



Universidade do Minho  
Escola de Engenharia

André Agostinho Granja da Silva Oliveira

Apoio à Decisão na  
Análise Inteligente de Reclamações





Universidade do Minho  
Escola de Engenharia

André Agostinho Granja da Silva Oliveira

Apoio à Decisão na  
Análise Inteligente de Reclamações

Dissertação de Mestrado  
Ciclo de Estudos Integrados Conducentes ao Grau de  
Mestre em Engenharia e Gestão de Sistemas de Informação

Trabalho efectuado sob a orientação do  
Professor Doutor Manuel Filipe Santos  
(Professor Associado DSI)  
Professor Doutor Álvaro Moreira da Silva  
(Professor Catedrático Convidado ICBAS)  
Professor Doutor Carlos Filipe Portela  
(Professor Auxiliar Convidado DSI)

## DECLARAÇÃO

Nome: André Agostinho Granja da Silva Oliveira

Endereço eletrónico: a61561@alunos.uminho.pt Telefone: 918223128

Bilhete de Identidade/Cartão do Cidadão: 14105589

Título da dissertação: Apoio à Decisão na Análise Inteligente de Reclamações

Orientadores:

Professor Doutor Manuel Filipe Santos

Professor Doutor Álvaro Moreira da Silva

Professor Doutor Carlos Filipe Portela

Ano de conclusão: 2015

Mestrado em Engenharia e Gestão de Sistemas de Informação

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA TESE/TRABALHO.

Universidade do Minho, \_\_\_\_/\_\_\_\_/\_\_\_\_\_

Assinatura:

## **AGRADECIMENTOS**

O desenvolvimento desta Dissertação de Mestrado marca o encerramento de mais um capítulo da minha vida acadêmica e desta grande jornada educativa e estudantil, que levou à formação da pessoa que sou hoje como estudante, cidadão e homem.

Desde já quero agradecer ao Professor Doutor Manuel Filipe Santos e ao Professor Álvaro Moreira da Silva, meu orientador e coorientador de mestrado respetivamente, aos quais devo a escolha deste tema e que sempre se mostraram disponíveis para apoiar o desenvolvimento do trabalho mas também por terem aproximado e encurtado distâncias entre Universidade do Minho e a Entidade Reguladora da Saúde.

Gostaria de agradecer também ao meu coorientador de mestrado, Professor Doutor Carlos Filipe Portela, por todo o apoio, interesse e ajuda que prestou a este projeto de investigação, orientando sempre o desenvolvimento do trabalho de forma adequado desde o ponto inicial do projeto até à sua conclusão.

Gostaria de agradecer à Entidade Reguladora da Saúde e a todas as pessoas que nesta instituição laboram por toda a cortesia e disponibilidade para comigo e para com o projeto.

Quero também agradecer a todos os meus amigos, colegas e a todas as pessoas que se cruzaram comigo neste caminho estudantil e que, de uma forma ou de outra, acabaram por me apoiar e acompanharam o desenvolvimento do trabalho nas várias etapas do mesmo.

Gostaria também de agradecer a toda a minha família que sempre me apoiou e apoiará independentemente dos caminhos que venha a percorrer.

Quero agradecer à Joana Silva por toda a simpatia, por todo o carinho, por todo o apoio e por toda a força e interesse que demonstrou pelo projeto, incentivando-me a cada dia fazer mais e melhor.

Em último lugar mas com o agradecimento maior e mais especial de todos, quero agradecer aos meus pais Agostinho Oliveira e Maria da Conceição Granja, que sempre me apoiaram sobre todas as formas e incentivaram a estudar, e aos quais devo toda a educação e formação que hoje detenho.



