

UNIVERSIDADE DE ÉVORA
ESCOLA DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE QUÍMICA

*Previsão da Classificação da Qualidade dos Efluentes da Indústria
Petrolífera com Recurso a Sistemas Inteligentes*

por

Ana Teresa Moreirinha Vila Fernandes



*Dissertação no Âmbito do Mestrado em Análises Químicas Ambientais sob a
orientação do Professor Doutor Henrique Vicente*

'Esta dissertação não inclui as críticas e sugestões feitas pelo júri

Évora

2009

*Previsão da Classificação da Qualidade dos Efluentes da Indústria
Petrolífera com Recurso a Sistemas Inteligentes*

Ana Teresa Moreirinha Vila Fernandes

anavilafernandes@gmail.com



171 351

*Dissertação submetida à Universidade de Évora para a obtenção do grau de Mestre em
Análises Químicas Ambientais e elaborada sob a orientação do Professor Doutor
Henrique Vicente*

Esta dissertação não inclui as críticas e sugestões feitas pelo júri.

Évora

2009

*Aos verdadeiros amigos;
Ao Rui;
Ao meu irmão;
Aos meus pais, que o souberam ser.*

Agradecimentos

Poderia enumerar aqui uma lista infindável de pessoas que muito me ajudaram a atingir este objectivo. Seria impossível nomeá-las todas, como gostaria, pois correria o risco de esquecer alguém. No entanto, não posso deixar de fazer referência a algumas delas, agradecendo primeiramente e em particular ao meu orientador, Professor Doutor Henrique Vicente, com o qual tive o privilégio de trabalhar e que me acompanhou de forma empenhada, ajudando-me, encorajando-me e inculcando-me o espírito de objectividade e clareza necessário a qualquer trabalho académico. A ele agradeço, não só o facto de ter influenciado o meu interesse pela área de Inteligência Artificial mas também todos os ensinamentos, aconselhamento e amizade partilhando comigo muitas e longas conversas: umas vezes de partilha de ensinamentos, outras de desabaços e motivação o que permitiu decisivamente que esta dissertação fosse uma realidade.

Ao Professor Doutor José Carlos Ferreira Maia Neves, do Departamento de Informática da Universidade do Minho, a quem devo um agradecimento muito especial pelos ensinamentos que me proporcionou.

Não posso esquecer os amigos e colegas da Refinaria de Sines - Galp Energia, nomeadamente o Engenheiro José Fernandes, Engenheiro Joaquim Letras e Engenheiro Cordeiro Catarino pela colaboração, pelo interesse e pela disponibilidade sempre manifestada. Também não poderia deixar de agradecer aos meus amigos Vítor Pio e Maria José Rocha, pelas oportunas manifestações de companheirismo, encorajamento e amizade. Expresso ainda à Doutora Conceição Fernandes e à Engenheira Adelaide Bernardino, do Porto de Sines, o meu agradecimento pela disponibilidade, apoio e estímulo.

As últimas palavras, nem por isso as menos importantes, são para a minha família. À minha mãe, por todo o apoio, dedicação, carinho e mimo. Ao meu pai por todos os ensinamentos, valores e honra transmitidos. Ao meu irmão, por toda a paciência e companheirismo. Ao Rui, por todo o carinho, amizade, compreensão e cumplicidade. À restante família, por todo o apoio e dedicação. A eles, agradeço especialmente a confiança que me inculcaram ao longo dos meus anos de vida, sei que é a vós que devo o facto de ser aquilo que sou hoje.

A todos, os meus sinceros agradecimentos.

Resumo

O presente trabalho realizou-se na Refinaria de Sines e teve como principal objectivo a utilização de ferramentas oriundas da Área Científica da Inteligência Artificial no desenvolvimento de modelos de previsão da classificação da Água Residual Industrial de acordo com a Legislação em vigor, com vista à minimização dos impactes ambientais e das tarifas aplicadas pela Concessionária (Águas de Santo André) à Refinaria.

Actualmente a avaliação da qualidade do efluente é realizada através de métodos analíticos após colheita de uma amostra do efluente final. Esta abordagem é muito restritiva já que não permite actuar sobre o efluente em questão pois apenas pode evitar que, no futuro, uma mistura semelhante volte a ser refinada. Devido a estas limitações, o desenvolvimento de modelos de previsão baseados em Data Mining mostrou ser uma alternativa para uma questão pró-activa da qualidade dos efluentes que pode contribuir decisivamente para o cumprimento das metas definidas pela Empresa.

No decurso do trabalho, foram desenvolvidos dois modelos de previsão da qualidade do efluente industrial com desempenhos muito semelhantes. Um deles utiliza a composição das misturas processadas e o outro, utiliza informações relativas ao crude predominante na mistura.

Predicting the Classification of the Quality of Wastewater Oil Industry using Intelligent Systems

Abstract

This study has taken place at the Sines Refinery and its main objective is the use of Artificial Intelligence tools for the development of predictive models to classify industrial residual waters according with the Portuguese Law, based on the characteristics of the mixtures of crude oil that arrive into the Refinery to be processed, to minimize the Environmental impacts and the application of taxes.

Currently, the evaluation of the quality of effluent is performed by analytical methods after harvesting a sample of the final effluent. This approach is very restrictive since it does not act on the intended effluent; it can only avoid that in the future a similar mixture is refined. Due to these limitations, the development of forecasting models based on Data Mining has proved to be an alternative on the important issue which is the quality of effluent, which may contribute to the achievement of targets set by the Company.

During this study, two models were developed to predict the quality of industrial effluents with very similar performances. One uses the composition of processed mixtures and the other uses information regarding the predominant oil in the mixture.

Conteúdo

<i>Agradecimentos</i>	iii
<i>Resumo</i>	v
<i>Abstract</i>	vii
<i>Conteúdo</i>	ix
<i>Lista de Figuras</i>	xiii
<i>Lista de Tabelas</i>	xv
<i>Notação e Terminologia</i>	xvii
<i>Acrónimos</i>	xvii
<i>Capítulo I – Introdução</i>	1
<i>I.1. Caracterização da Entidade Acolhedora</i>	4
<i>I.1.1. Breve Historial da Empresa</i>	4
<i>I.1.2. Caracterização do Local de Trabalho</i>	5
<i>I.1.3. Política de Qualidade</i>	8
<i>I.1.4. Política de Segurança, Saúde e Ambiente</i>	9
<i>I.2. A Gestão do Conhecimento</i>	11
<i>I.3. Inovação e Valor Acrescentado do Trabalho</i>	12
<i>I.4. Organização da Dissertação</i>	13
<i>I.5. Conclusões</i>	14
<i>Capítulo II – Química do Petróleo</i>	15
<i>II.1. Origem do Petróleo</i>	16
<i>II.2. Composição Química do Petróleo</i>	18
<i>II.2.1. Hidrocarbonetos Parafínicos</i>	20
<i>II.2.2. Hidrocarbonetos Nafténicos</i>	20

II.2.3. Hidrocarbonetos Aromáticos	21
II.2.4. Hidrocarbonetos Asfálticos	21
II.3. Impurezas do Petróleo.....	25
II.3.1. Compostos de Enxofre	25
II.3.2. Compostos Oxigenados	26
II.3.3. Compostos Azotados	26
II.3.4. Compostos Metálicos.....	26
II.4. Tipos de Petróleo Bruto.....	27
II.5. Densidade – Grau API	29
II.6. A Refinação	29
II.7. Conclusões.....	34
Capítulo III – Aspectos Ambientais na Indústria Petrolífera.....	35
III.1. Sistemas de Protecção Ambiental na Refinaria de Sines.....	37
III.2. Descrição dos Efluentes da Refinaria de Sines	38
III.3. Tratamento de Efluentes Industriais.....	39
III.3.1. Tratamento Preliminar	40
III.3.2. Tratamento Primário	41
III.3.3. Tratamento Secundário.....	42
III.3.4. Tratamento Terciário	44
III.4. Conclusões	45
Capítulo IV – Descoberta de Conhecimento em Bases de Dados e Data Mining	47
IV.1. Descoberta de Conhecimento em Bases de Dados	49
IV.2. Data Mining.....	51
IV.3. Modelos e Técnicas de Data Mining	53
IV.3.1. Árvores de Decisão.....	54
IV.3.2. Indução de Regras.....	57
IV.3.3. Avaliação dos Modelos	58
IV.4. Conclusões.....	59

<i>Capítulo V – Procedimento Experimental.....</i>	<i>61</i>
<i>V.1. Amostragem e Preservação das Amostras.....</i>	<i>61</i>
<i>V.1.1. Pontos de Amostragem.....</i>	<i>62</i>
<i>V.1.2. Frequência de Amostragem.....</i>	<i>64</i>
<i>V.2. Procedimento Analítico.....</i>	<i>64</i>
<i>V.3. Base de Dados.....</i>	<i>65</i>
<i>V.4. Conclusões.....</i>	<i>67</i>
<i>Capítulo VI – Apresentação, Interpretação e Análise de Resultados e Modelos.....</i>	<i>69</i>
<i>VI.1. Seleção, Pré-Processamento e Transformação de Dados.....</i>	<i>70</i>
<i>VI.2. Modelação.....</i>	<i>84</i>
<i>VI.3. Apresentação, Interpretação e Avaliação dos Modelos.....</i>	<i>85</i>
<i>VI.4. Conclusões.....</i>	<i>90</i>
<i>Capítulo VII – Conclusões Finais e Perspectivas Futuras.....</i>	<i>91</i>
<i>Referências Bibliográficas.....</i>	<i>95</i>
<i>Anexo I.....</i>	<i>101</i>
<i>Anexo II.....</i>	<i>105</i>

Lista de Figuras

<i>Figura I. 1– Localização Geográfica da Refinaria de Sines</i>	6
<i>Figura I. 2– Organograma da Refinaria de Sines</i>	7
<i>Figura II. 1– Refinaria de Sines</i>	30
<i>Figura II. 2– Esquema do Processo de Destilação Fraccionada</i>	30
<i>Figura II. 3– Esquema do Processo de Cracking Catalítico</i>	31
<i>Figura II. 4– Esquema do Processo de Cracking Térmico</i>	32
<i>Figura II. 5– Esquema do Processo de Reformaço</i>	32
<i>Figura II. 6– Esquema do Processo de Hidrogenação</i>	33
<i>Figura II. 7– Refinação, Processos de Transformação e Derivados do Petróleo</i>	34
<i>Figura III. 1– Qualidade dos Efluentes Industriais vs Poluição e Tarifaço</i>	39
<i>Figura III. 2– Separadores Gravíticos</i>	41
<i>Figura III. 3– Flotaço</i>	41
<i>Figura III. 4– Bacia de Águas Contaminadas</i>	42
<i>Figura IV. 1– Processo de Descoberta de Conhecimento em Base de Dados</i>	50
<i>Figura IV. 2– Algoritmo de Indução de Árvores de Decisào</i>	56
<i>Figura IV. 3– Validaço Cruzada com k Iteraço</i>	58
<i>Figura V. 1 – Diagrama Simplificado da Refinaria de Sines</i>	63
<i>Figura V. 2– Extracto da Base de Dados Utilizada</i>	65
<i>Figura VI. 1– Etapas da DCBD Agrupadas em Tarefas</i>	69
<i>Figura VI. 2– Distribuço do Número de Casos por Classe de Qualidade</i>	72
<i>Figura VI. 3– Distribuço do Número de Casos por Agrupamentos de Qualidade</i>	73
<i>Figura VI. 4– Percentagens de Crudes Presentes nas Misturas</i>	74

<i>Figura VI. 5– pH vs Crude Predominante na Mistura</i>	<i>77</i>
<i>Figura VI. 6– CQO vs Crude Predominante na Mistura</i>	<i>78</i>
<i>Figura VI. 7– Compostos Fenólicos vs Crude Predominante na Mistura</i>	<i>79</i>
<i>Figura VI. 8– SST vs Crude Predominante na Mistura.....</i>	<i>80</i>
<i>Figura VI. 9– Sulfuretos vs Crude Predominante na Mistura.....</i>	<i>81</i>
<i>Figura VI. 10– Óleos e Gorduras vs Crude Predominante na Mistura.....</i>	<i>81</i>
<i>Figura VI. 11– Classes de Qualidade dos Efluentes vs Crude Predominante.....</i>	<i>82</i>
<i>Figura VI. 12– Agrupamentos de Qualidade dos Efluentes vs Crude Predominante. 83</i>	
<i>Figura VI. 13– Interações Crudes Predominantes - Agrupamentos de Qualidade</i>	<i>84</i>
<i>Figura VI. 14– Extracto da Árvore de Decisão - Modelo I</i>	<i>86</i>
<i>Figura VI. 15– Extracto da Árvore de Decisão - Modelo II.....</i>	<i>87</i>
<i>Figura VI. 16– Extracto do Conjunto de Regras Obtido para cada Modelo</i>	<i>88</i>

Lista de Tabelas

<i>Tabela II. 1– Composição Elemental do Petróleo Bruto e do Gás Natural.</i>	<i>19</i>
<i>Tabela II. 2– Constituição de um Crude (Fracções de Hidrocarbonetos)</i>	<i>22</i>
<i>Tabela II. 3– Composição de um Petróleo em Condições de Subsuperfície</i>	<i>23</i>
<i>Tabela II. 4– Composição de um Petróleo Bruto</i>	<i>24</i>
<i>Tabela II. 5– Tipos de Petróleo Bruto.....</i>	<i>27</i>
<i>Tabela II. 6– Produtos Finais Típicos da Refinação do Petróleo Bruto.....</i>	<i>34</i>
<i>Tabela III. 1– Impactes dos Resíduos Provenientes das Refinarias</i>	<i>36</i>
<i>Tabela III. 2– Diferentes Tipos de Efluentes</i>	<i>38</i>
<i>Tabela IV. 1– Etapas do Processo de DCBD.....</i>	<i>50</i>
<i>Tabela IV. 2– Objectivos de Data Mining.....</i>	<i>52</i>
<i>Tabela IV. 3– Tarefas e Técnicas Usadas no Processo de Data Mining.....</i>	<i>54</i>
<i>Tabela V. 1– Recolha e Preservação das Amostras.....</i>	<i>62</i>
<i>Tabela V. 2– Determinação dos Parâmetros Analíticos.</i>	<i>64</i>
<i>Tabela V. 3– Campos Presentes na BD Utilizada</i>	<i>65</i>
<i>Tabela V. 4– Caracterização das Variáveis Utilizadas em Termos Estatísticos.....</i>	<i>66</i>
<i>Tabela VI. 1– Critérios para a Classificação de Água Residual Industrial</i>	<i>70</i>
<i>Tabela VI. 2– Matrizes de Coincidências para os Modelos I e II</i>	<i>89</i>
<i>Tabela VI. 3– Percentagem de Acertos por Classe para cada Modelos</i>	<i>89</i>

Notação e Terminologia

A notação em todo o documento que é aqui apresentado segue a convenção:

Texto em itálico - é utilizado não só para palavras em língua estrangeira (e.g., em Inglês) mas também para dar ênfase a um determinado termo ou expressão; e

Texto com letra maiúscula - é utilizado para realçar um conceito ou palavra no meio de um parágrafo.

Acrónimos

AA - Aprendizagem Automática

AG - Algoritmos Genéticos

API - American Petroleum Institute

BAC - Bacia de Águas Contaminadas

BD - Base de Dados

BPT - Bacia de Águas Pré-Tratadas

BTU - British Thermal Units

DCBD - Descoberta de Conhecimentos em Base de Dados

DM - Data Mining

EIA - Environmental Impact Assessment

ETAR - Estação de Tratamento de Águas Residuais

FCC - Fluid Catalytic Cracking

GALP - Petróleos e Gás de Portugal

GDP - Gás de Portugal

HSE - Health, Safety and Environment

IA - Inteligência Artificial

IRAR - Instituto Regulador de Águas Residuais

OPEP - Organização dos Países Produtores de Petróleo

OP-V - Separador Gravítico

Petrogal - Petróleos de Portugal

PPL - Produção Primária Líquida

RARISA - Regulamento de Recolha e Tratamento de Água Residual Industrial do Sistema de Santo André

SI - Sistemas Inteligentes

SMEWW - Standard Methods for the Examination of Water and Wastewater

SSA - Segurança, Saúde e Ambiente

TAN - Total Acid Number

VOC - Volatile Organic Compounds

Capítulo I – Introdução

Notas introdutórias; problemática ambiental; indústria do petróleo; descrição das contribuições e valor acrescentado do trabalho; metodologia de investigação; organização da tese e caracterização da entidade acolhedora.

O Ambiente é um fenómeno de dimensão universal que atravessa, hoje em dia, sérios problemas causados pelo elevado consumo de recursos naturais, assim como pela má utilização de alguns dos seus componentes como a água, o solo e o ar.

Devido ao carácter estratégico de certos recursos naturais, como por exemplo a água, o petróleo ou mesmo a pesca, os impactes directos ou indirectos sobre o Ambiente sendo, muitas vezes locais têm, no entanto, repercussões de cariz global.

A importância da água como elemento natural indispensável à vida foi reconhecida desde os primórdios da existência do Homem sobre a Terra. Ao longo dos tempos, os rios marcaram de forma determinante a radicação das populações e a evolução das grandes civilizações, cuja cultura e engenho aparecem muito ligados ao usufruto da água [Serralheiro, 2001] [Lehr & Keeley, 2005b].

Só em meados do século XX a Humanidade tomou consciência do valor do recurso renovável água como factor básico do desenvolvimento económico e social, em consequência da pressão crescente exercida pelo crescimento demográfico, pelo fomento da produção agrícola e industrial e pela progressiva concentração urbana [Lehr et al., 2005] [Lehr & Keeley, 2005a].

A prevalência da visão infraestrutural na oferta da água veio, assim, a dar lugar à noção de recurso hídrico, com utilizações alternativas susceptíveis de análise comparativa e de avaliação económico-financeira. Neste contexto, pode afirmar-se que as empresas enquanto agentes de extracção, transformação e venda de produtos, contribuem para alguns dos problemas ambientais existentes, na medida em que para produzirem necessitam de recursos naturais, cujo consumo supera a sua natural reposição. Actualmente as empresas enfrentam uma forte competição, sendo indispensável dar resposta à evolução e diversificação das necessidades e expectativas dos consumidores, assim como à inovação

tecnológica. Em consequência, os bens e os serviços têm um ciclo de vida mais curto, necessitando também de melhoria contínua [Cunha et al., 1980] [Braga & Morgado, 2007].

O desenvolvimento sustentável baseia-se no equilíbrio entre os seus três pilares, i.e. o desenvolvimento económico, o progresso social e a preservação do ambiente. Este equilíbrio constitui hoje o maior desafio global para o Mundo em que vivemos, apresentando cada país prioridades e questões que lhe são próprias [Cunha et al., 1980] [Braga & Morgado, 2007].

No que respeita à predisposição do tecido empresarial português em relação às questões ambientais, a situação é bem melhor que há uma década, conforme o prova o número crescente de empresas e organizações que obtêm a certificação ambiental. A aplicação de técnicas computacionais ao desenvolvimento de soluções económicas e ambientais em Indústrias situa-se no âmbito das preocupações e necessidades que se têm vindo a afirmar como uma das mais relevantes para o progresso das mesmas. Neste contexto, o desempenho das empresas e dos seus membros passa por produções ambientalmente responsáveis, que satisfaçam o consumidor final e não sejam nocivas nem para o ambiente, nem para a saúde das populações contribuindo, deste modo, para a melhoria da qualidade e do nível de vida dos cidadãos e da própria Terra [Braga & Morgado, 2007].

O papel que as empresas têm na problemática ambiental é forçosamente grande e até mesmo crucial, pois são elas que produzem uma elevada quantidade de resíduos devendo, por isso ter a responsabilidade de efectuar o seu acompanhamento e o seu reaproveitamento ou encaminhamento ao destino final [Braga & Morgado, 2007].

A Indústria, como actividade integrada no ciclo da satisfação das necessidades da sociedade, é inevitavelmente consumidora de recursos naturais e é um elemento transformador do meio em que se insere. Esta interacção tem aspectos claramente negativos que urge minimizar, na dupla consciência de que esta actividade é socialmente necessária e de que não é, possível a eliminação absoluta dos seus inconvenientes [Cunha et al., 1980] [Braga & Morgado, 2007].

Em Portugal, e no que respeita à relação Economia/Ambiente, à medida que aquela se moderniza, vão sendo respeitados de modo crescente os factores ambientais mas, por outro lado, à medida que as exigências ambientais crescem, são também introduzidas limitações à implantação e à expansão das actividades económicas, nomeadamente as de carácter industrial [Braga & Morgado, 2007].

Devido à crescente preocupação ambiental, a Indústria do Petróleo começou a ter um papel mais preponderante na minimização do impacto ambiental [Reis, 1996].

Num Mundo crescentemente globalizado em que as economias são cada vez mais interdependentes, o Petróleo continua a ser a fonte de energia privilegiada em que assentam as redes de transportes que asseguram, a um ritmo cada vez mais intenso e mais acelerado, as trocas comerciais e a mobilidade das pessoas e bens entre países, regiões e continentes [Braga & Morgado, 2007] [Gomes & Alves, 2007].

Os produtos derivados de petróleo são os que mais contribuem para os actuais padrões de vida. São usados nos mais variados sectores desde os transportes, passando, entre outros, pela medicina, pela electricidade e pela química [Gomes & Alves, 2007].

As actividades de descoberta e extracção de petróleo podem, contudo, causar impacto ambiental ao libertarem resíduos em concentrações consideráveis. Destes resíduos fazem parte hidrocarbonetos, sólidos contaminados com hidrocarbonetos, água contaminada com uma enorme variedade de sólidos suspensos e uma enorme variedade de produtos químicos [Gomes & Alves, 2007]. Estes resíduos têm efeitos adversos no ambiente, os quais devem ser minimizados ou eliminados através da implementação de uma adequada gestão de resíduos [Reis, 1996]. As empresas ligadas ao universo petrolífero devem, até por questões de competitividade, ser racionais não só no consumo de recursos naturais e de energia como também, nos bens que colocam no mercado, que devem ser amigos do ambiente e sem componentes nocivos à saúde dos consumidores. Como tal, é necessária a adopção de atitudes que minimizem os impactos ambientais não apenas ao nível das operações como, também, ao nível da optimização dos compostos que serão lançados no ambiente, os quais devem não somente respeitar a legislação em vigor como, acima de tudo, preservar o equilíbrio do Ecosistema Global. Deste modo, é necessário prevenir e actuar de uma forma pró-activa, pelo que o recurso a métodos de previsão, que têm subjacentes processos de tomada de decisão, oriundos da Área Científica da Inteligência Artificial pode contribuir para solucionar ou amenizar os problemas [Abelha, 2004] [Vicente, 2004] [Santos et al., 2005].

I.1. Caracterização da Entidade Acolhedora

A Galp Energia, SGPS, S.A. foi constituída em 22 de Abril de 1999, sob a denominação GALP - Petróleos e Gás de Portugal, em resultado da reestruturação do sector energético em Portugal, para operar no Sector Petrolífero e do Gás Natural [Galp-Sines, 1999].

Actualmente, a Galp Energia é a principal empresa integrada de produtos petrolíferos e gás natural do país, com uma presença crescente em Espanha e uma actividade em desenvolvimento no sector da produção e de fornecimento de energia eléctrica.

A Galp Energia agrupou a antiga Petrogal - Petróleos de Portugal, actualmente designada por Refinaria de Sines, a única Empresa refinadora e principal distribuidora de produtos petrolíferos em Portugal e a Gás de Portugal (GDP), sociedade responsável pela importação, transporte e distribuição de gás natural em Portugal [Galp-Sines, 1999].

I.1.1. Breve Historial da Empresa

De acordo com o Portal do Colaborador Galp [Galp, 2007a], em 1999 foi criada a Galp Energia como um veículo de reestruturação do sector energético em Portugal, nas áreas do Petróleo e de Gás Natural.

Em 2000 foram obtidos os primeiros resultados na produção do Petróleo em Angola. Na mesma altura, a empresa participou no Brasil no 2º *License Round* com a obtenção, juntamente com a Petrobrás - Petróleo Brasileiro, S.A. da concessão de duas participações de 10% em blocos localizados no *deep offshore* da Bacia de Santos. Em 2001 foi concluída a substituição do gás de cidade por gás natural em Lisboa e a construção de uma caverna de armazenamento de gás propano em Sines.

Em 2002 foi adquirido 5% da CLH - *Compañia Logística de Hidrocarburos, S.A.* o que facilitou o desenvolvimento de uma estratégia de logística no mercado espanhol.

Em 2003 foram concluídos os investimentos no Terminal de Gás Natural Liquefeito em Sines o que permitiu ligar esta infra-estrutura à rede nacional, através do gasoduto Sines - Setúbal. No mesmo ano, concretizou-se a operação de *Swaps*, processo que envolveu a troca com a *Compañia Española de Petroleos, S.A.*, a *Cepsa Portuguesa S.A.*, a *Total España S.A.* e a *Total Portuguesa S.A.* de 79 postos em Espanha por cedência de 78 postos em Portugal.

Em 2004, foi reforçada a presença no mercado espanhol com a aquisição da *BP Enérgica* que permitiu a integração de uma base de clientes atractiva e de uma plataforma logística que possibilitou o alargamento da cadeia de valor em Espanha, chegando mais próximo do cliente final.

Em 2005, a Galp Energia através da Petrogal adquiriu à Total a totalidade do capital social da Empresa Petróleos de Valência, proprietária do Parque de Armazenamento Terminal de Valência.

Em 2006, concretizou-se o processo de separação das actividades reguladas no sector do gás natural.

Em 2007 e 2008 foram construídas novas unidades fabris.

1.1.2. Caracterização do Local de Trabalho

A entidade onde foi desenvolvido o presente trabalho, denomina-se Refinaria de Sines e pode ser definida como uma instalação industrial de processamento de petróleo bruto, produzindo uma gama diversificada de produtos refinados, nomeadamente:

- Propano, Butano e Propileno;
- Gasolinas com Chumbo e Sem Chumbo;
- Nafta Química (Matéria-Prima para a Indústria Petroquímica de Base);
- Jet´s;
- Petróleos;
- Gasóleo;
- Fuel-Óleo;
- Betume; e
- Enxofre.

A Refinaria de Sines está situada a 2,5 km para Leste da cidade costeira de Sines, na região da Serra de Grândola tal como se pode observar Figura I.1, sendo uma das maiores unidades da Europa com capacidade de destilação de 10,8 milhões ton/ano (223 000 barris/dia). Constitui-se como uma unidade industrial estratégica, muito importante na actividade económica do país. Ocupa uma área de 320 hectares, com uma capacidade de

armazenagem de 3 milhões de m³, dos quais 1,5 milhões são de petróleo bruto e o restante de produtos intermédios e finais.

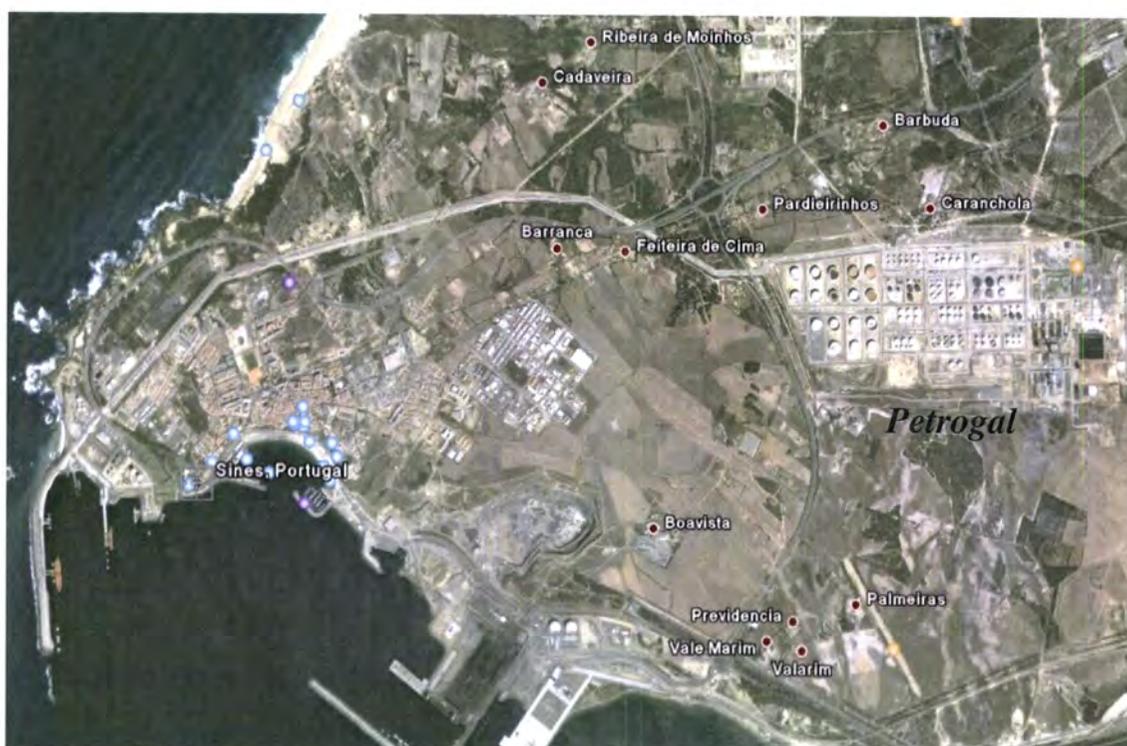


Figura I. 1 - Localização Geográfica da Refinaria de Sines. Fonte: <http://www.googlearth.pt/>

A Empresa é constituída por seis unidades principais, i.e. Operações; Programa e Controlo de Produção; Técnica; Manutenção e Engenharia; Ambiente, Qualidade e Segurança e Informação de Gestão, as quais são constituídas por várias subunidades. Na Figura I.2 apresenta-se o Organograma da Refinaria de Sines onde estão representadas as suas diferentes Unidades Funcionais.

O trabalho experimental que suporta a presente tese foi realizado no sector da Refinaria de Sines designado por Movimentação de Produtos pertencente à Área de Operações. O sector de Movimentação de Produtos é responsável pela operação do Sistema de Tratamento de Efluentes, área que está intimamente ligada à Área Técnica e à Área de Ambiente. Na Refinaria de Sines protege-se o domínio hídrico através da operação de um sistema de drenos e de separadores gravíticos que recolhem e tratam águas oleosas provenientes dos seus processos. O efluente industrial, depois de pré-tratado, é enviado para a Estação de Tratamento de Águas Residuais (ETAR) da Ribeira de Moinhos.

