

ORIGINAL



**Universidade Católica Portuguesa  
Faculdade de Engenharia**

**Melhoria de Questionários de Satisfação Empresariais**  
Criação e Aplicação de Metodologias

**André Vilela de Freitas Ornelas**

**Dissertação para obtenção do Grau de Mestre em  
Engenharia Informática**

**Júri**

Prof. Doutor Rui Jorge Correia Mendes Alves Pires (Presidente)

Prof. Doutor Fernando Manuel Fernandes Melício

Prof. Doutor Tito Lívio dos Santos Silva (Orientador)

**Maio, 2012**

# Resumo

Um dos mais populares métodos de recolha de informação sobre uma dada população é o preenchimento de inquéritos e questionários. O uso deste tipo de angariação de informação tem a vantagem de abranger uma área geográfica praticamente ilimitada com o uso das novas tecnologias. No entanto, o conceito de questionário actualmente utilizado não garante a qualidade dos dados obtidos.

Os resultados extraídos de questionários de satisfação estão directamente relacionados com a tomada de decisão de uma empresa, nomeadamente na alocação de recursos e iniciativas de forma a dar resposta aos índices de satisfação dos serviços por ela prestada. Surge então a necessidade de garantir o mais elevado grau de qualidade dos dados obtidos pelos questionários.

Esta dissertação de mestrado visa apontar uma metodologia que i) determine e pondere critérios de qualidade, que variam consoante as necessidades; ii) aponte formas de desenhar esses critérios e de criar métricas para testar até que ponto esses critérios são atingidos ou melhorados comparando com estado anterior; iii) sugira formas de implementar as melhorias que respondam aos critérios de qualidade criados. A adopção da metodologia criada e testada nesta dissertação poderá ter a capacidade de garantir essa qualidade essencial. A aplicação de técnicas de endereçamento de perguntas, ponderação de respostas e a utilização de técnicas de *Data Mining* na criação de um método que averigua a veracidade das respostas dadas, contribuem de forma decisiva para a evolução do conceito actual de questionários e a forma como estes são administrados.

Esta dissertação apresenta ainda uma aplicação prática da metodologia criada, que ilustra a metodologia processual de melhoria de questionários.

**Palavras-chave:** Questionários *Online*, *Data Mining*, Regras de Associação, Ponderação de Respostas, Endereçamento de Perguntas

# Abstract

One of the most popular methods for collecting data about a population is filling out surveys. This type of data-gathering technique has the benefit of reaching a virtually unlimited geographic area with the use of new technology. However, today's concept of survey does not guarantee the quality of the data.

The results gathered from satisfaction surveys are directly related to a company's decision-making process which has direct impact in the allocation of resources and initiatives to meet improve the satisfaction rates for the services it provides. Therefore, the need to ensure the highest standards of data quality is of the utmost importance.

The aim of this master thesis is the definition of a methodology that i) determines and applies weights to quality criteria, which vary depending on the needs; ii) suggests ways to design these criteria and create metrics to test the extent to which these criteria are met or improved in comparison to previous questionnaires; iii) suggests ways to implement the improvements that meet the quality criteria. The adoption of the methodology created and tested in this master thesis may be able to provide this essential quality. The combined use of techniques for dynamically directing questions, weighted answers and the use of *Data Mining* techniques to create a response validation system, contribute to the evolution of the current concept of questionnaires and how they are administered.

This master thesis also presents a practical application of the methodology created, which illustrates the procedural methodology for improving questionnaires.

**Keywords:** e-Surveys, Data Mining, Association Rules, Weighted Answers, Directing Questions

# Agradecimentos

Agradeço ao meu Orientador, Prof. Doutor Tito Santos Silva que ao longo dos 6 meses me incentivou e me aconselhou quando mais precisava, contribuindo de forma crucial para a conclusão e entrega desta dissertação. Tendo sido uma presença constante ao longo de todo o percurso académico na Faculdade de Engenharia, tanto como professor como orientador, não tenho dúvidas que graças à sua presença e orientação aprendi e cresci como pessoa.

Tendo sido a dissertação efectuada na Tabaqueira II, S.A., queria agradecer a todos os colegas do departamento de *Information Services* pela forma fantástica como me receberam e por todo o apoio prestado sempre que necessitei. Queria deixar um agradecimento especial ao meu Co-Orientador Paulo Cardoso pelo apoio, sugestões, críticas e disponibilidade na hora em que mais precisava.

Aos meus pais e ao meu irmão pelo apoio e palavras de incentivo.

Por fim, mas não com menos importância, à minha namorada Joana por ter sido sem dúvida alguma uma das principais razões pela qual completei a dissertação. Por toda a força e incentivo que me deu e por ter estado sempre comigo nos melhores e piores momentos.



# Conteúdo

ÍNDICE DE FIGURAS .....	VII
ÍNDICE DE TABELAS .....	IX
LISTA DE ABREVIATURAS .....	XI
INTRODUÇÃO .....	1
1.1 MOTIVAÇÃO.....	1
1.2 OBJECTIVOS.....	1
1.3 ESTRUTURA.....	2
REVISÃO DA LITERATURA .....	3
2.1 ENQUADRAMENTO HISTÓRICO .....	3
2.2 O INQUÉRITO .....	5
2.2.1 <i>Introdução</i> .....	5
2.2.2 <i>Tipos e Estruturas de Inquéritos</i> .....	6
2.2.3 <i>Questionários Online</i> .....	8
2.2.4 <i>Qualidade das Perguntas</i> .....	10
2.3 QUESTIONÁRIO: PARTICIPAR OU NÃO? .....	11
2.3.1 <i>Introdução</i> .....	11
2.3.2 <i>Principais Factores</i> .....	12
2.4 ERROS E LIMITAÇÕES NOS DADOS.....	14
2.4.1 <i>Enquadramento</i> .....	14
2.4.2 <i>Dados Errados</i> .....	14
2.4.3 <i>Datasets Incompletos</i> .....	17
2.5 TRATAMENTO DE DADOS .....	18
2.5.1 <i>Detecção de Outliers</i> .....	19
2.5.2 <i>Tratamento de Missing Values</i> .....	21
2.5.3 <i>Últimas Considerações</i> .....	22
2.6 INQUÉRITOS DINÂMICOS.....	22
2.6.1 <i>Lógica Condicional</i> .....	23
2.6.2 <i>Seleção de Perguntas</i> .....	25
2.7 DESCOBERTA DE CONHECIMENTO EM BASES DE DADOS.....	26
2.7.1 <i>Introdução</i> .....	26
2.7.2 <i>Fases do Processo</i> .....	26
2.8 DISCUSSÃO .....	32
METODOLOGIA PROPOSTA .....	35
3.1 INTRODUÇÃO .....	35
3.2 LEVANTAMENTO DE REQUISITOS .....	36
3.2.1 <i>Requisitos Directos</i> .....	36
3.2.2 <i>Requisitos Indirectos</i> .....	37
3.3 AVALIAÇÃO DO QUESTIONÁRIO.....	37
3.3.1 <i>Introdução</i> .....	37
3.3.2 <i>Metodologia de Avaliação</i> .....	38
3.4 DESENHO.....	45
3.4.1 <i>Introdução</i> .....	45
3.4.2 <i>Endereçamento de Perguntas</i> .....	46

3.4.3	<i>Ponderação de Respostas</i> .....	50
3.4.4	<i>Regras de Associação</i> .....	55
3.5	<b>IMPLEMENTAÇÃO</b> .....	57
3.5.1	<i>Metodologia de Avaliação</i> .....	57
3.5.2	<i>Endereçamento de Perguntas</i> .....	58
3.5.3	<i>Ponderação de Respostas</i> .....	59
3.5.4	<i>Regras de Associação</i> .....	59
<b>RESULTADOS</b>	.....	<b>63</b>
4.1	<b>BACKGROUND</b> .....	<b>63</b>
4.2	<b>APLICAÇÃO DA FERRAMENTA DE AVALIAÇÃO</b> .....	<b>66</b>
4.2.1	<i>Criação de Escala</i> .....	66
4.2.2	<i>Extracção de Pontos</i> .....	67
4.3	<b>APLICAÇÃO DE ENDEREÇAMENTO DE PERGUNTAS</b> .....	<b>68</b>
4.3.1	<i>Recolha de dados</i> .....	68
4.3.2	<i>Aplicação de Filtros</i> .....	69
4.3.3	<i>Conjunto de Aplicações Identificadas</i> .....	75
4.4	<b>APLICAÇÃO DE PONDERAÇÃO DE RESPOSTAS</b> .....	<b>76</b>
4.4.1	<i>Características Identificadas</i> .....	76
4.4.2	<i>Atribuição de Pesos</i> .....	77
4.4.3	<i>Peso por Característica</i> .....	77
4.4.4	<i>Aplicação de Respostas Ponderadas</i> .....	80
4.5	<b>criação de Regras de Associação</b> .....	<b>83</b>
4.5.1	<i>Descrição dos Dados</i> .....	83
4.5.2	<i>Pré-Tratamento dos Dados</i> .....	84
4.5.3	<i>Utilização do Algoritmo</i> .....	85
<b>CONCLUSÕES</b>	.....	<b>88</b>
5.1	<b>DISCUSSÃO</b> .....	<b>88</b>
5.2	<b>LIMITAÇÕES</b> .....	<b>89</b>
5.3	<b>TRABALHO FUTURO</b> .....	<b>91</b>
<b>APÊNDICE A</b>	.....	<b>96</b>
<b>APÊNDICE B</b>	.....	<b>101</b>
<b>APÊNDICE C</b>	.....	<b>102</b>
<b>APÊNDICE D</b>	.....	<b>106</b>

# Índice de Figuras

Figura 2.1 – <i>Flow-chart</i> de etapas .....	5
Figura 2.2 – Componentes de um Questionário <i>Online</i> .....	9
Figura 2.3 – Exemplo de <i>outliers</i> num <i>dataset</i> bidimensional .....	15
Figura 2.4 – Fases cruciais de um projecto de Análise de Inquéritos .....	18
Figura 2.5 – Exemplo de <i>Branching Logic</i> .....	24
Figure 2.6 – Processo de Descoberta de Conhecimento em Bases de Dados .....	26
Figura 2.7 – Classificação: exemplo.....	29
Figura 2.8 – <i>Clustering</i> : exemplo .....	30
Figura 2.9 – Árvores de Decisão: exemplo.....	31
Figura 3.1 – Metodologia Adoptada .....	35
Figura 3.2 – Visão alto nível da aplicação da ferramenta de análise de qualidade.....	37
Figura 3.3 – Escala pontuação utilizando 4 pontos.....	42
Figura 3.4 - Exemplo <i>fitting</i> da Gaussiana .....	43
Figura 3.5 – Curva de critérios de elevada exigência .....	44
Figura 3.6 – Arquitectura do Sistema .....	45
Figura 3.7 – Filtros de Informação .....	49
Figura 3.8 – Visão Alto-Nível do Processo de Endereçamento de Perguntas .....	58
Figura 3.9 – Processo de Selecção de Aplicações Alvo de Perguntas .....	59
Figura 3.10 – <i>Flowchart</i> do algoritmo <i>Apriori</i> modificado .....	60
Figura 4.1 – Número de respostas ao questionário .....	66
Figura 4.2 – Apresentação Parcial da Escala Criada .....	67
Figura 4.3 – Informação extraída de IMDL sobre cargos dos utilizadores.....	68
Figura 4.4 – Informação extraída de IMDL sobre utilizadores dos sistemas .....	69
Figura 4.5 – Cargos do utilizador A (ID = 1000023535) .....	70
Figura 4.6 – Cargos do utilizador B (ID = 1000131225).....	70
Figura 4.7 – Cargos do utilizador C (ID = 1000023617).....	70
Figura 4.8 – Histórico de acessos do Utilizador A em SAP P1 .....	72
Figura 4.9 – Histórico de acessos do Utilizador B em SAP P1 .....	72
Figura 4.10 – Histórico de acessos do Utilizador C em SAP P1 .....	72
Figura 4.11 – Histórico de utilização de impressoras .....	75
Figura 4.12 – Funcionamento do Endereçamento de Perguntas.....	76
Figura 4.13 – Solução de questionários <i>web</i> adoptada pela Tabaqueira.....	80

Figura 4.14 – Índices Finais de Satisfação SAP CRM .....	83
Figura 4.15 – Folha de Cálculo com Respostas .....	84
Figura 4.16 – Folha de Cálculo com Respostas Tratadas .....	85
Figura 4.17 – Regras Extraídas pelo Algoritmo .....	85
Figura 4.18 – Regras Violadas no Questionário de 2010 .....	86
Figura A.1 – Selecção de Universos .....	96
Figura A.2 – Informações Disponíveis .....	98
Figura A.3 – Informações para Criação de Relatório .....	98
Figura A.4 – Condições Aplicadas à Informação .....	99
Figura A.5 – Relatório Criado .....	99
Figura A.6 – Relatório utilizando operador <i>Group By</i> .....	100
Figura C.1 – Processo em RapidMiner .....	102
Figura C.2 – Análise de importação de dados .....	103
Figura C.3 – Análise de transformação Nominal to Binominal.....	104

# Índice de Tabelas

Tabela 3.1 – Exemplo de atribuição de pontos .....	39
Tabela 3.2 – Avaliação <i>Manager A</i> .....	41
Tabela 3.5 – Avaliação Final <i>Managers</i> .....	41
Tabela 3.3 – Avaliação <i>Manager B</i> .....	41
Tabela 3.4 – Avaliação <i>Manager C</i> .....	41
Tabela 3.6 – Exemplo Classificação de Critério.....	42
Tabela 3.7 – Características de Utilizadores.....	52
Tabela 3.8 – Exemplo Genérico Atribuição de Pesos a Características .....	53
Tabela 3.9 – Exemplo Genérico Atribuição de Pontos a Característica .....	54
Tabela 4.1 – Escala de Pontuação de Taxas de Questionários Completos .....	67
Tabela 4.2 – Relatório de Pedidos de Suporte Utilizador A .....	73
Tabela 4.3 – Relatório de Pedidos de Suporte Utilizador B .....	74
Tabela 4.4 – Relatório de Pedidos de Suporte Utilizador C .....	74
Tabela 4.5 – Atribuição de pesos a características .....	77
Tabela 4.6 – Pontuação Característica A .....	77
Tabela 4.7 – Pontuação Característica B .....	78
Tabela 4.8 – Pontuação Característica C .....	78
Tabela 4.9 – Pontuação Característica D .....	78
Tabela 4.10 – Exemplo de características de dois utilizadores.....	79
Tabela 4.11 – Resultados SAP CRM 2010.....	81
Tabela 4.12 – Respostas com Pesos Fictícios.....	81
Tabela 4.13 – Resultados SAP CRM de Utilizadores de Sales .....	82
Tabela 4.14 – Resultados SAP CRM com Ponderação de Respostas.....	82



# Lista de Abreviaturas

BD – Base de Dados

CAI – *Computer-Assisted Interviewing*

DM – *Data Mining*

DW – *Data Warehouse*

ERP – *Enterprise Resource Planning*

IS – *Information Services*

KDD – *Knowledge Discovery in Databases*

SDLC – *Software Development Life Cycle*



# Capítulo 1

## Introdução

### 1.1 Motivação

O uso de ferramentas como inquéritos e questionários com o objectivo de obter *feedback* da população não é novidade. São ferramentas eficazes com a capacidade de recolher grandes quantidades de informação.

Ao longo dos anos, têm ocorrido mudanças e evoluções contínuas no que toca à análise de inquéritos e questionários. No entanto, o sucesso de um questionário tem forte dependência da sua estrutura e a forma como é administrado à população alvo. Alguns dos problemas mais recorrentes de uma estrutura fraca e limitada são:

- Baixas taxas de participação
- Fraca qualidade dos dados
- Pouca fiabilidade das respostas
- Questionários longos e desmotivantes

No sentido de resolver ou minimizar estes problemas, é necessário estudar e criar novas metodologias de elaboração de inquéritos, que não só motivem a pessoa a preencher mas que também gerem resultados fidedignos de forma a apoiar a tomada de decisão.

### 1.2 Objectivos

Face ao problema apresentado, esta dissertação tem como objectivo a criação de uma nova metodologia de elaboração de questionários para contribuir para a resolução dos problemas apresentados na secção anterior. Pretende-se melhorar o questionário de satisfação actualmente utilizado no mercado, tornando-o mais *user-directed*, propondo técnicas de apresentação e selecção de perguntas sobre serviços/aplicações que a população utiliza dentro da organização, e que esta metodologia possa ser aplicada em qualquer organização que disponibilize serviços.

Esta dissertação tem como objectivos principais:

- Criação de um método objectivo e quantitativo de avaliar a qualidade de questionários
- Criação de técnicas de valorização de respostas mediante características únicas de inquiridos
- Detecção de respostas potencialmente incorrectas com determinado grau de confiança associado
- Adaptação do questionário ao inquirido tendo em conta hábitos de trabalho

### **1.3 Estrutura**

Este documento está organizado da seguinte forma:

- No capítulo 2 *Revisão da Literatura* será efectuada uma introdução ao tópico de criação e investigação de questionários e inquéritos onde será feito um enquadramento histórico, seguido por um levantamento das diversas técnicas de criação de questionários e as suas limitações, bem como as soluções actuais respeitando o dinamismo do questionário
- No capítulo 3 *Metodologia Proposta* será apresentada a metodologia adoptada e as técnicas e métodos criados de forma a dar resposta aos problemas enunciados acompanhado de proposta de implementação
- No capítulo 4 *Resultados* serão apresentados os resultados provenientes da aplicação das técnicas explicadas no capítulo 3
- No capítulo 5 *Conclusões* é efectuada uma análise discussão dos resultados obtidos ao longo da dissertação, seguido de trabalho futuro e recomendação de melhorias na metodologia

# Capítulo 2

## Revisão da Literatura

### 2.1 Enquadramento Histórico

Actualmente, o uso de inquéritos e questionários é uma prática comum nas áreas científicas, sociais, económicas e outras. São efectuados tendo em conta um objectivo específico, seja ele a medição de qualidade de um produto, satisfação de um cliente ou estudos de mercado.

Um inquérito “providencia uma descrição quantitativa ou numérica de uma fracção da população – a amostra – através do processo de angariação de informação de efectuar questões a pessoas” (Fowler, 1988; Creswell, 1993). Tornou-se das mais importantes ferramentas ao dispor de uma organização e outras entidades, que transforma algo qualitativo, como uma opinião ou sugestão, em algo mensurável. É relevante termos uma visão da evolução histórica dos inquéritos para conseguirmos compreender a sua importância como instrumento de análise.

Desde o tempo dos Egípcios ao final do século XVII, usava-se a Estatística para fins diferentes da realidade quotidiana. O investimento neste tipo de disciplina era alimentado pelo Exército onde os interesses passavam pela quantificação do número de “cabeças”, estatísticas de pós-guerra e todo o tipo de informações relacionadas. Somente em 1885, após diversas reuniões de partes interessadas, foi fundado o *International Statistics Institute* (ISI) e por conseguinte, a entrada na era da estatística mundial (Heer, 1999). No entanto, foi no século XIX que se deram os maiores avanços na área.

Foi em 1912 que Bowley, primeiro professor de estatística na *London School of Economics and Political Sciences*, começou um estudo das condições da classe-trabalhadora em cinco capitais britânicas: Reading, Northampton, Stanley, Warrington, e Bolton. Utilizando uma estrutura de entrevista cuidada, foram seleccionadas amostras de cidadãos para coleccionar dados acerca das suas condições financeiras e hábitos de gastos (De Mast, 1997). Bowley teve especial cuidado na selecção das amostras fazendo deste estudo o primeiro inquérito científico com uso de amostragens.

O método de inquéritos tornava-se cada vez mais popular e a sua utilização mais frequente um pouco por todo o mundo, em especial nos Estados Unidos da América (EUA). Entre 1940 e começos de 1960, os EUA foram os principais pioneiros em novas técnicas e metodologias de inquéritos e análise de dados (Heer, 1999). Após a Segunda Guerra Mundial, novas

metodologias chegavam à Europa e diversas firmas foram criadas motivadas e inspiradas por centros de investigação conceituados nos EUA. Até meados do ano de 1970, inquéritos com base em entrevistas eram o método mais utilizado e o número de entrevistas cara-a-cara cresceu exponencialmente (Heer, 1999). Seguiu-se um período de reflexão e posteriormente o desenvolvimento e reavaliação de métodos alternativos.

Após um estudo efectuado na Alemanha, estimou-se que um em cada três indivíduos era alvo de um inquérito (Heer, 1999). Inevitavelmente, surgiu uma fadiga por parte da população e incerteza quanto à disponibilização de dados pessoais sem ganho próprio. A análise de inquéritos também revelou que não existia um *standard* de inquérito mas sim uma adaptação para cada pessoa. Limitações quanto aos dados obtidos, tendo em conta o objectivo dos inquéritos, emergiam de uma forma incontornável. A construção e modelo do questionário, o local de entrevista e a qualidade das perguntas são apenas alguns dos problemas mais complicados que iriam atormentar os especialistas.

Os anos 80 foram de enorme contribuição no que diz respeito a novas técnicas de efectuar inquéritos. A utilização de tecnologias emergentes influenciou a tipologia de inquéritos até à data. A utilização do telefone tornava menos dispendiosa a tarefa de efectuar estudos em áreas demográficas muito extensas. Com um maior acesso por parte da população a telefones e o estabelecimento de ligações para números aleatórios como técnica de amostragem, este tipo de inquérito tornava-se muito popular nos EUA (De Leeuw, 1992). Um dos sistemas mais populares na década era o CATI (*Computer-Assisted Telephone Interviewing*) que unia a utilização do telefone com o computador. Posteriormente, a criação de novos modelos, tirando partido da Internet como plataforma, começaram a surgir.

Não há dúvida que os anos 80 foram os anos de ouro no que diz respeito ao volume de dados angariados. No entanto, no final do século, tornou-se evidente que quantidade não era suficiente. A qualidade dos dados teria de ser garantida para se conseguir retirar conclusões dos estudos efectuados. Foi em 1990 que se realizaram duas importantes conferências. Realizou-se em Tucson, Arizona, uma conferência apenas para abordar dois temas: estimação de erros e redução de erros. Uma importante monografia sobre a conferência foi publicada em 1991 (Biemer *et al.* 1991). No mesmo ano, a primeira “*Internacional Workshop on Household Survey Nonresponse*” tomou lugar em Stockholm pela *Statistics Sweden* (primeira de muitas *workshops*). O conceito de “*Total Quality Management*” (TQM) introduziu metodologias de criação de inquéritos através de uma monitorização de todo o processo e disseminação dos resultados (Heer, 1999). Na década de 1990, houve também uma sensibilização quanto aos aspectos cognitivos inerentes à criação de questionários. A importância do processo cognitivo

dos indivíduos no acto de preenchimento de inquéritos, a forma como interpretam e abordam as questões, são elementos extremamente importantes para gerar respostas com valor.

A realidade em que vivemos actualmente, mantém em grande parte a sua base em metodologias que foram introduzidas e apresentadas como “último grito” mas tirando partido da Internet de uma forma eficiente.

## 2.2 O Inquérito

### 2.2.1 Introdução

Não há dúvida que os avanços tecnológicos ao longo da história têm sido os principais catalisadores de progressos nas mais variadas áreas da ciência. Recordando o uso do telefone como meio para efectuar inquéritos como um sucesso (De Leeuw, 1992), o mesmo se pode aplicar ao uso emergente da Internet como meio de unificação global.

Muitos investigadores dedicam o seu tempo ao *design* do inquérito. Uma metodologia adoptada por Walonick (1997), sugere que todo o processo de desenvolvimento de um inquérito deve seguir um *flow-chart* de etapas onde não deverão haver desvios do processo para manter a qualidade do inquérito (Figura 2.1).

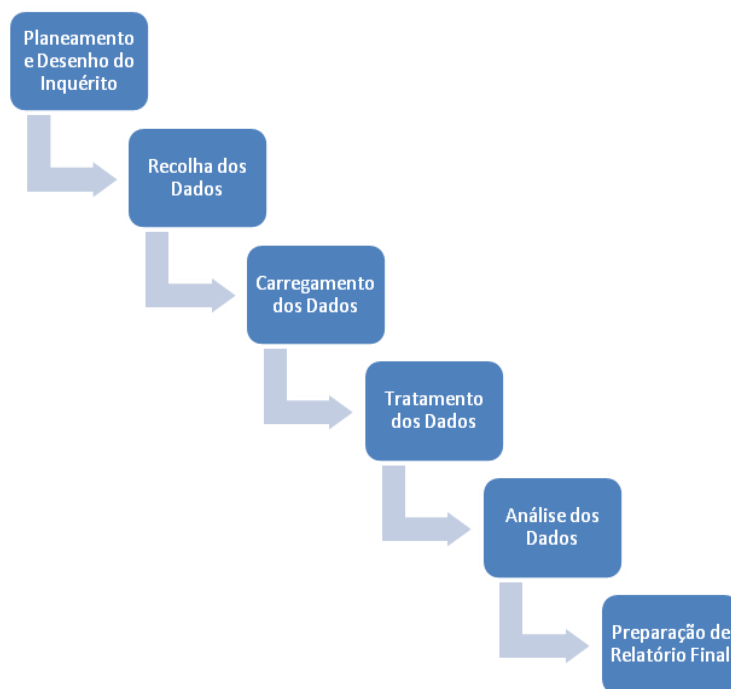


Figura 2.1 – *Flow-chart* de etapas

Seguindo as etapas identificadas em cima, o inquérito terá uma estrutura melhorada e cuidada. Por vezes, a qualidade dos inquéritos e dos seus resultados não é a esperada sendo as razões mais frequentes a fraca qualidade das perguntas e o processo de entrada dos dados no sistema.

Actualmente, o uso de inquéritos efectuados em papel tem perdido terreno face às novas gerações de inquéritos. Mas é importante referir que, ainda que represente uma pequena percentagem, o indivíduo ao receber o inquérito por *e-mail* ou por acesso a sites, continua a ter a possibilidade de imprimir, preencher e enviar o inquérito.

Analisemos agora os diferentes tipos de inquéritos mais comuns nos dias que correm.

### **2.2.2 Tipos e Estruturas de Inquéritos**

Independentemente do tipo de inquérito escolhido, o modelo e o tipo de perguntas mantêm-se os mesmos. Este terá de ter uma boa apresentação e estrutura para que se reduza o número de não-respostas e erro de medida.

O inquérito deverá ser construído e organizado de forma a:

- Motivar os inquiridos a preenchê-lo
- Facilitar a sua compreensão usando linguagem fácil e apropriada
- Capturar informação relevante, alinhando as perguntas efectuadas com os objectivos do questionário

Existe um leque de tipos de perguntas que podem ser feitas a um indivíduo. Em cada abordagem, será apresentado o seu propósito e as suas vantagens.

#### **2.2.2.1 Perguntas Abertas vs Perguntas Fechadas**

A escolha entre estes dois tipos de perguntas desde cedo gerou diferentes opiniões. Na década de 1940, ocorreu a disputa entre duas divisões do *Bureau of Intelligence* dos EUA (ver Converse, 1987) onde uma divisão defendia o uso de perguntas abertas enquanto a outra apoiava o uso de perguntas fechadas.

Perguntas abertas caracterizam-se pela liberdade que é dada ao indivíduo de responder à questão através do uso de frases de texto corrido onde pode expressar a sua opinião utilizando as suas próprias palavras.

#### **Vantagens Perguntas Aberta**

- Respostas contêm mais informação, incluindo sentimentos, atitudes e compreensão do tópico em questão

- Possibilidade de captura de conhecimento realmente inovador e não previsto inicialmente
- Não dá hipótese ao indivíduo de responder sempre com a mesma resposta (responder a todas as perguntas com “Talvez” ou “Não”) com o objectivo de preencher rapidamente

Perguntas fechadas caracterizam-se pela obrigatoriedade do indivíduo seleccionar uma das opções disponibilizadas pelo questionário quando responde a uma questão. O indivíduo perde a liberdade de responder em texto corrido e é confinado a um certo número de respostas (“sim/não”, “verdade/falso” e outras).

#### Vantagens Perguntas Fechadas

- Facilidade de análise dos dados em termos computacionais
- Preenchimento do inquérito torna-se, em média, mais rápido
- Facilidade em comparar respostas de diferentes indivíduos
- Apresentação de respostas pode clarificar o significado da pergunta

Com o passar do tempo, as perguntas fechadas tornaram-se populares e as perguntas abertas cada vez menos utilizadas (Smith, 1987).

#### **2.2.2.2 Escalas**

Tipicamente, as perguntas estruturadas desta forma oferecem uma escala apenas com o valor inicial e final devidamente identificados (índices de satisfação) onde o meio é preenchido apenas com números, letras ou palavras. Foi, no entanto, provado que o uso de palavras em toda a escala melhora a experiência e facilita a decisão do indivíduo ao responder às perguntas (Krosnick e Berent, 1993). Existe sempre o problema sistemático das presunções dos indivíduos, que se caracteriza por estes, por vezes, acreditarem que uma pessoa “normal” ou típica acaba por “cair” no meio da escala e respondem de modo a se inserirem nesse intervalo independentemente da escala adoptada (Schwarz *et al.* 1985).

Este tipo de estrutura é útil por conseguir capturar opiniões de uma forma mais ampla graças à escala apresentada que, normalmente, apresenta 5 valores escolhas (Ex: Muito Insatisfeito, Insatisfeito, Neutro, Satisfeito e Muito Satisfeito). A escolha das frases que identificam as escolhas é importante e deve facilitar a compreensão da pergunta. A utilização de escalas não só partilha algumas das vantagens-chave das perguntas fechadas – fácil de analisar e preenchimento rápido – mas também introduz a capacidade de compreender o estado de espírito do indivíduo.

### 2.2.3 Questionários Online

Com o avanço tecnológico, a forma de administrar e preencher questionários tem sofrido contínuas alterações. *Computer-assisted Interviewing* (CAI) é o método actual de administrar questionários ou inquéritos utilizando um computador. Este método, como foi referido no início do capítulo, não é recente, no entanto tem existido contínua evolução e métodos que tiram partido do computador têm sido criados e adaptados.

Ao método CAI são atribuídas diversas vantagens face aos questionários clássicos de papel-e-lápis. Algumas das vantagens apontadas são (Bethke, 2008):

- Validação de respostas durante o preenchimento
- Redução de erros de origem humana
- Encaminhamento de perguntas complexo e sofisticado

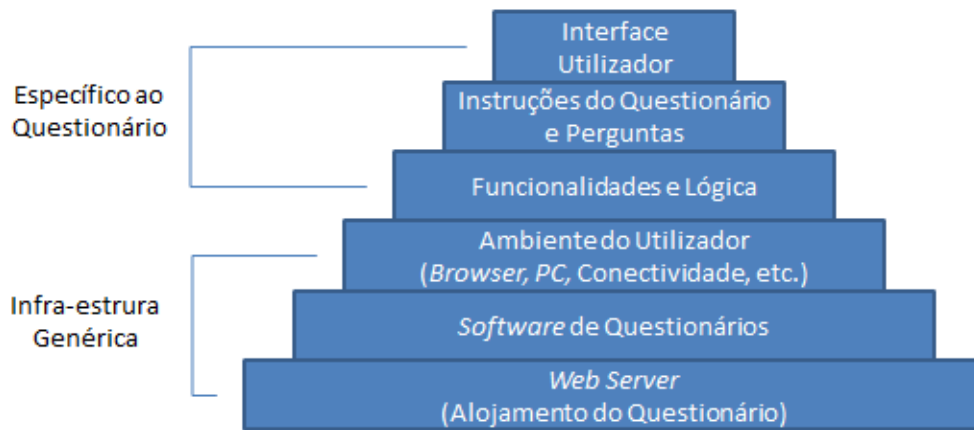
A utilização deste tipo de questionários introduz funcionalidades que garantem, a um primeiro nível, a integridade das respostas. Perguntas de escolha múltipla, onde o utilizador tem que escolher apenas uma das diversas opções, são computacionalmente fáceis e garantem que apenas uma resposta seja seleccionada através de verificações e limitações impostas pelo próprio questionário. No caso de questionários de papel-e-lápis, o indivíduo pode seleccionar mais que uma resposta ou assinalar respostas de uma forma ilegível.

No que toca ao armazenamento de respostas, os sistemas CAI são mais vantajosos pelo facto de não necessitarem de esforço humano para armazenar. Após todos os questionários terem sido preenchidos, no caso dos questionários CAI as respostas são automaticamente guardadas em Bases de Dados (BD), enquanto nos questionários mais antigos é necessário a introdução manual de toda a informação num computador para efectuar futura análise de resultados.

Nos sistemas CAI é possível enriquecer as perguntas posteriores ao ter a capacidade de utilizar respostas a perguntas anteriores, tornando a experiência mais positiva e mais próxima do utilizador.

Questionários em papel utilizam notas e instruções quando o utilizador não precisa de responder a certas perguntas. No entanto, é comum o utilizador não compreender e não seguir o caminho que o investigador pretende. Estes casos geram situações em que os utilizadores não respondem a perguntas quando na realidade tinham capacidade para o fazer. No caso dos questionários CAI, implementa-se lógica complexa para que o caminho a percorrer seja efectuado sem desvios inesperados.

