

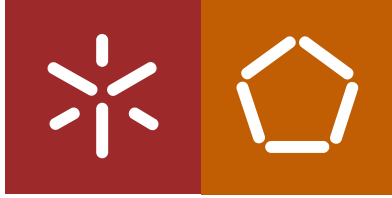


Universidade do Minho
Escola de Engenharia

Mariana Reimão Queiroga Valério de Carvalho
Descoberta de Padrões num Sistema de Certificação
de Empresas

Mariana Reimão Queiroga Valério de Carvalho

Descoberta de Padrões num Sistema de
Certificação de Empresas



Universidade do Minho
Escola de Engenharia

Mariana Reimão Queiroga Valério de Carvalho

Descoberta de Padrões num Sistema de
Certificação de Empresas

Dissertação de Mestrado
Mestrado em Engenharia Informática

Trabalho efectuado sob a orientação do
Professor Doutor Orlando Manuel de Oliveira Belo

Outubro de 2011

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Descoberta de Padrões num Sistema de
Certificação de Empresas

Mariana Reimão Queiroga Valério de Carvalho

Dissertação de Mestrado

2011

Descoberta de Padrões num Sistema de Certificação de Empresas

Mariana Reimão Queiroga Valério de Carvalho

Dissertação apresentada à Universidade do Minho para obtenção do grau de Mestre em Engenharia Informática, na especialidade de Sistemas de Suporte à Decisão, elaborada sob orientação do Professor Doutor Orlando Manuel de Oliveira Belo.

2011

Aos meus pais e irmã.

Agradecimentos

A realização desta dissertação não teria sido possível sem o apoio de diversas pessoas que me ajudaram durante o decorrer da mesma. Um agradecimento especial a todas que me acompanharam durante esta fase e sem as quais não teria conseguido.

Em primeiro lugar gostaria de agradecer ao meu orientador, professor doutor Orlando Belo, por todo o apoio e orientação. Por ser incansável e me ter ajudado sempre que precisava e me fez crescer bastante a nível científico. E ao professor doutor Paulo Sampaio, que forneceu os dados, sem os quais não teria sido possível a realização deste trabalho, e que me deu a conhecer a problemática da certificação de empresas, queria também manifestar o meu agradecimento.

À minha colega e amiga, Ana Sequeira, por toda a paciência, pelos conselhos e por se ter demonstrado um verdadeiro apoio durante todo o processo.

Queria agradecer ao André Rodrigues que me acompanhou durante todo este processo da realização da dissertação e foi incansável comigo, ajudando-me quando me deparava com alguma dificuldade, dando conselhos e por me dar força para nunca desistir.

Por fim, queria agradecer aos meus pais e irmã por todo o apoio e força que me deram e sem os quais não teria conseguido. Por toda a dedicação e por toda a paciência, estou muito agradecida.

Resumo

Descoberta de Padrões num Sistema de Certificação de Empresas

A mineração de dados tem como principal objectivo descobrir padrões escondidos num conjunto de dados, isto é, informação útil que ajude a tomar decisões sobre um determinado tema. Nesta dissertação usa-se a mineração numa base de dados que descreve os certificados adquiridos pelas empresas portuguesas no período 2008-2010. A certificação é um processo voluntário, que, apesar de ser bastante demorado, custoso e envolver demasiada burocracia, pode ser crucial para a sobrevivência das empresas. Os certificados demonstram o compromisso da empresa com a qualidade dos produtos e dos serviços, e com o meio ambiente, a saúde e a segurança dos trabalhadores, entre outros. Quais os certificados que uma empresa deve adquirir é uma das principais questões que se colocam a uma nova empresa no mercado. A resposta a esta pergunta pode variar com vários factores, como a região onde a empresa se encontra localizada ou o seu sector de actividade. É necessária uma análise prévia que forneça um conhecimento do estado do mercado para tomar as melhores decisões. A aplicação de técnicas de mineração de dados permite obter uma descrição do estado do mercado e uma previsão de quais os certificados a adquirir consoante as características de uma empresa. A informação extraída facilita a tomada de decisão relativamente ao conjunto de certificados que melhor se adapta às características da empresa. Este conjunto de certificados varia com a competitividade resultante do número de empresas na região e do número de empresas no sector de actividade em que a nova empresa se encontra inserida. Os resultados apresentados nesta dissertação fornecem às novas empresas uma orientação inicial no mercado competitivo.

Palavras-Chave: Mineração de Dados, Processos de certificação de Empresas, Classificação e Segmentação de Dados, Descoberta de Padrões de Certificação.

Abstract

Discovery of patterns in a system of certification of companies

The main objective of data mining is to discover patterns in a database, i.e., useful information to support decision making on a given subject. In this dissertation data mining is used in a database that describes the certificates acquired by Portuguese companies in the period 2008 – 2010. Certification is a voluntary process, which, despite the time taken in the process, the costs and the bureaucracy, may be crucial for the survival of a company. Certificates are pledges of the commitment of the company with the quality of services and products, and with the environment, the health and the security of the workers, among others. Which are the certificates that a company must acquire is one of the main questions that is placed to a new company in the market. The answer to this question may depend on several factors, as the region where the company is located or its activity sector. A previous analysis is necessary to get the knowledge about the state of the market that enables to take the best decisions. The application of data mining techniques allows a description of the state of the market and a forecast of which are the certificates to acquire depending on the characteristics of a company. The information extracted supports the decision making relatively to the set of certificates that better suit the characteristics of the company. This set of certificates varies with the competitiveness resulting from the number of companies in the region and with the number of companies in the activity sector to which the new company belongs. The results presented in this dissertation provide an initial orientation to the new companies in the competitive market.

Keywords: Data Mining, Process of Certification of Companies, Data Classification and Data Clustering, Discovery of Patterns of Certification

Índice

Introdução	1
1.1 O tecido empresarial português	1
1.2 Motivações e Objectivos	3
1.3 Estrutura da dissertação	5
A Certificação de Empresas	7
2.1 O Processo de Certificação.....	7
2.2 Certificação de Sistemas de Gestão de Qualidade	9
2.3 Certificação de Sistemas de Gestão Ambiental.....	12
2.4 Certificação de Sistemas da Segurança e Saúde no Trabalho.....	14
2.5 Outros Certificados.....	16
2.5.1 Verificação EMAS II	16
2.5.2 Certificação HACCP - Segurança Alimentar.....	16
2.5.3 Certificação Segurança Alimentar	17
2.5.4 Certificação de Sistemas de Gestão de Qualidade no Sector Automóvel	17
2.5.5 Certificação de Sistemas de Gestão de Investigação, Desenvolvimento e Inovação ...	18
2.5.6 Certificação de Sistemas de Gestão da Responsabilidade Social.....	18
2.5.7 Certificação de Sistemas de Gestão de Responsabilidade Social.....	18
2.6 Uma Nota Final sobre Certificação	19
O Sistema de Dados Alvo	21
3.1 Certificação e Informação	21

3.2	O Modelo de Dados.....	25
3.3	Análise Exploratória dos dados.....	26
3.4	Construção, Integração e Formatação dos dados	38
	O Processo de Descoberta de Conhecimento	45
4.1	Metodologia do CRISP-DM	46
4.2	Mineração de Dados.....	48
4.2.1	Técnicas de Mineração de Dados	48
4.2.2	Classificação.....	48
4.2.3	Segmentação	54
4.2.4	Escolha da Técnica de Mineração de Dados	57
4.2.5	Importância dos métodos na obtenção dos resultados.....	58
	Análise dos Resultados	60
5.1	Caracterização dos Sectores usando Classificação.....	60
5.1.1	Por Norma	61
5.1.2	Por Região	64
5.1.2.1	Caracterização dos Sectores de Actividade da Região Norte	65
5.1.2.2	Caracterização dos Sectores de Actividade da Região de Lisboa e Vale do Tejo	68
5.1.2.3	Caracterização dos Sectores de Actividade da Região Centro.....	72
5.1.2.4	Caracterização dos Sectores de Actividade da Região Alentejo.....	75
5.1.2.5	Caracterização dos Sectores de Actividade da Região Algarve.....	77
5.1.2.6	Caracterização dos Sectores de Actividade da Região Madeira.....	79
5.1.2.7	Caracterização dos Sectores de Actividade da Região Açores.....	81
5.2	Caracterização dos Sectores usando Segmentação	83
	Conclusões e Trabalho Futuro	88
6.1	Processos e Resultados	88
6.2	Linhas de Orientação para Trabalho Futuro.....	91
	Bibliografia	94
	Referências WWW	102
	Anexo I	104

Índice de Figuras

Figura 1 - Número de Empresas Certificadas por Referencial.....	22
Figura 2 - Número de Empresas Certificadas com a Norma ISO 9001:2000 implementada por Região	23
Figura 3 - Número de Empresas Certificadas com a Norma ISO 14001:2004 implementada por Região	23
Figura 4 - Número de Empresas Certificadas com a Norma OHSAS 18001:1999 implementada por Região	24
Figura 5 - Distribuição das Empresas Certificadas pelas Diversas Regiões Nacionais.....	28
Figura 6 - Número de Empresas Nacionais Certificadas por EAC	28
Figura 7 - Top 5 dos Sectores de Actividade da Região Norte.....	30
Figura 8 - Top 5 de Sectores de Actividade da Região Lisboa e Vale do Tejo	31
Figura 9 - Top 5 dos Sectores de Actividade da Região Centro	32
Figura 10 - Top 5 dos Sectores de Actividade da Região Alentejo	34
Figura 11 - Top 5 dos Sectores de Actividade da Região Algarve	35
Figura 12 - Top 5 dos Sectores de Actividade da Região Madeira.....	36
Figura 13 - Top 5 dos Sectores de Actividade da Região Açores	37
Figura 14 - Top 5 dos Sectores de Actividade em Portugal	37
Figura 15 - Modelo Relacional do <i>Data Mart</i> construído.....	42
Figura 16 - Fases da Metodologia CRISP-DM.....	47
Figura 17 - Tratamento dos valores discretos pelo algoritmo C4.5	51
Figura 18 - Tratamento dos valores contínuos pelo algoritmo C4.5	51
Figura 19 - Representação dos Centróides gerados	84

Índice de Tabelas

Tabela 1 - Densidade das empresas nas regiões nacionais	2
Tabela 2 - Conjunto de Dados Inicial.....	26
Tabela 3 – A Matriz de Decisão utilizada	39
Tabela 4 - Constituição da Dimensão Certificado	40
Tabela 5 - Constituição da Dimensão Região.....	40
Tabela 6 - Constituição da Dimensão Empresa	41
Tabela 7 - Constituição da Dimensão EAC.....	41
Tabela 8 - Constituição da Dimensão Tempo	42
Tabela 9 - Sectores de Actividade com Norma ISO 9001:2000 predominante	62
Tabela 10 - Sectores de Actividade com as Normas ISO 9001:2000 e ISO 14001:2004 predominantes.....	62
Tabela 11 - Sectores de Actividade com as Normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999 predominantes	63
Tabela 12 - Sectores de Actividade com Diversas Normas predominantes	63
Tabela 13 - Sectores de Actividade com Diversas Normas predominantes (continuação)	64
Tabela 14 - Precisão e taxa de erro de cada árvore de decisão obtida.....	65
Tabela 15 - Caracterização dos Sectores de Actividade da Região Norte.....	67
Tabela 16 - Caracterização dos Sectores de Actividade da Região Norte (continuação).....	68
Tabela 17 - Caracterização dos Sectores de Actividade da Região de Lisboa e Vale do Tejo.....	70
Tabela 18 - Caracterização dos Sectores de Actividade da região de Lisboa e Vale do Tejo (continuação)	71
Tabela 19 - Caracterização dos Sectores de Actividade da Região Centro	74
Tabela 20 - Caracterização dos Sectores de Actividade da Região Centro (continuação)	75

Tabela 21 - Caracterização dos Sectores de Actividade da Região Alentejo.....	76
Tabela 22 - Caracterização dos Sectores de Actividade da Região Alentejo (continuação).....	77
Tabela 23 - Caracterização dos Sectores de Actividade da Região Algarve.....	79
Tabela 24 - Caracterização dos Sectores de Actividade da Região Autónoma da Madeira.....	80
Tabela 25 - Caracterização dos Sectores de Actividade da Região Autónoma da Madeira (continuação)	81
Tabela 26 - Caracterização dos Sectores de Actividade da Região Autónoma dos Açores.....	82
Tabela 27 - Caracterização dos Sectores de Actividade da Região Autónoma dos Açores (continuação)	83
Tabela 28 - Resultados de valores de <i>Davies Bouldin</i>	83
Tabela 29 - Resultados de valores de <i>Davies Bouldin</i> (continuação).....	84
Tabela 30 - Composição dos segmentos obtidos	84
Tabela 31 - Anexo I - Correspondência entre o Sector de Actividade e o EAC.....	104

Lista de Siglas e Acrónimos

PME	Micro, Pequena ou Média empresa
INE	Instituto Nacional de Estatística
ISO	<i>International Organization for Standardization</i>
OHSAS	<i>Occupational Health and Safety Assessment Series</i>
EMAS	<i>Eco-Management and Audit Scheme</i>
HACCP	<i>Hazard Analysis Critical Control Point</i>
APPCC	Análise de Perigos e Pontos Críticos de Controle
SA	<i>Social Accountability</i>
EAC	<i>Economic Activity Code</i>
CAE	Código de Actividade Económica
NUTS	<i>Nomenclature commune des Unités Territoriales Statistiques</i> (Unidades Territoriais Estatísticas de Portugal)
SGQ	Sistema de Gestão de Qualidade
SGA	Sistema de Gestão Ambiental
SGSSO	Sistema de Gestão e Segurança e Saúde Ocupacional
SGSA	Sistema de Gestão de Segurança Alimentar
SGIDI	Sistema de Gestão de Investigação, Desenvolvimento e Inovação
SGRS	Sistema de Gestão de Responsabilidade Social
KDD	<i>Knowledge Discovery in Databases</i>
CRIPS-DM	<i>Cross Industry Standard Process for Data Mining</i>
SEMMA	<i>Sample, Explore, Modify, Model and Access</i>
DB	<i>Davies Bouldin</i>

Capítulo 1

Introdução

1.1 O tecido empresarial português

Nos dias que correm, é difícil para uma empresa subsistir sem um bom sistema de gestão que lhe assegure um lugar no mercado competitivo. Para sobreviver é necessário ter consciência do estado do mercado, ou seja, ser capaz de identificar quais as empresas que representam uma competição directa e indirecta. Um dos aspectos importantes para a sobrevivência é a aquisição de certificados. A certificação é um processo que consiste na avaliação dos processos de uma determinada empresa e verifica se esses estão de acordo com regras já estipuladas por organizações competentes. No final deste processo a empresa adquire o certificado pretendido. A certificação permite, também, às empresas melhorarem os seus sistemas de gestão, satisfazendo os requisitos dos clientes e assegurando o seu lugar no mercado competitivo. Cada empresa tem interesse em saber quais os certificados que as empresas que lhe são directa ou indirectamente concorrentes possuem. Com esta informação é mais fácil decidir qual ou quais certificados é mais benéfico adquirir no sector de actividade em que está inserida.

As Micro, Pequenas e Médias Empresas (PME) representam uma importante parcela das empresas em Portugal traduzindo uma importante contribuição para o país a nível económico e de criação de

emprego, sendo responsáveis por uma percentagem muito significativa (72,5%) de postos de trabalho (INE, 2010b). O número de empresas tem vindo a aumentar gradualmente ao longo dos anos, com uma taxa de crescimento médio de 2,7%, entre 2004 e 2007 (INE, 2009a). Neste intervalo, os sinais da crise que se vive actualmente ainda não se notavam: as taxas de natalidade e de mortalidade de empresas demonstravam-se bastante elevadas, apresentando cerca de 15,7% de taxa de natalidade e cerca de 15,5% de taxa de mortalidade. A criação e o encerramento de empresas é uma realidade económica normal, devendo-se ao facto de as novas empresas criadas serem muitas vezes preferidas pelos clientes às empresas com maior tempo de vida, que se mostram muito menos eficientes do que as com menos anos de vida (INE 2008, 2009a).

Tabela 1 - Densidade das empresas nas regiões nacionais

Densidade das empresas (Nº/Km ²)	2006	2007	2008
Norte	16,6	16,8	16,7
Lisboa e Vale do Tejo	8,4	8,5	8,4
Centro	112,7	114,7	113,5
Alentejo	2,1	2,2	2,1
Algarve	11,2	11,7	11,9
Região Autónoma dos Açores	8,1	8,4	8,6
Região Autónoma da Madeira	27,6	27,4	27,3

(Fonte: INE 2008, 2009b, 2010c)

Na Tabela 1 é possível verificar que entre os anos 2006 e 2007 o número de empresas por Km² apresenta um aumento significativo, ao contrário do que se verifica na transição 2007 e 2008, em que este número diminui em média pelas várias regiões nacionais, apenas se verificando um aumento do número de empresas por Km² na região do Algarve e Região Autónoma dos Açores.

Em 2007 começam-se a denotar os primeiros sinais da crise que se iria viver em Portugal, apesar de os indicadores relativos a esse ano parecerem insignificantes quando comparados com o que iria acontecer nos anos seguintes. A taxa de natalidade neste ano diminui para 15,2% e verifica-se que cerca de 30% das empresas existentes em Portugal não sobrevivem mais de um ano de vida (INE, 2009a). A questão da competição no mercado ainda se torna mais relevante em cenários de crise económica, como aquele que o país está a atravessar. No período que se seguiu a 2007, a economia portuguesa desacelerou de forma acentuada: em 2008, menos 0,5% empresas activas

do que no ano anterior (INE, 2010a) e a taxa de natalidade de empresas decresce para 14,17% (INE, 2010c). Em 2009, os sinais de crise não apresentaram melhorias, continuando a diminuir os valores das principais variáveis, como por exemplo um decréscimo de 3,2% do total do sector empresarial face ao ano anterior. O volume de negócios apresenta uma descida de 10,2% comparando com o valor do ano anterior (INE, 2011c). Actualmente, representando as PME 99,7% do total das empresas no sector não financeiro (INE 2011a, 2011b), o país ainda vive num cenário de crise económica, e é crucial para as empresas conseguirem sobreviver neste ambiente.

Como se poderá verificar, o número de certificados adquiridos pelas empresas aumentou gradualmente ao longo do período de tempo analisado, e em todas as regiões. Em períodos de crise, quando a actividade se reduz e existem folgas de capacidade, as empresas economicamente saudáveis tipicamente usam os seus recursos para se prepararem para as novas fases do ciclo económico. Essa é a justificação para o aumento da procura pela obtenção de certificados por parte de empresas que, tendo boas perspectivas de sobreviver no mercado, sabem que a certificação é essencial para esse efeito.

1.2 Motivações e Objectivos

A decisão de escolha dos certificados a adquirir por uma determinada empresa tem de ser bastante ponderada e é necessário ter vários factores em consideração, como por exemplo: tamanho da empresa, número de empregados, sector de actividade em que a empresa está inserida e a região onde a empresa está localizada. Desta forma, é uma vantagem para os responsáveis pela escolha dos certificados a adquirir pela empresa ter noção e conhecimento do estado do mercado: quais os certificados mais requisitados pelas empresas, por sector de actividade e por região. Desta forma, será uma mais-valia para a empresa e ganhará assim uma vantagem competitiva no mercado, tendo uma maior probabilidade de conseguir sobreviver.

Para facilitar a decisão da escolha de certificados foi decidido a aplicação de duas técnicas de mineração de dados. Estas técnicas são aplicadas a uma base de dados que foi elaborada por dois professores Paulo Sampaio e Pedro Saraiva e que desde 2008 colaboram para o "Guia de

Empresas Certificadas". Esta colaboração consiste num estudo realizado anualmente sobre a evolução dos sistemas de certificação, em que apenas são apresentados resultados sobre o total dos principais certificados obtidos pelas empresas ao longo dos anos, como por exemplo, as normas de certificação de:

- Sistemas de Gestão de Qualidade (ISO 9001:2000).
- Sistemas de Gestão Ambiental (ISO 14001:2004).
- Sistemas de Gestão de Saúde e Segurança Ocupacional (OHSAS 18001:1999).

As técnicas de mineração de dados aplicadas a estes dados recolhidos desde 2008 irão possibilitar obter uma descrição do estado do mercado que seja credível e que permita uma previsão que ajude os responsáveis a tomar a decisão de quais os certificados será melhor para uma empresa obter tendo em conta vários factores. A mineração de dados pode ser vista como um processo em que ocorre a procura por padrões e relacionamentos escondidos em grandes volumes de dados, aplicando técnicas e algoritmos adequados (Adriaans & Zantinge, 1996; Cabena, et al., 1998). Para isto e tendo em conta os vários requisitos foram escolhidas as técnicas de mineração Classificação e Segmentação. O objectivo da aplicação de um processo de Classificação é obter uma previsão do conjunto de certificados que será melhor adquirir tendo em conta diversas características da nova empresa. E com os resultados da Segmentação será possível obter uma descrição do estado do mercado, ajudando a ter uma noção dos certificados que são mais adquiridos quer por sector de actividade, quer por região.

Através da utilização destas técnicas de mineração de dados será possível então, analisar o estado do mercado competitivo. Esta análise vai possibilitar a uma nova empresa uma visão sobre a disposição e preferência de certificados consoante as várias características de uma nova empresa. Para isso, é necessário para cada organização adquirir um conjunto de referenciais para melhorar os seus sistemas de gestão, que lhes forneçam vantagens no mercado competitivo. Este conjunto de normas varia com o sector de actividade em que a organização está inserida e com a região onde esta se encontra localizada. As normas requeridas por um determinado sector de actividade são importantes para uma organização se esta pretende conseguir a confiança dos clientes e das entidades envolvidas, satisfazendo assim requisitos exigidos por estes e melhorando a imagem organizacional, entre outras vantagens. Esta tomada de decisão é bastante crítica, não só em

cenários de crise, como aquele que de momento se vive em Portugal, como também é crucial para a sobrevivência de empresas num cenário comum da economia.

Com este estudo será possível obter uma descrição do estado do mercado, ajudando novas empresas a iniciar o negócio escolhendo quais os melhores certificados a adquirir por uma empresa para conseguir sobreviver perante o cenário que se vive actualmente na economia portuguesa, através da caracterização dos vários sectores de actividade, indicando quais as principais normas adquiridas em cada um, e identificando quais os principais grupos de empresas com características semelhantes, ajudando assim, novas empresas a uma orientação inicial no mercado competitivo.

1.3 Estrutura da dissertação

Além do presente capítulo, esta dissertação está organizada da seguinte forma:

- Capítulo 2 - A Certificação de Empresas - Aqui será feita uma abordagem teórica introduzindo o tema dos certificados, em que consistem, quais as motivações para as empresas escolherem um determinado certificado, quais as vantagens ou desvantagens que advém da sua aquisição. Todos estes factores dependem do tamanho da empresa, sector de actividade e região onde a empresa se encontra localizada.
- Capítulo 3 - O Sistema de Dados Alvo - Neste capítulo será feita uma apresentação do caso de estudo e em que contexto é que surgiu. Será realizada, também, uma descrição exhaustiva do conteúdo da base de dados fornecida pelos Professores Paulo Sampaio e Pedro Saraiva, juntamente com os passos que foram executados para preparar a mesma para a mineração de dados.
- Capítulo 4 - O Processo de Descoberta de Conhecimento - Uma breve abordagem teórica sobre as técnicas de mineração existentes será realizada neste capítulo, apresentando-se

também as vantagens e desvantagens correspondentes, ajudando assim à escolha das técnicas de mineração pretendidas para obter os padrões escondidos nos dados.

- Capítulo 5 – Análise dos Resultados - Este é o capítulo dos resultados do trabalho. Nele apresentaremos o que obtivemos a partir dos processos de mineração de dados e faremos uma caracterização dos sectores usando técnicas de classificação e segmentação, quer por certificado, quer por região.
- Capítulo 6 – Conclusão e Trabalho Futuro - Aqui terminaremos os trabalhos desta dissertação, apresentando alguns comentários e conclusões acerca do trabalho realizado, bem como apontaremos algumas ideias sobre algum trabalho futuro que poderá ser realizado sobre este tema.

Capítulo 2

A Certificação de Empresas

Neste capítulo, serão apresentados alguns dos termos teóricos mais relevantes relacionados com os certificados analisados neste trabalho de dissertação, com o objectivo de familiarizar o utilizador com as normas existentes, tais como as principais vantagens e desvantagens da aquisição dos certificados por parte das empresas. Apresenta-se de seguida a forma como decorre a certificação e enumeram-se as características dos principais certificados.

2.1 O Processo de Certificação

É de extrema importância que as empresas tenham o conhecimento em que consistem e o que defendem as normas existentes, pois estas são uma mais-valia para a imagem projectada pela empresa no mercado e são essenciais para a sua sobrevivência. O processo em que uma empresa adquire um certificado, seguindo à risca os requisitos defendidos por estas, designa-se por processo de certificação.

A certificação consiste num processo no qual uma empresa se sujeita a um processo de avaliação, que consiste na análise dos protocolos dessa organização, verificando se a empresa segue um

determinado conjunto de normas previamente definidas a nível internacional ou nacional, normalmente tidas como boas práticas. Este processo inicia-se com uma análise da informação previamente recolhida pelos elementos da organização, como a identificação dos processos que caracterizam a actividade da organização, seguindo-se uma verificação se estes estão de acordo com as normas estipuladas. Os certificados normalmente possuem um prazo de validade de 3 anos, período após o qual a empresa pode decidir renovar o certificado passando pelo processo de avaliação novamente (Nahuz, 1995).

Como a certificação é um processo voluntário, cabe a uma organização decidir se pretende adquirir um determinado certificado. As motivações para obter um dado certificado podem ser tanto internas como externas, sendo que as primeiras estão relacionadas especialmente com o facto de a empresa ter uma estratégia de evolução e melhoria contínua ou outros objectivos e políticas organizacionais. As motivações externas estão relacionadas com a pressão feita pela concorrência para a empresa se colocar a par das exigências do mercado e dos clientes.

