente).

O CRM Colaborativo consiste em utilizar todo um conjunto de serviços, tais como, e-mail, fax, cartas, conferências, ou mesmo através de interacções directas, com o intuito de facilitar, e personalizar, as interacções entre clientes e empresas. Neste tipo de CRM a principal preocupação está em manter todo um conjunto de canais de contacto que permitam obter fluxos de dados entre os clientes e as organizações.

Este trabalho tem como base o CRM Analítico, já que foi feito todo um processo de análise de dados associados aos clientes, recorrendo a técnicas de *Data Mining*, com vista à obtenção do valor de cada cliente (CLTV) existente no banco.

¹ Este tipo de automatização engloba tarefas de gestão de campanhas e personalização de perfis.

² A automatização das forças de vendas permite toda uma redução da duração dos ciclos dos processos relacionados com os clientes.

³ Tais como segmentação, *customer lifetime value*, modelos de *cross* e *up sell*, modelos de aquisição de novos clientes, modelos de risco (incumprimentos), entre outros.

Com todo o conhecimento sobre os conceitos, aplicações e vantagens do CRM existente actualmente, as empresas, e muito em particular os bancos, têm apostado cada vez mais na implementação de soluções de CRM dinâmicas.

2.2 O Modelo de CLTV como elemento fundamental em todo o processo de CRM na Banca

As necessidades, características e desejos dos clientes existentes nos bancos são aspectos extremamente dinâmicos ao longo do tempo, já que lidam com questões de expectativas, gostos, vontades, vivências e experiências dos próprios clientes, mas também porque estamos perante toda uma agitação económica e social. Como tal, para que se consiga reagir a todas estas variáveis extremamente voláteis, os bancos têm de ser capazes de centralizar toda esta informação que recebem constantemente, e com isto saber retirar conclusões eficazes e adequadas sobre os clientes, e dar resposta, em termos de oferta de serviços ou produtos, às necessidades e expectativas dos mesmos. Uma forma de o conseguir é recorrendo à implementação de processos, e métricas, que permitam todo um processo de gestão de relacionamento com os clientes.

Com o desenvolvimento e implementação de soluções de CRM na Banca surgem diversos benefícios associados, tais como:

- acções estratégicas sobre os clientes de acordo com as suas necessidades, expectativas e desejos, o que permite uma maior satisfação;
- maior fidelização dos clientes;
- clientes menos sensíveis ao preço, visto estarem mais satisfeitos;
- campanhas mais focadas e direccionadas para alvos mais específicos e com maior retorno;
- menor propensão para o abandono;
- obtenção de maiores receitas e menores custos;
- identificação dos clientes mais rentáveis e de maior valor para o banco.

Face às constantes aquisições, evoluções, fusões, ou mesmo liquidações, que existem no mercado financeiro, os bancos têm, cada vez mais, de ser capazes de adaptar novas soluções que permitam todo um forte posicionamento estratégico e competitivo no Mercado. Uma forma de o conseguir é através da existência de uma plataforma tecnológica que possibilite

uma integração, processamento e disponibilização, dos dados dos clientes, e consequentemente todo um processo de análise de dados, descoberta de tendências, padrões e associações. Todo este processo ocorre com o intuito de maximizar a satisfação, lealdade, fidelização e rentabilidade associada a cada cliente. Toda esta análise assume especial relevância, uma vez que a lealdade, e os lucros, são os efeitos a longo prazo da satisfação dos clientes (Oliver, 1997).

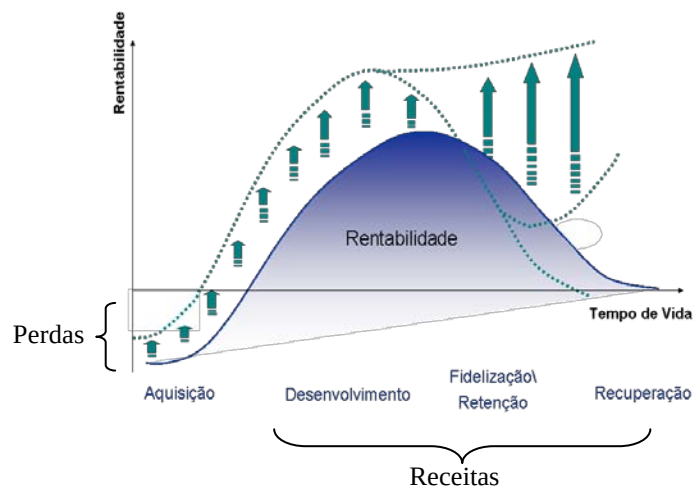
A selecção dos clientes com maior valor associado deve ser feita de forma a seleccionar os clientes cujo retorno do investimento, efectuado pelo banco, seja superior e rentável ao longo do tempo – sendo fundamental a obtenção do CLTV. Esta métrica tem vindo a assumir especial relevância na área do marketing e CRM [(Blattberg e Deighton, 1996); (Blattberg, Getz e Thomas, 2001); (Winer, 2001); (Rust, Zeithaml e Lemon, 2000)].

O cálculo do valor associado a cada cliente é uma questão considerada chave para avaliar, e determinar, quais os ganhos futuros associados, e com isto seleccionar os clientes considerados mais rentáveis para o banco. De salientar que, identificando tal valor, é útil agregar os clientes por segmentos de valor, e assim desenvolver estratégias que permitam maximizar o valor associado a cada um dos segmentos existentes. Assim, Reinartz e Kumar (2000), sugerem que toda esta análise de rentabilidade associada aos clientes deve ser feita o mais cedo possível no seu ciclo de vida. Com base neste ciclo é possível construir e delinear estratégias de acordo com o tempo de permanência no banco (ciclo de vida). Kalakota e Robinson (2001) sugerem que o ciclo de vida dos clientes traduz-se essencialmente em 3 fases distintas: (1) aquisição; (2) desenvolvimento da relação com a organização e (3) retenção. Todas estas fases têm impactos directos no cliente e no próprio banco, e estão muito interligadas e correlacionadas entre si.

Surgem assim três questões fundamentais para a obtenção de uma estratégia de CRM eficiente nos bancos: (1) captação de novos clientes, através a oferta de produtos e/ou serviços (campanhas de aquisição de clientes e campanhas de *cross sell*); (2) maximização do potencial dos clientes actualmente existentes no banco através das ofertas correctas para cada um (campanhas de *cross sell* e *up sell*) e (3) retenção dos clientes com maior valor para o banco.

A figura 1 representa o tipo de retorno (perdas ou receitas) que o banco tem, tendo em conta o ciclo de vida dos clientes.

Figura 1. Retorno que o Banco tem perante o ciclo de vida dos Clientes.



Durante a fase de aquisição/captação de clientes existem diversos custos associados. Por este motivo, e de modo a que exista rentabilidade, é de extrema importância desenvolver, nos primeiros meses de vinculação, campanhas de estreitamento da relação (por exemplo, campanhas de “Boas Vindas”) com os clientes e campanhas de desenvolvimento da relação existente com o próprio banco, através da ofertas de novos produtos ou serviços (*cross sell*).

Após todo o processo de captação de clientes, interessa aumentar o grau de relação/envolvimento existente entre estes e o banco, de forma a obter maior rentabilidade. Uma forma de o conseguir, para além da venda de novos produtos ou serviços, é através da venda destes com maior valor associado, de modo a aumentar as receitas inerentes (*up sell*).

Com o passar do tempo de permanência dos clientes no banco é importante efectuar todo um conjunto de campanhas: (1) de retenção e de fidelização, com o intuito de evitar, ou diminuir, a probabilidade de abandono, bem como evitar a diminuição da rentabilidade associada aos clientes e (2) de recuperação de clientes, ou seja, campanhas que permitam a captação de clientes considerados como sendo “bons” e que já abandonaram o banco.

2.3 Métricas tradicionais utilizadas no processo de identificação dos melhores Clientes

Venkatesan e Kumar (2004) consideram que, para que haja sucesso na implementação de soluções de CRM, é de extrema importância o desenvolvimento de abordagens que permitam seleccionar os melhores clientes, já que a percentagem destes que tende a produzir lucros significativos é relativamente pequena. Assim, é de enorme importância a identificação do valor e lealdade dos clientes.

Segundo Reinartz e Kumar (2003) existem três métricas (ditas tradicionais) que possibilitam medir a lealdade e o valor associado a cada cliente: (1) *Recency-Frequency-Monetary Value*; (2) *Past Customer Value* e (3) *Share of Wallet*. A partir destas métricas os bancos conseguem priorizar os clientes com base no seu valor e rentabilidade futura.

2.3.1 *Recency-Frequency-Monetary Value (RFM)*

A métrica RFM é um dos métodos mais utilizados para identificar quais os melhores clientes existentes num banco, tendo em conta três variáveis distintas: recência (R)⁴, frequência (F)⁵ e valor monetário (M)⁶.⁷ Usualmente é utilizado um processo de segmentar/dividir um conjunto de clientes em grupos heterogêneos, em relação às três variáveis anteriormente descritas. Assim, a métrica RFM baseia-se na hipótese de que o comportamento passado, no que diz respeito às aquisições (recência), movimentações (frequência) e receitas (valor monetário) geradas por cada cliente (Kumar, 2007).

O conhecimento do comportamento de cada segmento criado pode ser utilizado:

- na elaboração de actividades de marketing específicas para cada grupo de clientes;
- na identificação do melhor segmento;
- na aquisição de novos clientes, com perfil similar aos melhores;
- no desenvolvimento de novos produtos, com base nas necessidades de cada segmento;
- na diferenciação das oportunidades de *cross/up sell* no mercado;
- na melhoria da precisão de outros modelos, através da utilização do segmento como variável explicativa;
- no cálculo do valor de cada cliente.

Em suma, esta métrica possibilita a identificação dos melhores clientes face às três variáveis utilizadas (recência, frequência e valor monetário). Segundo Kumar (2007) os clientes com maior recência, frequência e valor monetário associados são aqueles com maior hipótese de resposta face às campanhas efectuadas.

⁴ A recência refere-se ao tempo decorrido desde o momento que o cliente efectuou a última acção/compra.

⁵ A frequência diz respeito à frequência de contacto entre o cliente.

⁶ O valor monetário corresponde às receitas geradas pelo cliente.

⁷ Como tal, este tipo de segmentação não utiliza muita informação para associar clientes a um determinado grupo, sendo por isso de fácil análise (o que poderá ser visto como uma vantagem).

2.3.2 Past Customer Value (PCV)

O método PCV é um método baseado na hipótese de que a rentabilidade passada dos clientes reflecte o nível de rentabilidade futura associada, ou seja, é um método que extrapola a rentabilidade dos clientes para o futuro (Bejou *et al*, 2006). Tendo em conta que os produtos e/ou serviços são comprados em diversos momentos de tempo pelos clientes, a contribuição de tais transacções efectuadas deve ser ajustada tendo em conta o valor do dinheiro (taxa de juro) ao longo do tempo. A contribuição acumulada de tais transacções, até ao momento presente, representa o valor presente de cada cliente.

A equação 1 reflecte a fórmula de cálculo do PVC para o cliente i.

$$PCV_i = \sum_{t=1}^T GC_{it} \times (1 + R)^{-t} \quad (1)$$

sendo que:

- i: Cliente i;
- R: Taxa de juro associada;
- T: Número de períodos de tempo desde o momento da aquisição/compra até ao momento da análise;
- GC_{it} : Diferença entre as receitas directas originadas pela venda dos produtos e/ou serviços e os custos directos (não incluindo impostos) com tal venda. Ou seja, reflecte a contribuição bruta das transacções efectuadas por cada cliente no período t;
- t = Período de tempo.

Em suma, esta métrica traduz a rentabilidade do banco associada a cada cliente com a venda dos produtos e/ou serviços até ao momento actual.

2.3.3 Share-of-Wallet (SOW)

Já o método SOW, ou carteira de cliente, é bastante utilizado nas instituições financeiras para identificar a preferência e lealdade dos clientes face a uma determinada família de produtos e/ou serviços. Assim, esta métrica é muito utilizada no estudo e análise do comportamento consumista, e necessidades dos clientes existentes no banco, face a um determinado produto existente numa dada família (Kumar e Reinartz, 2006).

O aumento da carteira de um dado cliente (SOW) é importante na medida que permite aumentar a quota de mercado associada a esse mesmo cliente. Uma forma de o conseguir é, por exemplo, através da oferta de novos produtos e/ou serviços (modelos de *cross-sell*) existentes no banco.

A equação 2 reflecte a fórmula de cálculo do SOW para cada cliente existente num determinado banco. A medida SOW é expressa em percentagem.

$$SWO(\%)_i = \frac{S_{ij}}{\sum_{j \in J} S_{ij}} \times 100 \quad (2)$$

sendo que:

- i: Cliente
- j: Produto ou serviço existente numa determinada família;
- J: Conjunto de todos os produtos ou serviços numa determinada família existentes no banco;
- S_{ij} : Total gasto por cada cliente com determinado produto ou serviço de uma determinada família;
- $\sum_{j \in J}$: Total gasto por cada cliente com todos os produtos ou serviços de uma determinada família.

2.3.4 Métricas tradicionais vs. CLTV

Segundo Reinartz e Kumar (2003) os métodos tradicionais apresentados (RFM, PCV e SOW) têm diversos inconvenientes associados, existindo duas desvantagens comuns: (1) não é tido em conta o facto de os clientes existirem futuramente no banco e (2) comportamento observado dos clientes é extrapolado para o futuro, o que num cenário de alterações de mercado origina diversos erros associados.

A tabela 1 reflecte os inconvenientes existentes tendo em conta o método utilizado para medir a lealdade e valor dos clientes.

Tabela 1. Desvantagens associadas às métricas tradicionais utilizadas na medição da lealdade e valor dos clientes

Método	Descrição das Desvantagens
RFM	• Agrega cada cliente a um determinado grupo de clientes homogéneo dentro de si, e o mais heterogéneo possível entre si, e não calcula qualquer pontuação (<i>score</i>) para cada um dos clientes; • Assume que as únicas variáveis essencialmente importantes para analisar o valor do cliente dizem respeito à recência, frequência e valor monetário; • Como efectuar o ponto de corte dos segmentos? Estatisticamente ou pelo conhecimento de negócio? • Orientado para a rentabilidade passada do cliente, e não tem em conta as acções futuras do cliente (não preditivo).
PCV	• Não tem em conta a aquisição futura de produtos e/ou serviços pelos clientes na obtenção do seu valor; • Não tem em conta os custos de retenção, e custos de manutenção, dos próprios clientes; • Não tem em conta a existência do cliente no banco no futuro.
SOW	• Falta de agilidade para explicar o que os clientes vão comprar a seguir, e qual a rentabilidade associada no futuro; • Não tem em conta a existência do cliente no banco no futuro; • Privilegia os produtos com maiores montantes associados ao cliente; • Foco no produto; • Apenas tem em conta os gastos com o produto e/ou serviço e não a transaccionalidade.

Tendo em conta todas as desvantagens associadas à utilização das métricas tradicionais para efectuar o cálculo do valor dos clientes, surge a necessidade de utilizar uma outra métrica – designada por CLTV.

O CLTV é uma métrica que engloba vários elementos para a obtenção do valor associado a cada cliente - receitas, custos e comportamento dos clientes (Kumar, 2007). Assim, esta

métrica engloba vários aspectos, que os chamados métodos tradicionais não têm em conta, tais como:

- custos de aquisição de clientes;
- custos de retenção de clientes;
- custo de manutenção de clientes;
- custos de marketing;
- taxa de juro existente ao longo do tempo;
- probabilidade dos clientes existirem no banco no futuro;
- probabilidade de cross/up sell.