## RESUMO

Este trabalho enquadra-se no desenvolvimento de um projeto de dissertação de mestrado em Engenharia e Gestão de Sistemas de Informação da Universidade do Minho, sobre o tema “Análise Inteligente de Reclamações”. O tema surgiu através de uma colaboração entre um grupo de investigação da Escola de Engenharia da Universidade do Minho e a Entidade Reguladora da Saúde, ERS. Esta Entidade pública e independente foi criada pelo governo português com a missão de regular toda a atividade desenvolvida em estabelecimentos prestadores de cuidados de saúde. No seu plano de atividades, a Entidade Reguladora da Saúde conta com o desenvolvimento de análises às reclamações dos utentes, na tentativa de melhorar a prestação de serviços médicos. Visto que atualmente recebe um grande volume de reclamações, sentiu-se a necessidade de adicionar ferramentas tecnológicas à execução da atividade de análise. Isto resulta no desafio central deste projeto, cujo seu principal objetivo se concentra na análise e estudo de grandes conjuntos de informações através de técnicas de exploração de dados, nomeadamente *Data e Text Mining*. Neste projeto, é também observável o desenvolvimento de uma ontologia de dados, e o desenvolvimento de *dashboards* como forma de analisar, estudar e visualizar a informação contida nas reclamações, e a exploração da possibilidade de criar modelos de classificação e sugestão automáticos de reclamações recolhidas em unidades prestadoras de cuidados de saúde. Todo o trabalho desenvolvido foi conduzido através de uma metodologia capaz de combinar ensinamentos da metodologia de Tomada de Decisão de Simon, da metodologia *Cross Industry Standard Process for Data Mining (CRISP-DM)* e da metodologia de investigação *Design Science Research*.

**Palavras-chave:** Sistemas de Apoio à Decisão, Text Mining, Análise Inteligente de Reclamações, Information Retrieval, Sistemas de Informação na Saúde;





## **ABSTRACT**

This work was written within the scope of a Master's Thesis project in the course of Engineering and Management of Information Systems at the University of Minho, about the thematic of "Complaints Intelligent analysis". The theme was originally created through a collaboration between a research group from the Engineering School at the University of Minho and the Health Regulation Authority. This public and independent entity was originally created by the Portuguese Government, with the mission of regulating all activities developed in health care provider establishments. The Health Regulation Authority, in its activity plan, relies on the development of analyzes to complaints from users, in an attempt to improve the provision of medical services. At the moment, the Health Regulation Authority receives a large volume of complaints, and it is necessary to add technological tools for the implementation of the analysis activity. The project aim is to analyze massive sets of information using the most innovative technologies and techniques of data exploration, including Data and Text Mining. In this project, it is also observed the development of an oriented complaints ontology, and the development of dashboards as tools to analyze, study and explore the possibility of information visualization. This study is focused on the possibility of developing automatic classification models and suggestion models of complaints collected in health care establishments. All work was developed by a combined methodology including acknowledgements of the Decision Making Methodology by Simon, the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology and the Design Science Research Methodology.

**KEYWORDS:** Decision Support Systems, Text Mining, Intelligent Analysis of Complaints, Information Retrieval, Health Information Systems;



# ÍNDICE

Agradecimentos.....	v
Resumo.....	vii
Abstract.....	ix
Índice.....	xi
Lista de Figuras.....	xv
Lista de Tabelas.....	xvii
Lista de Abreviaturas, Siglas e Acrónimos.....	xix
1. Introdução.....	1
1.1 Enquadramento e Motivação.....	1
1.2 Objetivos.....	3
1.3 Organização da Dissertação.....	4
2. Metodologias, Métodos e Materiais.....	7
2.1 Metodologias de Investigação.....	7
2.1.1 Design Science Research.....	7
2.1.2 <i>Cross Industry Standard Process for Data Mining</i> .....	9
2.1.3 Modelo de Tomada de Decisão.....	10
2.1.4 Metodologia Utilizada.....	11
2.2 Ferramentas Utilizadas.....	12
3. Estado de Arte.....	15
3.1 ERS - Entidade Reguladora da Saúde.....	15
3.2 Reclamações em Estabelecimentos de Saúde.....	16
3.3 Sistemas de Informação.....	18
3.4 Sistemas de Informação na Saúde.....	19
3.5 Descoberta de Conhecimento em Bases de Dados.....	22
3.6 <i>Data Mining</i> .....	24
3.7 <i>Text Mining</i> .....	26
3.7.1 Extração.....	29