Há várias vantagens que advêm da obtenção do certificado, tanto a nível interno como a nível externo. Porém, estas podem variar consoante a área da organização. A nível interno, a organização apresenta uma melhoria da produtividade, sendo importante salientar que normalmente também ocorre uma significativa melhoria do funcionamento da empresa com a correspondente diminuição das reclamações, o que aumenta a satisfação dos clientes e reduz custos e desperdícios. A nível externo, a organização projecta uma melhor imagem no mercado, em termos de qualidade, proporcionando uma melhor confiança a todos os interessados, como, por exemplo, actuais ou potenciais clientes (Sampaio, et al., 2008). Apesar das vantagens enumeradas, o processo de certificação apresenta algumas dificuldades para as empresas. Geralmente a falta de uma equipa na empresa especializada em processos de certificação, torna este processo muitas vezes difícil para algumas empresas. Após a aquisição do certificado, a certificação revela-se um processo custoso, demorado e complexo (Sampaio, et al., 2008; Douglas, et al., 2003).

Os certificados possuem um conjunto de requisitos que as empresas devem observar com o principal objectivo de melhorarem os seus sistemas de gestão. Existem vários tipos de certificados que as empresas podem adquirir baseando-se nos sistemas de gestão que possuem e que

pretendem melhorar. Há certificados para melhorar o sistema de gestão de qualidade, o sistema de gestão ambiental, o sistema de gestão de saúde e segurança ocupacional, o sistema de gestão de segurança alimentar, etc.

As empresas podem adquirir vários certificados, baseando a sua escolha essencialmente no sector de actividade em que está inserida. Os certificados mais adoptados pelas empresas, geralmente são os correspondentes às três principais normas, nomeadamente ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 – estas normas serão apresentadas de seguida. Os restantes certificados apresentados não possuem tanta importância para os restantes sectores de actividade existentes. Por isso, as empresas usualmente só os escolhem consoante o sector de actividade em que estão inseridas e a região em que a empresa se encontra localizada, como as normas ISO 22000:2005, EMAS II, HACCP, ISO/TS 16949, NP4457/IDI, NP 4469-1:2008 e SA 8000 (Sampaio & Saraiva, 2008).

2.2 Certificação de Sistemas de Gestão de Qualidade

O primeiro sistema de gestão a ser implementado foi o da qualidade, designado por norma NP EN ISO 9001:2000 (*International Organization for Standardization*). Este certificado reconhece o esforço da organização em assegurar a qualidade dos seus produtos e serviços, a satisfação dos seus clientes e a melhoria contínua do desempenho da empresa no mercado. É conhecido por ser o mais adoptado pelas empresas e muitas vezes visto como um requisito essencial pelos futuros clientes, sendo assim difícil a uma empresa sobreviver no mercado se não possuir este certificado. Esta norma fornece requisitos para o Sistema de Gestão de Qualidade (SGQ) de uma empresa, e indica a forma de obter melhorias em diversas áreas. Estas melhorias são fortemente verificadas em pequenas e médias empresas, ou em companhias que já estão certificadas por um longo período de tempo assegurando assim um sistema adequado, proporcionando uma maior notoriedade e melhorando a sua imagem (Gotzamani & Tsiotras, 2001; Ohashi & Melhado, 2004). As motivações para adquirir este certificado variam de área para área e a principal decisão de o adquirir cabe, obviamente, à empresa. Essencialmente, as empresas necessitam deste certificado para ajudar a melhorar os serviços e corresponder às exigências feitas pelos clientes, melhorar o

comportamento e eficiência da empresa, evoluindo no mercado, melhorando a imagem e mantendo uma melhoria contínua da qualidade (Casadesús, et al., 2001; Douglas, et al., 2003; Ohashi & Melhado, 2004).

As motivações tanto podem ser internas como externas. As primeiras são claros incentivos para a empresa atingir uma melhoria contínua, otimizando o desempenho da organização e a qualidade dos produtos e dos processos, melhorando a definição de procedimentos do trabalho e a definição de responsabilidade e obrigações dos trabalhadores. Todos estes factores irão permitir à empresa uma melhoria de procedimentos da produção dos produtos na empresa, reduzindo o imprevisto, que como se sabe pode levar a maus resultados, como por exemplo, acidentes ou desperdícios de matéria-prima, implicando maiores custos (Casadesús, et al., 2001). As motivações externas são relacionadas com a pressão feita pelos clientes e pelo mercado competitivo.

Após a aquisição da norma ISO 9001:2000 verifica-se uma melhor resposta às exigências dos clientes, melhorando o serviço prestado e conseqüentemente, o relacionamento com os clientes. Também se verifica um aumento da satisfação dos clientes e uma redução das reclamações por parte dos mesmos. O reconhecimento por parte dos clientes vai permitir às empresas penetração em novos mercados (Casadesús, et al., 2001; Douglas, et al., 2003).

As principais vantagens associadas à implementação de um sistema de gestão de qualidade são, como já referido, uma melhoria contínua da qualidade dos produtos e dos processos realizados pela empresa. Para além disso, a aquisição da norma também ajuda os trabalhadores a ter uma melhor percepção do trabalho, das suas responsabilidades e obrigações e do modo como estes se enquadram na actividade da empresa, e conseqüentemente a melhorar o seu desempenho. Outra das vantagens da aquisição é o aumento da satisfação dos clientes, uma melhoria contínua da qualidade do produto final e a obtenção da melhor relação (Corbett & Kirsch, 2000; Ohashi & Melhado, 2004; Sampaio, et al., 2008; Silva, et al., 1997).

O principal obstáculo da certificação, que ocorre normalmente em qualquer certificado, é o seu elevado custo de implementação. Efectuar um melhor controlo no processo de gestão de qualidade e aumentar a burocracia a ele associada traduzem-se num aumento de custos. Muitas empresas não possuem pessoal especializado para a implementação de uma norma e, no final, este processo

mostra ser demasiado complexo, custoso e demorado, e, além disso, pode não se desenvolver como esperado (Casadesús, et al., 2001; Douglas, et al., 2003; Sampaio, et al., 2008; Terlaak & King, 2006).

Apesar das inúmeras vantagens, a norma ISO 9001:2000 foi alvo de bastantes críticas. Uma das principais críticas é a de que, em alguns aspectos, a norma não valoriza nem beneficia algumas empresas, simplesmente pelo facto de ela não forçar as empresas a adoptar um padrão de comportamento de boas práticas, ou seja, o conjunto de regras definidas pela norma ISO 9001:2000 não é específica para as empresas de um determinado sector. De facto, a escolha do conjunto de regras que as empresas decidem seguir para atingir a qualidade exigida pela norma é da responsabilidade das mesmas (Douglas, et al., 2003; Seddon, 2001).

Como a certificação é um processo demorado e custoso, é aconselhável que as empresas só se certifiquem se os clientes assim o exigirem. Apesar disto, não se verificam grandes aumentos de lucros após a certificação, apesar do dinheiro investido na certificação ser facilmente recuperado nos anos que se seguem à certificação (Heras, et al., 2002).

Devido à grande importância deste certificado, já foram realizados vários estudos às empresas portuguesas, analisando o seu comportamento e desempenho após a aquisição desta norma. Furtado (2002) e Ribeiro (2006, 2007) são responsáveis por alguns dos estudos feitos às empresas portuguesas, com o propósito de analisar o desenvolvimento dos seus negócios e a evolução dos seus comportamentos e das possíveis alterações, antes e após a aquisição da norma ISO 9001:2000. As conclusões gerais retiradas desses dois estudos não divergem muito do esperado e escrito anteriormente nesta dissertação. Foi possível verificar que as empresas melhoraram o desempenho ao nível da evolução do volume de negócios (aumento das vendas) e ao nível da produtividade dos materiais e produtos da empresa (redução dos produtos defeituosos e desperdícios, com melhor aproveitamento dos materiais).

No entanto, Furtado (2002), no inquérito que realizou a empresas nacionais, verificou um pior desempenho a nível da produtividade de trabalho após a aquisição da norma, com um aumento de tempos não produtivos na empresa durante um período de adaptação. Furtado (2002) chegou a algumas conclusões sobre os motivos que levariam a este comportamento por parte das empresas.

Em primeiro lugar, uma maior exigência feita à empresa na fase inicial da certificação, como resultado de uma melhor definição de procedimentos e responsabilidades, e a necessidade dos trabalhadores, não habituados aos novos procedimentos, se adaptarem a eles. Em segundo lugar, a necessidade de uma reformulação da definição dos processos, em resultado de as empresas aderirem de uma forma superficial ao processo de certificação, visando obter o certificado rapidamente, sem proceder à avaliação necessária e correcta da definição dos processos.

Ribeiro (2006, 2007) teve como base o estudo feito por Furtado (2002). Com o seu estudo, Ribeiro chegou à conclusão de que a aquisição da norma não tem um efeito positivo nos primeiros anos da certificação, efeito este que se fica a dever às alterações de processos e de definições dos procedimentos das empresas, que são desencadeadas pelo cumprimento das regras impostas pela norma ISO 9001:2000. Verificou também que a maior parte dos efeitos positivos que provêm da norma ISO 9001:2000 só surgem após alguns anos da certificação, sendo provocados pelo cumprimento das regras impostas pela norma e à estabilização do desempenho das empresas. Esta variação pode depender também de características específicas do sector de actividade em que a empresa está inserida (Ribeiro 2006, 2007).

2.3 Certificação de Sistemas de Gestão Ambiental

A protecção ambiental é uma das preocupações tanto dos clientes como das organizações e da população em geral. Para adquirir um sistema de gestão ambiental eficaz, segundo a norma NP EN ISO 14001:2004, a empresa deve estar em condições de controlar os efeitos ambientais resultantes dos processos da empresa, como a poluição e os desperdícios, e consequentemente diminuir o impacto ambiental negativo.

A norma ISO 14001:2004 define requisitos para orientar a empresa num Sistema de Gestão Ambiental (SGA) eficaz, obedecendo a compromissos com a melhoria contínua dos procedimentos da empresa e com a protecção ambiental (Corbett & Kirsch, 2000; Delmas, 2001), tendo em especial atenção a prevenção da poluição e a diminuição dos danos ao meio ambiente, proporcionando um desempenho ambiental melhorado. Como estes factores são uma preocupação

para qualquer organização, a obtenção deste certificado é uma mais-valia, que assegura a confiança por parte das entidades envolvidas, como clientes, fornecedores e sociedade em geral (Darnall, et al., 2000; Santos, 2007).

Existe um procedimento específico que uma empresa tem que seguir com vista a obter um sistema de gestão ambiental eficaz (Darnall, et al., 2000):

- Adoptar uma política ambiental escrita.
- Identificar os impactos ambientais dos procedimentos da empresa.
- Definir prioridades, objectivos para o compromisso e melhoria contínua no desempenho ambiental.
- Atribuir responsabilidades para a implementação, treino, monitorização e acções correctivas.
- Avaliar e aperfeiçoar a implementação ao longo do tempo.

Surgem várias vantagens para a empresa da aquisição da norma ISO 14001:2004, sendo estas vantagens muito semelhantes às provenientes da aquisição da norma ISO 9001:2000: melhoria contínua do desempenho e da produtividade. À semelhança da norma ISO 9001:2000, verifica-se uma melhoria nos relacionamentos com os clientes e público em geral, devido ao compromisso com o ambiente, possibilitando uma vantagem competitiva à empresa e a penetração em novos mercados. Desta forma aumenta a confiança e satisfação de clientes actuais e futuros, e de possíveis investidores, bem como contribui para a melhoria da imagem da empresa no mercado. Com a optimização dos procedimentos da empresa, ocorre naturalmente uma redução dos desperdícios resultantes de processos realizados na empresa e a consequente redução de custos. Ao longo dos anos, verifica-se o aumento do lucro associado à poupança derivada do cumprimento dos requisitos da norma ISO 14001:2004 (Corbett & Kirsch, 2000; Darnall, et al., 2000; Maier, 2005; Rivera-Camino, 2001; Strachan, et al., 2003).

As desvantagens mais relevantes da implementação deste certificado são as mesmas que as da norma ISO 9001:2000. Trata-se também de um processo complexo, demorado e custoso, devido à dificuldade e má comunicação que pode surgir entre as entidades responsáveis, em relação à implementação da norma e à burocracia associada à aquisição da norma. Outra desvantagem são

os custos de manutenção anual. As principais dificuldades que surgem da implementação da norma ISO 14001:2004 são a dificuldade em determinar com eficácia os impactos ambientais provocados pelos procedimentos da empresa e a dificuldade em traduzir os requisitos e as regras impostas pela norma ISO 14001:2004 em ações. Apesar de só surgirem lucros a longo prazo associados ao sucesso, a empresa obtém algumas vantagens competitivas devido ao reconhecimento por parte dos clientes e público em geral, melhorando assim as vendas dos seus produtos e reduzindo os seus gastos. (Corbett & Kirsch, 2000; Strachan, et al., 2003; Darnall, et al., 2000; Darnall, 2001; Mohammed, 2000).

2.4 Certificação de Sistemas da Segurança e Saúde no Trabalho

A segurança e saúde no trabalho é um tema cada vez mais abordado e de grande preocupação das empresas, pois, como refere Araújo (2002): "a taxa de acidentes de trabalho representa perdas consideráveis tanto a nível económico, como a nível social, e estes acidentes normalmente resultantes de um ambiente de trabalho onde estão presentes riscos físicos, químicos, biológicos, ergonómicos e de acidentes." As empresas sabem que o controlo e a monitorização da segurança e a saúde ocupacional proporcionam condições para melhorar a produtividade e o funcionamento da empresa, de forma segura para os seus trabalhadores. A preocupação das empresas passa então, essencialmente, por diminuir o número de acidentes que surgem no trabalho protegendo os trabalhadores e os rendimentos da empresa.

A certificação OHSAS 18001:1999 (*Occupational Health and Safety Assessment Series*) fornece requisitos a um sistema de gestão de segurança e saúde ocupacional (SGSSO) eficaz, de forma a melhorar o desempenho e a eficácia das empresas inseridas em sectores de actividade em que os trabalhadores estão sujeitos a acidentes. Com um SGSSO eficaz, as empresas controlam melhor os riscos de acidentes no local de trabalho, possuindo como finalidade específica, a implementação, manutenção e melhoria contínua do SGSSO, além de assegurar conformidade com a política definida e demonstrar tal conformidade a terceiros (Romano, 2006).

Esta norma permite demonstrar o compromisso da empresa com a Segurança, Higiene e Saúde no Trabalho. Esta certificação permite também condições que permitem à organização: monitorizar, melhorar a saúde dos empregadores e a segurança no trabalho, permitindo minimizar os riscos de acidentes. Esta norma é totalmente compatível com as normas apresentadas anteriormente - a norma ISO 9001:2000 e 14001:2004 -, podendo estas serem integradas num único sistema de gestão (Bobsin, 2005; Neves, 2007).

Com este certificado, a empresa possui as ferramentas necessárias para controlar os riscos e reduzir os acidentes de trabalho de forma preventiva. Isso permite à empresa:

- reduzir custos, melhorando a produtividade da empresa e o desempenho organizacional, e consequentemente os seus lucros;
- minimizar ou mesmo eliminar os riscos de acidentes e doenças no trabalho;
- aumentar a satisfação do cliente e reconhecimento por parte de terceiros, permitindo assim penetração em novos mercados.

Com a implementação destas medidas, a empresa melhora a sua imagem no mercado, melhorando a confiança do cliente e da própria companhia. As vantagens apresentadas variam dependendo de vários factores, como o sector de actividade, a política da empresa, a complexidade dos procedimentos, entre outros (Araújo, 2002; Bobsin, 2005; Costella, 2008).

Apesar das diversas vantagens obtidas da certificação da norma OHSAS 18001:1999, este processo apresenta algumas dificuldades. Estas são bastante acentuadas em pequenas empresas, pois as empresas deste tipo são as que geralmente apresentam menos recursos para a implementação de certificações, nomeadamente (Costella, 2008):

- a falta de familiaridade com os sistema de gestão, o que pode tornar este processo bastante penoso;
- a escassez de recursos e falta de pessoal responsável para a gestão de segurança;
- o elevado custo para a correcta implementação da norma OHSAS 18001:1999, bem como o tempo consumido pela empresa e a burocracia envolvida neste processo.

2.5 Outros Certificados

Nesta secção serão apresentados os certificados considerados menos importantes devido à sua fraca aquisição por parte das empresas. Geralmente só são adoptados por empresas que estão localizadas em regiões em que o sector de actividade se encontra bastante evoluído, isto é, sectores de actividade que possuem um grande número de empresas e, conseqüentemente, um conjunto maior de certificados em relação a outras regiões onde a competitividade entre estas não é tão elevada.

2.5.1 Verificação EMAS II

O certificado EMAS II (*Eco-Management and Audit Scheme* ou), ou Sistema Comunitário de Eco Gestão e Auditoria é um sistema aplicável a qualquer organização que pretende melhorar o seu compromisso com a protecção ambiental, sendo totalmente compatível com a norma ISO 14001:2004. Apesar de ter princípios bastante semelhantes com a norma ISO 14001:2004, existem vários factores que diferenciam a Verificação EMAS II da norma ISO 14001:2004. A principal diferença é o facto da verificação EMAS II ser principalmente aplicada a nível de locais da empresa e não a nível da organização ou empresa, como acontece com a norma ISO 14001:2004. Por outras palavras, as duas normas possuem diferentes objectivos, a norma ISO 14001:2004 fornece à empresa requisitos para serem implementados de forma a que ocorra uma melhoria do desempenho e da gestão ambiental da empresa, enquanto a Verificação EMAS II, por outro lado, foi desenvolvida para introduzir mudanças no desempenho ambiental (Morrow & Rondinelli, 2002).

2.5.2 Certificação HACCP - Segurança Alimentar

A certificação HACCP (*Hazard Analysis Critical Control Point*) ou APPCC (sistema de análise de perigos e pontos críticos de controle) é vista como uma metodologia de gestão de segurança alimentar. Esta certificação contém vários requisitos para qualquer organização que pretenda

demonstrar um compromisso voluntário com a higiene dos géneros alimentícios e promove a adopção de boas práticas de gestão da segurança alimentar, garantindo a segurança dos produtos (Garcia, 2000). É considerado um sistema baseado em conceitos preventivos, o que significa que todas as etapas de preparação do alimento são controladas e monitorizadas (Garcia, 2000). Esta norma é bastante vantajosa para as empresas, uma vez que a realização de um controlo final da qualidade dos alimentos é bastante difícil para algumas empresas devido a vários factores como, por exemplo, o alto custo e a demora da obtenção dos resultados de análises feitas aos produtos (Garcia, 2000). A obtenção deste certificado irá certamente aumentar a confiança dos clientes, na medida em que, ao implementar este sistema, a empresa reduz significativamente os riscos para os consumidores (Magalhães, 2006).

2.5.3 Certificação Segurança Alimentar

A norma NP EN ISO 22000:2005 define requisitos para orientar a empresa num Sistema de Gestão de Segurança Alimentar (SGSA) eficaz, mantendo a conformidade com a metodologia HACCP, sendo ambas as normas compatíveis. Esta norma obedece a compromissos com a melhoria contínua e com a segurança alimentar (Neves, 2007), sendo aplicável a todos os sectores da cadeia alimentar, incluindo organizações que influenciem a segurança. (Magalhães, 2006; Queiroz, 2006). Como esta norma está ligada à segurança alimentar, é uma mais-valia para as empresas adquirem-na pois permitirá diminuir a possibilidade de ocorrência de riscos para os clientes, aumentando consequentemente a confiança na empresa dos mesmos (Neves, 2007).

2.5.4 Certificação de Sistemas de Gestão de Qualidade no Sector Automóvel

A norma ISO/TS 16949 foi desenvolvida tendo por base os requisitos das normas ISO 9000. Esta norma define requisitos de um sistema de gestão de qualidade, mas orientados ao sector

automóvel, obedecendo a compromissos com a melhoria contínua, como a norma ISO 9001:2000, implicando uma redução dos defeitos, riscos e recursos neste sector de actividade (Haro, 2001).

2.5.5 Certificação de Sistemas de Gestão de Investigação, Desenvolvimento e Inovação

A certificação NP 4457:2006 define requisitos para orientar uma organização num Sistema de Gestão de Investigação, Desenvolvimento e Inovação (SGIDI) mais eficaz, estabelecendo regras que contribuam para melhorar a imagem organizacional no mercado, permitindo também integrar o SGIDI com outros sistemas de gestão já implementados na organização. Pode ser implementada em qualquer tipo de organização que pretenda inovar não só o plano tecnológico, mas também outras áreas de actividade [1].

2.5.6 Certificação de Sistemas de Gestão da Responsabilidade Social

A certificação NP 4469-1:2008 especifica requisitos para orientar uma organização num Sistema de Gestão de Responsabilidade Social (SGRS). Pode ser implementado em qualquer organização independentemente da dimensão e do sector de actividade. Esta norma permite à organização uma vantagem competitiva no mercado, maior credibilidade e aumento da produtividade por parte da organização e aumento da confiança por parte dos clientes e entidades relacionadas (Rosa, 2009).

2.5.7 Certificação de Sistemas de Gestão de Responsabilidade Social

A certificação SA 8000 (*Social Accountability*) define requisitos para orientar uma organização num Sistema de Gestão de Responsabilidade Social (SGRS). A principal diferença entre esta norma e a norma NP 4469-1:2008 é o facto de a norma SA 8000 ser reconhecida internacionalmente e de se focar mais na forma como as organizações actuam no mercado, valorizando a sua dimensão social

e garantindo a componente ética. Com esta norma, a imagem da organização melhora significativamente, o que se traduz numa vantagem sobre a concorrência, a produtividade aumenta e a confiança dos clientes e das entidades envolvidas cresce [1].

2.6 Uma Nota Final sobre Certificação

Todos os certificados apresentados neste capítulo trazem variadas vantagens às empresas. Estas vantagens podem variar consoante o tamanho da empresa, sector de actividade ou região onde esta se encontra localizada. Apesar das vantagens associadas existem diversas desvantagens que impedem as empresas de adquirir todos os certificados existentes. A burocracia associada e o elevado custo do processo de implementação do certificado são duas das importantes razões que tornam difícil a certas empresas adquirir certificados e assim conseguirem obter as diversas vantagens competitivas.

A norma de Certificação de SGQ (ISO 9001:2000) é sem dúvida a norma que os clientes consideram "obrigatória" nas empresas. A reputação desta norma deve-se ao facto de fazer com que a empresa que quer adquirir o certificado, siga um conjunto de normas que demonstrem o compromisso da empresa com a qualidade dos produtos e dos serviços, e uma melhoria contínua. Desta forma é a que mais agrada aos clientes, sendo as empresas certificadas com a norma ISO 9001:2000 as principais eleitas pelos clientes no mercado.

As outras duas normas, norma de Certificação de SGA (ISO 14001:2004) e norma de Certificação de SGSSO (OHSAS 18001:1999) são duas normas que são muito requisitadas pelas empresas, na sua maioria por exigência de certos clientes. Dois casos em que é necessário possuírem estes dois tipos de normas são por exemplo no caso de empresas inseridas em sectores de actividade que geralmente causam poluição ambiental (ex: fábricas têxtil) e empresas que possuem actividade que envolve mão-de-obra humana e pode envolver os empregados em acidentes (ex: construção).

As outras normas referidas ainda são ligeiramente recentes e a adesão por parte da maioria das empresas ainda não é grande. Apenas as empresas que precisam de determinadas normas, as adquirem devido à competição dentro do sector de actividade ou da região que assim o exige.

Capítulo 3

O Sistema de Dados Alvo

3.1 Certificação e Informação

A certificação das empresas é cada vez mais um tema abordado pelos órgãos especializados, existindo cada vez mais informação, recolhida e divulgada por organismos certificados como a entidade CemPalavras - Comunicação Empresarial Lda [2], com o especial objectivo de informar as empresas nacionais do estado dos sistemas de gestão das empresas nacionais. Este aspecto assume particular relevância no ambiente que se vive actualmente de crise financeira mundial. Sampaio e Saraiva contribuíram com um estudo para o cempalavras.pt [2] que consistiu na apresentação de informação sobre os sistemas de certificação e a sua evolução ao longo dos anos. Essa informação foi fornecida por várias entidades, como o *ISO Survey*, INE e organismos certificadores. Estas organizações possuem toda a informação sobre os sistemas gestão das empresas de todo o mundo, bem como informação sobre a evolução da certificação ao longo dos anos. Estes dados contêm detalhes sobre as empresas certificadas em Portugal e permitem uma visualização e uma análise da evolução da certificação e dos sistemas de gestão nas empresas nacionais, quer por região, quer por sector de actividade.

O trabalho que se está a desenvolver nesta dissertação é um complemento à análise dos resultados do estudo de Sampaio e Saraiva, que tinha como principal finalidade a análise da evolução do número de empresas e do número de certificados, ao longo dos três anos para os quais foram recolhidos dados. A título de exemplo, foi analisada a evolução dos três principais certificados (ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999), bem como a evolução de cada um dos certificados por região (Figuras 1, 2, 3 e 4).

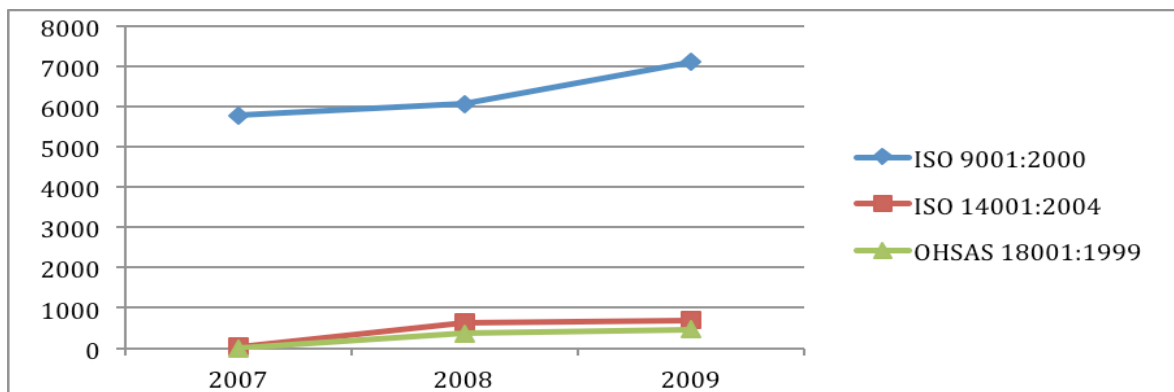


Figura 1 - Número de Empresas Certificadas por Referencial

No gráfico apresentado na Figura 1 é possível verificar a importância da norma ISO 9001:2000 em relação às outras duas normas. A adesão deste certificado por parte das empresas sempre foi bastante elevada devido à importância dada pelos clientes a esta norma, empresas certificadas com esta norma demonstra que estas possuem um bom sistema de gestão de qualidade implementado e são sempre a prioridade por parte dos clientes. Por esta razão existe uma grande motivação para as empresas para a aquisição desta norma.