Face a todas as componentes utilizadas no cálculo do CLTV, podemos afirmar que este tipo de métrica traz diversas vantagens associadas, comparando com os métodos tradicionais de obtenção do valor de cliente. A tabela 2 reflecte as vantagens associadas à métrica CLTV.

Tabela 2. Vantagens associadas à métrica CLTV

Método	Descrição das Vantagens
CLTV	• É obtido um <i>score</i> para cada cliente; • Não é feita uma extrapolação do valor actual do cliente para o futuro; • É considerado o facto de o cliente ser activo, ou seja, existir futuramente no banco; • É analisado todo o comportamento do cliente ao longo do tempo; • É tido em conta todo um conjunto de proveitos e custos associados a cada cliente ao longo do tempo.

Após percebermos que a métrica CLTV é conceptualmente mais indicada para efectuar o cálculo do valor de cada cliente existente no banco, é importante perceber as abordagens existentes, e possíveis de se utilizar para obter o valor de cliente.

2.4 Abordagens utilizadas no processo de obtenção do CLTV

Desde os anos 80 o cálculo do CLTV tem vindo a despertar cada vez mais atenção, podendo ser calculado em diversas indústrias, assumindo actualmente particular interesse na Banca.

Existem duas abordagens distintas que devem ser consideradas aquando do cálculo do CLTV: (1) abordagem agregada e (2) abordagem individual. Ambas as abordagens assumem a existência de 4 componentes básicas para o cálculo do CLTV (Reinartz e Kumar, 2000):

- receitas (*cross e up sell*, marketing, etc);
- custos (aquisição, marketing, produção dos produtos e/ou serviços vendidos, administrativos);
- taxa de retenção (probabilidade do cliente se manter leal ao banco);
- taxa de juro.

2.4.1 Abordagem Agregada

A abordagem agregada foi sugerida por Blattberg e Deighton (1991), posteriormente reforçada por Berger e Nasr (1998) e por Kumar, Ramani e Bohling (2004). Segundo esta abordagem o CLTV de um dado cliente deriva do valor do segmento (grupo) em que esse mesmo cliente se insere. Cada segmento é obtido de acordo com as características demográficas e com o envolvimento perante o banco.

Numa primeira fase o banco deve posicionar, e integrar, os seus clientes em determinados grupos homogéneos dentro de si, e o mais heterogéneo entre si (clusters). Numa segunda fase, é calculado o valor para cada um dos segmentos criados. Assim sendo, os clientes existentes num dado segmento têm igual valor. Esta abordagem pode ser vista como uma forma de minimizar os esforços na obtenção do valor para cada segmento. A equação 3 reflecte a forma de obter o valor de cada segmento.

$$CLTV_{Segi} = \sum_{t=1}^T \left[\frac{GC - M}{(1 + d)^t} \times R^t \right] - A \quad (3)$$

sendo que:

- i: Segmento em análise;
- t: Período de tempo;
- T: Horizonte de tempo para a obtenção do CLTV;

- GC: Margens médias⁸ obtidas com os produtos e/ou serviços que os clientes existentes em cada segmento têm;
- M: Custos médios de marketing existentes no segmento;
- d: Taxa de juro existente em cada período de tempo;
- R: Taxa de retenção do segmento ao longo do tempo;
- A: Custo médio de aquisição dos clientes existentes no segmento.

Assim, segundo esta abordagem, o valor de cada cliente é dado pelo valor associado ao segmento em que este se insere no banco, sendo esse valor calculado previamente.

2.4.2 Abordagem Individual

Na abordagem individual, cada cliente irá ter associado o seu próprio valor tendo em conta o seu ciclo de vida no banco. Existem diversas formas de modelar este valor que cada cliente tem no banco. Reinartz e Kumar (2000, 2002) sugerem que o cálculo do valor associado a cada cliente engloba duas componentes distintas, tal como a equação 4 reflecte.

$$CLTV_i = \sum_{t=1}^T \frac{[Contribuição_futura_margens_{it} - Custos_Futuros_{it}]}{(1+d)^t} - A_i =$$

$$= \left[\sum_{t=1}^T p(activo)_{it} \times \frac{AMGC_{it}}{(1+d)^t} \right] - \left[\sum_{t=1}^T M_{it} \times \left(\frac{1}{(1+d)} \right)^t \right] - A_i \quad (4)$$

sendo que:

- i: Cliente;
- t: Período de tempo;
- T: Horizonte de tempo para a obtenção do CLTV;
- $p(activo)_{it}$: Probabilidade do cliente ser activo no futuro, ou seja, probabilidade de cada cliente não abandonar o banco em cada período de tempo;
- $AMGC_{it}$: Média das margens (receitas – custos) de cada cliente por ter produtos e serviços em cada período de tempo;
- d: Taxa de juro existente em cada período de tempo;

⁸ As margens médias de cada segmento correspondem à diferença entre as receitas médias e os custos médios associados ao segmento em questão.

- M_{it} : Custos de marketing por cliente e por cada período de tempo (engloba custos de retenção e desenvolvimento da relação entre o cliente e o banco);
- A_i : Custos de aquisição de cada cliente.

Esta equação apresenta alguns inconvenientes, uma vez que a probabilidade do cliente existir no banco no futuro só é tida em conta no cálculo da contribuição futura das margens, e não nos próprios custos de marketing. Ou seja, perante esta abordagem, caso o cliente abandone o banco, então existirá sempre custos de marketing associados ao longo do tempo, o que significa que serão feitas campanhas de marketing sobre esse mesmo cliente. Assim sendo, cada uma das componentes deveria ter em conta a probabilidade do cliente existir no banco em cada período de tempo, o que não acontece (traduzindo-se numa desvantagem).

Depois de Reinartz e Kumar (2000, 2002) surgiram outros três autores, Gupta, Lehmann e Stuart (2004), que propõem uma outra forma de obter o CLTV de cada cliente (equação 5).

$$CLTV_i = \sum_{t=1}^T \left[\frac{(P_{it} - C_{it}) \times R_{it}}{(1 + d)^t} \right] - AC_i \quad (5)$$

sendo que:

- i : Cliente;
- t : Período de tempo;
- T : Horizonte de tempo para a obtenção do CLTV;
- P_{it} : Proveito gerado com cada cliente pelo pagamento de produtos e/ou serviços em cada período de tempo;
- C_{it} : Custos directos de servir, e reter, cada cliente em cada período de tempo;
- d : Taxa de juro existente em cada período de tempo;
- AC_i : Custos de aquisição de cada cliente;
- R_{it} : Probabilidade de cada cliente ser activo, ou seja, não abandonar o banco em cada período de tempo.

A equação acima definida tem em conta: (1) flutuações existentes ao longo do tempo relativamente ao valor das compras efectuadas pelo cliente; (2) os custos de servir/reter o mesmo e (3) o facto do cliente poder ser, ou não, activo em cada período de tempo. Contudo,

a equação não tem em conta a probabilidade de *cross/up sell* ou mesmo de abandono de posse de produtos.

Face a todas as desvantagens associadas às formas anteriormente apresentadas de modelar o CLTV tem surgindo, ao longo do tempo, e de acordo com a indústria onde se pretende calcular o valor de cada cliente, várias abordagens de modelação, tais como:

- Modelo de Migração de Clientes

O modelo de migração de clientes, sugerido por Dwyer (1997), não tem em conta a possibilidade de cada cliente existir no futuro, o que significa que o valor do mesmo é dado pelo cálculo da rentabilidade actual (CE) e não pela sua potencialidade (o que se traduz numa grande desvantagem).

A equação 6 reflecte o valor de um cliente (rentabilidade):

$$CE = \sum_{t=1}^T \left[\frac{MM_t C_t P_t}{(1+d)^t} \right] \quad (6)$$

sendo que:

- t: Período de tempo;
- T: Horizonte de tempo para a obtenção do CLTV;
- MM_t : Matriz que contém as probabilidades de cada cliente se mover de um segmento para um outro segmento em cada período de tempo;
- C_t : Número de Clientes existentes em cada segmento;
- P_t : Rentabilidade de cada segmento em cada período de tempo;
- d: Taxa de juro existente em cada período de tempo.

Assim, neste tipo de modelo, os clientes são previamente agrupados em segmentos e tem-se em conta, no cálculo da rentabilidade actual de cada cliente, a probabilidade do mesmo mudar o segmento a que pertence (migração do cliente entre segmentos).

- Modelo Econométrico

A utilização do modelo econométrico como forma de determinar o valor de cada cliente foi sugerida por Dràze e Bonfrer (2001). Segundo estes dois autores o CLTV pode ser definido como uma função de intervalo de tempo entre acções realizadas sobre o cliente (acções estas que provocam um aumento ou diminuição do valor associado ao mesmo). Ou seja, o principal objectivo deste tipo de modelo consiste em prever o intervalo de tempo considerado óptimo para a realização de determinada acção sobre o cliente (tais como, envio de campanhas de *cross sell*, *up sell* ou retenção). O modelo em causa pode ser definido pela equação 7.

$$CLTV_i = \frac{(1+d)^T}{(1+d)^T - P(T)} \times A(T) \quad (7)$$

sendo que:

- i: Cliente;
- T: Horizonte de tempo entre as acções realizadas sobre cada cliente, ou seja, intervalo de tempo utilizado para a obtenção do CLTV;
- A(T): Retorno obtido pelas acções realizadas sobre cada cliente no período t;
- P(T): Probabilidade de retenção de cada cliente dado o período T;
- d: Taxa de juro no horizonte T.

Este tipo de modelo não requer técnicas de modelação muito complicadas ou base de dados complexas, podendo por isso ser um modelo vantajoso para organizações de pequena dimensão.

- Modelo de Probabilidade

Este tipo de modelo tem em conta: (1) as receitas; (2) custos; (3) taxa de retenção; (4) taxa de juro existente em cada período de tempo; (5) as probabilidades desse mesmo cliente adquirir, reter, ou abandonar, determinados produtos e/ou serviços, bem como (6) a probabilidade de existir no banco (Ogden, 2009). Face a todas as componentes que integra, é um modelo que reflecte o comportamento do cliente ao longo do seu ciclo de vida, já que permite ter noção da sua rentabilidade actual (*Past Customer Value* (PCV)) e da sua potencialidade, isto é, permite ter em conta o futuro próximo (*Recent Future Customer Value*

(RFCV)) ou o futuro considerado mais longínquo (*Future Customer Value* (FCV)) de cada cliente na obtenção do valor do mesmo.

Existem várias formas de integrar estas componentes com o intuito de obter o valor de cada cliente. Neste trabalho de projecto propomos a utilização da equação 8 como forma de obter o valor de cliente.

$$CLTV_i = PCV_i + RFCV_i + FCV_i \quad (8)$$

Tendo em conta todas as abordagens definidas, e os seus elementos constituintes, podemos concluir que um modelo de probabilidades é mais adequado, e que por isso mesmo será o utilizado no caso prático desenvolvido e apresentado posteriormente. A utilização destes elementos constituintes fornece uma previsão realista, e robusta, do valor associado aos clientes. De salientar que a partir do desenvolvimento, e implementação, deste tipo de modelo, é possível obter um *score* que traduza o valor de cada um dos clientes existentes no banco, *score* este que resulta da uma análise comportamental ao longo do ciclo de vida dos mesmos (potencialidade e rentabilidade futura). Assim, hierarquizando o *score* (valor) obtido é possível identificar e seleccionar os clientes com maior rentabilidade e potencialidade, e por sua vez criar estratégias específicas que permitam maximizar o ciclo de vida no banco.

3. Caso Prático

3.1 Introdução

O principal objectivo deste projecto consistiu na determinação e análise do valor associado a cada cliente existente num banco, de modo a permitir uma análise de quais os clientes considerados mais rentáveis e de maior valor futuro.

Os dados utilizados são dados reais pertencentes a um banco, disponibilizados pelo SAS Portugal, e cujo nome por motivos de confidencialidade não será divulgado.

O modelo desenvolvido, e implementado, traduz-se num modelo de *Customer Lifetime Value* (CLTV) probabilístico. A partir deste modelo foi possível identificar quais os clientes com maior valor associado, e com isto tornar os investimentos em marketing e pós-venda mais eficientes, o que origina um aumento da rentabilidade do banco (uma vez que este estabelece acções adaptadas ao perfil de cada cliente) e a uma maior duração do ciclo de vida do cliente dentro do próprio banco.

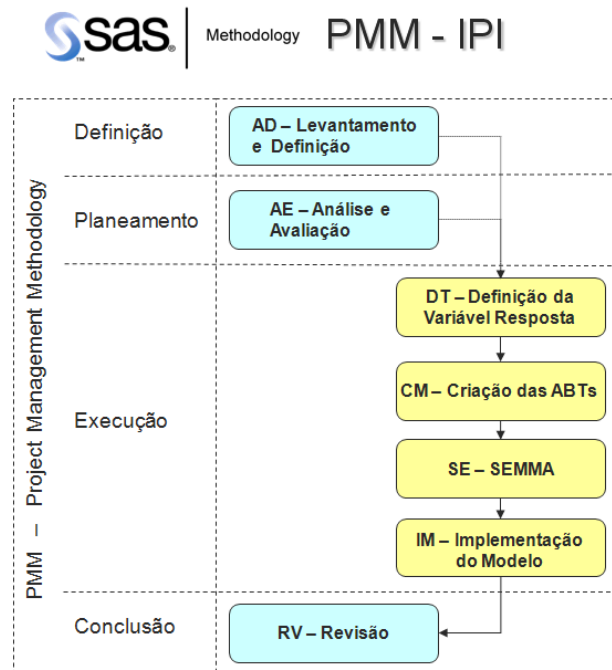
O objecto em estudo consistiu num conjunto de clientes de retalho bancário com as seguintes características no banco: (1) clientes particulares; (2) clientes não colaboradores; (3) clientes não falecidos e (4) clientes considerados activos⁹.

3.2 Metodologia

A metodologia utilizada no desenvolvimento deste projecto baseou-se na metodologia que o SAS utiliza para o desenvolvimento de projectos de *Data Mining* – Metodologia “Implementação de Plataformas Inteligentes” (IPI). Esta metodologia é constituída por diversas fases, sendo que este projecto foi desenvolvido de acordo com sete dessas fases (figura 2). Cada uma destas fases tem características e funções distintas.

⁹ Um cliente é considerado activo caso nos últimos 3 meses tenha, no mínimo, 5 movimentos efectuados ou tenha um volume de negócios superior a 100 euros.

Figura 2. Metodologia adoptada no desenvolvimento do Trabalho de Projecto



Fonte: SAS® Intelligence Platform Implementation Methodology by SAS Institute Inc., Cary, NC, USA, 2006.

Fase I: Levantamento e definição

Na primeira fase foi efectuado todo um levantamento e definição (AD) da situação actual do banco, de modo a definir claramente quais os objectivos a alcançar, e qual a estratégia a adoptar no processo de obtenção do valor de cada cliente. Para que seja possível determinar este valor é necessário efectuar duas grandes etapas: (1) cálculo da rentabilidade actual de cada cliente e (2) cálculo da potencialidade futura do mesmo, visto que o valor deste será dado pela equação 8 anteriormente definida.