3.7.2	Análise de Distribuição.....	29
3.8	<i>Information Retrieval</i> .....	29
3.9	Trabalhos Relacionados na Área.....	31
3.9.1	<i>FreshDesk</i> .....	31
3.9.2	<i>Zoho Support</i> .....	31
3.9.3	<i>NABD</i> .....	32
3.9.4	<i>UsefulFeedback</i> .....	32
3.9.5	<i>Complaints Management System</i> .....	32
4.	Aquisição de Conhecimento.....	33
4.1	Recolha e Estudo dos Dados.....	33
4.2	Caracterização dos Intervenientes.....	36
4.3	Classificação do Problema.....	38
5.	Desenho da Solução.....	39
5.1	Orientação de Tarefas.....	39
5.2	Desenvolvimento da Ontologia.....	39
5.3	Descrição da Ontologia.....	42
5.4	Seleção de Dados para Análise.....	46
6.	Desenvolvimento da Solução.....	49
6.1	Desenvolvimento do Estudo no Knime.....	49
6.2	Análise de Palavras com o Knime.....	51
6.2.1	Aplicação ao Total de Palavras.....	51
6.2.2	Aplicação à Tipificação “Outros Assuntos”.....	53
6.2.3	Aplicação à Tipificação “Qualidade nos Cuidados de Saúde”.....	55
6.2.4	Aplicação à Tipificação “Direitos dos Utentes”.....	57
6.2.5	Aplicação à Tipificação “Tempos de Espera”.....	59
6.2.6	Aplicação à Tipificação “Livro de Reclamações”.....	60
6.2.7	Aplicação à Tipificação “Taxas Moderadoras”.....	61
6.2.8	Aplicação à Tipificação “Acesso”.....	62
6.2.9	Aplicação à Tipificação “Rejeição Infundada”.....	63

6.2.10	Aplicação à Tipificação “Qualidade” .....	64
6.2.11	Aplicação à Tipificação “Faturação” .....	65
6.3	Processo de Criação de <i>Dashboards</i> .....	66
6.4	Análise da Informação Contida nas <i>Dashboards</i> .....	66
6.4.1	Análise do Número de Reclamações por Ano de Criação .....	67
6.4.2	Análise de Numero de Reclamações por Ano e Tipologia .....	68
6.4.3	Análise da Criação de Reclamações por Meses .....	69
6.4.4	Análise do Número de Reclamações por Tipologia e Horas.....	70
6.5	Estudo das palavras mais usadas nas reclamações .....	71
6.5.1	Estudo de Palavras por Frequência .....	72
6.5.2	Estudo de Palavras por Tipologia.....	73
6.5.3	Estudo da Frequência de Palavras por Tipologia.....	73
6.6	Elaboração de modelos de Classificação de Reclamações .....	74
7.	Monitorização do Projeto.....	77
7.1	Tarefas de Monitorização do Projeto .....	77
7.2	Propostas de Melhoria de Dados e Condições do Estudo.....	78
8.	Conclusão .....	81
8.1	Considerações Finais .....	81
8.2	Limitações e Dificuldades .....	82
8.3	Análise de Riscos .....	84
9.	Referências.....	87
Anexo I	– Publicação Científica .....	93