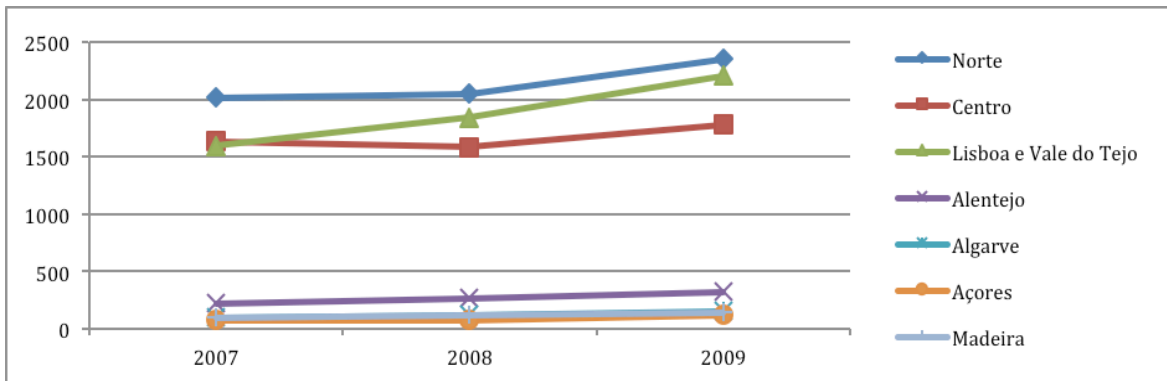


Figura 2 - Número de Empresas Certificadas com a Norma ISO 9001:2000 implementada por Região

No gráfico representado na Figura 2 é possível verificar que as regiões Norte, Lisboa e Vale do Tejo e Centro apresentam o maior número de empresas certificadas no país. Estas três regiões possuem assim a maior concentração de empresas e conseqüentemente são as regiões que apresentam a maior competitividade no mercado nacional.

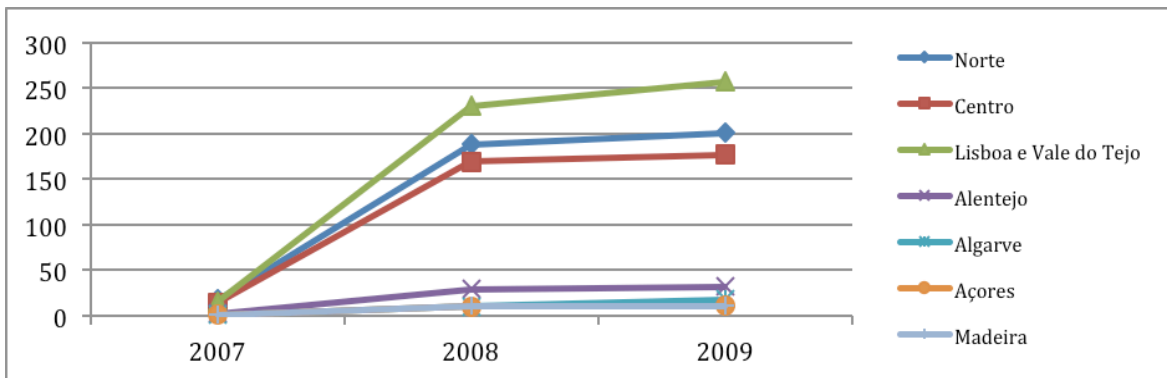


Figura 3 - Número de Empresas Certificadas com a Norma ISO 14001:2004 implementada por Região

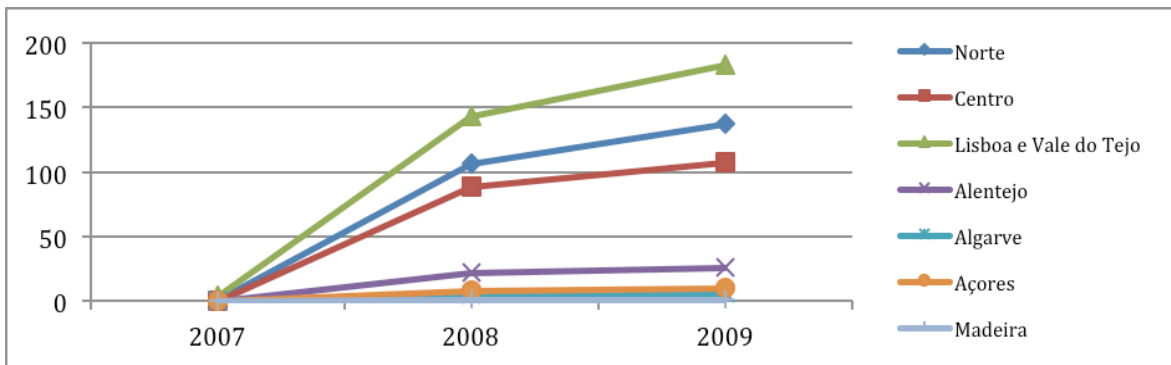


Figura 4 - Número de Empresas Certificadas com a Norma OHSAS 18001:1999 implementada por Região

Nas Figuras 3 e 4 é possível observar a grande adesão por parte das empresas na aquisição das normas ISO 14001:2004 e OHSAS 18001:1999, entre 2007 e 2008 especialmente das regiões que apresentam a maior quantidade de empresas, como a região Lisboa e Vale do Tejo, Norte e Centro. Uma das justificações para tal acontecimento é o facto de se começarem a notar os primeiros sinais de crise e as empresas decidirem adquirir mais certificados com o principal objectivo de agradar aos clientes.

Neste estudo, não foi feita uma análise da evolução ao longo do tempo, tendo os dados respeitantes aos três anos sido agregados para fazer uma análise conjunta. Dado que os dois estudos têm perspectivas diferentes, mas complementares, em conjunto eles fornecem uma melhor e mais precisa análise do mercado.

A análise feita com os resultados do estudo e da mineração de dados permitirá ajudar o cliente a tomar decisões sobre o futuro de uma nova empresa, tais como, por exemplo:

- i) Qual a melhor região para iniciar um negócio?
- ii) Qual o melhor sector de negócio, tendo em conta o número de empresas desse sector?
- iii) Quais os melhores certificados a obter nesse mesmo sector e nessa região?

Esta última questão é considerada a mais importante, pois o conjunto de certificados a adquirir pela empresa varia com o sector de actividade. A título ilustrativo, as empresas na área da

alimentação deverão possuir normas como as HACCP ou ISO 22000:2005, pois estas são referenciais que contêm requisitos para melhorar o Sistema de Gestão da Segurança Alimentar da organização.

3.2 O Modelo de Dados

Os dados disponíveis para este trabalho contêm informação recolhida desde 2008 a 2010 sobre os sistemas de gestão das empresas nacionais certificadas. Contendo um total de 6577, 6772 e 7432 registos em cada um dos ficheiros, e informações sobre as empresas como nome, localização, EAC (código de actividade económica), representando o sector de actividade em que a empresa está inserida e os certificados que possuem ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP, ISO/TS 16949, NP4457/IDI, NP 4469-1 e SA 8000 (Sampaio & Saraiva, 2008).

Sampaio e Saraiva desenvolveram um estudo bastante profundo ao longo dos anos, e juntando a informação das bases de dados é possível visualizar a evolução dos sistemas de gestão mais implementados pelas empresas nacionais (Sistema de Gestão de Qualidade, Sistema de Gestão Ambiental e Sistema de Gestão de Segurança e Saúde no Trabalho) e no intervalo de tempo considerado, quer por sector de actividade quer por região. O objectivo desse estudo foi demonstrar o estado do mercado nacional, com especial atenção aos sistemas de gestão adquiridos pelas empresas nacionais. Estes dados foram fornecidos por entidades como o *ISO Survey* e o INE.

Os ficheiros com os dados analisados continham a informação de algumas empresas nacionais, correspondendo a um total de 20 285 empresas. Os dados das empresas contêm elementos como: o nome da empresa; se tem na sua posse um determinado certificado; a localização da empresa segundo a agregação NUTS II e o código EAC da empresa, como é possível observar na Tabela 2. O código EAC consiste num código atribuído a cada sector de actividade, denominado por código de actividade económica (CAE), ou *Economic Activity Code* (EAC).

Tabela 2 - Conjunto de Dados Inicial

Atributo	Descrição	Tipo de Dados	Domínio	Nulo
Nome	Nome da empresa	Texto	Tamanho variável (até 100 caracteres)	Não
ISO 9001	Certificação de Sistemas de Gestão de Qualidade (NP EN ISO 9001:2000)	Texto	Possibilidade: {TRUE/FALSE}	Não
ISO 14001	Certificação de Sistemas de Gestão Ambiental (NP EN ISO 14001:2004)	Texto	Possibilidade: {TRUE/FALSE}	Não
OHSAS 18001	Certificação de Sistemas da Segurança e Saúde no Trabalho (OHSAS 18001:1999)	Texto	Possibilidade: {TRUE/FALSE}	Não
ISO 22000	Certificação Segurança Alimentar (NP EN ISO 22000:2005)	Texto	Possibilidade: {TRUE/FALSE}	Não
EMAS II	Verificação EMAS II	Texto	Possibilidade: {TRUE/FALSE}	Não
HACCP	Certificação HACCP – Segurança Alimentar	Texto	Possibilidade: {TRUE/FALSE}	Não
ISO/TS	Certificação de Sistemas de Gestão de Qualidade no Sector Automóvel (ISO/TS 16949)	Texto	Possibilidade: {TRUE/FALSE}	Não
NP4457/IDI	Certificação de Sistemas de Gestão de Investigação, Desenvolvimento e Inovação (NP 4457:2006)	Texto	Possibilidade: {TRUE/FALSE}	Não
NP4469-1(2008)	Certificação de Sistemas de Gestão da Responsabilidade Social. (NP 4469-1:2008)	Texto	Possibilidade: {TRUE/FALSE}	Não
SA 8000	Certificação de Sistemas de Gestão de Responsabilidade Social (SA 8000)	Texto	Possibilidade: {TRUE/FALSE}	Não
NUTS II	Região correspondente à localização da empresa	Texto	Tamanho variável (até 10 caracteres)	Sim
EAC	Código de actividade económica da empresa	Número	Número de dois dígitos (entre 1 e 99)	Sim

3.3 Análise Exploratória dos dados

Os dados fornecidos por Sampaio & Saraiva são referentes a 20 265 empresas nacionais, e estão contidos num total de 202 971 registos. Cada registo corresponde a um sistema de gestão que uma determinada empresa decidiu ou não adquirir através do processo de certificação, pelo que,

dado que há dez certificados diferentes, seria expectável haver cerca de 202650 registos. No entanto, o número total de registos é maior, porque há registos na base de dados relativos a uma mesma empresa com actividade em mais do que uma região e porque a mesma empresa aparece, por vezes, identificada com várias designações diferentes.

Dado que, neste estudo, se fez uma análise de certificados por região, foi decidido manter a duplicação de registos de uma mesma empresa quando ela tinha actividade em mais do que uma região. No entanto, no que diz respeito à ocorrência de registos de uma mesma empresa com designações diferentes, dado o elevado número de registos não permitir uma limpeza de dados adequada, e por se considerar que essa limpeza de dados não traria alterações significativas aos resultados da análise, foi decidido não alterar as designações.

Após a devida selecção dos dados, em que apenas se seleccionaram os dados sobre empresas que contêm algum tipo de sistema de gestão, já que apenas interessavam os dados sobre empresas certificadas, obteve-se um total de 23 932 registos. As empresas nacionais analisadas encontram-se localizadas por todo o território nacional. O território nacional no conjunto de dados foi classificado segundo as sub-regiões estatísticas baseando-se na classificação NUTSII:

- Portugal Continental (Norte, Centro, Lisboa e Vale do Tejo, Alentejo e Algarve).
- Região Autónoma dos Açores.
- Região Autónoma da Madeira.

A maior parte das empresas certificadas nacionais encontra-se dividida entre a região do Norte (33%), Lisboa e Vale do Tejo (31%) e Centro (26%), correspondendo a cerca de 90% das empresas nacionais. O resto das empresas analisadas encontram-se localizadas no Alentejo (5%), Algarve (2%), Região Autónoma dos Açores (2%) e Região Autónoma da Madeira (2%). No gráfico apresentado na Figura 5 é possível ver a distribuição das empresas pelas diversas regiões nacionais.

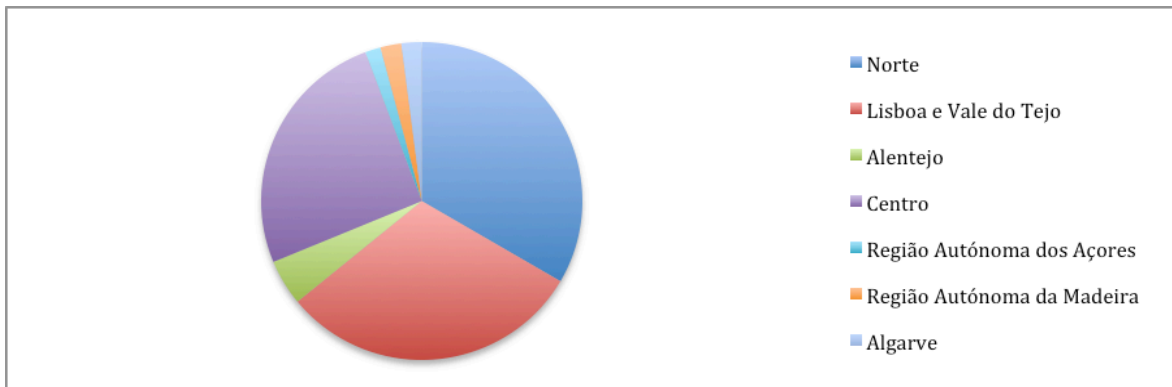


Figura 5 - Distribuição das Empresas Certificadas pelas Diversas Regiões Nacionais

As empresas nacionais encontram-se divididas por vários sectores de actividade. Estes podem ser classificados usando a Classificação Portuguesa de Actividades Económicas, Revisão 3 (CAE-Rev.3) – que pode ser consultada no Decreto-Lei nº 381/2007, de 14 de Novembro. Esta classificação agrupa as unidades de sectores de actividade segundo a actividade económica em diversas secções (exemplo: a Secção A corresponde as actividades de agricultura e pesca). No entanto, nesta dissertação, para facilitar a compreensão dos resultados, iremos seguir a terminologia usada no estudo realizado por Sampaio e Saraiva, e os sectores de actividade serão tratados de forma independente. A Figura 6 apresenta os diversos sectores de actividades representados pelo EAC e o número de empresas nacionais certificadas de cada sector. A correspondência entre o sector de actividade e o respectivo EAC podem ser consultados na Tabela 31 (Anexo I).

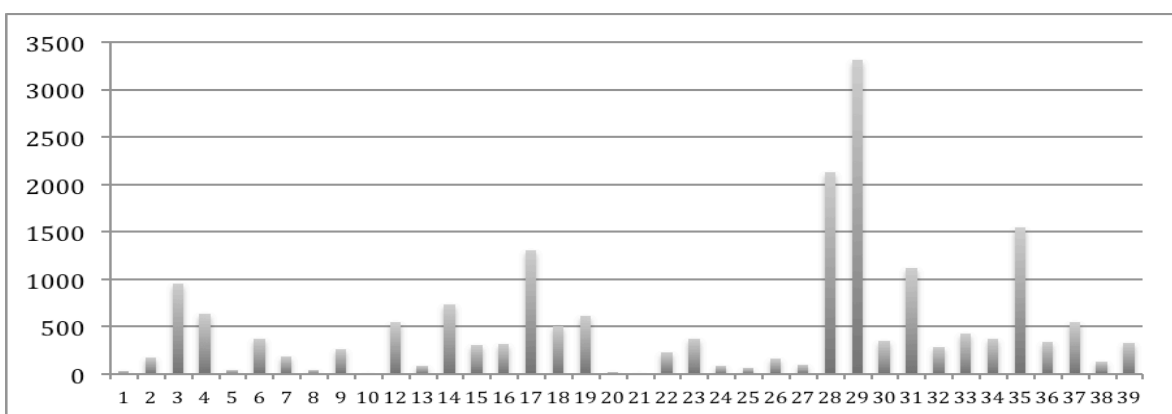


Figura 6 - Número de Empresas Nacionais Certificadas por EAC

Apresenta-se de seguida uma análise exploratória do conjunto de dados fornecido. Esta análise visa identificar e facilitar a compreensão do conteúdo da base de dados, identificando os sectores de actividades principais em cada uma das regiões do país, isto é, o top 5 dos sectores de actividade em que há mais empresas certificadas. Foi escolhido o valor de 5, visto que é suficiente para se ter uma noção dos sectores de actividade possuem o maior número de empresas certificadas. Iremos, nas secções que se seguem, designar estes sectores de actividade como sectores principais da região.

A região Norte é a zona que possui o maior número de empresas nacionais certificadas, possuindo um total de 6780 empresas. Nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 48% do total de certificados nesta região. Como se pode observar no gráfico da Figura 7, 14% de todos os certificados adquiridos pelas empresas do Norte correspondem a certificados adquiridos por empresas inseridas no sector de actividade Comércio. Neste sector, a norma ISO 9001:2000 representa a maioria dos certificados adquiridos pelas empresas, apresentando uma percentagem de 91% de todos os certificados neste sector na região Norte.

Segue-se o sector de actividade Construção, representando um total de 13% de todos os certificados nesta região. Este sector apresenta uma característica particular, possuindo uma maior percentagem de certificados da norma OHSAS 18001:1999 em relação aos outros sectores de actividade, o que acontece devido à grande importância desta norma, relativa à Segurança e Saúde no Trabalho, para as empresas neste sector.

O terceiro sector de actividade possuindo o maior número de normas nesta região é o sector Têxteis e Produtos Têxteis, que possui na sua maioria a norma ISO 9001:2000, representando um total de 78%. A norma ISO 14001:2004 e a Verificação EMAS II são certificados de grande importância para este sector, pois é do interesse das empresas neste sector assumir um compromisso com o ambiente, reduzindo os desperdícios e resíduos para o mesmo. O sector de Fabricação Metalúrgica de Base e Produtos Metálicos possui na sua maioria a norma da gestão de qualidade, seguindo-se da norma ISO 14001:2004, OHSAS 18001:1999 e dos restantes certificados.

O sector de actividade Outros Serviços possui, como os restantes sectores, a norma ISO 9001:2000 na sua maioria, representando um total de 85% de todos os certificados neste sector.

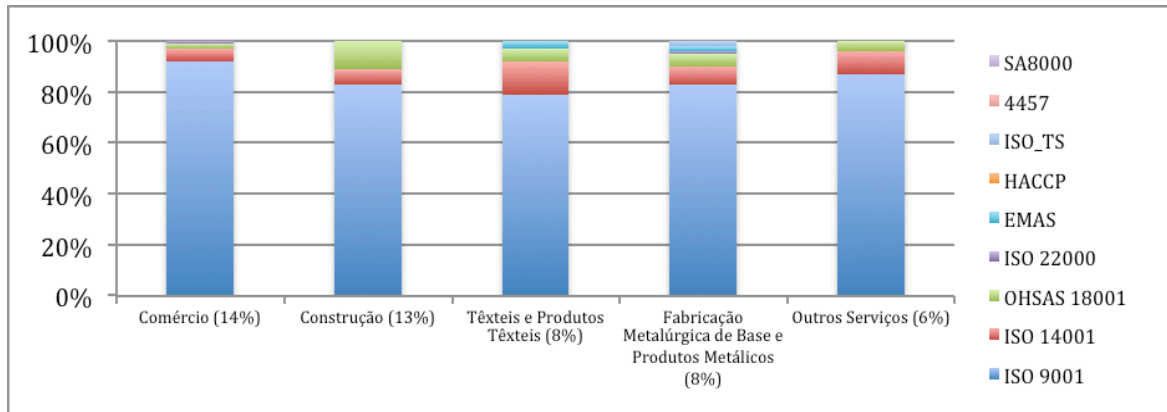


Figura 7 - Top 5 dos Sectores de Actividade da Região Norte

A região Lisboa e Vale do Tejo é a zona que possui a segunda maior quantidade de certificados adquiridos pelas empresas, possuindo um total de 7510 sistemas de gestão divididos por 6221 empresas nesta região. Como se pode observar no gráfico da Figura 8, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 53% do total de certificados nesta região. O número de sectores de actividade com o maior número de certificados nesta região difere do da região Norte: Comércio (16%), Outros Serviços (13%), Construção (10%), Transporte, Armazenamento e Comunicações (8%) e Saúde e Serviços Sociais (6%).

O sector de actividade com a maior percentagem de normas adquiridas pelas empresas é o sector de actividade Comércio, possuindo, à semelhança do que acontece com a região Norte, na sua maioria a norma ISO 9001:2000. O sector de actividade Outros Serviços possui, em relação ao sector representado na secção da região do Norte, uma maior percentagem das normas ISO 14001:2004 e OHSAS 18001:1999, demonstrando o interesse das empresas desta região pela importância da gestão ambiental e saúde e segurança ocupacional. As empresas do Sector de Transporte, Armazenamento e Comunicações demonstram também esta preocupação, possuindo principalmente estes três certificados demonstrando a preocupação com a qualidade dos seus produtos, com o ambiente e segurança e saúde ocupacionais. As empresas inseridas no sector de actividade Construção na região de Lisboa e Vale do Tejo possuem, em relação às empresas da

região Norte inseridas neste mesmo sector, um maior número de normas ISO 14001:2004 (13% contra os 6% das empresas da região Norte) e OHSAS 18001:1999 (14% contra os 11% das empresas da região Norte). O sector de actividade Saúde e Serviços Sociais, representando um total de 6% de todos os certificados adquiridos pelas empresas desta região, apresenta uma percentagem de certificados ISO 9001:2000 igual a 97%.

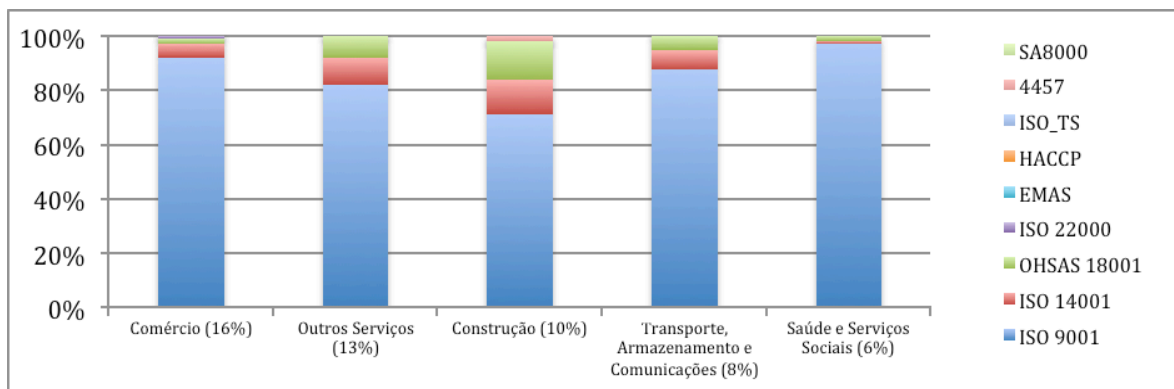


Figura 8 - Top 5 de Sectores de Actividade da Região Lisboa e Vale do Tejo

As empresas da região do Centro correspondem a um total de 6098 certificados adquiridos pelas 5181 empresas desta região. Como se pode observar no gráfico da Figura 9, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 44% do total de certificados nesta região. As empresas possuem uma classificação por sectores de actividade muito semelhante à classificação do Norte, possuindo o sector de actividade Comércio, Fabricação Metalúrgica de Base e Produtos Metálicos, Construção representados pelas empresas que adquiriram o maior número de certificados na região do Centro. Estes sectores apresentam uma disposição de certificados muito semelhante à apresentada anteriormente na secção da região do Norte.

O sector de actividade Comércio representa o sector com o maior número de empresas certificadas. À semelhança das duas regiões apresentadas anteriormente, a norma ISO 9001:2000 corresponde a uma percentagem bastante elevada (91% de todos os certificados adquiridos pelas empresas neste sector) e a norma ISO 14001:2004 representada por 4% e a norma OHSAS 18001:1999 representada por 3%. As empresas inseridas no sector de actividade Fabricação

Metalúrgica de Base e Produtos Metálicos possuem uma lista de certificados muito semelhantes à da região do Norte, possuindo ISO 9001:2000 (80%), ISO 14001:2004 (10%), OHSAS 18001:1999 (4%), EMAS II (<1%), ISO/TS 16949 (4%) e NP 4457 (<1%), apesar destes últimos apresentarem percentagens mais baixas em relação aos da região Norte.

O sector de actividade Construção possui uma disposição de certificados mais semelhante à da região Norte que em relação à da região de Lisboa e Vale do Tejo, possuindo a percentagem de normas OHSAS 18001:1999 (8%) e ISO 14001:2004 (6%) bastante menor que as adquiridas pelas empresas da região de Lisboa e Vale do Tejo, apresentando desta forma, uma distribuição semelhante à apresentada pelas empresas da região do Norte, apresentando 11% e 6% respectivamente.