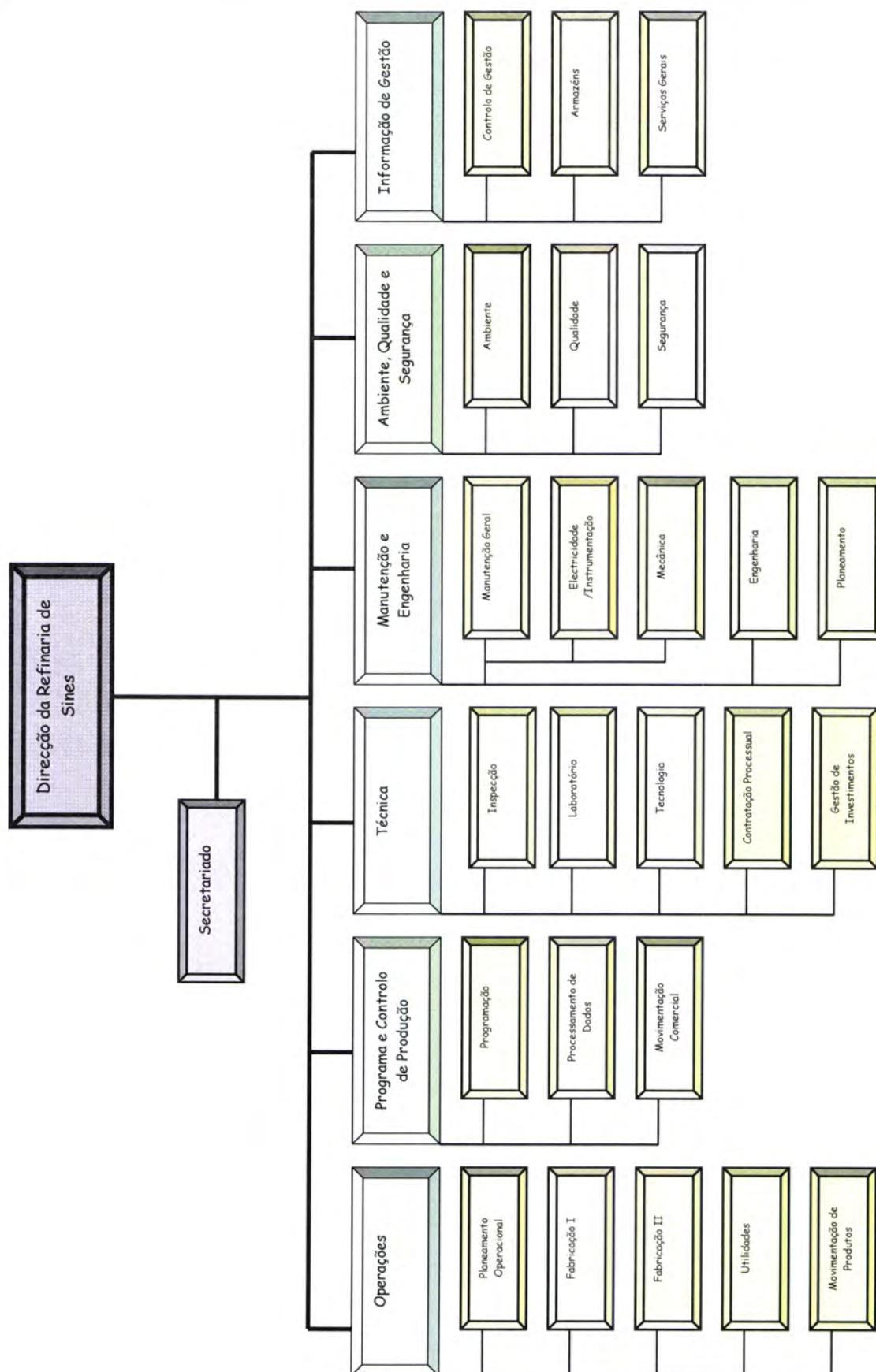


Figura I. 2 - Organograma da Refinaria de Sines. Fonte: Recursos Humanos da Refinaria de Sines, 2007

1.1.3. Política de Qualidade

A Galp Energia enquanto empresa de referência no sector da energia está consciente da sua responsabilidade empresarial e da necessidade de contribuir, a par de outros agentes, para o permanente desenvolvimento da Sociedade. Entende, por isso, que é seu dever assumir o compromisso da melhoria contínua da qualidade dos seus processos, produtos e serviços como uma prioridade da gestão e pilar fundamental na consolidação do prestígio da marca Galp. A Galp Energia compromete-se a:

- Conceber, implementar e desenvolver modelos de relacionamento com os seus clientes que lhe permitam conhecer e compreender as suas necessidades e expectativas actuais e futuras, de forma a garantir a sua plena satisfação e a criação de valor sustentável;
- Fomentar uma liderança eficaz, a todos os níveis, criando as condições necessárias para que os responsáveis conduzam as suas equipas numa base de confiança e entusiasmo, transmitindo-lhes a indispensável visão de futuro;
- Estabelecer uma gestão por processos integrados em sistemas que sejam capazes de assegurar o desenvolvimento das estratégias da empresa, de permitir a tomada de decisão baseada em informação factual e de antecipar as necessidades de todas as partes interessadas;
- Satisfazer as necessidades do mercado e das partes interessadas, fornecendo produtos e serviços que cumpram ou excedam os requisitos legais e contratuais;
- Aplicar as boas práticas de gestão em todas as suas actividades, assegurando a utilização eficiente dos recursos e investindo em tecnologias inovadoras e nas melhores técnicas operacionais disponíveis, tendo em vista a contínua adaptação à evolução dos mercados em que opera;
- Promover a sensibilização e formação dos colaboradores para as suas obrigações individuais e colectivas quanto ao cumprimento dos requisitos e procedimentos estabelecidos, bem como à permanente melhoria da qualidade, assegurando o desenvolvimento das suas competências, o seu envolvimento na obtenção dos resultados planeados e o respeito pelos valores de cidadania;
- Incluir objectivos quantificáveis de qualidade nos planos de negócio, procedendo à medição e avaliação dos resultados obtidos, à identificação de áreas de melhoria e ao estabelecimento de programas de actuação;

- Desenvolver e manter um relacionamento profícuo com os seus fornecedores e parceiros apostando no reforço da confiança mútua, em moldes que permitam desenvolver competências e alcançar objectivos partilhados; e
- Aplicar os requisitos de gestão da qualidade estabelecidos em normas internacionalmente reconhecidas, avançando de forma progressiva rumo à excelência.

Em suma, a Galp Energia compromete-se a actuar em todas as circunstâncias para que seja possível alcançar e manter os mais elevados padrões de qualidade em todas as suas actividades de modo a gerar valor para os accionistas, satisfazer os clientes e contribuir para o bem-estar da Sociedade, reconhecendo que a competência, criatividade, talento e motivação dos seus colaboradores é o seu recurso mais valioso [Galp, 2007b].

1.1.4. Política de Segurança, Saúde e Ambiente

A segurança e o impacte ambiental constituem dois dos aspectos mais importantes de qualquer operação industrial. Conhecidos por HSE (do inglês *Health, Safety and Environment*), são uma preocupação fundamental de qualquer indústria. No entanto, a história do petróleo e as dificuldades inerentes aos projectos de exploração e de produção, tornam esta indústria muito mais susceptível a problemas e questões levantadas por governos, autoridades locais e grupos ambientalistas.

De tal modo se desenvolveu a componente HSE nos projectos petrolíferos que a política e os procedimentos nesta área se tornaram parte integrante da gestão das melhores companhias. É habitualmente aceite que as companhias petrolíferas com melhores níveis de HSE são também as que apresentam melhores resultados económicos [Gomes & Alves, 2007].

A Galp Energia entende que a protecção do ambiente, a segurança e a saúde dos seus colaboradores, clientes e comunidade em geral, são valores essenciais para a sustentabilidade da Empresa e, nessa medida, está consciente da sua responsabilidade na gestão do impacte das suas actividades, produtos e serviços na Sociedade em que se insere. Estabelece assim um compromisso de integrar a Segurança, Saúde e Ambiente (SSA) na estratégia e actividades da empresa, fazendo destes pilares da gestão e

contribuindo, dessa forma, para alcançar o desenvolvimento sustentável e a excelência empresarial.

A Galp Energia compromete-se a:

- Consagrar a SSA como valores fundamentais da empresa;
- Assumir que a gestão da SSA é uma responsabilidade directa dos líderes e a prevenção de riscos uma responsabilidade de todos;
- Promover a formação e o treino de todos os colaboradores nesta matéria, envolvendo parceiros e demais partes interessadas, comprometendo-os com as questões de SSA para que actuem proactivamente dentro e fora do ambiente de trabalho;
- Aplicar as melhores práticas de gestão e soluções técnicas disponíveis, para além do cumprimento da legislação, nas estratégias de prevenção contínua mediante a identificação, controlo e monitorização de riscos para garantir a SSA;
- Criar condições para que a organização, como um todo, se mantenha permanentemente preparada para responder a emergências;
- Assegurar a sustentabilidade de projectos, empreendimentos e produtos, ao longo do seu ciclo de vida, mediante a utilização de tecnologias, instalações, recursos e práticas que previnem ou minimizam consequências adversas;
- Estabelecer metas e objectivos desafiadores, medindo e avaliando os resultados obtidos e tomando as acções necessárias à sua prossecução;
- Assegurar a utilização eficiente da energia e recursos e a incorporação de tecnologias seguras e inovadoras na gestão das suas actividades, minimizando o impacte, de forma a garantir a sustentabilidade da empresa e a protecção do ambiente; e
- Informar e divulgar a presente política, de forma responsável e transparente, às partes interessadas, comunicando o desempenho da empresa a nível de SSA.

A Galp Energia assume-se, portanto, como uma empresa social e ambientalmente responsável, constituída por uma equipa motivada, competente e inovadora, empenhada em proteger o ambiente, a segurança e a saúde dos seus colaboradores, clientes, parceiros e da comunidade, contribuindo activamente para o bem-estar da Sociedade [Galp, 2007a].

1.2. A Gestão do Conhecimento

A Gestão do Conhecimento tem a ver com o aumento do desempenho das organizações, ao permitir-lhes capturar, partilhar e aplicar o conhecimento que emana de um ou mais colectivos, de forma a tomar decisões em tempo real. A Gestão do Conhecimento é a pedra angular para que as empresas, organizações e instituições sobrevivam numa sociedade e economia competitivas. Por outro lado, o fornecimento da informação, do conhecimento e da experiência, de forma a tomar as melhores decisões, deve ser concretizado em tempo real. Por tempo real entende-se o tempo disponível para tomar a decisão, i.e. desenvolver uma acção que afecte positivamente os resultados. A Gestão do Conhecimento está ligada a questões de eficiência, de produtividade, de qualidade e do recurso à experiência e a formas de meta conhecimento. Outros pressupostos, com implicações a médio e longo prazo, são necessários para produzir resultados. Não é suficiente apostar na tecnologia, há que melhorar processos internos e externos em termos de gestão e organização, motivar e formar as pessoas, olhar para o seu comportamento dentro das organizações e, sustentar uma maior capacidade de inovação [Abelha, 2004] [Vicente, 2004].

Uma das tarefas mais árduas dentro de uma organização tem naturalmente a ver com a partilha de conhecimento entre indivíduos, quer estes estejam próximos ou separados, para que seja possível:

- Partilhar o que sabem;
- Aplicar o conhecimento de terceiros;
- Capturar e partilhar o conhecimento num formato passível de ser aplicado dentro da organização;
- Melhorar os canais de comunicação, sejam estes locais ou remotos;
- Interagir com os colegas, em qualquer local, em qualquer altura e de diferentes formas; e
- Utilizar novas aplicações, tecnicamente avançadas, assim como o uso de ferramentas de suporte à decisão, como a mineração de dados, a extracção de conhecimento e a simulação.

A informação pode, então, ser definida como um conjunto de dados que são interpretados e o conhecimento como a informação em acção, i.e. a informação transformada em capacidade para dar corpo a acções efectivas. Uma acção pressupõe uma

alteração ao estado do mundo em que se desenvolve. O conhecimento é visto como o recurso mais importante da sociedade contemporânea, muitas vezes referida por sociedade da informação ou do conhecimento, tal como aconteceu com a qualidade e os serviços de apoio ao cliente em décadas anteriores. A capacidade de aprender rapidamente e continuamente é uma chave para o sucesso das organizações [Abelha, 2004] [Vicente, 2004].

1.3. Inovação e Valor Acrescentado do Trabalho

A relevância de qualquer tipo de informação para a resolução de um dado problema está na razão directa da frequência com que é considerada e do modo como é aplicada nessas situações. A fonte de informação ideal seria aquela que facultasse apenas o conhecimento e os dados relevantes para a resolução do problema com o mínimo de esforço por parte do utilizador. As deficiências com que as actuais fontes externas de informação se apresentam torna claro que algo mais é necessário, e este parece ser o mote para o aparecimento de alternativas com base na electrónica e na robótica e que sejam portáteis e fáceis de utilizar. Tais alternativas, não poderão ter uma atitude passiva, ou seja, limitarem-se a responder a questões colocadas pelos utilizadores, mas subscrever um comportamento pró-activo, o que passa por interrogar o decisor e explicar-lhe as suas opções, procurando ajudar, encontrando uma solução técnica eficiente, desenvolvendo a alternativa mais amiga do ambiente e já mais se assumindo como uma entidade indesejada.

Neste trabalho aplicam-se técnicas oriundas da Área Científica da Inteligência Artificial ao desenvolvimento de modelos de previsão da classificação da água residual industrial de acordo com o Regulamento de Recolha e Tratamento de Água Residual Industrial do Sistema de Santo André (RARISA).

O presente trabalho não pretende fornecer soluções definitivas, mas sim contribuir com uma pequena parcela de conhecimento que possibilite algum avanço científico nesta área de estudo. Desta forma, a presente dissertação possui os seguintes objectivos principais:

- Conhecer o estado actual da problemática ambiental relacionada com a indústria petrolífera;

- Fazer a revisão bibliográfica sobre os conceitos ligados à Indústria Petrolífera;
- Minimizar os custos relacionados com a aplicação de tarifas pela Concessionária (Sociedade denominada por Águas de Santo André) à Refinaria de Sines;
- Utilizar as ferramentas da Área Científica da Inteligência Artificial no desenvolvimento de modelos de previsão da classificação da água residual industrial em função dos programas de fabrico de Petróleo Bruto na Refinaria de Sines e das suas diferentes características;
- Realçar a importância de tais ferramentas e dos modelos construídos quando aplicadas a uma empresa do sector petrolífero;
- Realçar a capacidade dos modelos na descrição do estado da qualidade dos efluentes em tempo real e fazer previsões sobre cenários futuros;
- Tratar os efluentes de forma a aumentar a qualidade de descarga, cumprindo o regulamento de descarga em vigor;
- Apurar quais as melhorias que a Refinaria de Sines pode introduzir no seu relacionamento com o Ambiente; e
- Construir modelos potencialmente aplicáveis à gestão de uma empresa do sector petrolífero, constituindo um caminho para a Gestão de Excelência.

1.4. Organização da Dissertação

A dissertação encontra-se estruturada e organizada em sete Capítulos, não se contabilizando nem as Referências Bibliográficas nem os Anexos. No Capítulo I são efectuadas algumas considerações sobre a problemática ambiental relacionada com o universo do petróleo, expondo-se o enquadramento, a fundamentação, o objectivo do trabalho e a caracterização da entidade acolhedora.

No Capítulo II apresenta-se a origem e a formação do Petróleo, os principais processos envolvidos na sua refinação e os produtos finais típicos.

No Capítulo III referem-se os aspectos ambientais a ter em conta na indústria petrolífera nomeadamente a análise e o tratamento dos efluentes industriais produzidos.

O Capítulo IV trata do processo de Descoberta de Conhecimento em Bases de Dados (DCBD) sendo apresentados os objectivos, os conceitos, os tipos de abordagem, as áreas relacionadas, as metodologias e as especificações.

O Capítulo V está dividido em duas partes. A primeira diz respeito à amostragem, ao programa de amostragem e à frequência de amostragem cumprido pela Refinaria de Sines bem como às técnicas analíticas utilizadas na determinação dos parâmetros utilizados neste estudo. Na segunda parte descreve-se a construção da base de dados que irá ser utilizada para o desenvolvimento dos modelos.

No Capítulo VI são fundamentadas as opções tomadas durante o processo de DCBD. Apresentam-se, discutem-se e avaliam-se quer os resultados quer os modelos construídos.

Por fim, no Capítulo VII, apresentam-se as conclusões finais e perspectiva-se o trabalho que se pretende desenvolver no futuro.

1.5. Conclusões

Neste Capítulo para além de se explicitar o objectivo e a necessidade do trabalho, caracterizou-se o local onde se realizou o presente estudo, o seu funcionamento e a sua organização. Apresentou-se de forma muito sucinta a Refinaria de Sines e o seu objectivo de maximização de produção de gasolinas a partir da matéria-prima, cuja química associada será explicitada no Capítulo II. Demonstrou-se a existência de medidas de adequação ambiental, segurança e saúde.

Capítulo II – Química do Petróleo

A origem e a formação do petróleo; o processo de refinação e produtos finais típicos.

A maior parte dos combustíveis líquidos são derivados do petróleo bruto. Se o seu consumo continuar a seguir os padrões do passado recente, as reservas de petróleo bruto actualmente identificadas serão suficientes somente para as próximas duas/três décadas. Face a este cenário, é previsível que, cada vez mais, os combustíveis líquidos venham a ser obtidos a partir do carvão, xistos betuminosos, areias asfálticas e biomassa. Porém, devido às actuais dificuldades técnicas e aos elevados custos associados à extracção de combustíveis líquidos destas matérias-primas, é expectável que o petróleo convencional permaneça como a principal matéria-prima das refinarias nos próximos anos [Gary & Handwerk, 2001] [Ramage, 2003].

O mundo dos nossos dias gira à volta do petróleo. Todos nós, de uma maneira ou de outra, dependemos desse produto precioso, que se esconde no espaço poroso de uma rocha a muitas centenas ou milhares de metros de profundidade [Gary & Handwerk, 2001] [Gomes & Alves, 2007].

Em sentido lato, podemos considerar que o termo petróleo abrange todas as ocorrências ou concentrações naturais de hidrocarbonetos, qualquer que seja o estado físico em que se encontrem. Em sentido mais restrito, reflectindo já uma perspectiva comercial, o termo é usado para referir hidrocarbonetos que ocorrem exclusivamente no estado líquido, constituindo o produto designado por Petróleo Bruto ou *Crude Oil*. Os hidrocarbonetos que ocorrem sob a forma de gás constituem o gás natural e os que se apresentam no estado sólido são conhecidos por betumes e asfaltos.

De uma maneira geral, todos nós conhecemos o petróleo e estamos conscientes da sua utilidade, em particular como fonte de energia. Na verdade, a sua utilização remonta aos tempos bíblicos, embora nessa altura não fosse explorado activamente. Pequenas nascentes naturais do petróleo foram, desde há séculos, uma fonte de betume, material utilizado na construção das cidades antigas. Para além disso, o petróleo empregava-se para

impermeabilizar navios, para embalsamar, para tratar cabedal e, até mesmo na medicina, tinha utilidade [Gary & Handwerk, 2001] [Gomes & Alves, 2007].

Em meados do século XIX aparecem as refinarias e foi então possível separar vários componentes do petróleo bruto permitindo, desta forma, o seu emprego em novas utilizações. Durante várias décadas o petróleo foi usado na iluminação. Contudo, isto só foi possível quando se conseguiu separar o querosene dos outros compostos do crude. A invenção do motor de combustão interna, já no final do século XIX, e a sua rápida adopção à maioria dos meios de transporte, tornaram o petróleo indispensável como combustível e, a partir de então, o seu consumo aumentou de forma constante. Hoje o petróleo é responsável por cerca de 40% da produção de energia a nível mundial e, na sua forma refinada, continua a ter bastantes utilizações, tais como a manufactura de materiais sintéticos, de que os plásticos são um bom exemplo [Gary & Handwerk, 2001] [Ramage, 2003].

II.1. Origem do Petróleo

O petróleo é formado a partir de material orgânico que sedimentou no fundo dos mares, há milhões de anos atrás, tendo sido coberto por camadas de sedimentos e enterrado a grandes profundidades. O processo de formação é muito lento e, embora não se saiba exactamente como se formaram os depósitos que hoje encontramos, uma sequência de acontecimentos poderá ser aquela que a seguir se descreve, nomeadamente a Sedimentação do Material Orgânico, a Metamorfose e a Migração [Gomes & Alves, 2007].

1. Sedimentação

Material orgânico diverso que, por algum motivo, não foi decomposto deposita-se no fundo dos oceanos e é incorporado nos sedimentos. Uma das razões pela qual nem todo o material orgânico é decomposto, levando ao início do processo de formação do petróleo, pode ser o facto de, em sistemas altamente produtivos, a Produção Primária Líquida (PPL) exceder a velocidade de oxidação do carbono, provocando o aumento da velocidade de sedimentação. A PPL designa a taxa de armazenamento da matéria orgânica nos tecidos, i.e. mede a energia armazenada em excesso numa planta relativamente às suas necessidades

respiratórias e, portanto, potencialmente disponível para os organismos heterotróficos. Noutros sistemas, onde existam águas profundas estagnadas, podem desenvolver-se condições anóxicas que retardam bastante a oxidação do carbono dos detritos. Uma outra hipótese será, ainda, a produção de compostos refractários de carbono que, por isso mesmo, são de oxidação difícil. Qualquer que seja a razão, o primeiro passo na génese do petróleo prende-se com o facto de este material orgânico se depositar no fundo dos oceanos e aí permanecer durante um período de tempo suficiente para ser coberto pelos sedimentos. Após ser envolvido nos sedimentos, a oxidação aeróbia desse material torna-se praticamente impossível, uma vez que, de um modo geral, o nível de oxigénio desce para zero abaixo de uma profundidade de alguns centímetros [Gomes & Alves, 2007].

2. Metamorfose

Uma vez enterrado nos sedimentos o carbono orgânico acabará por ser incorporado em rochas sedimentares como os xistos argilosos, os arenitos ou as rochas carbonatadas. Sob determinadas condições, durante longos períodos de tempo, o carbono orgânico é lentamente transformado, originando uma complexa mistura de hidrocarbonetos. Deste modo, o petróleo forma-se na ausência de oxigénio, sob a influência de condições físicas (temperaturas e pressões elevadas) e químicas (é possível que a presença de catalisadores seja importante) que são peculiares às rochas sedimentares, enterradas a grandes profundidades [Gomes & Alves, 2007].

3. Migração

Pela análise de rochas sedimentares verifica-se que, na sua composição, a concentração de carbono orgânico é muito baixa. Esta observação levanta alguns problemas quanto ao local da génese de depósitos de petróleo que apenas pode ser explicada pelo facto deles não serem formados nos mesmos locais onde são encontrados. Quer isto dizer que haverá algum mecanismo que leva o petróleo recém-formado a migrar para locais afastados das rochas de origem, formando depósitos concentrados. Os mecanismos desta migração não são ainda conhecidos, embora se pense que, uma vez formado, o petróleo líquido tenha tendência para subir, atravessando os materiais porosos que formam a crosta terrestre. Esta ascensão terminaria quando o petróleo encontrasse uma camada rochosa impermeável,

ficando assim aprisionado. O mecanismo de migração parece bastante razoável, sendo ainda reforçado pelo facto de os depósitos de petróleo estarem, muitas vezes, cobertos por bolsas de gás e, invariavelmente terem por baixo um lençol de água. A associação de gás e água com petróleo reflecte a tendência que as substâncias fluidas apresentam para ascender, através de materiais porosos, até encontrarem um substrato impermeável que as detenha. Este tipo de migração não é, portanto, uma propriedade exclusiva do petróleo [Gomes & Alves, 2007].

O petróleo é, talvez, a mistura natural mais complexa no nosso planeta. Na sua constituição predominam os hidrocarbonetos (50 a 98%) sendo a restante parte constituída por outras substâncias como o enxofre, o azoto, compostos contendo oxigénio ou metais como o níquel, o vanádio, o ferro e o cobre [Gomes & Alves, 2007].

II.2. Composição Química do Petróleo

Embora todo o petróleo seja formado da mesma maneira, a partir do material orgânico, a sua composição pode apresentar grandes variações nas diferentes zonas do Mundo onde é encontrado. Estas variações dependem de diversos factores, podendo salientar-se, entre outros, a idade do petróleo, as condições da sua formação e os materiais que nele são naturalmente incorporados. Os seus componentes podem ocorrer na fase gasosa - gás natural; na forma líquida - crude ou petróleo em rama; na forma sólida ou semi-sólida - alcatrão ou pez, muitas vezes associado a arenitos. No seu conjunto, estes três materiais compreendem centenas de espécies químicas que variam em complexidade, em massa e em composição (para além do carbono e do hidrogénio podem conter outros elementos químicos, como o enxofre, o azoto e o oxigénio).

A Tabela II.1 indica as percentagens médias dos elementos no petróleo bruto (crude) e no gás natural.

Tabela II. 1- Composição Elemental do Petróleo Bruto e do Gás Natural. Adaptado de [Gomes & Alves, 2007]

ELEMENTO	(%) NO PETRÓLEO BRUTO (EM MASSA)	(%) NO GÁS (EM MASSA)
C	82-87	65-80
H	12-15	1-25
S	0,1-5,5	0-0,2
N	0,1-1,5	1-15
O	0,1-4,5	--

O petróleo é uma mistura complexa de muitas moléculas de hidrocarbonetos que variam de 1 a 60 ou mais átomos de carbono. A razão básica desta diversidade reside na capacidade única que o átomo de carbono tem de formar ligações químicas quer com átomos do mesmo elemento, quer com átomos de outros elementos, nomeadamente o hidrogénio, como no caso dos hidrocarbonetos. Um outro aspecto a salientar, está relacionado com a capacidade dos átomos de carbono formarem cadeias lineares ou cíclicas mais ou menos longas que podem exibir ramificações. A abundância relativa dos diversos compostos varia de acordo com o tipo de petróleo e determina as suas propriedades físico-químicas.

A quase totalidade dos hidrocarbonetos que constituem o petróleo pertence a quatro grandes séries ou famílias: hidrocarbonetos parafínicos, nafténicos, aromáticos e asfálticos. Embora os vários tipos de petróleo, definidos pela presença dominante de uma destas famílias, possam existir numa qualquer região petrolífera, é habitual encontrar-se um tipo mais abundante. Trata-se de um facto relacionado com as características químicas da rocha geradora com a matéria orgânica presente, bem como com a temperatura do meio. Como tal, de acordo com a predominância do tipo de hidrocarbonetos encontrados no petróleo bruto, este pode ser classificado como parafínico, nafténico, aromático e asfáltico [Ali et al., 2005] [Gomes & Alves, 2007]

II.2.1. Hidrocarbonetos Parafínicos

Os hidrocarbonetos parafínicos, também designados por alifáticos, possuem uma estrutura molecular do tipo linear, onde os átomos de carbono possuem ligações simples (saturadas) com outros átomos de carbono. A fórmula geral destes hidrocarbonetos é C_nH_{2n+2} , sendo "n" o número de átomos de carbono. O primeiro membro desta série, o metano é o componente principal do gás natural. Em condições normais de pressão e temperatura, os hidrocarbonetos parafínicos podem ser gasosos, líquidos ou sólidos consoante o número de átomos de carbono que possuam. Assim, os hidrocarbonetos cujas moléculas possuem menos de cinco átomos de carbono aparecem sob a forma de gases e são os principais constituintes do gás natural. Estes mesmos compostos, em combinação com os hidrocarbonetos mais pesados, formam a base dos petróleos brutos e de um grande número de produtos petrolíferos, desde as gasolinas até aos lubrificantes. Os hidrocarbonetos que contêm dezoito ou mais átomos de carbono são os principais constituintes da parafina sólida [Ali et al., 2005] [Gomes & Alves, 2007].

II.2.2. Hidrocarbonetos Nafténicos

Os hidrocarbonetos nafténicos, também conhecidos pela designação de alifáticos cíclicos, possuem uma composição química que é expressa pela fórmula geral C_nH_{2n} . São hidrocarbonetos saturados, cujo arranjo estrutural das suas moléculas se caracteriza por serem cíclicos. A estabilidade dos três primeiros membros desta família aumenta do ciclopropano para o ciclopentano, sendo o ciclohexano muito estável e frequentemente usado como solvente em reacções químicas. Muitos dos hidrocarbonetos nafténicos têm propriedades similares às dos hidrocarbonetos de cadeia aberta. Por exemplo, o ciclohexano é um líquido incolor à temperatura ambiente, inflamável e não é miscível com a água - propriedades partilhadas com o seu homólogo de cadeia linear, o n-hexano.

Os ácidos nafténicos são os principais responsáveis pela corrosão em fase líquida durante a refinação do petróleo. Alguns estudos demonstram que existe uma relação entre a estrutura molecular do ácido nafténico e a sua corrosividade, demonstrando que os ácidos com maior peso molecular são menos corrosivos [Slavcheva, 1998] [Lewis et al., 1999].

O número de acidez total (TAN) determinado segundo o método ASTM D 664 aprovado pela "American Society for Testing and Materials", é uma medida internacionalmente aceite do conteúdo de ácidos totais do petróleo [Ali et al., 2005] [Gomes & Alves, 2007].

II.2.3. Hidrocarbonetos Aromáticos

Os hidrocarbonetos aromáticos são compostos cíclicos, tal como os nafténicos, mas que se distinguem por exibir pelo menos um anel benzénico.

Os hidrocarbonetos aromáticos mais simples, como o benzeno, o tolueno e o xileno, encontram-se em pequenas quantidades na maior parte dos petróleos brutos e apenas em relativa abundância em alguns deles.

Existem ainda outros grupos de hidrocarbonetos insaturados, conhecidos por alcenos e alcinos, que não figuram, todavia, entre os constituintes do petróleo bruto, ou quando figuram é apenas em proporções ínfimas. Formam-se por conversão molecular durante as operações de refinação e são produtos de grande interesse na indústria petroquímica, visto serem hidrocarbonetos não saturados e, por isso, bastante reactivos. Possuem estruturas moleculares que formam cadeias lineares ou ramificadas com uma ou mais ligações duplas (alcenos) ou triplas (alcinos) [Gomes & Alves, 2007].

II.2.4. Hidrocarbonetos Asfálticos

Os hidrocarbonetos designados por asfálticos encontram-se, à temperatura ambiente, no estado sólido ou semi-sólido (betume). Têm um ponto de ebulição elevado e apresentam elevada massa molar, contendo entre 40 a 60 átomos de carbono. Os hidrocarbonetos asfálticos não têm grande valor comercial, não sendo portanto um produto valorizado pelas refinarias [Ali et al., 2005] [Gomes & Alves, 2007].

A Tabela II.2 mostra um exemplo da composição de um petróleo em função das fracções parafínica, nafténica, aromática e asfáltica.

Tabela II. 2- Constituição de um Crude (Fracções de Hidrocarbonetos). Adaptado de [Gomes & Alves, 2007]

TIPO MOLECULAR	% (EM MASSA)
Parafínicos	30
Nafténicos	49
Aromáticos	15
Asfálticos	6

As refinarias utilizam principalmente dois tipos de crudes, o crude asfáltico e o crude parafínico. O primeiro apresenta uma cor negra, poucas parafinas e, quando refinado, produz gasolinas de alto valor (*high-grade*) e asfaltos. O crude parafínico contém pouco ou nenhum asfalto, possui uma cor esverdeado e, quando refinado, produz parafinas, óleos lubrificantes de grande qualidade e petróleo (querosene).

O tipo de produtos resultantes da refinação do petróleo bruto é função das suas características. De facto, os petróleos parafínicos são, como o nome indica, constituídos principalmente por parafinas, enquanto os nafténicos têm uma percentagem maior de ciclo-parafinas. Os petróleos asfálticos têm mais hidrocarbonetos pesados, sendo usados para a produção de asfaltos e de combustíveis. A partir dos petróleos aromáticos é possível extrair compostos aromáticos como o benzeno, o tolueno e o xileno, com aplicação na indústria dos solventes.