Fase II: Análise e avaliação

Na fase de análise e avaliação (AE) foram avaliados os dados existentes, quanto ao seu histórico, e possíveis transformações dos mesmos, tendo-se optado por um período de desenvolvimento de Julho de 2007 a Setembro de 2009. A figura 3 reflecte os horizontes temporais utilizados no cálculo de cada componente que engloba o modelo de CLTV.

Figura 3. Horizontes Temporais de cada componente que engloba o modelo de CLTV.



Nesta fase foi, ainda, realizada a escolha do software a utilizar, tendo-se optado por:

- (1) *SAS Base*: manipulação, transformação de dados, obtenção da rentabilidade e potencialidade associada a cada cliente e processo de criação do modelo de CLTV através de código SAS;
- (2) *SAS Enterprise Miner*: desenvolvimento dos modelos preditivos e modelos de análise de sobrevivência, cujos resultados obtidos são utilizados como *input* no cálculo da potencialidade futura associada a cada cliente.

Fase III: Definição da variável resposta associada a cada componente que servirá de *input* para o cálculo do valor de cada cliente

Nesta fase procedeu-se à definição das variáveis resposta (DT) de cada uma das componentes utilizadas como *input* para o cálculo do valor dos clientes:

- (1) *Past Customer Value*: cuja variável resposta retrata a rentabilidade actual dos clientes;
- (2) *Future Recent Customer Value*: em que a variável resposta traduz a potencialidade futura nos próximos 3 meses de cada cliente;
- (3) *Future Customer Value*: variável resposta que traduz a potencialidade futura de cada cliente existente no banco durante os 9 meses seguintes.

A definição de cada variável resposta associada a cada componente será apresentada no tópico 3.5.

Fase IV: Definição da *Analytical Base Table*

A *Analytical Base Table* (ABT) é uma tabela, de desenvolvimento e aferição dos modelos analíticos, com todas as variáveis de entrada necessárias para o cálculo de cada componente utilizada no processo de obtenção do modelo de CLTV. Cada registo existente na ABT representa um cliente e contém os dados que descrevem o seu comportamento.

Assim, nesta etapa foram preparadas as tabelas analíticas com os dados e variáveis relevantes para a obtenção de cada componente utilizada no processo de obtenção do valor de cada um dos clientes.

Fase V: Desenvolvimento dos modelos analíticos utilizados como *input* do cálculo do CLTV

A fase V consistiu na realização de modelos analíticos que permitem a obtenção do valor futuro (recente e longínquo) de cada um dos clientes. Como referido anteriormente o valor associado a cada cliente existente no banco foi obtido tendo em conta o cálculo de três elementos distintos: (1) PCV; (2) RFCV e (3) FCV.

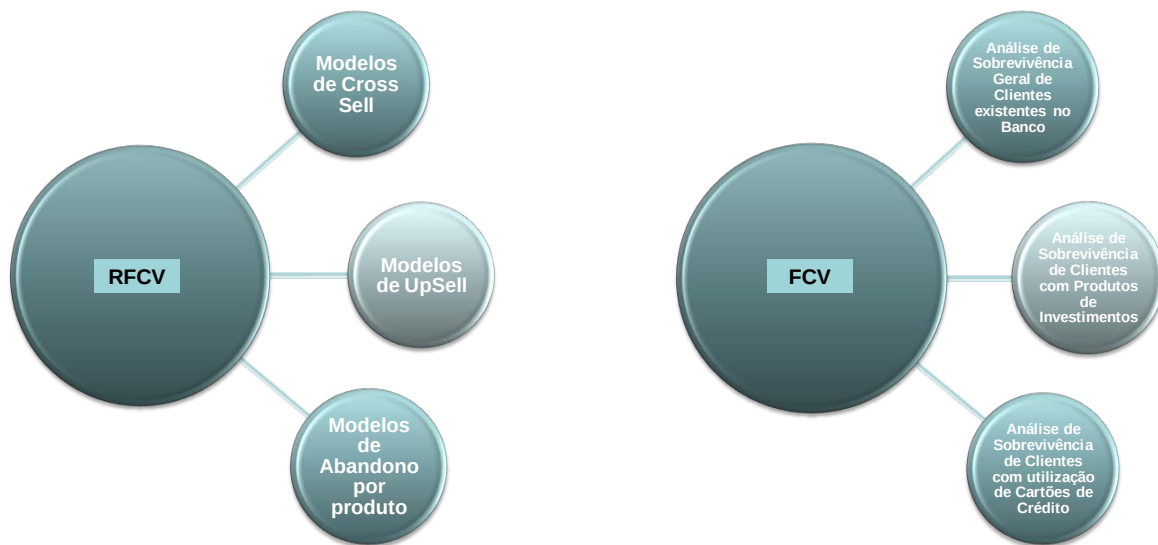
O cálculo do PCV foi obtido através da análise do comportamento do cliente até ao momento presente (Setembro de 2008), não sendo por isso necessária qualquer extrapolação para o futuro.

Já o RFCV foi calculado prevendo o valor que cada cliente terá nos próximos 3 meses (futuro este considerado como sendo mais próximo), ou seja, de Outubro a Dezembro de 2008. Para tal, foram desenvolvidos e implementados diversos modelos preditivos (*cross sell*, *up sell* e abandono), consoante os produtos utilizados na modelação do CLTV, com o intuito de utilizar os *scores* de tais modelos analíticos na fórmula de obtenção do RFCV.

O futuro considerado mais longínquo (FCV) de cada cliente foi calculado entre Janeiro e Setembro de 2009, tendo-se utilizado como um dos elementos constituintes as probabilidades obtidas através do desenvolvimento e implementação de modelos de análise de sobrevivência.

A figura 4 traduz os modelos analíticos desenvolvidos para a modelação do CLTV. No tópico 3.4 encontram-se descritos os produtos utilizados na modelação do CLTV.

Figura 4. Modelos Analíticos desenvolvidos para a modelação do CLTV.



Os modelos analíticos em causa foram desenvolvidos através do processo SEMMA, processo este que reflecte a forma de desenvolver modelos de *Data Mining* utilizando o software SAS. O processo SEMMA é constituído por cinco estágios distintos (*Sample*¹⁰, *Explore*¹¹, *Modify*¹², *Model*¹³ e *Assess*¹⁴), sendo que cada estágio tem funções específicas distintas e é guiado por uma sequência lógica do processo de actividades.

Fase VI: Desenvolvimento e Implementação do Modelo de CLTV (IM)

Na sexta fase foi realizada o desenvolvimento e implementação do modelo de CLTV, através da conjugação das componentes obtidas na fase anterior: (1) PCV; (2) RFCV e (3) FCV. Este desenvolvimento, e conjugação, foram realizados através de código SAS.

A disponibilização do modelo de CLTV em produção foi feita recorrendo a código SAS, código este que foi incluído nos mecanismos rotinados do banco. Desta forma, um cliente que satisfaça as condições do desenvolvimento do modelo de CLTV, será scorizado, o que implicitamente significa que lhe será atribuído um valor. Este processo será feito periodicamente, revisando os *scores* associados aos clientes.

¹⁰ Estágio que permite efectuar a criação de tabelas de *input*, Partição e Amostragem.

¹¹ No estágio *Explore* é possível realizar uma análise exploratória dos dados, de forma a identificar tendências e uma melhor compreensão dos dados, bem como efectuar selecção de variáveis.

¹² Estágio que permite a criação de variáveis, modificação dos dados e transformação de variáveis.

¹³ Onde é possível gerar modelos analíticos que respondam à questão que se pretende analisar;

¹⁴ Estágio que permite uma avaliação de resultados e comparação dos modelos, de forma a identificar e seleccionar o melhor modelo.

Fase VII: Revisão (RV).

Na última fase foi efectuada toda uma revisão, quer do processo de desenvolvimento, como de implementação do modelo de CLTV. Esta revisão mede a eficiência da metodologia utilizada no projecto, do suporte à gestão, do impacto na organização, e das ferramentas e técnicas lançadas durante a execução do projecto.

3.3 Dados

O modelo de CLTV probabilístico foi desenvolvido para 28881 clientes, todos eles considerados activos, não colaboradores e não falecidos em Setembro de 2008 (mês considerado como sendo o actual na análise).

Antes de proceder ao desenvolvimento do modelo de CLTV interessou perceber algumas das características dos clientes, no que diz respeito a características sócio-demográficas (sexo, idade e antiguidade dos clientes) como em relação ao produto bancário que os mesmos têm associado. A tabela 3, existente no anexo 1, ilustra algumas estatísticas descritivas dos dados em causa.

Os clientes cujo comportamento foi analisado têm idades compreendidas entre os 10 e os 99 anos, sendo que 50% dos clientes ordenados segundo a idade têm até 54 anos. A idade média dos clientes é de cerca de 55 anos, podendo haver um desvio em termos de idade de 14 anos, o que significa a existência de clientes maioritariamente em idade activa e no final desta.

Em relação ao sexo dos clientes utilizados na obtenção do CLTV concluímos que cerca de 57% dos casos dizem respeito a indivíduos do sexo masculino e 43% do sexo feminino.

No que diz respeito à antiguidade de cliente podemos concluir que os sujeitos analisados são considerados como sendo relativamente recentes no banco, visto que têm, no mínimo, 10 meses de antiguidade e no máximo 64 meses (cerca de 5 anos) de antiguidade, sendo a antiguidade média dos clientes existentes no banco de cerca de 17 meses.

Depois de analisar os clientes existentes no banco quanto à idade, sexo, e antiguidade interessou analisá-los quanto ao produto bancário em termos de número de produtos, montante total de recursos, montante total de activos e saldo médio dos depósitos à ordem efectuados nos últimos 12 meses.

Os clientes analisados possuem, nos últimos 12 meses, um mínimo de 1 produto e no máximo 15 produtos distintos, sendo que 50% dos dados ordenados têm até 5 produtos distintos. O número médio de produtos que os clientes adquiriram nos últimos 12 meses é de

6, podendo ocorrer um desvio de 2 produtos, o que significa que os clientes em estudo possuíam, em média, 4 a 8 produtos distintos.

Relativamente ao montante total de recursos¹⁵ que os clientes possuem nos últimos 12 meses, podemos concluir que existe uma grande dispersão, sendo o valor mínimo de 0 euros, e o valor máximo de cerca de 8362896 euros. Interessa realçar que 50% dos clientes ordenados, de acordo com o montante total de recursos, apresentam, nos últimos 12 meses, um montante em recursos até 5359 euros, e que, em média, o montante total de recursos dos clientes é de 31109 euros.

No que diz respeito ao montante total de activos¹⁶ nos últimos 12 meses verificou-se uma grande dispersão, uma vez que grande parte dos clientes em análise tinham 0 euros de activos (metade dos clientes ordenados possuíam 0 euros de montante total de activos), sendo que o máximo de activos corresponde a cerca de 1657900 euros. O montante médio de activos que os clientes apresentam, nos últimos 12 meses, é de 16413 euros.

Os clientes em análise apresentam uma grande dispersão no que diz respeito ao saldo médio em depósitos à ordem ao longo dos últimos 12 meses, uma vez que existem clientes com 0 euros de saldo médio (mínimo) e clientes com 1344839 euros (valor máximo). Interessa realçar que, em média, os clientes possuem um saldo médio em depósitos à ordem de cerca de 4181 euros, e que 50% dos dados ordenados apresentam um saldo médio até 918 euros, o que demonstra que grande parte dos clientes em análise apresentam baixo montante de saldo médio.

Em suma, após o processo de identificação e análise das características dos clientes e grau de envolvimento dos mesmos no banco, concluímos que são clientes: (1) maioritariamente do sexo masculino; (2) maioritariamente em idade activa ou no final desta; (3) com antiguidade média no banco de cerca de 17 meses; (4) cujo número médio de produtos adquiridos no banco é de 6; (5) com montante médio de recursos associados superior ao montante médio de activos, o que significa que são clientes investidores e (6) com saldo médio em depósitos à ordem de cerca de 4181 euros.

¹⁵ Traduz o somatório dos montantes de depósitos a prazo, fundos, seguros e poupanças.

¹⁶ Consiste no somatório dos montantes de crédito à habitação, crédito especializado e crédito pessoal.

3.4 Produtos utilizados no desenvolvimento do modelo de CLTV

De modo a construir um modelo de CLTV mais eficiente resolveu-se utilizar os produtos considerados como sendo chave para uma maior rentabilidade para o banco. Assim, os produtos abaixo referidos são os produtos que originam 80% da rentabilidade do banco. Os produtos utilizados na obtenção de cada uma das componentes que envolvem o cálculo do CLTV foram:

- depósitos à ordem (DO);
- cartões de crédito (CC);
- seguros (SE);
- fundos (FU);
- depósitos a prazo (DP);
- poupanças (PO);
- créditos à habitação (CH);
- créditos pessoais (CP);
- créditos especializados (CE).

3.5 Elementos envolvidos no processo de obtenção do CLTV

O valor associado a cada cliente existente no banco foi obtido através do cálculo de duas componentes distintas: (1) rentabilidade do cliente e (2) potencialidade do cliente. Assim sendo, genericamente considera-se que o valor do cliente i é dado pela equação 9, o que significa que o valor do cliente i é dado pela soma do valor que esse mesmo cliente tem até ao momento actual, e o valor do mesmo no futuro, tal como sugerem Peppers e Rogers (1993).

$$CLTV_{it} = Rentabilidade_{it} + Potencialidade_{it} \quad (9)$$

Segundo os autores Brito e Ramos (2000) a rentabilidade de cada cliente (i) reflecte o valor real do negócio desse mesmo cliente até à data actual (o que no caso prático em questão se traduz até Setembro de 2008), sendo este valor calculado para todos os momentos do tempo até à data actual (t). Já o valor futuro de negócio de cada cliente traduz a potencialidade deste cliente após o momento actual (t).

Primeiro que tudo interessa perceber a forma de cálculo de cada uma das componentes que é necessário existir para a sua obtenção.

3.5.1 Rentabilidade associada aos Clientes

A rentabilidade associada a cada um dos clientes (i) até ao momento actual (Setembro de 2008) foi denominada de *Past Customer Value* (PCV). Interessa salientar que este conceito, apesar de já ter sido referenciado na tese anteriormente como sendo um dos métodos “tradicionais” de obtenção do valor do cliente, é, neste caso, utilizado para reflectir o cálculo da rentabilidade do cliente até a data actual, não sendo feita qualquer extrapolação do seu valor para o futuro (como acontecia no conceito referido anteriormente).

O PCV correspondeu ao somatório das margens (diferença entre receitas e custos) associadas a cada um dos nove produtos em análise¹⁷, que cada cliente possui num determinado momento de tempo, sendo posteriormente subtraído os custos de manutenção dos clientes¹⁸. O resultado obtido foi ajustado com a taxa de juro (que não é mais do que um indicador económico) para os períodos de tempo em questão (t), uma vez que o valor do dinheiro não é constante ao longo do tempo. Assim sendo, o PCV calculado pode ser dado pela equação 10. O PCV de cada um dos clientes (i) foi obtido mensalmente, no período de Julho de 2007 a Setembro de 2008 (T).

$$PCV_{it} = \frac{\sum_{j \in J} m_{ijt} - cc_{it}}{(1 + d_t)^t} \quad (10)$$

Os parâmetros utilizados no cálculo do PCV dizem respeito a:

- mm_{ijt} : Margens de cada cliente por produto, num dado momento;
- cc_t : Custo de manutenção de cada cliente, num dado momento;
- d_t : Taxa de juro no momento t;
- i: Cliente;
- t: Período de tempo;
- j: Cada um dos produtos utilizados na modelização;
- J: Produtos utilizados na modelização.