As empresas inseridas no sector de actividade Fabricação de Artigos de Borracha e Matérias Plásticas, apresentam na sua maioria (83%) a norma ISO 9001:2000, os restantes 17% são representados pelos vários certificados: ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II e ISO/TS 16949. O sector de actividade Saúde e Serviços Sociais apresenta uma percentagem de ISO 9001:2000 de 99%.

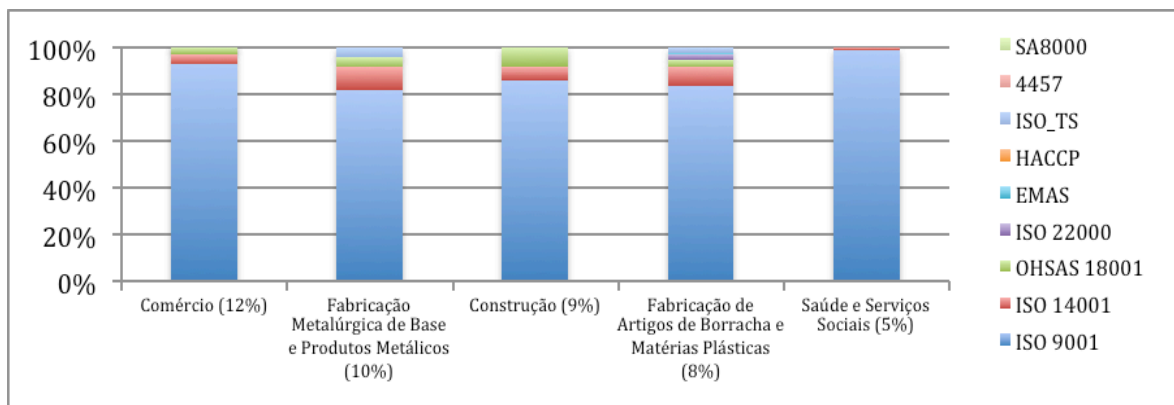


Figura 9 - Top 5 dos Sectores de Actividade da Região Centro

As regiões que serão apresentadas seguidamente: Alentejo, Algarve, e Açores e Madeira, são as zonas que possuem o menor número de empresas nacionais certificadas, e em que o número e o

tipo de certificados que estas possuem diminui bastante em relação às regiões descritas anteriormente.

A região Alentejo possui 1137 certificados adquiridos pelas 944 empresas distribuídas por vários sectores de actividade. Como se pode observar no gráfico da Figura 10, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 51% do total de certificados nesta região.

Esta região apresenta um sector de actividade que ainda não foi referido nas regiões descritas anteriormente: o sector de actividade Alimentação, Bebidas e Tabaco, que representa cerca de 12% de todos os certificados adquiridos pelas empresas nesta região. Este sector de actividade tem a particularidade de não apresentar uma percentagem da norma ISO 9001:2000 tão elevada como no resto dos sectores de actividade e possuir uma grande percentagem da norma ISO 22000:2005 e HACCP em relação às outras normas. Isto deve-se ao facto destas duas normas estarem associadas ao compromisso da empresa com a gestão alimentar, demonstrando a preocupação e o interesse pelos clientes e pela qualidade dos produtos da empresa.

Os restantes sectores apresentados no gráfico correspondente à região do Alentejo não apresentam nenhuma particularidade diferente das já referidas anteriormente. As empresas inseridas no sector de actividade Comércio possuem na sua maioria a norma ISO 9001:2000 correspondendo um total de 98% de todos os certificados adquiridos pelas empresas deste sector; o sector de actividade Construção apresenta uma elevada percentagem, em relação aos outros sectores, da norma OHSAS 18001:1999 correspondendo a uma percentagem de 10%; o sector de Saúde e Serviços Sociais apresenta uma percentagem de 97% de certificados ISO 9001:2000 e o sector de actividade de Fabricação Metalúrgica e Produtos Metálicos apresenta uma percentagem de 89% de certificados ISO 9001:2000 e os restantes 11% estão distribuídos pelas normas ISO 14001:2004 (5%) e ISO/TS 16949 (6%).

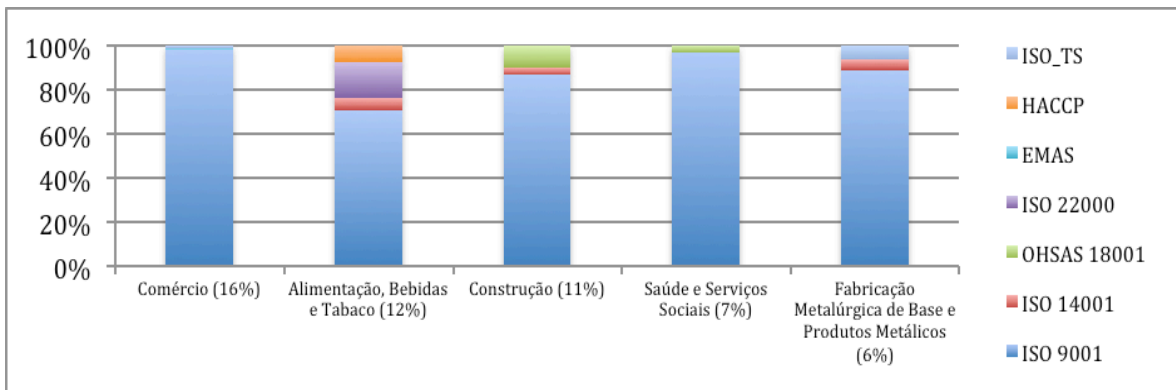


Figura 10 - Top 5 dos Sectores de Actividade da Região Alentejo

A região do Algarve é representada por 474 certificados adquiridos pelas 414 empresas nesta região. Como se pode observar no gráfico da Figura 11, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 64% do total de certificados nesta região. Os sectores de actividade que possuem o maior número de certificados adquiridos pelas empresas nesta região são Comércio (23%), Construção (20%), Hotéis e Restaurantes (8%), Saúde e Serviços Sociais (7%) e Outros Serviços (6%). Como se pode observar no gráfico da Figura 11, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 44% do total de certificados. O sector de actividade Hotéis e Restaurantes apresentado nessa classificação, ainda não foi mencionado anteriormente, evidenciando a importância do turismo na região do Algarve.

Estas empresas apresentam uma maioria de norma ISO 9001:2000, como no resto dos sectores de actividade, apresentando um total de 74%. Os restantes 26% estão divididos pelas normas ISO 14001:2004 (16%) e HACCP (10%). As empresas desta região inseridas no sector de actividade Comércio, correspondendo a um total de 23% de todos os certificados adquiridos pelas empresas desta região, demonstram a sua preocupação adquirindo os diversos certificados como ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005 em relação às empresas das outras regiões, possuindo a norma ISO 9001:2000, desta forma uma percentagem mais baixa que a das outras regiões, apresentando um total de 83% de certificados ISO 9001:2000. As empresas inseridas no sector de actividade Saúde e Serviços Sociais apresentam, em semelhança às empresas inseridas neste sector de outras regiões, uma percentagem exageradamente elevada da norma 9001

apresentando um total de 97%. Ao contrário das regiões Norte e Lisboa e Vale do Tejo, em que as empresas destas regiões inseridas no sector de actividades Outros Serviços apresentam os três certificados principais (ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999), na região do Algarve as empresas apenas apresentam os dois principais certificados: ISO 9001:2000 (86%) e ISO 14001:2004 (14%).

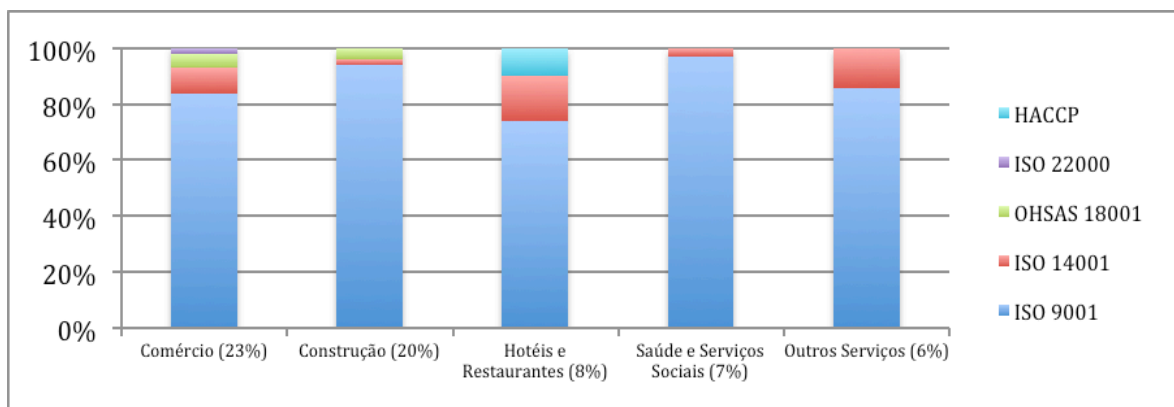


Figura 11 - Top 5 dos Sectores de Actividade da Região Algarve

Na região da Madeira, as 427 empresas possuem ao todo 457 certificados. Como se pode observar no gráfico da Figura 12, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 50% do total de certificados. Os sectores de actividade que possuem o maior número de certificados são: Comércio (20%), Alimentação, Bebidas e Tabaco (9%), Construção (9%), Outros Serviços (8%) e Administração Pública (5%). Estes sectores de actividade apresentam o menor número de variedade de normas em relação às outras regiões apresentadas: as empresas inseridas no sector de actividade Outros Serviços apresentam apenas os dois principais certificados: ISO 9001:2000 com 80% e ISO 14001:2004 com 20%; as empresas inseridas nos sectores de actividade Construção e Administração Pública apenas apresentam a norma ISO 9001:2000 e os sectores de actividade Comércio e Alimentação, Bebidas e Tabaco apresenta as normas ISO 9001:2000, ISO 14001:2004 e HACCP.

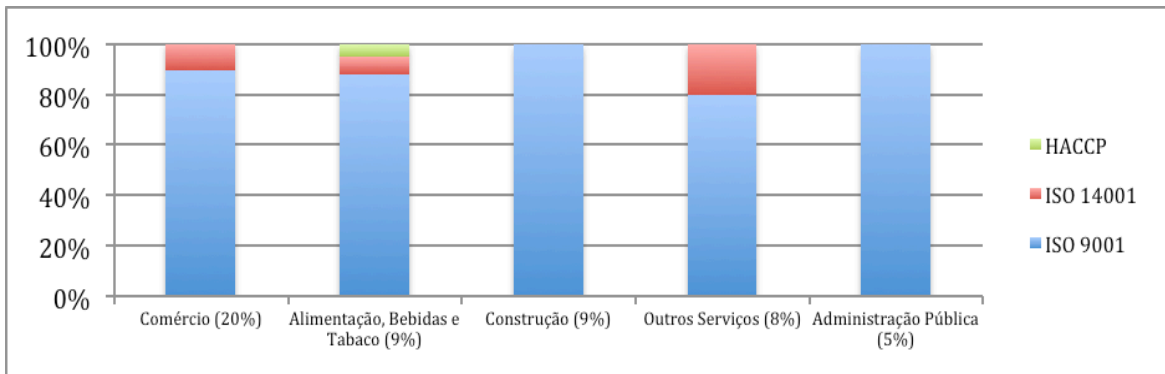


Figura 12 - Top 5 dos Sectores de Actividade da Região Madeira

As 318 empresas da região dos Açores possuem 381 certificados. Como se pode observar no gráfico da Figura 13, nesta região, os 5 principais sectores de actividade correspondem a uma percentagem de 57% do total de certificados. Os sectores de actividade desta região possuem algumas particularidades ainda não discutidas anteriormente. Os certificados adquiridos pelas empresas desta região inseridas no sector de actividade Comércio correspondem a um total de 17% de todos os certificados adquiridos nesta região. As empresas inseridas neste sector apresentam somente as normas ISO 9001:2000, OHSAS 18001:1999 e HACCP. O sector de actividade Hotéis e Restaurantes apresenta uma particularidade ainda não discutida, a norma ISO 9001:2000 adquirida pelas empresas neste sector apresenta uma percentagem bastante mais baixa do que esperado, verificando-se apenas 39% da norma ISO 9001:2000 nestas empresas; segue-se a norma ISO 14001:2004 com cerca de 26%, OHSAS 18001:1999 com cerca de 15%, EMAS II com cerca de 15% e HACCP com cerca de 4%. Estes dados mostram, à semelhança do que ocorre na região do Algarve, uma grande importância do turismo nesta região.

As empresas da região dos Açores inseridas no sector de actividade Construção apenas apresentam as normas ISO 9001:2000 e OHSAS 18001:1999, com respectivamente 95% e 5%. À semelhança da região do Norte, as empresas inseridas no sector de actividade Transporte, Armazenamento e Comunicações apresentam as três normas principais, ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999. O sector de actividade Saúde e Serviços é representado apenas pela norma ISO 9001:2000, tendo todas as empresas neste sector apenas a norma ISO 9001:2000.

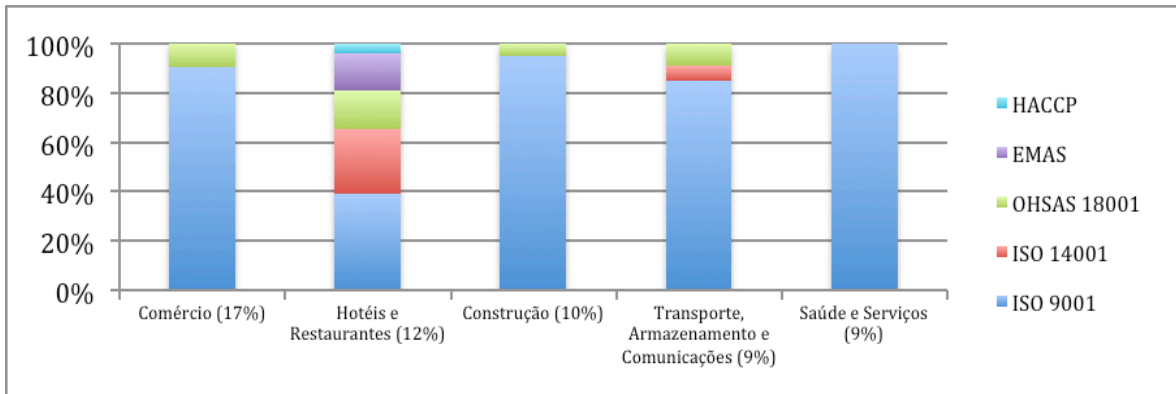


Figura 13 - Top 5 dos Sectores de Actividade da Região Açores

As regiões que apresentam o maior número de certificados adquiridos pelas empresas dos diversos sectores são, respectivamente, as regiões Norte, Lisboa e Vale do Tejo e Centro. E os sectores de actividade que apresentam empresas com o maior número de certificados são Comércio, Construção, Outros Serviços, Fabricação Metalúrgica de Base e Produtos Metálicos e Saúde e Serviços Sociais como se pode verificar no gráfico da Figura 14.

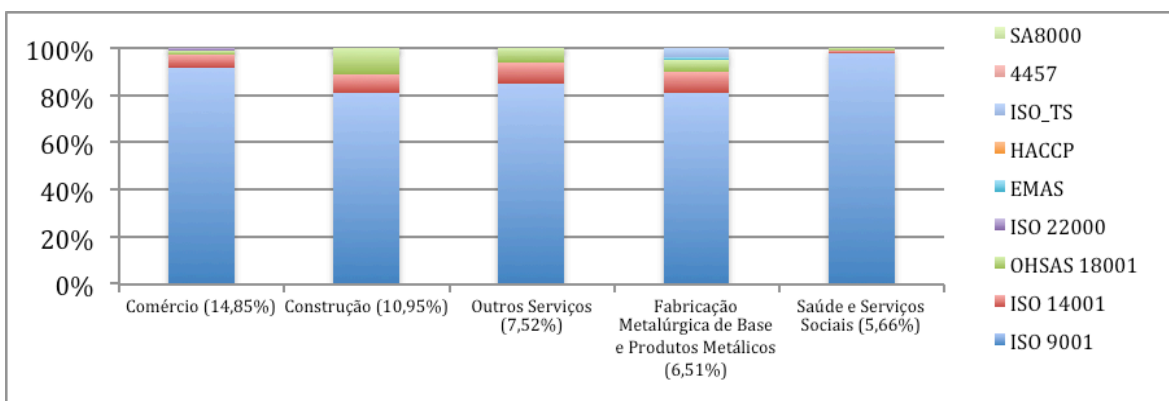


Figura 14 - Top 5 dos Sectores de Actividade em Portugal

O conjunto de dados apresentado possui informação suficiente para a análise pretendida ser realizada com sucesso. Todos os dados foram usados para a obtenção dos resultados. Esta descrição dos dados, obtida com a aplicação de *queries* simples ao conjunto de dados fornecido, já

é uma análise interessante e que permite ao utilizador a caracterização de alguns dos principais sectores de actividade.

3.4 Construção, Integração e Formatação dos dados

Os dados disponíveis possuem algumas características como a existência de conflitos de codificação do texto, a formatação diversa do texto ou a existência de alguns campos sem valores, que dificultaram o processo de limpeza de dados. Contudo, todos esses problemas foram ultrapassados sem grandes dificuldades. Após a resolução dos conflitos de codificação e formatação do conteúdo dos dados, foi necessário seleccionar os dados necessários à extracção do conhecimento. Todos os dados fornecidos são relevantes para esta etapa.

Para tornar o sistema de *data warehousing* enriquecido em termos de informação, é necessária uma investigação mais aprofundada para que fosse possível completar alguns detalhes importantes como as designações completas dos certificados. Isto porque, só possuímos a designação breve dada pelas empresas (exemplo: ISO9001 quando a designação detalhada é NP EN ISO 9001:2000). Será necessário adicionar também as designações dos sectores de actividade, pois no conjunto de dados da fonte só temos informação sobre os códigos de actividade económica e não aos sector de actividade a que cada código corresponde. Esta fase de investigação adicional é essencial porque os sistemas de *data warehousing* são bastante completos em termos de informação. Esta informação adicional vai ser bastante útil na extracção de conhecimento, permitindo ao futuro utilizador ter acesso a informação, que caso a investigação não fosse realizada, não existia.

Após a selecção e a limpeza dos dados, foi necessário extrair os dados para um modelo que facilitasse a extracção realizada pelas técnicas de mineração de dados. Para a construção deste modelo foi necessário definir os eixos de análise, tendo em conta as necessidades do cliente e o que se pretende com o estudo. Mas antes disso foi necessário homogeneizar os dados na área de retenção. A área de retenção (ou *Data Staging Area*) consiste numa estrutura computacional para suporte aos processos de preparação desses dados, para depois estarem em condições para serem

carregados para o *data mart*. Nesta fase de construção do sistema de *data warehousing* os dados foram tratados, e foi feita a gestão das chaves de substituição, usando uma chave homogénea de forma a gerir os dados que provem de fontes de diferentes e a homogeneizar as suas estruturas. O tratamento da gestão de chaves de substituição foi necessário visto que como possuímos três fontes de dados diferentes, à mesma empresa estão atribuídas chaves diferentes. Depois da fase da área de retenção, dos dados foram carregados para o *data warehouse*. Um *data warehouse* consiste num repositório de dados, orientado ao assunto, integrado, consistente e não-volátil que tem como principal objectivo apresentar a evolução dos dados ao longo do tempo, ajudando nos processos de tomada de decisão do negocio (Golfarelli & Rizzi, 2009).

Para construir o modelo relacional do *data mart*, representado na Figura 15, foi necessário construir a matriz de decisão e definir o grão, bem como definir o número de dimensões e as tabelas de facto que serão necessárias adicionar ao modelo. Após a análise dos requisitos e das especificações do cliente, decidiu-se pela seguinte definição de grão: um determinado **certificado** adquirido por uma determinada **empresa**, localizada numa determinada **região**, inserida num determinado sector de actividade (**EAC**), numa determinada **data**.

Tabela 3 – A Matriz de Decisão utilizada

Dimensões	Data Mart: Barómetro
Certificado	X
Empresa	X
Tempo	X
EAC	X
Região	X

Após a construção da matriz, representada na Tabela 3, bastante simples para a situação que estamos a tratar, decidiu-se pela criação de 5 dimensões e uma tabela de factos (designada no modelo por **tf_barometro**). Um facto representado pela tabela de facto é o que pretendemos analisar, e é constituído pelos diversos eixos de análise, ou mais especificamente as chaves correspondentes a cada uma das dimensões. Como uma empresa pode possuir vários certificados, e um tipo de certificado pode ser adquirido por várias empresas, foi necessário criar uma tabela ponte para resolver este relacionamento. Esta tabela designada no modelo por **tp_empresa_certificado**, vai ligar a tabela de factos e a dimensão correspondente aos

certificados e tem como principal objectivo associar uma determinada empresa a um conjunto de certificados. Em suma, decidiu-se criar um esquema para o *data mart* composto por 5 dimensões, uma tabela-ponte e uma tabela de factos.

A partir dos dados da fonte, foram criadas dimensões tendo em conta os diversos eixos de análise:

1. **Dimensão com dados sobre os Certificados** – designada por **Dim_certificado**, representada na Tabela 4, esta dimensão possui informações detalhadas sobre os certificados. Como as fontes não possuíam informações relevantes para o enriquecimento do data warehouse, foi necessária uma investigação mais profunda para obter esses dados.

Tabela 4 - Constituição da Dimensão Certificado

Atributo	Descrição	Tipo de Dados	Domínio
codigo	Código atribuído ao certificado, correspondente à chave primária	Número	Até 10 dígitos
Norma	Designação breve atribuída a um certificado	Texto	Até 20 caracteres
Estado	Estado de um certificado (ex: Activo/Inactivo)	Texto	Até 10 caracteres
TipoServico	Designação extensa de um certificado	Texto	Até 100 caracteres

2. **Dimensão com os dados sobre as Regiões** – designada por **Dim_Regiao**, representada na Tabela 5, esta dimensão possui informação relativa às regiões onde as empresas estão localizadas. Neste caso, as regiões estão agrupadas segundo a classificação NUTS II.

Tabela 5 - Constituição da Dimensão Região

Atributo	Descrição	Tipo de Dados	Domínio
codigo	Código atribuído à região, correspondente à chave primária	Número	Até 10 dígitos
Região	Designação breve de uma região	Texto	Até 50 caracteres

3. **Dimensão com os dados sobre as Empresas** – designada por **Dim_Empresa**, representada na Tabela 6, esta dimensão possui informações sobre as empresas. Como os dados das fontes não continham mais informação sobre as empresas usadas no estudo por Sampaio e Saraiva, apenas foi usada a informação disponível, contendo apenas o nome das empresas.

Tabela 6 - Constituição da Dimensão Empresa

Atributo	Descrição	Tipo de Dados	Domínio
codigo	Código atribuído à empresa, correspondente à chave primária	Número	Até 10 dígitos
Nome	Designação de uma empresa	Texto	Até 100 caracteres

4. **Dimensão com os dados sobre os Sectores de Actividade** – designada por **Dim_eac**, representada na Tabela 7, esta dimensão possui informações sobre os sectores de actividade em que as empresas em questão estão inseridas. Como os dados das fontes não continham mais informação sobre os sectores de informação, apenas continham o código de actividade económica (EAC), foi realizada uma investigação mais aprofundada, permitindo obter descrições sobre cada um dos sectores de actividade.

Tabela 7 - Constituição da Dimensão EAC

Atributo	Descrição	Tipo de Dados	Domínio
codigo	Código atribuído à empresa, correspondente à chave primária	Número	Até 10 dígitos
EAC	Código de actividade económica de um sector de actividade	Número	Até 10 dígitos
Descricao	Descrição detalhada de um sector de actividade	Texto	Até 100 caracteres

5. **Dimensão com os dados sobre o Tempo** – designada por **Dim_tempo**, representada na Tabela 8, esta dimensão corresponde a um eixo de análise de grande importância. Esta característica permite realizar análises temporais sobre os dados, permitindo analisar a evolução de um sistema ao longo dos anos. Esta dimensão vai ser uma dimensão degenerada, pois os dados nas fontes só permitiam uma consulta por ano.

Tabela 8 - Constituição da Dimensão Tempo

Atributo	Descrição	Tipo de Dados	Domínio
codigo	Código atribuído ao ano, correspondente à chave primária	Número	Até 10 dígitos
Ano	Ano de uma determinada data	Ano (AAAA)	Exemplo: AAAA

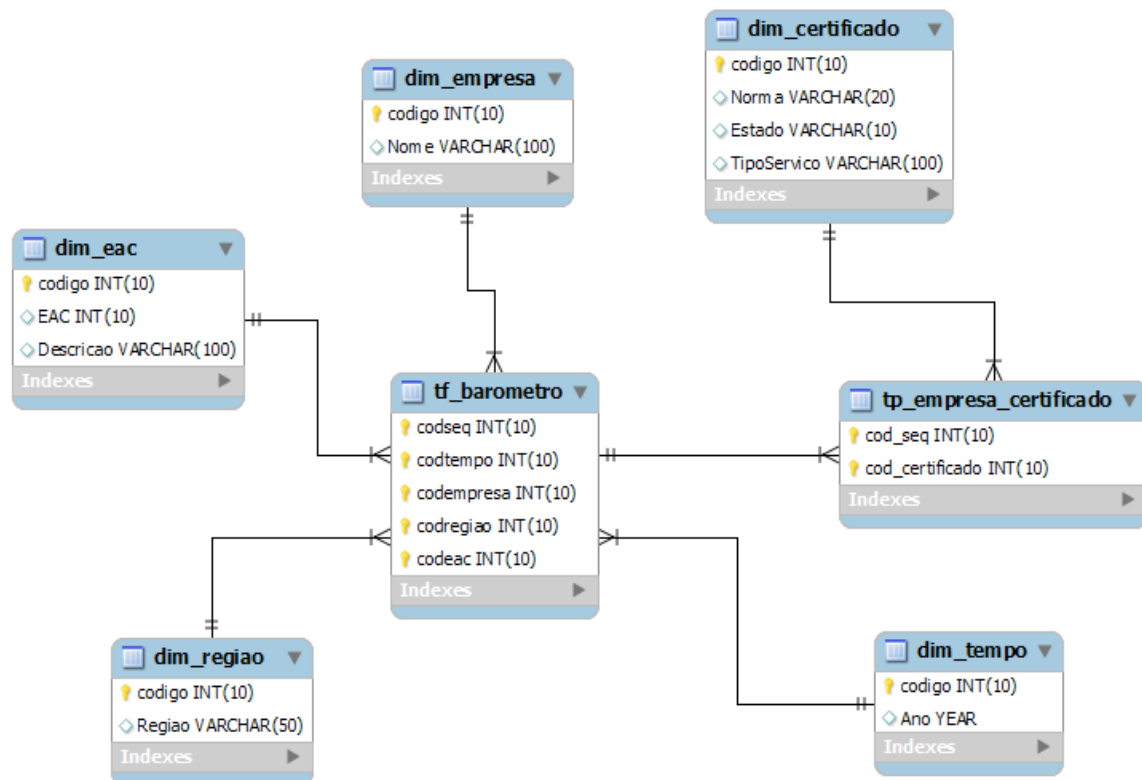


Figura 15 - Modelo Relacional do Data Mart construído

6. **Tabela de factos** – designada por **tf_barometro**, consiste numa tabela relacional que inclui a unidade designada por facto, este conjunto de chaves é constituído pelas chaves estrangeiras das dimensões ligadas à tabela.