Na Tabela II.3 mostra-se a composição média de um petróleo, tal como é retirado do reservatório petrolífero através do poço de produção. Por outras palavras, trata-se de uma composição representativa de petróleo nas condições de subsuperfície, ou seja, no próprio reservatório onde se acumulou há milhões de anos.

Após a separação da água e do gás associado, operação conduzida nas instalações de separação e armazenamento do campo petrolífero, a composição de um petróleo bruto, tal como ele é transportado e comercializado, tem um padrão claramente diferente do anterior, tal como mostrado na Tabela II.4. Os hidrocarbonetos leves (metano, etano e propano) praticamente não se encontram presentes, devido à remoção do gás associado [Ali et al., 2005] [Gomes & Alves, 2007].

Tabela II. 3- Composição de um Petróleo em Condições de Subsuperfície. Adaptado de [Gomes & Alves, 2007]

COMPOSTO	FAMÍLIA	% MOLAR
Azoto	Compostos não Hidrocarbonetos	0,730
Dióxido de Carbono		0,153
Ácido Sulfídrico		0,000
Metano	Metano	23,736
Etano	Etano	5,517
Propano	Propano	3,227
Iso-butano	Butanos	1,312
n-butano		3,461
Iso-pentano	Pentanos	1,876
n-pentano		2,109
Hexanos parafínicos	Hexanos	3,549
Hexanos Nafténicos		0,104
Aromáticos C6 (benzeno)	Pseudo-Heptanos	0,130
Heptanos parafínicos		2,751
Heptanos nafténicos		1,885
Aromáticos C7 (tolueno)	Pseudo-Octanos	0,530
Octanos parafínicos		2,923
Octanos nafténicos		2,777
Aromáticos C8 (xileno)	Pseudo-nonanos	0,754
Nonanos Parafínicos		2,560
Nonanos Nafténicos		0,846
Decanos	Pseudo-decanos	39,070
		100,000

Tabela II. 4- Composição de um Petróleo Bruto. Adaptado de [Gomes & Alves, 2007]

COMPOSTO	FAMÍLIA	% MOLAR	% MASSA
Azoto	Compostos não Hidrocarbonetos	0,00	0,00
Dióxido de Carbono		0,00	0,00
Ácido Sulfídrico		0,00	0,00
Metano	Metano	0,00	0,00
Etano	Etano	0,01	0,00
Propano	Propano	0,29	0,06
Iso-butano	Butanos	0,41	0,12
n-butano		1,68	0,49
Iso-pentano	Pentanos	1,91	0,69
n-pentano		2,93	1,07
Hexanos parafínicos	Hexanos	3,6	1,56
Hexanos Nafténicos		1,94	0,82
Aromáticos C6 (benzeno)	Pseudo-Heptanos	0,42	0,17
Heptanos parafínicos		6,92	3,31
Heptanos nafténicos		3,16	1,59
Aromáticos C7 (tolueno)	Pseudo-Octanos	1,35	0,63
Octanos parafínicos		6,13	3,53
Octanos nafténicos		0,29	0,17
Aromáticos C8 (xileno, etil-benzeno)	Pseudo-nonanos	3,02	1,62
Nonanos Parafínicos		5,50	3,55
Nonanos Nafténicos		0,15	0,10
Aromáticos C9 Decanos	Pseudo-decanos	5,93	4,28
	Pseudo-undecanos	6,21	4,60
	Pseudo-dodecanos	5,09	4,13
	Pseudo-tridecanos	4,30	3,79
	Pseudo-tetradecanos	3,35	3,21
	Pseudo-pentadecanos	3,53	3,67
	Pseudo-hexadecanos	2,93	3,28
	Pseudo-heptadecanos	2,49	2,98
	Pseudo-octadecanos	2,29	2,90
	Pseudo-nonadecanos	2,21	2,93
	Pseudo-eicosanos	1,94	2,69
	Pseudo-heneicosanos	1,70	2,49
	Pseudo-docosanos	1,50	2,31
	Pseudo-trocosanos	1,33	2,12
	Pseudo-tetracosanos	1,16	1,94
	Pseudo-pentacosanos	1,07	1,86
	Pseudo-hexacosanos	0,93	1,68
	Pseudo-heptacosanos	0,89	1,67
	Pseudo-octacosanos	0,80	1,56
	Pseudo-nonacosanos	0,75	1,52
Pseudo-triacontanos	8,54	24,10	
		100,00	100,00

II.3. Impurezas do Petróleo

A análise elemental de um combustível é determinada experimentalmente em laboratório e quantifica as percentagens dos elementos individuais de um combustível, especificamente, as percentagens mássicas de carbono, hidrogénio, azoto, enxofre, oxigénio e cinzas.

Excluindo os hidrocarbonetos, vários tipos de compostos ocorrem nos óleos de crudes. Os mais importantes são os compostos de enxofre, os compostos oxigenados e os compostos azotados, por exemplo, o fuel contendo elevadas quantidades de enxofre e azoto causa, por um lado, problemas ambientais e, por outro, problemas ligados à natureza corrosiva dos seus produtos de oxidação [Meyers, 1996] [Reis, 1996].

Os compostos metálicos, embora em quantidades vestigiais, também são encontrados em todos os crudes. A presença deste tipo de impurezas causa alguns problemas em certos processos catalíticos [Meyers, 1996] [Matar & Hatch, 2000].

Um petróleo bruto com percentagens elevadas de impurezas tem menos procura por parte das refinarias, sendo transaccionado a um preço inferior aos preços de referência do mercado. Por exemplo, um petróleo com elevado teor de compostos de enxofre é normalmente classificado como ácido, sendo menos procurado por originar problemas relacionados com a segurança da operação e com fenómenos de corrosão.

Seguidamente, apresenta-se uma breve caracterização dos tipos de impurezas mais comuns.

II.3.1. Compostos de Enxofre

Os compostos de enxofre são os mais importantes entre as impurezas presentes no petróleo. Estes compostos são normalmente instáveis termicamente e transformam-se, por craqueamento, em hidrocarbonetos mais leves gás sulfídrico e mercaptans.

A presença destes compostos nos produtos de petróleo é muito prejudicial, pois são responsáveis pela corrosividade desses produtos e pelo envenenamento dos catalisadores utilizados nos processos de transformação. São tóxicos, produzindo, por combustão, óxidos de enxofre que são poluentes atmosféricos. Outros problemas causados prendem-se com a



formação de gomas, diminuição de estabilidade dos combustíveis, deterioração da cor e odor desagradável [Meyers, 1996] [Speight, 2001].

II.3.2. Compostos Oxigenados

Os compostos oxigenados são quase tão abundantes quanto os compostos de enxofre. De um modo geral, são responsáveis pela acidez, coloração, odor, formação de gomas e corrosividade dos derivados de petróleo [Meyers, 1996] [Speight, 2001].

II.3.3. Compostos Azotados

Os compostos azotados existentes no petróleo podem ser classificados como básicos e não básicos. Em geral a concentração de compostos de azoto no óleo cru é baixa mas é de grande importância nas refinarias. A presença desses compostos podem ser responsáveis pelo envenenamento dos catalisadores no processo de craqueamento e também podem contribuir para a formação de gomas nos produtos de petróleo, tal como, o óleo combustível doméstico [Meyers, 1996] [Speight, 2001].

II.3.4. Compostos Metálicos

Os constituintes metálicos são encontrados em vários óleos crus e a sua concentração deve ser reduzida antes da conversão do óleo em combustíveis de transporte. A presença de metais afecta muitos processos e causa problemas particulares como o envenenamento dos catalisadores usados para remover o enxofre e o azoto e também outros processos, tal como, o craqueamento catalítico [Speight, 2001]. Além disso, metais como o cobre, o ferro, o níquel e o chumbo dificultam o armazenamento dos produtos, potencializando reacções indesejáveis, causando a formação de goma e grumos no interior dos recipientes [Cole, 1994] [Speight, 2001].

Os petróleos com maiores teores de enxofre e asfaltenos apresentam maiores concentrações de metais [Gary & Handwerk, 2001]. Os referidos petróleos, de acordo com American Petroleum Institute (API), são classificados como sendo petróleos ácidos por

apresentarem 2,5% de enxofre na sua composição. Os metais de maior relevância nos petróleos são ferro, níquel, cobre e vanádio, mas somente uma fracção dos metais encontrados nos destilados provém do petróleo [Cole, 1994].

Normalmente, na refinação do petróleo os metais, mesmo em quantidades vestigiais, não deveriam estar presentes. Contudo, alguns deles podem ser adicionados como catalisadores, anticorrosivos ou dispersantes, como é o caso do chumbo, do bário e do cálcio [Cole, 1994]. Os metais cuja presença não está relacionada com nenhuma das formas mencionadas anteriormente podem ter a sua origem na contaminação por recipiente, pelo ar ou pelos líquidos refrigerantes usados nos equipamentos. Alguns compostos presentes nos combustíveis, tais como pirróis, mercaptans e ácidos nafténicos agem como iniciadores de reacções de oxidação. Estes compostos reagem com metais permitindo a sua solubilização, o que facilita a acção catalítica destes nas reacções de oxidação, principalmente a altas temperaturas [Cole, 1994].

É de salientar que a quantidade de metais pesados presente nos produtos de petróleos serve para avaliar a quantidade desses metais que serão lançados na atmosfera pelo seu uso como combustível cujos efeitos podem ser bastante danosos para o meio ambiente.

II.4. Tipos de Petróleo Bruto

Em termos da sua utilização em refinaria, o petróleo bruto pode ser classificado de muitas formas, mas o mais habitual é agrupá-lo em vários tipos, baseados não só na composição, mas também na densidade e na presença de enxofre, tal como é exemplificado na Tabela II.5 [Gomes & Alves, 2007].

Tabela II. 5- Tipos de Petróleo Bruto. Adaptado de [Gomes & Alves, 2007]

TIPOS DE PETRÓLEO BRUTO	CONTEÚDO DE ENXOFRE (% EM MASSA)	DENSIDADE (GRAU API)
Leve Doce	0,10 - 0,40	33 - 55
Leve Ácido	1,00 - 2,50	30 - 45
Pesado Doce	0,20 - 0,80	19 - 29
Pesado Ácido	1,70 - 2,80	11 - 29

Trata-se de uma classificação intimamente ligada ao projecto da refinaria. De facto, as refinarias são projectadas para o processamento de certos tipos de petróleo, aqueles a que têm acesso pela disponibilidade do mercado. Este mercado distribui-se regionalmente, com os petróleos do Médio Oriente a serem classificados, na sua grande maioria, no tipo leve ácido, enquanto a África e a Ásia-Pacífico têm petróleos do tipo leve doce.

A qualidade do petróleo bruto, para efeitos de mercado, depende essencialmente de dois factores: da densidade (grau API) e do conteúdo de enxofre. Quanto menos denso (petróleo leve) e quanto menor for a percentagem de enxofre na sua composição (petróleo doce) mais elevado será o seu valor comercial. Em termos gerais, uma percentagem de enxofre acima de 1% leva a classificar o petróleo bruto como ácido, sendo considerado pesado para um grau API inferior a 30, leve entre 30 e 45 graus API e condensado para um grau API acima de 45 [Speight, 2001] [Gomes & Alves, 2007].

O petróleo bruto de melhor qualidade assume preços mais elevados no mercado, mas é fundamental destacar que a definição de qualidade varia de cliente para cliente, em função das especificações das refinarias que, por seu lado, estão dependentes dos mercados que abastecem, nomeadamente dos tipos de produtos refinados mais procurados na área de influência que cobrem.

Importa realçar que, para o mesmo petróleo bruto, é possível produzir diferentes produtos refinados, de acordo com o projecto e os processos aplicados em cada refinaria. Estes devem adaptar-se não só aos petróleos brutos de mercado de mais fácil acesso, mas também ao tipo de produtos que o mercado local pretende. O objectivo do refinador é, portanto, maximizar o seu lucro através da aquisição do petróleo bruto mais adequado às especificações da sua refinaria e aos produtos que pretende colocar no mercado. O seu problema não é tanto o preço do petróleo bruto, mas sim o lucro que pode gerar a partir dos produtos refinados. Desta forma, é possível existir em refinadores a valorizar de forma diferente o mesmo petróleo bruto. Assim, as margens de lucro dos refinadores é diferente, tendo, nos últimos anos, mostrado uma tendência para aumentar acompanhando a alta do preço do barril.

II.5. Densidade – Grau API

A densidade é uma medida fundamental do valor do petróleo bruto. Como foi referido anteriormente, na Indústria petrolífera é usado o grau API proposto pelo *American Petroleum Institute*, que se relaciona com a densidade (g/cm^3) pela fórmula:

$$\text{Grau API} = (141,5/\text{densidade}) - 131,5$$

Assim, o petróleo de densidade inferior a $0,83 \text{ g/cm}^3$ ronda os 40 graus API, o que torna o petróleo de elevado teor económico. Os petróleos muito pesados aproximam-se da densidade da água e dos correspondentes 10 graus API [Gary & Handwerk, 2001] [Gomes & Alves, 2007].

II.6. A Refinação

O petróleo bruto é de pouca utilidade na sua forma original. O seu valor está nos produtos que se podem obter a partir dele, tal como combustíveis, solventes, naftas e gases associados. A partir destes últimos, é possível gerar produtos petroquímicos, tais como os polietilenos e polipropilenos, que são usados na fabricação de plásticos. A indústria de adubos e a farmacêutica vão também buscar aos derivados do petróleo alguma da sua matéria-prima [Gary & Handwerk, 2001] [Gomes & Alves, 2007].

Uma refinaria, Figura II.1, é uma instalação industrial extremamente complexa do ponto de vista tecnológico, onde se fabricam produtos combustíveis e matérias-primas para outras indústrias. Pode dizer-se que a função da refinaria é receber o petróleo bruto, dividi-lo em várias fracções, que ainda podem ser submetidas a processos físico-químicos, de onde resultam os muitos produtos de interesse comercial tão comuns no nosso dia-a-dia.

A refinação consiste numa série de tratamentos físicos e químicos que visam a separação do petróleo bruto em numerosos componentes, denominados derivados. De acordo com as características do petróleo bruto, escolhe-se um entre os vários processos de refinação. Contudo, há passos obrigatórios seguidos por qualquer processo. As Figuras II.2 a II.6 mostram esquematicamente vários processos que ocorrem numa refinaria.



Figura II. 1- Refinaria de Sines. Fonte [Galp, 2007b]

No processo de destilação fraccionada (Figura II.2), o petróleo bruto é pré-aquecido antes de entrar numa coluna de destilação atmosférica onde se processa a separação das diversas fracções, aproveitando a diferença das respectivas volatilidades. As fracções mais leves saem pelo topo, as intermédias lateralmente e as mais pesadas pelo fundo da coluna.

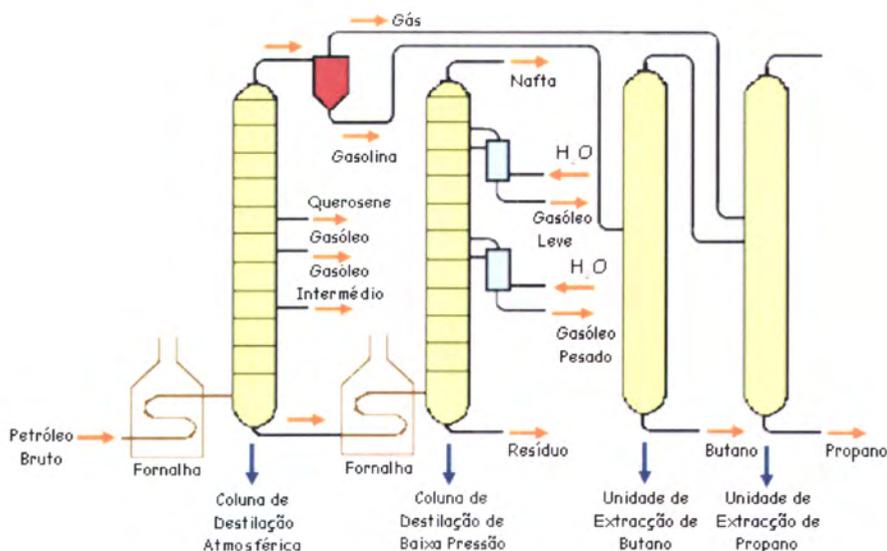


Figura II.2- Esquema do Processo de Destilação Fraccionada. Adaptado de [Coelho & Costa, 2007]

Após a destilação atmosférica é comum seguir-se uma outra a baixa pressão para otimizar o processo de separação dos componentes do petróleo. Posteriormente, as diferentes fracções são purificadas, para retirar o enxofre e alguns hidrocarbonetos perigosos [Gary & Handwerk, 2001] [Gomes & Alves, 2007]. A destilação fraccionada permite obter, em termos mássicos, cerca de 1% de metano, 4% de gases de Refinaria, 12% de gasolina leve, 18% de nafta, 17% de querosene, 7% de gasóleo pesado e 41% de resíduos. Os resíduos são geralmente sujeitos a uma nova destilação a baixa pressão para retirar o alcatrão (cerca de 11% da massa original de petróleo bruto). Os restantes materiais que constituem o resíduo são então tratados numa unidade de cracking (Figura II.3) orientada para a obtenção de gasolina e de outros produtos leves [Meyers, 1996] [Gary & Handwerk, 2001] [Gomes & Alves, 2007] [Coelho & Costa, 2007].

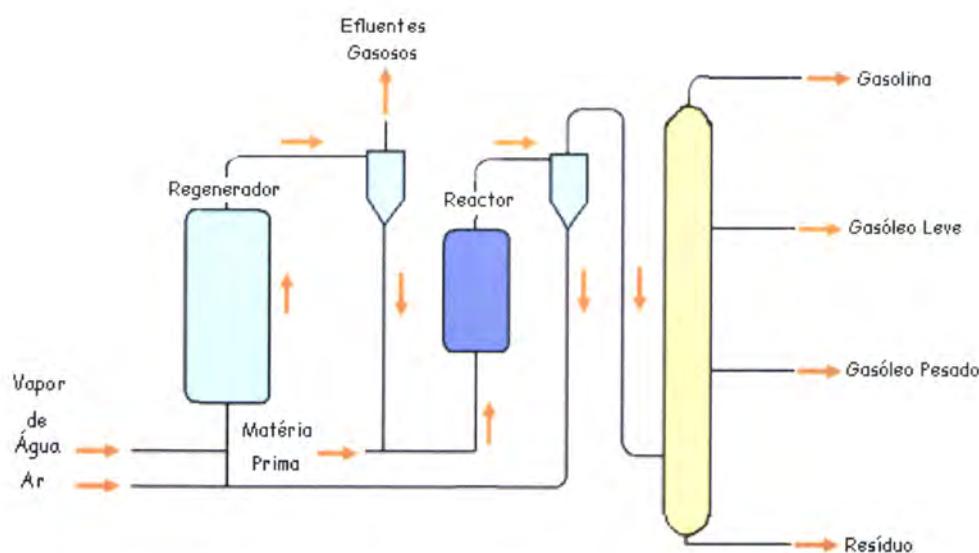


Figura II. 3- Esquema do Processo de Cracking Catalítico. Adaptado de [Coelho & Costa, 2007]

Destes materiais (cerca de 30% da massa original de petróleo bruto) obtém-se, em termos mássicos, cerca de 2% de gás de petróleo, 3% de butano, 10% de gasolina, 8% de gasóleo leve, 5% de gasóleo pesado e 2% de coque [Meyers, 1996] [Gary & Handwerk, 2001] [Gomes & Alves, 2007] [Coelho & Costa, 2007].

O gasóleo pesado resultante do processo de cracking catalítico é combinado com o alcatrão (no total cerca de 16% da massa original de petróleo bruto) e sujeito a um

processo de "cracking" térmico (Figura II.4) obtendo-se, em termos mássicos, cerca de 1% de gás de petróleo, 0,5% de butano, 3% de gasolina, 1,5% de gasóleo e 10% de *fuel-óleo* residual [Meyers, 1996] [Gary & Handwerk, 2001] [Gomes & Alves, 2007] [Coelho & Costa, 2007].

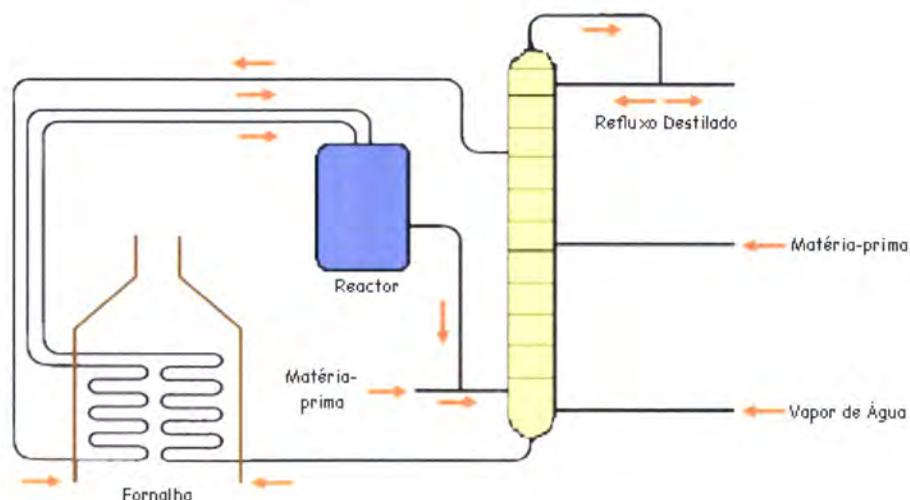


Figura II. 4- Esquema do Processo de *Cracking* Térmico. Adaptado de [Coelho & Costa, 2007]

O processo de reformação (Figura II.5) permite aumentar o índice de octano quer da nafta quer da gasolina obtida nas unidades de cracking, originando, deste modo, gasolinas de qualidade elevada [Meyers, 1996] [Gary & Handwerk, 2001] [Gomes & Alves, 2007] [Coelho & Costa, 2007].

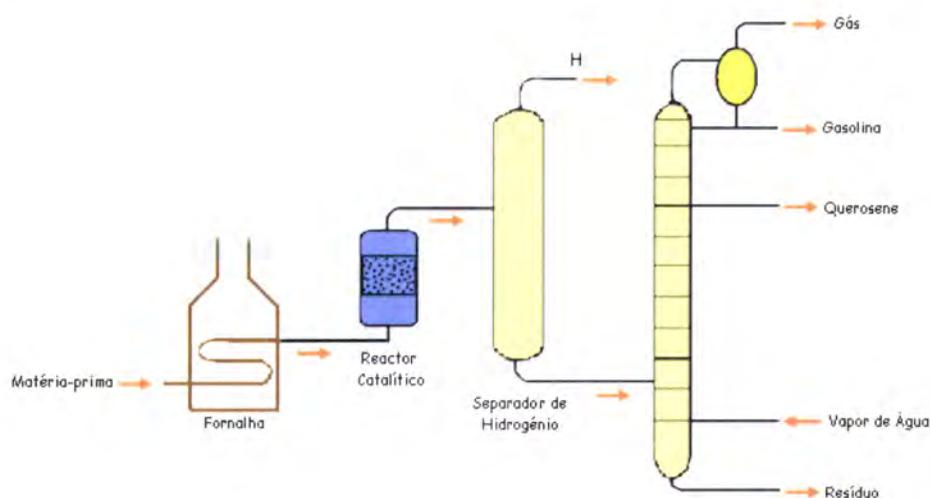


Figura II. 5- Esquema do Processo de Reformação. Adaptado de [Coelho & Costa, 2007]

A hidrogenação (Figura II.6) permite aumentar o teor em hidrogénio de um combustível. No processo representado na Figura II.6 são retirados, cataliticamente, compostos de azoto ao combustível antes do processo de hidrogenação. Este tipo de procedimento é adequado para o tratamento de xistos betuminosos e de areias asfálticas [Meyers, 1996] [Gary & Handwerk, 2001] [Gomes & Alves, 2007] [Coelho & Costa, 2007].

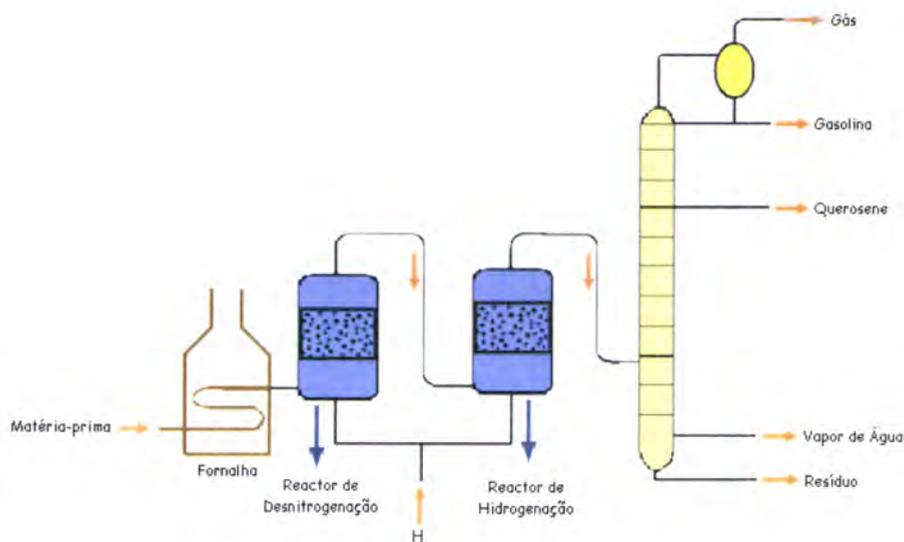


Figura II. 6- Esquema do Processo de Hidrogenação. Adaptado de [Coelho & Costa, 2007]

É importante reforçar que a separação é realizada em termos de fracções de interesse comercial. Na realidade, a destilação do petróleo não é mais do que uma simples separação física baseada nos diferentes pontos de ebulição dos seus componentes.

A Figura II.7 sintetiza a refinação, os processos de transformação e os derivados do petróleo. Na Tabela II.6 apresentam-se os combustíveis finais típicos obtidos a partir da refinação do petróleo bruto e as suas percentagens mássicas. Contudo, tal como foi referido, o tipo de produto final depende do interesse do mercado, sendo possível regular os processos e os pontos de saída dos vários produtos de modo a satisfazer as necessidades.

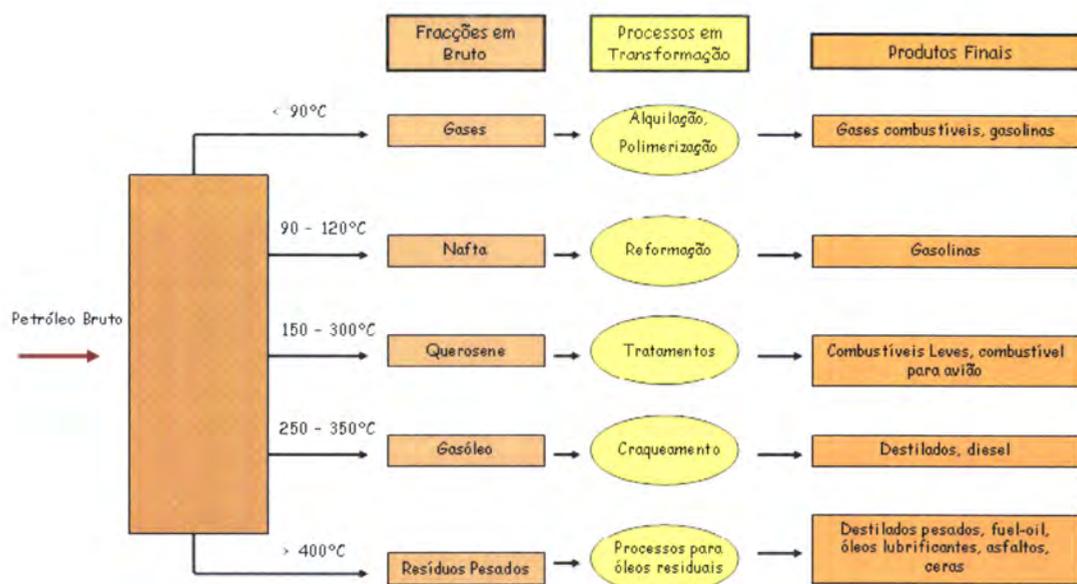


Figura II. 7- Refinação, Transformação e Derivados do Petróleo. Adaptado de [Gomes & Alves, 2007]

Tabela II. 6- Produtos Finais Típicos da Refinação do Petróleo Bruto. Fonte [Gomes & Alves, 2007]

COMBUSTÍVEL	PERCENTAGEM MÁSSICA
GPL	3,7
Gases de Refinaria	2,9
Nafta	1,3
Gasolina	38,9
Gás de Ar	0,2
Combustível para Aviação	5,6
Querosene	0,8
Gasóleo	18,2
Óleo Lubrificante	18,2
Fuel Óleos	16,6
Outros	11,8

II.7. Conclusões

A formação das acumulações de petróleo, ao longo de milhões de anos, depende da conjugação de vários factores, os quais, foram apresentados neste Capítulo.

No presente Capítulo também se abordou a composição química do petróleo, mais precisamente as grandes séries ou famílias de compostos que o constituem, os tipos de petróleo bruto bem como alguns dos processos físicos e químicos em que consiste a refinação do petróleo.

Capítulo III – Aspectos Ambientais na Indústria Petrolífera

Medidas de adequação ambiental na indústria petrolífera; tipos de emissões; efluentes; desperdícios e respectivos impactes. Descrição dos efluentes na Refinaria de Sines e tipos de tratamento efectuados.

Do ponto de vista ambiental, a actividade da Indústria Petrolífera não tem sido bem vista, devido às emissões gasosas de dióxido de carbono provenientes da queima para a produção de energia. O Protocolo de Quioto pretendeu criar limites e condições para investimentos dirigidos para o desenvolvimento de tecnologias destinada à redução das emissões, sem pôr em causa a sustentabilidade do fornecimento de energia [Gary & Handwerk, 2001] [Gomes & Alves, 2007].

Nas últimas décadas têm-se dado alguns passos importantes na redução das emissões. Das várias medidas tomadas destacam-se a proibição total, por praticamente todos os países, da queima do gás associado ao petróleo nos projectos de desenvolvimento petrolífero e a obrigação da utilização do gás, especialmente em projectos de produção de energia e na indústria petroquímica.

A preservação do Ambiente que nos rodeia é cada vez mais uma preocupação constante da Humanidade. É nesse sentido que os estudos de impacte ambiental (*Environmental Impact Assessment* - EIA), subjacentes a qualquer projecto petrolífero, merecem cada vez mais a atenção das companhias petrolíferas, dos governos e do público em geral.

Hoje em dia esses estudos não são feitos apenas para cumprir obrigações legais, mas também porque fazem sentido no contexto de uma política ambiental a nível global. Para além disso, podem trazer vantagens económicas para as partes e promover um espírito de confiança entre governos, populações e companhias. Os critérios de rigor nos EIA variam de país para país, de região para região, consoante as condições geográficas.

São considerados resíduos das refinarias as emissões, os efluentes, bem como desperdícios vários. Na Tabela III.1, enumeram-se alguns exemplos e referem-se os principais impactes sobre as populações, sobre os ecossistemas e sobre o ambiente [Nemerov, 1977].