Após a apresentação de algumas vantagens dos questionários CAI sobre os questionários clássicos, surge a necessidade de introduzir os principais componentes de um questionário *online*. Identificam-se 6 componentes agrupadas em 2 categorias: Infra-estrutura Genérica e Especifico de Questionário (Baker *et al.* 2004). Na Figura 2.2 são apresentados os diversos componentes onde na base se encontram os componentes genéricos e no topo os específicos.



**Figura 2.2 – Componentes de um Questionário Online**

O *Interface Utilizador* consiste na formatação gráfica e na estrutura utilizada para mostrar os ecrãs com a informação do questionário: *layout* do ecrã, cores, fontes, gráficos, botões de navegação, etc.

A componente *Instruções do Questionário e Perguntas* inclui a apresentação das perguntas e respectivas escolhas de respostas, instruções acerca das perguntas gerais e específicas e textos introdutórios do questionário.

*Funcionalidade e Lógica* refere-se à forma como o utilizador é guiado durante o preenchimento do questionário. Algumas das técnicas que traçam o caminho do utilizador são: *branching logic* condicional, aleatoriedade, possibilidade de voltar a uma pergunta anterior, etc. Esta componente também abrange diversas funcionalidades como validação de respostas, apresentação de itens, inserção de texto, etc.

A componente *Ambiente do Utilizador* consiste nas configurações de *hardware* e *software* necessárias para o correcto funcionamento do questionário. Algumas configurações importantes são a compatibilidade do *browser* com o *software* do questionário, propriedades do monitor, velocidades de *download*, capacidade de memória e processamento, entre outros.

*Software de Questionários* é a ferramenta ou a linguagem de programação utilizada na codificação do questionário. Todas as actividades acima referidas são implementadas e geridas

pelo *software* incluindo a lógica condicional, interações com o utilizador e a forma de guardar as respostas e dados do utilizador numa base de dados ou qualquer outra forma de armazenar dados.

Por fim, a componente *Web Server* define em termos de *hardware* e *software* a plataforma onde o questionário será alojado. É esta componente que gere possíveis sobrecargas garantindo *performance* e disponibilidade do questionário. Também gere configurações de segurança, *firewalls*, sistemas de *backup*, entre outros.

Qualquer inconsistência ou problema num dos 6 componentes implica uma degradação na experiência do utilizador no acto de preenchimento do questionário. Entre os possíveis problemas destacam-se:

- Uma fraca interface dificulta a leitura e compreensão das perguntas
- Uma fraca implementação pode gerar diferentes interpretações de perguntas em *browsers* diferentes
- Erros no estabelecimento da lógica e funcionalidades do questionário podem levar à não apresentação de perguntas ao utilizador ou à apresentação de perguntas por erro
- Uma fraca implementação do *Web Server* a nível de tempo de carregamento da página e deficiências ao guardar informação causa frustração ao utilizador podendo este no limite abandonar o questionário prematuramente

Existem diversas vantagens para o formato digital de questionários ser utilizado, no entanto acarreta sérias responsabilidades na fase de criação e implementação do mesmo. Identificaram-se em cima algumas novas ameaças na utilização deste formato que, caso não sejam prevenidas, poderão resultar num questionário destinado ao fracasso.

#### **2.2.4 Qualidade das Perguntas**

Fowler (1995) sugeriu que “independentemente do tamanho da amostra, da quantidade de dinheiro investido na recolha de dados e o tempo de resposta, a qualidade dos resultados obtidos de um inquérito nunca será melhor que a qualidade das perguntas”. Esta é uma questão de extrema importância mas que por vezes é descurada pelos investigadores. Não existe utilidade para dados que não têm qualquer valor. Tendo em mente a importância da forma como apresentamos uma questão, é possível obter qualidade se esta obedecer a algumas regras essenciais (Walonick, 1997):

1. Deve ser usada apenas uma dimensão de análise. É importante saber formular uma pergunta para obter a resposta que se espera. Por exemplo, um investigador que

pretenda saber o nível de satisfação de um utilizador quando este usa os ecrãs de uma aplicação, poderá colocar a pergunta “Está satisfeito com a apresentação dos menus e das janelas?”. Se a resposta for “não”, é impossível o investigador saber se o utilizador desgosta dos menus ou das janelas, ou de ambos.

2. Devem ser disponibilizadas todas as respostas possíveis. Este ponto é especialmente importante em perguntas de escolha múltipla. O indivíduo deve ter ao seu dispor uma lista exaustiva de respostas que deve incorporar qualquer tipo de resposta possível (por vezes é impossível prever a totalidade de respostas).
3. Deve ter opções exclusivas. Em cada pergunta, salvo por opção do investigador, existe apenas uma resposta certa ou apropriada para cada pergunta. É importante não introduzir ambiguidades nas perguntas com o objectivo de não confundir ou frustrar o indivíduo.
4. Deve induzir variabilidade de respostas. Para obtermos conclusões e informação útil, uma pergunta deverá ser colocada para que os inquiridos não escolham sempre a mesma opção. Uma pergunta apenas deve ser feita se houver necessidade e se as respostas trouxerem nova informação. Por exemplo, a pergunta “É contra o racismo?”. Obviamente, as respostas irão cair de uma forma esmagadora para o “sim”, não trazendo qualquer informação adicional sobre o assunto.
5. Deve ser aplicável a qualquer indivíduo. Não devem ser feitas perguntas que se baseiem em suposições. Perguntar o nível de satisfação de um serviço/produto a um indivíduo, que no momento não usufruiu dele, poderá gerar dados incorrectos no estudo. O indivíduo sente-se na obrigação de “adivinhar” a resposta que o investigador está à espera, ou irá deixar a resposta em branco.
6. Deve ser de rápida compreensão e toda a população deverá conseguir responder. Um questionário com perguntas pouco claras torna o seu preenchimento lento e poderá não só causar confusão, mas também causar desinteresse e frustração por parte do indivíduo.

## **2.3 Questionário: Participar ou não?**

### **2.3.1 Introdução**

Questionários concedem ao investigador informação valiosa que descreve uma população alvo com erros estimados de amostragem. No entanto, a descrição depende directamente da vontade dos indivíduos seleccionados participarem no estudo. Nos subcapítulos seguintes, serão introduzidos os efeitos nefastos causados pela não participação de indivíduos e as respostas

incompletas que são fornecidas. No entanto, é interessante aprofundar e perceber as razões para tal fenómeno acontecer.

Existem diversos factores que influenciam um indivíduo a participar num questionário podendo estar directamente relacionados com o indivíduo ou com o próprio questionário. Ao longo deste subcapítulo, serão abordados os factores que mais influenciam os participantes a responder.

### **2.3.2 Principais Factores**

Para totalmente compreender as decisões que levam o indivíduo a participar num questionário, é necessário um levantamento dos principais factores influenciadores. Por um lado, as influências sociodemográficas e o *design* do questionário e, por outro, a influência de factores psicológicos do indivíduo (Groves *et al.* 1992).

#### **2.3.2.1 Factores Sociodemográficos**

Dos primeiros factores destacam-se os seguintes:

- Factores Sociais – certas características do meio social em que o questionário é efectuado influenciam a taxa de participação. O sentido de responsabilidade social dos indivíduos, o número de questionários administrados nesse meio e a própria legitimidade do questionário são factores importantes.
- Estrutura do Questionário – como foi referido no subcapítulo 2.2, a estrutura e formulação de perguntas de um questionário deverá seguir uma estrutura reflectida para maximizar a taxa de participação. Para além de obedecer a regras de boa conduta na criação de questionários, outro factor que influencia a participação é a dimensão do mesmo e o tempo que é necessário dispensar para o preencher na totalidade.
- Características do Indivíduo – as próprias características sociodemográficas dos participantes estão relacionadas com a adesão ao questionário. Entre as mais importantes destaca-se a idade, sexo, situação económica e estado de saúde.

Os factores apresentados são os percebidos pelo investigador. No entanto, uma investigação mais profunda leva-nos à compreensão das questões e decisões a nível psicológico e cognitivo. A análise e compreensão destes factores contribuem para traçar a estratégia que será adoptada pelo administrador do questionário para maximizar a aceitação e preenchimento do mesmo.

### 2.3.2.2 Factores Psicológicos Sociais

Tipicamente um indivíduo toma a decisão de realizar uma actividade com base no interesse e relevância da actividade juntamente com o tempo, energia e recursos necessários para a fazer. Esta decisão segue um processo heurístico composto por diversos princípios (Groves *et al.* 1992):

- Reciprocidade – qualquer sociedade humana rege-se por normas de reciprocidade que levam um indivíduo a agir da mesma forma que um outro indivíduo se dirigiu a ele. Exemplos desta relação encontram-se diariamente: um indivíduo que seja abordado de uma forma positiva responderá da mesma maneira positiva e vice-versa. Segundo Groves, estas normas apontam para uma situação causa-efeito: um indivíduo será mais recíproco a pedidos se a sua participação gerar um tipo de recompensa que crie algo de positivo ao indivíduo (prémio, ofertas, favores, etc.).
- Consistência – investigadores de psicologia social em diversas alturas deram conta que a maioria das pessoas desejava ser o mais consistente possível nas suas atitudes, decisões, valores e crenças, e que dificilmente mudariam a sua posição no futuro. Deduz-se que se o administrador do questionário conseguir efectuar uma ligação entre o conjunto de valores e crenças do indivíduo, o administrador aumentará a probabilidade do indivíduo participar.
- Validação Social – o Homem sendo um animal social, tende a procurar outros com um conjunto de valores e crenças parecidas para tomar uma determinada decisão. Como resultado, indivíduos tendem a abordar um problema da mesma forma que um outro indivíduo abordou um problema semelhante. Mencionar uma elevada adesão da população ao questionário é uma das técnicas que tira partido desta predisposição de procurar acções de outros para validar a sua.
- Autoridade – as pessoas estão mais dispostas a aceitar um pedido se este tiver como origem uma entidade detentora de uma autoridade legítima aos olhos de toda a sociedade. A aplicação deste princípio a questionários passa pela enfatização do patrocinador do questionário quando este está a ser administrado. No entanto, o uso desta técnica pode revelar-se negativa podendo prejudicar a participação. Por um lado, se o patrocinador for uma organização positivamente conotada (governo, educação, etc.) pode surtir efeitos positivos. Por outro lado, se o patrocinador pertencer a uma classe de organizações comerciais dedicadas exclusivamente a esta actividade por vezes pode ter conotações negativas, baixando desse modo a taxa de participação.

- Escassez – quando oportunidades se mostram escassas, estas começam a ser encaradas como mais valiosas. É com esta premissa que o administrador do questionário tenta induzir um sentido de escassez à oportunidade de participar. Na tentativa de cativar o indivíduo a participar, o administrador passa a ideia que a sua opinião é extremamente valorizada pelo facto de apenas 1 em cada 300,000 pessoas é seleccionada para responder, sendo esta uma oportunidade única para ele.

A compreensão dos factores psicológicos sociais que influenciam a decisão de participar, juntamente com análises custo-benefício do nível de adesão a questionários, dota o administrador de conhecimento valioso, provado ao longo dos anos, para maximizar a taxa de adesão ao questionário. No entanto, nem todos os indivíduos passam por todo o processo heurístico descrito em cima, podendo a decisão de participar depender apenas de um ou dois processos (ex: dimensão do questionário ou autoridade do *sponsor*).

## **2.4 Erros e Limitações nos Dados**

### **2.4.1 Enquadramento**

Desde cedo foram identificadas inúmeras limitações e erros nos dados recolhidos através de inquéritos. O investigador e analista depara-se, inevitavelmente, com diversos tipos de limitações que não estão sob o seu controlo. Problemas como a validade dos dados, dados incompletos e qualidade dos dados são alguns exemplos.

O desenvolvimento de metodologias e técnicas até à data, não se resumem apenas a estruturas de inquéritos e à formulação de perguntas, mas também à antecipação de erros. Neste capítulo serão apresentados alguns dos problemas que mais comprometem a qualidade dos dados finais e conclusões retiradas.

### **2.4.2 Dados Errados**

Tendo sido apresentados nos capítulos iniciais os diversos problemas com que o investigador se depara, é importante perceber que poderão existir erros nos dados. Os erros podem ter origem no inquirido ou no sistema informático.

Os erros nos dados de origem humana por vezes são impossíveis de tratar e é por isso que devem ser tomadas medidas preventivas. Tomemos como exemplo uma série de respostas a um inquérito que utiliza uma escala de 1 a 5 em cada pergunta. Diferentes respostas geram diferentes erros. Se o inquirido responder um valor que não faça parte da escala (ex: 8), a sua identificação como erro é imediata levando à sua não inclusão no conjunto dos dados finais. No entanto, existe outro erro que não é possível ser identificado por computador. Tomemos agora o

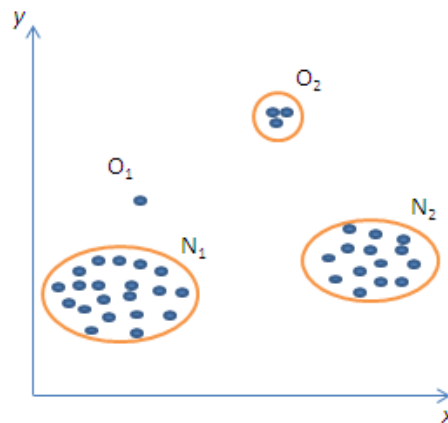
exemplo de um acto de venda ao público. No acto do pagamento, para ser finalizada uma compra, vamos supor que o sistema necessita que seja introduzido o número de contribuinte do comprador para a finalizar. Se o comprador não souber o número ou não o quiser facultar, a pessoa que está a processar o pagamento é forçada a introduzir um número aleatório da mesma dimensão para que possa finalizar a compra. Este tipo de erro é impossível ser identificado por um computador, pois o valor introduzido está de acordo com o exigido.

Deverão ser tomadas medidas preventivas para evitar este tipo de erros. Uma medida a ser aplicada poderá passar pela não inclusão de perguntas de resposta obrigatória num questionário. A utilização deste tipo de pergunta não só desmotiva o inquirido pela forma “grosseira” com que obriga a ser dada uma resposta, mas também pela criação de uma forma de recolher dados errados, como foi exemplificado acima.

Como foi anteriormente referido, nem todos os erros são impossíveis de tratar. Analisemos agora um dos erros que mais compromete os resultados de um inquérito: *Outliers*.

### ***Outliers***

*Outliers* são padrões existentes em dados que não se enquadram num comportamento considerado normal (Chandola *et al.* 2007). Podemos então considerar que uma amostra de uma população cai num padrão normal, com fronteiras bem definidas, ou cai fora sendo considerado um *outlier*.



**Figura 2.3 – Exemplo de *outliers* num *dataset* bidimensional**

A Figura 2.3 ilustra *outliers* num *dataset* bidimensional onde são identificadas duas regiões normais,  $N_1$  e  $N_2$ .  $O_1$  é uma instância de *outlier*, enquanto  $O_2$  é uma região de *outliers* (região que agrupa diversas instâncias “anormais”).

A existência de *outliers* nos dados pode ser induzida de diversas formas: fraude, intrusão, diferenças no ambiente, erro humano, etc. São vistos, à primeira vista, como anomalias nos dados sendo a sua resolução trivial se considerarmos a sua remoção da amostra. No entanto, estes dados têm vindo a ganhar o interesse do analista. Na verdade, podemos caracterizar um *outlier* como ruído ou como dados válidos. Diversas técnicas de detecção foram criadas e aperfeiçoadas para efectuar esta mesma distinção.

Como sabemos se um conjunto de dados, que aparentemente não segue um padrão normal, não é algo de realmente novo e merecedor de investigação? São com perguntas como estas que a detecção de *outliers* se distingue de técnicas de remoção de ruído (Teng *et al.* 1990) ou acomodação de ruído (Rousseeuw e Leroy, 1987), ambas técnicas de eliminação ou aplicação de modelos estatísticos para acomodar *outliers*.

Como ilustrado na Figura 2.3, a definição de uma fronteira que delimita os dados normais, aparentemente seria suficiente para detectar *outliers* na amostra. No entanto, vários factores tornam esta abordagem difícil (Chandola *et al.* 2007).

- É extremamente difícil estabelecer uma fronteira que englobe todos os comportamentos normais
- Tendo em conta que vivemos num mundo extremamente dinâmico, a fronteira que hoje engloba todos os comportamentos normais pode não reflectir os comportamentos no futuro
- Se uma observação estiver muito próxima da fronteira entre normal e *outlier*, existe a hipótese de um dado normal ser considerado um *outlier* e vice-versa
- A natureza dos dados e o domínio em que se insere, altera a definição e as características que tornam dados em *outliers*, o que torna mais complexo o seu tratamento

Torna-se crucial efectuar o tratamento de *outliers* antes de iniciar a análise de dados. Se não forem despendidos esforços neste sentido, a inclusão deste tipo de dados em análise pode comprometer gravemente o estudo e a qualidade dos resultados.

No entanto, existem casos em que uma entrada aparenta ser um *outlier* mas que não deve ser tratado como tal. A sua baixa ocorrência, juntamente com o seu valor aparentemente fora do normal, pode ser interpretado como não tendo qualquer significado, quando na verdade deverá ser contabilizado e deverá ser alvo de estudo.

Para compreendermos melhor estas excepções, tomemos como exemplo as medições sísmicas em Portugal. Diariamente ocorrem sismos em todo o país mas a esmagadora maioria não são

sentidos pela população e apenas são registados pelo Instituto de Meteorologia de Portugal. Em média, os sismos têm uma magnitude de 1 a 2,5 na escala de Richter. Se uma análise simples de detecção de *outliers* for aplicada e se forem descartados os valores que não obedecem ao intervalo proposto pelo investigador (tomemos por exemplo o intervalo 0 a 6), um dos maiores acontecimentos da história de Portugal nunca aconteceu. Segundo a análise de sismos de Portugal, o Terramoto de Lisboa de 1755, cuja magnitude estimada chegou a 9 na escala de Richter, nunca terá acontecido segundo o estudo efectuado. Este é apenas um dos exemplos em que dados que caem fora de um certo padrão têm informação extremamente útil mas que podem ser encarados como *outliers*.

### 2.4.3 Datasets Incompletos

Um *Dataset* é o *input* primário para o processo de tratamento e análise. É o conjunto de dados, ainda em bruto, recolhidos através dos processos de captura de informação (inquéritos, questionários, etc.).

Em análise de inquéritos, é comum o aparecimento de dados incompletos. As razões são das mais variadas podendo ser por esquecimento, por opção ou por erro computacional. Diversos métodos têm sido propostos para o tratamento de dados incompletos em amostras de inquéritos (Mundfrom *et al.* 1998). Abordagens mais recentes estão a ser estudadas no domínio de *Data Mining* (Grzymala-Busse e Hu, 2000). Um exemplo deste tipo de abordagem é o uso de classificação supervisionada e classificação não-supervisionada.

*Datasets* incompletos são extremamente penalizadores para o trabalho/estudo desenvolvido pelo investigador. É então importante tentar minimizar o não preenchimento de questões num inquérito. Para isso, é necessário perceber algumas das causas (Young *et al.* 2011):

- Por vezes, a não resposta a uma ou mais perguntas pode dever-se à falta de dados. Se não existirem dados para responder a essa pergunta, certamente esta irá ser deixada em branco. Por exemplo, numa investigação num dado país revelou-se a inexistência de informações acerca de atributos físicos da população feminina (van Schoor e Konz, 1996). Neste caso os investigadores sabiam a causa da falta de informação e puderam estimar os valores por ratios, que foram obtidos através da informação da população masculina do país.
- Algumas das perguntas efectuadas em inquérito podem ser de natureza muito sensível. Perguntas como a etnia ou situação financeira causam desconforto ao inquirido, mesmo sabendo que está protegido por anonimato.

- É importante ter em mente que quanto mais perguntas tiver o inquérito, mais tempo irá ser necessário dispensar para o responder. O investigador deve manter um número razoável de perguntas mas que garantam a captura de informação pretendida.
- Em inquéritos baseados na Web, uma boa apresentação do inquérito e a inclusão de instruções de usabilidade, facilitam e motivam a pessoa a responder ao inquérito.

Outras questões prendem-se com a estrutura do inquérito. Na secção 2.2 foram apresentadas considerações importantes na criação de um inquérito. É de realçar, novamente, que a qualidade das perguntas e a forma pouco clara com que podem ser apresentadas influenciam bastante o não preenchimento de perguntas

## 2.5 Tratamento de Dados

Na secção 2.4, foram introduzidas duas grandes ameaças no que toca à análise de dados de inquéritos. *Datasets* incompletos, juntamente com dados errados, obrigam os investigadores a adoptar diversas técnicas para efectuar o tratamento dos dados. Em qualquer projecto que trate da análise de inquéritos, existem sempre três fases cruciais (Figura 2.4).



**Figura 2.4 – Fases cruciais de um projecto de Análise de Inquéritos**

A fase de Recolha caracteriza-se pelas diversas actividades relacionadas com a captura e recolha de informação. As respostas dos inquéritos realizados *online* ou por *e-mail* são guardadas em Bases de Dados Relacionais, folhas de cálculo ou outros formatos.

A fase de Tratamento é das fases mais cruciais de todo o projecto, onde inconsistências nos dados e falta de informação são tratados para que os resultados da fase de Análise estejam menos sujeitos a erros.

A última fase, de Análise (após o devido tratamento, organização e consistência do *dataset* ter sido garantido) é a fase em que se aplicam técnicas estatísticas aos dados tendo em conta o objectivo do estudo, e são gerados os resultados e conclusões.

Nas secções seguintes, serão apresentadas algumas técnicas para efectuar o Tratamento dos dados, tais como técnicas de detecção de *datasets* com erros e técnicas de tratamento de *datasets* incompletos.

### 2.5.1 Detecção de *Outliers*

Este tipo de problema é resolvido através da aplicação de técnicas de detecção. Cada técnica de detecção recebe um importante *input*: o tipo de *outlier*. Estes podem ser classificados em três categorias distintas (Tipo I, Tipo II e Tipo III) baseadas na sua composição e na relação com o resto dos dados.

#### 2.5.1.1 Tipos de *Outliers*

##### Tipo I

É o tipo mais simples e alvo da maioria das técnicas actuais de detecção. Consiste no processamento dos dados como uma distribuição estática, identificação dos pontos mais afastados e sinalização desses pontos como potenciais *outliers*. Cada entrada do *dataset* tem valores únicos em cada atributo sendo esses valores inconsistentes face aos valores “normais”, é considerado um *outlier* (Hodge *et al.* 2004). Este tipo caracteriza-se por não ter qualquer conhecimento prévio dos dados.

Existem duas sub-técnicas que se podem aplicar, *diagnosis* e *accommodation* (Rousseeuw *et al.* 1987). Uma abordagem *diagnosis* identifica e marca os potenciais *outliers* que, após serem identificados, o sistema opta por descartar deixando de fazer parte do conjunto de dados que posteriormente será analisado. A abordagem *accommodation*, após identificar os possíveis *outliers*, tenta acomodá-los na análise de dados. Incluem-se esses dados mas são criadas protecções contra eles. Para tal, são efectuadas modificações no modelo base e/ou nos modelos de análise. Uma possível resolução é a atribuição de um peso reduzido a essa instância. Ao atribuir menos importância a estas observações, o impacto nos resultados é atenuado.

##### Tipo II

Tal como *outliers* do Tipo I, trata-se de entradas individuais do *dataset* mas no entanto, distinguem-se pelo sistema ter conhecimento prévio dos dados. São aplicados rótulos nas entradas como sendo normais ou anormais. Pode ser feita uma análise mais pormenorizada através da rotulação prévia dos dados. Imaginemos que os dados considerados normais estão agrupados numa área A, que está dividida em três zonas distintas (A1, A2 e A3). Quando um novo dado aparecer, podemos efectuar uma simples análise normal ou *outlier*, ou podemos ir mais longe e classificar com mais pormenor que tipo de dados normais foram encontrados.

### Tipo III

Este tipo foca-se na modelação da fronteira de normalidade (Japkowicz *et al.* 1995). Consegue processar dados em ambientes estáticos ou dinâmicos. Este tipo de detecção é por vezes chamado de *novelty detection*. Tal como o Tipo II, esta abordagem necessita de rotulação prévia dos dados mas apenas se foca nos dados rotulados como normais. Através da entrada de novos dados, o modelo é aprendido de forma incremental, conseguindo “afinar-se” para conseguir organizar melhor os dados.

É importante ter em mente que *outliers* do Tipo I podem ser detectados em qualquer tipo de dados. O Tipo II e Tipo III requerem a presença de estruturas sequenciais ou espaciais.

#### 2.5.1.2 Técnicas de Detecção

A maioria das técnicas existentes provém de várias disciplinas distintas (Estatística, Teoria da Informação, *Data Mining*, etc.).

A técnica de classificação tem como principal objectivo a aprendizagem de um grupo de instâncias devidamente classificadas (treinar) e posteriormente classificar novas instâncias numa das classes aprendidas (testar). Neste método podem ser identificadas duas classes: *normal* e *outlier*.

São então identificadas duas fases distintas nesta abordagem: fase de treino e fase de teste. A fase de treino cria um modelo de classificação através dos dados que o sistema tem na altura. Através das características de cada instância e a sua classificação, o sistema consegue criar um modelo. Após ter sido criado, o modelo é então testado ao receber novos dados – não classificados – e é efectuada a classificação dos mesmos. Estas técnicas são consideradas **técnicas de detecção de outliers supervisionadas**.

Em contraste com problemas de classificação normais que tentam separar os dados em duas (ou mais) classes de objectos, existem técnicas que identificam apenas uma classe e tentam separá-la de todos os outros tipos de dados. Estas técnicas são consideradas **técnicas de detecção de outliers semi-supervisionadas**. Através destas técnicas, é isolado com sucesso um certo tipo de dados como sendo normal. Quando estiver criado e devidamente testado, o modelo é alterado para rejeitar este objecto e dessa forma rotular como *outlier*.

As técnicas acima apresentadas, partem de um ponto onde inicialmente existe uma rotulação das classes ou dos diversos valores que o resultado pode tomar. No entanto, existem **técnicas de detecção de outliers não-supervisionados** que apenas recebem os *inputs*. A forma como o sistema aprende a detectar *outliers* baseia-se na representação dos *inputs* de forma a tomar decisões, prever valores futuros, etc. Em contraste com as técnicas apresentadas anteriormente,

estas técnicas não são influenciadas minimamente pelo ser humano, fazendo com que o algoritmo retire conclusões e faça escolhas por si próprio. Um exemplo clássico deste tipo de técnica é *clustering* (agrupamento).

### 2.5.2 Tratamento de *Missing Values*

O tratamento de *outliers* por si só não é suficiente para garantir resultados fiáveis. O aparecimento de *missing values* pode ser causado por factores externos ao inquirido ou pode dever-se ao inquirido.

Existem diversos métodos para tratar este tipo de limitação tendo em conta a percentagem de *missing values* no *dataset*: 1% incompleto é considerado trivial e de 1-5% são tratáveis. No entanto, de 5-15% requerem métodos sofisticados, e mais de 15% impacta severamente qualquer análise (Acuna e Rodriguez, 2004).

Analisemos alguns métodos que podem ser utilizados para o seu tratamento (Acuna e Rodriguez, 2004).

- **Case Deletion (Apagar Selectivo)** – Também conhecido como *complete case analysis*, é o método *default* em muitos *softwares* informáticos. Este método consiste em descartar as entradas (instâncias) que contêm valores não preenchidos num ou mais atributos. Uma variação deste método consiste em analisar as instâncias e atributos tendo em conta valores incompletos. As instâncias e/ou as classes que tiverem mais valores incompletos são descartados. Na aplicação deste método é necessário ter especial atenção na eliminação de atributos. Por vezes, mesmo que haja um elevado grau de dados incompletos, esses dados poderão ser relevantes para a análise em questão.
- **Mean Imputation (Introdução por Média)** – É um dos métodos mais utilizados que consiste no preenchimento dos valores incompletos de um atributo. O valor é estimado através da média de todos os valores das entradas que estão preenchidos, referentes ao atributo em questão.
- **Median Imputation (Introdução por Mediana)** – Tendo em conta que o tratamento de *outliers* não é 100% eficaz e que podem não ser identificados todos os *outliers*, a introdução por média pode não ser o melhor método. Daí surge um método alternativo que estima os valores incompletos através da mediana dos dados existentes de um dado atributo.

Existem inúmeros métodos e algoritmos de preenchimento de *missing values* onde são identificadas duas classes de algoritmos de introdução:

1. *Single imputation* (Introdução Única) - Este tipo de algoritmos escolhe apenas um valor para preencher o campo a vazio. Alguns exemplos destes algoritmos são:
  - *Hot Deck*: Este algoritmo procura no *dataset* completo, um exemplo (instância) que tenha características idênticas ao exemplo incompleto e preenche os campos vazios com os mesmos valores do exemplo encontrado. Se não houver um exemplo que esteja completo, então utiliza-se o exemplo que tiver menos valores incompletos.
  - *Naive-Bayes*: Naive-Bayes é uma técnica de classificação probabilística simples (Farhangfar *et al.* 2008). É um algoritmo computacionalmente eficiente porque apenas necessita de uma passagem pelos dados de treino para inferir o valor.
2. *Multiple imputation* (Introdução Múltipla) - Nesta abordagem, são seleccionados múltiplos valores para serem introduzidos. O funcionamento deste tipo de algoritmo passa pela criação de bases de dados paralelas com os *missing values* preenchidos com as várias alternativas de valores. Tem a vantagem de posteriormente serem comparadas entre si e de se conseguir escolher a que melhor se adapta ao problema em questão (Buuren e Groves, 1999). Buuren e Oudshoorn criaram uma plataforma de introdução múltipla designada MICE (*Multivariate Imputation by Chained Equations*), que disponibiliza um conjunto de distribuições condicionais e métodos baseados em regressões.

### 2.5.3 Últimas Considerações

É importante ter presente que métodos de introdução de dados devem ser utilizados com especial cuidado. Por um lado o investigador fica com o *dataset* completo que irá enriquecer os resultados finais, mas por outro é necessário ter em mente que os valores foram de certo modo “falsificados” e que isso deverá ser considerado no final. Por vezes é melhor trabalhar com uma amostra reduzida mas que descreva o suficiente o problema do que trabalhar uma amostra de grande dimensão mas com demasiados valores introduzidos *a posteriori*.

## 2.6 Inquéritos Dinâmicos

Foram abordadas técnicas no domínio da Inteligência Artificial no tratamento de erros, nomeadamente do tratamento de *outliers* e *missing values*. A área da Inteligência Artificial poderá ter mais aplicações no que toca ao aumento da qualidade do inquérito e da forma como os inquiridos respondem às questões.

A realização de um inquérito estático num mundo actualmente tão dinâmico é algo desapropriado. É especialmente interessante a criação e o desenvolvimento de inquéritos *online* que sejam únicos para cada utilizador. Actualmente ainda são desenvolvidos inquéritos onde as perguntas são previamente seleccionadas e a estrutura do inquérito fixa para qualquer indivíduo, independentemente do seu *background*, situação actual e respostas a perguntas anteriores.

Hoje em dia, existem inúmeros *sites* que disponibilizam ferramentas para criar questionários dinâmicos aplicando funcionalidades que, quando implementadas correctamente, transmitem uma experiência positiva e pessoal a cada utilizador. Para o dinamismo ser alcançado, utilizam-se diversas técnicas entre elas:

- *Branching*
- *Piping*
- *Looping*

Todas as técnicas acima enunciadas assentam na correcta utilização de Lógica para alcançar o dinamismo.

### **2.6.1 Lógica Condicional**

Lógica Condicional caracteriza-se pela investigação das propriedades lógicas e semânticas de certas frases que ocorrem em linguagem natural (Nute, 2002). As declarações condicionais contêm as palavras ‘se’ e ‘então’. Apontam-se como exemplos de declarações desta natureza:

1. Se estivesse com calor, então tiraria o casaco
2. Se o João jogasse no Euro milhões, então poderia ganhar o prémio

Os dois exemplos apresentados dividem-se em duas frases ou cláusulas, onde uma introduz uma condição e outra uma afirmação que depende da condição. Esta técnica é a base adoptada por métodos bastante usados em questionários *online*: *Branching Logic* e *Skip Logic*. Tipicamente estas duas técnicas são utilizadas em simultâneo, tirando dessa forma partido das vantagens e funcionalidades das duas.

#### ***Branching Logic***

Este método, através de lógica condicional, selecciona a próxima pergunta a mostrar mediante uma análise da resposta que o utilizador forneceu na pergunta anterior, assemelhando-se a uma árvore de decisão onde cada nó é uma pergunta e cada ramo uma escolha/resposta. A Figura 2.5 apresenta um exemplo simples da aplicação desta técnica.

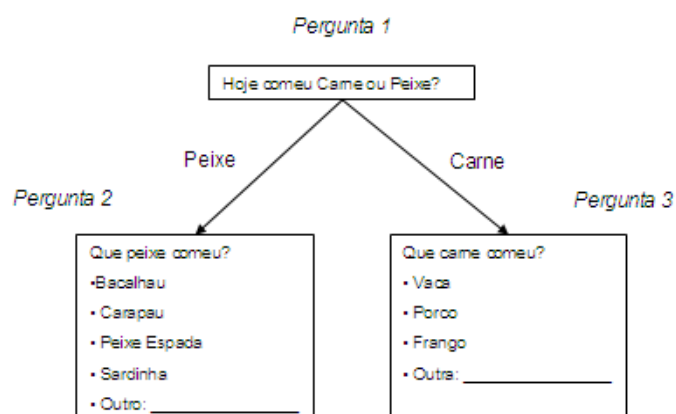


Figura 2.5 – Exemplo de *Branching Logic*

A resposta à Pergunta 1 condiciona a apresentação da próxima pergunta uma vez que, mediante a resposta do utilizador, uma das perguntas seguintes nunca lhe será apresentada. Em termos de lógica condicional, no exemplo descrito identificam-se duas regras: (1) Se a resposta à Pergunta 1 for “Peixe”, então mostrar Pergunta 2; (2) Se a resposta à Pergunta 1 for “Carne”, então mostrar Pergunta 3. Este método visa, por um lado, diminuir o número de perguntas do questionário e por outro, não apresentar perguntas para as quais o utilizador não consegue fornecer uma resposta válida.