## LISTA DE FIGURAS

Figura 1 - Ciclos do DSR.....	7
Figura 2 - Fases do CRISP-DM .....	9
Figura 3 - Fases da Tomada de Decisão.....	11
Figura 4 - Processo de DCBD.....	24
Figura 5 - Processo de TM .....	28
Figura 6 - Modelo Relacional de Base de Dados .....	34
Figura 7 - Diagrama das Classes da Ontologia Reclamações.....	41
Figura 8 - Taxonomia da Ontologia Reclamações.....	42
Figura 9 - Diagrama da Ontologia Reclamações.....	43
Figura 10 - Classe Documento da Ontologia.....	44
Figura 11 - Classe Pessoas da Ontologia.....	44
Figura 12 - Classe Medidas da Ontologia.....	45
Figura 13 - Classe Estado da Ontologia .....	46
Figura 14 - Campos do Dataset Desenvolvido.....	47
Figura 15 - Dataset Utilizado no Knime .....	49
Figura 16 - Modelo Utilizado no Knime.....	49
Figura 17 - Nuvem das 20 Palavras mais Frequentes .....	53
Figura 18 - Nuvem de Palavras da Tipificação "Outros Assuntos" .....	54
Figura 19 - Nuvem de Palavras da Tipificação "Qualidade nos Cuidados de Saúde" .....	56
Figura 20 - Nuvem de Palavras da Tipificação "Direitos dos Utentes".....	58
Figura 21 - Quadro Geral de Dashboards .....	67
Figura 22 - Dashboards Distribuição de Reclamações por Ano.....	67
Figura 23 - Dashboard Distribuição de Reclamações por Ano e Tipologia .....	68
Figura 24 - Distribuição de Reclamações por Meses em 2014.....	69
Figura 25 - Distribuição de Reclamações por Meses em 2015.....	69
Figura 26 - Distribuição de Reclamações por Horas e Tipologia .....	70
Figura 27 - Quadro Geral de Dashboards para Estudo de Palavras.....	71
Figura 28 - Dashboard Frequência Total de Palavras .....	72

Figura 29 - Dashboard de Palavras por Tipologia.....	73
Figura 30 - Dashboard Distribuição Geral de Palavras por Tipologia.....	74



## LISTA DE TABELAS

Tabela 1 - Metodologia Combinada .....	11
Tabela 2 - Ferramentas Utilizadas .....	12
Tabela 3 - Modelos de DM .....	26
Tabela 4 - Top 20 Palavras mais Frequentes .....	51
Tabela 5 - Top 20 Palavras da Tipificação "Outros Assuntos" .....	53
Tabela 6 - Top 20 Palavras da Tipificação "Qualidade nos Cuidados de Saúde" .....	55
Tabela 7 - Top 20 Palavras da Tipificação "Direitos dos Utentes" .....	57
Tabela 8 - Top 20 de Palavras da Tipificação "Tempos de Espera" .....	59
Tabela 9 - Top 20 de Palavras da Tipificação "Livro de Reclamações" .....	60
Tabela 10 - Top 20 de Palavras da Tipificação "Taxas Moderadoras" .....	61
Tabela 11 - Top 20 de Palavras da Tipificação "Acesso" .....	62
Tabela 12 - Top 20 de Palavras da Tipificação "Rejeição Infundada" .....	63
Tabela 13 - Top 20 de Palavras da Tipificação "Qualidade" .....	64
Tabela 14 - Top 20 de Palavras da Tipificação "Faturação" .....	65
Tabela 15 - Análise de Riscos .....	84



## **LISTA DE ABREVIATURAS, SIGLAS E ACRÓNIMOS**

BI – Business Intelligence

CRISP-DM – Cross Industry Standard Process for Data Mining

DCBD – Descoberta de Conhecimento em Bases de Dados

DM – Data Mining

DSC – Design Science Research

DW – Data Warehouse

ERS – Entidade Reguladora da Saúde

HW – Hardware

ID – Identificador Numérico

IE – Information Extraction

IR – Information Retrieval

KMIS – Knowledge Management and Information Sharing

ML – Machine Learning

NLP – Natural Language Processing

REC - Reclamação

SI – Sistemas de Informação

SIS – Sistemas de Informação na Saúde

SNS – Serviço Nacional de Saúde

SQL – Structured Query Language

SW – Software

TDS – Tomada de Decisão de Simon

TM – Text Mining



# 1. INTRODUÇÃO

## 1.1 Enquadramento e Motivação

Portugal, este país situado no extremo sul europeu constituinte da Península Ibérica, conta com uma população de cerca de 11 milhões de habitantes (AKINúmeros, 2014). Como país desenvolvido que se assume, uma das grandes preocupações dos seus governantes prende-se com a saúde dos seus habitantes e a qualidade dos serviços prestados aos utentes. Devido a este facto surgiu a necessidade de se criar um plano nacional para cuidar e regular o sistema de saúde em Portugal.