Após a construção do modelo, foi necessário povoar as dimensões, a tabela de factos e finalmente, a tabela ponte. Para isto foi necessário extrair os dados das fontes, tratar os dados, torná-los homogéneos usando chaves de substituição (todas estas acções são realizadas na área de retenção) e só depois é que se tratou do processo de povoamento do *data mart*.

As primeiras tabelas a serem povoadas foram as dimensões, visto que é necessário manter a integridade referencial das chaves primárias das dimensões e as chaves estrangeiras da tabela de factos. No caso do povoamento das dimensões, a extracção dos dados foi automática: fez-se a selecção da informação pretendida e copiou-se para as dimensões em questão. Após o povoamento das dimensões estar completo realizou-se o povoamento da tabela de factos, através da extracção dos dados alojados na área de retenção, relacionando os dados de todas as dimensões. Finalmente, povoou-se a tabela ponte, relacionando a informação da tabela de factos e da dimensão correspondente aos certificados.

A parte do trabalho prático envolvendo a gestão de bases de dados, criação e tratamento na área de retenção, extracção dos dados para o *data warehouse* e a criação do modelo do *data warehouse* foi realizada usando *MySQL* versão 5.2 [3]., e a ferramenta de manuseio utilizada foi o *MySQLWorkBench* [4]. Esta ferramenta permite administrar um servidor de base de dados, modelar dados e desenvolver e gerir um sistema de bases de dados, usando a linguagem *SQL (Structured Query Language)*.

Depois da construção do sistema de *data warehousing*, a aplicação das técnicas de mineração de dados fez-se através duma ferramenta de *open-source* para *data mining* designada por *RapidMiner* [5]. Esta ferramenta permite importar o modelo de *data warehousing* com os dados implementado anteriormente, aplicar as diferentes técnicas de *data mining*, alterar os parâmetros consoante a necessidade do utilizador, obter os resultados num tipo de linguagem completamente compreensível pelo utilizador. Estas características são essenciais numa ferramenta de *data*

mining, pois é necessário que este saiba o que pretende com o conjunto de dados inicial e interpretar os resultados.

Capítulo 4

O Processo de Descoberta de Conhecimento

O processo de extrair padrões em grandes volumes de dados designa-se usualmente por Extração de Conhecimento (*Knowledge Discovery in Databases* ou KDD). Os padrões escondidos ou conhecimento correspondem a informação previamente desconhecida e potencialmente útil. Este processo engloba também outras tarefas, nomeadamente como é que os dados são acedidos e armazenados, como é que os algoritmos são escalados de maneira a que se obtenha o melhor desempenho possível e como é que os resultados são interpretados e analisados. O principal problema do processo de Extração de Conhecimento é tornar a informação viável e compreensível para as pessoas que irão tomar as decisões baseadas nos relatórios (Fayyad, et al., 1996; Frawley, 1991). Algumas das principais características deste processo são (Frawley, 1991):

- **Linguagem de alto nível** – os resultados da descoberta de conhecimento são representados por uma linguagem de alto nível, que pode ser compreendida por utilizadores humanos.
- **Eficiência** – o processo de descoberta de conhecimento é um processo eficiente, possuindo tempos de execução e desempenho aceitáveis e previsíveis.
- **Precisão** – o processo de descoberta retrata com bastante precisão o conteúdo de uma base de dados.
- **Resultados Interessantes** – os resultados obtidos neste processo são de interesse do utilizador, sendo o conhecimento obtido potencialmente útil para o utilizador.

Existem alguns problemas nas bases de dados que podem reflectir-se facilmente nos resultados obtidos e impedem, por vezes, que a descoberta de conhecimento seja precisa, apesar de os resultados da mineração de dados serem considerados probabilísticos e não correctos. Alguns destes problemas das bases de dados são (Frawley, 1991):

- **Dados dinâmicos** – a alteração dos dados ao longo do tempo, que é uma das características fundamentais da maior parte das bases de dados existentes, pode afectar a descoberta de conhecimento.
- **Dados irrelevantes** – a irrelevância de grande parte dos dados contidos numa base de dados obriga a que o analista tenha conhecimento sobre o negócio e que se foque apenas nos dados que importam para a descoberta de conhecimento.
- **Dados em falta** – a falta de dados relevantes para a descoberta de conhecimento pode afectar este processo.
- **Ruídos nos dados** – o ruído nos dados pode afectar a descoberta de conhecimento e muitas vezes levar a resultados não certos.

Todavia, existem algumas metodologias que se podem seguir para conseguir resolver com sucesso problemas que envolvam mineração de dados, como a metodologia de Fayyad (KDD), a CRISP-DM (*CRoss Industry Standard Process for Data Mining*) ou a SEMMA (*Sample, Explore, Modify, Model and Assess*) (Kurgan & Musilek, 2006).

4.1 Metodologia do CRISP-DM

Neste projecto optámos por seguir a metodologia CRISP-DM. Esta metodologia é usada para resolver problemas que se enquadram no contexto de mineração de dados e divide-se, essencialmente, em seis fases (Chapman, et al., 2000):

1. **Compreensão do Negócio**, em que se analisa o problema e se identificam os principais objectivos e os resultados que se pretendem obter com a mineração de dados.

2. **Compreensão dos Dados, na qual** ocorre uma familiarização com os dados, que envolve a verificação da qualidade dos dados e a identificação dos conjuntos de dados onde será possível encontrar padrões escondidos.
3. **Preparação dos Dados**, em que se limpam, adaptam e transformam os dados de forma a que seja construído o conjunto de dados final para a aplicação de técnicas de mineração de dados.
4. **Modelação, na qual** se procede ao estudo, selecção e aplicação das técnicas de mineração de dados adequadas ao caso em estudo.
5. **Avaliação**, em que os modelos criados na fase anterior são testados e se verifica se estão de acordo com os objectivos estipulados na fase inicial.
6. **Desenvolvimento**, em que são gerados os relatórios com a análise dos resultados e entregues ao cliente.

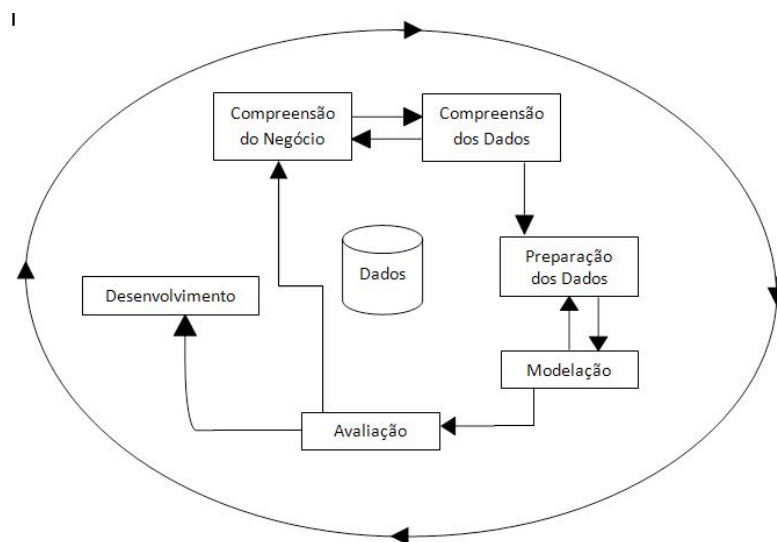


Figura 16 - Fases da Metodologia CRISP-DM

baseada na figura contida em (Chapman, et al., 2000)

Como se pode verificar na imagem apresentada na Figura 16, a metodologia CRISP-DM não é um processo rígido, é possível retornar a uma fase anterior a partir de uma fase mais avançada, dependendo do que se obtém em cada fase e de surgir a necessidade de melhorar o resultado obtido (Chapman, et al., 2000).

4.2 Mineração de Dados

NA CRISP-DM, a fase da mineração de dados é aquela em que ocorre a procura de padrões e relacionamentos escondidos em grandes volumes de dados, aplicando técnicas de mineração e algoritmos adequados. É a fase em que são utilizadas algumas técnicas de *machine learning* e algoritmos de reconhecimento de padrões para extrair informação de dados pré-processados, ou seja, dados obtidos após as fases necessárias de remoção do ruído e de tratamento dos valores nulos (Adriaans & Zantinge, 1996; Cabena, et al., 1998).

4.2.1 Técnicas de Mineração de Dados

As técnicas de mineração de dados tornaram-se uma revelação ao longo dos anos devido à sua importância na extracção de padrões e conhecimento de grandes bases de dados. Classificação e segmentação são dois exemplos de técnicas de mineração de dados, sendo também as técnicas que melhor se adequam para extrair padrões das bases de dados do caso apresentado anteriormente. A classificação permite classificar um novo caso segundo a uma pré-classificação de instâncias da base de dados, enquanto que a segmentação divide os dados em segmentos (*clusters*) baseada na similaridade entre os objectos. A escolha da técnica de mineração de dados a ser aplicada varia com o que se pretende obter do conjunto de dados, quer seja uma previsão ou uma sua simples descrição. Para o caso apresentado foram escolhidos as técnicas de classificação (na vertente de previsão) e segmentação (na vertente de descrição).

4.2.2 Classificação

A classificação é uma técnica que tem como objectivo descobrir conjuntos de modelos ou funções que distinguem as diferentes classes no conjunto de dados baseadas nas características de outros

objectos. Posteriormente, estes modelos irão ser aplicados a novos casos ainda não classificados (Chen, et al., 1996). Devido às suas características, esta técnica é usada em diversas áreas de aplicação, como a detecção de fraudes, os diagnósticos médicos, a previsão de comportamentos ou as estratégias de *marketing* (Han & Kamber, 2001).

Com esta técnica é possível criar o modelo de classificação, analisando e usando as características do conjunto de treino e desenvolver uma correcta descrição do mesmo, usando as características disponíveis dos exemplos na base de dados. As classes descritas são usadas para classificar o conjunto de teste da base de dados com o objectivo de refinar a descrição das classes (Chen, et al., 1996). O principal objectivo deste método é classificar novas instâncias tendo como base esse modelo. Todos estes objectos do conjunto de dados devem ser classificados baseados nas suas propriedades, isto é, cada classe deve possuir objectos com características em comum (Cabena, et al., 1998; Quintela, 2005).

Esta técnica é composta por duas fases (Han & Kamber, 2001). A primeira delas, denominada de treino ou de aprendizagem, consiste em construir o modelo de classificação ou classificador, que tem como objectivo classificar novos casos (Chen, et al., 1996). Para a construção do classificador é usada uma amostra da base de dados original, denominada por conjunto de treino (Langaas, 1995). Todos os exemplos (instâncias) do conjunto de treino possuem o mesmo número de atributos que os registos da base de dados. Nesta primeira fase, é definida, usando o algoritmo pretendido, uma regra de classificação que permite associar o atributo classe-alvo a uma determinada classe, tendo em conta os valores dos restantes atributos do exemplo.

Na segunda fase, é testado o conjunto de teste (restantes exemplos da base de dados original). Esta fase serve essencialmente para classificar os casos ainda não classificados e determinar a precisão do classificador. A precisão de um classificador consiste na percentagem de tuplos do conjunto de teste que são bem classificados, segundo o classificador. O algoritmo de classificação que mais se adequa ao caso apresentado, permitindo obter os melhores resultados, tendo em conta o que se pretende, são as árvores de decisão usando o algoritmo C4.5.

O princípio básico de um modelo de árvore é particionar, de uma maneira recursiva, o conjunto de dados até que os elementos de uma partição pertençam maioritariamente à mesma classe (Hand,

et al., 2001). Uma árvore de decisão consiste numa estrutura em grafo, direccionada e acíclica, constituída por um conjunto de nodos e folhas. Os nodos internos da árvore contêm um teste lógico feito ao valor do atributo (caso) a ser classificado, cada ramo contém as respostas possíveis ao teste lógico contido no nodo anterior e os nodos terminais (folhas) representam a classe-alvo (Craven, 1996; Han & Kamber, 2001). Desta forma, todos os exemplos vão ser divididos de forma a que pertençam a grupos mutuamente exclusivos (Kirkos, et al., 2007).

O processo da classificação de um exemplo inicia-se no nodo correspondente à raiz da árvore, que depois é percorrida consoante os resultados obtidos nos testes feitos nos nodos internos da árvore ao atributo a classificar. Este processo é feito até ser atingida alguma das folhas das árvores. Quando um exemplo atinge um nodo folha, a classe associada a esse nodo folha é a previsão feita pela árvore a esse exemplo (Craven, 1996; Han & Kamber, 2001; Quinlan, 1986). Quinlan nos anos 80s desenvolveu o algoritmo de árvores de decisão conhecido como ID3, e como seu sucessor melhorado o algoritmo C4.5 (Quinlan, 1996).

O algoritmo C4.5 permite a geração de árvores de decisão a partir do conjunto de treino e torna possível separar os dados em classes distintas. Tendo em consideração os termos: nodo N, partição D consiste numa amostra de tuplos/exemplos do conjunto de treino:

- Criar o nodo N.
- Se os tuplos da partição em causa (tuplos a serem avaliados pelo nodo N) são todos/maioria da classe C: tornar o nodo N, nodo folha representando a classe C.
- Se não se verificar a condição anterior, utilizar um critério para a divisão dos tuplos, com o objectivo de separar os tuplos em diferentes classes e de se adicionar os ramos ao nodo N (baseando essa decisão nos diferentes resultados obtidos pelos testes aos tuplos e do tipo de dados desses resultados).

Para construir a árvore, o algoritmo usa todos os exemplos do conjunto de treino, para inicialmente definir o teste lógico que será atribuído ao nodo-raiz. De cada vez que um nodo interno é adicionado à árvore é escolhido apenas um conjunto de exemplos que está ser classificado para escolher o teste atribuído a esse nodo. Depois do teste atribuído ao nodo ter sido escolhido, este é usado para dividir os exemplos do conjunto de treino e este processo decorre de

forma recursiva (Craven, 1996). Os ramos associados a um determinado nodo são criados consoante o tipo de resultados desse nodo, podem ser resultados contendo variáveis de valores contínuos ou discretos (Han & Kamber, 2001). Para lidar com resultados com valores discretos, o algoritmo C4.5 cria um ramo para cada possível resultado, como é possível verificar na Figura 17.



Figura 17 - Tratamento dos valores discretos pelo algoritmo C4.5

Baseado na figura em (Han & Kamber, 2001)

Para conjuntos de treino com valores contínuos, o algoritmo C4.5 considera uma divisão binária para cada nodo interno, isto é, duas saídas (ou ramos) para cada nodo interno e cada uma destas saídas possui um teste que será realizado aos atributos. O teste correspondente a uma das saídas é definido como 'menor ou igual que um valor de divisão (*threshold*)' e da outra saída é 'maior ou igual que o valor de divisão', como mostra a Figura 18.



Figura 18 - Tratamento dos valores contínuos pelo algoritmo C4.5

Baseado na figura em (Han & Kamber, 2001)

O atributo que é escolhido para a divisão da partição não é escolhido aleatoriamente. Para o escolher, o algoritmo ID3 usa uma medida de avaliação chamada *Information Gain* ou *Entropy*

Reduction, que permite minimizar o número de testes num determinado nodo, sendo possível obter uma árvore mais simples (Han & Kamber, 2001; Larose 2005, 2006).

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

A Equação (1) é usada para classificar um determinado tuplo existente em D , em que o valor de p_i corresponde ao valor da probabilidade de um tuplo pertencer à classe C_i . O uso da função logarítmica (\log_2) deve-se ao facto de a informação ser codificada em *bits*.

$$Info_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j) \quad (2)$$

A Equação (2) é usada para classificar um determinado tuplo existente em D , mas tendo em consideração uma partição realizada por um determinado atributo A (Han & Kamber, 2001). Esta equação combinada com a Equação (1), permite calcular o ganho de informação caso a partição seja realizada com o atributo A . O ganho de informação pode ser calculado com a Equação (3).

$$Gain(D) = Info(D) - Info_A(D) \quad (3)$$

Por outras palavras, esta medida consiste numa heurística que permite seleccionar o atributo de divisão para uma partição de um determinado nodo interno de uma árvore, que maximiza o ganho de informação, minimizando o grau de "impureza" das partições resultantes com o objectivo de facilitar a classificação destas no próximo nodo. Um dos problemas desta medida é o facto de não ter em conta o número de partições que são criadas, pois só interessa a "pureza" das partições, podendo levar à geração de partições apenas com um tuplo ou exemplo. O algoritmo C4.5 utiliza uma extensão desta medida de avaliação, *Gain Ratio*. Esta medida aplica um tipo de normalização ao ganho de informação na divisão da informação, permitindo "combater" o grande número de partições geradas pelo *information gain* (Han & Kamber, 2001).

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right) \quad (4)$$

Com a Equação (4) é possível obter a informação obtida pela divisão de D , em v partições, obtendo um total de v resultados de um teste ao atributo A . O atributo que obtiver o maior valor de *Gain Ratio* (obtido pela Equação (5)) é escolhido para o atributo de divisão do nodo N (Han & Kamber, 2001).

$$Gain\ Ratio\ (A) = \frac{Gain\ (A)}{SplitInfo(A)} \quad (5)$$

Um aspecto importante deste algoritmo é o seu critério de paragem. Se os exemplos do subconjunto que atingem um nodo são todos da mesma classe, o algoritmo não irá continuar a divisão e criará um nodo folha com a classe em questão. Caso contrário, se o algoritmo não conseguir criar, no mínimo, dois ramos, tendo em consideração o número mínimo de cada nodo, então o algoritmo não continuará a divisão. E ainda, no caso de não existirem exemplos relativos à aplicação do teste lógico, o algoritmo não continua a divisão (Craven, 1996).

Depois de ser construída a árvore, o algoritmo tenta simplificar a árvore, uma vez que quando é construída uma árvore muito extensa, são identificadas anomalias no conjunto de treino, como *outliers* e ruído nos dados, o que é considerado um problema na medida em que pode levar a um pobre previsão (Craven, 1996). Para eliminar este problema é feito *pruning* à árvore. Este processo consiste em remover nodos sem alterar significativamente a precisão do classificador (Kirkos, et al., 2007). Para isto o algoritmo C4.5 utiliza uma medida de avaliação, *cost complexity*, que é um tipicamente um *postpruning approach*. Isto significa que só é realizado o *pruning* à árvore após a sua construção. Esta aproximação considera o custo de complexidade (função do número de folhas da árvore) e o *error rate* da árvore, e por estas razões é considerado um *pruning* pessimista (Craven, 1996; Han & Kamber, 2001).

O algoritmo C4.5 não se restringe apenas à formação de dois ramos por folha, permitindo que sejam formados mais ramos consoante o resultado do teste à folha. Uma das desvantagens deste algoritmo é, que para conjuntos de dados muito extensos, a árvore apresentada pode ser muito grande e conter ruído, o que leva a uma baixa precisão quando classificados novos casos. A precisão é a capacidade do método para prever correctamente os casos. (Kirkos, et al., 2007; Sousa, et al., 1998; Han & Kamber, 2001).

Para a aplicação do algoritmo aos dados obtido foi utilizado o software *RapidMiner* e o conjunto de dados total obtido pelo *data warehouse* criado. O atributo *norma* da dimensão Dim_certificado foi definido como *label*, pois este atributo é a classe que se pretende analisar e relacionar com os sectores de actividade e as regiões. Para a aplicação deste algoritmo foi utilizada validação cruzada (*cross validation*) com o operador "*XValidation*" do *RapidMiner*, com um total de 10 iterações e *sampling_type=shuffled sampling*. A precisão do classificador é calculada automaticamente pelo *RapidMiner*.

4.2.3 Segmentação

A técnica de segmentação consiste em dividir um conjunto de dados em segmentos (*clusters*), baseado na similaridade entre os objectos. Os segmentos obtidos desta técnica de aprendizagem não supervisionada correspondem a padrões escondidos numa grande quantidade de dados (Berkhin, 2002). Este processo é um processo de aprendizagem não supervisionada. Ao contrário do processo da classificação, a segmentação permite construir um modelo descritivo dos dados e não tem como base para o processo uma pré-classificação de uma amostra dos dados da base de dados original (Han & Kamber, 2006; Jain, et al., 1999; Quintela, 2005). A segmentação é muito útil em várias situações: tomadas de decisão, situações de *machine-learning* e análise e exploração de padrões em conjunto de dados com conteúdo desconhecido (Jain, et al., 1999).

Para a divisão dos objectos do conjunto de dados, a técnica de segmentação ajuda a construir um particionamento com significado de um grande conjunto de objectos baseado na metodologia *divide and conquer* que decompõe um sistema de grande escala num mais pequeno e a organização dos dados em grupos de forma a se verificar alta similaridade intra-segmentos e baixa similaridade inter-segmentos (Chen, et al., 1996). É necessário ter conhecimento das

características dos algoritmos de segmentação para que seja mais fácil (e efectiva) a escolha do algoritmo que melhor se adapta às necessidades do utilizador e ao que este pretende. As propriedades dos algoritmos de segmentação são (Berkhin, 2002):

- Tipo de dados que processa.
- Escalabilidade em conjuntos de dados muito grandes.
- Capacidade de trabalhar com dados dimensionais.
- Capacidade para encontrar segmentos com formas irregulares.
- Capacidade de lidar com *outliers* e ruído nos dados.
- Complexidade do algoritmo/ tempo de execução razoáveis.
- Dependência da ordem dos dados.
- Rotulagem.
- Parâmetros disponíveis para o utilizador definir.
- Interpretação dos resultados de fácil interpretação.

O algoritmo de segmentação escolhido e geralmente o mais utilizado é o *k-means* (MacQueen, 1967). Este é um algoritmo de segmentação que separa um conjunto de dados em grupos representados pelo valor médio dos valores de um grupo (centróide), agrupando os dados de acordo com as suas características, maximizando a similaridade entre os objectos de cada grupo minimizando a similaridade entre os objectos fora do segmento. Na prática, o algoritmo *k-means* minimiza a soma dos erros quadrados, representada na Equação (6) (Han & Kamber, 2001).

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (5)$$

A Equação (5) representa a soma dos erros quadrados para cada objecto do conjunto; p é a posição desse objecto, m_i representa o centróide do segmento C_i .

O algoritmo *k-means* recebe como *input* o número k de segmentos pretendido. Este corresponde ao número de segmentos em que o utilizador pretende que seja dividido o conjunto de dados. Inicialmente, estes k pontos vão ser escolhidos aleatoriamente pelo algoritmo e considerados os

centróides dos segmentos. Seguidamente, o algoritmo atribui cada objecto ao segmento que é mais similar (baseando-se no valor médio dos objectos já contidos no segmento), e recalcula a média de cada segmento, isto é, novo centróide. Estes últimos dois passos são executados recursivamente até não existir qualquer mudança na escolha dos objectos para os segmentos. O resultado obtido por este algoritmo depende bastante do número k dado como *input* ao algoritmo, o que demonstra ser um grande problema porque é necessário recorrer a vários testes ao conjunto de dados e critérios para determinar qual o melhor k a usar. (Berkhin, 2002; Jain, et al., 1999):

O algoritmo *k-means* é um algoritmo escalável, só trabalha com atributos numéricos e pode ser facilmente afectado por *outliers* e ruídos nos dados. Apesar disso, é um algoritmo bastante eficiente e tem a capacidade de fornecer resultados estatísticos com significado para o utilizador. Este algoritmo é bastante utilizado para uso científico e industrial devido às suas características (Berkhin, 2002; Huang, 1997).

Uma desvantagem deste algoritmo é o facto de ter que ser fornecido o número de segmentos em que é necessário dividir o conjunto de dados. Para escolher o número de segmentos em que ocorre a divisão óptima, isto é, para qual número de k segmentos o algoritmo obtém os melhores resultados, é necessário observar os valores de *Davies Bouldin* (DB) (Ray & Tury, 1999; Halkidi, et al., 2001). Nestes trabalhos é possível analisar várias formas de efectuar segmentação. Os autores sugerem usar o critério da minimização do valor de DB. Ao encontrar o menor valor desta variável, significa que os diferentes segmentos apresentam a menor similaridade que podia existir entre eles.

Para a aplicação do algoritmo *k-means* foi utilizado o software *RapidMiner*. Os atributos utilizados foram: o atributo *Norma* da dimensão Dim_certificado, o atributo *Regiao* da dimensão Dim_regiao e o atributo *Descricao* da dimensão Dim_EAC. Todos estes atributos são nominais, logo foi necessário usar o método *NominalToNumerical* para que o algoritmo *k-means* consiga processar estes atributos. Para obter o valor de k é necessário analisar os valores obtidos pelo critério de *Davies Bouldin*. Este valor é calculado automaticamente pelo *RapidMiner* e a melhor aproximação que se obteve deste critério foi com $k=2$. O critério usado para calcular a média dos centróides foi soma dos quadrados (*SumOfSquares*).