### ***Skip Logic***

Este método assemelha-se ao *Branching Logic* no sentido em que também usa lógica condicional para efectuar decisões. A resposta de uma primeira pergunta determina se uma questão ou um grupo de questões seguintes será apresentado ao utilizador. Um exemplo deste método é possível encontrar-se numa avaliação de um item segundo uma escala de satisfação, onde mediante a sua avaliação poderão surgir perguntas mais específicas sobre o item em questão (Ex: caso a avaliação seja negativa).

Uma das lacunas destes métodos é a elevada dependência das perguntas e respostas dadas pelo utilizador. O facto de não existir informação de *background* do utilizador e utilizar exclusivamente as respostas para efectuar decisões, o questionário não tem forma de endereçar as perguntas de forma realmente direccionada limitando-se apenas a introduzir bifurcações no questionário.

Este método torna o questionário mais pessoal e ganha um carácter mais individual mesmo que não seja inteiramente perceptível ao utilizador. No entanto, têm sido feitos esforços para melhorar o dinamismo dos inquéritos. A questão da escolha das perguntas foi alvo de estudo e

desenvolvimento de novas metodologias de criação de inquéritos, nomeadamente de selecção das perguntas.

A criação de aplicações que tirem partido de regras torna o sistema mais robusto e inteligente. Um sistema deste tipo tem em conta três aspectos vitais num questionário (Urbano, 2008):

- As dependências entre as várias perguntas e respostas
- Decisão quanto à selecção da próxima pergunta a ser apresentada
- Detecção de inconsistências entre as várias respostas

Com o sistema concebido de modo a ter em conta estas questões, o questionário torna-se mais flexível e adaptável ao utilizador, garantindo também a consistência das respostas.

### 2.6.2 Selecção de Perguntas

Na maioria das soluções actuais de inquéritos *online*, a selecção da pergunta baseia-se apenas na resposta pontual do utilizador, que dita a próxima pergunta a ser seleccionada. No entanto, existe outro tipo de informação que deve ser considerada na decisão. Essas informações adicionais podem então ser utilizadas para a criação de regras para seleccionar próximas perguntas. Algumas informações relevantes para a escolha das perguntas são:

- O número de vezes que uma questão foi colocada ao utilizador e deixada em branco
- No acto de fazer *login*, saber se é a primeira vez que acede ao *site*
- Registrar a última pergunta não respondida pelo utilizador (para não cair no erro de mostrar novamente e desmotivar o inquirido)

As informações acima referidas são utilizadas para criar um sistema de selecção de perguntas dinâmico. Um exemplo de regras lógicas que utilizam informações do tipo acima referido poderá ser (Urbano, 2008):

*Se o utilizador efectuou login e é a primeira vez então mostrar pergunta 1.*

A regra seguinte ilustra a complexidade e graus de liberdade que poderá ter este tipo de abordagem:

*Se um utilizador efectuou login e não é a primeira vez então escolher uma pergunta que ainda não tenha sido respondida e que seja diferente da última pergunta apresentada.*

Esta última regra utiliza as três informações apresentadas anteriormente para tomar a decisão. A utilização de regras nos questionários possibilita uma personalização e um automatismo que antes não era possível com simples análise das respostas dos utilizadores enquanto estes preenchiem.

## 2.7 Descoberta de Conhecimento em Bases de Dados

### 2.7.1 Introdução

No capítulo anterior foram apresentadas soluções e técnicas utilizadas actualmente para a concepção de um questionário dinâmico. No entanto, existem técnicas que, após correcta implementação, levam o termo dinamismo um passo à frente. A utilização de DCBD e *Data Mining* vem introduzir novas possibilidades e vem revolucionar a forma de construção e utilização de inquéritos ou questionários.

Descoberta de Conhecimento em Bases de Dados (DCBD), é o processo que permite a aplicação de técnicas de *Data Mining* para identificar padrões ou tendências nos dados (Santos e Ramos, 2009). Ao longo deste subcapítulo serão descritas as várias fases deste processo com maior incidência na fase de *Data Mining*.

### 2.7.2 Fases do Processo

O processo DCBD é constituído por várias fases, como ilustrado na Figura 2.6, que são executadas de forma **iterativa** – uma vez que podem existir retrocessos a etapas anteriores – e também **interactivo** pelo facto de requerer interacção do utilizador sempre que for necessário tomar uma decisão.



Figure 2.6 – Processo de Descoberta de Conhecimento em Bases de Dados

### 2.7.2.1 Selecção

A primeira fase visa delimitar o espaço de pesquisa e a redução do número de atributos existentes na fonte de dados. A redução do número de atributos reflecte-se na minimização da complexidade e poder de processamento necessário para criar os modelos de DM, isto deve-se ao facto dos atributos serem interpretados como uma **dimensão** de análise de dados. Quanto menos dimensões de análise existirem, menos complexo se torna o problema. É então necessário efectuar uma triagem dos atributos que serão essenciais para a análise que iremos fazer. Por norma, existem atributos que não trazem qualquer informação a ser utilizada pelas técnicas de DM (atributos meramente informativos, com nomes, números de contribuinte, etc.).

### 2.7.2.2 Pré-Processamento

Após a delimitação do espaço de pesquisa, é necessário limpar os dados para assegurar a sua qualidade antes de serem alimentados ao DM. Existem diversas tarefas a cumprir nesta fase, entre elas: identificação de duplicação, preenchimento de campos vazios e tratamento de *outliers* (eliminação/acomodação).

### 2.7.2.3 Transformação

A fase de Transformação consiste no processo de redução do espaço de pesquisa, onde na prática se reflecte na diminuição do número de linhas/colunas a analisar, aplicando reduções dimensionais e transformação de variáveis. Procura também as características que melhor representam os dados tendo em conta o objectivo do projecto.

### 2.7.2.4 Data Mining

O crescimento do volume de dados tem gerado a urgente necessidade de novas técnicas e ferramentas que sejam capazes de lidar com tal volume de informação e que consigam transformar a informação em conhecimento de uma forma automática e inteligente. É neste contexto que surge o *Data Mining* (DM) como fase do processo DCBD.

*Data Mining* caracteriza-se pela identificação de padrões, relacionamentos ou modelos interessantes existentes nos dados que não fazem parte dos dados explicitamente (Witten e Frank, 2005). Na realidade, podemos definir *Data Mining* como (Chiu e Tavella, 2008):

- O uso de técnicas estatísticas ou analíticas para processar e analisar dados em bruto com o objectivo de encontrar padrões interessantes
- A extracção e uso de informação e conhecimento de modo a apoiar a tomada de decisões de uma empresa

Sendo *Data Mining* a fase mais complexa do projecto, efectua-se uma análise das aplicações e técnicas existentes.

#### **2.7.2.4.1 Aplicações de *Data Mining***

O uso de *Data Mining* vai para além das técnicas existentes que se focam na análise de eventos passados onde retiram a informação que necessitam. Este tipo de análise inovadora não tem apenas a vertente de analisar eventos passados e retirar informações sobre esses dados, mas também tem a capacidade de prever tendências futuras. A forma de aplicar está intimamente relacionada com o objectivo, que pode permitir aumentar o conhecimento acerca dos dados (descrição) ou suportar o processo de tomada de decisão (previsão) através de modelos capazes de prever o valor de uma variável (Berry e Linoff, 2000).

No entanto, a necessidade de aplicação de técnicas de DM parte sempre da necessidade do negócio. Recorre-se apenas à sua utilização caso outras técnicas não consigam satisfazer essas necessidades. Algumas das aplicações destas técnicas são:

- Gestão de Mercado (Marketing Direccionado, *Cross-Selling*, Segmentação de Mercado)
- Gestão de Risco (Previsão, Retensão de Clientes, Análise de Concorrência)
- Gestão de Processos (Optimização de Inventário, Controlo de Qualidade, Previsão de Procura de Mercado)

#### **2.7.2.4.2 Tarefas de *Data Mining***

*Data Mining* disponibiliza ao utilizador diversas tarefas, entre elas duas das mais utilizadas: classificação e segmentação. A análise que se segue apresenta as principais tarefas e como se aplicam.

##### **Classificação**

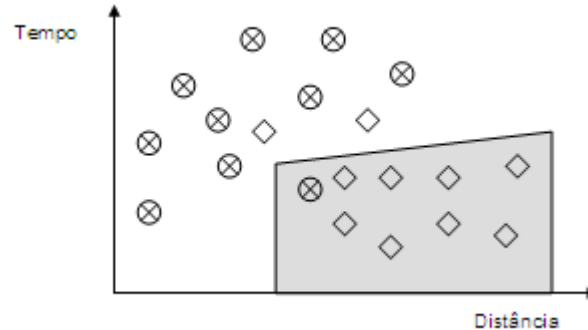
A classificação assenta na alocação de todos os elementos de um conjunto de dados a uma classe predefinida. A classificação de dados é um processo que passa por duas etapas (Han e Kamber, 2001).

A primeira fase consiste na criação de um modelo que será aplicado aos dados que queremos classificar. Este modelo é obtido a partir de um conjunto de dados caracterizados por certas propriedades e também por um rótulo ou *label*, que é o conjunto de valores de saída possíveis que o registo pode ter. A este conjunto dá-se o nome de dados de treino. Através da exploração

do conjunto de valores possíveis que integram os atributos dos registos, são encontrados os valores comuns entre eles de forma a atribuí-los a classes.

A segunda fase caracteriza-se pela aplicação do modelo criado a um conjunto de dados de teste. É a fase em que o modelo é testado no sentido de apurar o seu desempenho no que toca a classificar dados desconhecidos. A precisão do modelo determina-se com base no número de classificações correctas, comparando o valor real de cada registo com o valor previsto pelo modelo. Se o desempenho for considerado aceitável, o modelo pode ser aplicado ao problema em causa. Se no entanto o desempenho não for o esperado, devem ser feitos os ajustes considerados necessários para melhorar o modelo.

O método de classificação é uma tarefa de aprendizagem supervisionada. Isto quer dizer que o atributo e as classes que vão conduzir o processo de classificação dos dados são conhecidos à partida. Na Figura 2.7 estão representados 20 exemplos (pontos na figura) de uma base de dados fictícia, na qual os exemplos são classificados em corredor experiente e corredor amador, onde a distinção é efectuada mediante o valor de duas características (Distância percorrida e Tempo decorrido). No eixo dos *xx* está representada a *Distância* que o indivíduo percorreu e no eixo dos *yy* o *Tempo* que o indivíduo demorou a percorrer.



**Figura 2.7 – Classificação: exemplo**

Os dados foram classificados segundo os dois valores possíveis: o símbolo ‘ $\otimes$ ’ representa os corredores amadores e o símbolo ‘ $\diamond$ ’ representa os corredores experientes. Após uma análise da Figura 2.7, é possível identificar a existência de duas regiões, que permitem dividir os indivíduos nas duas classes disponíveis para o atributo de saída: experiente e amador. Na figura também está representada a qualidade do modelo e a sua limitação. Verifica-se na zona sombreada a presença de um indivíduo amador que foi classificado como experiente e na zona não sombreada o mesmo sucede. Este modelo exemplificado consegue classificar a maioria dos exemplos, no entanto existem alguns que não foram correctamente classificados.

## Segmentação

Segmentação, ou *clustering*, tem como objectivo organizar objectos em grupos, ou *clusters*, onde os objectos partilham semelhanças de alguma maneira. É uma tarefa de aprendizagem não supervisionada, uma vez que o utilizador não tem qualquer influência na definição dos *clusters* que são identificados. Ou seja, as *labels* que identificam cada elemento não são determinadas no início da tarefa, cabendo ao algoritmo identificar os diferentes *labels* dos dados não-rotulados. Um *cluster* é então um conjunto de exemplos que são “semelhantes” entre eles e “não-semelhantes” face aos exemplos pertencentes a outros *clusters*.

Na Figura 2.8 são apresentados diversos exemplos com certas características seguindo duas variáveis (Rendimentos no eixo dos *xx* e Dívida no eixo dos *yy*). Os algoritmos que calculam automaticamente os agrupamentos utilizam a distância entre os diversos exemplos (segundo as várias variáveis/dimensões) para determinar a sua semelhança. No exemplo apresentado, são identificados três *clusters* distintos, onde a identificação destes agrupamentos podem servir para diversos fins (atribuição de risco, alvo de campanhas de marketing, etc.). Tal como sucede com algoritmos de classificação, a identificação de um grupo único para cada elemento é uma tarefa complexa. Poderão então surgir exemplos que se inserem simultaneamente em dois *clusters* diferentes, visto as suas características o tornarem um elemento “fronteiriço”.

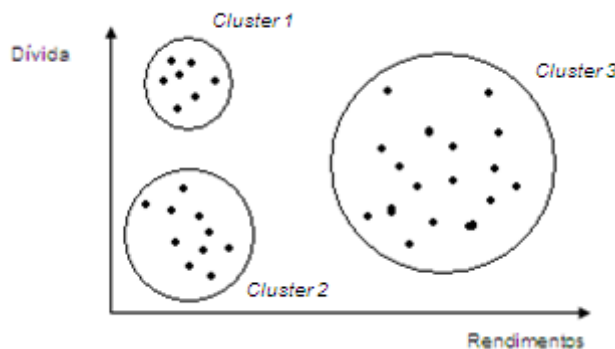


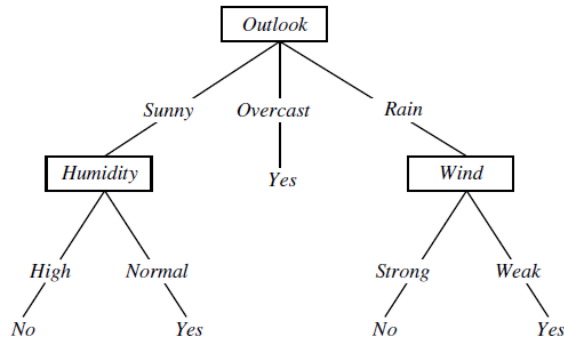
Figura 2.8 – *Clustering*: exemplo

### 2.7.2.4.3 Técnicas de *Data Mining*

#### Árvores de Decisão

Esta técnica, tal como o nome sugere, caracteriza-se por seguir uma estrutura em árvore onde estão representadas todas as decisões (Figura 2.9). É uma técnica com representação simplista mas poderosa para a geração de regras baseadas na informação armazenada na base de dados.

A capacidade de lidar com grandes quantidades de dados aliada à fácil interpretação da informação, torna a técnica uma das mais usadas na área de DM.



**Figura 2.9 – Árvores de Decisão: exemplo**

Uma árvore de decisão é composta por **nós**, **ramos** e **folhas**. Cada nó interno corresponde a um atributo a ser classificado, enquanto os ramos correspondem aos valores que os atributos podem tomar. As folhas da árvore correspondem às diferentes classificações que os atributos podem ter.

A árvore apresentada da figura descreve as situações climáticas que condicionam um indivíduo a ir jogar Ténis. Existem 3 regras positivas e 2 negativas na árvore. Um exemplo das condições necessárias para se praticar Ténis é:

*Se Outlook = “Sunny” e Humidity = “Normal” Então “Yes”*

Cada folha da árvore cria uma regra como a expressa anteriormente. Como foi possível verificar, a utilização de Árvores de Decisão é uma forma simples de criação de regras.

### **Regras de Associação**

A utilização de regras de associação aplica-se a grandes bases de dados de transacções – supermercados, lojas *online* entre outras – para encontrar relações entre os diversos itens transaccionados. Um exemplo clássico é a ordem de exposição de produtos numa cadeia de supermercados. Através da exploração da BD de transacções, é possível identificar padrões de aquisição, isto é, identificar conjuntos de produtos que são comprados em conjunto e desse modo reorganizar a disposição dos itens para maximizar o lucro.

Esta necessidade levou ao desenvolvimento de técnicas que identifiquem automaticamente associações entre itens guardados em grandes BD. Esta técnica é umas das mais comuns de DM

para reconhecimento de padrões dentro de sistemas de aprendizagem não-supervisionados (Han e Kamber, 2006).

Seguindo a definição de Agrawal *et al.* (1993), o problema de criação de regras de associação é definido como: Dada uma base de dados  $T$ , que armazena um conjunto de atributos  $A = \{ A_1, A_2, \dots, A_n \}$ , uma regra de associação existente em  $T$  é representada por  $X \Rightarrow Y$ , onde  $X, Y \subseteq T$  e  $X \cap Y = \emptyset$ . A regra é verificada no conjunto de transacções de  $T$  com um factor de suporte e confiança. O suporte de uma regra indica o subconjunto de registos que satisfazem a união dos atributos que integram a parte antecedente ( $X$ ) e consequente ( $Y$ ) da regra, ou seja, é o número de transacções que contêm um determinado conjunto de atributos. O factor confiança indica a percentagem das transacções de  $T$  que satisfazem  $X$  também satisfazem  $Y$ .

A força de uma regra é então medida através do suporte e da confiança, onde o suporte determina a quantidade de vezes que a regra é aplicável a um determinado *dataset*, enquanto a confiança determina a frequência com que o atributo  $Y$  surge nas transacções que contêm  $Y$ .

#### **2.7.2.5 Interpretação de Resultados**

Esta é a fase onde se interpretam os padrões originados pela aplicação de algoritmos e técnicas de DM. Pode surgir a necessidade de, após interpretar os padrões, retroceder a qualquer uma das fases anteriores de forma a incluir novos dados na análise. É nesta fase que são visualizados os padrões extraídos, remoção de padrões redundantes ou irrelevantes, e a transformação dos interessantes numa forma perceptível para utilizadores.

### **2.8 Discussão**

Ao longo deste capítulo, foi apresentado o tópico de investigação por questionários e inquéritos. Apresentou-se a definição e a estrutura de um questionário, bem como os problemas existentes. Foi abordada a forma como os indivíduos encaram a participação e preenchimento do questionário e os diversos erros e limitações que estas podem criar (dados errados e dados incompletos).

Foi efectuado um levantamento de soluções e metodologias existentes que tentam dar resposta aos diversos problemas, no entanto não respondem e não dão solução a todos os objectivos desta dissertação. Soluções como o *branching logic* e metodologias como a proposta por Urbano (2008) não dão resposta às limitações identificadas no início.

Foi também realizada uma introdução de técnicas da área de *Data Mining* que podem ser aplicadas a esta realidade, podendo no final contribuir de maneira positiva à resolução dos problemas levantados.

Face à ineficácia que as metodologias apresentam, foi criada uma metodologia e conjunto de técnicas que dão respostas aos objectivos propostos.



# Capítulo 3

## Metodologia Proposta

### 3.1 Introdução

Identificados os objectivos da dissertação bem como uma revisão da literatura do tópico em questão, procede-se à elaboração da metodologia que traça as etapas que deverão ser seguidas de forma a garantir o cumprimento dos objectivos.

Atendendo à natureza da dissertação, esta é encarada como um projecto e como tal, deve seguir determinadas fases que se inserem num ciclo iterativo. A metodologia adoptada no projecto baseou-se num dos principais mecanismos para obter produtos de qualidade na área de Engenharia de Software: *Software Development Life Cycle* (SDLC).

O SDLC é um conjunto de actividades encadeadas e ordenadas que garantem a correcta criação de *software* garantindo os mais elevados padrões de qualidade ( Necco *et al.*, 1987). No ciclo adoptado para este projecto, identificam-se 5 actividades principais que são executadas segundo uma sequência predefinida: Levantamento de Requisitos, Metodologia de Avaliação, Desenho, Implementação e Testes. Na Figura 3.1 é apresentada a sequência de actividades representada num modelo cascata onde o *output* de cada actividade alimenta a próxima.

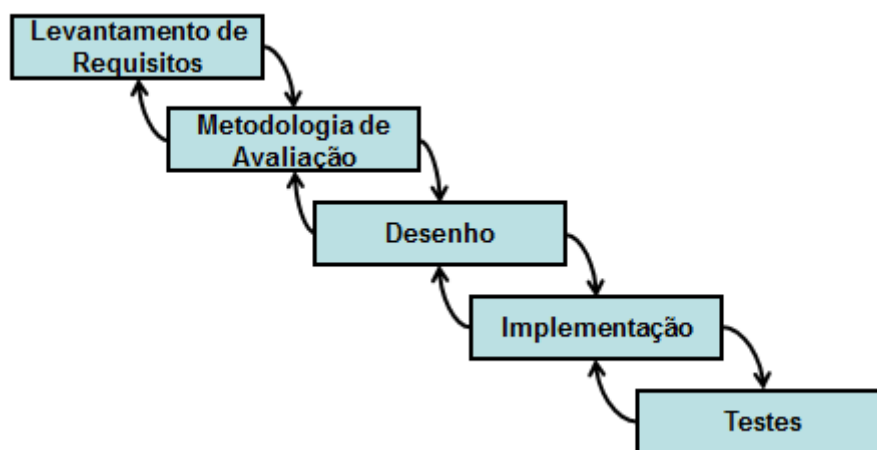


Figura 3.1 – Metodologia Adoptada

Nos subcapítulos seguintes, é explicada cada etapa do processo começando com a documentação dos requisitos do projecto, passando pela criação de uma ferramenta de avaliação de questionários através de uma nova metodologia. Prossegue-se para o desenho da

solução baseada nos requisitos e objectivos do projecto, passando pela implementação das técnicas abordadas. No final, são efectuados testes através de exemplos práticos onde é demonstrado o funcionamento das técnicas e o valor que estas trazem.

## 3.2 Levantamento de Requisitos

A primeira fase do projecto é crucial pois define à partida o que é expectável obter do projecto, bem como os índices de qualidade do *output*.

Definem-se duas classes de requisitos: Directos e Indirectos. **Requisitos Directos** são aqueles que advêm de reuniões e análise directa dos questionários anteriores, e são facilmente identificáveis pelo cliente como por exemplo, taxa de participação do questionário. **Requisitos Indirectos** são aqueles que tendo uma substancial carga subjectiva, são dificilmente transformáveis em critérios específicos e mensuráveis como por exemplo, a fiabilidade das respostas (probabilidade do inquirido estar a mentir ou a enganar-se).

Os requisitos têm dupla aplicação. Por um lado, definem objectivos para o *output* do projecto, por outro, os requisitos são utilizados para avaliar a qualidade e *performance* dos questionários.

### 3.2.1 Requisitos Directos

Como em qualquer projecto, é essencial capturar as necessidades e requisitos que a empresa gostaria de satisfazer. Com o objectivo de traçar o futuro e o âmbito do projecto, realiza-se uma reunião com o líder do projecto e a chefia para obter detalhes sobre a natureza do projecto. Entre os diversos tópicos, deverão ser abordados:

- Origem do Projecto
- Funcionamento do Questionário
  - Destinatários
  - Estrutura
  - Repositório de Perguntas
- Fontes de Informação
- Limitações Existentes
- Expectativa para o Projecto

Ao ser apresentado todo o conhecimento da empresa sobre o assunto e através de troca de ideias e sugestões, são definidos os pilares nos quais o projecto irá assentar. Estabelece-se o que é desejado e o que é alcançável para definir os objectivos e metas a cumprir no prazo acordado.

Definem-se os requisitos directos durante esta fase que deverão ser devidamente documentados (Ex: Aumentar a taxa de resposta do questionário em 10%).

### 3.2.2 Requisitos Indirectos

Os requisitos indirectos exigem geralmente uma análise mais complexa dos dados para serem satisfeitos, assim como para medir o seu grau de satisfação. Iremos usar técnicas de Aprendizagem automática, tais como criação de regras de associação, para identificar, implementar e validar os requisitos.

## 3.3 Avaliação do Questionário

### 3.3.1 Introdução

Após terem sido estabelecidos os requisitos do projecto, é necessário criar uma ferramenta para avaliar o desempenho e qualidade dos questionários. Aplicando esta avaliação aos questionários antes e depois da implementação da metodologia proposta nesta dissertação, é possível verificar se houve melhoria de um questionário para outro.

No entanto, a forma como medimos a melhoria de qualidade de um questionário para outro não é trivial. Existem inúmeros critérios e métricas que podem ser utilizadas mediante cada situação. Existem **critérios** gerais que se podem aplicar a qualquer questionário, seja de satisfação ou não, e critérios usadas para determinada classe de questionários.

---

### Definição

Critérios de Avaliação são indicadores que medem, de forma objectiva, o quão longe a qualidade actual do questionário está face ao que é desejado.

---

Apenas com a aplicação de um bom conjunto de critérios será possível concluir que um questionário é objectivamente melhor que outro. O objectivo desta fase passa pela atribuição de pontos aos questionários para que estes possuam uma determinada pontuação numa determinada escala. A Figura 3.2 ilustra o objectivo da criação desta ferramenta.

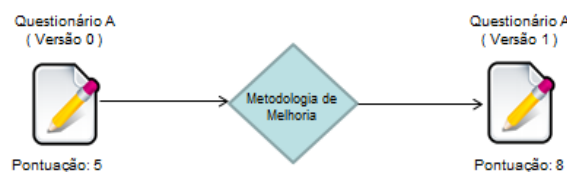


Figura 3.2 – Visão alto nível da aplicação da ferramenta de análise de qualidade

Na figura é ilustrado o *output* da etapa de avaliação, caracterizando-se pela atribuição de uma pontuação (valor numérico) que serve como medida de comparação entre questionários.

### **3.3.2 Metodologia de Avaliação**

Tendo em conta a dificuldade que apresenta a avaliação de um questionário de forma objectiva, a utilização de uma metodologia é essencial para tornar o processo mais perceptível e aplicável a qualquer questionário no futuro.

A metodologia proposta assenta no cumprimento de 4 fases:

1. Criação de critérios objectivos de avaliação
2. Criação de uma escala de pontuação
3. Identificação de expectativas e grau de exigência da empresa
4. Atribuição de pontos ao questionário

A aplicação destas fases irá resultar num valor numérico que será usado para comparar a qualidade dos questionários.

Consideramos um questionário objectivamente melhor que outro quando, após a aplicação da metodologia, a pontuação obtida for mais elevada que outro.

#### **3.3.2.1 Criação de critérios objectivos de avaliação**

Nesta primeira fase, o conjunto de requisitos identificados após a execução do subcapítulo 4.2 é utilizado como critérios de avaliação. Analisando cada requisito, efectua-se a adaptação para um critério que seja mensurável.

Os critérios variam mediante os questionários e realidades empresariais, onde a importância e aplicação dos critérios vai de encontro à utilidade dos mesmos na avaliação dos questionários. Ou seja, mesmo que alguns critérios possam ser aplicáveis em qualquer realidade, existem alguns que apenas fazem sentido mediante a natureza do questionário e o que se pretende avaliar. Para compreendermos melhor o tipo de critério que são utilizados para avaliar apresenta-se o seguinte exemplo:

#### **Exemplo**

“Número de questionários preenchidos” – a melhoria do novo questionário é medida através do número de respostas que se obteve, se for maior que o ano anterior obtém-se um indicador de melhoria segundo este critério e uma pontuação associada.

### 3.3.2.2 Criação de uma Escala de Pontuação

Após a documentação dos diversos critérios, a fase seguinte passa pela criação de uma escala pontual que consiga avaliar os questionários segundo os critérios. A escala deverá ser uniforme e constituída por valores numéricos.

Qualquer escala poderá ser utilizada desde que seja utilizada a mesma para avaliar cada um dos questionários. Caso seja utilizada uma escala diferente para cada questionário, estes tornam-se incomparáveis pelo facto de cada critério ser pontuado com escalas diferentes. A escala pontual adoptada deverá ser discutida e acordada com a chefia onde são discutidas as vantagens e as desvantagens da utilização de cada uma.

Independentemente da escala numérica adoptada, é necessário atribuir significado a cada um dos valores da escala. Efectua-se o mapeamento de uma escala numérica para uma escala subjectiva, isto é, atribuir valores numéricos a índices de satisfação. A escala subjectiva surge para mais facilmente se entender a escala numérica, que deverá ser também definida em conjunto com a chefia. Uma escala subjectiva possível poderá ser: Ideal, Excelente, Aceitável, Razoável e Mau. Qualquer escala subjectiva poderá ser criada desde que seja perceptível para qualquer indivíduo.

Com intuito de simplificar o processo, poderá estabelecer-se a escala subjectiva da mesma forma como é efectuada a escala utilizada no próprio questionário (Ex: Muito Satisfeito, Satisfeito, Neutro, Insatisfeito e Muito Insatisfeito). A Tabela 3.1 apresenta um exemplo de uma escala onde se adoptaram valores numéricos de 0 a 1 (codificação de percentagens) e a classificação subjectiva Ideal, Excelente, Aceitável, Razoável e Mau.

Tabela 3.1 – Exemplo de atribuição de pontos

<b>CLASSIFICAÇÃO</b>	<b>PONTOS</b>
<b>Ideal</b>	<b>1</b>
<b>Excelente</b>	<b>0.8</b>
<b>Aceitável</b>	<b>0.6</b>
<b>Razoável</b>	<b>0.3</b>
<b>Mau</b>	<b>&lt; 0.3</b>

Surge agora a dúvida sobre as situações em que os requisitos são classificados como Excelente ou Aceitável, ou qualquer outra classificação. Esta classificação é abordada na próxima fase da metodologia onde serão apresentados os passos necessários para o fazer.

### 3.3.2.3 Identificação de Expectativas e Grau de Exigência da Empresa

O grau de satisfação de uma pessoa, seja sobre um serviço, um ambiente ou uma infraestrutura, é único e muito subjectivo tornando a sua medição uma tarefa complexa. Tomemos como exemplo uma empresa A, onde se registou uma média de 5% de pessoas a chegar atrasadas ao posto de trabalho e foi considerado “Excelente”. Para uma empresa B, os mesmos 5% foram considerados “Aceitável”. O exemplo ilustra de forma clara a subjectividade do índice de satisfação, ou seja, o que para um indivíduo é excelente, poderá não ser para outro.

É com este problema em mente que esta terceira fase se torna importante no sentido de identificar as expectativas e a exigência da empresa.

Tomando como referência a escala subjectiva apresentada, traça-se como objectivo perceber aos olhos da empresa, o que é considerado “Ideal”, “Excelente”, “Aceitável”, “Razoável” e “Mau”. Para capturar o grau de exigência da entidade, é necessário perguntar a indivíduos que nos possam facultar essa informação da forma mais precisa possível. Os indivíduos que podem facultar uma opinião formal são os *managers*, que supervisionam a informação e actividade da empresa, logo estão ao corrente dos objectivos e graus de exigência. Para capturar este conhecimento, deve ser realizada uma reunião com os mesmos.

Na reunião deverá ser explicada a importância desta para assegurar a qualidade do modelo de avaliação que irá ser criado e implementado, como também dos critérios e a forma como serão aplicados aos diversos questionários.

É feito um mapeamento da escala subjectiva de cada um para uma escala numérica podendo ser percentagem (0 a 100%) ou outra que faça sentido ao critério em causa. São apresentados os critérios e de seguida, será solicitado aos *managers* que estabeleçam o seu nível de exigência para cada um deles utilizando a escala proposta.

#### **Exemplo**

Critério: Número de questionários recebidos

Escala: 100% => Número de respostas recebidas igual ao número de questionários enviados

0% => Nenhuma resposta recebida

Tabela 3.2 – Avaliação *Manager A*

CLASSIFICAÇÃO	RESULTADO
Ideal	100%
Excelente	95%
Aceitável	90%
Razoável	85%
Mau	< 85%

Tabela 3.3 – Avaliação *Manager B*

CLASSIFICAÇÃO	RESULTADO
Ideal	100%
Excelente	97%
Aceitável	93%
Razoável	89%
Mau	< 89%

Tabela 3.4 – Avaliação *Manager C*

CLASSIFICAÇÃO	RESULTADO
Ideal	100%
Excelente	90%
Aceitável	85%
Razoável	80%
Mau	< 80%

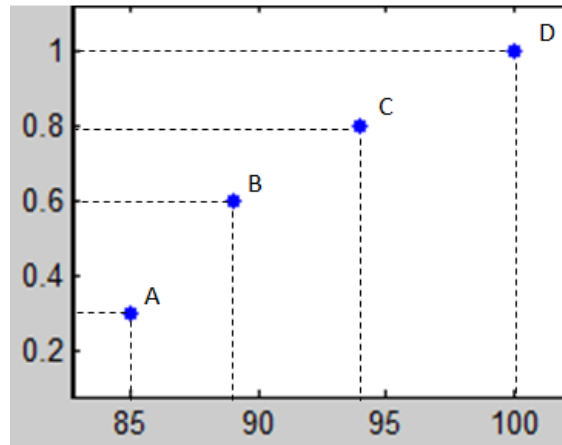
Tabela 3.5 – Avaliação Final *Managers*

CLASSIFICAÇÃO	RESULTADO
Ideal	100%
Excelente	94%
Aceitável	89%
Razoável	85%
Mau	< 85%

A Tabela 3.5 passa a ser uma escala com intervalos definidos que classificam os resultados obtidos no questionário segundo cada critério, onde mediante o resultado, é classificado segundo a escala subjectiva. Através da tabela sabe-se que caso sejam recebidos 89% de questionários preenchidos é considerado Aceitável aos olhos da Empresa. É agora necessário atribuir pontuação a cada critério utilizando a escala de pontuação criada.