O Serviço Nacional de Saúde (SNS) é um setor do ministério da Saúde, em que o seu objetivo fundamental é descrito como a responsabilidade governativa de proteger a saúde individual e coletiva da comunidade através de cuidados integrados de saúde, nomeadamente a promoção e vigilância da saúde, a prevenção de doenças e epidemias, o diagnóstico e tratamento de doenças e reabilitação médica e social (Saúde, 2014). Na persecução do seu objetivo, o SNS integra em Portugal Continental uma rede hospitalar de 212 hospitais, sendo que 91 destes são de carácter Privado e 363 centros de saúde espalhados por todo o território nacional (PORDATA, 2012).

Visto que em Portugal Continental há uma vasta rede hospitalar constituída tanto por hospitais públicos, hospitais privados como também por centros de saúde, surgiu a necessidade de criar uma entidade para regular e controlar a qualidade do serviço prestado aos utentes. Surge então assim a Entidade de Reguladora da Saúde (ERS) caracterizando-se por ser uma entidade de cariz público independente que tem como missão regular a atividade dos estabelecimentos prestadores de cuidados de saúde em todo o território nacional. Uma das responsabilidades mais trabalhosas desta entidade prende-se com a análise de reclamações, sendo que diariamente recebe um grande volume de reclamações que são classificadas e tratadas casuisticamente com base no seu conteúdo. De forma a suavizar e agilizar todo o processo, melhorando também a sua eficiência, surgiu a necessidade de recorrer a tecnologias disponíveis, nomeadamente na área do tratamento de dados, para avaliar, analisar e correlacionar informações contidas nas reclamações, de forma a criar modelos de sugestão de respostas ou recomendações de forma semiautomática, através de técnicas de *Data e Text mining*.

A área principal denominada por *Data Mining (DM)* surgiu nos anos 80 e demonstrou uma grande evolução durante os anos 90 com o surgimento de novas técnicas, sistemas, áreas de aplicação e

pesquisas científicas. Nomeadamente, o seu surgimento advém de uma popular adoção de tecnologias, bases de dados relacionais e do desenvolvimento de atividades dedicadas à pesquisa de novos e poderosos sistemas de bases de dados. Um dos principais motivos que despoletou o interesse geral nesta área científica foi o aparecimento de grandes conjuntos de dados organizados e estruturados que excediam a capacidade de análise e tratamento por parte dos recursos humanos das empresas (Han, et al., 2006). Hoje em dia, o *DM* é visto como uma oportunidade para adquirir vantagens competitivas sobre a concorrência (Hand, et al., 2001). Esta técnica mostra também vantagens na utilização eficiente dos recursos das empresas e no desenvolvimento de novos produtos e serviços que vão de encontro às necessidades dos clientes (Fayyad, et al., 1996). A exploração de bases de dados e aquisição de conhecimento tem sido reconhecido como um dos tópicos principais em sistemas de informação e inteligência artificial (Chen, et al., 1996). O processo consiste em recolher e agregar dados de um qualquer sistema de produção num *Data Warehouse (DW)*, seguindo-se um limpeza, correção de erros e tratamento de dados, de forma a estes se encontrarem estruturados e consistentes ao nível de formatação. Posteriormente os dados serão submetidos a *queries* estatísticas, redes neuronais e outras técnicas de aquisição de conhecimento de forma a descobrir padrões entre dados, relações e prever eventos futuros (Prather, et al., 1997). Torna-se interessante referir que posteriormente à fase de tratamento de dados, poder-se-á utilizar o conjunto de informações tratado para análise, submetendo-o a estratégias de visualização de informação, geralmente resultando daqui um conjunto de análises e informações interessantes e que poderão melhorar a eficácia da aplicação de outras tecnologias e auxiliar a delineação de tarefas a realizar consoante as informações recolhidas.