¹⁷ Depósitos à Ordem (DO), Cartões de Crédito (CC), Seguros (SE), Fundos (FU), Depósitos a Prazo (DP), Poupanças (PO), Créditos à Habitação (CH), Créditos Pessoais (CP), Créditos Especializado (CE).

¹⁸ Custos estes que são distintos consoante o tipo de cliente.

3.5.2 Potencialidade associada aos Clientes

A Potencialidade associada a cada um dos clientes (i), e tendo em conta que traduz o valor futuro de cada cliente, foi obtida em duas etapas distintas: (1) cálculo do valor dos clientes num futuro considerado mais recente (*Recent Future Customer Value*) e (2) cálculo do valor dos clientes num futuro considerado mais longínquo (*Future Customer Value*).

3.5.2.1 Valor dos clientes num futuro considerado mais recente (*Recent Future Customer Value*: RFCV)

O valor do futuro recente (RFCV) de cada cliente (i) traduz a potencialidade desse mesmo cliente no que diz respeito ao valor de negócio num futuro considerado como próximo, sendo que no caso em questão foi considerado os 3 meses consecutivos (Outubro a Dezembro de 2008) face ao período actual (Setembro de 2008), por ser o horizonte de previsão dos modelos preditivos utilizados como *input* no processo de obtenção do RFCV.

O RFCV associado a cada cliente foi obtido através do somatório das margens médias futuras que cada cliente terá no período de tempo em análise, multiplicando tais margens pela probabilidade de *cross sell*, probabilidade de *up sell* e probabilidade de não abandono associada a produtos (produtos estes apresentados na secção 3.4). De realçar que todas estas probabilidades foram obtidas através do desenvolvimento, e implementação, dos modelos preditivos em questão (modelos estes que iremos apresentar posteriormente). Já as margens médias foram obtidas de duas formas distintas: (1) através de médias móveis de 4 meses para cada mês em análise, caso cada um desses 4 meses tenha valor de margens associado ou (2) através das margens médias de todos os clientes, tendo em conta o tempo de posse de cada produto, casos as margens associadas aos clientes sejam desconhecidas. Ao resultado obtido foi subtraído todos os custos de retenção e manutenção de cada cliente em cada período de tempo, sendo que posteriormente tal resultado foi ajustado à taxa de juro (indicador económico) para os períodos de tempo em questão. O resultado final obtido foi combinado com a probabilidade de cada cliente não abandonar o banco em cada período de tempo. A equação 11 traduz a fórmula de cálculo do RFCV utilizada. O RFCV foi calculado mensalmente, entre Outubro de 2008 e Dezembro de 2008 (T).

$$RFCV_{it} = \left(\frac{\sum_{j \in J} \overline{m_{ij}} * xs_{ijt} + \sum_{j \in J} \overline{m_{ij}} * us_{ijt} + \sum_{j \in J} \overline{m_{ij}} * (1 - c_{ijt}) - cc_{it}}{(1 + d_t)^t} \right) * (1 - gc_{it}) \quad (11)$$

Os elementos envolvidos na obtenção do RFCV são:

- $\overline{m_{ij}} = \frac{\sum_{j \in J} m_{ij}}{\sum_{j \in J} j}$: Margens médias futuras dos produtos que cada cliente tem num momento de tempo;
- xs_{ijt} : Probabilidade de *cross sell* associada a cada produto e/ou serviço num determinado momento do tempo e para cada um dos clientes;
- us_{ijt} : Probabilidade de *up sell* associada a cada produto e/ou serviço em cada período de tempo e para cada um dos clientes;
- c_{ijt} : Probabilidade de abandono de cada produto e/ou serviço em cada período de tempo e para cada um dos clientes;
- cc_{it} : Custos de manutenção e custos de retenção de cada um dos clientes, num dado momento;
- d_t : Taxa de juro associada a cada período de tempo;
- gc_{it} : Probabilidade do cliente abandonar o banco em cada período de tempo;
- i : Cliente;
- t : Momento de tempo;
- j : Cada um dos produtos utilizados na modelização;
- J : Produtos utilizados na modelização.

3.5.2.1.1 Modelos analíticos desenvolvidos

Como referido anteriormente o RFCV associado a cada cliente foi obtido utilizando como *input* os *scores* de modelos de *cross sell*, modelos de *up sell* e modelos de abandono associados aos produtos utilizados na obtenção do CLTV.

Modelos de Cross Sell

Segundo Deighton e outros (1994) os modelos de *cross sell* são modelos que permitem maiores receitas, potenciar a lealdade e *share-of-wallet* dos clientes existentes no banco, através da oferta de produtos adicionais. Todos estes produtos serão adequados às respectivas necessidades de cada cliente, com o intuito de se obter maior satisfação. Assim sendo, este tipo de modelos assume especial importância, uma vez que, regra geral, a venda de produtos a clientes já existentes no banco apresenta custos mais baixos do que a aquisição de um novo cliente.

Foram desenvolvidos oito modelos distintos de *cross sell*: (1) cartões de crédito; (2) depósitos a prazo; (3) fundos; (4) poupanças; (5) seguros; (6) crédito especializado; (7) crédito à habitação e (8) crédito pessoal. Para cada um destes modelos foi necessária a criação da respectiva variável resposta.

Com excepção do modelo de *cross sell* de cartões de crédito, as variáveis resposta foram definidas pela seguinte regra:

- target=1: Caso o cliente tenha igual ou superior a 200 euros do produto nos últimos 3 meses (cliente com aquisição do produto);
- target=0: Caso o cliente tenha menos do que 200 euros do produto nos últimos 3 meses (cliente sem aquisição do produto).

Já para o modelo de *cross sell* de cartões de crédito a variável resposta foi definida da seguinte forma:

- target=1: Caso o cliente efectue, pelo menos, dois movimentos nos últimos 3 meses;
- target=0: Caso o cliente efectue menos de dois movimentos nos últimos 3 meses.

Como horizonte de análise¹⁹, para cada um dos modelos, foram utilizados 12 meses, e como horizonte de previsão 3 meses²⁰.

Modelos de *Up Sell*

Os modelos de *up sell* têm como objectivo apoiar na venda de produtos com maior valor, com o intuito de aumentar as receitas inerentes. Segundo Cardoso e Gonçalves (2001), através deste tipo de modelos é possível “efectuar a actualização dos produtos que os clientes já possuem”, uma vez que é possível perceber qual a propensão para cada cliente vir a aumentar o grau de relação com determinado produto. Este tipo de modelos, em conjunto com os modelos de *cross sell*, têm ainda outro benefício, que passa pela percepção de comportamentos que inspiram a criação de novas ofertas, novos produtos, ou a justificação de pacotes de produtos.

Com vista à obtenção do RFCV foram desenvolvidos quatro modelos de *up sell*: (1) seguros; (2) depósitos a prazo; (3) fundos e (4) poupanças. Para cada um destes modelos foi necessária a criação da respectiva variável resposta, tendo-se definido as mesmas através da seguinte regra:

- target=1: caso o montante do produto em análise tenha um saldo superior a 50 euros à 3 meses atrás, e tenha sofrido um aumento igual ou superior 7% nos últimos 3 meses (cliente realizou *up sell* do produto);
- target=0: caso o montante do produto em análise tenha um saldo superior a 50 euros à 3 meses atrás, e tenha sofrido um aumento inferior a 7% nos últimos 3 meses (cliente não realizou *up sell* do produto).

De salientar que como horizonte de análise foram utilizados 12 meses, e como horizonte de previsão 3 meses.

Modelos de Abandono

Os modelos de abandono tiveram como objectivo a determinação de quais os clientes com maior propensão para abandonar os produtos em análise ou mesmo o próprio banco, com o intuito de realizar campanhas específicas que permitam a não ocorrência desse abandono.

¹⁹ O horizonte de análise consiste no período temporal, prévio ao mês considerado como sendo o actual, que contem as informações comportamentais do Cliente.

²⁰ O horizonte de previsão traduz o horizonte temporal, utilizado para efectuar a previsão do evento.

Foram desenvolvidos oito modelos de abandono: (1) banco; (2) cartões de crédito; (3) crédito à habitação; (4) crédito pessoal; (5) depósitos a prazo; (6) fundos; (7) poupanças e (8) seguros.

Para cada um destes modelos foi necessária a construção da respectiva variável resposta, sendo que, para os modelos cujo objectivo é detectar o abandono de produtos, tal construção foi feita de acordo com a seguinte regra:

- target=1: caso o cliente possua o produto nos 5 meses anteriores ao mês N e no mês N deixe de o ter (cliente abandonou do produto);
- target=0, caso o cliente tenha o produto nos últimos 6 meses (cliente não abandonou o produto).

Já a variável resposta do modelo de abandono do banco foi definida de acordo com a seguinte regra:

- target=1: caso o cliente seja activo nos últimos 9 meses e nos últimos 3 meses deixe de ser activo (cliente abandonou o banco);
- target=0, caso o cliente seja activo nos últimos 12 meses (cliente não abandonou o banco).

Assim sendo, o horizonte de análise utilizado, para cada um dos modelos, foi de 12 meses, e como horizonte de previsão de 3 meses.

De salientar que o horizonte de análise utilizado no desenvolvimento dos modelos dependeu da informação existente e das definições das *targets* (definições estas desenvolvidas pelo Banco).

Cada um destes modelos acima representados (*cross sell*, *up sell* e abandono) foi desenvolvido através do processo de *Data Mining* denominado SEMMA (anteriormente definido), tendo-se utilizado como ferramentas preditivas árvores de decisão, regressões logísticas, redes neuronais, ou mesmo modelos de mistura, e diversas parametrizações.

Após o desenvolvimento de vários modelos (devido às várias técnicas e parametrizações utilizadas) procedeu-se à avaliação dos mesmos, e à selecção dos melhores modelos, por tipo de produto. Os modelos considerados como vencedores correspondem àqueles com menor taxa de clientes mal classificados no conjunto de validação.

Assim, interessa perceber os conceitos que envolvem as técnicas utilizadas no desenvolvimento dos modelos de *input* subjacentes à obtenção do RFCV.

Árvores de Decisão

Segundo Santos e Azevedo (2005) as árvores de decisão são representações gráficas de regras de classificação. Este tipo de técnica consiste num método de aprendizagem supervisionado que assenta na filosofia de “dividir para conquistar”, uma vez que as observações são divididas com base em determinadas características, o que significa que um dos objectivos das árvores de decisão é a obtenção de regras proposicionais. De salientar que todo este processo de dividir as características ou variáveis é feito com o intuito de diminuir o risco de previsão, ou o risco de classificação, e engloba três elementos distintos: (1) nó raiz (nó que contém o conjunto inicial dos dados); (2) ramificações (conjunto de regras criadas) e (3) nós folha (nós a partir dos quais se realiza previsões, ou seja, os nós finais).

Este tipo de método é bastante utilizado já que: (1) tem boa capacidade explicativa, uma vez que fornecem bons resultados; (2) é fácil de interpretar devido à sua representação lógica; (3) não exige o tratamento de valores omissos; (4) pode ser visto como um método para o tratamento de outliers e (5) podemos utilizar variáveis explicativas de diferentes naturezas (métricas e não métricas).

Todos os modelos desenvolvidos utilizando árvores de decisão foram construídos recorrendo ao algoritmo *Classification and Regression Trees* (CART). Este algoritmo, proposto por Breiman, Friedman, Olshen e Stone (1984), constrói árvores de decisão binárias, o que leva a que as árvores sejam de grande extensão. O algoritmo CART é um exemplo de algoritmo de partição binária recursiva, uma vez que fornece árvores binárias e é aplicado, recursivamente, a cada um dos subconjuntos originados até que não seja possível efectuar partições, ou sejam atingidos critérios de paragem previamente estabelecidos. De salientar que a construção de árvores binárias traduz uma simplicidade e legibilidade de análise.

Regressões Logísticas Binárias

As regressões logísticas efectuadas traduzem modelos bastante utilizados devido à fácil interpretação, compreensão e capacidade explicativa. O seu principal objectivo consiste na descrição das probabilidades associadas a cada valor da variável dependente binária (probabilidade de *cross sell*, *up sell* ou abandono), a partir de um conjunto de “variáveis explicativas, variáveis estas que podem ser categóricas ou contínuas” (Sharma, 1996). Os modelos construídos foram estimados a partir de máxima verosimilhança, sendo que, em alguns, foi efectuada a selecção das variáveis recorrendo a três critérios distintos: (1)

*forward*²¹; (2) *backward*²² e (3) *stepwise*²³. A partir deste tipo de modelação foi possível obter a estatística *Odds Ratio*, que permite efectuar a comparação de grupos face a um determinado evento (evento este que poderá ser o acto de *cross-sell*, *up-sell* e/ou abandono).

Redes Neurais

As redes neurais são modelos que podem ser de aprendizagem supervisionada²⁴ ou não supervisionada²⁵ inspirados na constituição do cérebro humano, já que a combinação de vários neurónios artificiais em camadas interligadas entre si originam este tipo de modelos (Mitchell, 1997). Assim, as redes neurais podem ser vistos como um instrumento de cálculo, já que o seu objectivo consiste em ajustar os pesos associados às ligações existentes entre os neurónios.

Apesar de ser uma técnica com muito boa capacidade preditiva, muito em especial para casos raros, existe uma grande desvantagem associada à utilização deste tipo de modelos – difícil interpretação de quais as variáveis explicativas que mais explicam a variável resposta em análise, uma vez que apenas conseguimos estimar o valor dos pesos associados às ligações dos neurónios existentes nas camadas.

Todos os modelos desenvolvidos neste projecto através de redes neurais utilizaram como configuração o perceptrão multicamada, mais concretamente o algoritmo *backpropagation*²⁶ (Rumelhart *et al*, 1986). O treino de cada rede neuronal desenvolvida terminou quando se obtém o mínimo de classificações erradas, visto tratarem-se de redes de classificação. Assim, as redes desenvolvidas são constituídas por três camadas distintas: (1) uma camada de entrada²⁷; (2) uma camada intermédia²⁸ e (3) uma camada de saída²⁹. Segundo Bishop (1995)

²¹ Abordagem do particular para o geral, em que se adiciona variáveis uma a uma, e analisa-se a sua significância estatística individual e a validade global do modelo.

²² Consiste numa abordagem do geral para o particular, em que se insere todas as variáveis explicativas possíveis, e que achamos relevantes para o modelo, e é analisada a significância estatística individual de cada variável, e a validade global do modelo.

²³ Abordagem que resulta de uma combinação dos métodos de selecção *forward* e *backward*, o que significa que testa cada inserção ou eliminação de variáveis, sendo por isso o método mais utilizado.

²⁴ Caso exista uma camada de saída, o que leva a que a aprendizagem seja feita recorrendo a dados de entrada e dados de saída.

²⁵ Caso não exista uma camada de saída, sendo que a aprendizagem é feita descobrindo padrões nos dados de *input*.

²⁶ A aprendizagem de uma rede neuronal, segundo o algoritmo *backpropagation*, funciona de acordo com duas fases distintas: (1) *forward*, onde, no sentido entrada-saída da rede se calcula as estimativas e o erro associado à variável resposta e (2) *backward*, em que de depois de calculado o erro associado à variável resposta, volta-se atrás na rede, para que se possa ajustar os pesos de forma a diminuir o erro (fase de aprendizagem).