### 3.3.1.4 Atribuição de Pontos ao Questionário

Para serem atribuídos pontos ao questionário segundo os diversos critérios, a primeira escala apresentada na Tabela 3.1 não é suficiente pela falta de flexibilidade que apresenta. A escala apenas possibilita a atribuição de 5 pontuações distintas numa escala de 0 a 1.



**Figura 3.3 – Escala pontuação utilizando 4 pontos**

Na Figura 3.3 são apresentadas as 4 pontuações disponíveis até ao momento onde o eixo dos  $xx$  define os valores registados no questionário e o eixo dos  $yy$  os pontos associados a cada uma delas. Por exemplo, o ponto A define a classificação razoável (85%) onde são atribuídos 0.3 pontos.

No entanto, os valores registados em cada critério pode ser qualquer número entre 0 e 100, ou seja, cada nível de satisfação deverá ter um intervalo de pontos atribuído a ele e não apenas um valor. Por exemplo, se for verificado um nível de 92% num dado critério, qual a classificação a atribuir? Nesta fase, apenas é possível afirmar que será uma pontuação entre 0.6 e 0.8.

Para obtermos a pontuação exacta de cada pergunta é necessário utilizar os dados disponíveis sobre a mesma, iremos interpolar os dados através de uma função Gaussiana, que nos apresenta a relação dos valores com as classificações.

### Função Gaussiana

Comecemos por atribuir valores numéricos, por exemplo 0 a 1, à classificação subjectiva tal como foi feito anteriormente na Tabela 3.1. A aplicação de uma função gaussiana possibilita a conversão de um valor obtido directamente num critério na escala numérica de pontuação que desejamos (no exemplo, 0 a 1). Designamos esta função por função de Transferência.

**Tabela 3.6 – Exemplo Classificação de Critério**

CLASSIFICAÇÃO	RESULTADO	PONTOS
Ideal	100%	1
Excelente	94%	0.8
Aceitável	89%	0.6
Razoável	85%	0.3
Mau	< 85%	< 0.3

A Tabela 3.6 agrupa toda a informação acerca da relação entre o Resultado e os Pontos no entanto ainda não existem pontos suficientes para se extrair a pontuação de qualquer resultado na escala 0 a 100. Para se obter todas as pontuações, obtém-se a função de Transferência por determinação da função Gaussiana que minimiza o erro quadrático face aos 4 pontos referidos (despreza-se o último por ser um intervalo de 0 ao último valor). A este processo dá-se o nome de *fitting* que na Figura 3.4 pode ser visualizado.

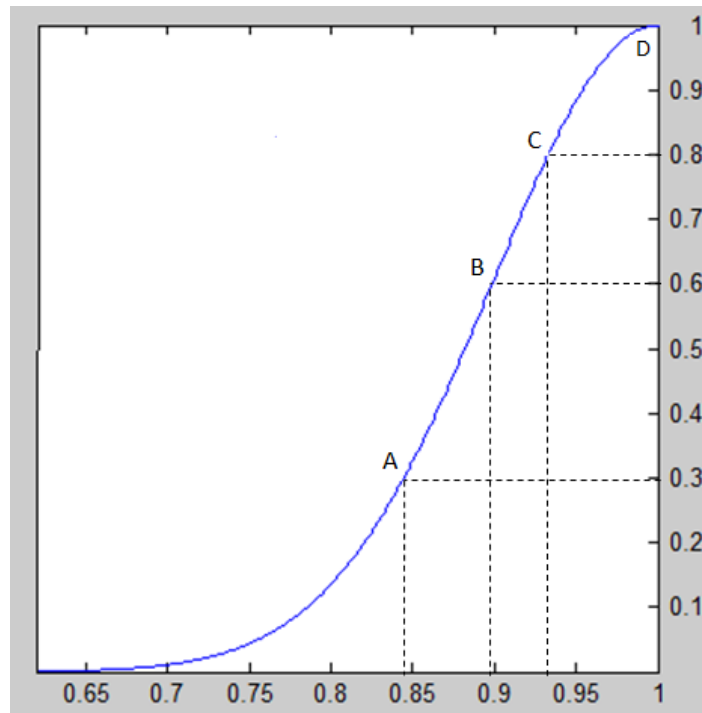


Figura 3.4 - Exemplo *fitting* da Gaussiana

No eixo das abcissas ( $xx$ ), está representado o resultado obtido do questionário num dado critério e no eixo das ordenadas ( $yy$ ), está representada a escala de 0 a 1 da pontuação que é atribuída. A função de Transferência possibilita a obtenção da pontuação quando indicamos o resultado do critério aplicado a um questionário. Através da sua utilização, facilmente se retira a sua pontuação do critério em questão. Este é apenas um exemplo e a curva e os valores não representam a realidade e serve apenas para nos inteirarmos da utilidade e aplicação da função de Transferência.

Para a definir, é necessário um conjunto de 3 variáveis:

- Valor ideal do critério
- Valor obtido do critério no questionário em questão
- Variância

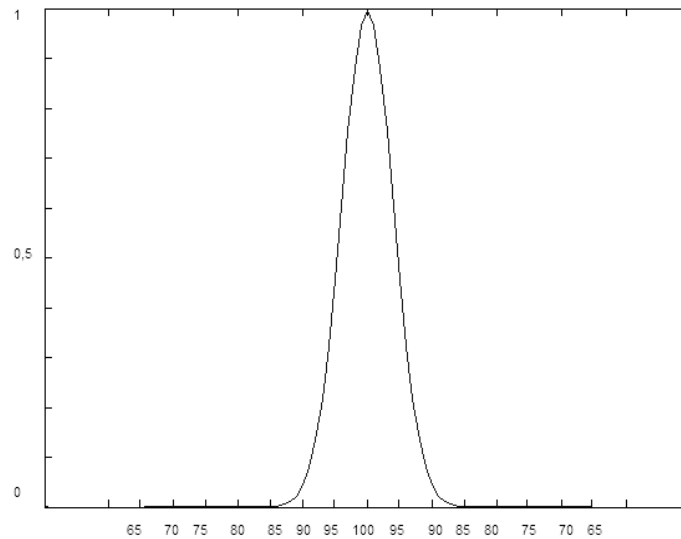
A função de Transferência é apresentada na Eq. 3.1, onde ‘x’ é o valor obtido, ‘μ’ é o valor ideal e ‘σ<sup>2</sup>’ é a variância.

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}, -\infty < x < \infty, \sigma > 0.$$

**Eq. 3.1 - Função de Transferência**

Para construir a função de Transferência de cada critério, são necessários os valores acima indicados. Para cada critério, sabemos à partida o valor ideal (máximo da escala escolhida), o valor de cada nível de satisfação e a pontuação para cada nível de satisfação, ficando apenas por calcular a variância.

A variância é o factor que influencia a “abertura” da curva, ou seja, trabalha directamente com a correspondência entre os resultados da aplicação dos critérios e as pontuações atribuídas a esses critérios. Utilizamos os níveis de exigência dos critérios calculados anteriormente, para conseguirmos extrapolar a variância. A “abertura” da curva reflecte os níveis de exigência da empresa para cada critério no sentido em que, quanto mais esguia for a curva, mais exigente é a empresa para esse critério em específico. Podemos visualizar na Figura 3.5 a forma como é afectada a forma da curva mediante a exigência.



**Figura 3.5 – Curva de critérios de elevada exigência**

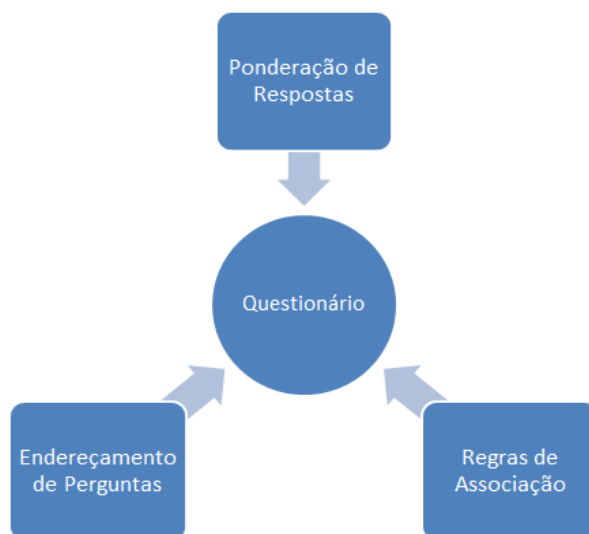
Para calcular a variância do critério – variável que influencia a abertura – é necessário testar valores σ<sup>2</sup> e escolher o que define a abertura que fique mais perto dos 4 pontos anteriores. Ou seja, não existe um valor de σ<sup>2</sup> que defina uma curva que passe exactamente nos 4 pontos mas existe um valor que minimiza esse desvio.

Após a construção da função densidade para cada critério, prossegue-se ao cálculo da pontuação. Nesta fase, a escala está devidamente criada para o critério em questão e poderá ser utilizada em qualquer questionário. Com o resultado do critério aplicado ao questionário (eixo do *xx*) obtém-se a pontuação associada a esse resultado (eixo do *yy*) tendo sido criada uma ferramenta que averigua de forma objectiva a qualidade de questionários.

### 3.4 Desenho

#### 3.4.1 Introdução

No Desenho do sistema devemos criar uma arquitectura que dê resposta aos requisitos identificados na fase inicial. Neste trabalho é proposta uma arquitectura baseada em 3 componentes (Figura 3.6): (1) um componente destinado a melhorar o endereçamento de perguntas respondendo aos requisitos directos; (2) um componente relacionado com a ponderação de respostas, também destinado a responder e valorizar os requisitos directos que foram estabelecidos; (3) um terceiro componente destinado a melhorar o requisito indirecto da fiabilidade do questionário.



**Figura 3.6 – Arquitectura do Sistema**

Nos subcapítulos seguintes, cada técnica é aprofundada de forma a demonstrar os seus benefícios e forma como funcionam. Cada uma destas três técnicas tem a particularidade de poder ser implementada em qualquer empresa ou entidade e por essa razão, cada subcapítulo aborda o tema de forma generalista.

## **3.4.2 Endereçamento de Perguntas**

### **3.4.2.1 Introdução**

No subcapítulo 2.5, foram apresentadas algumas formas de tornar o questionário mais *user-directed* procurando apresentar as perguntas que, mediante as respostas a perguntas anteriores, fossem seleccionadas para tornar o questionário único para o utilizador. Técnicas como lógica condicional e aplicação de regras são exemplos de abordagens para atingir esse objectivo. Estas abordagens focam-se na estrutura e encadeamento das perguntas mediante as respostas do utilizador. Estas soluções influenciam pela positiva a experiência do utilizador mas também existem outras formas complementares que optimizam todo o processo.

### **3.4.2.2 Motivação**

Uma das principais preocupações quando se pretende criar um questionário é a qualidade das perguntas. Diversas formas de garantir qualidade foram introduzidas no subcapítulo 2.2 para que a experiência de preenchimento do questionário seja positiva para o utilizador. No entanto, o tamanho de um questionário está inevitavelmente dependente do volume de perguntas à qual o investigador pretende obter resposta. Mesmo que todas as perguntas tenham qualidade, a dimensão do questionário aliada à colocação de perguntas que nada dizem respeito ao utilizador, desmotivam-no a finalizar o seu preenchimento.

A própria forma como as perguntas são colocadas e as opções de respostas disponibilizadas deverão ser analisadas. O endereçamento de perguntas deverá ser feito de forma a não necessitar de perguntar ao utilizador se este usa ou não um serviço ou se está em condições de responder à pergunta. O questionário deverá ser capaz de identificar os sistemas/aplicações e serviços que o utilizador está em condições de responder.

### **3.4.2.3 Abordagem**

Após a informação estar agrupada e organizada para ser eficientemente usada, tira-se partido dela para endereçar as perguntas aos utilizadores. Com a elevada quantidade de informação disponível, é necessário decidir que informação deverá ser usada e com que fim.

O endereçamento de perguntas específicas a cada utilizar é conseguido através da exploração da informação, nomeadamente através da intersecção de diversos tipos de informação. Um dos principais focos do questionário é o levantamento dos índices de satisfação dos utilizadores quanto às aplicações e sistemas disponíveis. Endereçar ao utilizador apenas as perguntas de aplicações e sistemas que de facto utiliza revela-se um desafio. O correcto endereçamento não só encurta, com grande probabilidade, a dimensão média do questionário, como também

garante uma experiência mais positiva ao utilizador enquanto este preenche apenas perguntas que lhe dizem respeito.

Para endereçar as perguntas, existe sempre um conjunto de informações que permite alcançar esse objectivo. A seguinte questão chave deverá ser respondida:

- Que sistemas e aplicações o utilizador tem **acesso** e **usa** durante a sua actividade?

A lista de sistemas e aplicações que resulta desta questão, define as perguntas que serão endereçadas ao utilizador, isto é, apenas serão efectuadas perguntas ao utilizador sobre os sistemas e aplicações que constam da lista.

Para obter a resposta a esta pergunta, aplicam-se filtros a toda a informação disponível. A abordagem é simples mas poderosa, no sentido em que resolve o problema através da utilização de filtros. Na questão anterior, identificam-se duas informações chave: Acesso e Utilização.

#### **Acesso vs Utilização**

É importante a distinção entre acesso e utilização de um sistema. A informação que é obtida dos sistemas que gerem as permissões e níveis de acesso que cada utilizador tem apenas indica se esse utilizador tem acesso a uma aplicação e com que privilégios. No entanto, esta informação por si não é suficiente para tomar a decisão de colocar uma pergunta sobre um sistema a um utilizador. Na maioria dos casos, existe uma *pool* predefinida de perguntas e todas são colocadas aos utilizadores.

Tomemos o exemplo de perguntas relacionadas com um sistema A onde, através do sistema que gere as permissões, sabemos que o utilizador com ID 10001 tem acesso à aplicação. Não existe informação suficiente para assumir com elevado grau de confiança que o utilizador usa o sistema e que está em condições de responder a questões sobre o uso do sistema. O correcto seria considerar o utilizador 10001 como candidato a questões sobre o sistema A pelo facto de ter acesso e permissões à aplicação.

#### **3.4.2.4 Filtros**

Foi apresentada a abordagem baseada na aplicação de filtros à informação disponível. Sugerem-se 3 etapas de modo a identificar os sistemas/aplicações:

1. Selecção de sistemas/aplicações utilizando tabela de níveis de acesso
2. Selecção de sistemas/aplicações utilizando históricos de acesso (*login*)
3. Selecção de sistemas/aplicações utilizando registos de pedidos de suporte

### **Filtro 1: Acessos Aplicações/Sistemas**

Num ambiente empresarial, como em qualquer ramo de actividade profissional, existe uma clara divisão a nível funcional que não só delimita as funções que o utilizador deverá desempenhar mas também delimita as aplicações/serviços que, com grande probabilidade, irá utilizar para o desempenho das funções.

As funções variam consoante o ambiente em que o questionário se insere, no entanto existe sempre um conjunto de funções, normalmente essenciais à actividade: Vendas, Recursos Humanos, Serviços de Informação, Financeira, etc. Para cada umas das enunciadas, existem ferramentas informáticas que apoiam o utilizador no desempenhar o cargo. Na área Financeira, o utilizador irá certamente trabalhar com uma aplicação que registe e organize todas as aquisições de bens e serviços. Na área de Recursos Humanos, o utilizador necessitará de uma ferramenta central que permita a gestão todos os recursos humanos e que esta mantenha toda a informação relevante para a entidade empregadora.

O primeiro filtro isola os sistemas/aplicações mencionadas anteriormente através de consultas ao sistema que gere os acessos de todos os utilizadores, guardando as aplicações que o utilizador tem acesso numa lista que será utilizada para escolher as perguntas.

### **Filtro 2: Histórico de Acessos a Sistemas**

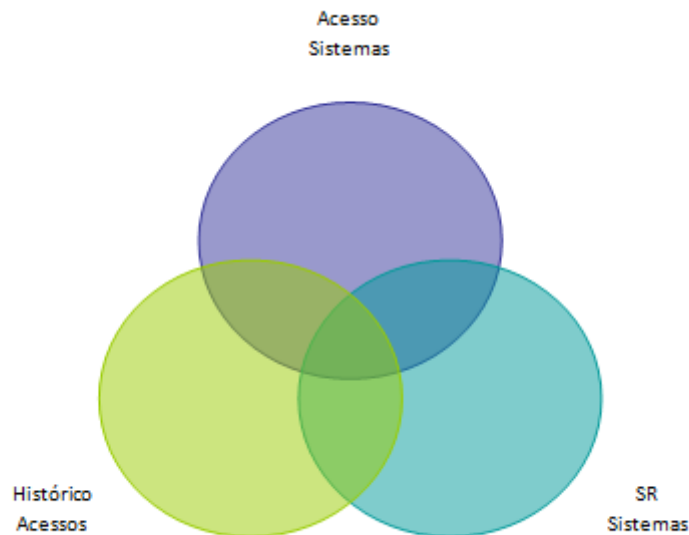
Como segundo filtro, são utilizados os Históricos de Segurança (*logs*) existentes nas aplicações. Recorre-se aos *logs* dos serviços e aplicações de modo a confirmar as que até ao momento foram identificadas e também identificar as que não têm registos no sistema de gestão de acessos. Percorrem-se históricos de acessos de todos os sistemas e aplicações sendo guardadas todas as aplicações/sistemas que o utilizador tenha acedido no ano em questão.

### **Filtro 3: Pedidos de Suporte**

Não sendo a informação extraída pelos dois primeiros filtros suficiente para eficientemente isolar as aplicações que utiliza, recorre-se aos registos de Pedidos de Suporte que o utilizador efectuou até ao momento. Os Pedidos de Suporte têm a vantagem de indicar, com certeza absoluta, os sistemas que o utilizador usou até ao momento. Ou seja, é identificado o conjunto de serviços ao qual o utilizador está em condições de expor a sua opinião. De realçar o facto de não havendo Pedidos de Suporte, este não implica que o utilizador não o tenha utilizado visto não haver obrigatoriedade de solicitar um Pedido de Suporte durante a normal execução das suas funções.

Para seleccionar os sistemas/aplicações utilizam-se três filtros, onde cada filtro pretende extrair a informação das 3 fases anteriormente identificadas (Figura 3.7). A necessidade de três fases advém das limitações que cada uma apresenta:

- A primeira fase identifica os sistemas/aplicações que o utilizador tem acesso mas não existe certeza quanto à utilização dos mesmos
- A segunda fase identifica os sistemas/aplicações que o utilizador efectuou *login* recorrendo ao histórico de cada sistemas mas está condicionado pela própria existência dos históricos dos sistemas
- A terceira fase identifica os sistemas/aplicações que geraram pedidos de suporte que o utilizador resolveu mas nem todos os sistemas necessitam de pedidos de suporte



**Figura 3.7 – Filtros de Informação**

Ao interceptar os resultados provenientes de cada filtro, serão identificadas quase na sua totalidade os sistemas/aplicações que o utilizador usa para que possa responder legitimamente às perguntas.

#### **3.4.2.5 Apresentação de Perguntas Extra**

Durante o processo de preenchimento, é solicitado ao utilizador que exponha a sua satisfação sobre os serviços e aplicações em questão. Sendo o questionário uma ferramenta de apoio à tomada de decisão de uma empresa, é fundamental obter o máximo de informação sobre o que aos olhos dos utilizadores não está conforme o expectável.

A colocação de perguntas extra que apenas são apresentadas mediante certas condições poderá ajudar a empresa a entender melhor o que foi avaliado como insatisfatório. No sentido de captar mais informação, são efectuadas perguntas extra quando a avaliação é negativa (Insatisfeito e Muito Insatisfeito). Nesses casos, deverão ser colocadas perguntas específicas sobre o assunto da pergunta em questão.

#### Exemplo

Se um utilizador responde que está insatisfeito com as Funcionalidades de um sistema, então é colocada uma pergunta de resposta livre para que o utilizador possa especificar a natureza da sua insatisfação ou é colocada uma outra pergunta de escolha múltipla ou utilizando escalas particularizando algumas das funcionalidades chave do sistema, onde o utilizador deve novamente assinalar o seu grau de satisfação.

A implementação desta funcionalidade baseia-se na utilização de condições lógicas a cada pergunta de avaliação (*branching logic*).

### **3.4.3 Ponderação de Respostas**

Um questionário que tenha uma taxa de adesão elevada e outros indicadores é importante para a empresa efectuar o correcto levantamento dos índices de satisfação dos seus utilizadores, no entanto a fase posterior (análise de respostas), que tem como objectivo a extracção de novo conhecimento sobre o assunto em questão, deverá também ser encarada com elevada atenção.

Ao longo do subcapítulo, será apresentada a importância da aplicação de diferentes ponderações a respostas de utilizadores baseada no seu tipo e relevância para aumentar a qualidade dos resultados extraídos no final.

#### **3.4.3.1 Introdução**

No processo de análise de um questionário de satisfação, todas as respostas são analisadas e retiram-se conclusões acerca dos índices de satisfação utilizando representações gráficas e médias aritméticas simples das respostas.

Partindo do princípio que todas as respostas sobre um determinado serviço foram preenchidas por utilizadores que de facto utilizam o serviço em questão para desenvolver o seu trabalho, retiram-se os índices de satisfação desse serviço recorrendo à soma de respostas de cada índice, onde no final se apresentam os resultados em tabelas ou gráficos. Este é de modo geral o método convencional no que toca à extracção e apresentação de resultados de um inquérito de satisfação. No entanto, propõe-se uma forma que com elevada probabilidade irá melhorar a qualidade dos resultados obtidos. Para o concretizar é necessário separar os utilizadores

mediante a sua **importância**, isto é, identificar os utilizadores que fornecem respostas que trazem mais valor aos resultados finais.

### 3.4.3.2 Tipos de Utilizadores

Após uma breve introdução no subcapítulo anterior, identificou-se a necessidade de separar utilizadores em diferentes tipos ou classes. A utilização destas diferentes classes tem como grande vantagem a capacidade de isolar as respostas ao questionário dos utilizadores que teoricamente têm mais valor.

As respostas sobre uma aplicação que tipicamente têm mais valor provêm de utilizadores mais experientes designados os *core users* da aplicação. Utilizadores que não têm tanta relevância para as perguntas são designados *occasional users* que utilizam a aplicação/sistema em regime pontual. Existem diversas razões para se recorrer a esta separação de utilizadores podendo-se afirmar que um *core user* de uma aplicação tem mais capacidades de avaliar pelo facto de:

- O utilizador estar mais familiarizado com os problemas e limitações a nível de *performance* e disponibilidade da aplicação
- O utilizador conhecer as diversas funcionalidades da aplicação
- O utilizador ter elevada experiência e horas de uso das aplicações

Apresentadas algumas das vantagens de classificar utilizadores, é necessário perceber como cada um é classificado numa das classes existentes. O utilizador deverá ter um determinado conjunto de características estabelecidas previamente.

### 3.4.3.3 Características de Utilizadores

As características utilizadas para diferenciar os utilizadores nas suas respectivas classes variam de caso para caso dependendo da forma como se processa toda a actividade da entidade. No entanto, as características identificadas devem diferenciar os utilizadores tendo em conta os pontos enunciados no subcapítulo anterior.

A criação de características que eficazmente definam a relevância e importância de um utilizador para uma certa aplicação é um processo que deve integrar o responsável do questionário e os *service owners* dos diversos sistemas e aplicações. Promovendo a discussão com os principais interessados numa reunião, são documentadas as características que irão definir a importância do utilizador para cada sistema ou aplicações.

A discussão deverá ter como início a transferência de informação do responsável do questionário para as chefias para que estas possam correctamente identificar as características. Inicia-se a discussão com uma breve apresentação que deverá abordar os seguintes pontos:

- Importância do uso de características como mecanismo de diferenciação de respostas relevantes e não relevantes e os impactos positivos sobre o resultado final
- Definição de *Core User* e *Occasional User* apresentando exemplos de características que os poderão definir
- Importância da atribuição de pesos às características identificadas e explicação do cálculo de índices de satisfação através de médias ponderadas

Após a apresentação do problema em questão, são discutidas as características que definem um *core user* na realidade em questão. Uma possível característica poderá ser a taxa de utilização de cada aplicação por parte do utilizador, isto é, o tempo que o utilizador despende a trabalhar com a aplicação. Desta fase, surge como *output* uma tabela com um certo número de características (Tabela 3.7).

**Tabela 3.7 – Características de Utilizadores**

Característica
A
B
...
Z

Não existindo um número limite de características de modo a definir as fronteiras entre os tipos de utilizadores, não é aconselhada a escolha de um elevado número podendo introduzir dificuldades desnecessárias tornando o processo moroso e mais complexo. Deverão ser definidas características que possam identificar claramente os utilizadores mais experientes.

Todas as características adoptadas deverão ser mensuráveis recorrendo a informação disponibilizada na BD da empresa, ou seja, a característica deverá ser quantitativa.

#### **3.4.3.4 Características Ponderadas**

Como foi referido, a análise de respostas é efectuada utilizando uma média aritmética simples. No entanto, este é um método que não tem em conta aspectos importantes como a experiência e a própria capacidade de um utilizador avaliar uma aplicação ou sistema.

A aplicação de pesos a respostas dadas pelos utilizadores introduz o conceito de importância a cada resposta registada. Relembrando os tipos de utilizador existentes, são consideradas mais

importantes as respostas de um *core user* do que um *occasional user* pelas razões apontadas no subcapítulo 3.4.3.2.

A utilização de pesos altera a forma como as respostas são analisadas passando a ser utilizada uma média aritmética ponderada no momento de extrapolação de conclusões sobre cada aplicação individual.

Tendo em conta as características que influenciam o estatuto de um utilizador, o processo de atribuição de pesos às respostas sobre um sistema/aplicação baseia-se na soma de percentagem obtidas em cada uma das características identificadas no subcapítulo anterior. Sabendo que a utilização conjunta das características providencia toda a informação necessária para classificar um utilizador, a importância de cada característica para cada entidade é diferente. Para que esta importância se reflecta na valorização do utilizador, estabelecem-se diferentes pesos para cada uma das características. Estes pesos devem novamente ser discutidos entre o responsável do questionário e as chefias de modo a reflectir as características que são mais valorizadas para a entidade.

Independentemente do número de características, o somatório dos pesos de cada característica deverá garantidamente ser 1 ou 100, caso se utilize percentagens. Considera-se como classificação total (soma dos pesos das características) de cada utilizador um valor real entre 0 e 100 (percentagem), onde 0 implica que as respostas do utilizador sejam menosprezadas e 100 implica que as respostas do utilizador consideram-se essenciais para a análise em questão. Na Tabela 3.8 a distribuição de pesos por 4 características (A, B, C e D) ilustra um possível *output* desta fase, lembrando que o número de características poderá ser maior ou menor.

**Tabela 3.8 – Exemplo Genérico Atribuição de Pesos a Características**

Característica	Peso
A	30%
B	15%
C	40%
D	15%

Estabelecido o peso que cada característica representa, é necessário definir a forma como os mesmos são atribuídos. Cada utilizador é único e as suas diferenças são definidas pelos valores que toma em cada característica. O cálculo do peso final de cada utilizador é constituído pela soma da pontuação obtida em cada característica ponderada pelo respectivo peso.

Por sua vez, cada característica tem um valor que é calculado mediante o levantamento dos comportamentos e situação actual de cada utilizador.

### 3.4.3.5 Critérios de Avaliação de Características

O processo de ponderação de características ganha mais complexidade com a introdução de classificação ao peso de cada característica.

Para calcular o valor de cada característica, recorre-se a um levantamento dos comportamentos e a situação actual de cada utilizador como já foi referido. Para cada característica, atribui-se uma percentagem que irá influenciar os pesos estabelecidos na Tabela 3.8 apresentada na secção anterior. Cada uma das 4 características é avaliada mediante diferentes critérios.

**Tabela 3.9 – Exemplo Genérico Atribuição de Pontos a Característica**

Critério	Pontuação
1	100%
2	50%

A Tabela 3.9 apresenta um exemplo genérico onde uma característica é avaliada segundo um dado critério. No exemplo, o critério identifica dois valores possíveis onde a pontuação da característica segundo esse critério atribui a pontuação máxima (100%) caso o utilizador cumpra o critério com valor igual a '1' e atribui 50% caso o utilizador cumpra o critério com valor igual a '2'.

### 3.4.3.6 Cálculo de Peso Final

Após definição das características e critérios das mesmas, o método de atribuição de pesos ao utilizador de uma aplicação está concluído. Através das informações dos utilizadores no momento do preenchimento do questionário, é calculado o peso do utilizador para cada sistema ou aplicação que responder.

Note-se que esta técnica trabalha directamente para a satisfação de um dos objectivos/requisitos directos identificados: valorizar respostas de utilizadores experientes (*core users*).

De modo a facilitar a compreensão do método, a sua utilização é apresentada através de um exemplo prático no capítulo Resultados.

### **3.4.4 Regras de Associação**

#### **3.4.4.1 Introdução**

Com a elevada dimensão de dados proveniente dos questionários, não é possível através de análise directa extrair todo o conhecimento dos dados. Relações entre perguntas e respostas é uma tarefa que, caso fosse levado a cabo por humanos, iria revelar-se morosa, complexa e sujeita a erro.

No subcapítulo 2.7.2.4.3, apresentou-se a componente teórica da criação de regras de associação entre itens e as vantagens da sua aplicação. No âmbito do projecto, a aplicação desta técnica surge como uma das medidas de avaliação do questionário, podendo também ser utilizada para modelar a estrutura do questionário.

A técnica é aplicada às respostas dos questionários onde são inferidas regras de associação, cuja sua aplicação visa encontrar dependências entre perguntas e respostas através da aplicação do algoritmo *Apriori*. As regras extraídas também definem conjuntos de respostas que não deveriam surgir, isto é, dada uma resposta não seria expectável obter uma certa resposta numa segunda pergunta. Dependendo da dimensão do questionário e dos índices de confiança e suporte das regras geradas, o número de regras extrapoladas pode variar bastante sendo na maioria das vezes na ordem das centenas ou milhares.

As Regras Associativas também desempenham um papel fundamental no que diz respeito à qualidade dos resultados dos questionários efectuados em anos anteriores. Visam de certo modo reabilitar as respostas dos questionários de forma a torná-los mais comparáveis com o novo formato do questionário. É interessante efectuar novo tratamento aos questionários anteriores no sentido de encontrar incoerências nos dados onde dessa forma se obteriam resultados mais fiáveis e mais comparáveis com os mais recentes.

#### **3.4.4.2 Relação entre Perguntas e Respostas**

A utilização desta técnica pretende averiguar as relações existentes entre perguntas e respostas, sendo que o método é aplicado directamente às respostas dos questionários. A técnica pode ser utilizada de diferentes formas, podendo ser aplicada no questionário enquanto o utilizador está a responder ou como tratamento de respostas pós-preenchimento. A aplicação no primeiro cenário passava pelo envio de mensagens de alerta chamando o utilizador à atenção numa certa pergunta, sendo encorajado a rever a sua resposta por não ter sido preenchida de uma forma normal. A aplicação do segundo caso, que funciona como uma medida de qualidade do questionário, passa pela aplicação das regras às respostas do questionário registando o número de regras violadas que serão posteriormente medidas numa escala.

## Criação e Aplicação de Regras

As regras clássicas que se extraem seguem a seguinte forma:  $A \Rightarrow B$ , onde A e B são respostas ao questionário. Este tipo de regras trabalha directamente com as relações do tipo positivo, onde uma resposta a uma pergunta A implica uma resposta a uma outra pergunta B. A forma de averiguar se o indivíduo respondeu de forma fidedigna passa pelo cumprimento de todas as regras geradas.

No entanto, estas regras da forma como estão construídas, não conseguem avaliar a veracidade das respostas submetidas mas é possível identificá-las através da alteração na forma como são construídas. Para sabermos se a resposta é fidedigna, observemos o seguinte:

Se o inquirido “x” responde a uma pergunta “i” com a resposta “a” (denotando por  $\text{Responde}(x,i,a)$ ), então o inquirido está a mentir ou de alguma forma ter-se-á enganado (denotado por  $\text{Mente}(x,i)$ ) ou é verdade que x,i,a (denotado por  $\text{Facto}(x,i,a)$ ).

Temos então:

$$\text{Responde}(x,i,a) \Rightarrow \text{Mente}(x,i) \vee \text{Facto}(x,i,a)$$

Note-se que as respostas estão acessíveis ao sistema de avaliação do inquirido, pelo que é possível saber se  $\text{Responde}(x,i,a)$ . Se  $\text{Responde}(x,i,a)$ , então é possível inferir que  $\text{Mente}(x,i)$  se  $\neg \text{Facto}(x,i,a)$ . É necessário um método que determine se  $\neg \text{Facto}(x,i,a)$ .

Dado que existe acesso às respostas, então as regras que são úteis são a do tipo  $\text{Resposta}(x,j,t) \Rightarrow \neg \text{Facto}(x,i,a)$ ; ou seja, regras que determinam que escolher uma determinada opção noutra pergunta determine que  $\neg \text{Facto}(x,i,a)$ .