4.2.4 Escolha da Técnica de Mineração de Dados

Existem diversos factores que se deve ter em conta na escolha da técnica de mineração de dados que se irá aplicar no conjunto de dados disponível com o objectivo de se obterem padrões e conhecimento útil (Dias, 2002): conhecimento do negócio em causa, conhecimento do que faz cada tarefa; e técnicas e algoritmos de mineração de dados disponíveis. Estas são 3 medidas importantes que se devem ter em consideração antes de se escolher as técnicas de mineração de dados a usar.

Tendo em conta o caso apresentado, e o tipo de dados com que iremos lidar, decidiu-se optar pela aplicação das tarefas Classificação e Segmentação. Os nossos objectivos (já enunciados anteriormente) são: prever quais os certificados que irão ser úteis para uma nova determinada empresa, tendo em conta a sua localização e sector de actividade, obter análises descritivas e quantitativas do mercado competitivo actual, como por exemplo, quais os certificados, zona de localização e sector de actividade, que predominam nas empresas nacionais actualmente. A Classificação será útil para obter uma caracterização de sectores de actividade, pois será possível dividir o conjunto de dados em classes distintas, tendo em conta as suas características. Isto é, dadas as características duma empresa, por exemplo, zona de localização, **Norte**, e sector de actividade, **Alimentação, Bebidas e Tabaco**, será possível obter uma lista de certificados, os mais úteis a obter pela empresa tendo em conta as condições, para que a empresa consiga sobreviver no mercado actual. A Segmentação será útil para obter uma descrição do mercado competitivo actual, podendo ser possível observar quais os sectores de actividade predominantes em Portugal, a evolução dos certificados e quais os que são mais requisitados pelas empresas e mesmo quais as zonas de Portugal que representam a maior actividade empresarial do país. Juntando o estudo realizado por Sampaio e Saraiva, será possível também obter uma conclusão mais detalhada sobre a evolução dos certificados e das empresas em Portugal. Depois de definidos os objectivos e as tarefas de mineração de dados a usar, é necessário escolher os algoritmos que melhor se adaptam ao que se pretende dentre os algoritmos existentes das técnicas classificação e segmentação.

Tendo em conta as características dos algoritmos e das técnicas, decidiu-se pela aplicação do algoritmo C4.5 (Classificação – *Decision Tree*) e do *k-means* (Segmentação – Partições). O

algoritmo escolhido de Classificação, o C4.5, é o mais adequado ao caso pois este contém dados categóricos (*categorical data*). Este algoritmo permite a geração de árvores de decisão, em que cada nodo corresponde a um teste lógico feito ao atributo e os ramos correspondem ao resultado desse mesmo teste. Este algoritmo não se restringe apenas à formação de dois ramos por folha, permitindo que sejam formados mais ramos consoante o resultado do teste à folha. É necessário fornecer a classe-alvo e os atributos que irão fazer parte da árvore; e é útil na descoberta de características das empresas que possuam o mesmo tipo de certificado, sendo possível prever se será útil para uma determinada empresa com certas características adquirir um determinado certificado. Quanto ao *k-means*, este é, como vimos, um algoritmo que agrupa os dados de acordo com as suas características, maximizando a similaridade entre os objectos de cada segmento, e minimizando a similaridade entre os objectos fora do segmento. É necessário fornecer o número de *k* segmentos pretendidos, isto é, o número de segmentos em que serão divididos os dados. Este algoritmo pode ser útil para ter conhecimento dos grupos formados baseado na similaridade entre as empresas tendo em conta os certificados que possuem.

4.2.5 Importância dos métodos na obtenção dos resultados

As técnicas de mineração de dados permitem a extracção de conhecimento de grandes quantidades de dados, permitindo ao utilizador tomar certas decisões baseado na informação extraída. O conhecimento extraído pelas técnicas de mineração de dados poderá ser uma vantagem competitiva no mercado, melhorando a eficiência da organização e permitindo melhores serviços aos clientes. Sem a aplicação destas técnicas, a tomada de decisões do utilizador será mais difícil, pois é necessária uma análise manual dos dados, o que pode ser demorado e menos eficaz devido ao grande volume dos dados.

Com as duas técnicas apresentadas, a classificação e a segmentação, é possível obter uma previsão e uma descrição do conjunto de dados em questão, que são uma mais-valia na medida em que, após a análise dos resultados da aplicação das técnicas, será possível sem qualquer percepção ou conhecimento do estado do mercado competitivo, ter uma noção do mesmo, ajudando as novas organizações a terem uma orientação inicial e conseguirem sobreviver nestas condições.

Através da aplicação da classificação é possível ter noção do conjunto de certificados que será melhor adquirir qualquer que seja a região ou o sector de actividade em que a organização se encontra localizada ou inserida. E com os resultados da segmentação será possível analisar o estado do mercado, tendo uma melhor noção dos sectores de actividade e regiões que possuem o maior número de organizações e quais os certificados que adquiriram durante os três anos da análise.

Capítulo 5

Análise dos Resultados

5.1 Caracterização dos Sectores usando Classificação

Em secções anteriores fizemos já uma descrição das características dos sectores de actividade que apresentam o maior número de certificados por região, mostrando também características específicas de cada região em relação ao mercado competitivo que se vive actualmente. Nesta secção serão apresentados em tabelas os resultados obtidos pela aplicação da técnica de mineração de dados, classificação. Estes resultados correspondem ao certificado predominante ou à lista de certificados predominantes por sector de actividade e qual a região em que as empresas inseridas nesse sector possuem o maior número de certificados. Os resultados apresentados estão organizados por número de EAC.

Através da aplicação da classificação é possível identificar os certificados predominantes, quer por região, quer por sector de actividade. Neste estudo, foram então criadas várias árvores de decisão para a exploração do conjunto de dados em questão. Uma árvore de decisão originada pela aplicação da técnica a todo o conjunto de dados e depois, foram geradas árvores de decisão por região, com intuito de ajudar à descrição de cada sector de actividade em cada região.

Devido à variedade de resultados obtidos, consolidou-se a informação em diferentes gráficos nesta secção, sendo apenas apresentado em cada gráfico os sectores de actividade que apresentam o maior número de certificados dentro da região. Nos gráficos apresentados seguidamente é possível observar a importância de cada norma em diferentes sectores e por região.

5.1.1 Por Norma

A árvore de decisão gerada pela aplicação do algoritmo de classificação ao conjunto de dados permite analisar a preferência dos sistemas de gestão das empresas nacionais consoante o sector de actividade em que estão inseridas. Os resultados obtidos foram bastante precisos apresentando um valor de acurácia relativamente elevado, cerca de 81.53%.

Os dados apresentados seguidamente foram os resultados obtidos pela aplicação da técnica de mineração de dados, a classificação, aplicando o algoritmo C4.5 aos dados em estudo. Uma particularidade que se observou foi o facto do certificado da norma ISO 9001:2000 ser adquirido por grande parte das empresas nacionais, apresentando uma percentagem bastante elevada em todos os sectores sem excepção. Os resultados obtidos permitem-nos uma visão geral sobre o mercado competitivo actual e, com isto, concluímos que a norma mais adquirida pelas empresas certificadas nacionais, para todos os sectores de actividade e para todas as regiões, a prioridade das empresas deve ser adquirir o certificado de gestão da qualidade, a norma ISO 9001:2000. Com a árvore de decisão obtida foi possível verificar quais os certificados predominantes para cada sector.

Todos os resultados apresentados nas tabelas que se seguem são referentes às empresas certificadas existentes em Portugal. Na prática, os sectores de actividade que se refere nas tabelas apresentavam como classe final o(s) sistema(s) de gestão. A Tabela 9 mostra os sectores de actividade em que as empresas nacionais possuem a norma ISO 9001:2000 como o sistema de gestão predominante, mas isto não significa que apenas possuam este sistema de gestão.

Tabela 9 - Sectores de Actividade com Norma ISO 9001:2000 predominante

Norma ISO 9001:2000	EAC	Descrição de Sector de Actividade
	5	Couro e Produtos de Couro
	38	Saúde e Serviços Sociais
	18	Equipamentos e Máquinas
	23	Outras Fabricações Não Especificadas
	20	Construção e Reparação Naval
	21	Indústria Aeroespacial
	29	Comércio
	31	Transporte, Armazenamento e Comunicações
	32	Mediação Financeira, Imobiliária e Aluguer
	37	Educação

Como já foi referido a norma ISO 14001:2004 é a segunda mais adquirida pelas empresas nacionais, sendo bastante importante para certos sectores de actividade. Na Tabela 10 é possível comprovar que para sectores de actividade que podem prejudicar de algum modo o ambiente (poluição, resíduos tóxicos) é essencial para as empresas possuírem esta norma, sendo muitas das vezes exigidas pelos clientes.

Tabela 10 - Sectores de Actividade com as Normas ISO 9001:2000 e ISO 14001:2004 predominantes

Normas ISO 9001:2000 e ISO 14001:2004	EAC	Descrição de Sector de Actividade
	4	Têxteis e Produtos Têxteis
	8	Editoras
	9	Empresas Gráficas
	12	Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais
	14	Fabricação de Artigos de Borracha e Matérias Plásticas
	17	Fabricação Metalúrgica de Base e Produtos Metálicos
	19	Equipamento Eléctrico e de Óptica
	30	Hotéis e Restaurantes
	33	Tecnologias de Informação
	34	Serviços de Engenharia
	35	Outros Serviços
	36	Administração Pública

No capítulo 3 foi referida a importância da norma OHSAS 18001:1999, realçando-se o facto de esta ser uma norma essencial para proteger os trabalhadores, caso estes estejam propensos a acidentes de trabalho. Na Tabela 11 verifica-se esta afirmação, na qual são apresentados os sectores de actividade que possuem as normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999 predominantes.

Tabela 11 - Sectores de Actividade com as Normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999 predominantes

Normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999	EAC	Descrição de Sector de Actividade
	2	Minas e Exploração Mineira
	7	Pasta, Papel e Produtos de Papel
	10	Fabricação de Coque e Produtos Petrolíferos Refinados
	13	Indústria Farmacêutica
	15	Produtos Minerais Não Metálicos
	24	Reciclagem
	25	Produção e Distribuição de Energia Eléctrica
	26	Produção e Distribuição de Gás
	27	Fornecimento de Água
	28	Construção
	39	Outros Serviços Sociais

Existem alguns sectores de actividade, como os apresentados nas Tabelas 12 e 13 que possuem um conjunto de normas diferente do resto dos sectores de actividade. Com estes resultados, conclui-se que para a maior parte dos sectores de actividade existentes e independente da região onde as empresas estão localizadas, a norma ISO 9001:2000 é sempre a principal prioridade das empresas certificadas nacionais.

Tabela 12 - Sectores de Actividade com Diversas Normas predominantes

Normas predominantes	EAC	Descrição de Sector de Actividade
Normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005	1	Agricultura e Pescas

Tabela 13 - Sectores de Actividade com Diversas Normas predominantes (continuação)

Normas predominantes	EAC	Descrição de Sector de Actividade
Normas ISO 9001:2000, ISO 22000:2005 e HACCP	3	Alimentação, Bebidas e Tabaco
Normas ISO 9001:2000, ISO 14001:2004 e ISO 22000:2005	6	Madeira e Produtos de Madeira
Normas ISO 9001:2000 e ISO 18001	16	Betão, Cimento, Cal e Gesso
Normas ISO 9001:2000, ISO 14001:2004, ISO/TS 16949	22	Outro Equipamento de Transporte

5.1.2 Por Região

Devido à predominância da Norma ISO 9001:2000, decidiu-se fazer uma investigação mais profunda e analisar os tipos de normas que os sectores de actividade possuem nas diversas regiões.

Os resultados apresentados nesta secção consistem na aplicação da técnica de mineração de dados, classificação, sobre os todos os dados do *data warehouse*. Estes resultados são diferentes dos outros na medida em que é possível analisar as diferentes combinações de certificados para cada sector de actividade em cada região analisada. Se existem sectores de actividade que não são enumerados em cada uma das regiões significa que não existem empresas certificadas inseridas nesses sectores nessa região. Adicionaram-se tabelas com os respectivos resultados para facilitar a compreensão dos resultados obtidos.

A precisão (*Accuracy*) e a taxa de erro (*Classification Error*) de cada uma das árvores obtidas, das regiões em estudo, Norte, Lisboa e Vale do Tejo, Centro, Alentejo, Região Autónoma da Madeira, Região Autónoma dos Açores e Algarve, estão apresentadas na Tabela 14.

Tabela 14 - Precisão e taxa de erro de cada árvore de decisão obtida

<i>Decision Tree</i>	Norte	Lisboa e Vale do Tejo	Centro	Alentejo	Madeira	Açores	Algarve
<i>Accuracy</i>	88, 80%	98,68%	97,73%	93,06%	72,00%	66,53%	95,48%
<i>Classification Error</i>	11,20%	1,32%	2,27%	6,94%	28,00%	33,47%	4,52%

5.1.2.1 Caracterização dos Sectores de Actividade da Região Norte

A região **Norte** é a zona que possui mais certificados e empresas certificadas nacionais. Os sectores que não são enumerados apenas possuem empresas certificadas com a norma ISO 9001:2000, como os sectores: Agricultura e Pescas (EAC=2), Construção e Reparação Naval (EAC=20), Couro e Produtos de Couro (EAC=6), Mediação Financeira, Imobiliária e Aluguer (EAC=32).

As empresas inseridas nos sectores de actividade Comércio (EAC=29), Minas e Exploração Mineira (EAC=3) e Produção e Distribuição de Energia (EAC=25) possuem as normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999. O sector de actividade Comércio é neste estudo o sector de actividade com empresas a adquirirem a maior parte dos certificados no mercado competitivo. As empresas da região Norte inseridas nos sectores de actividade Administração Pública (EAC=36) e Transporte, Armazenamento e Comunicações (EAC=31) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e HACCP; nos sectores de actividade Alimentação Bebidas e Tabaco (EAC=3) e Saúde e Serviços Sociais (EAC=38), as empresas desta região geralmente contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP, sendo estes os dois sectores com a maior combinação de tipos de certificados. Os sectores de actividade Betão, Cimento, Cal e Gesso (EAC=16) e Construção (EAC=28) possuem empresas que adquiriram as normas ISO 9001:2000 e OHSAS 18001:1999. Nos sectores de actividade Editoras (EAC=8), Equipamento Eléctrico e de Óptica (EAC=19), Produção e Distribuição de Gás (EAC=26), Reciclagem (EAC=24) e Têxteis e Produtos Têxteis (EAC=4), as empresas inseridas nestes sectores de actividade contêm as normas ISO 9001:2000 e ISO 14001:2004, as empresas

inseridas nos sectores de actividade Educação (EAC=37) e Hotéis e Restaurantes (EAC=30) possuem um conjunto de normas muito semelhante aos dos sectores apresentados anteriormente (EAC=8, EAC=26, EAC=19, EAC=24 e EAC=4) apresentando apenas no seu conjunto mais a norma ISO 22000:2005. As empresas inseridas nos sectores de actividade Empresas Gráficas (EAC=9) e Outros Serviços (EAC=35) possuem as normas ISO 9001:2000, 14001, 18001 e EMAS II. No sector de actividade Equipamentos e Máquinas (EAC=18) apenas possuem empresas que contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO/TS 16949 e em grande semelhança as empresas inseridas no sector de actividade Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e ISO/TS 16949. No sector de actividade Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14), as empresas desta região apenas possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e ISO/TS 16949.

As empresas inseridas no sector de actividade Fornecimentos de Água (EAC=27) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000. No sector de actividade Indústria Farmacêutica (EAC=13), as empresas desta região possuem as normas ISO 9001:2000, ISO 14001:2004 e NP 4457:2006 e no sector de actividade Madeira e Produtos de Madeira (EAC=6), as empresas desta região apenas apresentam as normas ISO 9001:2000, ISO 14001:2004, ISO 22000:2005 e HACCP. As empresas inseridas no sector de actividade Outras Fabricações Não Especificadas (EAC=23) possuem as normas ISO 9001:2000 e ISO/TS 16949. No sector de actividade Outro Equipamento de Transporte (EAC=22), as organizações possuem as normas ISO 9001:2000, ISO 14001:2004 e ISO/TS 16949 e no sector de actividade Outros Serviços Sociais (EAC=39), as empresas apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4469-1:2009 e SA 8000. Nos sectores de actividade Pasta, Papel e Produtos de Papel (EAC=7) e Tecnologias de Informação (EAC=33), as empresas desta região possuem as normas ISO 9001:2000, ISO 14001:2004 e EMAS II. No sector de actividade Produtos Minerais Não Metálicos (EAC=15), as empresas inseridas neste sector possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005 e as empresas inseridas no sector de actividade Químicos e Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO/TS 16949 e NP 4457:2006. E finalmente, as empresas desta região inseridas no sector de actividade Serviços de Engenharia (EAC=34) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e

NP 4457:2006. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 15 e 16.

Tabela 15 - Caracterização dos Sectores de Actividade da Região Norte

Sector de Actividade (EAC)	Normas
Agricultura e Pescas (EAC=2) Construção e Reparação Naval (EAC=20) Couro e Produtos de Couro (EAC=6) Mediação Financeira, Imobiliária e Aluguer (EAC=32)	ISO 9001:2000
Administração Pública (EAC=36) Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e HACCP
Alimentação, Bebidas e Tabaco (EAC=3) Saúde e Serviços Sociais (EAC=38)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP
Betão, Cimento, Cal e Gesso (EAC=16) Construção (EAC=28)	ISO 9001:2000 e OHSAS 18001:1999
Comércio (EAC=29) Minas e Exploração Mineira (EAC=2) Produção e Distribuição de Energia Eléctrica (EAC=25)	ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999
Editoras (EAC=8) Equipamento Eléctrico e de Óptica (EAC=19) Produção e Distribuição de Gás (EAC=26) Reciclagem (EAC=24) Têxteis e Produtos Têxteis (EAC=4)	ISO 9001:2000 e ISO 14001:2004
Educação (EAC=37) Hotéis e Restaurantes (EAC=30)	ISO 9001:2000, ISO 14001:2004 e ISO 22000:2005
Empresas Gráficas (EAC=9) Outros Serviços (EAC=35)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e EMAS II
Equipamentos e Máquinas (EAC=18)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO/TS 16949
Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e ISO/TS 16949
Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e ISO/TS 16949
Fornecimento de Água (EAC=27)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000
Indústria Farmacêutica (EAC=13)	ISO 9001:2000, ISO 14001:2004 e NP 4457:2006
Madeira e Produtos de Madeira (EAC=6)	ISO 9001:2000, ISO 14001:2004, ISO 22000:2005 e HACCP

Tabela 16 - Caracterização dos Sectores de Actividade da Região Norte (continuação)

Sector de Actividade (EAC)	Normas
Outras Fabricações Não Especificadas (EAC=23)	ISO 9001:2000 e ISO/TS 16949
Outro Equipamento de Transporte (EAC=22)	ISO 9001:2000, ISO 14001:2004 e ISO/TS 16949
Outros Serviços Sociais (EAC=39)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4469-1:2009 e SA 8000
Pasta, Papel e Produtos de Papel (EAC=7) Tecnologias de Informação (EAC=33)	ISO 9001:2000, ISO 14001:2004 e EMAS II
Produtos Minerais Não Metálicos (EAC=15)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005
Químicos e Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO/TS 16949 e NP 4457:2006
Serviços de Engenharia (EAC=34)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e NP 4457:2006

5.1.2.2 Caracterização dos Sectores de Actividade da Região de Lisboa e Vale do Tejo

A região de **Lisboa e Vale do Tejo** é a segunda zona com a maior concentração de certificados adquiridos pelas empresas divididas pelos diversos sectores de actividade. Os únicos sectores de actividade que apenas possuem empresas que só adquiriram a norma ISO 9001:2000 são Construção e Reparação Naval (EAC=20), Editoras (EAC=8), Fabricação de Coque e Produtos Petrolíferos Refinados (EAC=10), Indústria Aeroespacial (EAC=21) e Minas e Exploração Mineira (EAC=2).

O sector de actividade que possui o maior número de certificados nesta região e o maior conjunto de certificados adquiridos pelas empresas é o sector de actividade Comércio (EAC=29), no qual as empresas apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP, ISO/TS 16949, NP 4457:2006, SA 8000. O conjunto de sectores de actividade Administração Pública (EAC=36), Agricultura e Pescas (EAC=1), Indústria Farmacêutica (EAC=13), Produção e Distribuição de Energia Eléctrica (EAC=25), Produção e Distribuição de Gás

(EAC=26), Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12), Saúde e Serviços Sociais (EAC=38) são o maior conjunto de sectores de actividade nesta região que contêm empresas que adquiriram o mesmo tipo de certificados, as normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999. Seguem-se os sectores de actividade Equipamento Eléctrico e de Óptica (EAC=19), Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17), Outras Fabricações Não Especificadas (EAC=23), Outro Equipamento de Transporte (EAC=22), Produtos Minerais Não Metálicos (EAC=15) que contêm empresas que possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO/TS 16949.

As empresas desta região que se encontram inseridas no sector de actividade Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) e Transporte, Armazenamento e Comunicações (EAC=31) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO/TS 16949 e SA 8000. As empresas inseridas nos sectores de actividade Fornecimento de Água (EAC=27) e Mediação Financeira, Imobiliária e Aluguer (EAC=32) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000. Nos sectores de actividade Empresas Gráficas (EAC=9) e Equipamentos e Máquinas (EAC=18) as empresas possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e EMAS II. As empresas inseridas no sector de actividade Alimentação, Bebidas e Tabaco (EAC=3) apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, HACCP e NP 4469-1:2009. No sector de actividade Betão, Cimento, Cal e Gesso (EAC=16), as empresas apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP. O sector de actividade Construção (EAC=28) possui empresas que apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP, NP 4457:2006 e SA 8000. As empresas inseridas no sector de actividade Educação (EAC=37) apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e HACCP. O sector de actividade Hotéis e Restaurantes (EAC=30) contém empresas que possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e EMAS II. As empresas que se encontram inseridas no sector de actividade Outros Serviços (EAC=35) contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP, NP 4457:2006 e SA 8000. As empresas inseridas no sector de actividade Outros Serviços Sociais (EAC=39) possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP e NP 4469-1:2009. As empresas localizadas no sector de actividade Pasta, Papel e Produtos de Papel (EAC=7) possuem as normas ISO 9001:2000, ISO 14001:2004,

OHSAS 18001:1999, EMAS II e NP 4457:2006. No sector de actividade Reciclagem (EAC=24) as empresas apresentam as normas ISO 9001:2000, ISO 14001:2004 e EMAS II. No sector de Serviços de Engenharia (EAC=34), as empresas possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4457:2006 e SA 8000. As empresas que se encontram localizadas no sector de actividade Tecnologias de Informação (EAC=33) contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, HACCP e NP 4457:2006. E finalmente, os sectores de actividade Têxteis e Produtos Têxteis (EAC=4) e Madeira e Produtos de Madeira (EAC=6) são os dois sectores de actividade que contêm empresas com o menor conjunto de normas. Apenas apresentam as normas ISO 9001:2000, ISO/TS 16949 e as normas ISO 9001:2000 e ISO 22000:2005 respectivamente. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 17 e 18.

Tabela 17 - Caracterização dos Sectores de Actividade da Região de Lisboa e Vale do Tejo

Sector de Actividade (EAC)	Normas
Construção e Reparação Naval (EAC=20) Editoras (EAC=8) Fabricação de Coque e Produtos Petrolíferos Refinados (EAC=10) Indústria Aeroespacial (EAC=21) Minas e Exploração Mineira (EAC=2)	ISO 9001:2000
Comércio (EAC=29)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP, ISO/TS 16949, NP 4457:2006, SA 8000
Administração Pública (EAC=36) Agricultura e Pescas (EAC=1) Indústria Farmacêutica (EAC=13) Produção e Distribuição de Energia Eléctrica (EAC=25) Produção e Distribuição de Gás (EAC=26) Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) Saúde e Serviços Sociais (EAC=38)	ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999
Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO/TS 16949 e SA 8000

Tabela 18 - Caracterização dos Sectores de Actividade da região de Lisboa e Vale do Tejo (continuação)

Sector de Actividade (EAC)	Normas
Equipamento Eléctrico e de Óptica (EAC=19) Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) Outras Fabricações Não Especificadas (EAC=23) Outro Equipamento de Transporte (EAC=22) Produtos Minerais Não Metálicos (EAC=15)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO/TS 16949
Fornecimento de Água (EAC=27) Mediação Financeira, Imobiliária e Aluguer (EAC=32)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000
Empresas Gráficas (EAC=9) Equipamentos e Máquinas (EAC=18)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e EMAS II
Alimentação, Bebidas e Tabaco (EAC=3)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, HACCP e NP 4469-1:2009
Betão, Cimento, Cal e Gesso (EAC=16)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP
Construção (EAC=28)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP, NP 4457:2006 e SA 8000
Educação (EAC=37)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e HACCP
Hotéis e Restaurantes (EAC=30)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e EMAS II
Outros Serviços (EAC=35)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP, NP 4457:2006 e SA 8000
Outros Serviços Sociais (EAC=39)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP e NP 4469-1:2009
Pasta, Papel e Produtos de Papel (EAC=7)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e NP 4457:2006
Reciclagem (EAC=24)	ISO 9001:2000, ISO 14001:2004 e EMAS II
Serviços de Engenharia (EAC=34)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4457:2006 e SA 8000
Tecnologias de Informação (EAC=33)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, HACCP e NP 4457:2006
Têxteis e Produtos Têxteis (EAC=4)	ISO 9001:2000, ISO/TS 16949
Madeira e Produtos de Madeira (EAC=6)	ISO 9001:2000 e ISO 22000:2005

5.1.2.3 Caracterização dos Sectores de Actividade da Região Centro

A região **Centro** é a terceira zona a possuir a maior concentração de certificados adquiridos pelas empresas. É única região em que não existe nenhum sector de actividade possuindo empresas apenas com a norma ISO 9001:2000. O sector de actividade que apresenta o maior número de empresas certificadas é o sector de actividade Comércio (EAC=29), e as empresas inseridas neste sector desta região possuem geralmente ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP e ISO/TS 16949.