Tabela III. 1 - Impactes dos Resíduos Provenientes das Refinarias

EMISSIONES	IMPACTE DAS EMISSIONES
<ul style="list-style-type: none"> - Gás associado à produção do petróleo; - Emissão de CH₄, H₂S e VOC (<i>Volatile Organic Compounds</i>); - Queima de gás, com emissão de CO₂, CO, SO₂, CH₄, VOC e H₂O; - Emissões provenientes de turbinas e geradores (CO, CO₂, NO_x, VOC e H₂O); e - Gases de escape ou fugas nos equipamentos. 	<ul style="list-style-type: none"> - Dão origem a chuvas ácidas, que danificam construções e prejudicam o Ambiente; - Afectam a saúde do ser humano (problemas pulmonares, entre outros); - Algumas partículas podem causar problemas respiratórios; - Afectam o habitat ecológico da zona; e - Degradam a qualidade do ar.
EFLUENTES	IMPACTE DOS EFLUENTES
<ul style="list-style-type: none"> - Água de produção; - Fugas químicas; e - Óleos de lubrificação usados. 	<ul style="list-style-type: none"> - Produtos químicos tóxicos que podem causar queimaduras na pele; - Degradação da qualidade da água do mar e do subsolo (aquíferos), bem como do próprio solo arável; - Afectam o habitat da zona; e - Cheiros desagradáveis.
DESPERDÍCIOS	IMPACTE DOS DESPERDÍCIOS
<ul style="list-style-type: none"> - Amostras de sondagem (<i>drilling cuttings</i>); - Lama de sondagem; - Baterias e filtros de óleo; - Depósitos do fundo dos tanques; - Desperdícios metálicos; - Tambores; e - Garrafas, papel, cartões e sacos. 	<ul style="list-style-type: none"> - Degradação da qualidade do solo; - Atraem insectos, roedores ou outros animais indesejáveis; e - Cheiros desagradáveis.

III.1. Sistemas de Protecção Ambiental na Refinaria de Sines

A subunidade de Ambiente da Refinaria de Sines da Galp Energia zela pelo controlo dos parâmetros de qualidade ambiental, promovendo a minimização dos quantitativos de resíduos sólidos, efluentes líquidos e emissões gasosas. Para tal, está equipada com os mais modernos sistemas de protecção ambiental tais como:

- Sistema de monitorização em contínuo de efluentes gasosos, abrangendo a maior parte da instalação;
- Sistema de redução de emissões de partículas provenientes do regenerador do catalisador do FCC (*Fluid Catalytic Cracking*);
- Sistema de controlo analítico dos efluentes, salino industrial e águas pluviais;
- Sistema de monitorização de solos e águas subterrâneas, por intermédio de uma rede de piezómetros; e
- Equipamento dinâmico de baixo nível de ruído.

Possui vectores de orientação, que acompanham a legislação nacional e comunitária, com o objectivo de promover esforços de melhoria progressiva e constante para a redução do impacto ambiental.

A Refinaria para além de proteger o domínio hídrico, tal como já foi referido, também tem preocupação no domínio dos resíduos sólidos, através da redução e tratamento de lamas, as quais são depositadas num parque de resíduos sólidos e posteriormente tratadas.

Em relação à protecção do ar, a maioria dos efluentes gasosos são emitidos através de uma chaminé principal com 234 metros de altura, considerada adequada para uma boa dispersão dos fumos na atmosfera. Os valores são monitorizados continuamente e as caldeiras e fornalhas processuais têm instalados queimadores de baixo NO_x. A Refinaria conta também com um sistema para redução de partículas, na unidade de FCC.

III.2. Descrição dos Efluentes da Refinaria de Sines

Os efluentes da Refinaria de Sines subdividem-se em oleosos, acidentalmente contaminados, águas limpas (pluviais), sanitários, químicos e salinos. Na Tabela III.2 descreve-se cada um dos tipos de efluentes.

Tabela III. 2- Diferentes Tipos de Efluentes. Fonte: [Fernandes, 2007]

TIPO DE EFLUENTE	DESCRIÇÃO
OLEOSOS	A rede de efluentes oleosos é composta por um conjunto de separadores gravíticos, espalhados ao longo de toda a Refinaria, que recebem efluentes oleosos de várias proveniências (armazenagem e processo). Os óleos recolhidos nos separadores são enviados para <i>slop's</i> , enquanto o efluente é enviado para uma Bacia de Águas Contaminadas (BAC), cujo tempo de retenção é, aproximadamente, um dia. Nesta bacia é efectuada a última separação, sendo posteriormente o efluente enviado directamente para a ETAR de Ribeira de Moinhos.
ACIDENTALMENTE CONTAMINADOS	Os efluentes provenientes da fabricação considerados acidentalmente contaminados são enviados para um separador gravítico e, posteriormente, encaminhados para a BAC.
ÁGUAS LIMPAS (PLUVIAIS)	As águas das chuvas são recolhidas numa série de caixas que enviam essa água para uma Bacia de Águas Limpas. Parte dessa água é aproveitada na Bacia de Sistemas de Incêndio e a restante é enviada para a ETAR de Ribeira dos Moinhos.
SANITÁRIOS	As águas provenientes dos sanitários existentes na Refinaria são encaminhadas para um separador gravítico e posteriormente para a Bacia de Águas Contaminadas.
QUÍMICOS	As águas provenientes do Laboratório são enviadas para a BAC após a passagem por um separador gravítico.
SALINOS	Os efluentes salinos provenientes da unidade de desmineralização são enviados para um separador gravítico e posteriormente para o mar através de um emissário submarino. O separador gravítico está equipado com sistema de medição em contínuo de pH.

III.3. Tratamento de Efluentes Industriais

O tratamento das águas residuais é hoje encarado como uma necessidade urgente e primária para que se possa reutilizar a água sem prejudicar a saúde pública e ambiental, de modo a minimizar o impacto das fontes poluidoras nas linhas de água e ambiente envolvente.

Os efluentes industriais têm grande impacto sobre as ETARs uma vez que não são uniformes. Variam com o tipo e dimensão da indústria e são influenciados pela existência ou não de sistemas de tratamento pertencentes às próprias indústrias. Os referidos efluentes estão sujeitos a tarifação, a qual varia em função dos índices de poluição e qualidade dos efluentes (Figura III.1).

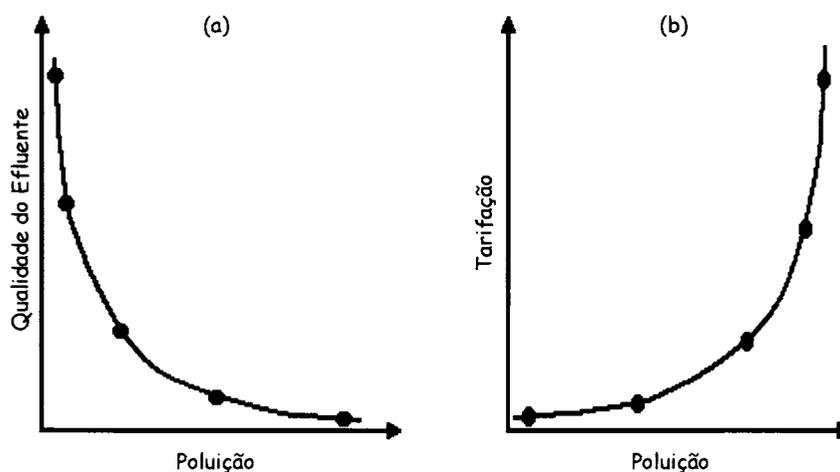


Figura III. 1 - Qualidade dos Efluentes Industriais vs Poluição (a). Tarifação aplicada vs Poluição (b)

A Refinaria de Sines realiza um pré-tratamento dos efluentes antes da sua descarga no sistema público de drenagem. Esse pré-tratamento ocorre dentro das instalações e é suportado pela empresa.

O referido pré-tratamento é composto por quatro fases designadas por tratamento preliminar, tratamento primário, tratamento secundário e tratamento terciário que se descrevem em seguida.

III.3.1. Tratamento Preliminar

Durante o processo de produção do petróleo é comum o aparecimento de gás e água associados ao *crude*. A separação dessas fases torna-se necessária uma vez que o gás apresenta relevante interesse económico. A água, por seu lado, deve ser removida pois afecta o dimensionamento do sistema de bombagem e transferência, compromete certas operações de processo nas refinarias, além de provocar problemas de incrustação e corrosão nos oleodutos de exportação [Fernandes, 2007].

A escolha do processo de desidratação depende fundamentalmente da quantidade da água que é necessário processar e das especificações do petróleo a exportar, em termos da percentagem máxima de água que pode conter. Em terra, onde o espaço não é um factor limitativo, a forma mais simples de desidratar o petróleo é usar tanques de desidratação. Com este sistema, o *crude* entra no tanque através de um dispersor onde gotículas de água saem do petróleo à medida que o nível da mistura sobe, permitindo reduzir o teor em água para valores abaixo dos 2% [Fernandes, 2007].

Para a desidratação dos crudes muito viscosos é necessário usar, para além dos tanques de desidratação, um sistema de aquecimento em que a temperatura é função da viscosidade do petróleo [Fernandes, 2007].

No caso da percentagem de água ser muito elevada utilizam-se tanques de lavagem concêntricos. Estes, são tanques de grandes dimensões equipados com dupla parede que obrigam a mistura a circular, aumentando assim o seu tempo de residência e permitindo a separação de grandes quantidades [Fernandes, 2007].

No caso do petróleo e a água aparecerem em forma de emulsão, a desidratação será muito mais complexa. Assim, utilizam-se produtos químicos para quebrar a emulsão e, muitas vezes, é necessário recorrer a em combinação com sistemas de aquecimento e de agitação suave [Fernandes, 2007].

A água separada do *crude* é geralmente devolvida ao oceano depois de tratada. Contudo, antes de se eliminar a água, é necessário baixar a percentagem de hidrocarbonetos presentes, para que a mesma não venha a poluir o ambiente e não sejam aplicadas à Refinaria tarifas resultantes do não cumprimento dos intervalos legalmente permitidos.

III.3.2. Tratamento Primário

Na Refinaria de Sines os efluentes oleosos resultantes dos tanques de desidratação anteriormente mencionados, são encaminhados para separadores gravíticos, denominados OP-V's, onde ocorre a primeira separação dos óleos e gorduras (Figura III.2).



Figura III. 2- Separadores Gravíticos. Fonte [Galp, 2007b]

Em todos os separadores, os óleos que se encontram à superfície são recolhidos por uma escumadeira de nível ajustável, depositados numa caixa de óleos e posteriormente reaproveitados para tanques de slops, sendo as águas encaminhadas para a flotação. A flotação é um processo de separação sólido-líquido utilizado em alternativa ao processo clássico de sedimentação e que consiste na introdução de um gás que agregará as partículas sólidas formando-se um agregado que flutua por ser menos denso que a água (Figura III.3) [Cheremisinoff, 2002].



Figura III. 3- Flotação. Fonte [Galp, 2007b]

Nesta fase, os óleos e gorduras presentes no efluente incorporam todo o ar introduzido em forma de micro-bolhas que os conduzem à superfície. A almofada de ar originada pelo sistema obriga a que todo o flotado se mantenha permanentemente à superfície até que seja removido pela pá raspadora.

Após a flotação, as águas são recolhidas num separador gravítico e daí são bombeadas para a BAC. Os óleos são recolhidos num separador gravítico e posteriormente bombeados para tanques de slops. Na BAC está implementado o sistema para a oxidação de sulfuretos e fenóis.

III.3.3. Tratamento Secundário

O tratamento secundário ocorre na BAC (Figura III.4) e consiste num conjunto de procedimentos em que se incluem processos biológicos, nos quais se utilizam bactérias aeróbias para remover cerca de 90% da matéria orgânica e diminuir a Carência Química de Oxigénio (CQO) do efluente. Este tipo de tratamento tem também como consequência a oxidação de compostos que se encontram na forma reduzida como, por exemplo, o azoto amoniacal e a redução dos nitratos.



Figura III. 4- Bacia de Águas Contaminadas. Fonte [Galp, 2007b]

Este sistema de tratamento foi projectado para reduzir a concentração de sulfuretos na água para valores inferiores a 20 mg/dm^3 e a de fenóis para valores inferiores a 40 ppm.

O tratamento secundário, como foi referido anteriormente, utiliza bactérias aeróbias. Contudo, como o enxofre presente no efluente se apresenta na forma reduzida é absolutamente necessário oxidá-lo antes de iniciar o tratamento biológico [Cheremisinoff, 2002]. Para o efeito, é promovida a recirculação do efluente a uma taxa constante de 600 m³/h utilizando uma electrobomba. A aspiração ocorre à saída da BAC e o retorno é feito através de tubagens e hidrojectores colocados no fundo da bacia. O sistema de recirculação permite a sucção e a dissolução no efluente de uma grande quantidade de ar, conseguindo-se, deste modo, a oxidação.

A actividade microbiológica promovida na BAC é a responsável pela diminuição da CQO, dos óleos, dos hidrocarbonetos, dos compostos fenólicos e até de uma parte de sulfuretos e outros compostos de enxofre [Fernandes, 2007].

A actividade microbiológica é significativamente aumentada pela presença, em suspensão, de carvão activado de porosidade adequada. O carvão actua também como protecção das bactérias face a subidas bruscas de concentração de algumas espécies que podem diminuir ou eliminar as colónias existentes na BAC [Fernandes, 2007].

O crescimento microbiano depende de diversos factores, salientando-se o pH, a temperatura, os nutrientes e a concentração de oxigénio dissolvido.

Na BAC estão instalados sistemas de introdução de oxigénio e dióxido de carbono, controlados independentemente por uma sonda de oxigénio dissolvido e por uma sonda de pH. Estão ainda instalados dois sistemas automáticos de recolha de matéria flotada os quais aspiram e encaminham o flotado para uma caixa de óleos, para posterior reutilização.

O sistema de arejamento tem dupla finalidade. Por um lado, fornece oxigénio suficiente para as necessidades dos microrganismos aeróbios e, por outro, provoca uma agitação e uma homogeneização suficiente para que no reactor ocorra uma mistura completa entre as lamas e a água residual a tratar.

A adição de oxigénio também é importante como meio de remoção de alguns poluentes por oxidação, eliminando compostos orgânicos que resistem aos processos biológicos.

A dose de dióxido de carbono é função do pH da mistura que deve ser mantido entre 7 e 8, intervalo no qual se situa o pH óptimo para o crescimento microbiano [Fernandes, 2007].

Um outro aspecto importante para a actividade microbiana prende-se, como foi referido anteriormente, com a presença de nutrientes nas lamas activadas. Os nutrientes

requeridos em maiores quantidades são o azoto e o fósforo que, normalmente, não são muito abundantes neste tipo de efluente. Para se obter uma boa estabilização das lamas e uma boa floculação bacteriana a relação CQO: N: P deve ser de aproximadamente 100:5:1. Para tal torna-se necessário adicionar ácido fosfórico.

A temperatura é outro factor importante uma vez que interfere tanto na taxa de crescimento dos microrganismos como na velocidade de remoção dos substratos. Com o aumento da temperatura ocorre um aumento da actividade biológica até um certo valor óptimo. A temperaturas mais elevadas ocorre a desnaturação térmica dos enzimas e, conseqüentemente, uma diminuição da actividade microbiológica. Os microrganismos presentes em processos de lamas activadas são, geralmente, mesofílicos cuja temperatura óptima está situada entre 20 e 40°C [Fernandes, 2007].

Após a conclusão do tratamento secundário os efluentes passam, por gravidade, para a Bacia de Pré-Tratamento (BPT) e depois para o colector de efluentes que os encaminha para a ETAR de Ribeira de Moinhos.

III.3.4. Tratamento Terciário

Antes da água ser descarregada nas linhas de água, as águas residuais sofrem um tratamento terciário que consiste na desinfecção, controlo de nutrientes e eliminação de bactérias. O tipo de tratamento utilizado depende da contaminação que se pretende remover. Entre as opções de desinfecção aplicadas nestas circunstâncias contam-se geralmente três tecnologias básicas i.e. a cloragem, a ozonização e o canal de ultravioletas.

A desinfecção por ozono é um processo mais dispendioso que a cloragem, e menos vulgar, mas apresenta menos desvantagens uma vez que não permanece muito tempo na água não formando subprodutos [Fernandes, 2007].

III.4. Conclusões

Com a importância crescente da Indústria Petrolífera no panorama energético mundial, foi fundamental atacar o problema ambiental de uma forma sistemática e transformar esta Indústria num contribuinte efectivo para o desenvolvimento sustentável do Mundo em que vivemos.

Neste Capítulo foram abordados os aspectos ambientais a ter em conta na Indústria Petrolífera tendo sido realizada uma descrição dos efluentes gerados na Refinaria de Sines e os tipos de tratamento efectuados.

Capítulo IV – Descoberta de Conhecimento em Bases de Dados e Data Mining

Objectivos, conceitos, tipos de abordagem, áreas relacionadas, metodologias e especificações da Descoberta de Conhecimento em Bases de Dados.

A capacidade de aprender através de exemplos e de generalizar a informação aprendida é fundamental na solução dos problemas de uma sociedade, onde a informação tem assumido um papel cada vez mais relevante. Torna-se, pois vital analisar os dados, extrair deles conhecimento, generalizar, sublinhar princípios e estruturar a informação. Estas características revelam-se decisivas em unidades industriais, onde a rapidez de resolução de problemas está fortemente correlacionada com a capacidade de seleccionar a informação certa para um problema específico [Turban et al., 2004].

A análise de dados não é um tema novo, tem sido feita desde há vários anos recorrendo, principalmente, a métodos estatísticos. Contudo, desde muito cedo que ficou claro que o cérebro humano analisa dados e trata a informação de um modo diferente. Na verdade, a aquisição de conhecimento nos seres humanos é feita através do processo de aprendizagem. Numa perspectiva sistémica, a aprendizagem pode ser definida como as alterações do sistema, que lhe permitem refazer as mesmas tarefas de uma forma mais eficaz e eficiente no futuro [Adriaans & Zantinge, 1996] [Russell & Norvig, 2003].

A constatação dos factos anteriormente expostos abriu caminho para o desenvolvimento de Sistemas Inteligentes (SI), os quais emanam da Área Científica da Inteligência Artificial (IA). O objectivo da IA passa pelo desenvolvimento de paradigmas e algoritmos para tentar realizar tarefas cognitivas as quais são executadas eficientemente pelos seres humanos. Apesar de ser recente, tem proporcionado desenvolvimentos relevantes em diversas áreas do conhecimento como, por exemplo, na compreensão da linguagem natural, no raciocínio lógico matemático, na teoria dos jogos ou nos sistemas de apoio à decisão [Russell & Norvig, 2003].

De acordo com Haykin [Haykin, 1999] um sistema IA deve preencher três requisitos, i.e. armazenar conhecimento, aplicar o conhecimento adquirido para resolver problemas e adquirir novo conhecimento através da experiência.

O primeiro requisito está relacionado com a representação do conhecimento que provém, fundamentalmente, de dois tipos de informação, o estado conhecido (representado por factos sobre o que é e o que se conhece) e observações ou medições sobre o ambiente (este tipo de informação pode conter ruído e está sujeito a erro).

O segundo requisito, a aplicação do conhecimento adquirido, diz respeito ao raciocínio, definido como a capacidade para a resolução de problemas, que pode ser encarada como um problema de procura.

O último requisito envolve a aprendizagem que permite aumentar o conhecimento através da experiência. A aprendizagem pode envolver duas fases distintas de processamento da informação - a forma indutiva e a forma dedutiva. Na primeira tem-se a aprendizagem baseada em casos, ou seja, determinam-se padrões e regras a partir dos dados e das experiências. Na segunda forma, as regras são utilizadas para criar novos factos.

No âmbito da aprendizagem um aspecto importante a realçar prende-se com a relação entre o modelo de aprendizagem e o seu ambiente. Neste contexto existem, fundamentalmente, três paradigmas em que se baseia a aprendizagem [Patterson, 1996].

- i) **Aprendizagem Supervisionada** - É uma técnica bastante divulgada, na qual são fornecidas as respostas correctas ao sistema. A aprendizagem é levada a cabo a partir de um conjunto de casos de treino ou exemplos onde cada um deles é formado por um vector de entrada e um vector de saída ou de resposta;
- ii) **Aprendizagem de Reforço** - Nesta técnica não se apresenta a resposta correcta ao sistema. Apenas se dá uma indicação sobre se a resposta apresentada pelo sistema é correcta ou não, sendo esta informação utilizada no melhoramento da sua eficácia; e
- iii) **Aprendizagem Não Supervisionada** - Neste caso não chega ao sistema qualquer indicação externa sobre a resposta correcta. A aprendizagem é feita a partir de descobertas de características nos dados de entrada, adaptando-se a regularidades estatísticas ou agrupamentos de padrões dos exemplos de treino.

IV.1. Descoberta de Conhecimento em Bases de Dados

O desenvolvimento tecnológico tem proporcionado, no que concerne ao armazenamento de dados, um crescimento exponencial, tanto em número de registos como em complexidade. Na verdade, as organizações acumulam informação, tendo como base o pressuposto de que grandes volumes de dados podem ser fonte de conhecimento, e que este pode ser extraído com ajuda de técnicas apropriadas. A existência desta mais valia pode contribuir para melhorar o comportamento das organizações, descobrindo tendências e especificidades, tornando mais célere e eficaz a acção dos gestores.

Como consequência deste aumento efectivo de informação, o seu processamento através dos métodos tradicionais tornou-se cada vez mais difícil e complexo. Deste modo, começaram a surgir aplicações vocacionadas para a tarefa de Descoberta de Conhecimento em Bases de Dados incorporando ferramentas de *Data Mining* [Michalski et al., 1998] [Goebel & Gruenwald, 1999] [Bramer, 2007] [Cios et al., 2007] [Taniar, 2008].

O termo Descoberta de Conhecimento em Bases de Dados (DCBD) foi formalizado em 1989 como uma referência ao conceito mais amplo de procura de conhecimento em dados sendo um processo que envolve a identificação e o reconhecimento de padrões ou tendências, numa Base de Dados (BD), de uma forma autónoma e porventura automática [Fayyad et al., 1996] [Thuraisingham, 1999] [Apte et al., 2002].

No início de qualquer projecto de DCBD é necessário definir o domínio sobre o qual se pretende executar o estudo, os objectivos a atingir e proceder à recolha dos dados e/ou variáveis necessárias à consecução dos objectivos, i.e. a construção da BD sobre a qual se irá actuar. Um outro aspecto importante a atender, prende-se com o estudo do domínio de aplicação. O conhecimento do domínio constitui um recurso essencial no desenvolvimento de qualquer projecto de DCBD, sendo o seu fio condutor.

O processo de DCBD depende de uma nova geração de ferramentas e técnicas de análise de dados que, como se pode observar na Figura IV.1, envolve diversas etapas [Adriaans & Zantinge, 1996] [Fayyad et al., 1996] [Michalski et al., 1998] [Bramer, 2007] [Cios et al., 2007] [Taniar, 2008]. As referidas etapas encontram-se sumariadas na Tabela IV.1, sendo explicitadas e exemplificadas no Capítulo VI.

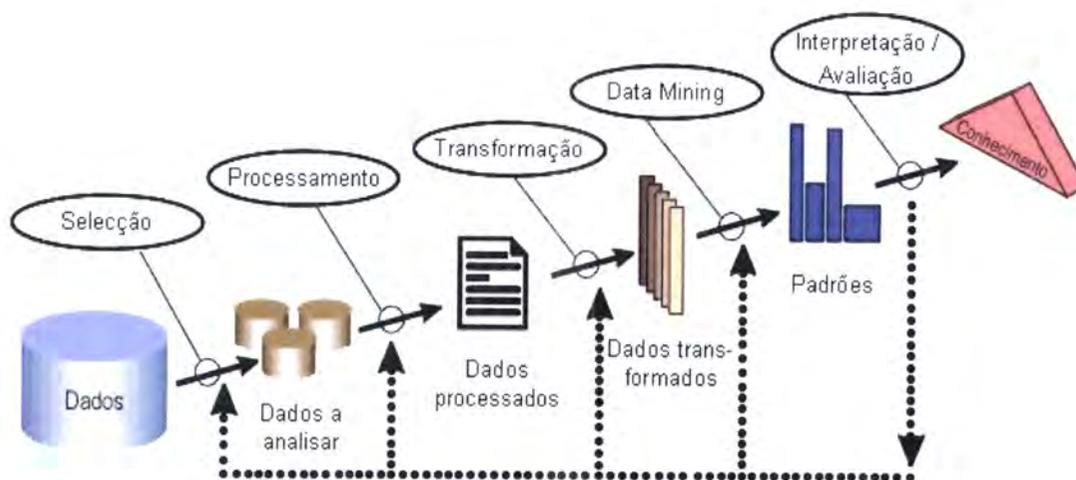


Figura IV. 1- Processo de Descoberta de Conhecimento em Base de Dados. Adaptada de [Sousa, 2004]

Tabela IV. 1- Etapas do Processo de DCBD

ETAPA DO PROCESSO DE DCBD	DESCRIÇÃO
SELECÇÃO	Consiste na segmentação dos dados de acordo com critérios que determinam subconjuntos.
PRÉ-PROCESSAMENTO	Comporta a limpeza dos dados (data cleaning) de forma a adequá-los aos algoritmos que, posteriormente, serão utilizados. Inclui, ainda a integração de dados heterogéneos e a eliminação de dados incompletos e inconsistentes. São estabelecidas estratégias para eliminar ou minimizar os problemas referentes à omissão de dados e à presença de ruídos e erros.
TRANSFORMAÇÃO	Visa armazenar os dados num formato que facilite a sua utilização por parte das ferramentas de <i>Data Mining</i> .
DATA MINING	Envolve a selecção dos métodos e técnicas a serem aplicados. A escolha depende dos objectivos do processo de <i>Data Mining</i> , podendo ser aplicadas técnicas distintas ao problema.
INTERPRETAÇÃO	Pretende-se construir uma representação do conhecimento que seja acessível e clara. Os padrões identificados pelo sistema são interpretados como conhecimento, o qual pode ser usado para o suporte à tomada de decisão. Este conhecimento traduz-se, por exemplo, em entendimento dos conteúdos da BD ou na explicação de um fenómeno observado.

Antes de se poder dar o processo de DCBD como concluído deve fazer-se a verificação da qualidade da informação obtida. Esta verificação pode ser encarada como uma fase de Avaliação do Desempenho do mecanismo de aprendizagem. A determinação dos valores de Confiança e Suporte são dados extremamente importantes para a validação dos modelos.

A Figura IV.1, que esquematiza a DCBD, pode sugerir a existência de um percurso linear, no entanto, isso nem sempre se verifica. Em cada etapa pode ser identificada a necessidade de retorno para cada uma das fases anteriores. Por exemplo, se na fase *Data Mining* se conclui que os dados não são plenamente coerentes, ou for verificada a necessidade da inclusão de um novo dado, isso pode levar ao retorno para fases anteriores. De acordo com Adriaans e Zantinge [Adriaans & Zantinge, 1996], nos projectos de DCBD 80% do tempo é dispendido nas etapas de preparação dos dados, enquanto que os restantes 20% referem-se à etapa de *Data Mining* propriamente dita.

IV.2. Data Mining

Encontram-se várias definições para o termo *Data Mining*, consoante a área de especialização, o autor e a abordagem [Fayyad et al., 1996] [Fayyad, 1997] [Chang et al., 2001] [Cortez, 2004] [Bramer, 2007] [Cios et al., 2007] [Taniar, 2008]. Contudo, existem características comuns que se podem sintetizar na seguinte definição: aplicação de métodos e técnicas em grandes Bases de Dados, para encontrar tendências ou padrões com o intuito de descobrir conhecimento.

Como foi anteriormente referido, a etapa de *Data Mining* é apenas uma etapa do processo de Descoberta de Conhecimento em Bases de Dados. Esta etapa, como foi referido anteriormente, consiste na escolha e utilização dos métodos e técnicas que melhor se enquadram no cumprimento dos objectivos estabelecidos. Na Tabela IV.2 apresentam-se e caracterizam-se alguns desses objectivos, divididos em duas tarefas (previsão e descrição). No caso da previsão pretende-se inferir comportamentos futuros com base nas experiências do passado, ao passo que na descrição pretende-se, de um modo geral, descrever de forma compacta conjuntos de dados ou associações variáveis.

Tabela IV. 2- Objectivos de Data Mining. Adaptada de [Cortez, 2004]

OBJECTIVO		DESCRIÇÃO
PREVISÃO	Classificação	Corresponde à descoberta de uma função que associa um caso a uma das diversas classes discretas de classificação, de forma a classificar um novo objecto de acordo com um padrão de classificação.
	Regressão ou Previsão	Consiste em prever valores futuros ou desconhecidos de uma variável dependente, a partir de exemplos. A ideia subjacente corresponde à concepção de um modelo capaz de mimetizar uma função desconhecida que se aproxime da função dada por um conjunto de vectores etiquetados.
DESCRIÇÃO	Associação ou Dependência	Pretende-se encontrar um modelo que descreva dependências significativas entre variáveis, através da identificação de grupos de dados fortemente associados. As associações surgem quando várias ocorrências estão ligadas num único evento, podendo surgir a nível estrutural (i.e. o modelo é representado de uma forma gráfica e com variáveis localmente dependentes em relação a outras) ou quantitativo (i.e., o modelo especifica o peso das dependências segundo uma escala numérica).
	Sumariação	Os métodos de sumariação derivam de regras de resumo e descobertas de relações funcionais entre variáveis.
	Segmentação	Permite identificar um conjunto finito de categorias ou segmentos para descrever os dados (e.g. identificação de grupos homogéneos de objectos). Pode ser um objectivo intermédio de DM, sendo realizado numa fase inicial para encontrar grupos homogéneos de dados para posterior aplicação de algoritmos de aprendizagem.
	Visualização	Trata da apresentação dos resultados de DM através de uma forma visual, geralmente através de gráficos ou diagramas. Pretende-se dar a conhecer informações complexas de uma forma acessível, evidenciando os padrões e tendências reveladas.

IV.3. Modelos e Técnicas de Data Mining

Por modelo entende-se uma representação matemática de um sistema (ou processo) com o objectivo de o estudar. Corresponde à descrição formal e simplificada de um sistema, ou seja, é a estruturação da representação do conhecimento com vista a atingir certo resultado. Os modelos podem ser estáticos, quando representam o sistema num determinado instante ou podem ser dinâmicos quando o representam ao longo do tempo, acompanhando as alterações inerentes ao seu funcionamento [Cortez, 2004] [Teixeira, 2005]. Um modelo é definido como uma função (mapa) que atribui a cada exemplo, no domínio definido pelos atributos de entrada, um valor contido no domínio dos atributos de saída, havendo um conjunto de parâmetros que têm de ser ajustados (ou estimados) a partir de um conjunto de dados, através de um algoritmo, na fase de aprendizagem. Após a aprendizagem, é possível extrapolar novas saídas, alimentando o modelo com novas entradas (utilização de um modelo) [Cortez, 2004] [Teixeira, 2005].

Na construção de um modelo definem-se as principais características do sistema, que devem representar o mais fielmente possível a realidade. Recolhem-se os dados necessários para a sua construção e para a sua validação. Torna-se, portanto, necessária uma divisão da BD em dois subconjuntos, mutuamente exclusivos, um para gerar o modelo, designado como **conjunto de treino**, e outro para a sua validação, denominado **conjunto de teste**. A cardinalidade dos referidos subconjuntos é variável e dependente, entre outros factores, da dimensão da base de dados.