Para ilustrar a forma como se conclui que um indivíduo está a responder a questões que não lhe dizem respeito, é demonstrado um processo simples de inferência.

---

### Exemplo Inferência

#### Regras:

i)  $\text{TemConta}(x, A1) \Rightarrow \neg \text{TemConta}(x,A2)$

**Textual: Se um indivíduo tem conta no sistema A1, então não tem conta no sistema A2.**

**Forma Normal Cognitiva (CNF):**  $\neg \text{TemConta}(x, A1) \vee \neg \text{TemConta}(x,A2)$

ii)  $\text{RespondeUsa}(x,y) \Rightarrow \text{Mentiu}(x) \vee \text{TemConta}(x,y)$

**Textual: Se um indivíduo responde que usa um sistema, então está a mentir ou tem conta nesse sistema.**

**Forma CNF:**  $\neg \text{RespondeUsa}(x,y) \vee \text{Mentiu}(x) \vee \text{TemConta}(x,y)$

Factos:

1) RespondeUsa (João, A1)

**Textual: O João respondeu que usa o sistema A1.**

2) RespondeUsa (João, A2)

**Textual: O João respondeu que usa o sistema A2.**

Efectuando resolução:

$\neg \text{RespondeUsa}(x,y) \vee \text{Mentiu}(x) \vee \text{TemConta}(x,y)$	<b>(Regra ii)</b>
$\Gamma \text{RespondeUsa}(\text{João}, \text{A1})$	
$\text{Mentiu}(\text{João}) \vee \text{TemConta}(\text{João}, \text{A1})$	
$\Gamma \neg \text{TemConta}(x, \text{A1}) \vee \neg \text{TemConta}(x, \text{A2}), \{x/\text{João}\}$	<b>(Regra i)</b>
$\text{Mentiu}(\text{João}) \vee \neg \text{TemConta}(x, \text{A2})$	
$\Gamma \text{Mentiu}(\text{João}) \vee \text{TemConta}(\text{João}, \text{A2}), \{x/\text{João}\}$	<b>(Regra ii.2)</b>
$\text{Mentiu}(\text{João}) \vee \text{Mentiu}(\text{João}) = \underline{\text{Mentiu}(\text{João})}$	

No exemplo foi demonstrada a utilidade do uso de regras do tipo  $A \Rightarrow \neg B$  para averiguar a veracidade das respostas submetidas pelos utilizadores.

## 3.5 Implementação

### 3.5.1 Metodologia de Avaliação

A metodologia proposta para avaliar um questionário consiste em grande parte por extracção de informação e troca de ideias entre o indivíduo que está a implementar o sistema e a chefia. No entanto, após documentação dos critérios e a sua tabela de pontuação, é necessário algo que automatize e cria a escala final. Foi por esta razão que foi criado um programa em MATLAB para esse efeito (código fonte em Apêndice B).

Para cada critério, é necessário utilizar o programa para obter a função de Transferência que mapeia todos os resultados aos respectivos pontos. O programa necessita de dois vectores: (1) vector com as 4 percentagens em formato decimal (Ex: resultados = [1; 0,9; 0,85; 0,8]); (2) vector com os 4 valores que definem os pontos de cada resultado (Ex: pontos = [1; 0,8; 0,6; 0,3]). Com estes vectores, o programa calcula a função Transferência que minimiza o erro quadrático, guardando no final todos os pontos da função que definem os pares Resultados-Pontos.

### 3.5.2 Endereçamento de Perguntas

O método proposto de endereçamento de perguntas consiste na execução de vários processos que seguem uma ordem específica. A Figura 3.8 apresenta uma visão alto-nível do funcionamento e tarefas que são executadas quando um utilizador pretende preencher o questionário.

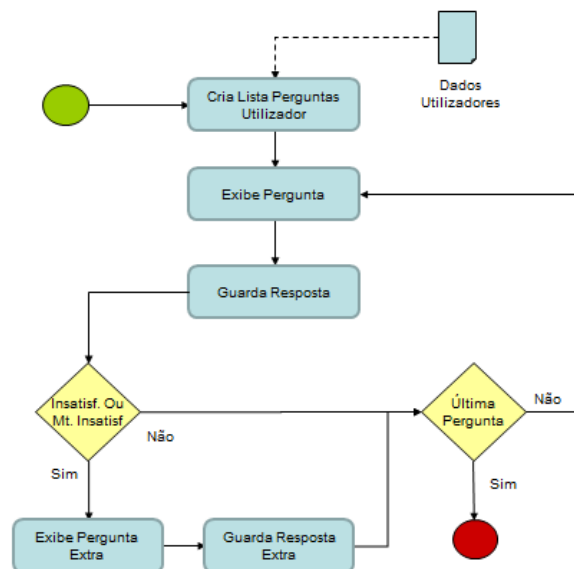
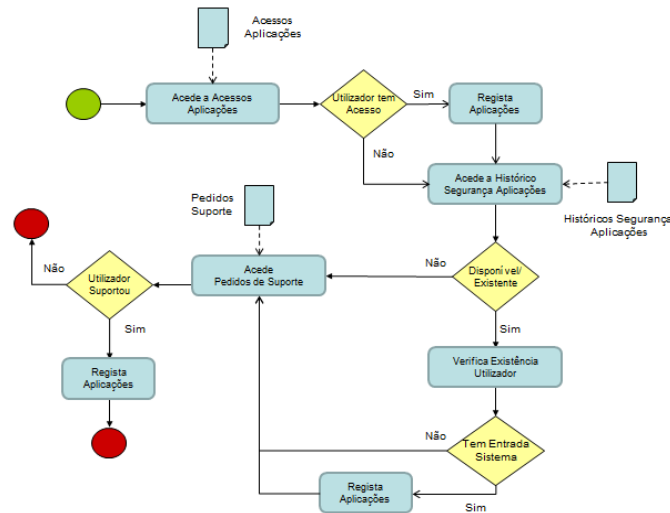


Figura 3.8 – Visão Alto-Nível do Processo de Endereçamento de Perguntas



**Figura 3.9 – Processo de Selecção de Aplicações Alvo de Perguntas**

Na Figura 3.9 é apresentado um diagrama que ilustra a forma como os diversos processos se relacionam no momento de identificação das aplicações e sistemas que o utilizador usa durante a sua actividade. Ao longo do processo, são registados os vários sistemas, aplicações e serviços que o utilizador usou ao longo do ano de actividade onde posteriormente é utilizada para escolher as perguntas relevantes ao utilizador.

O processo “Regista Aplicações” consiste numa lista ou tabela que é actualizada com novas aplicações ou sistemas no momento que passa cada filtro. São apenas adicionadas aplicações e sistemas que não estejam na lista, isto é, existem apenas registos únicos.

### 3.5.3 Ponderação de Respostas

O método de aplicação de ponderações às respostas do questionário foi testada manualmente, isto é, todos os cálculos foram efectuados utilizando folhas de cálculo Excel. Os critérios e as percentagens de cada característica foram facultados pelo *Service Owner*.

### 3.5.4 Regras de Associação

De forma a criar regras de associação, foi implementada uma versão modificada do algoritmo *Apriori*. Como explicado no subcapítulo 3.4.4.2, as regras idealizadas são do tipo  $A \Rightarrow \neg B$  ao contrário das clássicas  $A \Rightarrow B$ .

O algoritmo foi implementado utilizando o ambiente de desenvolvimento MATLAB e o programa *open-source RapidMiner* como auxílio.

A Figura 3.10 apresenta o *flowchart* que esquematiza todas as actividades executadas para se obter o conjunto de regras.

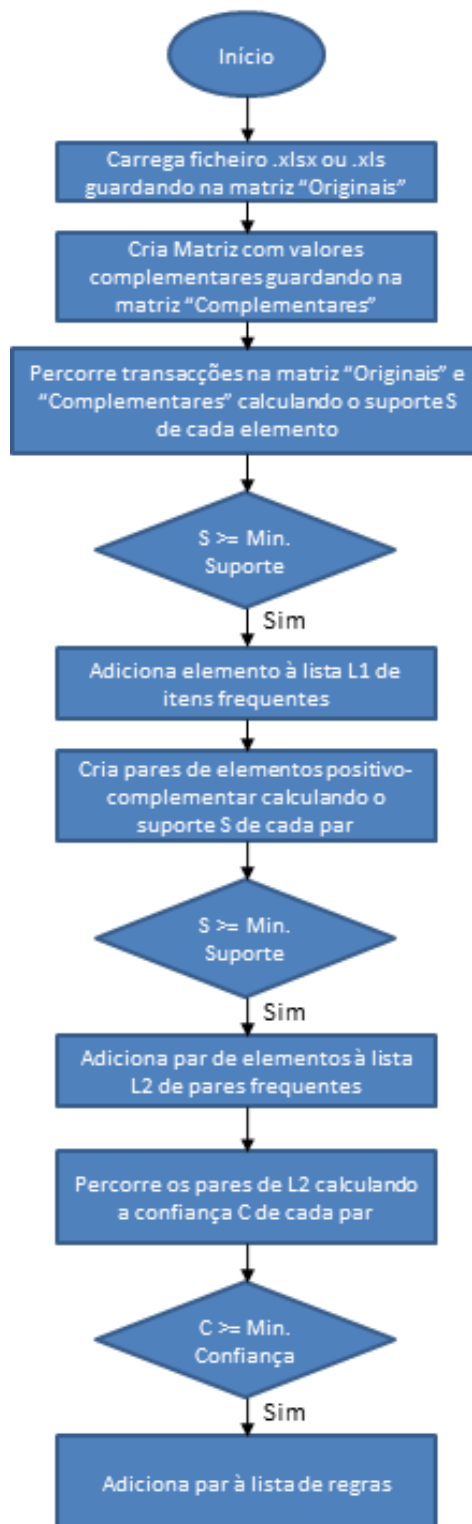


Figura 3.10 – *Flowchart* do algoritmo *Apriori* modificado

## Funcionamento do Algoritmo

Sendo o objectivo a obtenção de regras com premissa positiva e inferência negativa, o primeiro passo foi a criação de uma tabela de transacções com valores complementares. Na prática, caso na tabela original exista uma transacção com os itens A e C, na tabela complementar irá estar o par de itens  $\neg A$  e  $\neg C$ .

Após a criação da tabela complementar, é calculado o suporte de cada item nas duas tabelas e seleccionados os itens que apresentam um suporte superior ao mínimo estabelecido. As duas tabelas são actualizadas assinalando com *flags* 0 e 1 os diversos itens.

É calculado o suporte de pares de itens positivos-complementares excepto itens directamente complementares, ou seja, A e  $\neg A$  não é contabilizado pelo suporte ser sempre 0. São guardados os pares que obtiveram suporte superior ao valor mínimo estabelecido.

Encontrados os pares, é calculado a confiança de cada par que caso obtenham confiança superior ao mínimo exigido é adicionada à lista de regras fortes.



# Capítulo 4

## Resultados

### 4.1 Background

Esta dissertação é baseada num projecto da empresa Tabaqueira II, S.A. – Afiliada da Philip Morris International que, pela primeira vez, cooperou com Faculdade de Engenharia da Universidade Católica Portuguesa no sentido de auxiliar ao máximo e dentro do possível a elaboração de uma Dissertação para obtenção de grau de Mestre.

#### A Empresa

A Tabaqueira foi fundada em 1927 pelo carismático empresário Alfredo da Silva, tendo inaugurado a actual fábrica de Albarraque em 1962.

Em 1996 são aprovadas por Decreto-Lei as condições prévias para a reprivatização da totalidade do capital social da Tabaqueira, S.A., em três fases, terminando a última das quais em Dezembro de 2000, através de uma Oferta Pública de Venda. Terminado o processo de privatização, empresas da Philip Morris International, Inc. detêm mais de 99% do capital social da Tabaqueira.

Em 2008, a Tabaqueira passa a estar organizada em duas “empresas-irmãs”, a Tabaqueira – Empresa Industrial de Tabacos, S.A. responsável pela produção de cigarros e produtos afins, enquanto que a Tabaqueira II, S.A. é responsável pela comercialização de cigarros e produtos afins no mercado português, incluindo Madeira e Açores.

A Tabaqueira é a maior empresa tabaqueira de Portugal e também uma das maiores empresas do país. Localizada no Concelho de Sintra, a fábrica é uma das maiores instalações de produção da Philip Morris International da União Europeia, exportando cerca de 60% da sua produção anual para mais de 35 países.

#### Objectivos da Tabaqueira

A Tabaqueira, enquanto afiliada da Philip Morris International, tem como principais objectivos responder às expectativas de fumadores adultos, disponibilizando produtos de tabaco inovadores e de mais alta qualidade nas suas categorias de preço preferidas, reduzir a nocividade causada pelos produtos de tabaco, apoiando a adopção de regulamentação abrangente e desenvolvendo produtos que tenham o potencial de redução dos riscos de

desenvolvimento de doenças associadas ao tabaco, e ser uma entidade empresarial responsável, conduzindo os seus negócios com o mais elevado grau de integridade. Entre as diversas marcas produzidas pela empresa, encontram-se as mais bem-sucedidas do mercado, nomeadamente *Marlboro*, *L&M*, *SG* ou *Chesterfield*.

A Philip Morris Internacional (PMI), é a principal empresa internacional no sector do tabaco, com produtos vendidos em aproximadamente 180 países. Em 2010, tinha uma quota estimada de 16,0% do mercado mundial de cigarros fora dos E.U.A., ou 27,6% excluindo a República Popular da China e os E.U.A..

Para a Tabaqueira, a responsabilidade começa no produto, através do compromisso em divulgar, de forma aberta e transparente, informação pública sobre os riscos para a saúde decorrentes do acto de fumar e também em apoiar a adopção de uma regulamentação do tabaco onde quer que sejam vendidos os seus produtos. Para além disso, a Tabaqueira apoia iniciativas em comunidades locais onde vivem e trabalham os nossos funcionários. Faz parte da filosofia da empresa contribuir para questões sociais, maioritariamente em cinco áreas fundamentais: fome e pobreza extrema, educação, meio ambiente sustentável, violência doméstica e assistência em caso de catástrofe.

A PMI dispõe dos melhores recursos existentes na área da Investigação e Desenvolvimento, dedicado ao desenvolvimento de produtos que tenham o potencial para reduzir o risco de doenças associadas ao tabaco. Recentemente, foi inaugurado um novo centro de I&D de última geração em Neuchâtel, na Suíça

Actualmente, a PMI emprega mais de 78.000 pessoas em todo o mundo. Em Portugal, a Tabaqueira conta com mais de 700 colaboradores.

### **Situação Actual**

Anualmente é efectuado um questionário que abrange todos os serviços do IS (*Information Services*) e que é enviado a todos os trabalhadores da Tabaqueira com o objectivo de documentar e analisar os níveis de satisfação dos seus trabalhadores no que toca ao uso de serviços de informação e forma como são disponibilizados e suportados pelo IS.

Os questionários são preenchidos de forma anónima e têm a estrutura definida por um *Template* disponibilizado pela PMI (de modo a facilitar a análise dos resultados de todas as suas Afiliadas espalhadas por todo o Mundo) sendo implementado utilizando um *software* específico, denominado *Web Surveyor*.

As perguntas efectuadas abrangem diversas áreas consideradas cruciais para a prestação dos serviços. Todas as perguntas (à excepção de comentários) usam uma escala com 5 valores de

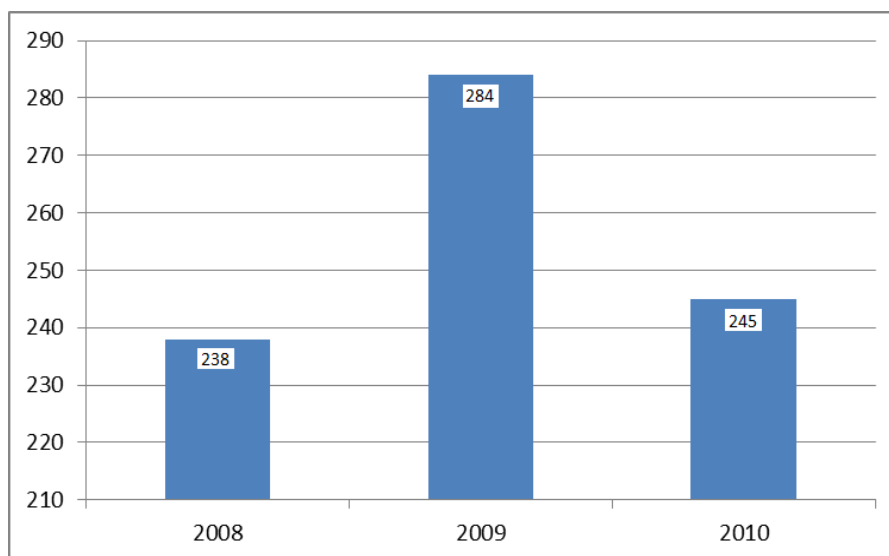
satisfação (*Very Dissatisfied, Dissatisfied, Neutral, Satisfied, Very Satisfied*) e um de abstenção (*Not Applicable*). As perguntas referem-se à:

- Classificação de aspectos presentes no ambiente de trabalho (*Base IT Services*)
- Classificação do apoio da *Help Desk* na resolução de problemas via telefone e assistência pessoal no escritório (*IS Help Desk and Desk-Side Support*)
- Classificação de ferramentas e serviços utilizados pela Tabaqueira (SAP BW, SAP CRM, iSMS, etc.)
- Classificação do serviço do IS no geral e área disponível para escrever a melhor experiência e sugestões de melhoria de serviço

Existem problemas quanto ao endereçamento das perguntas efectuadas, sendo estas direccionadas a todos os utilizadores e não apenas aqueles que de facto usam o serviço em questão (*target group*). Como resultado, o *output* não espelha a real satisfação dos utilizadores nos diversos serviços visto haver a possibilidade dos utilizadores responderem a qualquer questão. No entanto existem outras áreas com margem para progresso:

- Não existe diferenciação entre utilizadores experientes e utilizadores ocasionais
- As perguntas efectuadas não são baseadas em dados históricos – todas as perguntas têm o mesmo peso quando é efectuada a análise de dados e conclusões acerca dos índices de qualidade
- O questionário não é dinâmico – todos os utilizadores respondem às mesmas perguntas onde deverão seleccionar a opção “Not Applicable” caso não use esse serviço

Outro indicador é o número de questionários preenchidos e recebidos para análise e extracção dos índices de satisfação. Na Figura 4.1 apresenta-se o gráfico com o número de respostas recebidas nos últimos 3 anos de actividade.



**Figura 4.1 – Número de respostas ao questionário**

Para compreender melhor o gráfico e o significado dos valores, os inquéritos foram administrados a aproximadamente 800 utilizadores. Obtém-se então uma taxa média de participação na ordem dos 30%, valor que se encontra longe do óptimo (100% participação). Como objectivo do novo modelo de questionário resultante da aplicação da metodologia pretende-se maximizar a taxa de participação.

Surge a oportunidade de criar uma metodologia de análise de qualidade do questionário actual e aplicação de técnicas para, não só resolver os problemas referidos em cima, mas também para introduzir o conceito de questionário dinâmico e realmente adaptável a cada utilizador.

## **4.2 Aplicação da Ferramenta de Avaliação**

Para demonstrar a forma como a avaliação do questionário é feita utilizando critérios, apresenta-se a classificação dos questionários mediante um critério: Taxa de Questionários Completos (número de questionários totalmente preenchidos).

### **4.2.1 Criação de Escala**

Atendendo aos resultados verificados nos anos anteriores, estabelece-se uma escala de pontuação do critério em questão (Tabela 4.1).

Tabela 4.1 – Escala de Pontuação de Taxas de Questionários Completos

CLASSIFICAÇÃO	RESULTADOS	PONTOS
Ideal	100%	1
Excelente	95%	0,9
Aceitável	80%	0,6
Razoável	70%	0,5
Mau	< 70%	0,1

Utilizando o programa implementado em MATLAB, obtém-se o conjunto de pares Resultado-Pontos para extrair a pontuação do critério.

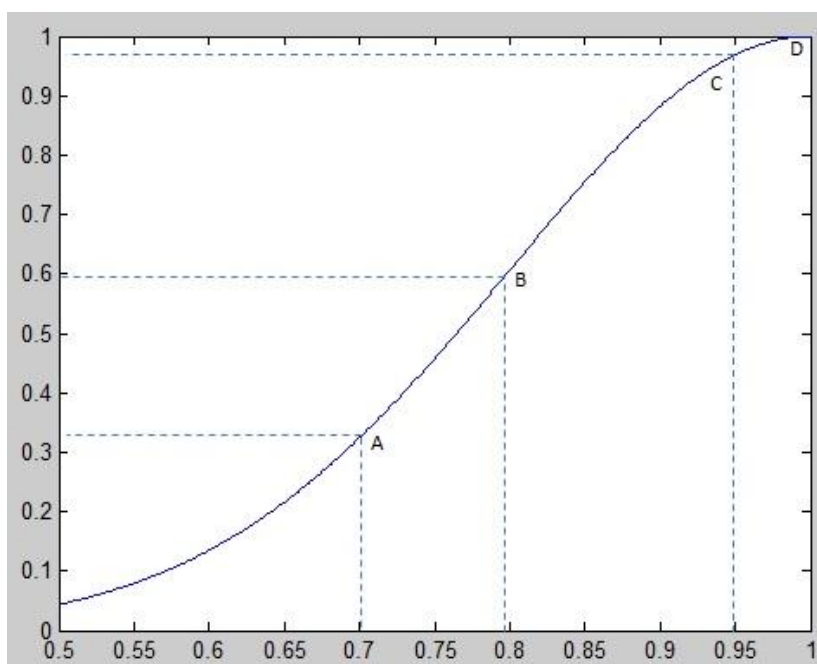


Figura 4.2 – Apresentação Parcial da Escala Criada

Na Figura 4.2 é apresentado gráfico resultante da função de Transferência calculada com base na informação da Tabela 4.1.

#### 4.2.2 Extracção de Pontos

Estando na posse da grelha de avaliação, esta é utilizada para extrair os pontos que são atribuídos aos questionários dos anos passados. Através da ferramenta *Web Surveyor*, extraiu-se a taxa de questionários totalmente preenchidos. Verificaram-se as seguintes taxas:

- Questionário 2010 – 82,2% Completos
- Questionário 2009 – 79,8% Completos
- Questionário 2008 – 72,3% Completos

Utilizando a informação presente na matriz que define os resultados e os pontos equivalentes, retira-se a pontuação final para o critério em questão:

- 2010 => 0,673 Pontos
- 2009 => 0,6 Pontos
- 2008 => 0,383 Pontos

## 4.3 Aplicação de Endereçamento de Perguntas

### 4.3.1 Recolha de dados

Da aplicação da técnica de endereçamento de perguntas resulta a criação de uma lista de aplicações e serviços que o utilizador usa durante a sua actividade, e que por essa razão se encontra em condições de responder a perguntas sobre esse conjunto. São apresentados 3 exemplos de aplicação do método. Foram considerados 3 utilizadores de departamentos diferentes de forma a serem identificadas diferentes aplicações/sistemas:

- Utilizador A – Departamento Vendas
- Utilizador B – Departamento Operações
- Utilizador C – Departamento Marketing

Na Tabaqueira, é usada uma aplicação denominada IMDL (*Identity Management Data Layer*) que gere os acessos dos utilizadores aos vários sistemas e o nível de acesso que cada utilizador tem a cada sistema. É solicitado ao administrador a extracção de toda a informação relevante para a aplicação da técnica, incluindo ID de utilizadores, ID de sistemas, níveis de acesso, etc. A extracção é feita para duas folhas de cálculo Excel: uma folha com os *roles* (cargos dos utilizadores mediante a sua função) e outra com toda a informação do utilizador e sistemas que tem acesso.

J	A	B	C	D	E	F	G	H	I	J	K
1	IMDLAcctID	Display name	Bus. role	Business role descr.	Start date	End date	AuthObjID	Authorization object	SyUID	FuncSyUID	SystAbbr
2	1000000270		10000000	Ereq Requestor	20.01.2011	31.12.9999	1000000000	ZGL_EMPLOYEE	10001	10001	HSP100
3	1000000270		10000003	Ereq Approver	20.01.2011	31.12.9999	1000000002	ZGL_MANAGER	10001	10001	HSP100
4	1000000270		10001961	MAS: Display Departmental	04.01.2011	31.12.9999	1500002335	WPTZLXXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
5	1000000270		10001961	MAS: Display Departmental	04.01.2011	31.12.9999	1500002652	WPTZMAAX00_D_XL_PT	10002	10002	WCP100
6	1000000270		10001961	MAS: Display Departmental	04.01.2011	31.12.9999	1500005241	WPTYFXXXXORGFCA	10002	10002	WCP100
7	1000000270		10001961	MAS: Display Departmental	04.01.2011	31.12.9999	1500026404	WPTZXXXX_F_EU_PT	10002	10002	WCP100
8	1000000270		10001962	MKT: Brand Manager	04.01.2011	31.12.9999	1500002335	WPTZLXXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
9	1000000270		10001962	MKT: Brand Manager	04.01.2011	31.12.9999	1500002623	WPTZMAMX05_P_MG_PT	10002	10002	WCP100
10	1000000270		10001962	MKT: Brand Manager	04.01.2011	31.12.9999	1500005241	WPTYFXXXXORGFCA	10002	10002	WCP100
11	1000000270		10001999	STA: Manager Role (BS)	04.01.2011	31.12.9999	1500002335	WPTZLXXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
12	1000000270		10001999	STA: Manager Role (BS)	04.01.2011	31.12.9999	1500002624	WPTZAX01T_XP_MG_PT	10002	10002	WCP100
13	1000000270		10001999	STA: Manager Role (BS)	04.01.2011	31.12.9999	1500005241	WPTYFXXXXORGFCA	10002	10002	WCP100
14	1000000270		10002011	PMI PT 1099 all line managers	01.01.2011	31.12.9999	1000000041	ZXDX8XXAL	10003	10003	GMP130
15	1000000270		10002011	PMI PT 1099 all line managers	01.01.2011	31.12.9999	1500000641	P-Z-GLBT-MSS-LINE MANAGERS	10010	10010	HRBW
16	1000000270		10002011	PMI PT 1099 all line managers	01.01.2011	31.12.9999	1500002369	PMIPT1099URAILineManagers_IMDL	10000	10000	AD
17	1000000270		10002011	PMI PT 1099 all line managers	01.01.2011	31.12.9999	1500002372	DLPMIPT1099AILineManagers_IMDL	10000	10000	AD
18	1000000270		10002011	PMI PT 1099 all line managers	01.01.2011	31.12.9999	1500002656	H-GLB-MSS-LINE MANAGER	10007	10007	HHP300
19	1000000270		10002012	PMI PT 1099 all employees	01.01.2011	31.12.9999	1500002335	WPTZLXXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
20	1000000270		10002012	PMI PT 1099 all employees	01.01.2011	31.12.9999	1500002370	PMIPT1099URAILEmployees_IMDL	10000	10000	AD
21	1000000270		10002012	PMI PT 1099 all employees	01.01.2011	31.12.9999	1500002373	DLPMIPT1099AILEmployees_IMDL	10000	10000	AD
22	1000000270		10002012	PMI PT 1099 all employees	01.01.2011	31.12.9999	1500002655	H-GLB-ESS-EMPLOYEE	10007	10007	HHP300
23	1000000270		10002012	PMI PT 1099 all employees	01.01.2011	31.12.9999	1500027382	ZEIC_WEBCRMTC	10008	10008	HKP100
24	1000000270		10003821	Approve Invoice via Workflow Portugal	04.01.2011	31.12.9999	1500002335	WPTZLXXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
25	1000000270		10003821	Approve Invoice via Workflow Portugal	04.01.2011	31.12.9999	1500004404	WPTZFXXXX_P_APPR_INV	10002	10002	WCP100
26	1000000270		10003821	Approve Invoice via Workflow Portugal	04.01.2011	31.12.9999	1500005241	WPTYFXXXXORGFCA	10002	10002	WCP100
27	1000000270		10004916	PMI Global All Marketing Users IMDL	08.07.2008	31.12.9999	1500005400	PMIGlobalURAILMarketing_IMDL	10000	10000	AD
28	1000000270		10004916	PMI Global All Marketing Users IMDL	08.07.2008	31.12.9999	1500005402	DLPMIIGlobalAILMarketing_IMDL	10000	10000	AD
29	1000000270		10006057	PT - 1099 PDF Generator Document Co	04.01.2011	31.12.9999	1500006479	PMIPTABURISPDFGenerator_IMDL	10000	10000	AD
30	1000000270		10006542	Extended Internet Access	01.09.2008	31.12.9999	1500006961	PMIGlobalURExtendedInternetAccess_IMDL	10000	10051	AD

Figura 4.3 – Informação extraída de IMDL sobre cargos dos utilizadores

Na Figura 4.3, é apresentado parte do conteúdo da folha com os cargos acompanhados dos identificadores dos utilizadores entre outras informações.

J	A	B	C	D	E	F	G	H	I
1	IMDL account ID	Login name	Display name	Function name	Subfunction name	Position name	Department name	System abbreviation	System description
2	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	AD	Active Directory
3	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	HSP100	EREQ (SRM) Pr
4	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	WCP100	P1WE Productor
5	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	GMP130	SMDLIMDL Prod
6	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	HHF300	HRZU Production
7	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	HKP100	EIC(CRM) Produc
8	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	HRBW	HRBW Productor
9	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	MCT_PROD	Marketing Collab
10	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	INTERNET	Internet Services
11	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	VCM	Value Content Ma
12	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	BIL_PROD	BIL & SmartPOST
13	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	GPS_PROD	Global integrated
14	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	BBI_PROD	Brand Building Se
15	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	BV_PROD	Brand Vision Proc
16	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	GIMS_PROD	GIMS Production
17	1000000270			Marketing	Brand Management	Group Brand Manager Premium Brands	Operations Center PMI	GSFR_PROD	GSFR Production
18	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	AD	Active Directory
19	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	HSP100	EREQ (SRM) Pr
20	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	WCP100	P1WE Productor
21	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	GMP130	SMDLIMDL Prod
22	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	HHF300	HRZU Production
23	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	HKP100	EIC(CRM) Produc
24	1000000441			Operations	Manufacturing / Producti	Supervisor Secondary Support	Operations Portugal	HRBW	HRBW Productor
25	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	AD	Active Directory
26	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	HSP100	EREQ (SRM) Pr
27	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	WCP100	P1WE Productor
28	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	GMP130	SMDLIMDL Prod
29	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	HHF300	HRZU Production
30	1000001242			Operations	Operations	Director Operations PT	Operations Portugal	HKP100	EIC(CRM) Produc

Figura 4.4 – Informação extraída de IMDL sobre utilizadores dos sistemas

Na Figura 4.4, é apresentado parte do conteúdo da folha com informações sobre os utilizadores. Entre a informação encontra-se: nome *login*, função, subfunção, departamento, sistemas com acesso, etc.

### 4.3.2 Aplicação de Filtros

Extraída a informação necessária, é efectuada a identificação das aplicações que serão alvo de perguntas ao utilizador em questão.

#### Filtro 1 – IMDL

O utilizador ao iniciar, o questionário é identificado pelo sistema segundo o seu *username* e é efectuada a sua identificação na tabela de informação de utilizadores após a aplicação de um filtro que isola o utilizador. Como é possível verificar nas figuras 4.3 e 4.4, o campo comum de identificação (Chave Primária) da tabela de utilizadores e tabela de cargos é o campo “*IMDLAcctID*” (na tabela de utilizadores tem o nome “*IMDL Account ID*”).

Sabendo o “*IMDL Account ID*” proveniente da tabela de utilizadores, é aplicado um filtro na tabela de cargos para isolar o utilizador. É identificado o utilizador segundo o seu ID único, onde se obtém os vários cargos.