²⁷ Camada constituída pelas variáveis explicativas.

²⁸ Camada que permite reconhecer padrões, e actua como um filtro das informações/sinais entre os neurónios, de forma a obter melhores previsões.

as ligações existentes entre os neurónios, e em cada uma destas camadas, consideram-se uni-dirigidas³⁰ e têm pesos distintos.

O cálculo das estimativas da variável dependente é baseado em funções de activação³¹ de combinações lineares, que combina o valor dos pesos que se associam às camadas intermédias e de saída.

Podemos pensar que todo o processo de aprendizagem das redes neuronais será cada vez melhor quanto maior o número de iterações executadas. Contudo, tal não é verdade, uma vez que a rede neuronal pode-se adaptar muito bem aos dados utilizados na aprendizagem e adaptar-se mal a outros dados (chama-se sobre-ajustamento ou *overfitting*). Assim, é fundamental a existência de regras de paragem. Zupan e Gasteiger (1993) sugerem os seguintes critérios de paragem: (1) erro mínimo; (2) gradiente do erro; (3) número de iterações ou (4) por validação cruzada. Os modelos desenvolvidos no caso prático utilizando redes neuronais utilizaram como critérios de paragem o mínimo de taxa de *missclassification* no conjunto de validação.

Modelos de Mistura

Segundo o *SAS Institute Inc.* (2003) os modelos de mistura correspondem a um tipo de modelos que resultam da combinação das previsões de dois ou mais modelos (utilizando árvores de decisão, regressões logísticas e/ou redes neuronais) que se possam utilizar num único modelo, sendo que o modelo resultante fornece normalmente melhores taxas de ajustamento. Existem duas formas de combinação de modelos, por forma a que o modelo final possa ser representado: (1) pela média dos valores previstos (probabilidades) para os diferentes modelos possíveis ou (2) pelo máximo dos valores previstos (probabilidades) para os diferentes modelos possíveis.

A tabela 4, existente no anexo 2, ilustra quais as técnicas de modelação seleccionadas nos modelos de *cross sell*, *up sell* e abandono considerados vencedores, e por sua vez utilizados no processo de obtenção do RFCV. De salientar que cada um dos modelos foi desenvolvido

29 Camada constituída pela variável resposta do modelo, e que é responsável por efectuar o processamento e produzir os outputs da rede.

30 Isto porque as saídas de cada camada correspondem a entradas de camadas posteriores.

31 As funções de activação são funções escalares, através das quais os neurónios transformam o *input*. O valor de activação produzido consiste no output do perceptrão.

de forma a minimizar os erros de classificação, ou seja, com o intuito de minimizar o número de clientes mal classificados, visto que desta forma obtém-se maior precisão do modelo.

3.5.2.1.2 Parametrização utilizada no desenvolvimento dos modelos e estratégia de selecção.

Após a identificação das técnicas utilizadas nos modelos preditivos considerados vencedores interessa perceber qual a parametrização utilizada para a obtenção de cada um dos modelos. De salientar que o desenvolvimento de cada um dos modelos foi efectuado de acordo com o processo SEMMA anteriormente apresentado, tendo-se realizado, em cada um dos modelos, as seguintes etapas:

- adequação do papel, tipo e nível das variáveis de *input* e variável resposta utilizadas em cada um dos modelos;
- amostragem estratificada pela variável resposta ³² de 30% (percentagem da amostragem), sendo que a proporção em cada um dos estratos é similar à proporção de casos existentes na população (critério de estratificação);
- partição dos dados estratificada pela variável resposta de cada um dos modelos, tendo-se definido a partição em 3 conjuntos: (1) 70% dos dados para o conjunto de treino ³³; (2) 20% para o conjunto de validação ³⁴ e (3) 10% para o conjunto de teste ³⁵;
- exploração de dados, com o intuito de avaliarmos a amostra obtida e verificar a existência de valores omissos;
- tratamento dos valores omissos anteriormente identificados, tendo-se utilizado como métodos de imputação para variáveis intervalares ou de classe: (1) árvores de decisão; (2) distribuição e (3) valores definidos mediante conhecimento de negócio. A utilização de cada um destes métodos diferiu consoante o modelo e variável em causa;
- aplicação de transformação às variáveis de *input* intervalares ou de classe utilizadas nos modelos. Nas variáveis intervalares foi utilizado, por exemplo, as seguintes transformações: (1) maximizar a normalidade da variável; (2) maximizar a

³² Visto ser um método que garante que a distribuição da variável resposta se mantenha.

³³ É o conjunto de dados a partir do qual é efectuada a estimação do modelo.

³⁴ Utilizado para ajustar o modelo durante a estimação e também para proceder à sua avaliação.

³⁵ Permite a obtenção de uma estimativa final fiável, já que não é mais do que um segundo teste feito aos dados.

correlação com a variável resposta; (3) logaritmizar a variável; (4) exponencial da variável e (5) divisão da variável em grupos com frequência aproximada (*quantile*). Já para as variáveis de classe utilizou-se o método de “*group rare levels*”, método este que permite a transformação de variáveis utilizando os níveis raros das mesmas;

- selecção de variáveis *input* para cada modelização, tendo-se utilizado o método “*Chi-Square*” ou o método “*R and Chi Square*”, o que significa que as variáveis de *input* utilizadas nos modelos preditivos foram seleccionadas de acordo com o critério da estatística de teste do *Chi-Square*³⁶ e Coeficiente de Determinação;
- utilização de diversas técnicas preditivas, tais como árvores de decisão, regressões logísticas, redes neuronais e modelos de mistura. A tabela 5, existente no anexo 2, ilustra a parametrização utilizada no que diz respeito às técnicas de modelação utilizadas nos modelos de *cross sell*, *up sell* e abandono considerados vencedores;
- comparação e avaliação dos melhores modelos, sendo que os modelos seleccionados foram aqueles com menor erro de má classificação (*Misclassification Rate*), maior Curva de ROC, maior Sensibilidade e Especificidade (estatísticas estas existentes na matriz de confusão) no conjunto de validação. A tabela 6, existente no anexo 2, permite verificar e analisar as estatísticas de avaliação da qualidade dos modelos desenvolvidos acima referidas. De salientar que os modelos vencedores foram modelos cuja performance e precisão era consideravelmente superior às restantes técnicas utilizadas na modelação;
- criação e disponibilização do processo de scorização dos modelos considerados vencedores para serem utilizados posteriormente em produção.

Após o desenvolvimento dos modelos preditivos de *cross sell*, *up sell* e abandono interessa perceber a qualidade e precisão de cada um dos modelos, através de algumas medidas acima referidas e cujos resultados se encontram na tabela 6, existente no anexo 2.

Ao analisarmos a tabela em causa podemos concluir que, globalmente, os modelos preditivos construídos são bons modelos, uma vez que, em todos os casos, a curva de ROC está acima de 0.5, ou seja, todos os modelos em causa têm uma precisão acima dos 50%.

³⁶ Sendo que o valor mínimo da estatística teste do *Chi-Square* foi de 3.84, já que o nível de significância utilizado foi de 5%. Caso o valor da estatística teste obtido para uma variável explicativa seja superior ao valor da estatística teste tabulado então significa que essa variável é estatisticamente relevante para explicar a variável resposta em causa.

De seguida, interessa perceber quais os modelos, que individualmente, têm maior e menor poder discriminante.

Analisando a curva de ROC, concluímos que os modelos com:

- maior performance e poder discriminante foram: (1) abandono de CH; (2) *cross sell* de CH e (3) abandono de DP, visto serem os modelos com maior valor de curva de ROC;
- menor performance e poder discriminante foram: (1) *cross sell* de PO; (2) abandono de PO e (3) abandono de CP, visto serem os modelos com menor valor de curva de ROC.

Um outro aspecto importante de se analisar, para além da performance e poder discriminante, consiste na sensibilidade e especificidade de cada modelo, onde se pode concluir que os modelos com maior precisão dos:

- eventos em análise (valor 1 associados às variáveis resposta) foram os modelos relativos ao: (1) abandono de CC; (2) abandono de FU e (3) abandono de CH, visto traduzirem os modelos com maior valor de sensibilidade;
- não eventos em análise (valor 0 associados às variáveis resposta) foram os modelos que analisam a propensão ao: (1) abandono dos clientes existentes no banco; (2) *cross sell* de CH e (3) abandono de CC, já que são os modelos com maior valor de especificidade.

Para além das medidas já referidas, é importante analisar a taxa de má classificação dos modelos, ou seja, a proporção de indivíduos mal classificados. Para tal recorreu-se à estatística *Misclassification Rate*, onde foi possível concluir que os modelos com:

- maior proporção de indivíduos mal classificados foram: (1) abandono dos clientes existentes no banco; (2) abandono de CP e (3) *up sell* de FU;
- menor proporção de indivíduos mal classificados foram: (1) abandono de CC; (2) *cross sell* de CH e (3) abandono de CH.

Posteriormente ao desenvolvimento e análise da qualidade dos modelos preditivos de *cross sell*, *up sell* e abandono, e com o intuito de perceber as características dos clientes existentes no banco, procedeu-se à análise descritiva dos resultados, cujos resultados encontram-se na tabela 7 e figura 5 existente no anexo 2. A figura 5, anexo 2, representa o *score* médio dos

modelos preditivos desenvolvidos para a obtenção da componente de RFCV, onde se pode concluir que, em média existia, entre Outubro e Dezembro de 2008, maior propensão para a aquisição de crédito pessoal (modelo de *Cross Sell* CP), depósitos a prazo (modelo de *Cross Sell* DP) e Abandono de Cartões de Crédito (modelo de Abandono CC). Em relação aos produtos com menor aceitação por parte dos clientes podemos concluir que, em média, existia menor propensão para o reforço de poupanças (modelo de *Up Sell* PO) e fundos (modelo de *Up Sell* FU). No que diz respeito ao modelo de abandono dos clientes existentes no banco conclui-se que, em média, verificou-se uma propensão para tal abandono de 27%, sendo um dos modelos com menor *score* médio associado.

Tal como anteriormente referido, as probabilidades obtidas através do desenvolvimento dos modelos preditivos foram utilizadas como *input's* para a obtenção da componente de RFCV.

3.5.2.2 Valor dos clientes num futuro considerado mais longínquo (*Future Customer Value: FCV*)

O valor do futuro considerado mais longínquo (FCV) de cada cliente *i* traduz a potencialidade desse mesmo cliente no que diz respeito ao valor de negócio num futuro considerado mais distante, tendo-se utilizado os próximos 9 meses consecutivos (Janeiro de 2009 a Setembro de 2009 (T)) face aos 3 meses utilizados na obtenção do RFCV.

O cálculo da potencialidade futura foi obtido através da utilização de diversos elementos: (1) margens médias futuras, tendo em conta o tempo de posse de cada produto ou serviço; (2) probabilidade associada à análise de sobrevivência de cada cliente existir no banco; (3) probabilidade associada à análise de sobrevivência da posse de produtos de investimentos e (4) probabilidade associada à análise de sobrevivência de utilização de cartões de crédito. De salientar que a obtenção destas probabilidades foi feita recorrendo ao desenvolvimento de modelos de Análise de Sobrevivência, modelos estes que iremos apresentar posteriormente.

Aos resultados obtidos foram subtraídos os custos de retenção e manutenção de cada cliente em cada período de tempo, sendo que posteriormente os resultados obtidos foram ajustados com a taxa de juro para os períodos de tempo em questão. O resultado final obtido foi combinado com a probabilidade de cada cliente não abandonar o banco em cada período de tempo.

Assim sendo, o FCV é dado pela equação 12.

$$FCV_{it} = \left(\frac{\sum_{j \in J} \overline{m_{ij}} * (1 - rs_{ijt}) + \sum_{j \in J} \overline{m_{ij}} * (1 - cs_{ijt}) + \sum_{j \in J} \overline{m_{ij}} * (1 - gs_{ijt}) - cc_{it}}{(1 + d_t)^t} \right) * (1 - gc_{it}) \quad (12)$$

Em suma, os elementos utilizados no cálculo do FCV são:

- $\overline{m_{ij}} = \frac{\sum_{j \in J} m_{ij}}{\sum_{j \in J} j}$: Margens futuras dos produtos que cada cliente tem num momento de tempo específico;
- rs_{ijt} : Probabilidade do cliente abandonar a posse de produtos de investimentos para cada período de tempo, sabendo que até então os possuía;
- cs_{ijt} : Probabilidade de abandono de utilização de cartões de crédito para cada período de tempo, sabendo que até ao momento tal cliente os utilizou;
- gs_{ijt} : Probabilidade de abandono do cliente existente no banco, para cada período de tempo, sabendo que até ao momento o cliente permaneceu no banco (modelo de análise de sobrevivência);
- cc_{it} : Custos de manutenção e custos de retenção de cada cliente num dado período de tempo;
- d_t : Taxa de juro associada a cada período de tempo (indicador económico);
- gc_{it} : Probabilidade do cliente abandonar o banco no horizonte de previsão utilizado (modelo preditivo de abandono);
- i : Cliente;
- t : momento de tempo;
- j : Cada um dos produtos utilizados na modelização;
- J : Produtos utilizados na modelização.

3.5.2.2.1 Modelos analíticos de sobrevivência desenvolvidos

Com vista a obter a componente FCV associado a cada cliente, e com o intuito de obter o valor de cada um dos clientes num futuro considerado como longínquo (Janeiro a Setembro de 2009), foi necessário utilizar os resultados, fornecidos pelo banco, de três modelos de análise de sobrevivência para o período de Julho de 2007 a Setembro de 2008.

Modelo de Análise de Sobrevivência de clientes existentes no banco

O principal objectivo deste modelo consistiu na observação do tempo médio de existência dos clientes no banco ao longo do tempo, através da análise da função de sobrevivência³⁷ e da função *hazard*³⁸. A partir destas funções foi possível obter, para cada mês da análise, a probabilidade de um cliente permanecer no banco.

A variável resposta deste modelo, sendo intervalar, traduziu a diferença, em termos de meses, entre o mês do primeiro abandono associado a cada um dos clientes, e o mês de activação do cliente. Para tal foi necessário, previamente, definir o conceito de abandono e o conceito de activação, com vista à identificação dos mesmos ao longo do tempo.

Assim sendo, o banco definiu que:

- um cliente abandonou o banco no mês N caso “o cliente deixe de ser activo, pelo menos, nos 2 meses posteriores a N, sendo que, pelo menos, nos 2 meses anteriores ao mês N se encontrava activo”;
- um cliente é activo no mês N, caso “o cliente realize, pelo menos, um movimento na DO, e tenha um envolvimento com o banco de pelo menos 15 euros”.

Os clientes alvos de modelização foram todos aqueles:

- que em Setembro de 2008 eram activos (pelo menos numa única vez), particulares, não colaboradores no banco e não falecidos;
- cuja entrada do cliente no banco foi entre Julho de 2007 e Setembro de 2008, visto ser o primeiro mês de análise;

³⁷ Dá-nos a probabilidade dos clientes sobreviverem no banco até ao intervalo de tempo em causa.

³⁸ Pode ser visto como um indicador natural da propensão ao abandono dos clientes existentes no Banco no intervalo $[t; t+\Delta t]$ dado que já sobreviveram até ao momento t .

Em suma, a partir do desenvolvimento deste modelo, foi possível ao banco retirar conclusões sobre a probabilidade dos clientes o abandonarem entre cada intervalo de tempo em análise, considerando que até então eram clientes activos.