Existem dois conjuntos compostos por diversos sectores de actividade que correspondem aos conjuntos de sectores de actividade que possuem a maior parte das empresas com a mesma combinação de tipos de certificados, os dois conjuntos de sectores de actividade são: Reciclagem (EAC=24), Produção e Distribuição de Energia Eléctrica (EAC=25), Outras Fabricações Não Especificadas (EAC=23), Indústria Farmacêutica (EAC=13), Betão, Cimento, Cal e Gesso (EAC=16); Construção (EAC=28), em que as empresas apresentam os certificados ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e EMAS II; e o conjunto de sectores de actividade: Couro e Produtos de Couro (EAC=5), Editoras (EAC=8), Empresas Gráficas (EAC=9), Serviços de Engenharia (EAC=34), Produção e Distribuição de Gás (EAC=26), Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) em que as empresas possuem as três principais normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999.

As empresas que apenas possuem as duas principais normas (ISO 9001:2000 e ISO 14001:2004) incluem-se nos sectores de actividade Mediação Financeira, Imobiliária e Aluguer (EAC=32), Tecnologias de Informação (EAC=33), Saúde e Serviços Sociais (EAC=38) e Fornecimento de Água (EAC=27). As empresas inseridas nos sectores de actividade Minas e Exploração Mineira (EAC=2), Equipamentos e Máquinas (EAC=18), Produtos Mineraiis Não Metálicos (EAC=15) geralmente nesta região possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e EMAS II.

Nos sectores de actividade Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17), Outro Equipamento de Transporte (EAC=22) e Têxteis e Produtos Têxteis (EAC=4), as empresas contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e ISO/TS 16949. Nos

sectores de actividade Madeira e Produtos de Madeira (EAC=6), Transporte, Armazenamento e Comunicações (EAC=31), as empresas desta região apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP. O sector de actividade Alimentação Bebidas e Tabaco (EAC=3) apenas apresentam empresas com as normas ISO 9001:2000, ISO 14001:2004, ISO 22000:2005, HACCP e SA 8000.

As empresas do sector de actividade Equipamento Eléctrico e de Óptica (EAC=19) apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II, ISO/TS 16949 e NP 4457:2006. No sector de actividade Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) as empresas desta região possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II e ISO/TS 16949.

O sector de actividade Hotéis e Restaurantes (EAC=30) apresenta empresas que apenas possuem as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e HACCP. As empresas inseridas no sector de actividade Outros Serviços (EAC=35) apenas apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4457:2006 e SA 8000. No sector de actividade Outros Serviços Sociais (EAC=39) as empresas geralmente contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000.

O sector de actividade Pasta, Papel e Produtos de Papel (EAC=7) apresenta empresas que contêm as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e NP 4457:2006. As empresas contidas no sector de actividade Administração Pública (EAC=36) possuem geralmente as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005. As empresas inseridas no sector de actividade Construção (EAC=28) apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e NP 4457:2006, um conjunto de tipos de certificados bastante alargado em relação às outras regiões. No sector de actividade Agricultura e Pescas (EAC=1) as empresas possuem as normas ISO 9001:2000, ISO 14001:2004, ISO 22000:2005, HACCP e ISO/TS 16949. E finalmente, o sector de actividade Educação (EAC=37) é o sector de actividade que apresenta empresas com o menor conjunto de tipos de normas: ISO 9001:2000 e OHSAS 18001:1999. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 19 e 20.

Tabela 19 - Caracterização dos Sectores de Actividade da Região Centro

Sector de Actividade (EAC)	Normas
Comércio (EAC=29)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II, HACCP e ISO/TS 16949
Reciclagem (EAC=24) Produção e Distribuição de Energia Eléctrica (EAC=25) Outras Fabricações Não Especificadas (EAC=23) Indústria Farmacêutica (EAC=13) Betão, Cimento, Cal e Gesso (EAC=16) Construção (EAC=28)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e EMAS II
Couro e Produtos de Couro (EAC=5) Editoras (EAC=8) Empresas Gráficas (EAC=9) Serviços de Engenharia (EAC=34) Produção e Distribuição de Gás (EAC=26) Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12)	ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999
Mediação Financeira, Imobiliária e Aluguer (EAC=32) Tecnologias de Informação (EAC=33) Saúde e Serviços Sociais (EAC=38) Fornecimento de Água (EAC=27)	ISO 9001:2000 e ISO 14001:2004
Minas e Exploração Mineira (EAC=2) Equipamentos e Máquinas (EAC=18) Produtos Minerais Não Metálicos (EAC=15)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e EMAS II
Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) Outro Equipamento de Transporte (EAC=22) Têxteis e Produtos Têxteis (EAC=4)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e ISO/TS 16949
Madeira e Produtos de Madeira (EAC=6) Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP
Alimentação Bebidas e Tabaco (EAC=3)	ISO 9001:2000, ISO 14001:2004, ISO 22000:2005, HACCP e SA 8000
Equipamento Eléctrico e de Óptica (EAC=19)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II, ISO/TS 16949 e NP 4457:2006
Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005, EMAS II e ISO/TS 16949
Hotéis e Restaurantes (EAC=30)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e HACCP

Tabela 20 - Caracterização dos Sectores de Actividade da Região Centro (continuação)

Sector de Actividade (EAC)	Normas
Outros Serviços (EAC=35)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, NP 4457:2006 e SA 8000
Outros Serviços Sociais (EAC=39)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e SA 8000
Pasta, Papel e Produtos de Papel (EAC=7)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e NP 4457:2006
Administração Pública (EAC=36)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005
Construção (EAC=28)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e NP 4457:2006
Agricultura e Pescas (EAC=1)	ISO 9001:2000, ISO 14001:2004, ISO 22000:2005, HACCP e ISO/TS 16949
Educação (EAC=37)	ISO 9001:2000 e OHSAS 18001:1999

5.1.2.4 Caracterização dos Sectores de Actividade da Região Alentejo

A região do **Alentejo** é a quarta região que apresenta o maior número de empresas. Existem empresas que só adquirirem a norma ISO 9001:2000, estas encontram-se inseridas nos sectores de actividade Agricultura e Pesca (EAC=1), Madeira e Produtos de Madeira (EAC=6), Mediação Financeira, Imobiliária e Aluguer (EAC=32), Produtos Minerais Não Metálicos (EAC=15) e Serviços de Engenharia (EAC=34). Apesar do sector de actividade Comércio (EAC=29) ser o sector de actividade com o maior número de empresas certificadas nesta região, as empresas inseridas neste sector desta região apenas possuem as normas ISO 9001:2000, EMAS II e ISO/TS 16949.

O conjunto de sectores de actividade que possui o maior número de sectores de actividade que têm empresas com o mesmo conjunto de tipos de certificados é composto pelos sectores de actividade: Reciclagem (EAC=24), Saúde e Serviços Sociais (EAC=38), Transporte, Armazenamento e Comunicações (EAC=31), Produção e Distribuição de Energia Eléctrica (EAC=25), Produção e Distribuição de gás (EAC=26), Minas e Exploração Mineira (EAC=2), Fabricação de Coque e Produtos Refinados (EAC=10), Educação (EAC=37), Administração Pública (EAC=36), Betão, Cimento, Cal e Gesso (EAC=16); Construção (EAC=28). E as empresas que se

encontram inseridas neste conjunto de sectores de actividade e localizadas neste região geralmente apresentam as normas ISO 9001:2000 e OHSAS 18001:1999.

As empresas inseridas no sector de actividade Pasta, Papel e Produtos de Papel (EAC=7) apresentam as normas ISO 9001:2000, OHSAS 18001:1999 e EMAS II. No sector de actividade Outros Serviços Sociais (EAC=39), as empresas possuem geralmente os certificados ISO 9001:2000, OHSAS 18001:1999, EMAS II e SA 8000. O sector de actividade Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) contém empresas que apresentam as normas ISO 9001:2000, OHSAS 18001:1999, ISO 22000:2005 e ISO/TS 16949. Os sectores de actividade enumerados seguidamente são os sectores de actividade que contêm o menor conjunto de tipo de certificados: Alimentação Bebidas e Tabaco (EAC=3), em que as empresas apenas apresentam as normas ISO 9001:2000 e ISO 22000:2005. No sector de actividade Empresas Gráficas (EAC=9), as empresas apenas contêm as normas ISO 9001:2000 e EMAS II. As empresas contidas nos sectores de actividade Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17), Outro Equipamento de Transporte (EAC=22) e Têxteis e Produtos Têxteis (EAC=4) apenas possuem as normas ISO 9001:2000 e ISO/TS 16949. Nos sectores de actividade Hotéis e Restaurantes (EAC=30) e Outras Fabricações Não Especificadas (EAC=23), as empresas desta região geralmente contêm as normas ISO 9001:2000 e HACCP. E finalmente as empresas inseridas no sector de actividade Outros Serviços (EAC=35) apresentam as normas ISO 9001:2000 e SA 8000. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 21 e 22.

Tabela 21 - Caracterização dos Sectores de Actividade da Região Alentejo

Sector de Actividade (EAC)	Normas
Agricultura e Pesca (EAC=1) Madeira e Produtos de Madeira (EAC=6) Mediação Financeira, Imobiliária e Aluguer (EAC=32) Produtos Minerais Não Metálicos (EAC=15) Serviços de Engenharia (EAC=34)	ISO 9001:2000
Comércio (EAC=29)	ISO 9001:2000, EMAS II e ISO/TS 16949

**Tabela 22 - Caracterização dos Sectores de Actividade da Região Alentejo
(continuação)**

Sector de Actividade (EAC)	Normas
Reciclagem (EAC=24) Saúde e Serviços Sociais (EAC=38) Transporte, Armazenamento e Comunicações (EAC=31) Produção e Distribuição de Energia Eléctrica (EAC=25) Produção e Distribuição de gás (EAC=26) Minas e Exploração Mineira (EAC=2) Fabricação de Coque e Produtos Refinados (EAC=10) Educação (EAC=37) Administração Pública (EAC=36) Betão, Cimento, Cal e Gesso (EAC=16) Construção (EAC=28)	ISO 9001:2000 e OHSAS 18001:1999
Pasta, Papel e Produtos de Papel (EAC=7)	ISO 9001:2000, OHSAS 18001:1999 e EMAS II
Outros Serviços Sociais (EAC=39)	ISO 9001:2000, OHSAS 18001:1999, EMAS II e SA 8000
Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14)	ISO 9001:2000, OHSAS 18001:1999, ISO 22000:2005 e ISO/TS 16949
Alimentação Bebidas e Tabaco (EAC=3)	ISO 9001:2000 e ISO 22000:2005
Empresas Gráficas (EAC=9)	ISO 9001:2000 e EMAS II
Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) Outro Equipamento de Transporte (EAC=22) Têxteis e Produtos Têxteis (EAC=4)	ISO 9001:2000 e ISO/TS 16949
Hotéis e Restaurantes (EAC=30) Outras Fabricações Não Especificadas (EAC=23)	ISO 9001:2000 e HACCP
Outros Serviços (EAC=35)	ISO 9001:2000 e SA 8000

5.1.2.5 Caracterização dos Sectores de Actividade da Região Algarve

A região do **Algarve** é a quinta zona com o maior número de empresas certificadas e normas adquiridas por estas. O sector de actividade que possui o maior número de empresas certificadas e normas adquiridas por estas é o sector de actividade Comércio (EAC=28), no qual as empresas apresentam as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP. O conjunto de sectores de actividade que contém empresas apenas com a norma ISO

9001:2000 é bastante vasto: Administração Pública (EAC=36), Educação (EAC=37), Empresas Gráficas (EAC=9), Equipamento Eléctrico e de Óptica (EAC=19), Equipamentos e Máquinas (EAC=18), Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14), Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17), Madeira e Produtos de Madeira (EAC=6), Mediação Financeira, Imobiliária e Aluguer (EAC=32), Minas e Exploração Mineira (EAC=2), Outras Fabricações Não Especificadas (EAC=23), Produção e Distribuição de Gás (EAC=26), Produtos Minerais Não Metálicos (EAC=15), Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) e Serviços de Engenharia (EAC=34).

O conjunto de sectores de actividade que possui o maior número de empresas com o mesmo tipo de certificados é o conjunto composto pelos sectores de actividade: Agricultura e Pescas (EAC=1), Outros Serviços (EAC=35), Outros Serviços Sociais (EAC=39), Saúde e Serviços Sociais (EAC=38), Transporte, Armazenamento e Comunicações (EAC=31) em que as empresas inseridas nestes sectores apresentam as normas ISO 9001:2000 e ISO 14001:2004. Os únicos sectores de actividade que possuem empresas que adquiriram os três certificados principais (ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999) são: Betão, Cimento, Cal e Gesso (EAC=16) e Construção (EAC=29).

As empresas inseridas no sector de actividade Fornecimento de Água (EAC=27) além das três principais normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, apresentam também a norma ISO 22000:2005. No sector de actividade Hotéis e Restaurantes (EAC=30) possui empresas que apresentam as normas ISO 9001:2000, ISO 14001:2004 e HACCP. E finalmente, o sector de actividade Alimentação, Bebidas e Tabaco (EAC=3) é o sector que possui empresas com o menor conjunto de tipos de certificados, em que as empresas apresentam apenas as normas ISO 9001:2000 e ISO 22000:2005. Todos os resultados apresentados nesta secção podem ser consultados na Tabela 23.

Tabela 23 - Caracterização dos Sectores de Actividade da Região Algarve

Sector de Actividade (EAC)	Normas
Administração Pública (EAC=36) Educação (EAC=37) Empresas Gráficas (EAC=9) Equipamento Eléctrico e de Óptica (EAC=19) Equipamentos e Máquinas (EAC=18) Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) Madeira e Produtos de Madeira (EAC=6) Mediação Financeira, Imobiliária e Aluguer (EAC=32) Minas e Exploração Mineira (EAC=2) Outras Fabricações Não Especificadas (EAC=23) Produção e Distribuição de Gás (EAC=26) Produtos Minerais Não Metálicos (EAC=15) Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) Serviços de Engenharia (EAC=34).	ISO 9001:2000
Comércio (EAC=29)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, ISO 22000:2005 e HACCP
Agricultura e Pescas (EAC=1) Outros Serviços (EAC=35) Outros Serviços Sociais (EAC=39) Saúde e Serviços Sociais (EAC=38) Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000 e ISO 14001:2004
Betão, Cimento, Cal e Gesso (EAC=16) Construção (EAC=29)	ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999
Fornecimento de Água (EAC=27)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999 e ISO 22000:2005
Hotéis e Restaurantes (EAC=30)	ISO 9001:2000, ISO 14001:2004 e HACCP
Alimentação, Bebidas e Tabaco (EAC=3)	ISO 9001:2000 e ISO 22000:2005

5.1.2.6 Caracterização dos Sectores de Actividade da Região Madeira

A região da **Madeira** é a sexta zona com o maior número de empresas certificadas e normas adquiridas por estas. O conjunto de sectores de actividade que contém empresas apenas com a

norma ISO 9001:2000 é bastante vasto: Administração Pública (EAC=36); Betão, Cimento, Cal e Gesso (EAC=16), Construção (EAC=28); Educação (EAC=37); Equipamento Eléctrico e de Óptica (EAC=19); Equipamentos e Máquinas (EAC=18); Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14); Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17); Fornecimento de Água (EAC=27); Minas e Exploração Mineira (EAC=2); Outras Fabricações Não Especificadas (EAC=23); Pasta, Papel e Produtos de Papel (EAC=7); Produção e Distribuição de Energia Eléctrica (EAC=25); Produção e Distribuição de Gás (EAC=26); Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12); Saúde e Serviços Sociais (EAC=38); Serviços de Engenharia (EAC=34).

Os sectores de actividade Mediação Financeira, Imobiliária e Aluguer (EAC=32), Outros Serviços (EAC=35), Tecnologias de Informação (EAC=33) e Transporte, Armazenamento e Comunicações (EAC=31), possuem empresas certificadas com as normas ISO 9001:2000 e ISO 14001:2004, representando o maior conjunto de sectores de actividade com o maior conjunto de mesmo tipo de certificados. Os sectores de actividade Alimentação, Bebidas e Tabaco (EAC=3), Hotéis e Restaurantes (EAC=30), Comércio (EAC=29) contêm empresas certificadas com as normas ISO 9001:2000, ISO 14001:2004 e HACCP. As empresas inseridas nos sectores de actividade Madeira e Produtos de Madeira (EAC=6) apresentam as normas ISO 9001:2000, HACCP, ISO 14001:2004 e ISO 22000:2005. E finalmente, o sector de actividade Outros Serviços Sociais (EAC=39) apenas possui empresas que apresentam as normas ISO 9001:2000 e OHSAS 18001:1999. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 24 e 25.

Tabela 24 - Caracterização dos Sectores de Actividade da Região Autónoma da Madeira

Sector de Actividade (EAC)	Normas
Administração Pública (EAC=36) Betão, Cimento, Cal e Gesso (EAC=16) Construção (EAC=28) Educação (EAC=37) Equipamento Eléctrico e de Óptica (EAC=19) Equipamentos e Máquinas (EAC=18)	ISO 9001:2000

Tabela 25 - Caracterização dos Sectores de Actividade da Região Autónoma da Madeira (continuação)

Sector de Actividade (EAC)	Normas
Fabricação de Artigos de Borracha e Matérias Plásticas (EAC=14) Fabricação Metalúrgica de Base e Produtos Metálicos (EAC=17) Fornecimento de Água (EAC=27) Minas e Exploração Mineira (EAC=2) Outras Fabricações Não Especificadas (EAC=23) Pasta, Papel e Produtos de Papel (EAC=7) Produção e Distribuição de Energia Eléctrica (EAC=25) Produção e Distribuição de Gás (EAC=26) Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) Saúde e Serviços Sociais (EAC=38) Serviços de Engenharia (EAC=34).	ISO 9001:2000
Mediação Financeira, Imobiliária e Aluguer (EAC=32) Outros Serviços (EAC=35) Tecnologias de Informação (EAC=33) Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000 e ISO 14001:2004
Alimentação, Bebidas e Tabaco (EAC=3) Hotéis e Restaurantes (EAC=30) Comércio (EAC=29)	ISO 9001:2000, ISO 14001:2004 e HACCP
Madeira e Produtos de Madeira (EAC=6)	ISO 9001:2000, HACCP, ISO 14001:2004 e ISO 22000:2005
Outros Serviços Sociais (EAC=39)	ISO 9001:2000 e OHSAS 18001:1999

5.1.2.7 Caracterização dos Sectores de Actividade da Região Açores

A região dos **Açores** é a zona analisada com o menor número de empresas certificadas e certificados adquiridos por estas. O sector de actividade que possui o maior número de empresas certificadas nesta região é o sector de actividade Comércio (EAC=29), apesar disto e ao contrário do resto das regiões, as empresas certificadas desta região inseridas neste sector apenas apresentam as normas ISO 9001:2000, OHSAS 18001:1999 e HACCP. O conjunto de sectores de

actividade que contém empresas apenas com a norma ISO 9001:2000 é bastante vasto: Educação (EAC=37); Equipamento Eléctrico e de Óptica (EAC=19); Equipamentos e Máquinas (EAC=18); Mediação Financeira, Imobiliária e Aluguer (EAC=32); Outra Fabricações Não Especificadas (EAC=23); Outros Serviços (EAC=35); Pasta, Papel e Produtos de Papel (EAC=6); Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12); Saúde e Serviços Sociais (EAC=38) e Serviços de Engenharia (EAC=34).

Os sectores de actividade Betão Cimento Cal e Gesso (EAC=16), Construção (EAC=28), Minas e Exploração Mineira (EAC=2) e Produção e Distribuição de Gás (EAC=26) contém empresas que apresentam somente as normas ISO 9001:2000 e OHSAS 18001:1999. As empresas inseridas no sector de actividade Hotéis e Restaurantes (EAC=30) possuem geralmente nesta região as normas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP e EMAS II, isto deve-se ao facto de este sector ser uma grande aposta nesta zona em relação às outras zonas, em que este sector apresenta empresas com um menor conjunto de certificados. As empresas inseridas no sector de actividade Administração Pública (EAC=36) apresentam as normas ISO 9001:2000, ISO 14001:2004 e EMAS II. No sector Alimentação Bebidas e Tabaco (EAC=3) as empresas possuem as normas ISO 9001:2000, HACCP e ISO 22000:2005. O sector de actividade Transporte, Armazenamento e Comunicações (EAC=31) contém empresas que apresentam as normas ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999. E finalmente, os sectores de actividade Empresas Gráficas (EAC=9) e Reciclagem (EAC=24), que possuem empresas que apresentam as normas ISO 9001:2000 e ISO 14001:2004. Todos os resultados apresentados nesta secção podem ser consultados nas Tabelas 26 e 27.

Tabela 26 - Caracterização dos Sectores de Actividade da Região Autónoma dos Açores

Sector de Actividade (EAC)	Normas
Educação (EAC=37) Equipamento Eléctrico e de Óptica (EAC=19) Equipamentos e Máquinas (EAC=18) Mediação Financeira, Imobiliária e Aluguer (EAC=32) Outra Fabricações Não Especificadas (EAC=23) Outros Serviços (EAC=35) Pasta, Papel e Produtos de Papel (EAC=6)	ISO 9001:2000

Tabela 27 - Caracterização dos Sectores de Actividade da Região Autónoma dos Açores (continuação)

Sector de Actividade (EAC)	Normas
Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais (EAC=12) Saúde e Serviços Sociais (EAC=38) Serviços de Engenharia (EAC=34)	ISO 9001:2000
Comércio (EAC=29)	ISO 9001:2000, OHSAS 18001:1999 e HACCP
Betão Cimento Cal e Gesso (EAC=16) Construção (EAC=28) Minas e Exploração Mineira (EAC=2) Produção e Distribuição de Gás (EAC=26)	ISO 9001:2000 e OHSAS 18001:1999
Hotéis e Restaurantes (EAC=30)	ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, HACCP e EMAS II
Administração Pública (EAC=36)	ISO 9001:2000, ISO 14001:2004 e EMAS II
Alimentação Bebidas e Tabaco (EAC=3)	ISO 9001:2000, HACCP e ISO 22000:2005
Transporte, Armazenamento e Comunicações (EAC=31)	ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999
Empresas Gráficas (EAC=9) Reciclagem (EAC=24)	ISO 9001:2000 e ISO 14001:2004

5.2 Caracterização dos Sectores usando Segmentação

Para a obtenção dos resultados foram executados vários testes variando o valor k , com objectivo de minimizar o valor de DB , para conseguir o número óptimo de divisão. Os valores obtidos são apresentados nas Tabela 28 e 29.

Tabela 28 - Resultados de valores de *Davies Bouldin*

Número de segmentos	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Valor de Davies Bouldin	-0,488	-0,501	-0,533	-0,700	-0,730	-0,794	-0,849

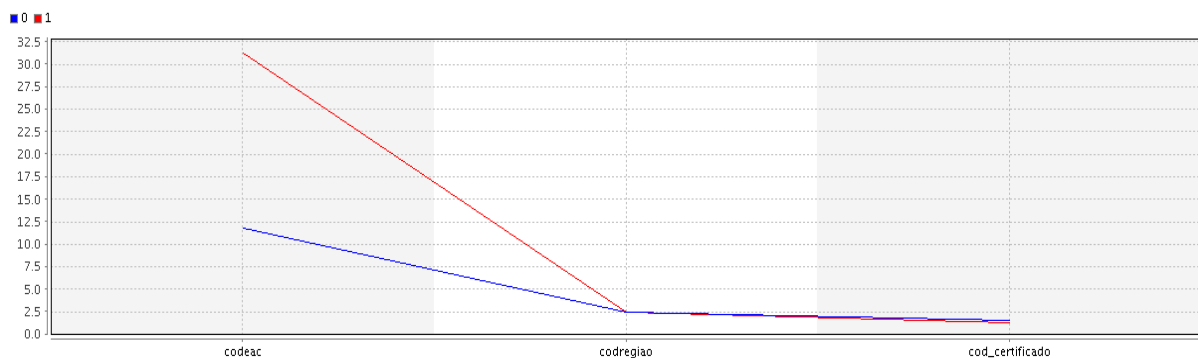
Tabela 29 - Resultados de valores de *Davies Bouldin* (continuação)

Número de segmentos	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Valor de Davies Bouldin	-0,789	-0,952	-0,972	-0,842	-0,912	0,877	-0,923

O número óptimo de segmentos obtido foi $k=2$, com $DB=-0,488$. Este valor pode ser explicado pela fraca variedade dos dados. Como foi dito nas secções anteriores, a quantidade de certificados ISO 9001:2000 adquiridos pelas empresas, corresponde a uma percentagem de 90% de todos os certificados obtidos pelas empresas nacionais. Devido a estas características do conjunto de dados, a aplicação desta técnica vai possibilitar ao utilizador uma noção do estado do mercado competitivo, podendo descobrir quais as regiões que possuem o maior número de sistemas de gestão adquiridos pelas empresas, quais os certificados mais adquiridos pelas organizações e quais os sectores mais concorridos no país.

Tabela 30 - Composição dos segmentos obtidos

Segmento 1	15084 registos
Segmento 2	8848 registos

**Figura 19 - Representação dos Centróides gerados**

Após a aplicação do algoritmo resultaram dois segmentos bastante diferentes a nível de tamanho, um com cerca de 63% dos registos e o outro com cerca de 37%, respectivamente com 15084 e 8848 registos como se pode verificar na Tabela 30. O segmento com mais dados, como é possível verificar no gráfico contido na imagem, envolve os sectores de actividade que apresentam o maior número de certificados obtidos pelas empresas certificadas nacionais, como Comércio e Construção. Os certificados mais obtidos por empresas inseridas nestes sectores são especialmente as normas ISO 9001:2000 (Certificação de Sistemas de Gestão de Qualidade), ISO 14001:2004 (Certificação de Sistemas de Gestão Ambiental) e OHSAS 18001:1999 (Certificação de Sistemas da Segurança e Saúde no Trabalho). Este último tem especial importância para as empresas inseridas no sector de Construção. Estes resultados são apresentados no gráfico da Figura 19. O gráfico não permite analisar quais as regiões que contêm a maior parte das empresas nacionais certificadas, permitindo apenas concluir que se encontram localizadas na região do Norte, Lisboa e Vale do Tejo, Centro e Alentejo.