Deve realçar-se que não existe um modelo universal de DM que resolva, de forma eficiente, todos os problemas. A escolha de um determinado algoritmo é de certa forma uma arte, uma vez que existem diferentes modelos para as mesmas tarefas com vantagens e desvantagens intrínsecas [Fayyad et al., 1996]. Na Tabela IV.3 é apresentado um mapeamento entre as diversas tarefas e as técnicas de DM que são normalmente utilizadas.

Como foi referido anteriormente, os modelos e as técnicas utilizadas na etapa de *Data Mining* dependem da natureza do problema e dos objectivos que se pretendem atingir. No caso do presente trabalho, as técnicas utilizadas foram as Árvores de Decisão e a Indução de Regras na tarefa de classificação.

Tabela IV. 3- Tarefas e Técnicas Usadas no Processo de Data Mining. Adaptada de [Cortez, 2004]

TAREFA TÉCNICA	PREVISÃO		DESCRIÇÃO			
	CLASSIFICAÇÃO	REGRESSÃO	ASSOCIAÇÃO	SUMARIAÇÃO	SEGMENTAÇÃO	VISUALIZAÇÃO
ÁRVORES DE DECISÃO	✓	✓		✓	✓	✓
INDUÇÃO DE REGRAS	✓	✓	✓		✓	
REDES NEURONAIS ARTIFICIAIS	✓	✓	✓		✓	
ALGORITMOS GENÉTICOS	✓	✓		✓	✓	
APROXIMAÇÃO DE VIZINHANÇAS				✓	✓	

IV.3.1. Árvores de Decisão

As Árvores de Decisão tiveram a sua origem na área da Aprendizagem Automática (*Automatic Interaction Detection*). Esta análise testa automaticamente todos os valores de um determinado atributo de forma a identificar aqueles que têm uma forte associação com os registos de saída.

Uma árvore de decisão é uma representação de um conjunto de regras, que classificam as instâncias desde o nó raiz até aos nós terminais ou folhas. Cada nó da árvore especifica um teste para os atributos da instância e, cada ramo descendente desse nó, correspondente a um dos valores possíveis para esse atributo. Uma instância é classificada começando por testar o atributo especificado pelo nó raiz e, depois, seguindo o ramo correspondente até ao nó terminal. O primeiro nó da árvore de decisão mostra o factor mais correlacionado com os objectos de saída. Os restantes factores são subsequentemente classificados como nós e relacionados com os nós anteriores, possibili-

tando uma visualização fácil e rápida dos factores que mais se relacionam com os objectos de saída [Quinlan, 1987] [Kufirin, 1997] [Han & Kamber, 2006] [Cios et al., 2007].

Existem dois tipos de árvores de decisão, i.e. árvores de classificação e árvores de regressão. As primeiras, aplicam-se a problemas de classificação, tendo como objectivo qualificar os registos e associá-los com uma determinada classe. As segundas, por seu lado, aplicam-se a problemas de regressão, visando estimar o valor de uma determinada variável [Quinlan, 1987] [Berry & Linoff, 2000]. Estes dois tipos de árvores apresentam a mesma estrutura. Quando os dados são apresentados ao modelo, cada registo é sujeito a uma série de testes até alcançar uma folha ou nó puro da árvore, onde lhe é atribuída uma classe ou um valor numérico consoante se tratem, respectivamente, de árvores de classificação ou de árvores de regressão [Berry & Linoff, 2000].

Na construção de uma árvore de decisão levantam-se dois problemas. O primeiro relaciona-se com o atributo a seleccionar para teste num determinado nó e o segundo está ligado à paragem da divisão dos exemplos. No que diz respeito à escolha dos atributos, existem várias medidas para avaliar a capacidade de um dado atributo para discriminar as classes, no entanto todas convergem em dois pontos:

- Uma divisão que mantém as proporções de classes em todas as partições é inútil; e
- Uma divisão onde em cada partição todos os exemplos são da mesma classe tem utilidade máxima [Teixeira, 2005] [Sousa, 2004] [Han & Kamber, 2006].

No que concerne à paragem da divisão dos exemplos, esta deve acontecer quando todos os exemplos pertencem à mesma classe, ou todos os exemplos têm os mesmos valores dos atributos mas classes diferentes, ou o número de exemplos é inferior a um certo limite, ou o mérito de todos os possíveis testes de partição dos exemplos é muito baixo [Sousa, 2004] [Teixeira, 2005] [Han & Kamber, 2006].

Os algoritmos de indução de árvores de decisão constroem árvores a partir dos dados do conjunto de treino, de uma forma recursiva, subdividindo o referido conjunto até à obtenção de nós "puros" em que cada nó represente apenas uma única classe ou a satisfação de um critério. Utilizam a aprendizagem supervisionada, i.e. são fornecidas as respostas correctas ao sistema, a partir de um conjunto de exemplos onde cada um deles é formado por um vector de entrada e um vector de saída ou resposta [Rich & Knight, 1991].

O algoritmo ID3, Figura IV. 2, desenvolvido em 1983 por Ross Quinlan [Quinlan, 1986], foi um dos algoritmos pioneiros que permitiu o desenvolvimento das árvores de decisão.

Desde então têm sido introduzidas novas funcionalidades que resultaram no aparecimento de evoluções deste algoritmo (e.g. algoritmos C4.5 e C5.0).

<p>Seja T a árvore de decisão a induzir Seja S o conjunto de exemplos para aprendizagem Se todos os exemplos de S pertencem à mesma classe C então a árvore tem um só nodo C Senão seleccionar o atributo A mais “informativo” cujos valores são V_1, \dots, V_n Particionar S em n subconjuntos S_1, \dots, S_n um para cada valor V_i de A construir (recursivamente) sub - árvores T_1, \dots, T_n para cada S_1, \dots, S_n</p>

Figura IV. 2- Algoritmo de Indução de Árvores de Decisão. Adaptado de [Teixeira, 2005]

Uma das limitações exibidas pelo algoritmo ID3 prende-se com a possibilidade de ocorrer sobreajustamento, que se manifesta na construção de árvores com um desempenho quase perfeito para os dados de treino, mas com um baixo desempenho nos dados de teste. No algoritmo C4.5 foram introduzidos melhoramentos que permitem ultrapassar o referido problema através da introdução da poda da árvore. Designa-se poda da árvore a redução de algumas subárvores a folhas, tendo em conta os erros nesse nó e a soma dos erros nos nós que dele descendem. O algoritmo C4.5 permite, ainda, ultrapassar problemas concretos e comuns do mundo real relacionados com atributos com valores numéricos, valores omissos, dados com ruído. Para além do exposto, o algoritmo C4.5, possibilita o recurso à validação cruzada, aumentando a qualidade da estimativa do erro cometido pelo classificador [Quinlan, 1993].

O algoritmo C5.0 [URL1] é o sucessor mais recente do algoritmo C4.5, otimizado para lidar com as exigências do mundo real, através do aumento de eficácia ao nível do tempo de processamento e da memória utilizada. Em determinadas situações, pode conduzir à redução dos erros de classificação até 40% [URL1]. Uma das características mais importantes deste algoritmo é a utilização da técnica de *Boosting*, que consiste em gerar vários classificadores a partir do mesmo conjunto de treino e depois combiná-los num único classificador final, no qual cada classificador inicial participa com um peso, ajustado durante a fase de treino [Quinlan, 1996] [Schapire, 2002].

IV.3.2. Indução de Regras

A Indução de Regras diz respeito à descoberta de tendências e padrões em grupos de dados. Esta técnica consiste na detecção de regras de previsão, do tipo **Condição... Acção**, onde a condição da regra especifica alguns atributos a acção da regra prevê um valor para um determinado atributo cujo conhecimento é desejado [Quinlan, 1993] [Fayyad et al., 1996] [Berson et al., 2000].

As regras devem, segundo Quinlan [Quinlan, 1993] e Kufirin [Kufirin, 1997], satisfazer três condições, nomeadamente previsões correctas, compreensíveis e úteis para a tomada de decisão.

A Indução de Regras é uma das técnicas de *Data Mining* mais populares e permitiu fazer uma das descobertas mais famosas na cadeia de supermercados *Wall-Mart*: "às sextas-feiras, quem compra fraldas também compra cerveja" [Quinlan, 1993] [Berson et al., 2000].

A Indução de Regras surge, muitas vezes, associada às Árvores de Decisão, sendo usadas para representar o conhecimento nelas contido. Na apresentação das regras, a regra mais importante corresponde ao primeiro nó da árvore e, as regras menos relevantes correspondem aos nós subsequentes de acordo com a sua relevância.

As regras têm dois graus associados, a **confiança** e o **suporte**. O suporte é o número de casos onde a regra é encontrada, reflectindo o número de vezes que X e Y aparecem associados. A confiança é a probabilidade condicional da regra. Por outras palavras, dá conta da probabilidade do segundo elemento estar presente, sendo calculada pelo cociente entre o número de casos onde X e Y aparecem juntos (suporte absoluto) e o número de casos onde somente X aparece [Quinlan, 1993]. As vantagens da técnica de Indução de Regras estão relacionadas com o modo directo de lidar com os dados, o desempenho, a facilidade de explicação e a compreensão das regras. As desvantagens prendem-se com a dificuldade no manuseamento de informação incompleta ou de valores fora do contexto [Langley & Simon, 1995].

IV.3.3. Avaliação dos Modelos

Após a obtenção dos modelos, é necessário avaliar o seu desempenho. Existem vários métodos para estimar a capacidade de generalização de um modelo (e.g. Estatística Simples, Validação com Divisão da Amostra, Validação Cruzada e *Bootstrapping*). Um dos métodos mais utilizados para estimar a capacidade de generalização é a validação com divisão da amostra, que se baseia numa divisão dos dados do problema em casos de treino, para a aprendizagem do modelo, e casos de teste para avaliar a sua capacidade de generalização. As principais vantagens deste método prendem-se com a sua simplicidade e rapidez, a sua principal desvantagem está relacionada com a redução efectiva dos dados disponíveis para treino [Han & Kamber, 2006] [Bramer, 2007].

A Validação Cruzada, Figura IV.3, é um melhoramento do método de Validação com Divisão da Amostra, que permite a utilização de todos os casos disponíveis. Na validação cruzada k -desdobrável, os dados (P) são divididos em k subconjuntos mutuamente exclusivos (P_1, P_2, \dots, P_k) de cardinalidade semelhante. Os modelos são treinados e testados k vezes. O erro final de generalização é dado pela média dos erros de validação obtidos durante os k testes. Os valores de k podem variar entre 2 e n , embora o valor 10 seja o mais utilizado. A Validação Cruzada é notavelmente superior à validação com divisão da amostra para pequenos conjuntos de exemplos de treino. A sua principal desvantagem prende-se com o facto de exigir um grande esforço computacional [Han & Kamber, 2006] e [Bramer, 2007].

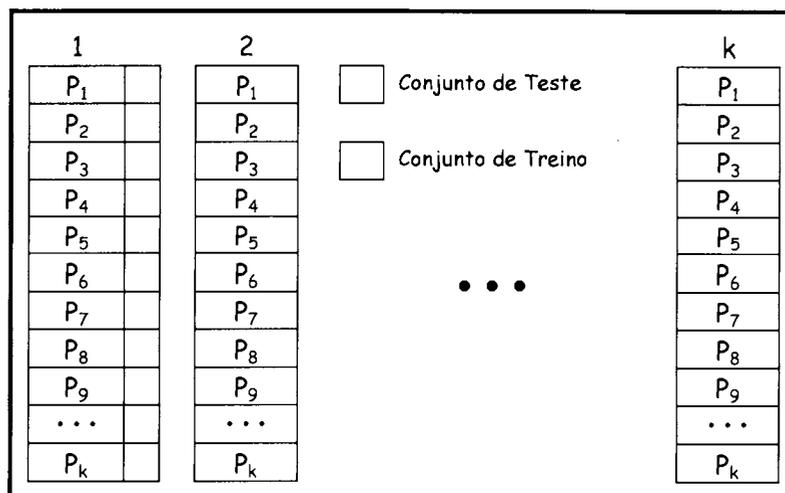


Figura IV. 3- Validação Cruzada com k Iterações. Adaptada de [Cortez, 2004]

IV.4. Conclusões

Neste Capítulo abordou-se o processo de Descoberta de Conhecimento em Bases de Dados. Este processo recorre a modelos e técnicas de *Data Mining* para extrair informação útil que, de outro modo, dificilmente seria revelada e representá-la de forma acessível e legível para o utilizador.

O processo de DCBD é essencialmente composto por três tarefas: o pré-processamento, o *Data Mining* e o pós-processamento, sendo cada uma destas etapas constituídas por várias sub-tarefas. Na etapa de pré-processamento, os dados são reconfigurados com vista à obtenção de uma BD consistente. A etapa de *Data Mining* visa a descoberta de conhecimento, enquanto na etapa de pós-processamento se procede à análise do desempenho, recorrendo-se as métricas de avaliação e se converte o conhecimento obtido numa linguagem perceptível para o utilizador final.

Capítulo V – Procedimento Experimental

Amostragem; programa de amostragem; procedimento de amostragem; técnicas analíticas utilizadas na determinação dos parâmetros químicos e construção da BD.

O laboratório da Refinaria de Sines é o responsável pelo controlo interno da qualidade dos efluentes. São determinados vários parâmetros, nomeadamente pH, SST, CQO, Óleos e Gorduras, Compostos Fenólicos e Sulfuretos.

V.1. Amostragem e Preservação das Amostras

Uma amostra é um conjunto de observações ou porção da população que é estudada de forma a compreender toda a população. De acordo com Zhang [Zhang, 2007], uma amostra representativa é "uma amostra de um universo ou um todo que seja previsível ter as propriedades médias do universo ou do todo".

A planificação de um programa de amostragem requer:

- A definição clara dos objectivos do programa;
- A pesquisa de estudos anteriores e de resultados obtidos sobre o local ou locais análogos;
- A estimativa das ameaças actuais e futuras do local;
- A inventariação do equipamento e pessoal disponível;
- Um orçamento disponível e um levantamento das limitações inerentes;
- A avaliação das componentes de erro; e
- A avaliação dos processos de documentação e bases de dados.

O procedimento de amostragem é um elemento importante do programa de controlo da qualidade dos efluentes industriais. Não se pode depositar confiança no resultado da uma análise laboratorial, mesmo que seja seguido um método analítico aconselhado e o equipamento mais adequado, se a amostra não for representativa do efluente a controlar.

O principal objectivo da amostragem é recolher uma amostra de efluente de volume suficientemente pequeno que possa ser convenientemente transportada e manuseada no

laboratório mas, que seja representativa do efluente. Deste modo, as proporções relativas e as concentrações das espécies que se querem dosear não devem variar significativamente até à conclusão das análises [Eaton et al., 2005] [Zhang, 2007].

A Refinaria de Sines possui um plano de amostragem que cumpre as exigências estabelecidas no Regulamento de Recolha e Tratamento de Água Residual Industrial do Sistema de Santo André (RARISA). As amostras de efluente são recolhidas directamente para frascos de vidro e preservadas com os reagentes apropriados para cada tipo de análise. São seguidas as recomendações do *Standard Methods for the Examination of Water and Wastewater* (SMEWW) [Eaton et al., 2005], encontrando-se na Tabela V.1 o tipo de recipiente utilizado, o volume recolhido e o modo de preservação da amostra.

Tabela V. 1 - Recolha e Preservação das Amostras

PARÂMETRO ANALÍTICO	FRASCOS DE COLHEITA	VOLUME DE AMOSTRA / cm ³	PRESERVAÇÃO
pH	Escuro	500	-
SST	Branco	1000	-
CQO	Escuro	500	Adicionar H ₂ SO ₄ até pH ≤ 2
Óleos e Gorduras	Branco	1000	Adicionar HCl 1:1 até pH ≤ 2
Compostos Fenólicos	Branco	1000	Adicionar 2 cm ³ de H ₂ SO ₄ concentrado
Sulfuretos	Escuro	500	Adicionar 1 cm ³ de (CH ₃ COO) ₂ Zn 2 mol dm ⁻³ e 0,5 cm ³ de NaOH 6 mol dm ⁻³

V.1.1. Pontos de Amostragem

Nas unidades de tratamento de efluentes é prática comum efectuar um controlo da qualidade do efluente à entrada e à saída de cada órgão. Este procedimento visa, por um lado, manter os níveis exigidos à saída e, por outro, possibilitar actuação quando se verifica uma situação fora do comum. Deste modo, é possível racionalizar a utilização de aditivos e

de reagentes, o que se traduz numa mais eficaz gestão de custos. Assim, é essencial definir correctamente a população alvo e definir as fronteiras espaciais que clarificam a representatividade dos resultados. No presente caso, para cumprir as metas anteriormente enunciadas, é necessário conhecer os dados relativos à qualidade do efluente nos seguintes pontos de amostragem:

- A) Entrada da Flotação;
- B) Saída da Flotação;
- C) Saída para a ETAR.

O ponto C (Figura V.1) corresponde ao ponto de fronteira entre o sistema de Santo André e a Refinaria. Neste ponto é realizada a recepção da água residual drenada pela refinaria, cujas amostras são recolhidas e analisadas pelo laboratório da Refinaria de Sines e por um laboratório acreditado, com certificação de qualidade, e aceite pelo concedente ou pelo Instituto Regulador de Águas Residuais (IRAR).

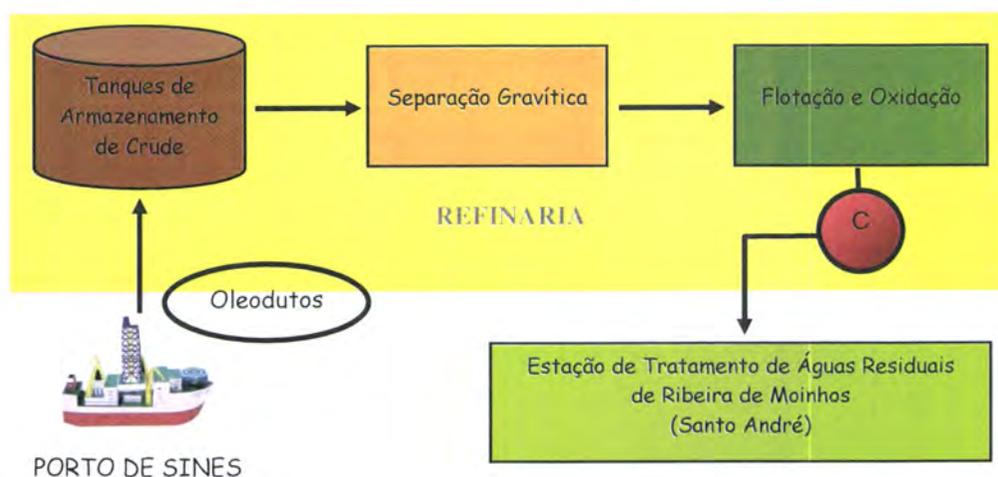


Figura V. 1 - Diagrama Simplificado da Refinaria de Sines onde se evidencia o ponto de colheita (C).

A Figura V.1 é um diagrama simplificado do trajecto dos crudes desde a sua chegada ao porto de Sines até à sua entrada na refinaria, cuja transferência é realizada através de oleodutos. Na Refinaria, o crude é armazenado em tanques sendo, posteriormente, submetido a processos de tratamento dos quais resultam efluentes que, após a realização dos primeiros tratamentos, são conduzidos para a ETAR de Ribeira de Moinhos.

V.1.2. Frequência de Amostragem

De acordo com o regulamento de recolha e tratamento de água residual industrial do sistema de Santo André e a legislação aplicável, a periodicidade da amostragem é bisemanal, correspondendo assim a dois períodos amostrais. Cada amostra é representativa do caudal dos três ou quatro dias anteriores à colheita.

V.2. Procedimento Analítico

A legislação aplicável a águas residuais deste tipo faz referência à determinação do pH, dos SST, da CQO, dos Óleos e Gorduras, dos Compostos Fenólicos e dos Sulfuretos [RARISA, 2007]. No laboratório da Refinaria de Sines são seguidas técnicas descritas no SMEWW [Eaton et al., 2005] apresentando-se, na Tabela V.2, as técnicas analíticas, os equipamentos e os métodos de ensaio utilizados na determinação dos referidos parâmetros.

Tabela V. 2 - Determinação dos Parâmetros Analíticos

PARÂMETRO ANALÍTICO	TÉCNICA ANALÍTICA	EQUIPAMENTO	MÉTODO DE ENSAIO
pH	Potenciometria Directa	pH Metrohm 780	SMEWW 4500-H ⁺ B
SST	Gravimetria	-	SMEWW 2540 D
CQO	Espectrometria de Absorção Molecular	UNICAM UV 300	SMEWW 5220 B
Óleos e Gorduras	Espectrometria de Infravermelho com transformada de Fourier	Mattson Galaxy FT-IR 7000	SMEWW 5520 C
Compostos Fenólicos	Espectrometria de Absorção Molecular	UNICAM UV 300	SMEWW 5530 D
Sulfuretos	Espectrometria de Absorção Molecular	UNICAM UV 300	SMEWW 4500 S ²⁻ D

V.3. Base de Dados

A Base de Dados (BD) que suporta este trabalho foi construída a partir de dados experimentais correspondentes ao período compreendido entre Abril de 2006 a Outubro de 2007. Os dados respeitantes aos parâmetros anteriormente referidos foram obtidos no período que decorreu o estágio na Refinaria de Sines, no âmbito do Ambiente e Higiene e Segurança do Trabalho. Os restantes, foram facultados pela empresa, que também facultou os dados referentes aos programas de fabrico.

A BD contém 56 campos e 153 registos. Os campos constantes na BD (Tabela V.3) estão relacionados com o processo de refinação, o crude predominante em cada mistura processada e as características do efluente resultante. Na Figura V.2 mostra-se um extracto da BD, a qual foi construída utilizando o *Microsoft Excel* e, na Tabela V.4, apresentam-se alguns dados estatísticos que permitem caracterizar as variáveis presentes na BD.

Tabela V. 3 - Campos Presentes na BD Utilizada

PROCESSO DE REFINAÇÃO	CRUDE PREDOMINANTE	CARACTERÍSTICAS DO EFLUENTE
1. Duração do Processo	1. Nome do Crude	1. pH
2. Carga	2. Densidade API	2. SST
3. Densidade API	3. Enxofre	3. CQO
4. Índice de Acidez Nafténica (TAN)	4. Acidez	4. Óleos e Gorduras
5. Percentagem de cada tipo de crude presente na mistura	5. País de Origem	5. Fenóis
	6. Região	6. Sulfuretos

	A	B	C	D	E	F	AJ	AO	AP	AQ
1	Duração	Caudal	Carga	Densidade	TAN	Alba Condensate	Predominante na Mistura	Região	pH	SST
2	4,2	35900	1429	0,8745	0,15	0,0	Iranian Light	Médio Oriente	8,1	64
3	5	35900	1389	0,8549	0,68	10,0	Marlim	América Latina	8	98
4	5	35900	1389	0,8549	0,68	10,0	Marlim	América Latina	8	63
5	3	35900	1376	0,8631	0,67	21,0	Antan	África	8	65
6	6,5	35900	1449	0,8628	0,12	3,0	Basrah Light	Médio Oriente	8	76
7	6,5	35900	1449	0,8628	0,12	3,0	Basrah Light	Médio Oriente	8	76
8	4,3	35900	1431	0,8541	0,67	9,0	Marlim	América Latina	7,4	52
9	3,6	35900	1428	0,861	0,77	23,0	Marlim	América Latina	7,6	65
10	4	35900	1360	0,8425	0,15	0,0	Urals	Europa	7,7	106

Figura V. 2 - Extracto da Base de Dados Utilizada

Tabela V. 4 - Caracterização das Variáveis Utilizadas em Termos Estatísticos

VARIÁVEL	UNIDADE	Nº DE REGISTOS	MÍNIMO	MÁXIMO	MÉDIA	DESVIO PADRÃO
Duração	dias	153	1,500	11,000	4,468	1,516
Carga	ton/dia	153	732	1449	1281,904	117,461
Densidade	-	153	0,824	0,878	0,853	0,011
TAN	-	153	0,050	0,800	0,431	0,289
Alba Condensate	%	153	0	37	9,037	9,925
Albacora	%	153	0	40	3,881	9,281
Antan	%	153	0	43	0,444	3,913
Arabian Heavy	%	153	0	53	1,378	6,420
Azeri Light	%	153	0	57	1,815	8,047
Basrah Light	%	153	0	70	7,844	17,209
Brega	%	153	0	65	2,570	9,852
Brass River	%	153	0	34	0,867	4,085
CPC	%	153	0	20	6,074	8,366
Dalia	%	153	0	59	6,111	14,015
Es Sider	%	153	0	67	4,044	13,403
Forties	%	153	0	16	0,348	1,729
Flotta	%	153	0	44	0,815	5,417
Iranian Light	%	153	0	77	6,259	16,230
Iranian Heavy	%	153	0	62	0,978	6,306
Kole	%	153	0	6	0,044	0,516
Kuito	%	153	0	8	0,119	0,970
Marlim	%	153	0	62	13,259	18,768
Maya	%	153	0	53	5,067	11,407
Mellitah	%	153	0	52	5,430	11,960
Mesa	%	153	0	89	2,311	11,523
Nemba	%	153	0	100	1,926	11,779
Oso Condensate	%	153	0	50	2,948	9,338
Okono	%	153	0	46	0,778	4,955
Okwori	%	153	0	44	1,733	7,314
Outros	%	153	0	6	2,622	2,568
Sarir	%	153	0	49	1,363	6,134
Saharan Blend	%	153	0	40	4,267	9,441
Sirtica	%	153	0	49	0,681	4,488
Slops	%	153	0	8	0,570	1,945
Souedia	%	153	0	7	0,096	0,791
Soroosh	%	153	0	36	1,170	5,743
Urals	%	153	0	48	1,126	6,222
Yoho	%	153	0	20	0,481	2,791
Zuetina	%	153	0	88	1,541	9,262
Densidade API	Graus	153	18,700	58,5	31,076	10,199
Enxofre	wt %	153	0,020	3,59	1,165	1,094
Acidez	mg koh/g	153	0,010	1,92	0,481	0,567
pH	Escala Sörensen	153	1,900	8,1	7,351	0,593
SST	mg dm ⁻³	153	6	380	69,141	45,886
CQO	mg (O ₂) dm ⁻³	153	32	917	314,141	165,195
Óleos e Gorduras	mg dm ⁻³	153	0	132	19	24,727
Compostos Fenólicos	mg (C ₆ H ₅ OH) dm ⁻³	153	0,032	12	0,858	1,584
Sulfuretos	mg dm ⁻³	153	0,020	38	3,760	6,096

V.4. Conclusões

No presente Capítulo explicitou-se o programa de amostragem, o ponto de amostragem e enumeraram-se as técnicas analíticas utilizadas na determinação dos parâmetros químicos considerados como indicadores de qualidade dos efluentes. Para além destes, também foram incluídas na BD dados relativos ao processo de refinação, à composição das misturas de crudes bem como o crude predominante.

Capítulo VI – Apresentação, Interpretação e Análise de Resultados e Modelos

Desenvolvimento do processo de DCBD e justificação das opções tomadas. Apresentação e avaliação dos modelos obtidos tendo em vista responder ao problema de previsão da qualidade dos efluentes em função das características da mistura processada.

O processo de DCBD comporta várias etapas, como se pode observar na Figura IV.1. Contudo, as referidas etapas podem ser agrupadas em três grandes tarefas, como se pode observar na Figura VI.1. A primeira tarefa reporta-se ao desenvolvimento do Processo de DCBD no que concerne à Selecção, Pré-Processamento e Transformação dos Dados. Por seu lado, a segunda tarefa, consiste na aplicação das ferramentas de Data Mining tendo em vista a descoberta de conhecimentos. A terceira e última etapa comporta a Visualização, Interpretação e Avaliação de Resultados, visando converter o conhecimento obtido numa linguagem perceptível.

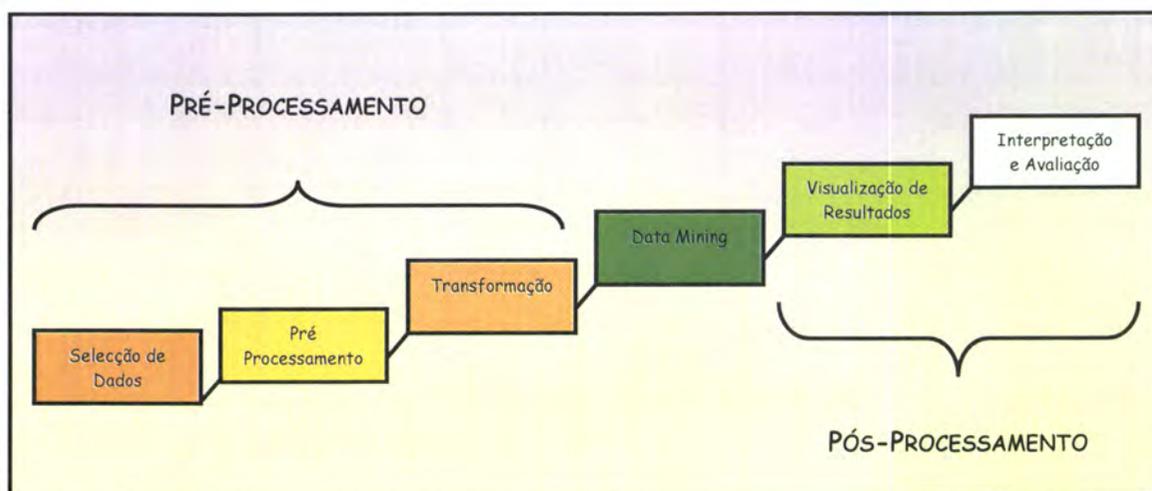


Figura VI. 1- Etapas de Desenvolvimento do Processo de DCBD Agrupadas em Tarefas

VI.1. Selecção, Pré-Processamento e Transformação de Dados

A etapa de Selecção tem como principal objectivo limitar o espaço de pesquisa, definindo-se o domínio sobre o qual se pretende executar o processo de DCBD. Nesta etapa é realizada a selecção do conjunto de dados ou variáveis necessárias ao estudo, bem como a segmentação dos dados de acordo com determinados critérios.

Para efeitos de tarifação, as descargas de águas residuais industriais são classificadas numa de cinco classes possíveis, consoante os intervalos apresentados na Tabela VI.1, sendo utilizados para o efeito os parâmetros pH, SST, CQO, Óleos e Gorduras, Compostos Fenólicos e Sulfuretos. A classificação final da Água Residual é determinada pelo parâmetro classificado na classe mais elevada [RARISA, 2007]. Deste modo, para cada um dos registos existentes na BD, foi-lhe atribuída a respectiva classe.

Tabela VI. 1- Critérios para a Classificação de Água Residual Industrial. Fonte: [RARISA, 2007]

PARÂMETROS	UNIDADE	CLASSE I	CLASSE II	CLASSE III	CLASSE IV	CLASSE V
pH	Escala Sørensen	≥ 6 ≤ 9	---	---	---	$\geq 4,5 \wedge < 6$ $> 9 \wedge \leq 10$
CQO	mg (O ₂) dm ⁻³	< 150	≥ 150 < 300	≥ 300 < 600	≥ 600 < 1000	≥ 1000 < 2000
SST	mg dm ⁻³	< 100	≥ 100 < 200	≥ 200 < 300	≥ 300 < 500	≥ 500 < 1000
Óleos	mg dm ⁻³	< 5	≥ 5 < 20	≥ 20 < 35	≥ 35 < 50	≥ 50 < 100
Sulfuretos	mg dm ⁻³	< 2	≥ 2 < 4	≥ 4 < 7	≥ 7 < 10	≥ 10 < 20
Compostos Fenólicos	mg (C ₆ H ₅ OH) dm ⁻³	< 5	≥ 5 < 10	≥ 10 < 15	≥ 15 < 20	≥ 20 < 40

As Classes de I a IV gradua a classificação das descargas de Água Residual Industrial estimulando a adopção, por parte dos utilizadores, de medidas voluntárias de redução das cargas poluentes no meio receptor. A Classe V é considerada uma classe de excepção, sendo a tarifa aplicável considerada uma penalização. No caso do valor de algum parâmetro exceder o máximo permitido para a Classe V, a descarga será classificada como proibida (DP).