Modelo de Análise de Sobrevivência de clientes com posse de produtos de investimento

Este modelo teve como principal objectivo observar a evolução dos clientes quanto ao tempo médio de posse de produtos de investimento e poupança (fundos, depósitos a prazo, poupanças e seguros) ao longo do tempo. Tal como no modelo acima referido, procedeu-se à análise da função *hazard* e da função de sobrevivência com o intuito de analisar a probabilidade de um cliente permanecer como “cliente investidor” do banco ao longo do período de estudo.

A variável resposta associada ao modelo traduz o número de meses até ao primeiro abandono da família de produtos de investimento e poupança (evento em estudo) desde o primeiro momento de aquisição de tais produtos. Para a construção da variável resposta foi necessário clarificar o conceito de abandono da família de produtos de investimento e poupança bem como o conceito de aquisição da família em causa. Definiu-se que:

- um cliente abandonou a família de produtos de investimento e poupança no mês N caso “a soma do montante total investido nesses produtos seja abaixo dos 100 euros, pelo menos nos 3 meses seguintes (N , $N + 1$ e $N + 2$), sendo que nos 3 meses anteriores ao mês N ($N - 1$, $N - 2$ e $N - 3$), o montante de produtos de investimento e poupança era superior a 100 euros”;
- um cliente adquire produtos de investimento e poupança no mês N caso “o somatório do montante total desses mesmos produtos seja superior ou igual a 100 euros nos 3 meses seguintes (N , $N + 1$ e $N + 2$)”.

Os clientes alvo utilizados na modelação foram todos os:

- que em Setembro de 2008 eram activos (pelo menos numa única vez), particulares, não colaboradores no banco e não falecidos;
- cuja entrada do cliente no banco foi entre Julho de 2007 e Setembro de 2008, visto ser o primeiro mês de análise;
- clientes com, pelo menos, uma posse de produtos de investimento e poupanças entre Julho de 2007 e Setembro de 2008, num mínimo de 3 meses;

Assim, com o desenvolvimento do modelo de análise de sobrevivência de produtos de investimento e poupança foi possível retirar conclusões sobre a probabilidade dos clientes abandonarem a posse de produtos de investimento e poupança (fundos, depósitos a prazo, poupanças e seguros) ao longo do tempo, ou seja, permitiu analisar a evolução do tempo médio de posse de tais produtos por parte dos clientes.

Modelo de Análise de Sobrevivência para Clientes que utilizavam Cartões de Crédito

O modelo de análise de sobrevivência de clientes com utilização de cartão de crédito teve como objectivo analisar o tempo médio de vida da utilização de cartões de crédito por parte dos clientes existentes no banco. Através da análise da função de sobrevivência e da função *hazard* foi possível, para cada momento no tempo, analisar qual a probabilidade de utilização de, pelo menos, um cartão de crédito.

A variável resposta associada a este modelo traduz o número de meses até ao primeiro “abandono de utilização de cartão de crédito” pelos clientes existentes no banco, isto é, a diferença entre o mês do primeiro abandono de utilização de cartão de crédito associada a cada um dos clientes e o mês da primeira utilização do mesmo. Para a criação da variável resposta deste modelo foi necessária a definição, por parte do banco, dos conceitos de “utilização de cartão de crédito” e “abandono de utilização de cartão de crédito”. Como tal, foram adoptadas as seguintes definições:

- Cliente utilizou cartão de crédito no mês N caso tenha um número de utilizações diferente de zero nos últimos 4 meses ($N-1$, $N-2$, $N-3$ e $N-4$);
- Cliente abandonou a utilização de cartão no mês N caso utilize o mesmo nos últimos 4 meses ($N-1$, $N-2$, $N-3$ e $N-4$) e não tenha qualquer utilização nos 3 meses seguintes ($N+1$, $N+2$ e $N+3$).

Os clientes considerados para a modelação foram todos os clientes:

- que em Setembro de 2008 eram activos (pelo menos numa única vez), particulares, não colaboradores no banco, e não falecidos;
- cuja entrada do cliente no banco foi entre Julho de 2007 e Setembro de 2008, visto ser o primeiro mês de análise;
- utilizaram, pelo menos, uma única vez o cartão de crédito no período em análise;
- clientes com idade superior a 18 anos.

Assim, a partir do modelo em análise, foi possível identificar qual a probabilidade de “abandono de utilização de cartões de crédito” pelos clientes em cada momento do tempo, dado que utilizaram os cartões até ao momento em causa.

Foram desenvolvidos os três modelos acima referidos através do procedimento SAS designado por “*Proc Lifetest*”, utilizando o método *LifeTable*. A partir deste método foi possível agrupar os meses em intervalos de tempo, o que é bastante útil quando se trabalha com uma grande quantidade de dados.

Um exemplo da sintaxe SAS necessária para o desenvolvimento destes modelos, tendo já as tabelas analíticas de *input* criadas, encontra-se na figura 7 existente no anexo 3.

Além da definição da variável resposta associada a cada um dos modelos, foi ainda necessário definir quais os clientes onde não existia qualquer noção do período de abandono em cada um dos modelos, limitação, esta, causada pela não existência de dados que permitam analisar o comportamento do cliente ao longo de todo o ciclo de vida no banco. Esta variável, sendo binária, definiu-se da seguinte forma:

- valor 0: caso o mês de abandono seja igual ao último dia do período de tempo em análise (Setembro de 2008)
- valor 1: caso contrário.

3.6 Obtenção do CLTV por cada um dos clientes

Após o cálculo de cada uma das componentes (PCV, RFCV e FCV) que permitem a obtenção do valor associado a cada cliente (CLTV), mensalmente, foi necessário efectuar o somatório das mesmas, por cliente (i) e por mês (t), tal como a equação 13 traduz.

$$CLTV_i = \sum_t (PCV_{it}, RFCV_{it}, FCV_{it}) \quad (13)$$

Como resultado final do desenvolvimento do modelo de CLTV obtemos o valor acumulado de cada cliente em Setembro de 2009 (último período de tempo em análise), tendo em conta a rentabilidade e potencialidade associada até então. A figura 5 ilustra o processo necessário para a obtenção do valor de cada cliente em Setembro de 2009.

De salientar que o valor de cada cliente neste período corresponde ao valor acumulado da rentabilidade e potencialidade acumulado ao longo do período de tempo em análise.

Valores superiores correspondem a clientes com maior valor e que os resultados obtidos em cada componente (PCV, RFCV e FCV) dizem respeito a um factor, não podendo por isso ser directamente interpretado, mas sim analisado quanto à ordem de grandeza e sinal.

Figura 5. Processo de obtenção do valor de cada cliente em Setembro de 2009.

ID Cliente	Mês	PCV	RFCV	FCV	CLTV Acumulado
1	200707	A	Não aplicável	Não aplicável	
1	...	...	Não aplicável	Não aplicável	
1	200809	O	Não aplicável	Não aplicável	
1	200810	Não aplicável	P	Não aplicável	
1	200811	Não aplicável	Q	Não aplicável	
1	200812	Não aplicável	R	Não aplicável	
1	200901	Não aplicável	Não aplicável	S	
1	...	Não aplicável	Não aplicável	...	
1	200909	Não aplicável	Não aplicável	Z	$AA = \text{Sum}(A, \dots, O, P, Q, R, S, \dots, Z)$
...	200707	A	Não aplicável	Não aplicável	
...	...	...	Não aplicável	Não aplicável	
...	200809	O	Não aplicável	Não aplicável	
...	200810	Não aplicável	P	Não aplicável	
...	200811	Não aplicável	Q	Não aplicável	
...	200812	Não aplicável	R	Não aplicável	
...	200901	Não aplicável	Não aplicável	S	
...	...	Não aplicável	Não aplicável	...	
...	200909	Não aplicável	Não aplicável	Z	$BB = \text{Sum}(A, \dots, O, P, Q, R, S, \dots, Z)$
N	200707	A	Não aplicável	Não aplicável	
N	...	...	Não aplicável	Não aplicável	
N	200809	O	Não aplicável	Não aplicável	
N	200810	Não aplicável	P	Não aplicável	
N	200811	Não aplicável	Q	Não aplicável	
N	200812	Não aplicável	R	Não aplicável	
N	200901	Não aplicável	Não aplicável	S	
N	...	Não aplicável	Não aplicável	...	
N	200909	Não aplicável	Não aplicável	Z	$CC = \text{Sum}(A, \dots, O, P, Q, R, S, \dots, Z)$

3.7 Resultados Finais

Após o desenvolvimento do modelo de CLTV interessa efectuar uma análise descritiva tanto das componentes utilizadas na obtenção do valor dos clientes como da própria métrica do CLTV (tabela 8). Desta forma foi possível retratar o nível de rentabilidade, nível de potencialidade e o valor dos clientes.

Tabela 8. Análise Descritiva das componentes utilizadas na obtenção do modelo de CLTV

Componente	Média	Desvio Padrão	Min	Max	1 Quartil	Mediana	3 Quartil
PCV	25.56	70.72	- 573.83	1609.45	- 191.78	5.13	45.23
RFCV	28.15	63.93	- 490.49	1999.29	- 100.09	23.61	27.11
FCV	34.62	57.23	- 245.37	2138.87	- 50.57	37.67	69.37
CLTV	29.44	63.96	-654.85	2873.81	- 114.15	49.14	76.24

Analisando a tabela 8 podemos concluir que, em média:

- existiu um nível de rentabilidade (PCV) positiva dos clientes (o que se traduz em lucro para o banco) entre o período de Julho de 2007 a Setembro de 2008, uma vez que $\overline{PCV} > 0$. De salientar que quanto maior for o valor de PCV maior será a rentabilidade associada aos clientes;
- o nível de potencialidade dos clientes num futuro considerado mais recente (RFCV), entre o período de Outubro a Dezembro de 2008, é positivo ($\overline{RFCV} > 0$) e superior ao nível de rentabilidade ($\overline{PCV} < \overline{RFCV}$), o que significa que ao longo do período em causa os clientes têm trazido maior valor e lucro para o banco. Tal como no caso do PCV, interessa realçar que quanto maior o valor de RFCV maior será a potencialidade num futuro considerado mais recente associada aos clientes;
- o nível de potencialidade dos clientes num futuro considerado mais longínquo (FCV), entre o período de Janeiro a Setembro de 2009, é positivo ($\overline{FCV} > 0$) e superior ao nível de RFCV ($\overline{FCV} > \overline{RFCV}$), o que significa que a potencialidade dos clientes será, em média, maior com o passar do tempo. Este facto pode-se explicar pelo crescente conhecimento dos padrões de conhecimento dos clientes existentes (segmentação e predição) no banco e pela adequação das estratégias de marketing sobre estes clientes;

- o valor esperado dos clientes (CLTV) no mês de Setembro de 2009 será positivo, o que significa que o banco terá clientes que, em média, são considerados como sendo valiosos e que contribuem para o lucro do banco.

Um outro aspecto importante de analisar diz respeito às estatísticas relativas ao 1Quartil, 2Quartil e 3Quartil, uma vez que estes dividem a distribuição das variáveis, ou seja, o conjunto ordenado de dados em 4 partes iguais, o que possibilita a compreensão dos dados em questão. Analisando a tabela 8, podemos concluir que:

- no que diz respeito à medida do PCV:
 - ✓ 25% da amostra ordenada tem uma rentabilidade negativa e inferior ou igual a -191.78;
 - ✓ 50% da amostra ordenada tem uma rentabilidade inferior ou igual a 5.13;
 - ✓ 75% da amostra ordenada tem uma rentabilidade inferior ou igual a 45.23;
- no que diz respeito à medida RFCV:
 - ✓ 25% da amostra ordenada tem uma potencialidade num futuro considerado mais recente negativa e inferior ou igual a -100.09;
 - ✓ 50% da amostra ordenada tem uma potencialidade num futuro considerado mais recente inferior ou igual a 23.61;
 - ✓ 75% da amostra ordenada tem uma potencialidade num futuro considerado mais recente inferior ou igual a 27.11;
- relativamente à medida FCV:
 - ✓ 25% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo negativa e inferior ou igual a -50.57;
 - ✓ 50% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo inferior ou igual a 37.67;
 - ✓ 75% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo inferior a ou igual 69.37;

- no que diz respeito ao valor dos clientes (CLTV):
 - ✓ 25% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo negativa e inferior ou igual a -114.15;
 - ✓ 50% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo inferior ou igual a 49.14;
 - ✓ 75% da amostra ordenada tem uma potencialidade num futuro considerado mais longínquo inferior ou igual a 76.24.

Após toda a análise dos resultados em termos de valores médios e de quartis importa perceber qual proporção de clientes com baixa, média e alta rentabilidade, potencialidade e valor. No caso prático desenvolvido sugere-se a realização de classes utilizando como valor de referência o valor 0 (tabela 9), onde se pode concluir:

- em Setembro de 2008 existia uma grande proporção de clientes com rentabilidade negativa (49%);
- em Dezembro de 2008 (futuro considerado mais recente) existia uma elevada proporção de clientes com potencialidade alta (61%);
- em Setembro de 2009 (futuro considerado mais longínquo) irá existir uma grande proporção de clientes com potencialidade alta (74%);
- em Setembro de 2009 (ultimo mês de cálculo do valor de cliente) cerca de metade da proporção de clientes analisados irão possuir elevado valor, tendo em conta o ciclo de vida dos mesmos até então (55%). Este facto pode ser explicado pela crescente utilização de técnicas estatísticas (por exemplo, através de modelos preditivos) que visam a descoberta de padrões, associações, tendências e previsões acerca do comportamento dos clientes existentes no banco, o que permite uma adequação das estratégias de marketing a efectuar sobre os clientes, e implicitamente, um aumento da potencialidade dos clientes e a obtenção de maiores lucros para o banco.

Tabela 9. Proporção de Clientes tendo em conta a rentabilidade, potencialidade e valor dos mesmos

	Componente	Valor de Referência	Descritivo	Proporção de Clientes	
	PCV	<0	Rentabilidade negativa	49%	
		=0	Rentabilidade Média	21%	
		>0	Rentabilidade Alta	30%	
	RFCV	<0	Potencialidade Recente Baixa	32%	
		=0	Potencialidade Recente Média Baixa	7%	
		>0	Potencialidade Recente Alta	61%	
	FCV	<0	Potencialidade Futura Baixa	11%	
		=0	Potencialidade Futura Média Baixa	15%	
		>0	Potencialidade Futura Alta	74%	
	CLTV	<0	Baixo Valor	31%	
		=0	Médio Valor	14%	
		>0	Alto Valor	55%	

Em suma, como resultado final do desenvolvimento do modelo de CLTV obtém-se uma tabela que retrata o tipo de rentabilidade, potencialidade e valor associado a cada um dos clientes no período de tempo em análise. A partir desta identificação e análise é possível seleccionar os clientes com baixo, médio ou alto valor, e efectuar campanhas de marketing específicas sobre os mesmos.

4. Conclusões e Investigação Futura

4.1 Conclusões

Tendo em conta toda a agitação, dinâmica, concorrência e competitividade existente actualmente, as organizações, e em especial as instituições financeiras, enfrentam alguns desafios relativamente à forma de como manter clientes fiéis, rentáveis e mais duradouros, já que os critérios de fidelidade dos clientes têm sofrido alterações. Com vista a contrariar esta tendência, as instituições estão, cada vez mais, interessadas no desenvolvimento e na implementação de processos e métricas que permitam a gestão de relacionamento com os clientes, e todo um conhecimento dos próprios clientes, muito em particular no que diz respeito ao valor da rentabilidade e potencialidade associado a cada cliente ao longo do tempo. Toda esta análise do valor dos clientes é de extrema importância, uma vez que para clientes com maior lealdade e valor, deverá existir um esforço adicional de retenção e satisfação dos mesmos em termos de ofertas de *cross* e *up sell*.