IMDLAccID	Display name	Bus. role	Business role descr.	Start date	End date	AuthObj	Authorization object	SyUID	FuncSyUID	SystAbbr
100023536		10001961	MAS: Display Departmental	20.12.2007	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
100023536		10001961	MAS: Display Departmental	20.12.2007	31.12.9999	1500002652	WPTZMAM00_D_XX_PT	10002	10002	WCP100
100023536		10001961	MAS: Display Departmental	20.12.2007	31.12.9999	1500005241	WPTYFXXXXXORGFCFA	10002	10002	WCP100
100023536		10001961	MAS: Display Departmental	20.12.2007	31.12.9999	1500026404	WPTZXXXXXX_F_EU_PT	10002	10002	WCP100
100023536		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
100023536		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002370	PMIPT1099URAIIEmployees_IMDL	10000	10000	AD
100023536		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002655	H-GLB-ESS-EMPLOYEE	10007	10007	HHP300
100023536		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500027382	ZEIC_WEBCRMTC	10008	10008	HKP100
100023536		10002257	MKS: BW Departmental Key User	30.10.2009	31.12.9999	1500001915	WXZLXBXW_A_BASIS_ALL_USERS	10015	10015	WBP100
100023536		10002257	MKS: BW Departmental Key User	30.10.2009	31.12.9999	1500002880	WPTZMAM00_D_KU_PT	10015	10015	WBP100
100023536		10003642	IMDL area/ external resource responsib	04.09.2008	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
100023536		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
100023536		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500004404	WPTZFXXXX_P_APPR_INV	10002	10002	WCP100
100023536		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500005241	WPTYFXXXXXORGFCFA	10002	10002	WCP100
100023536		10004917	PMI Global All Sales Users IMDL	08.07.2008	31.12.9999	1500005401	PMIGlobalURAIISales_IMDL	10000	10000	AD
100023536		10004917	PMI Global All Sales Users IMDL	08.07.2008	31.12.9999	1500005403	DLPMIGlobalAllSales_IMDL	10000	10000	AD
100023536		10000657	PT - 1099 PDF Generator Document Co	17.12.2010	31.12.9999	1500006479	PMIPTTABURISPDFGenerator_IMDL	10000	10000	AD
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011354	WDQZXXXXXX_D_VIEW	10134	10134	WBD300
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011355	WDQZXXXXXX_D_VIEW	10135	10135	WBD200
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011359	WXZLXBXW_A_BASIS_ALL_USERS	10134	10134	WBD300
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011360	WXZLXBXW_A_BASIS_ALL_USERS	10135	10135	WBD200
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011362	WDQZR02KX_XP	10134	10134	WBD300
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011366	WDQZR02XU_D_P1-ROLLIN	10134	10134	WBD300
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011367	WDQZR02XU_D_P1-ROLLIN	10135	10135	WBD200
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011368	WDQZR02XU_XP_P1-ROLLIN	10134	10134	WBD300
100023536		10010679	PIWE BW BUSINESS END USER QA/07.06.2010	07.06.2010	31.12.9999	1500011369	WDQZR02XU_XP_P1-ROLLIN	10135	10135	WBD200
100023536		10014568	PT 1099 SL Mantain VM data change	11.05.2009	31.12.9999	1500013513	ZPT	10189	10189	PCR_PROD
100023536		10014568	PT 1099 SL Mantain VM data change	11.05.2009	31.12.9999	1500013530	ZPTMMKVM	10189	10189	PCR_PROD

Figura 4.5 – Cargos do utilizador A (ID = 1000023535)

IMDLAccID	Display name	Bus. role	Business role descr.	Start date	End date	AuthObj	Authorization object	SyUID	FuncSyUID	SystAbbr
1000131225		10000000	Ereq Requestor	15.07.2009	31.12.9999	1000000000	ZGL_EMPLOYEE	10001	10001	HSP100
1000131225		10000059	IPSM CH Display specifications all mat	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000059	IPSM CH Display specifications all mat	15.07.2009	31.12.9999	1500000016	ZCHXXXXAD	10003	10003	GMP130
1000131225		10000063	IPSM DE Display specifications without	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000063	IPSM DE Display specifications without	15.07.2009	31.12.9999	1500000041	ZDEXXXXXAD	10003	10003	GMP130
1000131225		10000064	IPSM DE Display specifications solution	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000064	IPSM DE Display specifications solution	15.07.2009	31.12.9999	1500000040	ZDEXISLAD	10003	10003	GMP130
1000131225		10000125	IPSM NL ALL DISPLAY FOR NL09	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000126	IPSM NL ALL DISPLAY FOR NL09	15.07.2009	31.12.9999	1500000199	ZNLXXXXAD	10003	10003	GMP130
1000131225		10000185	IPSM CZ DISPLAY ALL MARETIALS F	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000185	IPSM CZ DISPLAY ALL MARETIALS F	15.07.2009	31.12.9999	1500000035	ZCZXXXXAD	10003	10003	GMP130
1000131225		10000200	IPSM LT Display all materials for LT01	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000200	IPSM LT Display all materials for LT01	15.07.2009	31.12.9999	1500000056	ZLTXXXXXAD	10003	10003	GMP130
1000131225		10000265	IPSM IT Display All Materials for IT40	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000265	IPSM IT Display All Materials for IT40	15.07.2009	31.12.9999	1500000339	ZITXXXXAD	10003	10003	GMP130
1000131225		10000524	IPSM CV MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000524	IPSM CV MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1500000063	ZCVXXXXAM	10003	10003	GMP130
1000131225		10000524	IPSM CV MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1500009676	XX_YYYY_IPSM_IMDL	10003	10003	GMP130
1000131225		10000525	IPSM CV DISPLAY ALL MATERIALS F	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000525	IPSM CV DISPLAY ALL MATERIALS F	15.07.2009	31.12.9999	1500000664	ZCVXXXXAD	10003	10003	GMP130
1000131225		10000526	IPSM CV Approving Manager level 1 for	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000526	IPSM CV Approving Manager level 1 for	15.07.2009	31.12.9999	1500000665	CVXXXX1PDM	10003	10003	GMP130
1000131225		10000526	IPSM CV Approving Manager level 1 for	15.07.2009	31.12.9999	1500009676	XX_YYYY_IPSM_IMDL	10003	10003	GMP130
1000131225		10000530	IPSM PT MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000530	IPSM PT MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1500000669	ZPTXXXXAM	10003	10003	GMP130
1000131225		10000531	IPSM PT MAINTAIN ALL MATERIALS F	15.07.2009	31.12.9999	1500009676	XX_YYYY_IPSM_IMDL	10003	10003	GMP130
1000131225		10000531	IPSM PT DISPLAY ALL MATERIALS F	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000131225		10000531	IPSM PT DISPLAY ALL MATERIALS F	15.07.2009	31.12.9999	1500000670	ZPTXXXXAD	10003	10003	GMP130
1000131225		10000532	IPSM PT Approving Manager level 1 for	15.07.2009	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130

Figura 4.6 – Cargos do utilizador B (ID = 1000131225)

IMDLAccID	Display name	Bus. role	Business role descr.	Start date	End date	AuthObj	Authorization object	SyUID	FuncSyUID	SystAbbr
1000023617		10000000	Ereq Requestor	02.01.2008	31.12.9999	1000000000	ZGL_EMPLOYEE	10001	10001	HSP100
1000023617		10000003	Ereq Approver	02.01.2008	31.12.9999	1000000000	ZGL_MANAGER	10001	10001	HSP100
1000023617		10001961	MAS: Display Departmental	14.12.2007	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
1000023617		10001961	MAS: Display Departmental	14.12.2007	31.12.9999	1500002652	WPTZMAM00_D_XX_PT	10002	10002	WCP100
1000023617		10001961	MAS: Display Departmental	14.12.2007	31.12.9999	1500005241	WPTYFXXXXXORGFCFA	10002	10002	WCP100
1000023617		10001961	MAS: Display Departmental	14.12.2007	31.12.9999	1500026404	WPTZXXXXXX_F_EU_PT	10002	10002	WCP100
1000023617		10001962	MKT: Brand Manager	14.12.2007	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
1000023617		10001962	MKT: Brand Manager	14.12.2007	31.12.9999	1500002623	WPTZMAM05_P_MG_FT	10002	10002	WCP100
1000023617		10001962	MKT: Brand Manager	14.12.2007	31.12.9999	1500006699	ZPTXXXXAM	10002	10002	WCP100
1000023617		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
1000023617		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002370	PMIPT1099URAIIEmployees_IMDL	10000	10000	AD
1000023617		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500002655	H-GLB-ESS-EMPLOYEE	10007	10007	HHP300
1000023617		10002012	PMI PT 1099 all employees	01.01.2008	31.12.9999	1500027382	ZEIC_WEBCRMTC	10008	10008	HKP100
1000023617		10002249	FIN: BW Departmental End User	11.01.2008	31.12.9999	1500001915	WXZLXBXW_A_BASIS_ALL_USERS	10015	10015	WBP100
1000023617		10002249	FIN: BW Departmental End User	11.01.2008	31.12.9999	1500002872	WPTZFAXX00_D_EU_PT	10015	10015	WBP100
1000023617		10003642	IMDL area/ external resource responsib	01.04.2011	31.12.9999	1000000041	ZXD98XXAL	10003	10003	GMP130
1000023617		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500002335	WPTZLXXX_A_BASIS_ALL_USERS	10002	10002	WCP100
1000023617		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500004404	WPTZFXXXX_P_APPR_INV	10002	10002	WCP100
1000023617		10003821	Approve Invoice via Workflow Portugal	28.04.2010	31.12.9999	1500005241	WPTYFXXXXXORGFCFA	10002	10002	WCP100
1000023617		10004907	Brand Building VCM MKT High Access	08.07.2008	31.12.9999	1500005391	PMIGlobalURBMMKTHAuto_IMDL	10000	10000	AD
1000023617		10004916	PMI Global All Marketing Users IMDL	08.07.2008	31.12.9999	1500005402	DLPMIGlobalAllMarketing_IMDL	10000	10000	AD
1000023617		10004916	PMI Global All Marketing Users IMDL	08.07.2008	31.12.9999	1500005402	DLPMIGlobalAllMarketing_IMDL	10000	10000	AD
1000023617		10000657	PT - 1099 PDF Generator Document Co	10.12.2010	31.12.9999	1500006479	PMIPTTABURISPDFGenerator_IMDL	10000	10000	AD
1000023617		10006931	PT MCT PORTUGAL Editor	01.07.2008	31.12.9999	1500005246	PMIGlobalURAMCTViewer_IMDL	10000	10000	AD
1000023617		10006931	PT MCT PORTUGAL Editor	01.07.2008	31.12.9999	1500007410	PMIGlobalURAMCTPORTUGALEditor_IMDL	10000	10000	AD
1000023617		10008728	MK BIL & BV Access Prod Viewer	18.02.2010	31.12.9999	1500009174	PMIGLOBALURBILPOSMPROD_IMDL	10000	10000	AD
1000023617		10008728	MK BIL & BV Access Prod Viewer	18.02.2010	31.12.9999	1500005298	PMIGLOBALURBILPOSMPROVIEWER_IMD	10000	10000	AD
1000023617		10008741	PLM/GIPS Packaging Read Only	28.11.2008	31.12.9999	1500009184	DLPMIGlobalGIPSPackagingUsers_IMDL	10000	10000	AD

Figura 4.7 – Cargos do utilizador C (ID = 1000023617)

No campo “SystAbbr” estão identificados os sistemas que os utilizadores têm acesso. Os sistemas de produção (versão utilizada para desenvolvimento das actividades da Empresa) são guardados num

Através dos dados provenientes das Tabelas 4.5, 4.6 e 4.7, são identificados os sistemas que cada utilizador tem acesso:

- Utilizador A – HHP300, HKP100, PCR\_PROD, WBP100, WCP100
- Utilizador B – HHP300, HKP100, HSP100, P53, SCP100, WBP100, WCP100
- Utilizador C – HHP300, HKP100, HSP100, WBP100, WCP100

O questionário de satisfação pretende apenas avaliar sistemas/aplicações disponibilizadas para uso final de utilizadores, isto é, apenas os sistemas de produção deverão ser avaliados. Na lista anterior, os utilizadores tem acesso a sistemas de Qualidade e de Desenvolvimento (identificados pela existência no nome do sistema das letra ‘D’ e ‘Q’ para Desenvolvimento e Qualidade respectivamente) que deverão ser desprezados.

Na folha de cálculo, existe outro campo de extrema importância pois apresenta as aplicações que estão alojadas nos sistemas identificados anteriormente. O campo “*Business role descr.*”, caracterizado por ser um campo de texto livre com a descrição dos cargos do utilizador, identifica as aplicações que o utilizador tem acesso. Analisando novamente a folha de cálculo, identificam-se as seguintes aplicações:

- Utilizador A – MAS, MKS, PDF *Generator*
- Utilizador B – Ereq, IPSM, PDF *Generator*, CMRT, QSMP, PSS4, ISS, STP
- Utilizador C – Ereq, BV, BIL, MAS, PDF *Generator*, MCT, GIPS, GIMS, GSPR

Isolados os sistemas e aplicações que o utilizador tem acesso, prossegue-se para a utilização do segundo filtro.

## **Filtro 2 – Históricos de Acesso**

Tendo presente a importância da diferença entre acesso e uso de um sistema, revela-se necessário o levantamento dos sistemas que o utilizador usa. No sentido de os determinar, recorre-se a históricos (*logs*) de acessos dos sistemas existentes na Empresa. Efectua-se uma análise baseada no número de acessos que o utilizador efectuou aos diversos sistemas, isto é, número de *logins* que o utilizador efectuou durante o último ano de actividade e com que frequência.

Para exemplificar o uso do segundo filtro, foi apenas analisado o sistema SAP P1 (solução ERP da SAP). O processo deverá ser aplicado a todas as aplicações que tenham históricos e também a serviços (Ex: Impressoras). O acesso é feito através de uma transacção específica de SAP (código de transacção SM20) para efeitos de auditoria.

**Analysis of Security Audit Log**

Period Requested 22.08.2011 09:00:00 - 26.08.2011 23:59:59  
 Period Selected 25.08.2011 12:01:42 - 25.08.2011 12:03:18  
 Server [redacted]  
 User [redacted]  
 Audit Classes Dialog Logon

Server Name	Creation Date	Date/Time	User Name	Terminal name	Transaction Code	Program	Message Text
[redacted]	25.08.2011	12:01:42	[redacted]	PMPITLISW09275	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	25.08.2011	12:03:18	[redacted]	PMPITLISW09275	ZWPPTS_REPORT_SALES	ZWPPTS_SALES_CUSTOMER_BRAND	User Logoff

Figura 4.8 – Histórico de acessos do Utilizador A em SAP P1

**Analysis of Security Audit Log**

Period Requested 04.07.2011 09:00:00 - 08.07.2011 23:59:59  
 Period Selected 04.07.2011 11:37:52 - 07.07.2011 19:36:45  
 Server [redacted]  
 User [redacted]  
 Audit Classes Dialog Logon

Server Name	Creation Date	Date/Time	User	Terminal name	Transaction Code	Program	Message Text
[redacted]	06.07.2011	17:57:03	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	06.07.2011	17:57:12	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	User Logoff
[redacted]	06.07.2011	19:00:59	[redacted]	PMPITLISL09674	ZVXOR M080 002	ZVXORP ZVXX MKAL MAINTAB	User Logoff
[redacted]	07.07.2011	18:42:11	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	07.07.2011	19:50:45	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPLSMTR_NAVIGATION	User Logoff
[redacted]	05.07.2011	12:58:42	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	05.07.2011	13:03:14	[redacted]	PMPITLISL09674	MB51	RM07DOCS	User Logoff
[redacted]	05.07.2011	17:17:38	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	05.07.2011	17:48:47	[redacted]	PMPITLISL09674	MB52	RM07MLDS	User Logoff
[redacted]	04.07.2011	13:55:06	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	04.07.2011	14:39:57	[redacted]	PMPITLISL09674	VL10X	RVV5OR10C	User Logoff
[redacted]	05.07.2011	20:59:08	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	05.07.2011	20:59:08	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	05.07.2011	21:00:11	[redacted]	PMPITLISL09674	PR05	SAPMP56T	User Logoff
[redacted]	05.07.2011	21:00:17	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPLSMTR_NAVIGATION	User Logoff
[redacted]	07.07.2011	15:07:04	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	07.07.2011	15:30:11	[redacted]	PMPITLISL09674	MB52	RM07MLDS	User Logoff
[redacted]	07.07.2011	19:03:40	[redacted]	PMPITLISL09674	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)

Figura 4.9 – Histórico de acessos do Utilizador B em SAP P1

**Analysis of Security Audit Log**

Period Requested 06.06.2011 09:00:00 - 10.06.2011 23:59:59  
 Period Selected 07.06.2011 13:31:59 - 09.06.2011 09:45:42  
 Server [redacted]  
 User [redacted]  
 Audit Classes Dialog Logon

Server Name	Creation Date	Date/Time	User	Terminal name	Transaction Code	Program	Message Text
[redacted]	07.06.2011	13:31:59	[redacted]	PMPITLISL09207	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	07.06.2011	13:32:03	[redacted]	PMPITLISL09207	SESSION_MANAGER	SAPMSYST	User Logoff
[redacted]	07.06.2011	13:51:29	[redacted]	PMPITLISL09207	MM01	SAPMMG01	User Logoff
[redacted]	09.06.2011	09:42:07	[redacted]	PMPITLISL09207	SESSION_MANAGER	SAPMSYST	Logon Successful (Type=A)
[redacted]	09.06.2011	09:45:42	[redacted]	PMPITLISL09207	MM01	SAPMMG01	User Logoff

Figura 4.10 – Histórico de acessos do Utilizador C em SAP P1

As Figuras 4.8, 4.9 e 4.10 apresentam o histórico dos utilizadores para o sistema. Existindo *logins* no sistema SAP P1, é seguro afirmar que não só têm acesso ao sistema através de IMDL (sistema com código WCP100 equivalente a SAP P1), como também usaram o sistema ao longo do ano. Estes dois factos combinados, asseguram a apresentação das perguntas referentes ao sistema SAP P1 aos utilizadores A, B e C.

### Filtro 3 – Pedidos de Suporte

Nesta fase as aplicações mais importantes deverão estar identificadas, no entanto existem aplicações que o terceiro filtro irá identificar. Através da aplicação SAP Business Objects é extraído um relatório em folha de cálculo com toda a informação relevante sobre os Pedidos de Suporte que o utilizador em questão realizou ao longo do ano (ver Apêndice A).

**Tabela 4.2 – Relatório de Pedidos de Suporte Utilizador A**

Numbe	Customer code	Name	Creation date
6815727		PMPT End User Computing	7/1/2010 11:33
6819896		PMPT Network	8/1/2010 17:03
6824036		iSMS Application for PMPT	11/1/2010 11:45
6923560		PMPT End User Computing	29/01/2010 05:13 PM
6923641		PMPT End User Computing	29/01/2010 06:04 PM
6930034		PMPT End User Computing	2/2/2010 9:44
6949637		PMPT End User Computing	9/2/2010 19:12
6961031		PMPT End User Computing	15/02/2010 09:49 AM
6973316		iSMS Application for PMPT	19/02/2010 11:32 AM
6974031		iSMS Application for PMPT	19/02/2010 03:38 PM
7008703		ITSC Data Warehouse for PMPT (SAMBA)	5/3/2010 16:16
7113643		PMPT End User Computing	19/04/2010 12:09 PM
7130241		PMPT End User Computing	26/04/2010 11:20 AM
7138332		PMPT End User Computing	28/04/2010 04:04 PM
7839129		PMPT End User Computing	7/6/2010 10:08
7839193		PMPT End User Computing	7/6/2010 10:11
8165946		PMPT End User Computing	17/06/2010 09:26 AM
9625608		PMPT Print	20/10/2010 06:51 PM
9637508		PMPT End User Computing	26/10/2010 08:58 AM
9704831		PMPT End User Computing	24/11/2010 01:15 PM

**Tabela 4.3 – Relatório de Pedidos de Suporte Utilizador B**

Numero	Customer code	Name	Creation date
6812443		PMPT End User Computing	6/1/2010 11:15
6812457		PMPT Network	6/1/2010 11:19
6815755		P1EU ERP Application for PMPT	7/1/2010 11:43
6898460		PMPT End User Computing	20/01/2010 11:13 AM
6926694		PMPT Network	1/2/2010 11:22
6968059		STP Application for PMPT	17/02/2010 04:03 PM
6976964		STP Application for PMPT	22/02/2010 11:00 AM
6990351		PMPT End User Computing	26/02/2010 04:40 PM
7007803		PMPT End User Computing	5/3/2010 11:29
7011393		SCIPIO Application for PMI	8/3/2010 11:20
7011409		PMPT End User Computing	8/3/2010 11:24
7046660		P1EU ERP Application for PMPT	22/03/2010 12:30 PM
7046676		ITSC SAP BW for PMIERP (P53)	22/03/2010 12:35 PM
7053887		SCIPIO Application for PMI	24/03/2010 02:57 PM
7068496		STP Application for PMPT	30/03/2010 02:41 PM
7077902		PMPT End User Computing	5/4/2010 9:55
7080541		PMPT End User Computing	6/4/2010 8:54
7096160		PMPT End User Computing	12/4/2010 13:59
7096951		P1EU ERP Application for PMPT	12/4/2010 20:05
7098737		STP Application for PMPT	13/04/2010 09:33 AM
7106891		SCIPIO Application for PMI	15/04/2010 03:27 PM
7114603		PMPT End User Computing	19/04/2010 05:26 PM
7117893		PMPT End User Computing	20/04/2010 06:05 PM
7126471		STP Application for PMPT	23/04/2010 11:18 AM
7187782		P1EU ERP Application for PMPT	20/05/2010 08:45 AM
7189157		PMPT End User Computing	20/05/2010 01:55 PM
7192925		P1EU ERP Application for PMPT	20/05/2010 06:24 PM

**Tabela 4.4 – Relatório de Pedidos de Suporte Utilizador C**

Numero	Customer code	Name	Creation date
6813185		PMPT End User Computing	6/1/2010 15:00
7022320		PMPT End User Computing	11/3/2010 10:58
7063901		GIPS Application for PMI	29/03/2010 10:42 AM
7064417		PMPT End User Computing	29/03/2010 12:49 PM
7065120		PMPT End User Computing	29/03/2010 03:58 PM
7065165		PMPT End User Computing	29/03/2010 04:18 PM
7067085		GIPS Application for PMI	30/03/2010 08:36 AM
7081180		GIPS Application for PMI	6/4/2010 11:34
7092137		PMPT End User Computing	9/4/2010 15:27
7092153		PMPT End User Computing	9/4/2010 15:35
7131530		PMPT End User Computing	26/04/2010 05:56 PM
7175846		PMPT Print	14/05/2010 04:08 PM
7185784		PMPT End User Computing	19/05/2010 01:47 PM
7390149		PMPT End User Computing	24/05/2010 04:26 PM
8343526		PMPT Print	24/06/2010 07:59 AM
8863626		PMPT End User Computing	13/07/2010 08:58 AM
8902233		PMPT Print	14/07/2010 07:32 PM
9462219		PMPT End User Computing	24/08/2010 03:40 PM
9515735		PMPT End User Computing	15/09/2010 05:11 PM
9520186		PMPT End User Computing	17/09/2010 10:57 AM
9523501		PMPT End User Computing	20/09/2010 09:48 AM
9591872		GIPS Application for PMI	6/10/2010 16:15
9613590		PMPT End User Computing	15/10/2010 08:32 AM
9614974		PMPT End User Computing	15/10/2010 05:45 PM
9628500		PMPT End User Computing	21/10/2010 05:14 PM
9694301		PMPT End User Computing	19/11/2010 11:22 AM
9783390		PMPT Print	3/12/2010 11:48

As Tabelas 4.2, 4.3 e 4.4 apresentam o relatório de Pedidos de Suporte onde o utilizador esteve envolvido. Identificam-se aplicações ou serviços que não foram identificados pelo primeiro filtro onde a lista de aplicações e sistemas deverá ser actualizada:

- Utilizador A – iSMS, SAMBA, Print
- Utilizador B – SAP P1, STP, SCIPPIO, SAP BW
- Utilizador C – GIPS, Print

### 4.3.3 Conjunto de Aplicações Identificadas

Após a aplicação dos três filtros, foram identificadas os seguintes sistemas e aplicações:

- Utilizador A – HHP300, HKP100, PCR\_PROD, WBP100, WCP100, MAS, MKS, PDF Convertor, Print, iSMS, SAMBA
- Utilizador B – HHP300, HKP100, HSP100, P53, SCIPPIO, WBP100, WCP100, Ereq, IPSM, PDF Convertor, CMRT, QSMP, PSS4, ISS, STP
- Utilizador C – HHP300, HKP100, HSP100, WBP100, WCP100, Ereq, BV, BIL, MAS, PDF Convertor, MCT, GIPS, GIMS, GSPR, Print

Nem todas as avaliações requeridas no questionário se referem a sistemas/aplicações, existindo perguntas sobre serviços como qualidade do atendimento do *Help Desk*, ou impressoras. O sistema de impressoras mantém um histórico para controlo de custos e de utilização, sendo esta outra fonte de dados a ser utilizada pelo Filtro 2.

Sendo o acesso às impressoras uma funcionalidade que abrange todos os utilizadores da Empresa, o primeiro filtro não será aplicado. Recorrendo ao segundo filtro, é extraído o histórico de utilização (Figura 4.11).

User	Color and B/W pages		Copy and print pages				Price
	B/W	Color	Copy	Print	Duplex	Total	
	198	0	177	21	0	198	0,00 €
	9	0	9	0	0	9	0,00 €
	36	0	0	36	0	36	0,00 €
	3	0	0	3	0	3	0,00 €
	1	0	0	1	0	1	0,00 €
	1	1	2	0	0	2	0,00 €
	33	0	0	33	0	33	0,00 €
	30	0	0	30	0	30	0,00 €
	80	0	0	80	0	80	0,00 €
	2	0	0	2	0	2	0,00 €
	79	0	0	79	0	79	0,00 €
	33	0	0	33	0	33	0,00 €
	2	0	0	2	0	2	0,00 €
	7	0	0	7	0	7	0,00 €
	4	0	0	4	0	4	0,00 €
	3	0	0	3	0	3	0,00 €

Figura 4.11 – Histórico de utilização de impressoras

No histórico estão presentes não só os utilizadores que usam as impressoras, mas também apresenta o número de impressões que efectuou. Esta última informação poderá também ser utilizada para o cálculo do peso das respostas do utilizador.

Após todo o processo de recolha das aplicações e sistemas que os utilizadores usam, a lista que é criada é utilizada para apresentar as perguntas relevantes. O questionário acede ao repositório com todas as perguntas e selecciona as que se referem às aplicações e sistemas presentes na lista do utilizador em questão (Figura 4.12) apresentando-lhes de seguida.



Figura 4.12 – Funcionamento do Endereçamento de Perguntas

## 4.4 Aplicação de Ponderação de Respostas

O exemplo que se apresenta em seguida é composto por duas fases distintas: (1) Definição de critérios e pesos de características dos utilizadores; (2) Aplicação de pesos a uma avaliação de um sistema do questionário de 2010. Na primeira fase do exemplo, o conjunto de dados que são apresentados é real tendo sido obtido pelo *Service Owner* do sistema SAP CRM. No entanto, os dados da segunda fase são **fictícios** pelo facto de não haver forma de identificar os utilizadores que responderam ao questionário do ano anterior. Logo, os resultados obtidos servem apenas para demonstrar a utilidade e o *output* da aplicação da técnica.

### 4.4.1 Características Identificadas

As quatro características enunciadas formam o conjunto que, aos olhos do *Service Owner*, melhor caracterizam um utilizador do sistema SAP CRM:

- Nível de Acesso – Nível de acesso que o utilizador tem para interagir com o sistema
- Experiência – Tempo (contabilizado em número de anos) que o utilizador tem acesso ao sistema
- Taxa de Utilização – Assiduidade com que o utilizador usa o sistema (diariamente, semanalmente, etc.)
- Pedidos Efectuados – Número de Pedidos de Suporte efectuados sobre o sistema

#### 4.4.2 Atribuição de Pesos

Foram identificadas 4 características chave: Nível de Acesso, Experiência, Taxa de Utilização e Pedidos Efectuados. Sabendo que as quatro influenciam a classificação de um utilizador, existem umas mais importantes que outras. Para isso, foi considerado como classificação total – peso final – de cada utilizador um valor real entre 0 e 100 (percentagem), onde 0 implica que o utilizador seja menosprezado e 100 implica que o utilizador é considerado essencial para a análise em questão.

Distribuindo a totalidade dos 100% que constitui o Peso entre as 4 características (Tabela 4.5), formaliza-se a atribuição de peso a cada característica na forma de uma tabela.

Tabela 4.5 – Atribuição de pesos a características

Característica	Peso
Nível de Acesso	30%
Experiência	10%
Taxa de Utilização	40%
Pedidos Concluídos	20%

A escolha dos pesos reflecte o que é mais valorizado para o *Service Owner* do sistema SAP CRM, sendo neste caso o nível de acesso e a taxa de utilização do utilizador.

#### 4.4.3 Peso por Característica

Para calcular o valor de cada característica, recorre-se a um levantamento dos comportamentos e a situação actual de cada utilizador. Para cada característica, atribui-se uma percentagem que irá influenciar os pesos estabelecidos na Tabela 4.5 apresentada no subcapítulo anterior. Cada uma das 4 características é avaliada mediante diferentes critérios.

##### Característica A

Foram identificados dois níveis de acesso fundamentais: Elevado e Baixo. Atribui-se a totalidade do peso ao utilizador que tenha acesso elevado ao sistema e metade do peso ao utilizador que tenha acesso baixo (Tabela 4.6). Ou seja, na prática o utilizador que se encontre na primeira situação terá a totalidade do peso da característica (30% pela Tabela 4.5), caso se encontre na segunda situação, terá 50% do peso ( $30/2 = 15\%$  pela Tabela 4.5).

Tabela 4.6 – Pontuação Característica A

Nível	Pontuação
Elevado	100%
Baixo	50%

### Característica B

A experiência de cada utilizador é contabilizada em número de anos de acesso ao sistema, onde foram definidos 4 intervalos e as respectivas pontuações (Tabela 4.7), atribuindo pontuação máxima a utilizadores com mais de 5 anos de experiência e pontuação mínima a utilização inferior a 1 ano.

**Tabela 4.7 – Pontuação Característica B**

Tempo (anos)	Pontuação
> 5	100%
2 - 5	75%
1 - 2	50%
< 1	25%

### Característica C

A taxa de utilização é medida através da assiduidade com que o utilizador acede ao sistema, podendo tomar qualquer um dos 4 valores distintos com a sua respectiva pontuação (Tabela 4.8). É atribuída pontuação máxima a utilizadores que acedam diariamente ao sistema e mínima a utilizadores que acedem anualmente.

**Tabela 4.8 – Pontuação Característica C**

Assiduidade	Pontuação
Diário	100%
Semanal	50%
Mensal	20%
Anual	0%

### Característica D

A última característica é o número de pedidos efectuados por ano sobre o sistema em causa. Novamente foram criados 5 intervalos com respectivas pontuações (Tabela 4.9).

**Tabela 4.9 – Pontuação Característica D**

Pedidos (p/ ano)	Pontuação
> 40	100%
25 - 40	75%
10 - 25	50%
5 - 10	25%
< 5	10%

#### 4.4.3.1 Exemplo de Aplicação

A apresentação de um exemplo aplicado ao sistema SAP CRM facilitará a compreensão da metodologia. Aplicar-se-á a dois utilizadores onde o utilizador A é experiente e o utilizador B inexperiente em relação ao sistema. Os dados apresentados baseiam-se em informação real referente ao ano de 2010.

##### 1) Características dos Utilizadores

As características dos utilizadores são apresentadas na Tabela 4.10. Levantadas as características, prossegue-se à atribuição dos pesos aos utilizadores.

Tabela 4.10 – Exemplo de características de dois utilizadores

	Utilizador A	Utilizador B
Nível de Acesso	Elevado	Baixo
Experiência	> 5	> 5
Taxa Utilização	Diário	Anual
Pedidos	43	1

##### 2) Pesos Finais

São efectuados os cálculos das variáveis de cada utilizador, obtendo-se o peso final de cada um deles após a soma das 4 variáveis.

- **Utilizador A**

$$\text{Peso\_Utilizador\_A} = [(1*30) + (1*10) + (1*40) + (1*20)] = \underline{\underline{100\%}}$$

- **Utilizador B**

$$\text{Peso\_Utilizador\_B} = [(0.5*30) + (1*10) + (0*40) + (0.1*20)] = \underline{\underline{27\%}}$$

Após o cálculo das ponderações dos utilizadores, é de reparo imediato a diferença entre ambos. No novo questionário, as respostas sobre o sistema SAP CRM do Utilizador A teriam a máxima ponderação, e por isso mesmo seriam encaradas com a máxima importância e relevância. Já as respostas do Utilizador B sobre o sistema irão impactar menos os resultados finais devido à sua baixa ponderação.

#### 4.4.4 Aplicação de Respostas Ponderadas

De forma a apresentar o impacto que a ponderação de respostas introduz face à abordagem clássica, o método foi aplicado às respostas do questionário de 2010. Foi analisado o impacto da ponderação na pergunta relativa ao sistema SAP CRM, um sistema utilizado pela área de Vendas.

#### Resultados *Web Surveyor*

A ferramenta actualmente utilizada na Tabaqueira é o *Web Surveyor*, uma solução de criação, manutenção e análise de questionários *web* (Figura 4.13). Ao publicar o questionário, os utilizadores podem aceder e preencher o mesmo, onde posteriormente as respostas são guardadas pelo *software* para no final analisar as respostas e publicar os resultados.