No desenvolvimento do processo de DCBD, a qualidade dos dados e a forma como estão estruturados condiciona o sucesso e a prossecução de uma análise inteligente dos mesmos. Na etapa de Pré-Processamento, são realizados procedimentos que têm por objectivo a melhoria da qualidade dos dados, contribuindo para que a fase de *Data Mining* seja efectuada da forma mais eficiente. Existem vários tipos de problemas, destacando-se os relacionados com o tamanho e dimensionalidade da BD, a volatilidade dos dados (i.e., a alteração rápida dos dados pode tornar os padrões anteriormente encontrados inválidos), o ruído, dados omissos e atributos relevantes não considerados.

Os mais comuns correspondem a informação insuficiente e ainda a dados corrompidos caracterizados por possuírem ruído ou estarem incompletos. Existem diversas técnicas para lidar com dados que contêm valores omissos ou desconhecidos com origens na estatística e matemática, e.g. Métodos Bayesianos [Gelman et al., 2003] e Métodos de Imputação [Little, 1992] [Rubin, 1996] [Schafer, 1999].

Um dos métodos de imputação mais comuns, sobretudo em BD com elevado número de registos, consiste na eliminação dos registos com dados omissos. No entanto, em algumas situações, este método pode conduzir a alterações significativas na representação dos dados relativamente ao universo em estudo e conduzir a amostras muito pequenas. Na BD utilizada no presente trabalho apenas alguns registos apresentavam valores desconhecidos, tendo sido descartados. Deste modo, número de exemplos foi reduzido para 135, sendo apenas estes considerados no desenvolvimento do trabalho. Após se terem descartado os registos incompletos e atribuído a cada registo uma classe de qualidade, fez-se a distribuição do número de casos por classe (Figura VI.2), tendo em vista avaliar a necessidade de adoptar estratégias que possam melhorar o desempenho das ferramentas de *Data Mining*.

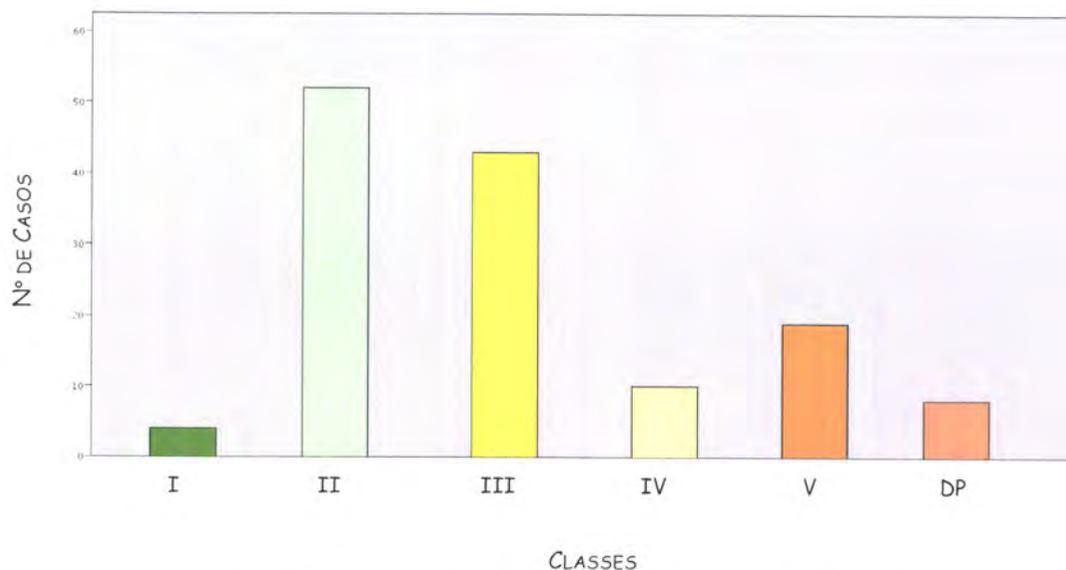


Figura VI. 2- Distribuição do Número de Casos por Classe de Qualidade

A análise da distribuição apresentada Figura VI.2 mostra que existe desproporção entre o número de casos de cada classe. Este desequilíbrio, de acordo com Weiss e Provost [Weiss & Provost, 2001], pode originar modelos com pior desempenho na previsão das classes minoritárias, sendo normalmente apontadas duas razões para esta situação. A primeira deriva do facto das regras geradas para a classe minoritária serem baseadas em menos exemplos e, conseqüentemente, mais sobreajustadas. Deste modo, o classificador tenderá a aprender limites mais rígidos do conceito. A segunda razão tem a ver com o facto de a classe mais frequentemente prevista ser a maioritária, existindo uma maior probabilidade de classificar incorrectamente exemplos da classe minoritária.

Existem, dois métodos para tornar a distribuição de classes mais equilibrada. Um dos métodos, designado por *under-sampling*, consiste em criar uma amostra mais pequena do conjunto de exemplos da classe maioritária. O outro, denominado por *over-sampling*, consiste em gerar casos a partir dos casos iniciais, de forma a aumentar o número de casos das classes minoritárias. Estes dois métodos têm desvantagens intrínsecas que se prendem com o desprezar dados potencialmente úteis no primeiro método e, no segundo, com o aumento do tamanho do conjunto de treino e, conseqüentemente, do tempo de computação podendo, ainda, proporcionar um maior sobre-ajustamento. No sentido de evitar as desvantagens associadas aos métodos de *under-sampling* e de *over-sampling*, neste trabalho optou-se por seguir uma outra metodologia que consistiu na redução do número de classes através de agrupamento.

Deste modo, as classes de qualidade foram agrupadas, passando-se do domínio {I, II, III, IV, V, DP} para o domínio {S, I, P} onde o agrupamento:

- **S (Satisfatório)** denota descargas menos poluentes, integrando as **classes I e II**;
- **I (Insatisfatório)** denota descargas mais poluentes integrando as **classes III e IV**; e
- **P (Proibitivo)** denota descargas extremamente poluentes, integrando as **restantes classes (V e DP)**.

Na Figura VI.3 encontra-se a distribuição do número de casos por agrupamento. A sua análise mostra que os desequilíbrios atrás referidos foram atenuados.

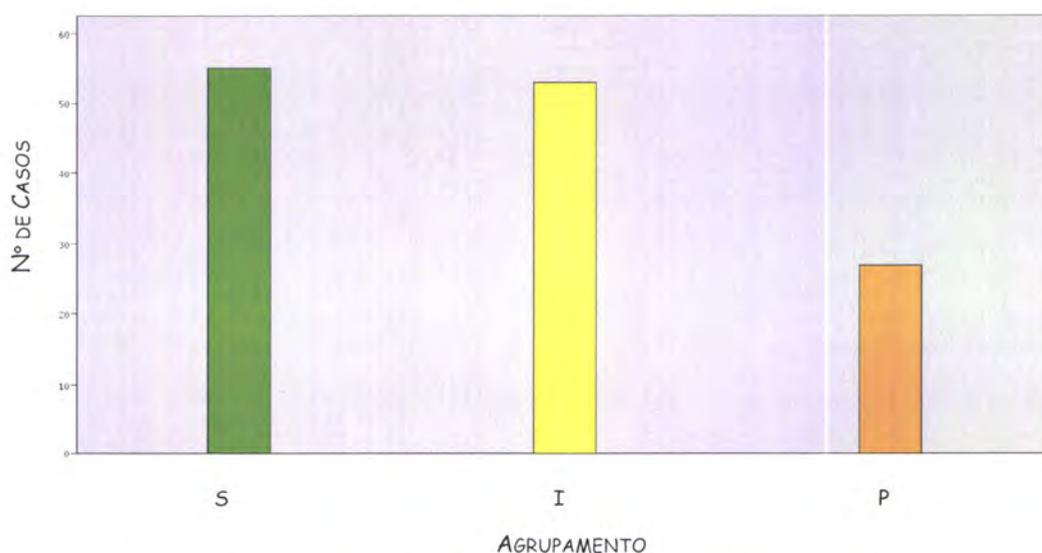


Figura VI. 3- Distribuição do Número de Casos por Agrupamentos de Qualidade

Outro aspecto importante, que deve ser atendido durante a fase de pré-processamento, prende-se com o escalonamento dos dados. Este procedimento visa transformá-los de modo a acelerar e a melhorar o processo de aprendizagem. Por exemplo, se o atributo $X \in [0; 100]$ e o atributo $Y \in [0,05; 0,80]$ o algoritmo de aprendizagem pode atribuir uma importância inadequada ao atributo X em consequência dele apresentar um domínio mais alargado [Pyle, 1999] [Han & Kamber, 2006] [Bramer, 2007]. Neste trabalho os dados foram escalonados para o domínio $[0, 1]$ tendo sido utilizada a equação:

$$X_{esc} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

onde, X , X_{esc} , X_{max} e X_{min} denotam, respectivamente, o valor do atributo, o valor do atributo escalonado para o domínio $[0, 1]$, o valor máximo que o atributo apresenta e o valor mínimo que o atributo exibe.

Como foi referido anteriormente, o conhecimento do domínio constitui-se como um recurso essencial em qualquer processo de DCBD, sendo utilizado para conduzir o processo. Assim, antes de se iniciar a etapa de Data Mining, procurou-se, por um lado, entender como é que as diversas misturas são constituídas e, por outro, obter algumas relações simples entre as variáveis. No que respeita ao estudo da composição das misturas, construíram-se gráficos nos quais se representa a percentagem dos diversos crudes presentes em cada uma delas.

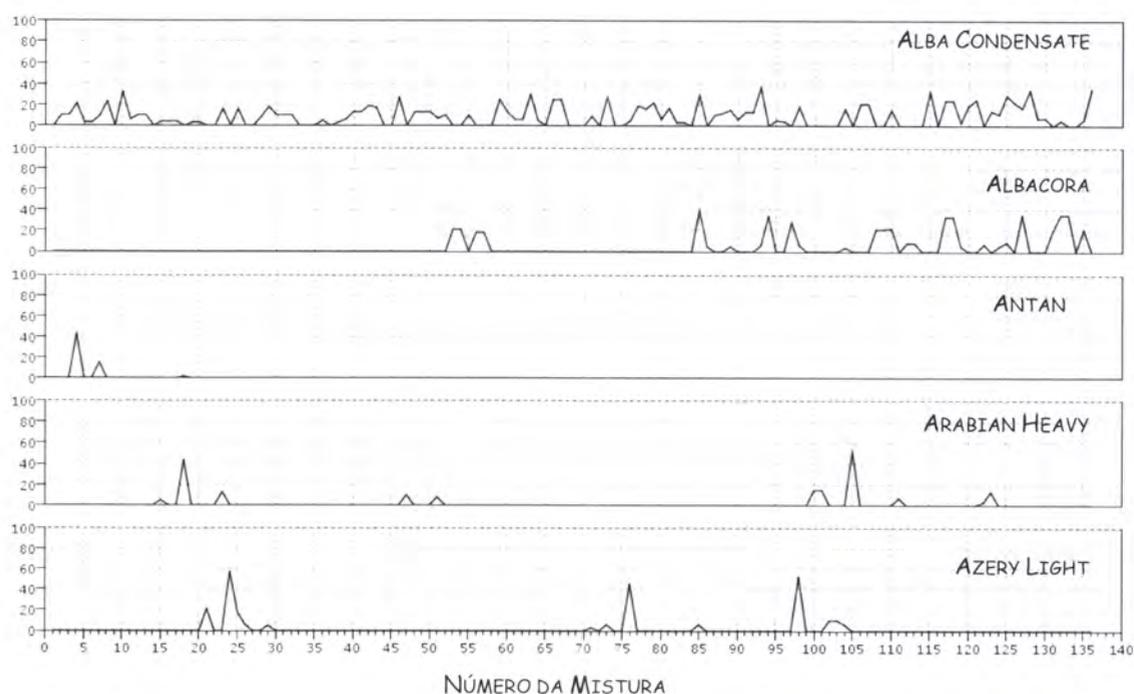


Figura VI. 4- Percentagens de Crudes Presentes nas Misturas

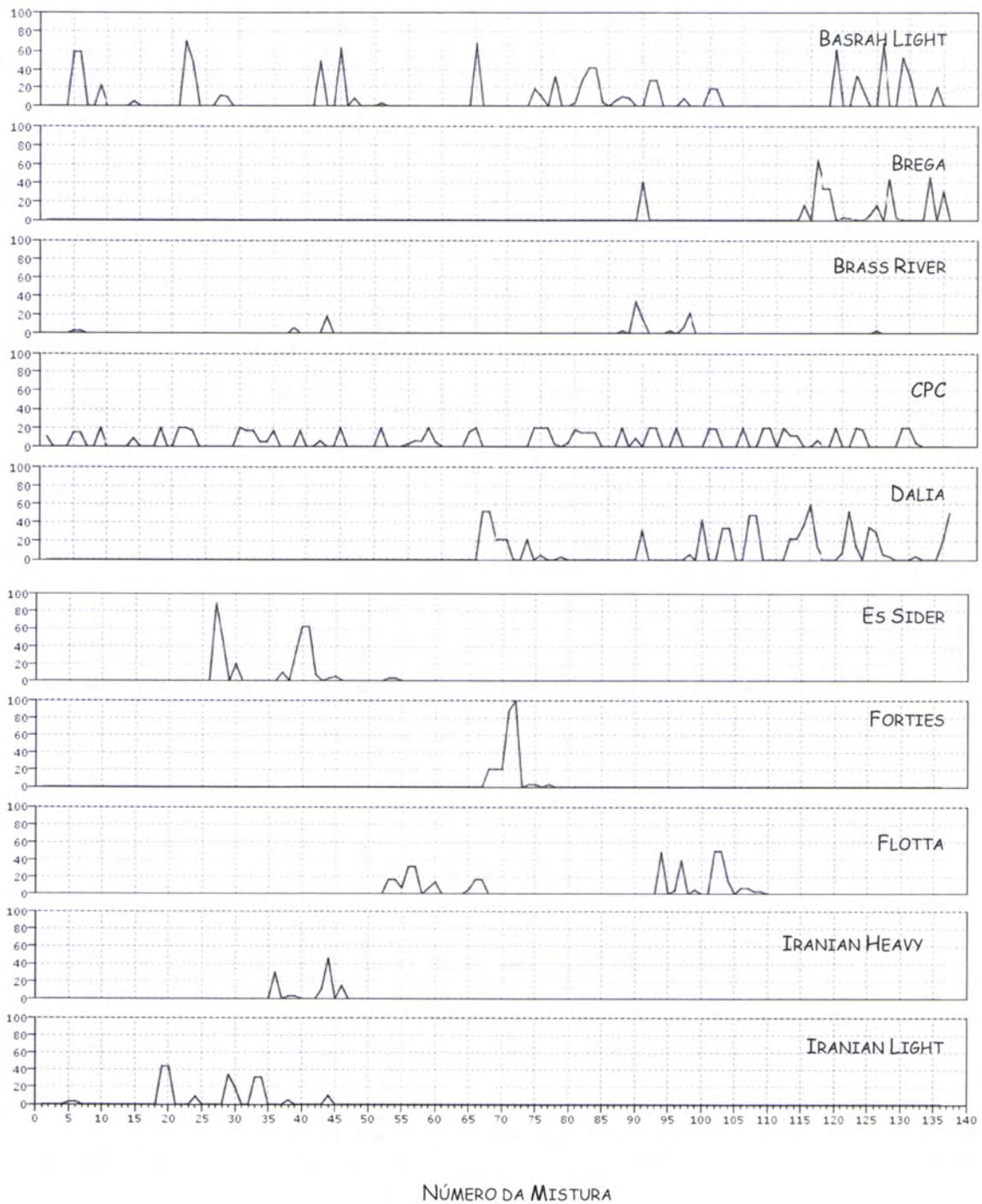


Figura VI.4 (cont.) - Percentagens de Crudes Presentes nas Misturas

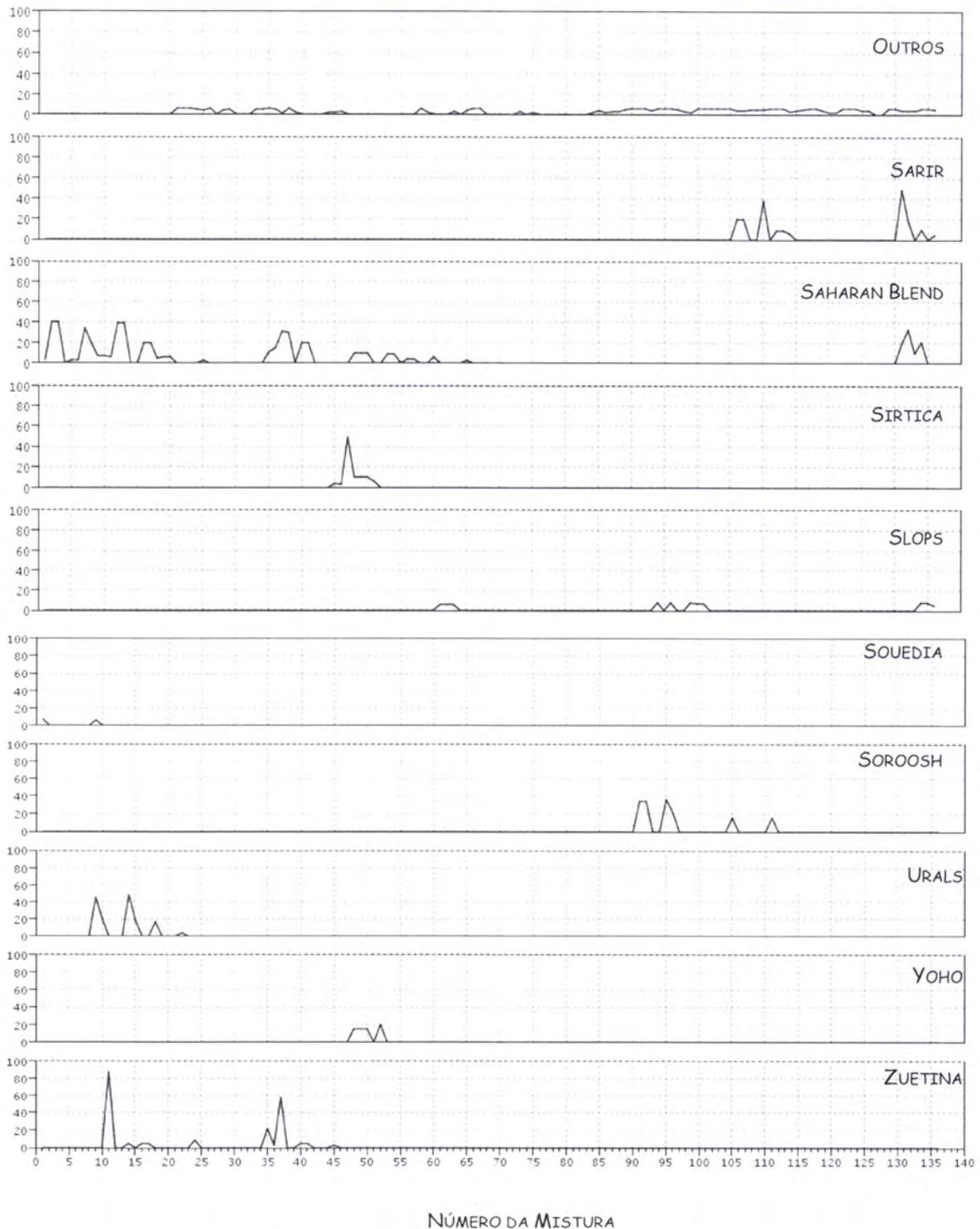


Figura VI.4 (cont.) - Percentagens de Crudes Presentes nas Misturas

A análise da Figura VI.4 revela a existência de alguns aspectos que importa salientar:

- Entre os diversos crudes utilizados é de referir que existem crudes que surgem com maior frequência nas misturas processadas, nomeadamente o *Alba Condensate* e o *CPC*, embora quase nunca excedam, 35% e 20%, respectivamente;

- Os crudes *Albacora*, *Basrah Light*, *Dalia* e *Iranian Heavy* embora não sendo tão frequentes como os anteriores, surgem em diversas misturas e, quase sempre, com percentagens elevadas, rondando os 40%, 70%, 60% e 80%, respectivamente;
- Os outros tipos de crudes são utilizados pontualmente nas diferentes misturas processadas embora, alguns deles o sejam em percentagens relativamente elevadas, como, por exemplo, o *Antan*, o *Flotta* e o *Sarir*.

Para além da pesquisa dos crudes predominantes nas misturas processadas, também se procuraram relações entre o crude predominante na mistura e os parâmetros químicos do efluente resultante tendo-se construído as representações gráficas que se apresentam em seguida. Na Figura VI.5 está representado o pH em função do crude predominante na mistura:

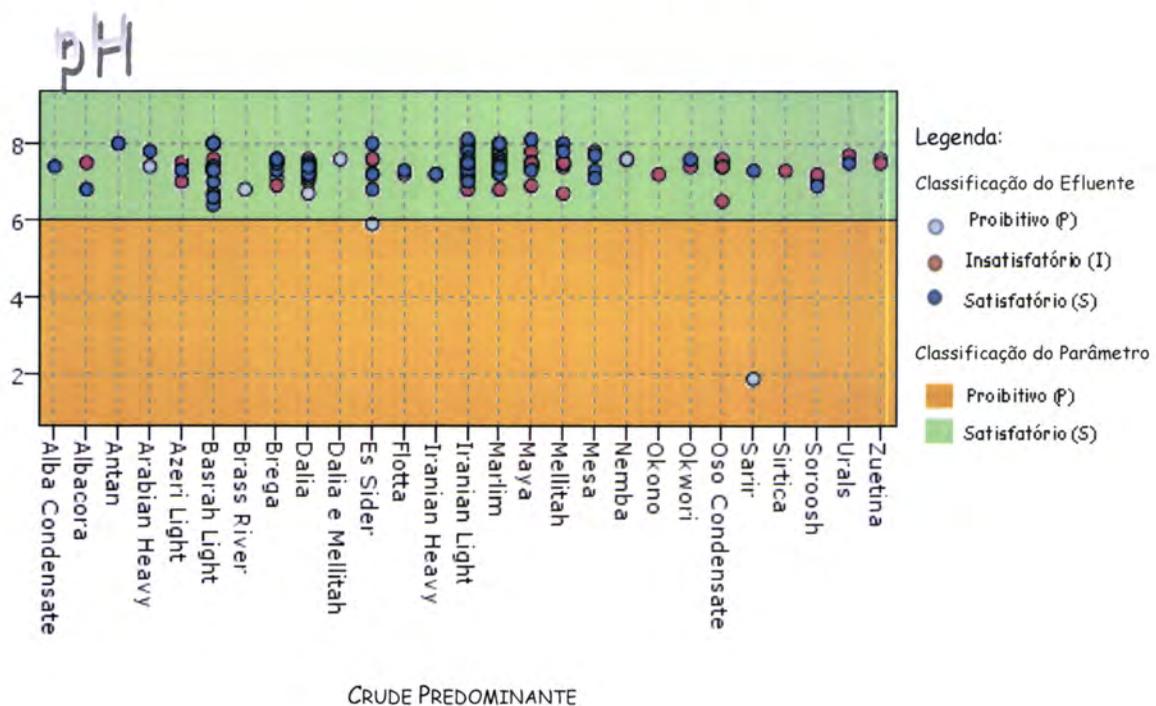


Figura VI. 5- pH vs Crude Predominante na Mistura

Através da análise da Figura VI.5 pode verificar-se que o intervalo de pH onde se localizam a maior parte das misturas processadas se situa entre 6 e 8. Este intervalo abrange o agrupamento S. Deste modo, pode-se afirmar que, por um lado, o pH não influencia a classificação do efluente e, por outro, que não existe uma relação observável entre o pH do efluente e o crude predominante na mistura que originou esse efluente. Porém, é de salientar

a existência de duas misturas cujo valor de pH pode ter influenciado a classificação do efluente.

Na Figura VI.6 pode observar-se que o intervalo onde existe maior predominância de valores, relativos ao parâmetro CQO, se situa entre 100 e 600 mg (O₂) dm⁻³, abrangendo os agrupamentos S e I.

Através da representação gráfica pode-se verificar que os teores em CQO não são responsáveis pela classificação do efluente na classe P. Contudo, existe um número relativamente elevado de casos em que este parâmetro pode ser responsável pela classificação dos efluentes como I. Tal como no parâmetro anterior, não existe uma relação observável entre os teores em CQO do efluente e o crude predominante na mistura.

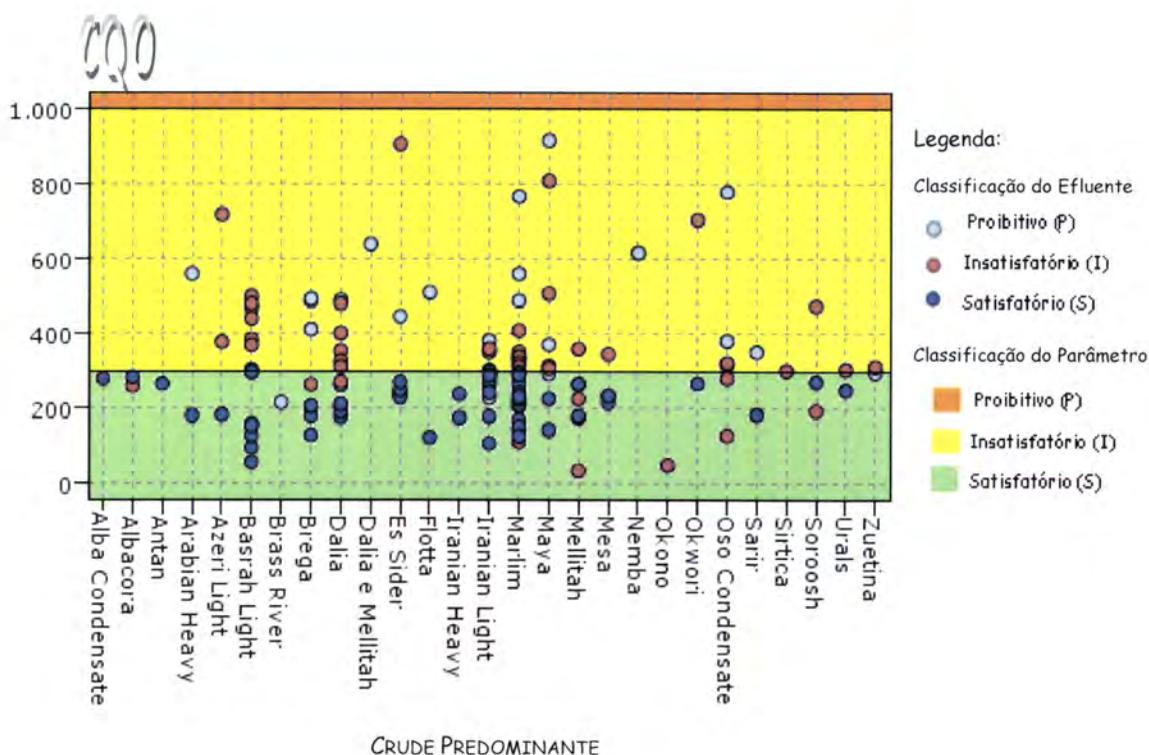


Figura VI. 6 - CQO vs Crude Predominante na Mistura

Relativamente aos Compostos Fenólicos, pode observar-se que o intervalo onde existe maior predominância de valores situa-se entre 0 e 2 mg (C₆H₅OH) dm⁻³, Figura VI.7. Este intervalo corresponde ao agrupamento S, o que permite afirmar que este parâmetro não influencia a classificação do efluente. Existe apenas uma situação, na qual o crude

predominante na mistura é *Mellitah*, e que o teor em compostos fenólicos pode ter influenciado a classificação do efluente como I.

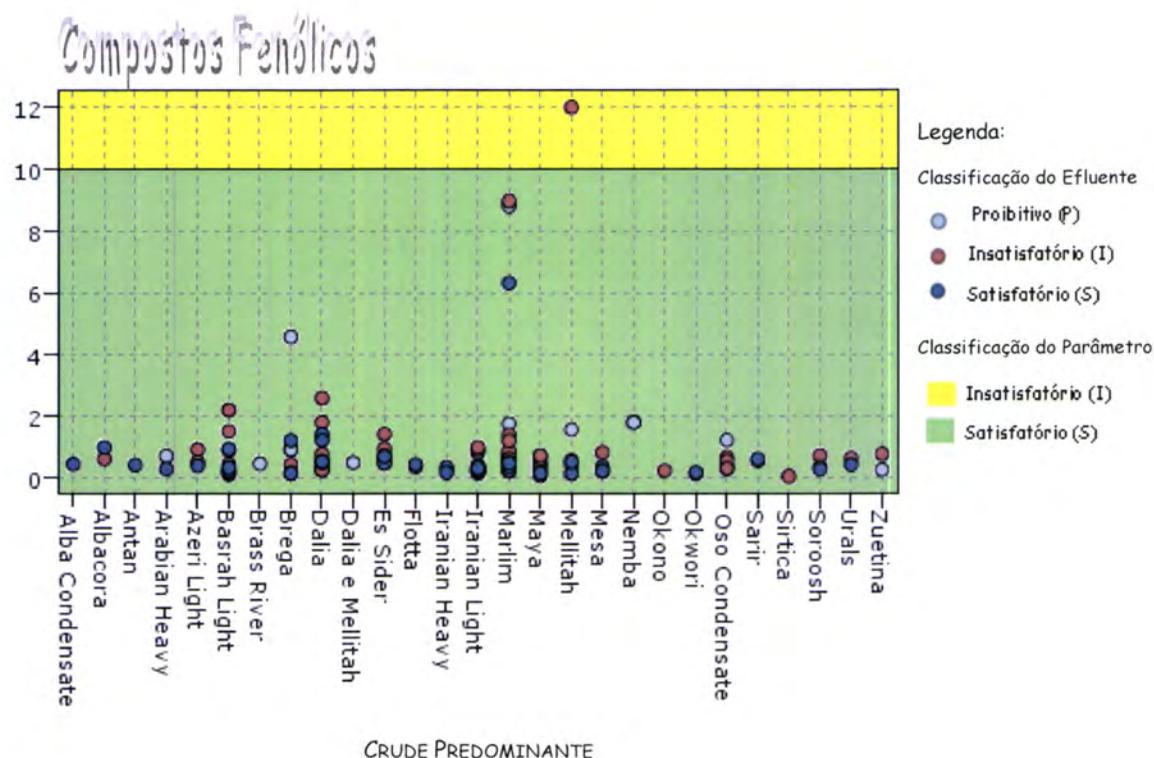


Figura VI. 7- Compostos Fenólicos vs Cruído Predominante na Mistura

Na Figura VI.8 pode observar-se que o intervalo onde existe maior predominância de valores, relativos aos SST, se situa entre 0 e 150 mg dm⁻³. Deste modo, pode afirmar-se que este parâmetro não influencia a classificação final da descarga industrial. Apenas em três situações este parâmetro poderá ter influenciado a classificação do efluente como I. Nestes casos, os cruídos predominantes nas misturas eram *Basrah Light*, *Maya* e *Mellitah*. Contudo, foram processadas outras misturas em que os referidos cruídos eram, também, predominantes e o teor em SST não influenciou a classificação do efluente.