Existem algumas métricas (ditas tradicionais) que permitem medir o valor associado a cada cliente: (1) *Recency-Frequency-Monetary Value*; (2) *Past Customer Value* e (3) *Share of Wallet*. Contudo, as três métricas em causa encontram-se orientadas para a rentabilidade passada do cliente, e não para o comportamento futuro do cliente, o que por si só constitui uma grande desvantagem. Tendo em conta as desvantagens associadas à utilização das métricas tradicionais para efectuar o cálculo do valor dos clientes, surge a necessidade de utilizar uma outra métrica – designada por *Customer Lifetime Value* (CLTV). Esta métrica tem diversas vantagens associadas, comparando com as métricas tradicionais de obtenção do valor de cliente: (1) não é feita uma extrapolação do valor actual do cliente para o futuro; (2) é considerado o facto de o cliente existir futuramente no banco; (3) é analisado todo o comportamento do cliente ao longo do tempo e (4) obtém-se um *score* para cada cliente.

No caso prático apresentado foi desenvolvido e implementado um modelo de CLTV probabilístico para um determinado banco, já que o modelo em causa fornece uma previsão mais realista, precisa e robusta, do valor associado aos clientes ao longo do ciclo de vida dos mesmos. Isto porque, engloba uma noção da rentabilidade actual (no período de Julho de 2007 a Setembro de 2008) e da potencialidade num futuro próximo (entre Outubro de Dezembro de 2008) e num futuro considerado mais longínquo (entre Janeiro e Setembro de 2009) de cada cliente. De salientar que: (1) o modelo foi desenvolvido para clientes com posse de produtos

considerados nucleares e que originam, pelo menos, 80% da rentabilidade para o banco e (2) para a obtenção da potencialidade associada a cada cliente foi utilizada como *input* o resultado de modelos analíticos de predição (*cross sell*, *up sell* e abandono) e de análise de sobrevivência.

O modelo de CLTV probabilístico foi criado recorrendo à metodologia que o *SAS Institute* aplica no desenvolvimento e implementação de projectos de *Data Mining* – Metodologia “Implementação de Plataformas Inteligentes” (IPI). Esta metodologia foi a seleccionada, visto ser bastante completa, sequencial e comprovadamente eficaz. Esta metodologia é composta por diversas fases, sendo que cada uma delas apresenta características e funções distintas.

A partir do desenvolvimento e implementação deste modelo, utilizando a metodologia IPI, foi possível obter um *score* que reflecte o valor de cada cliente existente no banco, *score* este que resulta de uma análise comportamental por cada mês de análise, bem como do respectivo valor acumulado ao longo do ciclo de vida dos mesmos (rentabilidade e potencialidade).

Analisando o valor associado a cada cliente com produtos de créditos e recursos é possível identificar e seleccionar aqueles cujo valor é superior e com isto tornar os investimentos em marketing e pós-venda mais eficientes (ou seja, com menores custos). Uma vez que não são efectuadas campanhas desnecessárias sobre os clientes, estabelecendo-se acções e estratégias específicas e adaptadas a cada perfil de cliente, de acordo com os vários canais de comunicação. Estas acções possibilitam a existência de uma relação mais eficaz e eficiente dos clientes com o Banco, o que implicitamente origina uma maior duração do ciclo de vida dos clientes dentro do próprio banco (já que estes vêem as suas necessidades, expectativas, vontades e gostos satisfeitos) e uma maior rentabilidade do Banco sustentada ao longo do tempo, com vista a alcançar uma vantagem competitiva sobre os concorrentes.

4.2 Investigação Futura

O projecto desenvolvido teve como base a formulação e o desenvolvimento de um modelo de CLTV probabilístico para um determinado banco, que permitisse atribuir o valor de cada cliente no período de Julho de 2007 a Setembro de 2009.

Sugere-se que em investigações futuras sejam tidas em conta algumas das limitações inerentes ao desenvolvimento do projecto em causa, no que diz respeito:

- ao desenvolvimento do modelo de CLTV: para uma maior precisão e eficiência do modelo, sugere-se o desenvolvimento do modelo ao longo de todo o ciclo de vida dos clientes no banco, ou seja, desde o momento de aquisição dos clientes até ao momento de abandono dos mesmos no banco. Desta forma, será possível obter o valor que um dado cliente tem em cada instante do seu ciclo de vida, o que permitirá criar estratégias que permitam maximizar o ciclo de vida dos mesmos no banco, de acordo como o tempo de permanência;
- à formulação do modelo de CLTV:
 - ✓ obtenção da rentabilidade de cada cliente (PCV): para além da utilização das margens por produto associadas aos clientes, custos de manutenção dos clientes e taxa de juro, deveria utilizar-se outros tipos de custos, tais como: (1) custos de aquisição³⁹; (2) custos de marketing⁴⁰ e (3) custos de vendas⁴¹. Para a obtenção destes custos será necessária uma análise de custeio por actividade (*Activity Based Management: ABM*) com custos específicos por cada cliente. Desta forma será possível analisar e identificar a rentabilidade dos clientes com maior precisão;
 - ✓ obtenção do valor dos clientes num futuro considerado mais recente (RFCV): para além da utilização das margens previstas por produto associadas aos clientes, custos de manutenção futuros dos próprios clientes, taxa de juro futura prevista e os *scores* de modelos preditivos de *cross sell*, *up sell* e abandono, futuramente, sugere-se a integração de modelos de incumprimento relativamente aos produtos de crédito utilizados no

³⁹ Custos de adquirir um novo cliente.

⁴⁰ Custos de campanhas de marketing realizadas sobre o cliente.

⁴¹ Inclui custos de produção dos produtos e custos administrativos.

desenvolvimento do modelo (crédito à habitação, crédito pessoal, crédito especializado e cartões de crédito). Assim, será possível, obter uma maior precisão do risco potencial e da potencialidade num futuro considerado mais recente;

- ✓ obtenção do valor dos clientes num futuro considerado mais longínquo (FCV): Para além de todos os elementos utilizados como *input* para a obtenção do RFCV no caso prático⁴² deverá ser integrado a probabilidade de posse de cada um dos produtos utilizados no desenvolvimento do modelo de CLTV, por cada período de tempo em análise. De salientar que no caso prático apresentado esta probabilidade apenas foi calculada para a família de produtos de investimento e cartões de crédito.

Depois de analisadas as limitações subjacentes à formulação e desenvolvimento do modelo de CLTV, e que podem ser utilizadas como oportunidades futuras, interessa perceber o que se poderá fazer em termos de investigação futura. Assim, a investigação em termos de CLTV, poderá centrar-se nos seguintes aspectos:

- desenvolvimento de um modelo de CLTV mais sofisticado que incorpore os efeitos: (1) das mudanças que os clientes enfrentam perante a existência de alterações da quantidade e qualidade do *marketing mix*⁴³; (2) e o impacto do “*word-of-mouth*”⁴⁴ por parte dos clientes, seja este efeito positivo ou negativo e (3) redes sociais e de relacionamento entre os clientes;
- análise de quais os valores de referência que devem ser utilizados na obtenção das classes que traduzem o valor de cliente, de modo a que haja maior eficiência na identificação e atribuição do tipo de valor de cada cliente;

⁴² Margens futuras por produto associadas aos clientes, custos de manutenção futuros dos clientes, taxa de juro prevista, probabilidade dos clientes abandonarem o próprio banco.

⁴³ O *marketing mix* é constituído por quatro variáveis controláveis que permitem desenvolver acções: (1) produto; (2) canais de distribuição e venda; (3) promoção e (4) preço.

⁴⁴ O “*word-of-mouth*” traduz todo um conjunto de comunicações ditas de boca-a-boca. Este tipo de comunicação é uma “arma” bastante poderosa de marketing, uma vez que as comunicações feitas boca-a-boca são as mais influenciáveis para a adopção de uma determinada acção sobre um produto e/ou serviço. Isto porque, os clientes são influenciáveis por experiências vivenciadas por outras pessoas existentes na sua rede social.

- segmentação dos clientes tendo em conta o valor dos mesmo e de acordo com determinadas características ao nível do: (1) dinamismo; (2) consumo; (3) volume de envolvimento no banco (número ou montante de produtos e serviços) e (4) características sócio-demográficas. A partir desta segmentação, será possível obter grupos de clientes heterogéneos entre si e homogéneos dentro de si que permitam perceber quais as características dos clientes com elevado valor e de clientes com baixo valor;
- elaborar estratégias diferenciadas segundo o valor dos clientes. Desta forma será possível identificar e responder às necessidades específicas de cada cliente consoante o seu valor e otimizar o número de campanhas efectuadas sobre os mesmos. Assim, para clientes com elevado valor será necessário ter especial atenção, já que serão eles o que mais irão contribuir para o retorno de marketing e investimento;
- analisar o impacto de decisões estratégias efectuadas no que diz respeito à rentabilidade e potencialidade dos clientes. Depois de se diferenciar as estratégias a efectuar sobre os clientes, de acordo com o valor dos mesmos, interessa perceber se, na realidade, estas acções tiveram impactos significativos ou não no que diz respeito à retenção dos clientes e mesmo ao nível de retorno existente;
- reporte de informação, visto que, actualmente, não existe conhecimento sobre quais as melhores metodologias ou métodos de reporte de informação sobre os resultados inerentes ao modelo.

5. Referências Bibliográficas

- Bayer, Judy A. (2002), Show me the ROI. Plan, Deliver and Measure Serious Returns on Serious CRM, *White Paper Teradata*.
- Blattberg, Robert C. e Deighton, John (1991), Interactive Marketing: Exploiting the Age of Addressability, *Sloan Management Review*, Vol. 33, No.1, 17-30.
- Blattberg, Robert C. e Deighton, John (1996), Manage Marketing by the Customer Equity Test, *Harvard Business Review*, 74 (July-August), 136-144.
- Blattberg, Robert C., Gary Getz e Jacquelyn C. Thomas (2001), Customer Equity: Building and Managing Relationships as Valuable Assets, *Boston: Harvard Business School Press*.
- Berger, P.D. e N.I. Nasr (1998), Customer Lifetime Value: Marketing models and applications, *Journal of Interactive Marketing* 12, 17-30.
- Berry, L.L. (1983), 'Relationship Marketing' in Berry, L.L. Shostack, GL and Upha, GD (Eds), *Emerging Perspectives in Service Marketing*, American Marketing Association, Chicago, II, 25-8.
- Bejou D., Keiningham, T. e Aksoy, L. (2006), *Customer Lifetime Value: Reshaping the way we manage to maximize profits*. Best Business Books.
- Bishop C. (1995), *Neural Networks for Pattern Recognition*. Oxford Univ. Press.
- Breiman L., Friedman J.H., Olshen R.A., and Stone, C.J. (1984), "Classification and Regression Trees". Belmont, CA: Wadsworth.
- Brito, C. e Ramos, C. (2000), *Comércio Eletrônico – Relação com Parceiros de Negócios*. Porto: Sociedade Portuguesa de Inovação.
- Cardoso, M. S. e Gonçalves., (2001), *Cid. CRM em ambiente E-business*. São Paulo, Atlas.
- Deighton, J., Peppers, D., and Rodgers, M. (1994), Consumer transaction databases: Present status and prospects, em Blattberg, R., Rashid G. and Little, J. (Eds.): *The marketing information revolution*, Boston, MA: Harvard Business School Press, 58-79.
- Dràze, X. e Bonfrer, A. (2001), To pester or leave alone: Lifetime value maximization through optimal communication timing, *Working-paper*, Wharton-SMU Research Center.
- Dwyer, F.R. (1997), Customer Lifetime Valuation to support marketing decision making, *Journal of Direct Marketing*, 11 (4), 6-13.
- Ferrão, Francisco (2003), *CRM – Marketing e Tecnologia*. Escolar Editora, Lisboa, 7-113.
- Fox, Tricia e Steve Stead (2001), Customer Relationship Management-Delivering the Benefits, *White Paper of Secor Consulting*.
- Gupta, S., Lehmann, D. e Stuart, J. (2004), Valuing Customers, *Journal of Marketing Research*, 41 (1), 7-18.
- Hawkes, V.A. (2000), The Heart of the Matter: The Challenge of customer lifetime value, *CRM Forum Resources*, 1-10.
- Jutla, D., Craig, J. e Bodorik, P. (2001), Enabling and measuring electronic customer relationship management readiness, *Proceedings of the 34th annual Hawaii international conference on system sciences organizational systems and technologies track*, 1-10.
- Kalakota, R. e Robinson, M. (2001), *E-business 2.0 – Roadmap for success*. New York: Addison-Wesley.
- Kotler, P. (2000), *Administração de marketing: a edição do novo milênio*. São Paulo: Prentice Hall, 30, 72, 278.
- Kumar, V. (2006), CLV: The Databased Approach, *Journal of Relationship Marketing*, Vol.5, 7-35.
- Kumar, V., Customer Lifetime Value – The Path to Profitability, *Foundations and Trends in Marketing*, vol 2, no 1, 1-96, 2007.

- Kumar, V., Ramani, G. e Bohling, T. (2004), Customer Lifetime Value Approaches and Best Practices Applications, *Journal of Interactive Marketing*, 3, 60-72.
- Kumar, V. e Reinartz, W. (2006), *Customer Relationship Management – A Databased Approach*. New York, John Wiley & Sons.
- Limera, T., *Administração das comunicações em marketing*, In: Dias, S. R, Gestão de marketing, São Paulo: Saraiva, 2003, 7, 301.
- McKenna, Regis (1991), *Relationship Marketing: Successful Strategies for the Age of the Customer*. MA, Addison –Westly.
- Mitchell, T. (1997), *Machine Learning*. McGraw-Hill Companies.
- Ogden, D. (2009), Customer Lifetime Value Methodology: A Step-by-Step Approach to Quantifying and Optimizing Economic Returns, *SAS Global Forum 2009 Working Paper*, SAS Institute Inc.
- Oliver, R. L. (1997), *Satisfaction: a behavioral perspective on the consumer*. New York, McGraw Hill.
- Peppers, D. e Rogers, M. (1993), *The One to One Future: Building Relationships One Customer at a Time*. Currency Doubleday.
- Prahalad, C.K., Ramaswamy e Venkatram (2002), Co-opting Customer Competence – Chapter 1, *Harvard Business Review on Customer Relationship Management*.
- Reichheld, F.F., (1990), Zero Defection: Quality Comes to Service, *Harvard Business Review September-October*.
- Reichheld, F.F. (1993), Loyalty Based Management, *Harvard Business Review*, 71(4), 64-73.
- Reinartz, Werner e V. Kumar (2000), On the Profitability of Long-Life Customers in a Non contractual Setting: An Empirical Investigation and Implications for Marketing, *Journal of Marketing* 64 (4), 17-35.
- Reinartz, Werner e V. Kumar (2002), The mismanagement of customer loyalty, *Harvard Business Review* 80 (7), 86.
- Reinartz, Werner e V. Kumar (2003), The impact of customer relationship characteristics on profitable lifetime duration, *Journal of Marketing* 67 (1), 77-99.
- Rumelhart, D.E.; Hinton, G.E.; William, R.J. (1986), ‘*Learning internal representation by error propagation*’ in *Parallel Distributed Processing*, Cambridge, MA, MIT Press.
- Rust, Roland T., Valarie A. Zeithaml e Katherine N. Lemon (2002), Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy, *New York: Free Press*.
- Santos M.F. e Azevedo C., “*Data Mining, Descoberta de Conhecimento em Bases de Dados*”. FCA – Editora de Informática, 2005.
- SAS Institute Inc. (2003), *Data Mining using SAS Enterprise Miner: A Case Study Approach*. Second Edition, SAS Publishing, Cary, NC, USA.
- SAS Institute Inc. (2006), *SAS® Intelligence Platform Implementation Methodology*, Cary, NC, USA.
- Sharma, S. (1996), *Applied Multivariate Techniques*, John Wiley & Sons, Inc.
- Sheth, J.N. e Parvatiyar A. (1995), Relationship Marketing in Consumer Markets: Antecedents and Consequences, *Journal of Academy of Marketing Sciences*, 23(Fall), 255-71.
- Stone, M., Woodcock, N. e Wilson, M. (1996), *Managing the change from marketing planning to customer relationship management*. Long Range Planning, 29, 675-683.
- Venkatesan, R. e V.Kumar (2004), A customer lifetime value framework for customer selections and resource allocation strategy, *Journal of Marketing* 68 (4), 1’6-125.
- Winer, Russel S. (2001), A Framework for Customer Relationship Management, *California Management Review*, 42 (4), 89-105.
- Zupan J. e Gasteiger J., 1993, “*Neural Networks For Chemists: An Introduction*”. VCH, NewYork.