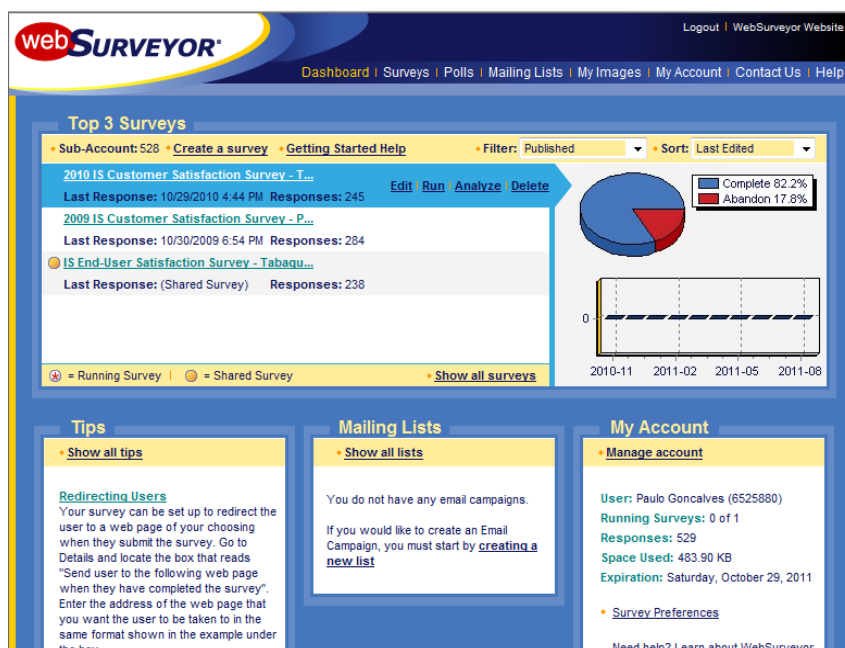


Figura 4.13 – Solução de questionários *web* adoptada pela Tabaqueira

O *Web Surveyor* como solução de criação de questionários, está limitado quanto ao dinamismo que os questionários podem ter. Algumas características dos questionários actuais são:

- Disponibilizado em HTML
- Utilização de *skip-logic*
  - Caso o utilizador responda que não utiliza uma aplicação, as perguntas sobre essa aplicação não são apresentadas
- Repositório de perguntas igual para todos os utilizadores
- Resultados calculados através de métodos simples (média, percentagens, etc.)

O questionário efectuado em 2010 obteve uma taxa de participação na ordem dos 30% (245 respostas), onde se verificou um decréscimo na participação. A pergunta de avaliação do sistema SAP CRM foi respondida por 24 utilizadores de diferentes departamentos.

**Tabela 4.11 – Resultados SAP CRM 2010**

CRM	Very Satisfied	Satisfied	Neutral	Disatisfied	Very Disatisfied	Not Applicable
Ease of use	12,5%	58,3%	16,7%	8,3%	0,0%	4,2%
Functionality	12,5%	58,3%	16,7%	8,3%	0,0%	4,2%
Performance and Availability	8,3%	58,3%	16,7%	12,5%	0,0%	4,2%
Efficiency of IS Support Organization	16,7%	41,7%	25,0%	4,2%	0,0%	12,5%
Frequency of Training	8,3%	33,3%	33,3%	4,2%	0,0%	20,8%
Quality of Training	12,5%	29,2%	33,3%	4,2%	0,0%	20,8%

Na Tabela 4.11 é apresentada a distribuição das respostas dos 24 utilizadores nos diversos índices de satisfação.

### Aplicação de Ponderação de Respostas

Sabendo que não existe forma de identificar os utilizadores que responderam aos questionários dos anos anteriores, é impossível aplicar as técnicas de ponderações idealizadas anteriormente. Foi então criada uma coluna extra com pesos fictícios escolhidos aleatoriamente de forma a demonstrar o funcionamento e os resultados por ele obtidos (Tabela 4.12).

**Tabela 4.12 – Respostas com Pesos Fictícios**

Your position	Business function where you work	SAP CRM : How important is this system for your daily work	SAP CRM : Ease of use	SAP CRM : Functionality	SAP CRM : Performance and availability	SAP CRM : Efficiency of IS support organization	SAP CRM : Frequency of training	SAP CRM : Quality of training	Peso Fictício
Non-managerial position	Sales	Useful, but I can live without it	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	90,0%
Non-managerial position	Sales	Important - I need this system to do my job properly	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	85,0%
Non-managerial position	Sales	Important - I need this system to do my job properly	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	70,0%
Non-managerial position	Sales	Important - I need this system to do my job properly	Satisfied	Satisfied	Satisfied	Satisfied	Neutral	Neutral	90,0%
Non-managerial position	Sales	Useful, but I can live without it	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	70,0%
Manager	Sales	Important - I need this system to do my job properly	Satisfied	Neutral	Neutral	Neutral	Neutral	Neutral	65,0%
Manager	Sales	Useful, but I can live without it	Satisfied	Satisfied	Satisfied	Not applicable	Neutral	Neutral	85,0%
Non-managerial position	Sales	Essential - I cannot work without this tool	Satisfied	Satisfied	Satisfied	Very satisfied	Not applicable	Not applicable	90,0%
Manager	Sales	Important - I need this system to do my job properly	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied	95,0%
Director & above	Sales	Essential - I cannot work without this tool	Neutral	Satisfied	Dissatisfied	Neutral	Neutral	Neutral	100,0%
Contracted / Consultant	Sales	Important - I need this system to do my job properly	Very satisfied	Very satisfied	Satisfied	Very satisfied	Satisfied	Very satisfied	75,0%

Para efectuar a nova análise, aplicou-se um filtro preliminar onde são isoladas apenas as respostas dos utilizadores da área de Sales. O filtro é aplicado assumindo que apenas utilizadores da área de *Sales* (Vendas) utilizam o sistema SAP CRM, informação obtida após discussão com o *Service Owner* do sistema.

Aplicado o filtro preliminar, o número de respostas diminui de 24 para 11 utilizadores e calculou-se novamente os índices de satisfação (Tabela 4.13). A utilização deste filtro tem como objectivo aproximar os resultados do questionário de 2010 ao que seria espectável que se verificasse no questionário de 2011 com o correcto endereçamento das perguntas (no novo questionário, as perguntas serão apenas apresentadas ao utilizadores quando estes de facto usam os sistemas e aplicações).

**Tabela 4.13 – Resultados SAP CRM de Utilizadores de Sales**

SAP CRM	V. S.	S.	N.	D.	V. D.	N/A
Ease of use	9,1%	72,7%	18,2%	0,0%	0,0%	0,0%
Functionality	9,1%	72,7%	18,2%	0,0%	0,0%	0,0%
Performance and Availability	0,0%	72,7%	18,2%	9,1%	0,0%	0,0%
Efficiency of IS Support Organization	18,2%	45,5%	27,3%	0,0%	0,0%	9,1%
Frequency of Training	0,0%	45,5%	45,5%	0,0%	0,0%	9,1%
Quality of Training	9,1%	36,4%	45,5%	0,0%	0,0%	9,1%

Utilizando os pesos definidos anteriormente para os 11 utilizadores que responderam à pergunta, foram calculados os novos índices de satisfação através da técnica definida nos capítulos anteriores (Tabela 4.14).

**Tabela 4.14 – Resultados SAP CRM com Ponderação de Respostas**

SAP CRM	V. S.	S.	N.	D.	V. D.	N/A
Ease of use	8,2%	73,2%	18,6%	0,0%	0,0%	0,0%
Functionality	8,2%	77,0%	14,8%	0,0%	0,0%	0,0%
Performance and Availability	0,0%	74,3%	14,8%	10,9%	0,0%	0,0%
Efficiency of IS Support Organization	18,0%	47,0%	25,7%	0,0%	0,0%	9,3%
Frequency of Training	0,0%	45,4%	44,8%	0,0%	0,0%	9,8%
Quality of Training	8,2%	37,2%	44,8%	0,0%	0,0%	9,8%

Os índices de satisfação dos utilizadores do sistema SAP CRM após aplicação do filtro preliminar e da técnica de ponderação são apresentados na Figura 4.14.

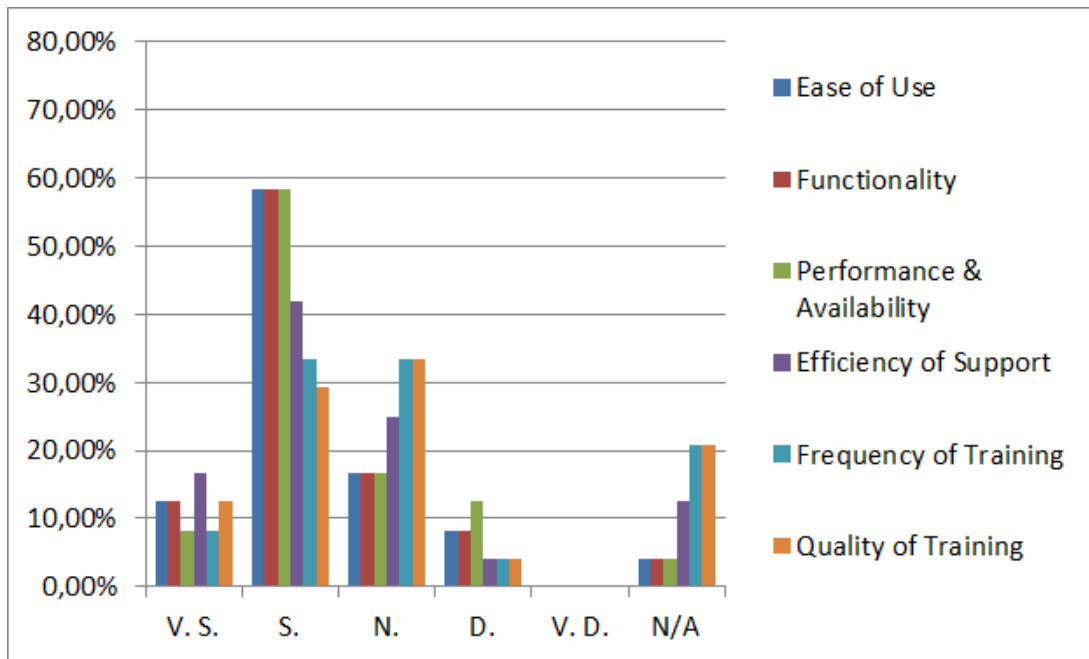


Figura 4.14 – Índices Finais de Satisfação SAP CRM

Os valores apresentados na Tabela 4.14 comparativamente aos da Tabela 4.11 sofreram profundas alterações. Tendo em conta o objectivo de realizar o questionário, após a aplicação do método, a avaliação do sistema SAP CRM demonstra ser mais positiva do que publicado em 2010. Houve ajuste nos índices de satisfação destacando-se o aumento na ordem dos 20% na Satisfação (codificado como “S.” nas tabelas) e o decréscimo na ordem dos 8% no Insatisfeito (codificado como “D.” nas tabelas).

## 4.5 Criação de Regras de Associação

O algoritmo modificado *Apriori* foi programado em MATLAB e testado com os dados disponibilizados pela empresa. A implementação foi efectuada com sucesso e foram geradas regras do tipo  $A \Rightarrow \neg B$  tal como idealizado. Como *input* do programa, utilizou-se a folha de cálculo com as respostas do questionário de 2010.

### 4.5.1 Descrição dos Dados

A folha de cálculo utilizada contém uma primeira linha de texto livre onde é apresentada a pergunta que foi efectuada. Cada linha contém as respostas de cada utilizador em campos multi-valor.

Em 2010 foram recebidas 245 respostas ao questionário, implicando que o ficheiro seja composto por 246 linhas (1 de *labels* + 245 de respostas). O ficheiro contém 92 colunas de informação, onde não só estão contidas as perguntas mas também informação sobre as datas recebidas, a língua recebida, etc.

#### 4.5.2 Pré-Tratamento dos Dados

Antes de alimentar a folha de cálculo ao algoritmo para este processar, é necessário efectuar um tratamento dos dados. A fase é de extrema importância pelo facto do algoritmo esperar uma folha de cálculo num formato pré-determinado.

O primeiro passo é a eliminação dos campos que não contêm informação relevante para a extracção de regras.

- Campos que solicitam comentários de texto livre ao utilizador são descartados
- Campos com informação irrelevante são retirados
  - Localização
  - Data recebida
  - Língua
  - Número de IP
  - *Email*

Após a eliminação dos campos irrelevantes, a folha passou a ser composta por 72 colunas e as mesmas 246 linhas (Figura 4.15).

	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	
1	Please rate the desk-side support group who assist you in yo...	Please rate the desk-side support group who assist you in yo...	Please rate the desk-side support group who assist you in yo...	Please rate the desk-side support group who assist you in yo...	Please rate the desk-side support group who assist you in yo...	SAP P1: How important is this system for your daily use?	SAP P1: Ease of use	SAP P1: Functionality	SAP P1: Performance and availability	SAP P1: Efficiency of IS organization	SAP P1: Frequency of training	SAP P1: Quality of training	ISMS: How important is this system for your daily work?	ISMS: Ease of use	ISMS: Functionality	ISMS: Performance and availability	ISMS: Efficiency of IS organization	ISMS: Frequency of training	ISMS: Quality of training	SAP BW: How important is this system for your daily use?	SAP BW: Ease of use	
2	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Useful, but I can live without it	Disatisfied	Satisfied	Neutral	Very satisfied	Neutral	Neutral	Not at all important - I never use it							Useful, but I can live without it	Disatisfied	
3	Very satisfied	Satisfied	Very satisfied	Very satisfied	Very satisfied	Useful, but I can live without it	Satisfied	Satisfied	Satisfied	Satisfied	Very satisfied	Not applicable	Not at all important - I never use it								Not at all important - I never use it	
4	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Useful, but I can live without it	Satisfied	Satisfied	Disatisfied	Disatisfied	Satisfied	Very satisfied	Not at all important - I never use it								Not at all important - I never use it	
5	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Useful, but I can live without it	Neutral	Satisfied	Satisfied	Very satisfied	Very satisfied	Very satisfied	Not at all important - I never use it								Useful, but I can live without it	Satisfied
6	Neutral	Neutral	Neutral	Neutral	Neutral	Important - I need this system to do my job properly	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Important - I need this system to do my job properly	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Very satisfied	Neutral	

Figura 4.15 – Folha de Cálculo com Respostas

Um dos requisitos para o correcto funcionamento do algoritmo é a forma como os campos estão formatados. Toda a informação que irá ser tratada e investigada de forma a encontrar relações tem de estar no formato binário (0 e 1). Foi então efectuado um tratamento utilizando a ferramenta *RapidMiner* que transforma os dados para a forma desejada (ver Apêndice C).

	FD	FE	FF	FG	FH	FI	FJ	FK	FL
1	ISMS : Eas	ISMS : Eas	ISMS : Ease of use = Satisfied	ISMS : Ease of use = Not applicable	ISMS : Ease of use = Dissatisfied	ISMS : Ease of use = Very dissatisfied	ISMS : Functionality = Very satisfied	ISMS : Fun	ISMS :
2	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
6	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE
7	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
8	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
9	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
10	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
11	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
12	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
13	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
14	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
15	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
16	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
17	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
18	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
19	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
20	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
21	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
22	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
23	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
24	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
25	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Figura 4.16 – Folha de Cálculo com Respostas Tratadas

A Figura 4.16 apresenta o novo formato da folha de cálculo com as respostas ao questionário. O tamanho da folha foi actualizado destacando-se o aumento significativo do número de colunas (devida à fragmentação de cada perguntas nas diversas respostas possíveis). A folha passou a ter 399 colunas de informação e as mesmas 246 linhas, e deste modo está no formato para que o algoritmo possa interpretar.

### 4.5.3 Utilização do Algoritmo

Para testar e extrair regras de associação, foi utilizada uma taxa de suporte de 50% e uma taxa de confiança de 97%. Com estes parâmetros e os resultados do questionário de 2010, foram extraídas 1922 regras.

1	2
605 SAP P1 : Functionality = Satisfied	Please rate the desk-side support group who assist you in yo...: Courtesy of the technician? = D...
606 SAP P1 : Functionality = Satisfied	Please rate the desk-side support group who assist you in yo...: Timeliness of the service provid...
607 SAP P1 : Functionality = Satisfied	Please rate the desk-side support group who assist you in yo...: Quality of the service provided?...
608 SAP P1 : Functionality = Satisfied	Please rate the desk-side support group who assist you in yo...: Overall desk-side support servic...
609 SAP P1 : Functionality = Satisfied	SAP P1&nbsp;How important is this system for your daily work... = Not at all important - I neve...
610 SAP P1 : Functionality = Satisfied	SAP P1 : Ease of use = Very dissatisfied
611 SAP P1 : Functionality = Satisfied	SAP P1 : Ease of use = Not applicable
612 SAP P1 : Functionality = Satisfied	SAP P1 : Functionality = Neutral
613 SAP P1 : Functionality = Satisfied	SAP P1 : Functionality = Dissatisfied
614 SAP P1 : Functionality = Satisfied	SAP P1 : Functionality = Not applicable
615 SAP P1 : Functionality = Satisfied	SAP P1 : Functionality = Very satisfied
616 SAP P1 : Functionality = Satisfied	SAP P1 : Functionality = Very dissatisfied
617 SAP P1 : Functionality = Satisfied	SAP P1 : Performance and availability = Not applicable
618 SAP P1 : Functionality = Satisfied	SAP P1 : Efficiency of IS support organization = Very dissatisfied
619 SAP P1 : Functionality = Satisfied	iSMS : Ease of use = Very dissatisfied
620 SAP P1 : Functionality = Satisfied	iSMS : Functionality = Very dissatisfied
621 SAP P1 : Functionality = Satisfied	iSMS : Performance and availability = Very dissatisfied
622 SAP P1 : Functionality = Satisfied	iSMS : Efficiency of IS support organization = Not applicable
623 SAP P1 : Functionality = Satisfied	iSMS : Quality of training = Very dissatisfied

Figura 4.17 – Regras Extraídas pelo Algoritmo

A Figura 4.17 apresenta os resultados obtidos no MATLAB onde na coluna 1 estão as premissas das regras (Ex: **Se** “SAP P1: *Functionality = Satisfied*”) e na coluna 2 as conclusões das regras (Ex: **então não** “SAP P1: *Performance and availability = Not applicable*”).

Analisando o conjunto de regras, identificam-se regras interessantes que podem ser aplicáveis com sucesso. Note-se por exemplo as seguintes:

- **Se** “SAP P1: *Functionality = Satisfied*” **então não** “SAP P1: *Performance and availability = Not applicable*”

Note-se a importância desta regra na verificação de inconsistência de respostas facultadas. No momento que o utilizador classifica o sistema SAP P1 numa característica, este deveria estar em condições de responder questões de “Performance and availability”

- **Se** “*Please rate your IS Help Desk Telephone: Overall Help Desk services? = Satisfied*” **então não** “*Please rate your IS Help Desk Telephone: Quality of the service provided? = Very dissatisfied*”

Mais uma vez, a regra criada averigua inconsistência nas respostas. Será estranho, ou pouco provável, um indivíduo responder que está satisfeito no geral com os serviços do *Help Desk* mas responder que está muito insatisfeito com a qualidade do serviço.

Extraídas as regras, é efectuada uma verificação das respostas e são anotadas as regras que são desrespeitadas e número de vezes que foram desrespeitadas (Figura 4.18). Na análise do ano 2010 com a aplicação das regras extraídas do mesmo questionário, foram registadas 3090 violações de regras.

	1	2
1448	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Other
1449	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Other
1450	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Other
1451	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = General Services (Admin)
1452	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = General Services (Admin)
1453	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1454	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1455	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1456	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1457	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1458	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Human Resources
1459	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Law
1460	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Corporate Affairs
1461	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Corporate Affairs
1462	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = R&D
1463	SAP CRM&nbsp;How important is this system for yo...	Business function where you work = Brand Integrity & Compliance
1464	SAP CRM&nbsp;How important is this system for yo...	How satisfied are you with the following aspects of your wor... : Desktop computer = Diss...
1465	SAP CRM&nbsp;How important is this system for yo...	How satisfied are you with the following aspects of your wor... : Desktop computer = Diss...

Figura 4.18 – Regras Violadas no Questionário de 2010

Utilizando o número de infracções, conclui-se que cada utilizador desrespeitou em média aproximadamente 13 regras ( $3090/245 = 12,61$  regras). Este valor médio poderá ser utilizado como critério de avaliação do questionário estabelecendo, por exemplo, um limite de 8 infracções por questionário recebido.

Sendo um processo automático baseado inteiramente em cálculos estatísticos, nenhuma acção precipitada deverá ser tomada. A técnica identifica questionários que contêm respostas aparentemente fora do normal mas peca pela inexistência de uma análise subjectiva. A decisão de descartar ou não o questionário ficaria à responsabilidade de uma equipa especializada, que ao analisar cada questionário poderá encontrar razões para as respostas dadas.

# Capítulo 5

## Conclusões

### 5.1 Discussão

Ao longo da dissertação foram abordados problemas existentes nos questionários administrados actualmente, tendo sido propostas técnicas e metodologias que visam resolver esses problemas. Face aos objectivos propostos inicialmente, em grande medida foram cumpridos. Foi criada uma ferramenta para avaliar objectivamente a qualidade de questionários, algo que até ao momento era uma tarefa subjectiva com conflito de critérios. Foram propostas técnicas que introduzissem um novo nível de dinamismo tanto a nível de endereçamento de perguntas aos utilizadores mas também no momento de análise de respostas.

#### Impactos e Contributos

Num ambiente empresarial, o *feedback* negativo sobre um determinado sistema/aplicação, pode ser motivo para alocar recursos e tempo ao desenvolvimento de soluções e programas para combater esse problema identificado no questionário. No entanto, pode acontecer os questionários estarem de facto a exagerar a gravidade do problemas, e dessa forma, os recursos estarão a ser alocados em algo com prioridade provavelmente menor que outros problemas e actividades, onde na realidade existe necessidade de melhoria. A técnica de endereçamento de perguntas proposta nesta dissertação aborda este problema ao captar apenas o *feedback* dos reais utilizadores das aplicações, contribuindo para melhorar a captação dos reais índices de satisfação dos colaboradores.

Ainda em relação à melhoria da qualidade dos dados gerados pelo questionário, a criação de regras de associação entre as respostas (aprendidas de forma automática com base no histórico dos inquéritos) tem a capacidade para, também de forma automática, encontrar respostas dadas pelos utilizadores que muito provavelmente são impossíveis face às restantes respostas que o mesmo utilizador deu. Esta pode ser uma contribuição para a detecção de respostas com elevada probabilidade de não reflectir a realidade (com origem, por exemplo, em lapsos, má compreensão da pergunta, respostas propositadamente enganosas, ou preenchimentos aleatórios do questionário).

As flutuações dos índices de satisfação alteram a avaliação final do sistema e conclusões retiradas do questionário. No capítulo dos Resultados, é de chamar à atenção o decréscimo de

8% no índice Insatisfeito para 0% após aplicação da metodologia. Este diferencial pode revelar-se crucial a nível de negócio e gestão de custos da Empresa no sentido em que recursos poderão ter sido alocados na tentativa de melhorar o sistema onde na realidade nunca necessitou de tal melhoria. É um impacto nos custos e alocação de recursos que tornam o processo de endereçamento de perguntas e ponderações tão importantes.

## **Ética**

A aplicação da metodologia utiliza, nalguns casos, dados sensíveis e informações que à partida retiram o conceito de anonimato. No sentido de clarificar a aparente ausência de anonimato durante o preenchimento do questionário, os dados sensíveis que são utilizados pelas diversas técnicas são apenas utilizados para o correcto funcionamento durante o próprio preenchimento, sendo os dados de identificação descartados após o preenchimento do questionário. Desta forma, não existe forma de identificar o utilizador e os resultados finais permanecerão anónimos podendo ser publicados sem qualquer restrição.

Na atribuição de ponderações aos utilizadores, nomeadamente na valorização das características que a empresa utiliza para classificar utilizadores, esses valores influenciam os índices de satisfação do questionário. Esta valorização das características e o peso que lhes é atribuído, podem afectar os resultados finais dando a possibilidade de alterá-los de forma a deturpar os resultados e minimizar ou maximizar os resultados face à necessidade e desejo da empresa. É então importante que à partida se estabeleçam os pesos das características e não alterar durante ou após o questionário. É uma questão de ética e é um ponto extremamente importante que deve ser cumprido e respeitado de forma a garantir resultados fidedignos e não deturpados.

## **5.2 Limitações**

### **Endereçamento de Perguntas**

Na metodologia foi proposta uma técnica de endereçamento de perguntas baseada na aplicação de filtros. Esta aplicação tem como objectivo a identificação dos sistemas/aplicações e serviços que o utilizador usou durante o ano em questão. Na aplicação da técnica foram identificadas limitações na abordagem, nomeadamente em cada um dos filtros.

No primeiro filtro (acessos de utilizadores), nem todos os sistemas estão inseridos na aplicação central que gere os acessos às mesmas (no caso da Tabaqueira é utilizado o sistema de acessos denominado IMDL que apresenta esta limitação).

No segundo filtro (histórico de acessos), nem todas as aplicações guardam históricos de *logins* e transacções (no caso da Tabaqueira os únicos sistemas que guardam o histórico por omissão são os sistemas SAP).

No terceiro filtro (históricos de pedidos de suporte), um Pedido de Suporte implica que haja um problema ou dúvida a ser esclarecida, no entanto não existe qualquer obrigatoriedade de um utilizador que usa uma aplicação de abrir um Pedido de Suporte durante a sua actividade (no caso da Tabaqueira é utilizado o *Service Desk* para registar e gerir Pedidos de Suporte).

Estas limitações acarretam consequência a nível de eficácia da técnica. Devido às limitações de cada filtro, existe a possibilidade de não serem identificadas todas as aplicações que o utilizador usou, resultando na apresentação de um questionário incompleto.

### **Regras de Associação**

O programa criado em MATLAB exige um conjunto de condições para poder ser aplicado, respeitantes às diferenças nas perguntas e forma como são colocadas as perguntas.

Sendo as *labels* (que armazenam o texto da pergunta) texto livre, qualquer alteração na pergunta entre questionários de anos diferentes, impossibilita a utilização das regras criadas, simplesmente porque a pergunta, tendo outro texto, passa a ser outra e a regras aprendida deixa de poder ser aplicada.

Na aplicação do algoritmo implementado, no meio do grande número de regras geradas surgem algumas cuja aplicação poderá não fazer sentido. Este fenómeno acontece porque nesta dissertação, tendo aplicado a metodologia para ilustrar o processo, não foi possível aplica-la em pleno porque havendo apenas dados respeitantes a um ano de aprendizagem, essas regras acabariam por encontrar contra-exemplos noutros anos. Com apenas um ano de inquéritos para aprender, o aparecimento de regras com falta de sentido é algo que já seria à partida expectável. Apenas com uma amostra maior é possível obter um conjunto de regras que possam ser utilizadas sem este fenómeno acontecer (ou reduzindo-o a um mínimo admissível).

Este problema seria resolvido com fases de treino e teste bem estruturadas. Para treinar devidamente esta técnica de aprendizagem, pode por exemplo, considerar-se uma amostra de 10 anos:

- 2000 – 2006

Utilização de 6 folhas de cálculo com respostas aos questionários onde o algoritmo será treinado e um conjunto de regras é criado.

- 2007 – 2010  
Fase onde o algoritmo é testado através da aplicação das regras inferidas dos questionários anteriores medindo a eficácia e o funcionamento do algoritmo.
- 2011 (presente)  
Aplicação definitiva ao questionário do ano corrente utilizando um conjunto de regras devidamente treinadas e testadas em exemplos anteriores.

### **Ponderação de Respostas**

A aplicação deste método exige que, durante o preenchimento do questionário, o sistema saiba identificar os utilizadores. Como esta informação é descartada no fim, não é possível aplicar esta parte da metodologia retroactivamente, nomeadamente para melhorar questionários passados. Esta técnica foi apenas desenhada em teoria e caracteriza-se por ser uma solução apenas implementável como funcionalidade do novo questionário.

### **5.3 Trabalho Futuro**

Esta dissertação comprometeu-se a criar a base de uma nova forma de criar questionários, nomeadamente através da criação de uma metodologia a aplicar. Todos os testes e resultados expostos ao longo do documento, foram efectuados e testados em ambiente de desenvolvimento para fins de investigação, sendo alguns resultados apresentados apenas como exemplos explicativos.

Como trabalho futuro, a implementação de um questionário piloto que incorpore as soluções aqui propostas será o passo natural. Esse questionário-piloto permitirá uma análise custo-benefício, e finalmente à decisão sobre a aplicação ou não da metodologia apresentada.

## Referências Bibliográficas

Acuna, E., Rodriguez, C. (2004) *The Treatment of Missing Values and its Effect in the Classifier Accuracy*. Banks, D. (Ed.), et al. *Classification, Clustering and Data Mining Applications*, Berlin, Heidelberg Springer-Verlag, pp. 639–648

Agrawal, R., Imielinski, T., Swami, A. (1993) *Mining Association Rules Between Sets of Items in Large Databases*, SIGMOD Conference pp. 207-216

Baker, R., Crawford, S., Swineheart, J. (2004) *Development and Testing of Web Questionnaires*. Presser, S. (Ed.) et al. *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley, pp. 361-384

Bethke, A. D. (2008) *Representing Procedural Logic in XML*. *Jornal of Software*, vol. 3, no. 2. pp: 33 - 40

Berry, M. J. A., G. Linoff (2000) *Mastering Data Mining: The Art and Science of Customer Relationship Management*. John Wiley and Sons, Inc.

Biemer, Robert M. Groves, Lars E. Lyberg, Naney A. Mathiowetz and Seymour Sudman (1991) *Measurement Errors in Surveys*. New York; Wiley

Buuren, S., Oudshoorn, C.G.M. (1999) *Flexible Multivariate Imputation by MICE*, Leiden: TNO Preventie en Gezondheid, TNO/VGZ/PG 99.045

Chandola, V., Banerjee A., Kumar V. (2007) *Outlier detection: a survey*, Technical Report. Univeristy of Minnesota, US

Chiu, S., Tavella, D. (2008) *Data Mining and Market Intelligence for Optimal Marketing Returns*. Elsevier, Amsterdam

Converse JM. (1987) *Survey Research in the United States: Roots and Emergence 1890.1960*. Berkeley, Los Angeles: Univ. Calif. Press

- Creswell, J. (1993) *Research Design – Qualitative and quantitative approaches*. Thousand Oaks, Sage Publications
- De Leeuw, E.D. (1992) *Data Quality in Mail, Telephone, and Face to Face Surveys*. Amsterdam: TT-publikaties
- De Maast, F. (1997) *Over Statistieken, Surveys en het CBS: Een Basispakket van Begrippen, Feiten en Namen*. Intern CBS rapport
- Farhangfar, A., Kurgan, L., Dy, J. (2008) *Impact of Imputation of Missing Values on Classification Error for Discrete Data*. Pattern Recognition 41: 3692-3705
- Fayyad, U., G. Piatetsky-Shapiro, *et al.*, Eds. (1996) *Advances in Knowledge Discovery and Data Mining*. Massachusetts, The MIT Press
- Fayyad, U., Uthurusamy R. (2002) Evolving Data Mining into Solutions for Insights, Communications of the ACM, Vol. 45, No. 8, pp. 28-31
- Fowler, F. (1988) *Survey research methods*. Newbure Park, Sage
- Fowler, F. (1995) *Improving Survey Questions – Design and Evaluation*, Sage Publications, Thousand Oaks
- Groves, R. M., Cialdini, R. B., Couper, M. P. (1992) *Understanding the Decision to Participate in a Survey*. *Public Opinion Quarterly* Volume 56: 475-495
- Grzymala-Busse, J.W., Hu, M. (2000) *A Comparison of Several Approaches to Missing Attribute Values in Data Mining*. *RSCTC'2000*, pages 340-347
- Han, J., Kamber, M. (2006) *Data Mining: Concepts and Techniques*. Elsevier
- Heer, de Wim, (1999) *Methodological Issues in Survey Research: A Historical Review*. *Bulletin de Méthodologie Sociologique*, 64: 25-48
- Hodge, V.J., Austin, J. (2004) *A Survey of Outlier Detection Methodologies*. *Artificial Intelligence Review* 22: 85-126.

- Japkowicz, N., Myers, C. & Gluck M. A. (1995) *A Novelty Detection Approach to Classification*. Proceedings of the 14th International Conference on Artificial Intelligence (IJCAI-95), 518–523
- Krosnick, J.A., Berent, M.K. (1993) *Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format*. Am. J. Polit. Sci. 37: 941-64
- Krosnick, J.A. (1999) *Survey Research*. Annu. Rev. Psychol. 50: 537-567
- Markou, M., Singh, S. (2003) *Novelty Detection: A Review-part 2: Neural Network Based Approaches*. Signal Processing 83, 12, 2499-2521
- Mundfrom, D.J, Whitcomb, A. (1998) *Imputing Missing Values: The Effect on the Accuracy of Classification*. Multiple Linear Regression Viewpoints 25(1): 13-19
- Necco, C.R., Gordon, C.L., and Tsai, N.W. (1987) *Systems Analysis and Design: Current Practices*. MIS Quarterly, December.
- Nute, D., and Cross, C. (2002) *Conditional logic*. Handbook of Philosophical Logic, vol. 4. Kluwer Academic Publishers, second edition. 1–98.
- Rousseeuw, P.J., Leroy, A.M. (1987) *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc., New York, NY, USA
- Santos, M. Y., Ramos, I. (2009) *Business Intelligence : Tecnologias da Informação na Gestão de Conhecimento*. 2ª ed., atualizada e aumentada. Lisboa : FCA - Editora de Informática
- Smith, T.W. (1987) *That which we call welfare by any other name would smell sweeter: an analysis of the impact of question wording on response patterns*. Public Opin. Q. 51: 75-83
- Sproull, L.S. (1986) *Using Electronic Mail for Data Collection in Organizational Research*. Academy of Management Journal, 29(1), 159-169

Synodinos, N.E., Papacostas, C. S., & Okimoto, G. M. (1994) *Computer-administered vs. paper-and-pencil surveys and the effect of sample selection*. Behavior Research Methods, Instruments, and Computers, 26(4): 395-401

Schwarz, N., Hippler, H.J., Deutsch, B., Strack, F. (1985) *Response Scales: Effects of Category Range on Reported Behavior and Subsequent Judgments*. Public Opin. W. 49: 388-95

Teng, H., Chen, K., Lu, S. (1990) *Adaptive Real-Time Anomaly Detection Using Inductively Generated Sequential Patterns*. Proceedings of IEEE Computer Symposium on Research in Security and Privacy. IEEE Computer Society Press, 278-84

van Schoor, H. and Konz, S., (1996) *Males/females: An Anthropometric Comparison for Modeling Missing Data*. International Journal of Industrial Ergonomics, 17, 437-440

Walonick, David S., (1997) *Survival Statistics*. Bloomington, MN, StatPac, Inc.