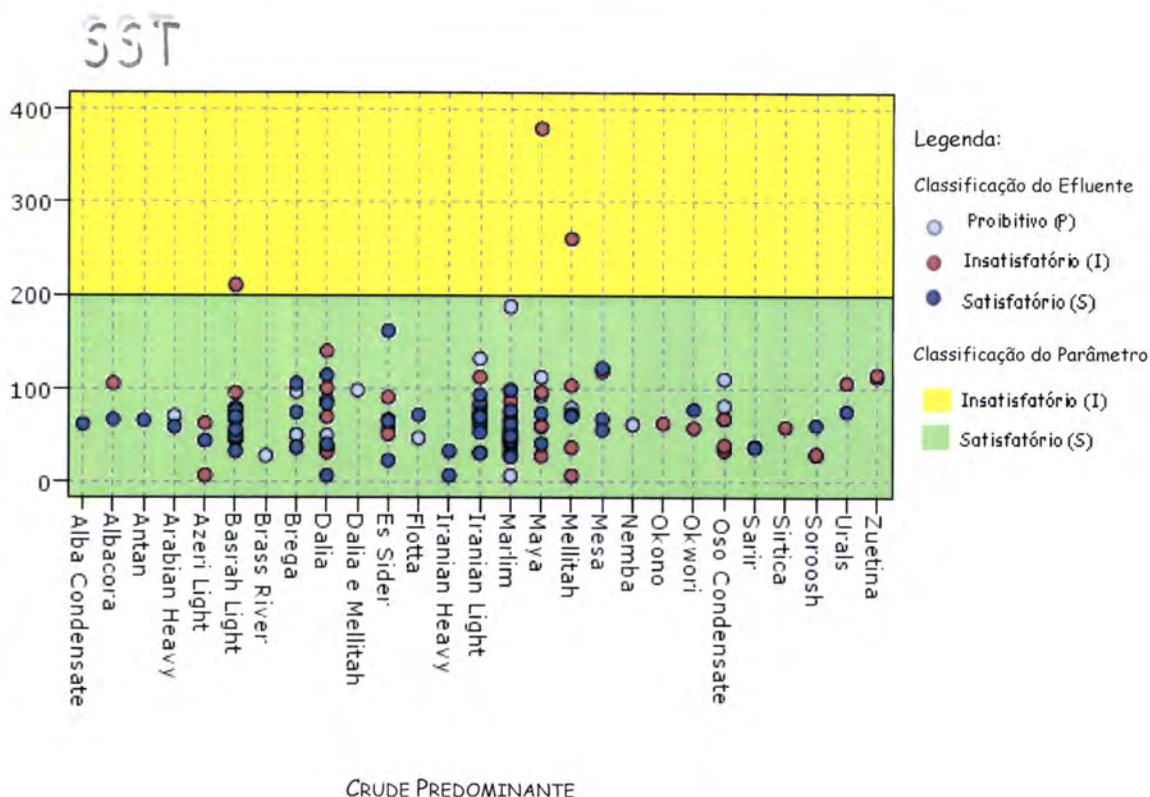


Figura VI. 8- SST vs Crude Predominante na Mistura

A análise da Figura VI.9 mostra que, na maior parte dos casos, a classificação final do efluente coincide com a classificação do parâmetro Sulfuretos. Este facto pode indicar este parâmetro como um dos principais responsáveis pela classificação final do efluente industrial. No entanto, existem alguns casos em que tal não se verifica, sendo outros os parâmetros que influenciam a classificação.

Relativamente ao parâmetro Óleos e Gorduras, tal como se pode observar na Figura VI.10, o intervalo onde existe maior predominância de valores situa-se entre 0 e 35 mg dm⁻³, abrangendo os agrupamentos S e I. Um outro aspecto a realçar prende-se com o facto da classificação final do efluente coincidir, nalguns casos, com a classificação deste parâmetro. Este facto, pode indicar este parâmetro esteja, tal como o anterior, relacionado com a classificação final do efluente industrial.

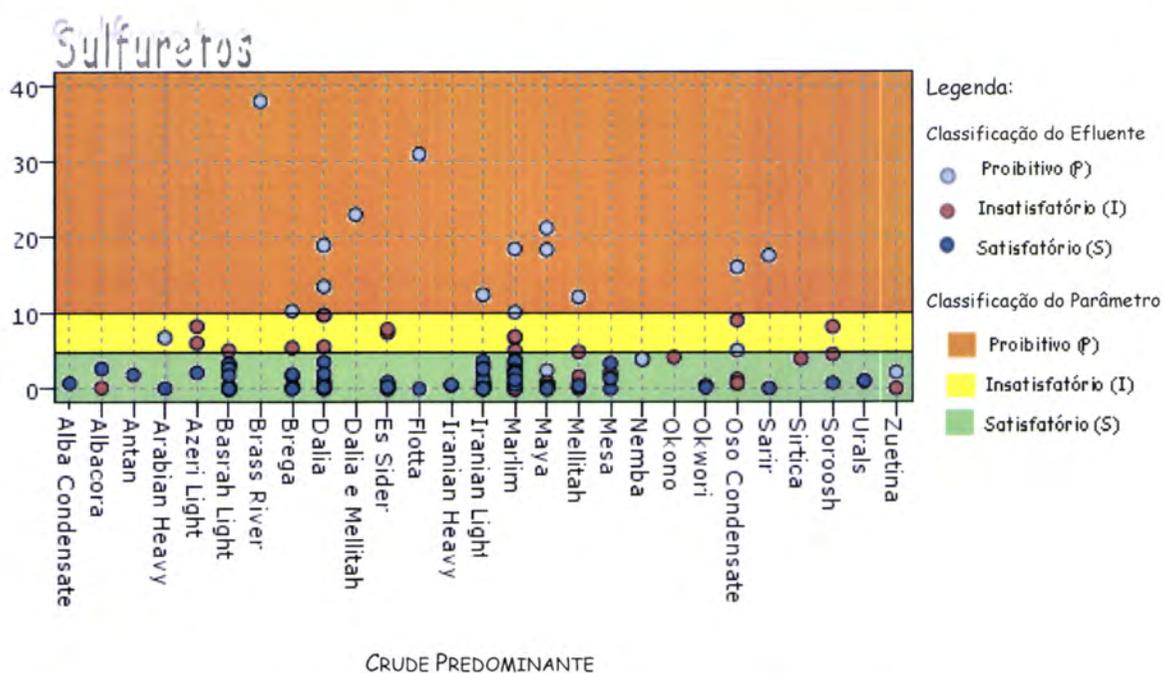


Figura VI. 9- Sulfuretos vs Crude Predominante na Mistura

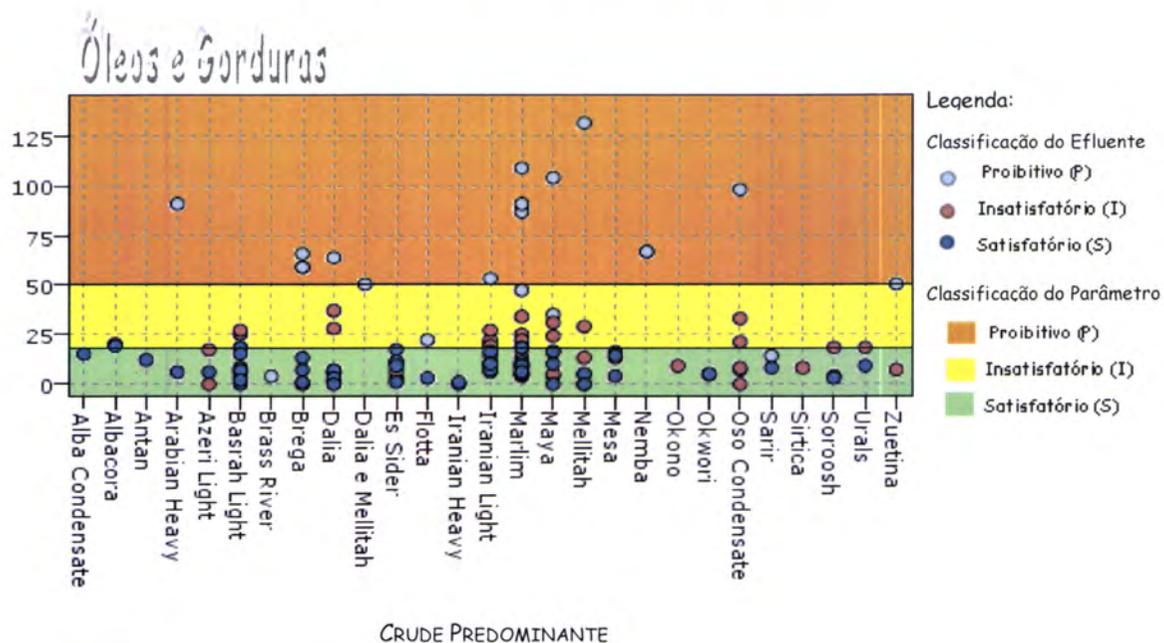


Figura VI. 10- Óleos e Gorduras vs Crude Predominante na Mistura

A análise das Figuras VI.5 a VI.10 mostra que não é possível estabelecer uma associação simples entre os crudes predominantes na mistura e os valores dos parâmetros químicos do efluente. Deste modo, pode afirmar-se que não existe uma relação observável entre os

resultados analíticos obtidos para o efluente e o crude predominante na mistura que originou esse efluente.

Não obstante o que foi afirmado anteriormente, esta análise evidencia que independentemente do crude predominante na mistura, os parâmetros CQO, SST, Compostos Fenólicos e pH não são responsáveis pela classificação do efluente no agrupamento P, contrariamente aos parâmetros Sulfuretos e Óleos e Gorduras, os quais parecem ser os responsáveis pela classificação do efluente na classe P.

Procurou-se ainda perceber se as classes de qualidade do efluente estariam relacionadas com o crude predominante na mistura. As Figuras VI.11 e VI.12 mostram a distribuição do número de casos em função do crude predominante. A Figura VI.11 diz respeito às seis classes (I, II, III, IV, V e DP) enquanto que a Figura VI.12 é relativa aos agrupamentos (S, I e P).

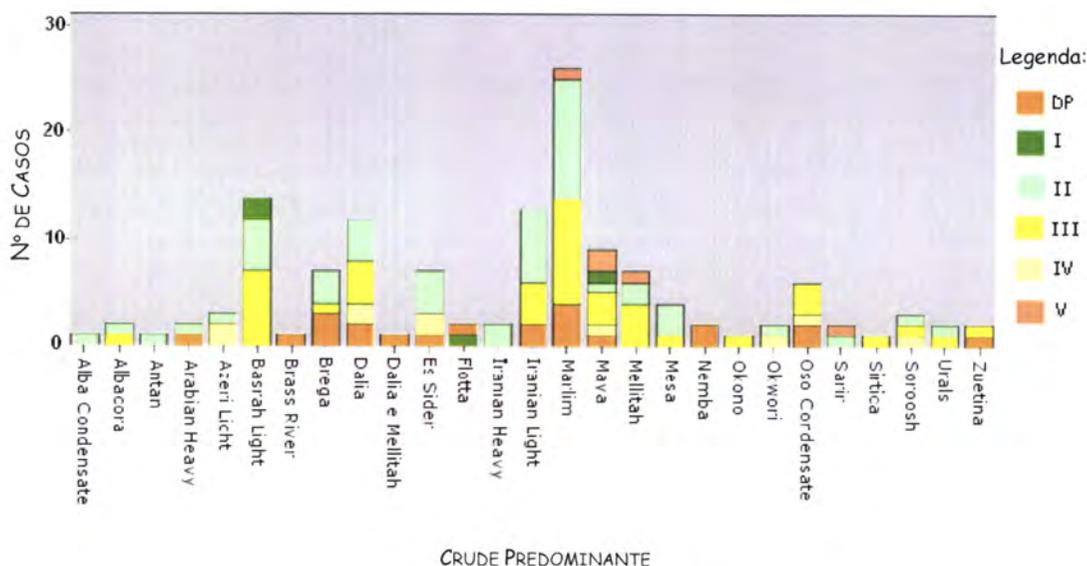


Figura VI. 11- Classes de Qualidade dos Efluentes vs Crude Predominante

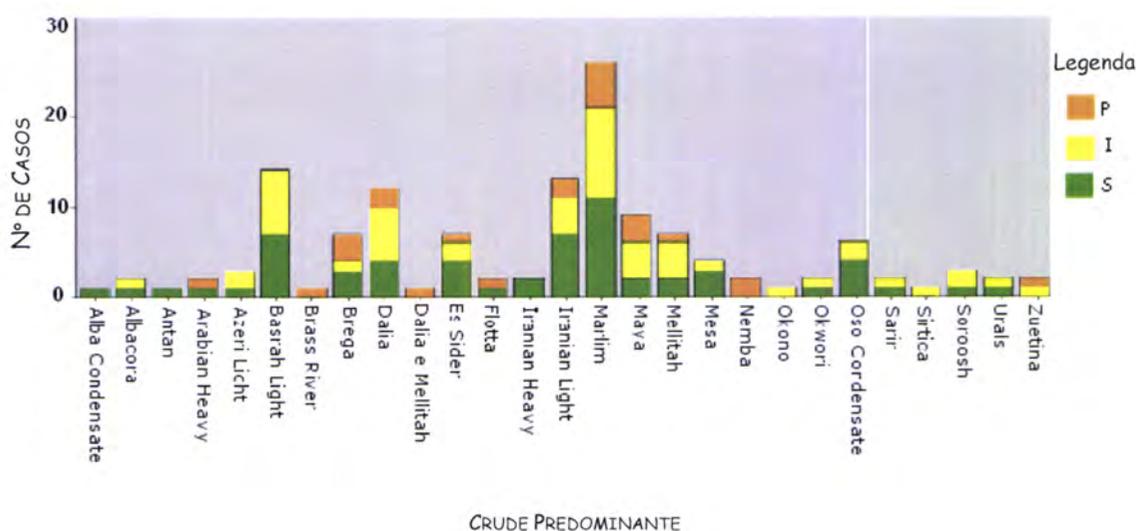


Figura VI. 12- Agrupamentos de Qualidade dos Efluentes vs Crude Predominante

A análise das distribuições apresentadas anteriormente mostra que não é possível associar uma classe ou um agrupamento de qualidade de efluente a um determinado crude predominante na mistura. Os crudes *Alba Condensate*, *Antan*, *Brass River*, *Dalia* e *Mellitah*, *Iranian Heavy*, *Nemba*, *Okono* e *Sirtica* possuem uma baixa frequência de predominância e, apesar de ser possível estabelecer a referida associação, esta não é significativa tendo em conta o número de casos.

Na procura de relações entre os crudes predominantes e os agrupamentos de qualidade estudou-se, ainda, a frequência de co-ocorrência relativa às referidas variáveis. A Figura VI.13 evidencia as ligações obtidas entre os crudes predominantes nas misturas e a classificação final do efluente resultante. Desta forma, pode verificar-se que aos crudes predominantes *Basrah Light* e *Marlim* estão fortemente associadas as classificações S e I, ao crude predominante *Dalia* está associada a classificação I e ao crude predominante *Iranian Light* a classificação S.

Não obstante a identificação das ligações fortes anteriormente referidas, existem ligações entre praticamente todos os crudes predominantes e os agrupamentos de qualidade dos efluentes não sendo, portanto, possível estabelecer relações simples entre estas variáveis.

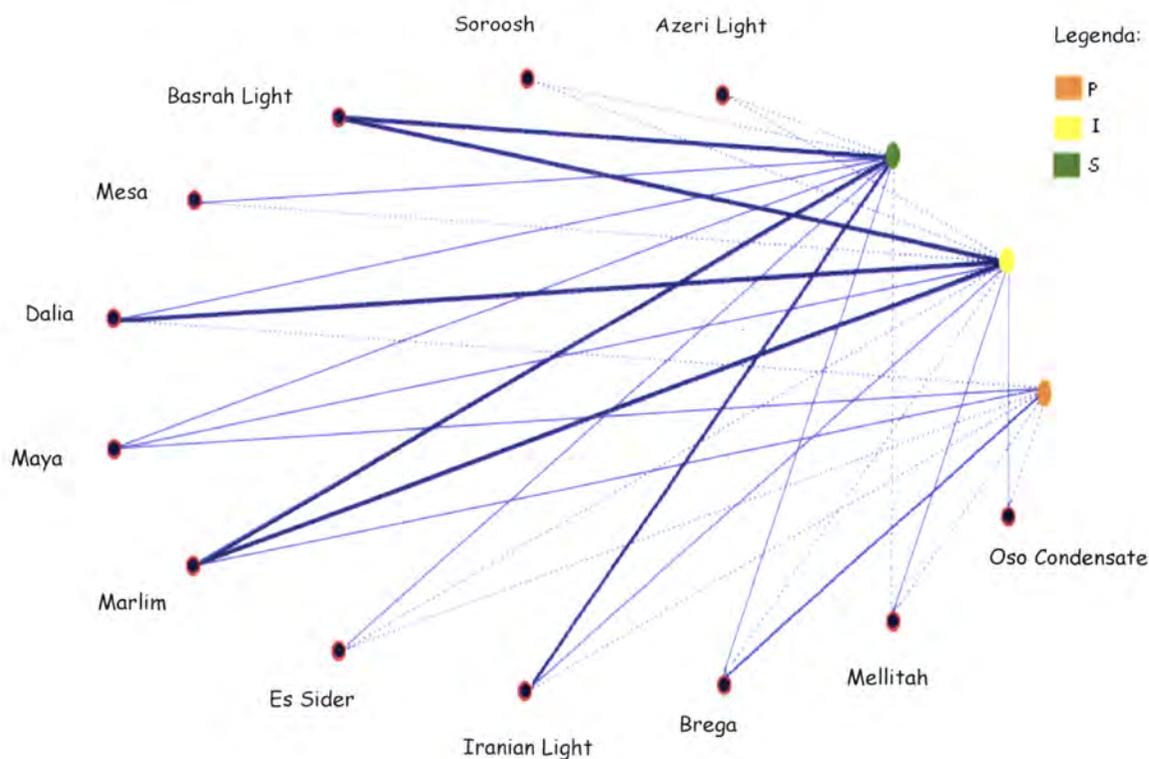


Figura VI. 13- Interações entre os Crudes Predominantes e os Agrupamentos de Qualidade

Nesta secção do trabalho foram colocadas e testadas diversas hipóteses. Contudo, apesar de terem contribuído para um melhor conhecimento do problema em estudo, todas elas se mostraram incapazes de relacionar as misturas processadas com a qualidade do efluente resultante.

Do exposto fica claro que se torna necessário recorrer a técnicas que permitam, de forma automática, descobrir padrões ou tendências nos dados e inferir regras para os mesmos. Deste modo, no seguimento do trabalho e na sequência da metodologia adoptada, passou-se para a etapa de Data Mining no sentido de descobrir relações entre as misturas processadas e a classe de qualidade dos efluentes originados.

VI.2. Modelação

Actualmente não se conhecem estudos de previsão dos parâmetros de qualidade deste tipo de efluentes industriais. Tal facto pode estar relacionado com a multiplicidade e com a

complexidade dos factores envolvidos no universo da Indústria Petrolífera que dificultam a abordagem deste tipo de problemas. De entre esses factores destacam-se os processos utilizados nas refinarias, o tipo de instalações, a carga utilizada em cada programa de fabrico e a variabilidade das misturas de crudes refinadas introduzindo, estas últimas, variáveis como a composição, as propriedades de cada componente e as características da mistura.

No presente trabalho, a previsão da classe do efluente foi definida como um problema de classificação tendo-se recorrido à Indução de Regras e Árvores de Decisão. A escolha das referidas técnicas prende-se com diversas razões. Por um lado, são umas das técnicas de *Data Mining* mais utilizadas neste tipo de problemas, Tabela IV.3, permitindo uma representação expressiva do conhecimento, de fácil utilização e compreensão. Por outro lado, têm sido aplicadas e originado resultados satisfatórios em estudos de previsão dos parâmetros de qualidade de água [Vicente, 2004] [Santos et al., 2005] [Atkins et al., 2007] [Pinto et al., 2009].

Tendo em vista assegurar a significância estatística dos resultados obtidos, cada simulação foi executada dez vezes. Em cada uma delas, os registos da BD foram aleatoriamente divididos em dois conjuntos mutuamente exclusivos. O conjunto de treino, composto por 67% dos exemplos e o conjunto de teste, constituído pelos restantes. Foi utilizada uma versão do algoritmo C4.5 disponibilizado pelo *software* do domínio público *Weka*, denominada por algoritmo J.48, para induzir as árvores de decisão e os respectivos conjuntos de regras.

VI.3. Apresentação, Interpretação e Avaliação dos Modelos

Atendendo ao facto da classe de qualidade do efluente ser apenas conhecida após a conclusão de todo o processo, pretendia-se desenvolver modelos para a previsão do tipo de efluente com base nas características dos crudes processados.

Com o intuito de se atingir este objectivo, desenvolveram-se dois modelos diferentes. Num deles, denominado por Modelo I, foram utilizadas, como variáveis de entrada, as que dizem respeito ao processo de refinação e ao crude predominante na mistura, Tabela V.3. No outro, designado por Modelo II, apenas foram utilizadas as variáveis respeitantes ao processo de refinação indicadas na Tabela V.3.

Na Figura VI.14 apresenta-se um excerto da árvore de decisão obtida para o Modelo I:

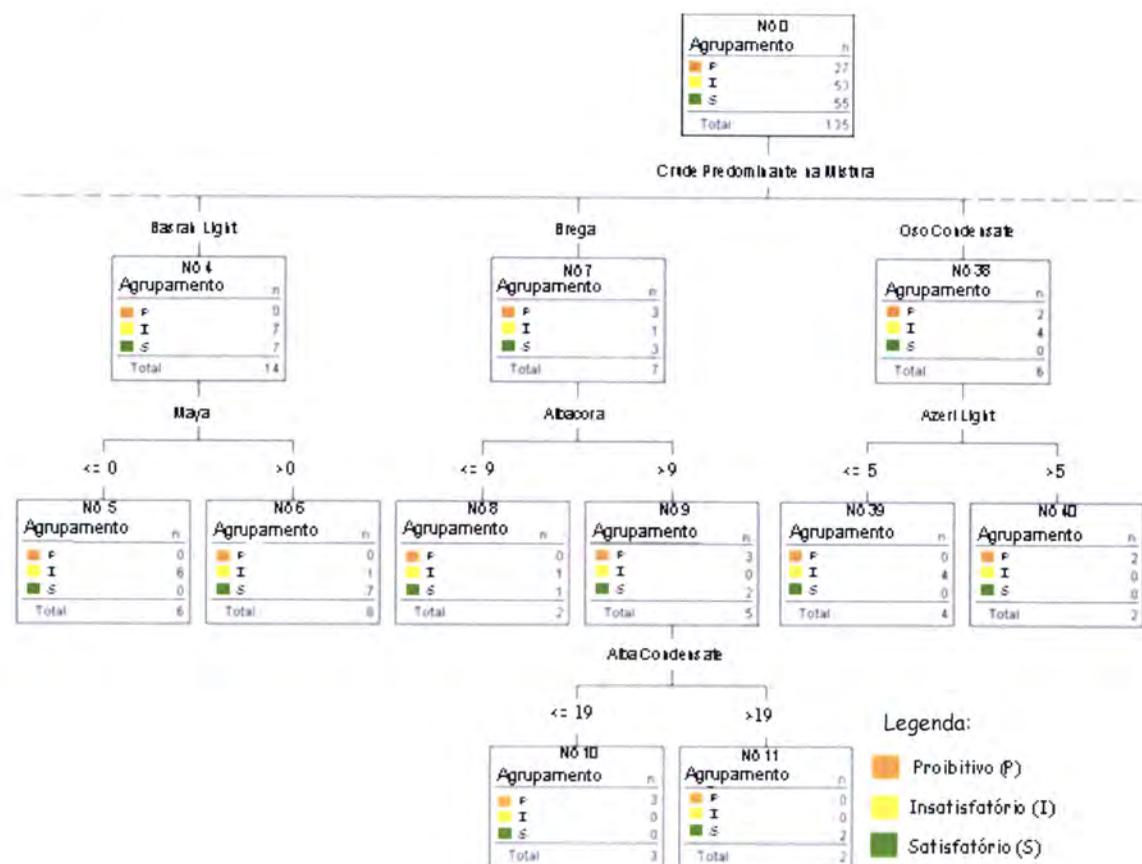


Figura VI. 14- Extracto da Árvore de Decisão - Modelo I

Através da análise da Figura VI.14 pode-se verificar que se por exemplo o crude predominante na mistura é o *Oso Condensate* e se existe *Azeri Light* numa quantidade inferior ou igual 5% então a classificação da qualidade do efluente será Insatisfatória. Porém, se a quantidade de *Azeri Light* for superior 5% então, a classificação da qualidade do efluente será Proibitiva.

Na Figura VI.15 apresenta-se um excerto da árvore de decisão obtida para o Modelo II.

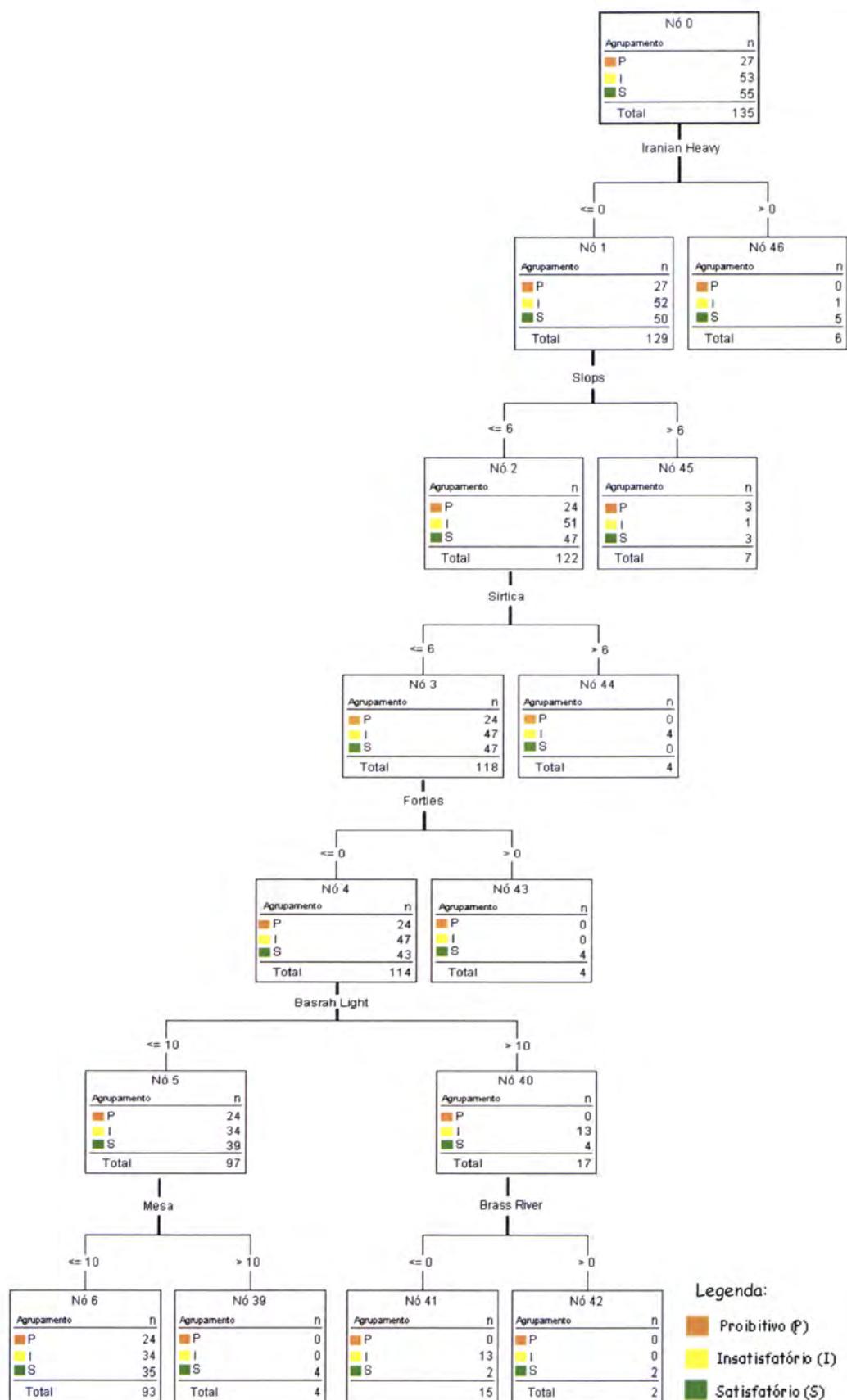


Figura VI. 15- Extracto da Árvore de Decisão - Modelo II

Através da análise da Figura VI.15 pode-se verificar que se por exemplo o *Iranian Heavy* existir nas misturas então a classificação da qualidade do efluente será Satisfatória. Porém, se não estiver presente na mistura a classificação estará dependente da quantidade e dos diferentes tipos de crude presentes. Deste modo, se estiverem presentes *Slops* em quantidades inferiores a 6% e a mistura tiver o crude *Sirtica* numa percentagem superior a 6%, a classificação da qualidade do efluente será Insatisfatória.

Em ambos os Modelos, as Árvores de Decisão utilizadas são do tipo de Classificação, nas quais os registos são classificados e associados a uma de várias classes discretas. As Árvores de Decisão obtidas tanto para o Modelo I bem como para o Modelo II são bastante extensas e complexas e como tal, de difícil apresentação. Deste modo, optou-se por apresentar, nos Anexos I e II, apenas o conjunto de regras extraído para cada um dos modelos. Na Figura VI.16 mostra-se um extracto das regras obtidas para cada um dos modelos, destacando-se em cada uma delas o suporte (i.e. número de casos onde a regra é encontrada) e a confiança (i.e. a probabilidade condicional dessa regra) que, como se pode observar, são elevadas.

MODELO I	MODELO II
...	...
Se Predominante = Marlim e Yoho <= 0 e Brass River <= 3 e Mellitah <= 4 e Albacora <= 0 e A.Condensate <= 27 e Marlim > 42	Se Iranian Heavy <= 0 e Slops <= 6 e Sirtica <= 6 e Forties <= 0 e Basrah Light > 10 e Brass River <= 0
Então → S	Então → S
n = 6 Confiança = 1,0	n = 15 Confiança = 0,867
...	...

Figura VI. 16- Extracto do Conjunto de Regras Obtido para cada Modelo

Uma ferramenta comum para a avaliação de modelos de classificação é a matriz de coincidências [Kohavi & Provost, 1998]. Trata-se de uma matriz de tamanho L x L, onde L denota o número de saídas possíveis (três, no presente caso). A matriz de coincidências avalia o resultado de uma classificação, mapeando os valores previstos pelo modelo com os valores

desejados. Na Tabela VI.2 apresentam-se as matrizes de coincidências referentes a cada um dos modelos obtidos.