Com os constantes desenvolvimentos e inovação tecnológica, surgiram novas variantes especializadas na área de *DM*. Uma dessas novas variantes é o *Text Mining (TM)*, cujo seu principal objetivo é analisar bases de dados não estruturadas com vista a descobrir informações desconhecidas do público em geral (Hearst, 2003). Esta nova e excitante área de pesquisa científica assume-se como a derradeira resposta ao problema da crise de sobrecarga de informação em *DW* (Feldman & Sander, 2007). Os profissionais da área consideram o *TM* como uma área de estudo objetiva, em comparação com o *DM*, no sentido em que os recursos utilizados para a realização de qualquer estudo são documentos que contêm informações legíveis sobre os dados, e não complicados conjuntos de dados numéricos, tornando a fase de compreensão dos dados um processo muito mais simples (Weiss, et al., 2005). O *TM* pode ser utilizado com vista a gerar diferentes tipos de outputs, sendo os principais padrões de dados, relações entre palavras e previsões (Feldman & Sander, 2007). Simultaneamente à utilização de técnicas de *TM*,

por vezes os dados utilizados em estudos do género são trabalhados e apresentados com recurso a plataformas de *BI*.

## 1.2 Objetivos

Os objetivos do trabalho foram definidos, de acordo com os requisitos levantados junto das entidades relacionadas com o projeto, tendo também em consideração as referências da metodologia *CRISP-DM*, sendo este um referencial das melhores técnicas e práticas de projetos de *Data Mining*. Assim, este trabalho pretende responder à seguinte questão de investigação:

*É possível categorizar e padronizar a informação recolhida em entidades de Saúde através da utilização de técnicas de extração e análise de dados?*

De modo a responder adequadamente a esta questão foram definidos um conjunto de objetivos principais e específicos.

O objetivo principal do projeto prende-se com a exploração de dados relativos às reclamações recebidas pela Entidade Reguladora de Saúde do Porto, e analisar a viabilidade de desenvolver modelos de visualização da informação, classificação e recomendação ou resposta, aos problemas constatados nessas mesmas reclamações.

Neste sentido, em seguida serão enunciados os objetivos específicos associados a este trabalho:

- Realizar estudo conceptual do negócio a ser estudado;
- Realizar estudo intensivo dos dados a trabalhar;
- Realizar tratamento de dados;
- Criar modelos de classificação de dados (ex. ontologias);
- Determinar quais as palavras-chave utilizadas em cada tipologia de reclamação;
- Desenvolver sistema para visualização e análise de informação (*Dashboards*);
- Desenvolver modelos de recomendação;
- Desenvolver protótipo do sistema;

### **1.3 Organização da Dissertação**

Esta dissertação encontra-se estruturada por capítulos acompanhando as fases do modelo de decisão de Simon. Este modelo é composto tradicionalmente por 3 fases principais sendo elas a fase de Inteligência, a fase de Conceção e a fase Escolha, tendo também mais duas fases de extensão ao modelo denominadas por Implementação e Monitorização (Simon, 1960).

Desta forma esta dissertação encontra-se organizada em 6 capítulos maioritários. O primeiro capítulo que compõe este documento é visto como um ponto introdutório ao desenvolvimento do trabalho, sendo possível observar simultaneamente um esclarecimento do problema que será tratado durante todo o trabalho, contando também com uma componente referente às motivações e condições que envolveram o desenvolvimento do trabalho.

Seguidamente poderá ser observado uma componente referente aos materiais e métodos que acompanharam e formataram o desenvolvimento do trabalho durante toda a realização do projeto, sendo esta secção totalmente descrita no segundo capítulo desta dissertação, denominado por Metodologias, métodos e materiais.