O outro segmento apenas contém os restantes 37% dos registos do conjunto de dados. Analisando o gráfico é possível verificar que as empresas certificadas que se encontram inseridas nos restantes sectores de actividade se encontram localizadas especialmente nas regiões já em cima enumeradas, como a região do Norte, Lisboa e Vale do Tejo, Centro e Alentejo, e estas possuem essencialmente as três normas já em cima enumeradas igualmente, que possibilitam um Sistema de Gestão de Qualidade, um Sistema de Gestão Ambiental e um Sistema de Gestão de Segurança e Saúde no Trabalho eficazes.

A análise exploratória feita no capítulo 3 possibilitou uma melhor compreensão do conteúdo da base de dados fornecida, sendo possível com ela e com os resultados obtidos pelo estudo de Sampaio & Saraiva (2008, 2009, 2010) justificar os resultados obtidos pela segmentação. Estes resultados não foram muito conclusivos devido à fraca diversidade dos resultados. Isto acontece devido à vasta quantidade de registos correspondentes às normas ISO 9001:2000 e de ISO 14001:2004, que permitem às empresas um sistema de gestão de qualidade e ambiental eficazes e que correspondem a um total de 95% de todas os certificados obtidos pelas empresas. Os outros 5% correspondem aos restantes certificados: OHSAS 18001:1999 (Certificação de Sistemas da Segurança e Saúde no Trabalho), ISO 22000:2005 (Certificação de Segurança Alimentar), EMAS II (Verificação EMAS II), HACCP (Certificação HACCP – Segurança Alimentar), ISO/TS 16949

(Certificação de Sistemas de Gestão de Qualidade no Sector Automóvel), NP 4457:2006 (Certificação de Sistemas de Gestão de Investigação, Desenvolvimento e Inovação), NP 4469-1:2009 (Certificação de Sistemas de Gestão da Responsabilidade Social) e SA 8000 (Certificação de Sistemas de Gestão de Responsabilidade Social).

Em relação à análise quantitativa dos registos divididos pelas diversas Regiões, é possível verificar que grande parte das empresas nacionais certificadas se encontra dividida pelas regiões Norte, Lisboa e Vale do Tejo e Centro. Em relação à última variante nos dados, o sector de actividade, apesar da grande quantidade de sectores de actividade, que ao todo são 38, cerca de 15% de todas as empresas nacionais certificadas estão inseridas no sector de actividade Comércio e 11% correspondem às empresas inseridas no sector de actividade Construção.

Capítulo 6

Conclusões e Trabalho Futuro

6.1 Processos e Resultados

Neste trabalho foram usadas técnicas de mineração de dados num sistema de certificação de empresas contendo informação sobre o mercado competitivo que se vive actualmente, tendo como principal objectivo descobrir padrões escondidos. Esta informação extraída é considerada muito importante para as empresas terem uma primeira orientação nos primeiros anos de existência. A informação extraída proporciona aos responsáveis das organizações uma noção sobre a caracterização dos sectores de actividade existentes em Portugal e sobre as características das empresas distribuídas pelas várias regiões nacionais. Os resultados obtidos neste trabalho contribuem assim para que as empresas consigam uma vantagem competitiva no mercado.

Quando se iniciou este trabalho, parecia haver evidências que se pretendiam analisar com as técnicas de mineração de dados. Por exemplo, a escolha de certificados pela empresa pode variar com o sector de actividade em que está inserida ou a região em que se encontra localizada. O conjunto de certificados necessário para uma empresa conseguir sobreviver varia de sector para sector de actividade, ou seja, existem sectores de actividade em que é necessário possuir certificados mais ligados com a protecção ambiental (exemplo: fábricas que provocam poluição

ambiental), certificados ligados à protecção alimentar (exemplo: empresas de produção alimentar) ou certificados relacionados com a segurança do trabalhador (exemplo: empresas de construção).

Por outro lado, existe uma variação de certificados por região, que se deve ao facto de as diversas regiões nacionais possuírem diferentes densidades de empresas. Uma das justificações para isto é a seguinte: as empresas localizadas em regiões com uma maior densidade de empresas possuem um maior conjunto de certificados, porque, para uma nova organização conseguir sobreviver neste cenário, é necessário que adquira os certificados necessários para que se torne uma preferência dos potenciais clientes.

Para a realização deste trabalho e consequente obtenção destes resultados foi seguida a metodologia *CRISP-DM*. Esta metodologia ajudou a uma orientação e acompanhamento ao longo deste problema de mineração de dados, incluindo um estudo sobre o negócio e os termos teóricos, um estudo sobre as técnicas de mineração de dados existentes e respectivas vantagens, um plano de quais seriam os resultados obtidos e sobre os dados que deviam ser seleccionados para o estudo. Esta metodologia tem a vantagem de ser independente da ferramenta que se usa para a resolução do problema.

Para facilitar a compreensão dos dados e decidir a melhor maneira para abordar e resolver o problema, foi necessário realizar uma investigação sobre o negócio e os termos teóricos relativos ao problema. A informação disponibilizada pelos principais responsáveis, como empresas de auditoria e certificação, e o facto de serem temas bastante abordados, é bastante rica, o que facilitou o estudo dos principais certificados, como o certificado ISO 9001:2000, ISO 14001:2004 e OHSAS 18001:1999. Em relação aos outros certificados verificou-se uma certa dificuldade em aprofundar o tema, porque, como são relativamente recentes, ainda não existe muita informação sobre estes.

A investigação inicial facilitou bastante a compreensão do conjunto de dados, permitindo perceber facilmente a importância de cada um dos certificados, que era evidenciada pelos próprios dados: os certificados considerados mais importantes pelas empresas responsáveis eram os certificados mais adquiridos pelas empresas. Para complementar a compreensão dos restantes campos contidos no conjunto de dados (*EAC*) foi realizada uma investigação adicional, envolvendo os artigos que o Professor Paulo Sampaio realizou juntamente com o Professor Pedro Saraiva. Estes

artigos ajudaram a corresponder os números de EAC às respectivas designações de sectores de actividade.

Durante a fase de limpeza dos dados e selecção do conjunto de dados a que iriam ser aplicados os algoritmos de mineração, surgiram algumas dificuldades. Os dados apresentavam uma codificação diferente daquela que é permitida na base de dados final. Este problema foi resolvido sem grande demora. Outro problema relevante foi o facto de existirem diferentes designações para uma mesma empresa no conjunto de dados inicial, criando registos duplicados de certas empresas no *data warehouse* final.

Não havia maneira de controlar as designações diferentes pois não existia qualquer campo no conjunto de dados que identificasse uma entidade, como, por exemplo, o número de contribuinte de uma empresa. O único campo que identifica uma empresa no conjunto de dados inicial é a sua designação. Sem qualquer tipo de chave primária para uma entidade, a única maneira de identificar uma empresa é comparando as designações destas, mas quando existem duas designações diferentes para a mesma empresa, como por exemplo, mais um carácter ou símbolo na designação, é criado outro registo no *data warehouse* final.

Apesar da importância deste problema e devido à natureza dos objectivos, isto é, como apenas são relevantes os campos de sector de actividade e região, e como estes dois campos são os mesmos para uma empresa com duas designações diferentes, não alterando os resultados finais, decidiu-se manter os registos duplicados do conjunto de dados inicial.

Depois de seleccionados os dados que se pretendia analisar, seguiu-se para o estudo das diferentes técnicas de mineração de dados com objectivo de se escolher as mais adequadas. Devido às características dos dados decidiu-se pela aplicação dos algoritmos *decision tree* C4.5 para obter uma previsão usando a classificação e *k-means* para obter uma descrição dos dados usando a segmentação. Estes dois algoritmos mostraram-se os mais adequados tendo em conta, como já foi dito, as características dos dados e também os objectivos pretendidos por este estudo.

Os resultados obtidos pela classificação (usando C4.5) foram bastante precisos e com estes é possível obter uma caracterização dos sectores, isto é, numa determinada região, quais os certificados que as empresas dessa região adquiriram para um determinado sector de actividade.

Estes resultados variam de região para região, e, por isso, este factor é bastante importante: a competitividade entre as empresas aumenta com o número de empresas na região.

Como é possível verificar nos resultados, as regiões Norte, Lisboa e Vale do Tejo e Centro apresentam um maior número de sistemas de certificação e uma maior variedade de tipos de certificados, ao contrário de, por exemplo, a região autónoma da Madeira em que as empresas desta região apenas apresentam um número bastante mais baixo de sistemas de certificação e uma menor diversidade de tipos de certificados (apenas ISO 9001:2000, ISO 14001:2004, OHSAS 18001:1999, EMAS II e HACCP). As regiões onde se verifica uma maior competitividade são também as regiões que apresentam o menor número de empresas que possuem apenas a norma ISO 9001:2000, isto é, é essencial para uma empresa adquirir o conjunto de tipo de certificados correspondente ao sector de actividade em que está inserida com o principal objectivo de obter uma vantagem competitiva, conseguindo sobreviver no mercado

Os resultados obtidos pela segmentação permitem obter uma descrição do estado do mercado e verificar assim a grande importância dos certificados ISO 9001:2000 e ISO 14001:2004, que são essenciais para uma empresa que pretenda sobreviver neste mercado competitivo. Com a aplicação deste método é também possível verificar quais os sectores de actividade que dominam o mercado nacional actual, o sector de actividade Comércio e Construção.

6.2 Linhas de Orientação para Trabalho Futuro

Este estudo pode ser considerado útil para qualquer empresa de um qualquer ramo, pois fornece uma primeira noção e conhecimento do estado do mercado, permitindo uma ajuda nos sistemas de certificação que devem adquirir, tendo em conta a região onde a empresa está localizada e também o sector de actividade em que se encontra inserida. Os certificados a adquirir variam segundo estes dois factores e poderão variar também com o tempo, tal como já referimos anteriormente.

Por isso, seria interessante prosseguir o estudo, adicionando novos dados para além dos relativos aos dos três anos analisados (2008, 2009 e 2010), obtendo, desse modo, uma nova previsão e uma nova descrição do estado do mercado competitivo nacional. A evolução do sistema de

certificados varia com as preferências e necessidades dos clientes, com a economia e um conjunto de factores que influenciam todo o mercado competitivo. Uma análise alargada iria com certeza fornecer novas perspectivas da evolução do mercado.

Para complementar a informação das empresas, seria útil utilizar outros dados sobre as empresas nacionais, como o volume de negócios gerado por cada empresa anualmente. Esta nova variável iria permitir relacionar o volume de negócios gerado com os sistemas de gestão adquiridos por uma determinada empresa, com um determinado sector de actividade ou uma determinada região. Os resultados obtidos neste caso seriam mais úteis para novas empresas, ajudando a definir quais os sectores de actividade que geram o maior volume de negócios ou quais as regiões que geram o maior volume de negócios. Desta forma, adiciona-se uma nova perspectiva ao estudo refinando bastante os resultados obtidos.

Bibliografia

Adriaans, P. e Zantinge, D., 1996. *Data Mining*. Addison-Wesley Longman Publishing Co.

Araújo, N., 2002. Proposta de sistema de gestão da segurança e saúde no trabalho, baseado na OHSAS 18001:1999, para empresas construtoras de edificações verticais. Tese de doutorado, Universidade Federal da Paraíba.

Berkhin, P., 2002. *Survey of clustering data mining techniques*, Technical report, Accrue Software, San Jose, CA.

Bobsin, M. A., 2005. Gestão de segurança, meio ambiente e saúde: Proposta de estrutura de sistema e metodologia de avaliação de desempenho. Dissertação de Mestrado Profissional em Sistemas de Gestão. Departamento de Engenharia de Produção. Universidade Federal Fluminense. Niterói.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. e Zanasi, A. 1998. *Discovering Data Mining - From Concept to Implementation*. Englewood Cliffs, NJ: Prentice-Hall.

Casadesús, M., Giménez, G. e Heras, I., 2001. Benefits of ISO 9000 implementation in Spanish industry. *European Business Review*, 13(6), pp. 327–335.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. e Wirth, R., 2000. *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Chen, M.-S., Han, J. e Yu, P. S., 1996. Data mining: An overview from a database perspective. *IEEE Trans. Knowledge Data Eng.*, 8 (6), pp. 866-883.

Corbett, C. e Kirsch, D., 2000. ISO 14000: An agnostic's report from the front line. *ISO 9000 + ISO 14000 News*, 9, pp. 4-17.

Costella, M. F., 2008. *Método de avaliação de sistemas de gestão de segurança e saúde no trabalho (MASST) com enfoque na engenharia de resiliência*. Tese (Doutorado em Engenharia de Produção). Programa de Pós-Graduação. Universidade Federal do Rio Grande do Sul. Porto Alegre.

Craven, M., 1996. *Extracting comprehensible models from trained neural networks*. Ph. D. Thesis, University of Wisconsin, Madison.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. e Wirth, R., 2000. *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS Inc.

Darnall, N., Gallagher, D., Andrews, R. e Amaral, D., 2000. Environment management systems: Opportunities for improved environmental and business strategy. *Quality Progress*, 9(3), pp. 1-9.

Darnall, N., 2001. Adopting ISO 14001:2004: Why Some Firms Mandate Certification while Others Encourage It. In: Third Annual Research Conference for the Association for Public Policy Analysis and Management Fall Conference, *Public Policy Analysis and Public Policy: Making the Connection*. November 1-3, Washington Monarch Hotel, Washington, DC.

Dias, M. M., 2002. Parâmetros na escolha de técnicas e ferramentas de mineração de dados. *Acta Scientiarum*, 24(6), pp. 1715-1725.

Douglas, A., Coleman, S. e Oddy, R., 2003. The case for ISO 9000. *The TQM Magazine*, 15(5), pp. 316-324.

Fayyad U., Piatetsky-Shapiro G. e Smyth P., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), pp. 37-54.

Furtado, A., 2002. *O Impacte de Sistemas de Qualidade Certificados no Desempenho das Empresas Portuguesas*. Dissertação de Mestrado em Gestão e Estratégia Industrial. Universidade Técnica de Lisboa, Instituto Superior de Economia e Gestão.

Garcia, M. D., 2000. *Uso integrado das técnicas de HACCP, CEP e FMEA*. Dissertação de Mestrado Profissionalizante em Engenharia. Universidade Federal do Rio Grande Do Sul. Escola de Engenharia, Porto Alegre, RS.

Golfarelli, M. e Rizzi, S., 2009. *Data Warehouse Design – Modern Principles and Methodologies*. McGraw-Hill.

Gotzamani, K. e Tsiotras, G., 2001. An empirical study of the ISO 9000 standards' contribution towards total quality management. *International Journal of Operations & Production Management*, 21(10), pp. 1326-1342.

Halkidi , M., Batistakis, Y. e Vazirgiannis, M., 2001. On Clustering Validation Techniques, *Journal of Intelligent Information Systems*, 17(2-3), pp- 107-145.

Han, J. e Kamber, M., 2001. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, San Francisco, CA.

Hand, D., Mannila, H. e Smyth, P., 2001. *Principles of Data Mining*. Cambridge: MIT Press.

Haro, D. G., 2001. *Sistemas da qualidade na indústria automobilística: uma proposta de auto-avaliação unificada*. Dissertação de Mestrado. Departamento de Engenharia de Produção. Universidade Federal do Rio Grande do Sul, Porto Alegre.

Heras, I., Dick G. e Casadesús, M., 2002. ISO 9000 registration's impact on sales and profitability: A longitudinal analysis of performance before and after certification. *International Journal Quality Reliability Management*, 19(6), pp. 774–791.

Huang, Z., 1997. A fast clustering algorithm to cluster very large categorical data sets in data mining. In *Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Dept. of Computer Science, The University of British Columbia, Canada, pp. 1–8.

Instituto Nacional de Estatística, I.P., 2008. "Anuário Estatístico de Portugal 2007", Lisboa, Portugal.

Instituto Nacional de Estatística, I.P., 2009a. Demografia de Empresas 2004-2007, O Empreendedorismo em Portugal - Indicadores sobre a Demografia das Empresas, *Destaque*.

Instituto Nacional de Estatística, I.P., 2009b. *Anuário Estatístico de Portugal 2008*, Lisboa, Portugal.

Instituto Nacional de Estatística, I.P., 2010a. Estatísticas das Empresas 2008 – Empresas em Portugal 2008, *Destaque*.

Instituto Nacional de Estatística, I.P., 2010b. Estudos sobre Estatísticas Estruturais das Empresas 2008 - Micro, Pequenas e Médias Empresas em Portugal, *Destaque*.

Instituto Nacional de Estatística, I.P., 2010c. *Anuário Estatístico de Portugal 2009*, Lisboa, Portugal.

Instituto Nacional de Estatística, I.P., 2011a. Estudos sobre Estatísticas Estruturais das Empresas 2007-2009 - O perfil exportador das PME em Portugal – 2007/2009, *Destaque*.

Instituto Nacional de Estatística, I.P., 2011b. Estudos sobre Estatísticas Estruturais das Empresas 2008 - Micro, Pequenas e Médias Empresas em Portugal, *Destaque*.

Instituto Nacional de Estatística, I.P., 2011c. *Empresas em Portugal 2009*, Lisboa, Portugal.

Jain, A. K., Murty, M. N. e Flynn, P. J., 1999. Data clustering: A review, *ACM Computing Surveys*, 31, pp. 264–323.

Kirkos, E., Spathis, C. e Manolopoulos, Y., 2007. Data mining techniques for the detection of fraudulent financial statements. *Expert with Systems with Applications*, 32, pp. 995–1003.

Kurgan, L. A. e Musilek, P., 2006. A survey of Knowledge Discovery and Data Mining process models, *The Knowledge Engineering Review*, 21, pp. 1-24.

Langaas, M., 1995. *Discrimination and classification*. Technical Report 1/95, Department of Mathematical Sciences, The Norwegian Institute of Technology, The University of Trondheim.

Larose D. T., 2005 *Discovering knowledge in data: an introduction to data mining*. New York: John Wiley & Sons.

Larose, D. T., 2006. *Data Mining Methods and Models*; Wiley-Interscience: Hoboken, NJ.

Magalhães, A, 2006. ISO 22000:2005 face a outros referenciais. *Segurança e Qualidade Alimentar*, 1, pp. 36-37.

Maier, S. e Vanstone, K., 2005. Do good environmental management systems lead to good environmental performance?, *Research Briefing*, Ethical Investment, Research Services (EIRIS), London.

MacQueen J., 1967. Some methods for classification and analysis of multivariate observations, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1, pp. 281–297, University of California Press, Berkeley, CA.

Mohammed, M., 2000. The ISO 14001:2004 EMS implementation process and its implications: A case study of Central Japan. *Eco-Management and Auditing*, 25(2), pp. 177-188.

Nahuz, M., 1995. O sistema ISO 14000 e a certificação ambiental. *Revista de Administração de Empresas*, 35(6), pp. 55–66.

Neves, L., 2007. Sistemas de gestão integrados. Segurança e Qualidade Alimentar. 2, pp. 30-1.

Ohashi, E. e Melhado, S., 2004. A importância dos indicadores de desempenho nas empresas construtoras e incorporadoras com certificação ISO 9001:2000. In *ENTAC (Encontro Nacional de Tecnologia do Ambiente Construído)*. São Paulo, Anais.

Quinlan, J. R., 1986. Induction of decision trees. *Machine Learning*, 1, pp. 81-106.

Quinlan, J. R., 1996. Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research* 4, pp. 77–90.

Quintela, H., 2005. *Sistemas de Conhecimento Baseados em Data Mining: Aplicação à Análise da Estabilidade de Estruturas Metálicas*, Dissertação de Mestrado, Departamento de Sistemas de Informação, Universidade do Minho, Portugal.

Ray S. e Turi R., 1999. Determination of number of clusters in k-means clustering and application in colour image segmentation. In *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*, pp. 137–143.

Ribeiro, A. M., 2006. *Certificação da qualidade e desempenho empresarial - Evidência empírica para Portugal*. Dissertação de Mestrado, Universidade do Minho, Braga, Portugal.

Ribeiro, A. M., 2007. Certificação da qualidade e desempenho empresarial: uma análise por quantis. *Revista de Estudos Politécnicos*, 5(8), pp. 201-214.

Rivera-Camino, J., 2001. What motivates european firms to adopt environmental management systems?. *Eco-management and auditing* 8, pp. 134–143.

Romano, C., 2006. *Gestão de Segurança e Saúde Ocupacional em Galvanoplastia – Aplicação do método Renault à OSHAS 18001*. Dissertação de mestrado. Porto Alegre: Universidade Federal do Rio Grande do Sul.

Rosa, A., 2009. *Proposta de aplicação de um sistema de gestão da responsabilidade social segundo a Norma NP 4469-1: 2008: O caso da ESTG*. Projecto aplicado no âmbito do Mestrado em Gestão, especialização em Administração Pública. Guarda.

Sampaio, P. e Saraiva, P., 2008. Barómetro da certificação 2008. Guia de Empresas Certificadas, Edição 2008, pp. 60–72.

Sampaio, P. e Saraiva, P., 2009. Barómetro da certificação 2009. Guia de Empresas Certificadas, Edição 2009, pp. 54–65.

Sampaio, P. e Saraiva, P., Barómetro da certificação 2010. CEMPALAVRAS
Disponível em: <http://www.qualidademadeira.com.pt/noticia/barometro-certificacao-10-cempalavras>

[Acedido a 14 de Dezembro de 2010]

Sampaio, P., Saraiva, P. e Rodrigues, A. 2008. ISO 9001:2000 certification research: questions, answers and approaches. *International Journal of Quality & Reliability Management*, 26(1), pp. 38–58.

Santos, N., 2007. *A implementação da NBR ISO 14001:2004 em organizações: um estudo multicaso*. M.S. Universidade Federal de Santa Maria.

Seddon, J., 2001. *The Case Against ISO 9000*. Dublin, Ireland: Oak Tree Press.

Silva, C. E. S., Tin, J. V. e Oliveira, V. C., 1997. Uma Análise da Aplicação da FMEA nas Normas de: Sistema de Gestão pela Qualidade (ISO 9000 e QS 9000), Sistema de Gestão Ambiental (ISO 14000), e Sistema de Gestão da Segurança e Saúde do Trabalho (BS 8800 . Futura ISO 18000). In *XVII ENEGEP*, Gramado, RS.

Sousa, M., Mattoso, M. e Ebecken, N., 1998. Data mining: A database perspective. In *Proc. Int. Conf. Data Mining*. pp. 413–431.

Strachan, P., Sinclair, I. e Lal, D., 2003. Managing ISO 14001:2004 implementation in the United Kingdom continental shelf. *Corporate Social Responsibility and Environmental Management*, 10(1), pp. 50–63.

Terlaak, A. e King, A.A., 2006. The Effect of Certification with the ISO 9000 Quality Management Standard: A Signaling Approach. *Journal of Economic Behavior and Organization*, 60(4), pp. 579-602.

Referências WWW

- [1] http://www.bureauveritas.pt/wps/wcm/connect/bv_pt/local/home/our-services/certification
"Bureau Veritas Portugal - Certificação" [Acedido a 15 de Dezembro de 2010]
Neste site é possível consultar informações adicionais sobre o processo de certificação e as normas disponíveis para as empresas adquirirem.
- [2] <http://www.cempalavras.pt/>
"cempalavras Comunicação Empresarial LDA" [Acedido a 10 de Janeiro de 2011]
Neste site é possível consultar as publicações e dados referentes às empresas nacionais, como a evolução dos sistemas de gestão ao longo dos anos.
- [3] <http://www.mysql.com/>
"MySQL The world's most popular open source database" [Acedido a 15 de Fevereiro de 2011]
Este site contém toda a informação relativa à *open-source database*, sendo também possível adquirir certos produtos e serviços.
- [4] <http://www.mysql.com/products/workbench/>
"MySQL The world's most popular open source database – MySQL Workbench 5.2"
[Acedido a 15 de Fevereiro de 2011]
Neste site é possível consultar informação sobre este programa de administração de bases de dados.

[5] <http://rapid-i.com/content/view/181/190/>

“RapidMiner – Report the Future” [Acedido a 10 de Fevereiro de 2011]

Neste site é possível consultar as funções do RapidMiner assim como fazer o download desta ferramenta *open-source*.

Anexo I

Tabela 31 - Anexo I - Correspondência entre o Sector de Actividade e o EAC

EAC	Sector
1	Agricultura e Pescas
2	Minas e Exploração Mineira
3	Alimentação, Bebidas e Tabaco
4	Têxteis e Produtos Têxteis
5	Couro e Produtos de Couro
6	Madeira e Produtos de Madeira
7	Pasta, Papel e Produtos de Papel
8	Editoras
9	Empresas Gráficas
10	Fabricação de Coque e Produtos Petrolíferos Refinados
12	Químicos, Produtos Químicos, Fibras Sintéticas e Artificiais
13	Indústria Farmacêutica
14	Fabricação de Artigos de Borracha e Matérias Plásticas
15	Produtos Minerais Não Metálicos
16	Betão, Cimento, Cal e Gesso
17	Fabricação Metalúrgica de Base e Produtos Metálicos
18	Equipamentos e Máquinas
19	Equipamento Eléctrico e de Óptica
20	Construção e Reparação Naval
21	Indústria Aeroespacial
22	Outro Equipamento de Transporte
23	Outras Fabricações Não Especificadas
24	Reciclagem
25	Produção e Distribuição de Energia Eléctrica
26	Produção e Distribuição de Gás

27	Fornecimento de Água
28	Construção
29	Comércio
30	Hotéis e Restaurantes
31	Transporte, Armazenamento e Comunicações
32	Mediação Financeira, Imobiliária e Aluguer
33	Tecnologias de Informação
34	Serviços de Engenharia
35	Outros Serviços
36	Administração Pública
37	Educação
38	Saúde e Serviços Sociais
39	Outros Serviços Sociais