Tabela VI. 2- Matrizes de Coincidências para os Modelos I e II

AGRUPAMENTO	MODELO I					
	CONJUNTO DE TREINO			CONJUNTO DE TESTE		
	S	I	P	S	I	P
S	29	11	1	9	3	2
I	4	33	1	2	12	1
P	3	3	14	0	1	6

AGRUPAMENTO	MODELO II					
	CONJUNTO DE TREINO			CONJUNTO DE TESTE		
	S	I	P	S	I	P
S	29	7	2	12	2	3
I	3	30	5	4	10	1
P	3	0	11	1	0	12

A acuidade dos modelos, medida em termos da percentagem de acertos relativamente ao número de casos apresentados é, no caso do Modelo I, de 76,77% (76 acertos/99 casos) para o conjunto de treino e de 75,00% (27 acertos/36 casos) para o conjunto de teste. Relativamente ao Modelo II, a acuidade foi de 77,78% (70 acertos/90 casos) para o conjunto de treino e de 75,56% (34 acertos/45 casos) para o conjunto de teste.

Na Tabela VI.3 apresenta-se a percentagem de acertos para cada um dos agrupamentos para ambos os modelos. A sua análise, conjugada com a acuidade global dos modelos, mostra que ambos têm um desempenho semelhante, embora o Modelo I seja preferível para prever o agrupamento S enquanto que o Modelo II dá melhores resultados na previsão dos agrupamentos I e P.

Tabela VI. 3- Percentagem de Acertos por Agrupamento para cada um dos Modelos

AGRUPAMENTO	CONJUNTO DE TREINO			CONJUNTO DE TESTE		
	S	I	P	S	I	P
Modelo I	80,5	70,2	87,5	81,8	75,0	66,7
Modelo II	82,9	81,0	61,0	70,6	83,3	75,0

VI.4. Conclusões

No presente Capítulo justificaram-se as opções tomadas em cada uma das etapas do processo de DCBD e procurou-se obter relações entre as variáveis em estudo. Apresentaram-se e avaliaram-se os modelos obtidos, tendo em vista responder ao problema de previsão da qualidade dos efluentes em função das características da mistura processada.

Capítulo VII – Conclusões Finais e Perspectivas Futuras

Conclusões do trabalho desenvolvido; identificação das principais contribuições; principais linhas orientadoras para o trabalho a desenvolver futuramente.

Um percurso efectuado na área da investigação não preconiza um fim em si mesmo, antes contextualiza o investigador numa amálgama labiríntica de caminhos que ele, deixando para trás a sua visão ciclópica do saber, se predispõe a questionar. Deste modo, não se fecham portas, delimitando de forma estanque áreas de investigação, pelo contrário, abrem-se outras que poderão conduzir a novos estádios do saber e proporcionar oportunidades de investigação inauditas.

O presente Capítulo enquadra e caracteriza de forma sinóptica o trabalho de pesquisa efectuado. Deste modo, clarifica-se o percurso a seguir para que o estado de conhecimento aqui atingido não permaneça suspenso, constituindo antes um ponto de partida para novos estádios de conhecimento.

As empresas que operam no sector petrolífero, dada a natureza das matérias-primas com que lidam, estão conscientes da sua responsabilidade na gestão do impacte das suas actividades. Deste modo, o cumprimento das metas ambientais por elas definidas, requer um comportamento pró-activo, que passa por antecipar problemas de modo a estar preparado para os enfrentar, sabendo exactamente como actuar nessa contingência.

Actualmente não se conhecem modelos de previsão/simulação da qualidade do efluente resultante em função das diferentes características das misturas de petróleos brutos que entram na Refinaria para serem processados. O presente trabalho dá uma contribuição para a resolução destes problemas, tendo-se recorrido a técnicas e metodologias da Área Científica da Inteligência Artificial e da Descoberta de Conhecimento em Base de Dados, utilizando, na fase de *Data Mining*, Árvores de Decisão e Indução de Regras.

Os modelos desenvolvidos permitiram relacionar o tipo de efluente resultante tanto com a composição das misturas de crudes, como com o crude predominante nas misturas

processadas. Este conhecimento, implícito na BD, pode ser utilizado como fonte de informação estratégica quer a nível ambiental quer a nível económico.

Do ponto de vista ambiental, os modelos obtidos permitem prever, antes da tomada de decisão, quais os crudes a adquirir para refinação, o tipo de efluente resultante e, com isso, minimizar o impacto das descargas sobre o ambiente. A nível económico, o conhecimento prévio do tipo de efluente resultante, permite evitar penalizações e situações de incumprimento, bem como fazer uma melhor gestão dos custos de produção, uma vez que permite incluir uma variável relacionada com os custos do tratamento de efluentes.

Os modelos obtidos permitem, por um lado, ultrapassar os tempos de latência entre a aquisição das matérias-primas, o momento de amostragem e os resultados das análises laboratoriais. Por outro lado, podem constituir-se como instrumentos de gestão de uma empresa do sector petrolífero, abrindo caminhos para uma Gestão de Excelência, levando à adopção de atitudes que minimizem os impactes ambientais, não apenas ao nível das operações como, também, ao nível da optimização das descargas que serão lançados no Ambiente, as quais devem não somente respeitar a legislação em vigor como, acima de tudo, manter o equilíbrio do Ecossistema.

O trabalho elaborado e apresentado na presente dissertação abre espaço para o desenvolvimento de ferramentas automáticas de suporte à decisão com recurso à programação em lógica estendida que se prevê que venha a reforçar respostas ecológicas.

O trabalho futuro que se perspectiva visa a criação de um Universo Virtual e Dinâmico, constituído por entidades complexas e sociabilizáveis tais como genes, genomas, organismos e populações. Tais entidades complexas e interactivas competem entre si, estando sujeitas a regimes de selecção, tendo em vista modelar e prever a qualidade dos efluentes resultantes do processo de refinação.

Os cenários de evolução do sistema, serão obtidos recorrendo à Computação Genética e Evolucionária, cujo regime de selecção assentará sobre um processo de quantificação da qualidade da informação emanada de um programa em lógica que descreve o sistema, em que o ambiente computacional será construído em torno de um processo de Raciocínio Baseado em Casos.

No final, uma agência para a integração, difusão e arquivo de dados irá emergir, a qual será utilizada para alicerçar um protótipo de um Sistema Multi-Agente para Modelação e Simulação da Qualidade dos Efluentes da Indústria Petrolífera, que se possa afirmar como

um Sistema de Apoio à Decisão e possa contribuir para a tomada da melhor decisão em tempo útil.

No decurso deste trabalho foi efectuada uma comunicação oral na "*Industrial Simulation Conference' 2009 - ISC' 2009*" que decorreu em Loughborough entre 1 e 3 de Junho, dando origem à seguinte publicação:

Fernandes, A. V., Vicente, H. & Neves, J., Solving Challenging Problems in the Oil Industry Using Artificial Intelligence Based Tools. Em Diganta B. Das, V. Nassehi & L. Deka Eds., *Proceedings of Industrial Simulation Conference' 2009 - ISC' 2009*, 325-331, Eurosis - ETI Publication, Loughborough, U.K. 2009.

Referências Bibliográficas

A

- [Abelha, 2004] Abelha, A., *Sistemas Multiagente como Suporte a Trabalho Cooperativo em Unidades Hospitalares*, Tese de Doutoramento, Departamento de Informática, Universidade do Minho, Braga, Portugal, 2004.
- [Adriaans & Zantinge, 1996] Adriaans, P. & Zantinge, D., *Data Mining*. Addison-Wesley, Edimburgo, 1996.
- [Ali et al., 2005] Ali, M., Ali, B., Speight, J., *Handbook of Industrial Chemistry - Organic Chemicals*, McGraw-Hill, New York, U.S.A., 2005.
- [Apte et al., 2002] Apte, C., Liu, B., Pednault, E. & Smith, P., *Business applications of Data Mining*. *Communications of the ACM*, 45: 49-53, 2002.
- [Atkins et al., 2007] Atkins, J. P., Burdon, D. & Allen, J. H., *An application of contingent valuation and decision tree analysis to water quality improvements*. *Marine Pollution Bulletin*, 55: 591-602, 2007.

B

- [Berry & Linoff, 2000] Berry, M. & Linoff, G., *Mastering Data Mining: The Art and Science of Customer Relationships Management*. John Wiley & Sons, New York, U.S.A., 2000.
- [Berson et al., 2000] Berson, A., Smith, S. & Thearling, K., *Building Data Mining Applications for CRM*. Mc-Graw Hill, USA, 2000.
- [Braga & Morgado, 2007] Braga, J. & Morgado, E., *Guia do Ambiente - Empresas, Competitividade e Desenvolvimento Sustentável*. Monitor, Lisboa, Portugal, 2007.
- [Bramer, 2007] Bramer, M., *Principles of Data Mining*. Springer-Verlag, Londres, U.K., 2007.

C

- [Carapeto, 1999] Carapeto, C., *Poluição das Águas*. Universidade Aberta, Lisboa, Portugal, 1999.
- [Chang et al., 2001] Chang, G., Healey, M., McHugh, J. & Wang, J., *Mining the Word Wide Web - An Information Search Approach*. Kluwer Academic Publishers, USA, 2001.
- [Cheremisinoff, 2002] Cheremisinoff, N., *Handbook of Water and Wastewater Treatment Technologies*. Butterworth-Heinemann, Boston, U.S.A., 2002.

- [Cios et al., 2007] Cios, K. J., Pedrycz, W., Swiniarski, R. W. & Kurgan, L. A., *Data Mining a Knowledge Discovery Approach*. Springer, New York, U.S.A., 2007.
- [Coelho & Costa, 2007] Coelho, P. & Costa, M., *Combustão*. 1ª Ed., Orion, Lisboa, Portugal, 2007.
- [Cole, 1994] Cole, G. M., *Assessment and Remediation of Petroleum Contaminated Sites*. CRC Press, Boca Raton, U.S.A., 1994.
- [Cortez, 2004] Cortez, P., *Aprendizagem e Avaliação de Modelos - Apontamentos Pedagógicos*. Departamento de Sistemas de Informação, Universidade do Minho, Guimarães, Portugal. 2004.
- [Cunha et al., 1980] Cunha, L. V., Gonçalves, A. S., Figueiredo, V. A. & Lino, M., *A Gestão da Água. Princípios fundamentais e sua aplicação em Portugal*. Fundação Calouste Gulbenkian, Lisboa, Portugal, 1980.

E

- [Eaton et al., 2005] Eaton, A., Clesceri, L., Rice, E. & Greenberg, A. Eds., *Standard Methods for the Examination of Water and Wastewater*. 21ª Ed., American Public Health Association, USA, 2005.

F

- [Fayyad, 1997] Fayyad, U., *Mining Databases: Towards Algorithms for Knowledge Discovery*. *IEEE Computer Society Technical Committee on Data Engineering*, 21: 39-48, 1997.
- [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., Smith, P. & Uthurusamy, R., *Advances in Knowledge Discovery and Data Mining*. AAAI Press/MIT Press, Massachusetts, USA, 1996.
- [Fernandes, 2007] Fernandes, A., *Manual de Operações do Sistema do Sistema de Tratamento de Efluentes*. Refinaria de Sines, Sines, Portugal. 2007.

G

- [Galp-Sines, 1999] Refinaria de Sines, *Plano de Emergência Interno*. Petrogal, S.A - Refinaria de Sines. 1999.
- [Galp, 2007a] Galp Energia, *Referencial Interno do Sistema de Gestão de Segurança, Saúde e Ambiente da Galp Energia*. Petróleos de Portugal, Petrogal SA. 2007a.
- [Galp, 2007b] Galp Energia, *Norma Regulamentar da Política da Qualidade*. Petróleos de Portugal, Petrogal SA. 2007b.

-
- [Gary & Handwerk, 2001] Gary, J. H. & Handwerk, G. E., Petroleum Refining Technology and Economics. 4ª Ed., Marcel Dekker, Inc, New York, U.S.A., 2001.
- [Goebel & Gruenwald, 1999] Goebel, M. & Gruenwald, L., A survey of Data Mining and knowledge discovery software tools. *ACM SIGKDD Explorations*, 1: 20-33, 1999.
- [Gomes & Alves, 2007] Gomes, J. S. & Alves, F. B., O Universo da Indústria Petrolífera. Fundação Calouste Gulbenkian, Lisboa, Portugal, 2007.

H

- [Han & Kamber, 2006] Han, J. & Kamber, M., Data Mining: Concepts and Techniques. 2ª Ed., Morgan Kaufmann Publishers, San Francisco, U.S.A., 2006.
- [Haykin, 1999] Haykin, S., Neural Networks - A Comprehensive Foundation. 2ª Ed., Prentice-Hall, New Jersey, U.S.A., 1999.

K

- [Kohavi & Provost, 1998] Kohavi, R. & Provost, F., Glossary of Terms. *Machine Learning*, 30: 271-274, 1998.
- [Kufirin, 1997] Kufirin, R., Generating C4.5 Productions Rules in Parallel. *Proceedings of 4th National Conference on Artificial Intelligence (AAAI-97)*, Rhode Island, U.S.A., pp. 565-570, AAAI Press, 1997.

L

- [Langley & Simon, 1995] Langley, P. & Simon, H., Applications of Machine Learning and Rule Induction. *Communications of the ACM*, 38: 54-64, 1995.
- [Lehr & Keeley, 2005a] Lehr, J. H. & Keeley, J. Eds., Water Encyclopedia: Oceanography; Meteorology; Physics and Chemistry; Water Law; and Water History, Art and Culture. John Wiley & Sons, Hoboken, New Jersey, USA, 2005a.
- [Lehr & Keeley, 2005b] Lehr, J. & Keeley, J. Eds., Water Encyclopedia: Domestic, Municipal, and Industrial Water Supply and Waste Disposal. John Wiley & Sons, Hoboken, U.S.A., 2005b.
- [Lehr et al., 2005] Lehr, J. H., Keeley, J. & Lehr, J. Eds., Water Encyclopedia: Water Quality and Resource Development. 1ª, John Wiley & Sons, Hoboken, New Jersey, USA, 2005.
- [Lewis et al., 1999] Lewis, K. R., Daane, M. L. & Schelling, R., Processing corrosive crude oils. *Proceedings of Corrosion 99, Research Topical Symposium*, Paper nº 377, NACE International, 1999.

[Little, 1992] Little, R., Regression with missing X's: a review. *Journal of the American Statistical Association*, 87: 1227-1237, 1992.

M

[Matar & Hatch, 2000] Matar, S. & Hatch, L., Chemistry of Petrochemical Processes. 2ª Ed., Gulf Publishing Company, Houston, U.S.A., 2000.

[Meyers, 1996] Meyers, R. A. Eds., Handbook of Petroleum Refining Processes. 2ª Ed., McGraw-Hill, New York, U.S.A., 1996.

[Michalski et al., 1998] Michalski, R., Bratko, I. & Miroslav, K., Machine Learning and Data Mining Methods and Applications. John Willey & Sons, Londres, U.K., 1998.

N

[Nemerov, 1977] Nemerov, L. N., Aguas Residuales Industriales - Teorias Aplicaciones y Tratamiento. H. Blume Ediciones, Madrid, Espanha, 1977.

P

[Patterson, 1996] Patterson, D., Artificial Neural Networks - Theory and Application. Prentice Hall, New York, U.S.A., 1996.

[Pinto et al., 2009] Pinto, A., Fernandes, A. V., Vicente, H. & Neves, J., Optimizing Water Treatment Systems Using Artificial Intelligence Based Tools. Em C.A. Brebbia Eds., Water Resources Management IV, WIT Transactions on Ecology and the Environment, Vol. 35, pp. 523-533, WIT Press, Southampton, UK, 2009.

[Pyle, 1999] Pyle, D., Data Preparation for Data Mining. Morgan Kaufmann, San Francisco, USA, 1999.

Q

[Quinlan, 1986] Quinlan, J. R., Induction of decision trees. *International Journal of Machine Learning*, 27: 221-234, 1987.

[Quinlan, 1987] Quinlan, J. R., Simplifying decision trees. *International Journal of Machine Studies*, 27: 221-234, 1987.

[Quinlan, 1993] Quinlan, J., C4.5 Programs for Machine Learning. Morgan Kaufmann Publishers Inc, USA, 1993.

[Quinlan, 1996] Quinlan, J. R., Bagging, boosting and C4.5. Em Eds., *Proceedings of AAAI'96 National Conference on Artificial Intelligence*, 725-730, AAAI Press, Menlo Park, USA, 1996.

R

[Ramage, 2003] Ramage, J., *Guia da Energia. Monitor - Projectos e Edições*, Lda, Lisboa, Portugal, 2003.

[RARISA, 2007] RARISA, Regulamento de Recolha e Tratamento de Água Residual Industrial do Sistema de Santo André. *Águas de Santo André - Grupo Águas de Portugal*. 2007.

[Reis, 1996] Reis, J. C., *Environmental Control in Petroleum Engineering*. Gulf Publishing Company, Houston, U.S.A., 1996.

[Rich & Knight, 1991] Rich, E. & Knight, K., *Artificial Intelligence*. 2ª Ed., McGraw-Hill Inc., New York, U.S.A., 1991.

[Rubin, 1996] Rubin, D. B., Multiple Imputation after 18 years. *Journal of the American Statistical Association*, 91: 473-489, 1996.

[Russell & Norvig, 2003] Russell, S. J. & Norvig, P., *Artificial Intelligence: A Modern Approach* 2ª Ed., Prentice Hall, New Jersey, U.S.A., 2003.

S

[Santos et al., 2005] Santos, M. F., Cortez, P., Quintela, H., Neves, J., Vicente, H. & Arteiro, J., Ecological Mining - A Case Study on Dam Water Quality. Em A. Zanasi, C.A. Brebbia & N.F.F. Ebecken Eds., *Data Mining VI - Data Mining, Text Mining and their Business Applications*, WIT Transactions of Information and Communication Technologies, Vol. 35, pp. 523-531, WIT Press, Southampton, UK, 2005.

[Schafer, 1999] Schafer, J., *Analysis of Incomplete Multivariate Data*. Chapman & Hall / CRC, Boca Raton, U.S.A., 1999.

[Schapire, 2002] Schapire, E. R., The boosting approach to machine learning: An overview. *Proceedings of MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[Serralheiro, 2001] Serralheiro, R., A Sustentabilidade do Regadio e a Conservação do Solo e da Água em Condições Mediterrâneas. *Anais da Universidade de Évora* 10-11: 209-248, 2001.

[Slavcheva, E. et al., 1998] Slavcheva, E., Shone, B., Turnbull, A., AAA Factors controlling naphthenic acid corrosion. *Corrosion* 98, 579: 19, 1998.

[Sousa, 2004] Sousa, C., *Data Mining: Metodologias, Tecnologias, Modelos e Aplicações*, Dissertação de Mestrado, Universidade do Minho, 2004.

[Speight, 2001] Speight, J. G., Handbook of Petroleum Analysis. John Wiley & Sons, New York, U.S.A., 2001.

T

[Taniar, 2008] Taniar, D. Eds., Data Mining and Knowledge Discovery Technologies. IGI Publishing, Hershey, New York, U.S.A., 2008.

[Teixeira, 2005] Teixeira, H. Q., Sistemas de Conhecimento Baseados em Data Mining: Aplicação à Análise da Estabilidade de Estruturas Metálicas, Dissertação de Mestrado, Universidade do Minho, Braga, Portugal, 2005.

[Thuraisingham, 1999] Thuraisingham, B., Data Mining Technologies, Techniques, Tools and Trends. CRC Press LLC, USA, 1999.

[Turban et al., 2004] Turban, E., Aronson, J. E. & Liang, T.-P., Decision Support Systems and Intelligent Systems. 7ª Edição, Prentice Hall, New Jersey, USA, 2004.

U

[URL1] <http://www.rulequest.com/see5-info.html> (Visitado em 4 de Janeiro de 2009).

V

[Vicente, 2004] Vicente, H., Especificação e Prototipação de Sistemas de Gestão e Controlo da Qualidade da Água de Albufeiras, Departamento de Química, Universidade de Évora, Portugal, 2004.

W

[Weiss & Provost, 2001] Weiss, G. & Provost, F., The effect of class distribution on classifier learning: an empirical study. Technical Report ML-TR-44, Department of Computer Science, Rutgers University. 2001.

Z

[Zhang, 2007] Zhang, C., Fundamentals of Environmental Sampling and Analysis. John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2007.

Anexo I

Conjunto de Regras Extraídas a partir da Árvore de Decisão - Modelo I

Neste Anexo apresenta-se o conjunto de regras gerado a partir da Árvore de Decisão obtida para o Modelo I. No final da apresentação de cada uma das regras destaca-se o suporte (i.e. o número de casos onde a referida regra é encontrada) e a confiança (i.e. é a probabilidade condicional dessa regra).

i) REGRAS PARA O GRUPO PROIBITIVO (P)

REGRA 1

Se Predominante na Mistura = Arabian Heavy
e Brass River
e Dalia e Mellitah
e Flotta
e Nemba
e Sarir
e Zuetina

Então → P
(12; 0,667)

REGRA 3

Se Predominante na Mistura = Marlim
e Yoho <= 0
e Brass River <= 3
e Mellitah <= 4
e Albacora <= 0
e Alba Condensate <= 27
e Marlim <= 42
e Es Sider <= 6
e Alba Condensate <= 7

Então → P
(2; 0,5)

REGRA 5

Se Predominante na Mistura = Marlim
e Yoho <= 0
e Brass River <= 3
e Mellitah <= 4

REGRA 2

Se Predominante na Mistura = Brega
e Albacora > 9
e Alba Condensate <= 19

Então → P
(3; 1,0)

REGRA 4

Se Predominante na Mistura = Marlim
e Yoho <= 0
e Brass River <= 3
e Mellitah <= 4
e Albacora <= 0
e Alba Condensate > 27

Então → P
(2; 1,0)

REGRA 6

Se Predominante na Mistura = Maya
e Saharan Blend > 5

Então → P

e Albacora > 0	(2; 1,0)
Então → P (2; 1,0)	
REGRA 7 Se Predominante na Mistura = Oso Condensate e Azeri Light > 5	
Então → P (2; 1,0)	
ii) REGRAS PARA O GRUPO INSATISFATÓRIO (I)	
REGRA 1 Se Predominante na Mistura = Albacora	REGRA 2 Se Predominante na Mistura = Basrah Light e Maya <= 0
e Azeri Light e Mellitah e Okono e Okwori e Sirtica e Soroosh e Urals	Então → I (6; 1,0)
Então → I (21; 0,619)	
REGRA 3 Se Predominante na Mistura = Brega e Albacora <= 9	REGRA 4 Se Predominante na Mistura = Dalia e Brega <= 4
Então → I (2; 0,5)	Então → I (10; 0,6)
REGRA 5 Se Predominante na Mistura = Es Sider e Nemba > 3	REGRA 6 Se Predominante na Mistura = Marlim e Yoho <= 0 e Brass River <= 3 e Mellitah <= 4 e Albacora <= 0 e Alba Condensate <= 27 e Marlim <= 42 e Es Sider > 6
Então → I (3; 0,667)	Então → I

REGRA 7

Se Predominante na Mistura = Marlim
 e Yoho \leq 0
 e Brass River \leq 3
 e Mellitah $>$ 4

Então \rightarrow I

(5; 0,8)

REGRA 9

Se Predominante na Mistura = Maya
 e Saharan Blend \leq 5

Então \rightarrow I

(7; 0,571)

iii) REGRAS PARA O GRUPO SATISFATÓRIO (S)**REGRA 1**

Se Predominante na Mistura = Alba Condensate
 e Antan
 e Iranian Heavy
 e Iranian Light
 e Mesa

Então \rightarrow S

(21; 0,667)

REGRA 3

Se Predominante na Mistura = Brega
 e Albacora $>$ 9
 e Alba Condensate $>$ 19

Então \rightarrow S

(2; 1,0)

REGRA 5

Se Predominante na Mistura é = Sider
 e Nemba \leq 3

Então \rightarrow S

(4; 0,75)

(2; 1,0)

REGRA 8

Se Predominante na Mistura = Marlim
 e Yoho $>$ 0

Então \rightarrow I

(3; 1,0)

REGRA 10

Se Predominante na Mistura = Oso
 Condensate
 e Azeri Light \leq 5

Então \rightarrow I

(4; 1,0)

REGRA 2

Se Predominante na Mistura = Basrah
 Light
 e Maya $>$ 0

Então \rightarrow S

(8; 0,875)

REGRA 4

Se Predominante na Mistura = Dalia
 e Brega $>$ 4

Então \rightarrow S

(2; 1,0)

REGRA 6

Se Predominante na Mistura = Marlim
 e Yoho \leq 0
 e Brass River \leq 3
 e Mellitah \leq 4
 e Albacora \leq 0

REGRA 7

Se Predominante na Mistura = Marlim

e Yoho ≤ 0

e Brass River ≤ 3

e Mellitah ≤ 4

e Albacora ≤ 0

e Alba Condensate ≤ 27

e Marlim > 42

Então $\rightarrow S$

(6; 1,0)

e Alba Condensate ≤ 27

e Marlim ≤ 42

e Es Sider ≤ 6

e Alba Condensate > 7

Então $\rightarrow S$

(2; 1,0)

REGRA 8

Se Predominante na Mistura = Marlim

e Yoho ≤ 0

e Brass River > 3

Então $\rightarrow S$

(2; 1,0)

Anexo II

Conjunto de Regras Extraídas a partir da Árvore de Decisão - Modelo II

Neste anexo apresenta-se o conjunto de regras gerado a partir da Árvore de Decisão obtida para o Modelo II. No final da apresentação de cada uma das regras destaca-se o suporte (i.e. o número de casos onde a referida regra é encontrada) e a confiança (i.e. é a probabilidade condicional dessa regra).

i) REGRAS PARA O GRUPO PROIBITIVO (P)

REGRA 1

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba ≤ 21
e Azeri Light ≤ 21
e Azeri Light ≤ 3
e Oso Condensate ≤ 3
e Alba Condensate ≤ 19
e Alba Condensate ≤ 11
e Brass River ≤ 0
e Es Sider ≤ 0
e Arabian Heavy ≤ 0
e Mellitah > 9

Então \rightarrow P
(8; 0,5)

REGRA 3

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0

REGRA 2

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba ≤ 21
e Azeri Light ≤ 21
e Azeri Light ≤ 3
e Oso Condensate ≤ 3
e Alba Condensate ≤ 19
e Alba Condensate > 11

Então \rightarrow P
(6; 0,833)

REGRA 4

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0

e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba ≤ 21
e Azeri Light ≤ 21
e Azeri Light > 3
e Es Sider ≤ 0

Então $\rightarrow P$
(5; 0,8)

REGRA 5

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina > 9

Então $\rightarrow P$
(3; 0,667)

REGRA 7

Se Iranian Heavy ≤ 0
e Slops > 6

Então $\rightarrow P$
(7; 0,429)

II) REGRAS PARA O GRUPO INSATISFATÓRIO (I)

REGRA 1

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0

e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba > 21

Então $\rightarrow P$
(2; 1,0)

REGRA 6

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate > 30

Então $\rightarrow P$
(3; 1,0)

REGRA 2

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0

e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light ≤ 21
 e Azeri Light ≤ 3
 e Oso Condensate ≤ 3
 e Alba Condensate ≤ 19
 e Alba Condensate ≤ 11
 e Brass River ≤ 0
 e Es Sider > 0

Então \rightarrow I
 (9; 0,667)

REGRA 3

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa ≤ 10
 e Slops ≤ 0
 e Alba Condensate ≤ 30
 e Zuetina ≤ 9
 e Mesa ≤ 0
 e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light ≤ 21
 e Azeri Light ≤ 3
 e Oso Condensate > 3

Então \rightarrow I
 (10; 0,8)

REGRA 5

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa ≤ 10
 e Slops ≤ 0
 e Alba Condensate ≤ 30
 e Zuetina ≤ 9
 e Mesa > 0

Então \rightarrow I
 (3; 0,667)

e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light ≤ 21
 e Azeri Light ≤ 3
 e Oso Condensate ≤ 3
 Alba Condensate ≤ 19
 e Alba Condensate ≤ 11
 e Brass River > 0

Então \rightarrow I
 (2; 1,0)

REGRA 4

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa ≤ 10
 e Slops ≤ 0
 e Alba Condensate ≤ 30
 e Zuetina ≤ 9
 e Mesa ≤ 0
 e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light > 21

Então \rightarrow I
 (2; 1,0)

REGRA 6

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa ≤ 10
 e Slops > 0

Então \rightarrow I
 (4; 0,75)

REGRA 7

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light > 10
e Brass River ≤ 0

Então $\rightarrow I$
(15; 0,867)

iii) REGRAS PARA O GRUPO SATISFATÓRIO (S)

REGRA 1

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba ≤ 21
e Azeri Light ≤ 21
e Azeri Light ≤ 3
e Oso Condensate ≤ 3
e Alba Condensate ≤ 19
e Alba Condensate ≤ 11
e Brass River ≤ 0
e Es Sider ≤ 0
e Arabian Heavy ≤ 0
e Mellitah ≤ 9

Então $\rightarrow S$
(15; 0,6)

REGRA 3

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9

REGRA 8

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica > 6

Então $\rightarrow I$
(4; 1,0)

REGRA 2

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9
e Mesa ≤ 0
e Zuetina ≤ 3
e Nemba ≤ 21
e Azeri Light ≤ 21
e Azeri Light ≤ 3
e Oso Condensate ≤ 3
e Alba Condensate ≤ 19
e Alba Condensate ≤ 11
e Brass River ≤ 0
e Es Sider ≤ 0
e Arabian Heavy > 0

Então $\rightarrow S$
(5; 0,6)

REGRA 4

Se Iranian Heavy ≤ 0
e Slops ≤ 6
e Sirtica ≤ 6
e Forties ≤ 0
e Basrah Light ≤ 10
e Mesa ≤ 10
e Slops ≤ 0
e Alba Condensate ≤ 30
e Zuetina ≤ 9

e Mesa ≤ 0
 e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light ≤ 21
 e Azeri Light ≤ 3
 e Oso Condensate ≤ 3
 e Alba Condensate > 19

Então $\rightarrow S$
 (8; 0,875)

REGRA 5

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa ≤ 10
 e Slops ≤ 0
 e Alba Condensate ≤ 30
 e Zuetina ≤ 9
 e Mesa ≤ 0
 e Zuetina > 3

Então $\rightarrow S$
 (4; 0,75)

REGRA 7

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light > 10
 e Brass River > 0

Então $\rightarrow S$
 (2; 1,0)

REGRA 9

Se Iranian Heavy > 0

Então $\rightarrow S$
 (6; 0,833)

e Mesa ≤ 0
 e Zuetina ≤ 3
 e Nemba ≤ 21
 e Azeri Light ≤ 21
 e Azeri Light > 3
 e Es Sider > 0

Então $\rightarrow S$
 (4; 1,0)

REGRA 6

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties ≤ 0
 e Basrah Light ≤ 10
 e Mesa > 10

Então $\rightarrow S$
 (4; 1,0)

REGRA 8

Se Iranian Heavy ≤ 0
 e Slops ≤ 6
 e Sirtica ≤ 6
 e Forties > 0

Então $\rightarrow S$
 (4; 1,0)