Witten, I.H., Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques* (2<sup>nd</sup> ed, Morgan-Kaufman Series of Data Management Systems). San Francisco: Elsevier

Young, W. , Weckman, G. and Holland, W.(2011) *A survey of methodologies for the treatment of missing values within datasets: limitations and benefits*, Theoretical Issues in Ergonomics Science, 12: 1, 15 — 43, First published on: 15 June 2010 (iFirst)

# Apêndice A

## Elaboração de Relatórios em SAP Business Objects

A exploração dos dados é feita através do sistema SAP Business Objects, onde é possível explorar os dados e criar relatórios de análise segundo diversas vistas. Temos então a oportunidade de explorar a *Data Warehouse* segundo as dimensões que pretendermos. Analisemos agora o processo de criação de um *Report* nesta ferramenta.

Ao criarmos um novo *Report*, é necessário escolher o acesso aos dados. No diagrama anterior, foi mencionado o Universo. O Universo é o conjunto de diversos *Data Marts* (subconjunto de dados do *Data Warehouse*) já criados, que estão disponíveis para analisar um determinado assunto. Entre os *Data Marts* disponíveis encontram-se:

- Alarm
- Change
- Project
- Root Cause Analysis
- Service Level Management
- Support Request

Na Figura A.1, podemos visualizar os diversos Universos à disposição.

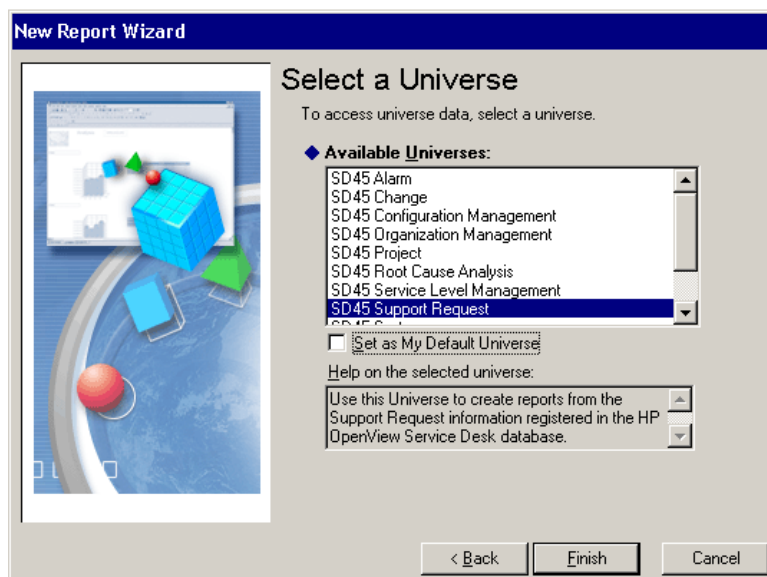


Figura A.1 – Seleção de Universos

Para compreendermos as capacidades da ferramenta, escolhemos um que nos convenha. Para esta explicação, consideremos como objectivo de exploração a recolha de informação sobre todos os Support Request registados em 2011.

Neste momento, estamos perante um sub-Data Warehouse onde no centro temos os Support Requests que podemos explorar segundo diversas dimensões.

---

### **Definição**

Support Request (SR) – é um pedido proveniente de um utilizador que pretende suporte do prestador de serviços da organização.

---

Um SR tem quatro categorias:

- Request for Change – pedidos de utilizadores para adicionar, modificar, mover, ou remover *hardware*, software, ou dados
- Request for Incident Resolution – pedidos de utilizadores para restaurar um serviço que não está a funcionar devidamente
- Request for Information – pedidos de utilizadores para obterem respostas sobre um determinado serviço
- Request for Support Improvement – pedidos de melhoria na forma como é prestado o suporte

Cada entrada em sistema de um SR tem de estar acompanhado de diversas informações: nome do utilizador, categoria do pedido, o serviço, o item, etc. No BO, é possível analisar todos os detalhes de todos os SR registados. Na Figura A.2 podemos ver os detalhes disponíveis.

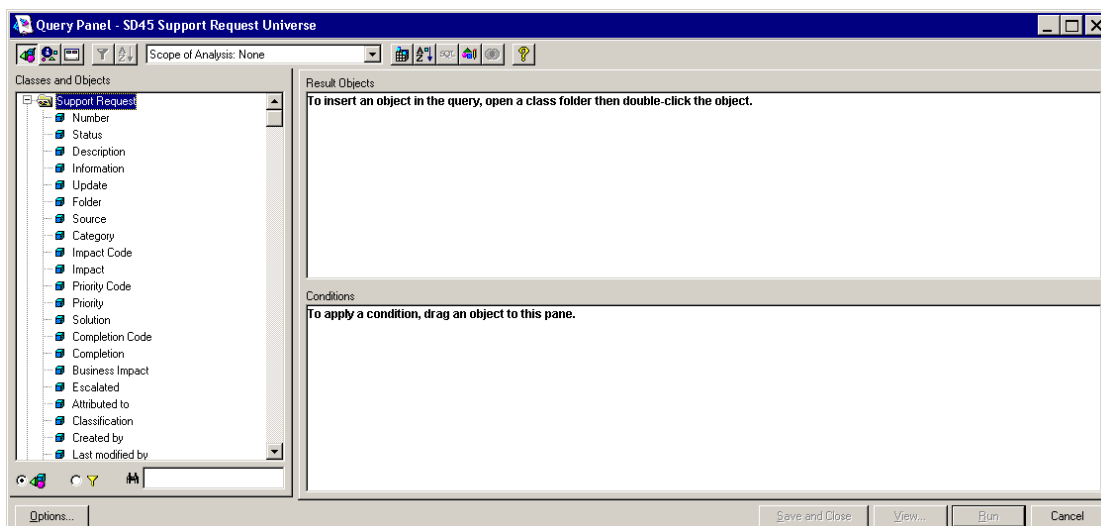


Figura A.2 – Informações Disponíveis

Relembrando que queremos obter informações sobre os SR da Tabaqueira em 2011, vamos definir a informação que pretendemos saber. Iremos analisar a informação segundo a Categoria, Nº do pedido, Descrição, Estado e Prioridade (Figura A.3).

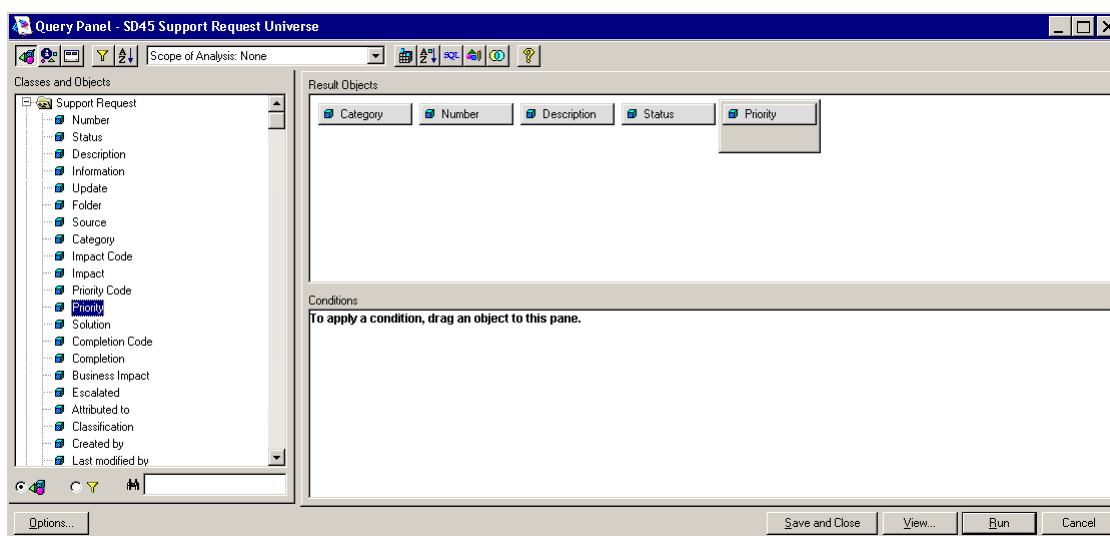


Figura A.3 – Informações para Criação de Relatório

No entanto, não é suficiente seleccionar as dimensões de análise. É necessário aplicar condições aos dados pelo facto destes serem referentes a todas as Afiliadas no mundo. Para isso, aplicamos a condição “Folder Equal to PMPT” (**Philip Morris PorTugal**) que filtra apenas os SR registados na Tabaqueira. Para o fazer, procura-se a classe “Folder” da lista existente e por fim arrasta-se para a janela “Conditions”. Para garantir que os resultados são de 2011, aplicamos uma segunda condição para restringir os SR criados em 2011. Para isso, utilizamos a condição “Creation year Equal to 2011”. Podemos visualizar o final do processo na Figura A.4.

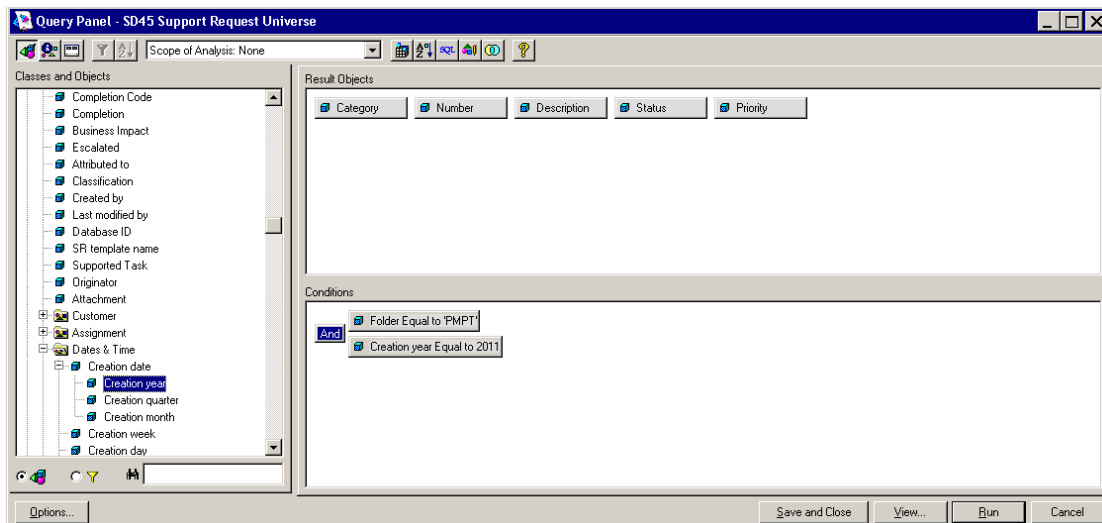


Figura A.4 – Condições Aplicadas à Informação

Após termos definido todas as condições que necessitamos, carregamos no botão “Run” e o relatório é executado. Na *taskbar* do lado esquerdo temos os vários atributos que escolhemos para explorar os SR que basta arrastar para a área principal do ecrã para termos acesso a eles. Na Figura A.5 podemos ver o relatório criado e a informação que pretendíamos.

Category	Number	Description	Status	Priority
Request for Change	10062708	Daily Backup procedures 31/12/201	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063009	Request to transfer AD Account	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063225	Daily Backup procedures 03/01/201	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063268	Request to Loan Hardware	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063441	Request to publish screensaver	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063561	Request to add FA Brand to FQA in	Closed	P120 - To Be Completed within 120 Service Hours
Request for Change	10063619	Request to configure user card	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063634	Request to give sfp access	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063641	Request to configure user card	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063651	Request to Remove Account - CH	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10063762	OSMP Tabaqueira (Albarraque) - T	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064123	OTC Repetitive Request - Define C	Closed	P160 - To Be Completed within 160 Service Hours
Request for Change	10064132	Request to send Message to all use	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064348	ISMS2GIS interface	Passed to Change	P40 - To Be Completed within 40 Service Hours
Request for Change	10064433	Request to do MES Changes - Urge	Closed	P120 - To Be Completed within 120 Service Hours
Request for Change	10064448	Request to do SAP changes	Closed	P160 - To Be Completed within 160 Service Hours
Request for Change	10064540	Request to do SAP changes	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064566	Request to Remove Hardware - Ch	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064618	Request to solve Issue with Workst	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064619	Request to send Message to all use	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064623	Request to solve Issue with Workst	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064645	Request to Remove Account - CH	Closed	P40 - To Be Completed within 40 Service Hours
Request for Change	10064701	Request to send Message to all use	Closed	P40 - To Be Completed within 40 Service Hours

Figura A.5 – Relatório Criado

Podemos escolher a forma como vemos os dados agrupando segundo uma variável (*Group By*) para que possamos entender melhor a informação. Para isso, basta seleccionar uma instância da coluna da variável que desejamos agrupar e arrastar para fora da tabela. A coluna desaparece da tabela e o atributo funciona agora como *Group By*. O output é o apresentado na figura A.6.

**Support Requests 2011**

**Request for Change**

Number	Description	Status	Priority
10062706	Daily Backup procedures 31/12/2011	Closed	P40 - To Be Completed within 40 Service Hours
10063009	Request to transfer AD Account	Closed	P40 - To Be Completed within 40 Service Hours
10063225	Daily Backup procedures 03/01/2011	Closed	P40 - To Be Completed within 40 Service Hours
10063288	Request to Loan Hardware	Closed	P40 - To Be Completed within 40 Service Hours
10063441	Request to publish screensaver	Closed	P40 - To Be Completed within 40 Service Hours
10063551	Request to add FA Brand to FGA in	Closed	P120 - To Be Completed within 120 Service Hours
10063619	Request to configure user card	Closed	P40 - To Be Completed within 40 Service Hours
10063634	Request to give stfp access	Closed	P40 - To Be Completed within 40 Service Hours
10063641	Request to configure user card	Closed	P40 - To Be Completed within 40 Service Hours
10063651	Request to Remove Account - CH: 7	Closed	P40 - To Be Completed within 40 Service Hours
10063762	OSMP Tabaqueira (Albaraque) - T	Closed	P40 - To Be Completed within 40 Service Hours
10064123	OTC Repetitive Request - Define Ca	Closed	P160 - To Be Completed within 160 Service Hours
10064132	Request to send Message to all use	Closed	P40 - To Be Completed within 40 Service Hours
10064348	SMS2GIS interface	Passed to Change	P40 - To Be Completed within 40 Service Hours
10064433	Request to do MES Changes - Urge	Closed	P120 - To Be Completed within 120 Service Hours
10064448	Request to do SAP changes	Closed	P160 - To Be Completed within 160 Service Hours
10064540	Request to do SAP changes	Closed	P40 - To Be Completed within 40 Service Hours
10064566	Request to Remove Hardware - Ch	Closed	P40 - To Be Completed within 40 Service Hours
10064618	Request to solve Issue with Worksta	Closed	P40 - To Be Completed within 40 Service Hours
10064619	Request to send Message to all use	Closed	P40 - To Be Completed within 40 Service Hours
10064623	Request to solve Issue with Worksta	Closed	P40 - To Be Completed within 40 Service Hours
10064645	Request to Remove Account - CH	Closed	P40 - To Be Completed within 40 Service Hours

**Figura A.6 – Relatório utilizando operador *Group By***

Utilizando o BO, podemos então elaborar diversos relatórios sobre diversos Universos que irão ajudar a quantificar as Medias de Satisfação adoptadas.

# Apêndice B

## Implementação de Método de Avaliação em MATLAB

### **Main: curva\_distribuição.m**

```
vector = [1 0.94 0.89 0.85];
d=[1 0.8 0.6 0.3];
cont=1;
total=0;
for sigma=0.01:0.01:0.2
    aux=0;
    r=gaussiana(sigma,vector);
    for a=1:4
        aux=aux+abs(r(a)-d(a));
    end
    res(cont)=aux;
    total=total+1;
    cont=cont+1;
end
indice=find(res==min(res));
sigmai=0.01*indice; % Sigma Ideal

x = 1:-0.001:0.5;
y = gaussiana(sigmai,x);

% Matriz com 2 colunas (1ª com valores registados e 2ª com
respectivos pontos)

v= [x' y'];

%res = pontuacoes(v,d);

% Gera gráfico da distribuição de pontos
plot(v(:,2))

% Gera ficheiro de texto com os valores possíveis na primeira
coluna e os respectivos pontos atribuídos a cada valor

dlmwrite('Pontuacoes.ascii',v,'delimiter',' ','');
```

### **Função: gaussiana.m**

```
function res = gaussiana(sigma,x)

p1 = -.5 * ((x-1)/sigma) .^2;
p2 = (sigma * sqrt(2*pi));
res = exp(p1) ./p2;

res=(1/res(1))*res;
```

# Apêndice C

## Pré-Tratamento de Dados Algoritmo *Apriori*

A folha de cálculo com as respostas às perguntas do questionário é o único *input* do programa em MATLAB. Para o programa criar as regras de associação, a informação presente na folha de cálculo deverá seguir um determinado formato:

- Primeira linha da folha deverá ser texto livre com a descrição da pergunta em questão
- As células da folha com as respostas às perguntas deverão estar no formato binomial, isto é, TRUE ou FALSE

A exportação das perguntas e respostas do *Web Surveyor* apresenta as perguntas na primeira linha onde cada coluna corresponde a uma pergunta. Cada linha corresponde a uma transacção, ou seja, cada linha apresenta as respostas que cada utilizador introduziu no momento de preenchimento. Todavia, cada célula é multi-valor havendo necessidade de manipular a folha de cálculo de forma a tornar cada célula em valores binomial.

Para efectuar a alteração, foi criado um pequeno processo no *RapidMiner* utilizando operadores para o efeito.

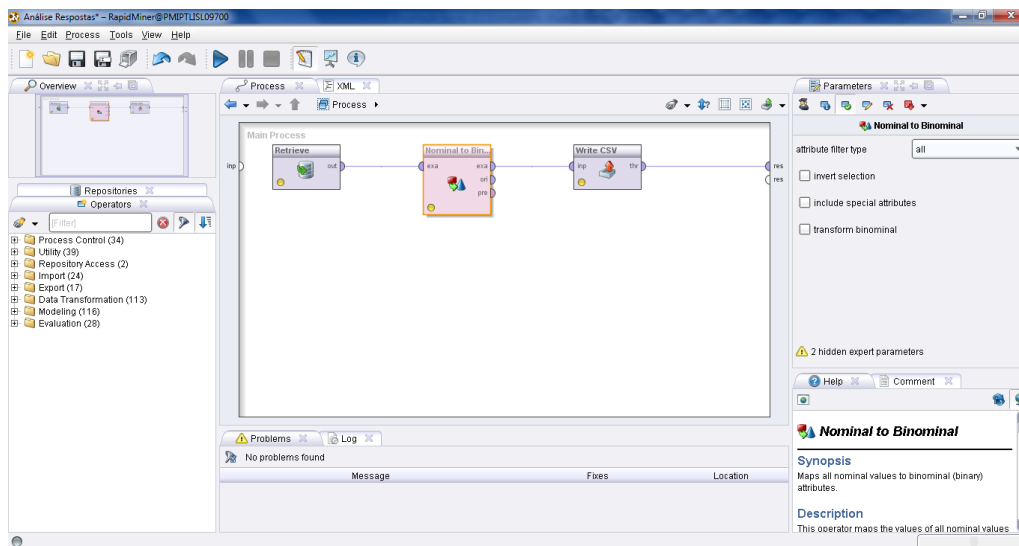


Figura C.1 – Processo em RapidMiner

Na Figura C.1 é apresentado não só o ambiente de desenvolvimento em RapidMiner mas também o processo com a utilização dos operadores. Identificam-se três operadores:

- Retrieve

- Nominal to Binominal
- Write CSV

O processo inicia-se com o carregamento da folha de cálculo, onde na aplicação existe o operador **Retrieve** que através de um *wizard* torna o processo de importação uma tarefa fácil. Nesta fase é possível verificar a correcta importação dos dados ao analisar o output do processo mesmo antes deste ser executado (Figura C.2). Informações como a o número de linhas, atributos (colunas), tipo dos dados são exibidas e intervalos que estes tomam. Note-se na figura que o tipo dos dados é *polynomial*, isto é, cada campo toma diferentes valores (Ex: Dissatisfied, Neutral, Manager, etc.).

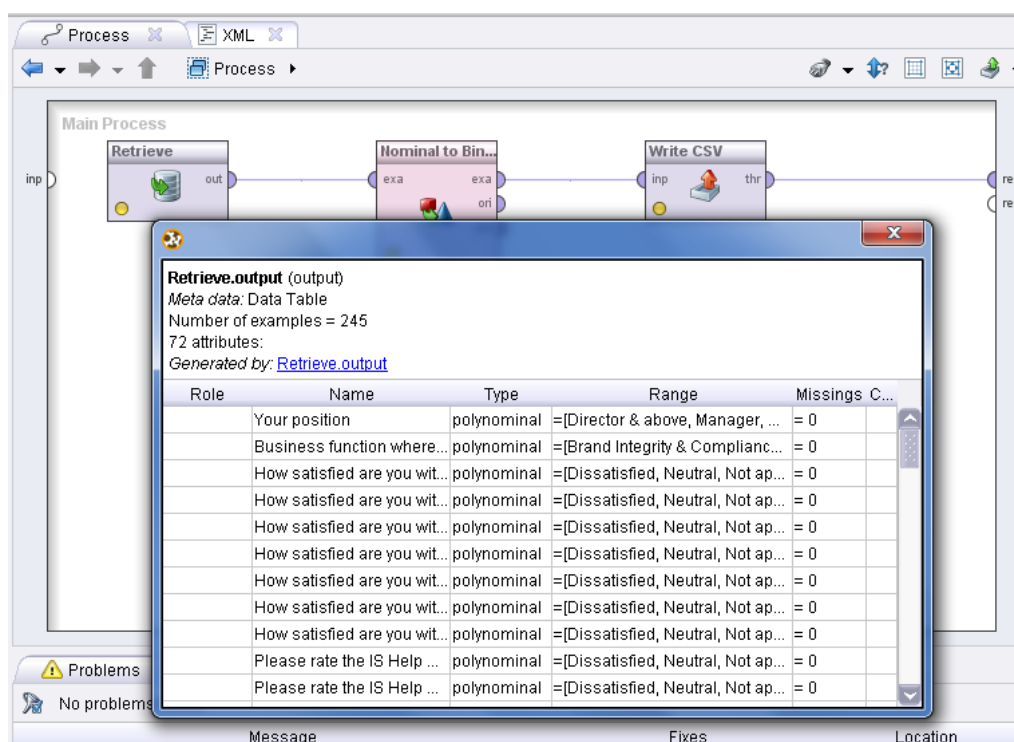


Figura C.2 – Análise de importação de dados

Importados os dados correctamente, aplica-se o operador **Nominal to Binomial** para transformar as respostas em TRUE e FALSE (Figura C.3).

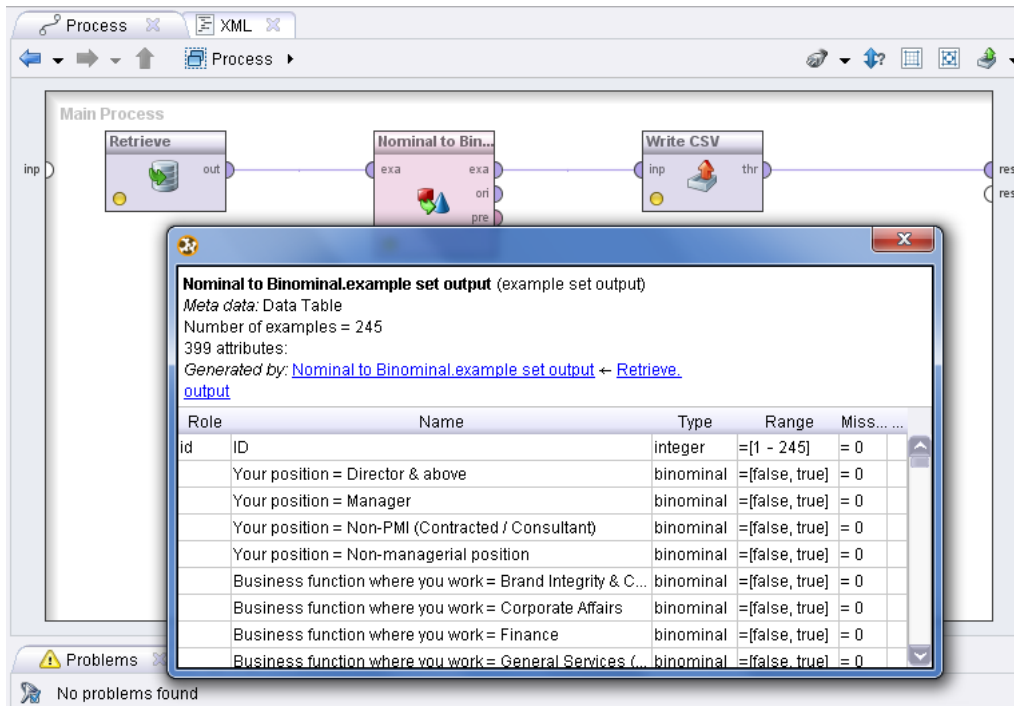


Figura C.3 – Análise de transformação Nominal to Binominal

Como se pode confirmar na figura, os dados sofreram modificações.

- Aumento do número de atributos de 72 para 399
- Tipo dos dados passou a binomial
- Intervalo de valores dos itens passou a ser “true” e “false”

O aumento do número de atributos está relacionado com o funcionamento do operador. De forma a tornar os dados binomiais, é necessário fragmentar cada pergunta nas diversas respostas e posteriormente classifica-las como “true” ou “false”.

De forma a compreender melhor o operador, tomemos a primeira pergunta do questionário “Your position”. Antes da aplicação do operador, a pergunta correspondia a uma coluna em que cada linha dessa coluna tomava o valor da resposta (Director & above, Manager, etc.). Após a aplicação, a coluna fragmentou-se em N colunas, sendo N o número de respostas possíveis que o utilizador pode escolher. Neste caso a pergunta podia tomar 4 valores, logo fragmentou-se em 4 colunas cada uma corresponde a uma resposta (Ex: “Your position = Manager”).

Após a transformação dos dados, é necessário exportar de forma a serem utilizados pelo MATLAB. O *RapidMiner* é dotado de diversos operadores de exportação, podendo exportar

directamente para folha de cálculo Excel no formato XLS, no entanto este apresenta uma limitação.

O formato XLS clássico está limitado a 256 colunas, no entanto a nova versão Excel 2007 e o novo formato XLSX não sofre esta limitação. É por esta razão que se utilizou o operador **Write CSV** que exporta no formato CSV (*Comma-Separated Values*). Após exportação neste formato, efectua-se a importação para o novo Excel 2007 no formato XLSX onde estão disponibilizadas as 399 colunas com a informação.

# Apêndice D

## Implementação do Algoritmo *Apriori* em MATLAB

### **Main:** regrasAssociacao.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Regras de Associação %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Gera regras do tipo A -> 'B
%

% Guardar nome do ficheiro a ser analisado (extensão .xls ou
.xls)
nomeFicheiro='teste2.xlsx';
%nomeFicheiro = 'Relacao_Aplicacoes.xls';
%nomeFicheiro = 'livro_teste.xlsx';

% Carregar folha excel com respostas para o Matlab
[valores, label] = xlsread(nomeFicheiro);
% Guardar numa nova variável as respostas complementares (ex:
'A, 'B, etc.)
valoresComplementares = criaMatrizComplementar(valores);

% Guardar dimensão da matriz com valores e dos labels
[lValores, cValores] = size(valores);
[lLabel, cLabel] = size(label);

% Nível de Suporte das Regras
suporte = 0.50;
% Nível de Confiança das Regras
confianca = 0.97;

% Rotina para encontrar as perguntas que passam o suporte mínimo
somaOriginais = sum(valores);
somaComplementares = sum(valoresComplementares);

% Para cada coluna da matriz somaOriginais, efectuar a soma de
cada linha
% para encontrar as colunas que apresentam suporte > 50% e
guarda numa
% matriz L1_positivos

L1_positivos = (somaOriginais/lValores)>=suporte;

% Para cada coluna da matriz somaComplementares, efectuar a soma
de cada linha
% para encontrar as colunas que apresentam suporte > 50% e
guarda numa
% matriz L1_complementares
```

```

L1_complementares = (somaComplementares/lValores)>=suporte;

% Guarda a dimensão da matriz com elementos que passaram o
suporte 50%
[lL1_Orig, cL1_Pos] = size(L1_positivos);
[lL1_Compl, cL1_Compl] = size(L1_complementares);

% Cálculo das variáveis que têm suporte maior que o threshold
[valores_positivos_L1, label_positivos_L1] =
criaMatrizL1(cL1_Pos, L1_positivos, valores, label);
[valores_complementares_L1, label_complementares_L1] =
criaMatrizL1(cL1_Compl, L1_complementares,
valoresComplementares, label);

selecionados = criaMatrizL2(valores_positivos_L1,
label_positivos_L1, valores_complementares_L1,
label_complementares_L1, suporte);
%sel = selecionados;
regras = criaRegras(selecionados, valores_positivos_L1,
label_positivos_L1, valores_complementares_L1,
label_complementares_L1, confianca);

% Identifica violações do questionário
[violacoes viol] = testaViolacoes(regras, label, lValores,
cValores, valores);

```

**Função: criaMatrizComplementar.m**

```

function vector = criaMatrizComplementar(numeros)
vector = numeros;
for i=1:numel(numeros)
    if(numeros(i)) == 1 vector(i) = 0;
    else vector(i) = 1;
    end
end

```

**Função: criaMatrizL1.m**

```

function [ valores_final, label_final ] = criaMatrizL1(
tamanhoIndices, matrizIndices, matrizValores, matrizLabels )
% criaMatrizL1 - Função que retorna os valores finais e os
labels
% associados às entradas que passaram o suporte mínimo definido

y = 1;
for i= 1:tamanhoIndices
    if(matrizIndices(1,i)) == 0
        matrizValores(:,y) = [];
        matrizLabels(:,y) = [];
    else
        if(matrizIndices(1,i)) == 1
            y=y+1;
        end
    end
end

```

```

        end
    end
end
valores_final = matrizValores;
label_final = matrizLabels;
end

```

### **Função: criaMatrizL2.m**

```

function [ selecionados ] = criaMatrizL2( valores_originais_L1,
label_originais_L1, valores_complementares_L1,
label_complementares_L1, suporte )
%UNTITLED Summary of this function goes here
% Detailed explanation goes here

[lL1_Orig, cL1_Orig] = size(valores_originais_L1);
[lL1_Comp1, cL1_Comp1] = size(valores_complementares_L1);

selecionados = zeros(0,2);
i=1;
j=1;
while(i<(cL1_Orig+1))
    while(j<(cL1_Comp1+1))
        if( strcmp(label_originais_L1(1,i),
label_complementares_L1(1,j) ) == 1)
            j = j+1;
        else
            if(
sum(valores_originais_L1(:,i).*valores_complementares_L1(:,j)) /
lL1_Orig ) >= suporte
                selecionados = [selecionados; [i j]];
                j = j+1;
            else j=j+1;
            end
        end
    end
    i = i+1;
    j = 1;
end
end

```

### **Função: criaRegras.m**

```

function [ regras ] = criaRegras(selecionados,
valores_originais_L1, label_originais_L1,
valores_complementares_L1, label_complementares_L1, confianca)
%UNTITLED Summary of this function goes here
% Detailed explanation goes here

[~, cL1_Orig] = size(valores_originais_L1);
[~, cL1_Comp1] = size(valores_complementares_L1);
[lSel, ~] = size(selecionados);
regras = zeros(0,2);

```

```

for i=1:lSel
    if
sum(valores_originais_L1(:,selecionados(i,1)).*valores_complementares_L1(:,selecionados(i,2)))/sum(valores_originais_L1(:,selecionados(i,1)))>=confianca

        regras = [regras;
[label_originais_L1(1,selecionados(i,1))
label_complementares_L1(1,selecionados(i,2))]];
        end
        %sum(valores_originais_L1(:,selecionados(8,1))

end

end

```

**Função: testaViolacoes.m**

```

function [ violacoes viol ] = testaViolacoes(regras, label,
lValores, cValores, valores)

[lRegras cRegras] = size(regras);
viol = zeros(0,2);
violacoes = 0;

for i=1:lRegras
    for j=1:cValores
        if( strcmp( regras(i,1), label(1,j) ) ) == 0
            else
                for k=1:lValores
                    if valores(k,j) == 1 % Encontrou questionário
com premissa verdadeira
                        for x=1:cValores
                            if ( strcmp ( regras(i,2), label(1,x) )
) == 0
                                else
                                    if valores(k,x) == 1 % Encontrou
questionário com conclusão verdadeira
                                        violacoes = violacoes + 1;
                                        viol = [viol; [regras(i,1)
regras(i,2)]];
                                            end
                                        end
                                    end
                                end
                            end
                        end
                    end
                end
            end
        end
    end
end

```