Anexo 1. Caracterização dos Dados

Tabela 3. Análise Descritiva das Características dos Clientes

Família	Característica	Estatística Descritiva	Valor
Socio-Demográfica do Cliente	Sexo do Cliente	Percentagem de Femininos	43,01%
		Percentagem de Masculinos	56,99%
	Antiguidade do Cliente	Desvio-Padrão	1,98
		Mínimo	10
		Média	17.3
		Máximo	64
		Mediana	17
	Idade do Cliente	Desvio-Padrão	13,8
		Mínimo	10
		Média	55.2
		Máximo	99
		Mediana	54
Produto Bancário	Número de Produtos	Desvio-Padrão	2,3
		Mínimo	1
		Média	5,6
		Máximo	15
		Mediana	5
	Montante de Recursos	Desvio-Padrão	120450
		Mínimo	0
		Média	31109,3
		Máximo	8362895,7
		Mediana	5359,1
	Montante de Activos	Desvio-Padrão	51518.7
		Mínimo	0
		Média	16412,5
		Máximo	1657900,2
		Mediana	0
	Saldo Médio em Depósitos à Ordem	Desvio-Padrão	15850.8
		Mínimo	0
		Média	4180.7
		Máximo	1344838.5
		Mediana	918.2

Anexo 2. Obtenção do Valor de Cliente num futuro considerado mais recente (RFCV)

Tabela 4. Técnicas utilizadas nos modelos vencedores de *Cross Sell*, *Up Sell* e Abandono

Família	Modelo	Árvore de Decisão	Regressão Logística	Rede Neuronal	Modelo de Mistura
<i>Cross Sell</i>	CC			✓	
	CE			✓	
	CH				✓
	CP			✓	
	FU		✓		
	PO	✓			
	SE			✓	
	DP		✓		
<i>Up Sell</i>	DP			✓	
	FU			✓	
	PO		✓		
	SE	✓			
Abandono	Banco				✓
	CC		✓		
	CH			✓	
	CP			✓	
	DP				✓
	FU		✓		
	PO	✓			
	SE			✓	

Tabela 5. Parametrização utilizada na modelação dos modelos vencedores de *Cross Sell*, *Up Sell* e Abandono.

Modelo	Técnica Vencedora	Parametrização
<i>Cross Sell</i> CC	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 15 Número Máximo de Iterações: 100 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
<i>Cross Sell</i> CE	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 3 Número Máximo de Iterações: 200 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
<i>Cross Sell</i> CH	Modelo de Mistura	Método de Previsão da Target: Obtida através da média dos valores previstos de um modelo de regressão logística e rede neuronal
<i>Cross Sell</i> CP	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 15 Número Máximo de Iterações: 75 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística

<i>Cross Sell</i> FU	Regressão Logística	Link Function: Logit Critério de Selecção de Variáveis: Backward Critério de Selecção do Modelo: Validation Misclassification Nível de Significância utilizado para a entrada de variáveis explicativas ('SLE'): 0.05 Nível de Significância utilizado para a saída de variáveis explicativas ('SLS'): 0.05
<i>Cross Sell</i> PO	Árvore de Decisão	Critério de splitting: ProbChisq Nível de significância: 0.05 Nível máximo de profundidade: 6 Número de nós descendentes após ramificação: 2 Número mínimo de observações que cada folha deve ter: 5 Método utilizado para a selecção da melhor árvore: Misclassification
<i>Cross Sell</i> SE	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 3 Número Máximo de Iterações: 50 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
<i>Cross Sell</i> DP	Regressão Logística	Link Function: Logit Critério de Selecção de Variáveis: Stepwise Critério de Selecção do Modelo: Validation Misclassification Nível de Significância utilizado para a entrada de variáveis explicativas ('SLE'): 0.05 Nível de Significância utilizado para a saída de variáveis explicativas ('SLS'): 0.05

<i>Up Sell DP</i>	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 15 Número Máximo de Iterações: 150 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
<i>Up Sell FU</i>	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 15 Número Máximo de Iterações: 80 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
<i>Up Sell PO</i>	Regressão Logística	Link Function: Logit Critério de Selecção de Variáveis: Stepwise Critério de Selecção do Modelo: Validation Misclassification Nível de Significância utilizado para a entrada de variáveis explicativas ('SLE'): 0.05 Nível de Significância utilizado para a saída de variáveis explicativas ('SLS'): 0.05
<i>Up Sell SE</i>	Árvore de Decisão	Critério de splitting: Gini Nível de significância: 0.05 Nível máximo de profundidade: 8 Número de nós descendentes após ramificação: 2 Número mínimo de observações que cada folha deve ter: 3 Método utilizado para a selecção da melhor árvore: Misclassification

Abandono Banco	Modelo de Mistura	Método de Previsão da Target: Obtida através da média dos valores previstos de um modelo de regressão logística e árvore de decisão
Abandono CC	Regressão Logística	Link Function: Logit Critério de Seleção de Variáveis: Backward Critério de Seleção do Modelo: Validation Misclassification Nível de Significância utilizado para a entrada de variáveis explicativas ('SLE'): 0.05 Nível de Significância utilizado para a saída de variáveis explicativas ('SLS'): 0.05
Abandono CH	Rede Neuronal	Arquitetura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 3 Número Máximo de Iterações: 50 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
Abandono CP	Rede Neuronal	Arquitetura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 10 Número Máximo de Iterações: 100 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística
Abandono DP	Modelo de Mistura	Método de Previsão da Target: Obtida através da média dos valores previstos de um modelo de regressão logística e rede neuronal
Abandono FU	Regressão Logística	Link Function: Logit Critério de Seleção de Variáveis: Stepwise Critério de Seleção do Modelo: Validation Misclassification

		Nível de Significância utilizado para a entrada de variáveis explicativas ('SLE'): 0.05 Nível de Significância utilizado para a saída de variáveis explicativas ('SLS'): 0.05
Abandono PO	Árvore de Decisão	Critério de splitting: ProbChisq Nível de significância: 0.05 Nível máximo de profundidade: 5 Número de nós descendentes após ramificação: 2 Número mínimo de observações que cada folha deve ter: 5 Método utilizado para a selecção da melhor árvore: Misclassification
Abandono SE	Rede Neuronal	Arquitectura: Multilayer Perceptron (MLP) Número de Unidades Escondidas: 8 Número Máximo de Iterações: 60 Tempo Máximo de Processamento: 2 Horas Técnica do Treino: Back Prop Função de Activação da Camada Intermédia: Logística Função de Activação da Camada Target: Logística

Tabela 6. Estatísticas de avaliação da qualidade dos modelos preditivos de *Cross Sell*, *Up Sell* e Abandono desenvolvidos.

Modelo	Curva de ROC⁴⁵	Sensibilidade⁴⁶	Especificidade⁴⁷	Misclassification Rate⁴⁸
Abandono Banco	0.77	37.41%	64.23%	0.38
Abandono CC	0.83	71.98%	45.82%	0.16
Abandono CH	0.97	52.98%	14.52%	0.07
Abandono CP	0.64	26.37%	44.79%	0.37
Abandono DP	0.90	43.05%	12.65%	0.19
Abandono FU	0.85	57.12%	23.01%	0.22
Abandono PO	0.69	25.07%	38.46%	0.35
Abandono SE	0.88	18.72%	27.29%	0.18
Cross Sell CC	0.79	21.56%	41.67%	0.28
Cross Sell CE	0.84	35.62%	22.12%	0.24
Cross Sell CH	0.92	18.91%	47.34%	0.15
Cross Sell CP	0.82	26.36%	31.98%	0.27
Cross Sell DP	0.76	38.89%	27.23%	0.29
Cross Sell FU	0.74	12.24%	27.89%	0.32
Cross Sell PO	0.70	27.83%	36.12%	0.35
Cross Sell SE	0.86	17.45%	38.34%	0.23
Up Sell DP	0.78	33.31%	29.67%	0.29
Up Sell FU	0.71	47.38%	15.29%	0.36
Up Sell PO	0.72	21.44%	42.89%	0.33
Up Sell SE	0.72	13.67%	19.56%	0.34

⁴⁵ A Curva de ROC é uma medida muito aplicada na avaliação da performance de modelos com resposta binária ilustra o *trade-off* existente entre o percentagem de respostas capturadas (sensibilidade) e a percentagem de falsos positivos (1- especificidade). Esta medida está associado ao poder discriminante de um teste de diagnóstico, uma vez que valores maiores deste índice representam maior poder discriminante, o que traduz melhor performance do modelo). Esta medida varia entre 0 e 1, sendo que valores acima de 0.5 é considerado como um bom modelo e valores abaixo de 0.5 um mau modelo.

⁴⁶ Medida da precisão com que se prevê os eventos (valor 1) do modelo.

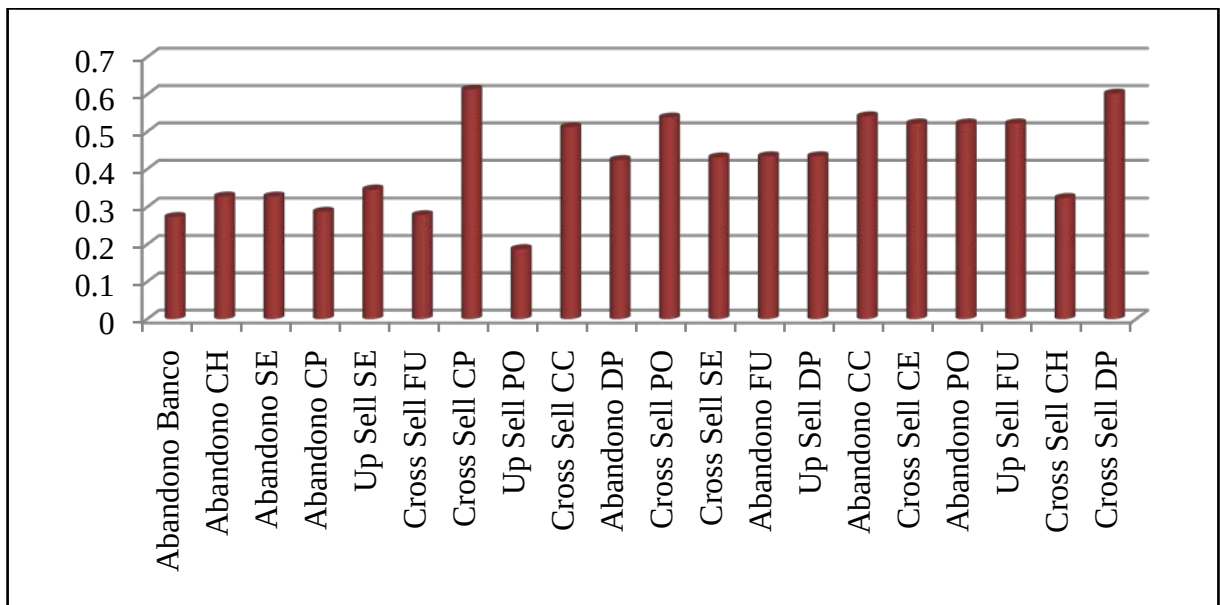
⁴⁷ Medida da precisão com que se prevê os não eventos (valor 0) do modelo

⁴⁸ Medida exprime a taxa de má classificação, ou seja, classificar o evento (definido como 1) como não evento (definido como 0) ou ainda, classificar o não evento (0) como evento (1).

Tabela 7. Análise Descritiva dos modelos de *Cross Sell*, *Up Sell* e Abandono desenvolvidos e considerados como vencedores.

Modelo	Média	Desvio Padrão	Mínimo	Máximo	1Quartil	2Quartil Mediana	3Quartil
Abandono Banco	0.273	0.303	0.000	1.000	0.014	0.117	0.503
Abandono CC	0.541	0.137	0.130	0.980	0.438	0.520	0.634
Abandono CH	0.328	0.299	0.110	0.880	0.116	0.147	0.514
Abandono CP	0.287	0.274	0.120	0.880	0.125	0.133	0.331
Abandono DP	0.424	0.211	0.020	1.000	0.305	0.323	0.490
Abandono FU	0.434	0.126	0.236	0.989	0.350	0.382	0.514
Abandono PO	0.522	0.090	0.440	0.880	0.462	0.488	0.544
Abandono SE	0.328	0.299	0.110	0.880	0.116	0.147	0.514
Cross Sell CC	0.512	0.103	0.104	0.921	0.298	0.395	0.603
Cross Sell CE	0.522	0.089	0.395	0.876	0.445	0.520	0.309
Cross Sell CH	0.324	0.289	0.02	0.990	0.475	0.559	0.789
Cross Sell CP	0.613	0.156	0.194	0.944	0.267	0.456	0.667
Cross Sell DP	0.602	0.078	0.521	0.910	0.543	0.587	0.634
Cross Sell FU	0.278	0.345	0.090	0.923	0.236	0.356	0.609
Cross Sell PO	0.538	0.089	0.103	0.878	0.321	0.498	0.723
Cross Sell SE	0.431	0.108	0.187	0.783	0.328	0.503	0.698
Up Sell DP	0.434	0.126	0.236	0.989	0.350	0.382	0.514
Up Sell FU	0.522	0.090	0.440	0.880	0.462	0.488	0.544
Up Sell PO	0.187	0.315	0.132	0.901	0.267	0.652	0.802
Up Sell SE	0.345	0.137	0.043	0.782	0.178	0.289	0.597

Figura 6. Score Médio dos Modelos Preditivos desenvolvidos para a obtenção do RFCV



Anexo 3. Obtenção do Valor de Cliente num futuro considerado mais longínquo (FCV)

Figura 7. Sintaxe utilizada no desenvolvimento dos modelos de Análise de Sobrevida

```
proc lifetest data="Nome_Tabela" Method=life intervals= "intervalos de tempo em que  
se pretende analisar" OUTSURV="tabela de output" plots=(s,ls,h) ;  
time "variável dependente"*"variável que identifica as observações censuradas" (0);  
run;
```