Posteriormente no terceiro capítulo é efetuada uma análise literária para fundamentação das ideias e conceitos envolvidos no projeto. Poderemos observar neste ponto um estudo do “Estado de Arte” envolvido no projeto, proporcionando ao leitor uma clarificação de conceitos e temas envolvidos nesta dissertação.

O quarto capítulo denominado por aquisição de conhecimento está intimamente envolvido com a primeira fase do modelo de Simon referido anteriormente, a fase de Inteligência. Assim como no projeto esta fase constitui-se como importantíssima sobre o prisma de tomada de decisão, pois aqui já começamos a falar propriamente do trabalho prático de todo o estudo. É efetivamente neste ponto que poderemos observar a realização de um estudo sobre os dados informacionais que compõe todo o exercício, é realizada também uma análise de requisitos e dos intervenientes no trabalho.

Após a fase de aquisição de conhecimento, segue-se o quinto capítulo da dissertação sendo este composto pela fase de Conceção do Modelo de Simon e denominado por desenho da solução. Neste ponto, como o próprio conceito indica, é uma altura de desenho de possibilidades e estudo de várias alternativas que poderão alterar o rumo do projeto, relativamente ao estudo implícito nesta dissertação esta é uma fase de avaliação de dados e estudo de alternativas, onde é possível verificar o



desenvolvimento de uma ontologia sobre os dados e os intervenientes implícitos deste projeto, e um estudo e análise de resultados prévios do projeto.

No capítulo 6 desta dissertação, referente ao desenho da solução, poderá ser observado que este equivale à etapa de Escolha do Modelo de Simon, e adaptando a definição desta fase este consagra-se como sendo um ponto intermédio em que após serem estudadas diversas alternativas que poderiam alterar o rumo do projeto é escolhida uma que seja capaz de garantir fiabilidade e que seja a mais indicada para conduzir o projeto ao sucesso. Nesta dissertação poderá ser observado que este ponto consagra diversas análises sobre os resultados obtidos das decisões tomadas em fases anteriores.

O capítulo seguinte desta dissertação está inteiramente relacionado com a fase de Avaliação do Modelo de Simon ou seja a fase de monitorização assumindo a designação neste documento de monitorização do projeto, e como o próprio nome indica esta fase tece diversas análises sobre o estado do trabalho realizado, contando também com propostas de melhorias para realização de trabalhos futuros na mesma área e com condições semelhantes.

Esta dissertação é finalizada pelo capítulo de conclusão, onde são efetuadas as considerações finais sobre o trabalho desenvolvido durante todo o projeto, contendo também referências às dificuldades e limitações e riscos verificados durante a realização do projeto.



## 2. METODOLOGIAS, MÉTODOS E MATERIAIS

### 2.1 Metodologias de Investigação

No desenvolvimento deste projeto de dissertação foram abordadas três metodologias. A metodologia de investigação que será adotada é o *Design Science Research* (DSR) juntamente com a metodologia adequada a projetos de Data Mining, *Cross Industry Standard Process for Data Mining* (CRISP-DM), sendo inclusivamente utilizada a metodologia de Tomada de Decisão de Simon. Todas as metodologias anteriormente referidas serão utilizadas neste projeto de forma combinada, garantindo assim uma metodologia de desenvolvimento capaz de agrupar as melhores práticas de todas, por forma a respeitar os ensinamentos incutidos em cada uma delas e de forma a garantir o sucesso do projeto.

#### 2.1.1 Design Science Research

Para este projeto será utilizada como metodologia de investigação a *Design Science Research*. Esta metodologia junta um vasto conjunto de técnicas e perspetivas que conduzem a pesquisa científica em sistemas de informação, e tem como especial particularidade o facto do seu principal objetivo prender-se com o desenvolvimento de artefactos que possam vir a ser aplicados pelos profissionais da área em projetos semelhantes (Peppers, 2008). Esta metodologia pode ser vista em ciclos de informação, como observado na seguinte imagem (figura 1).

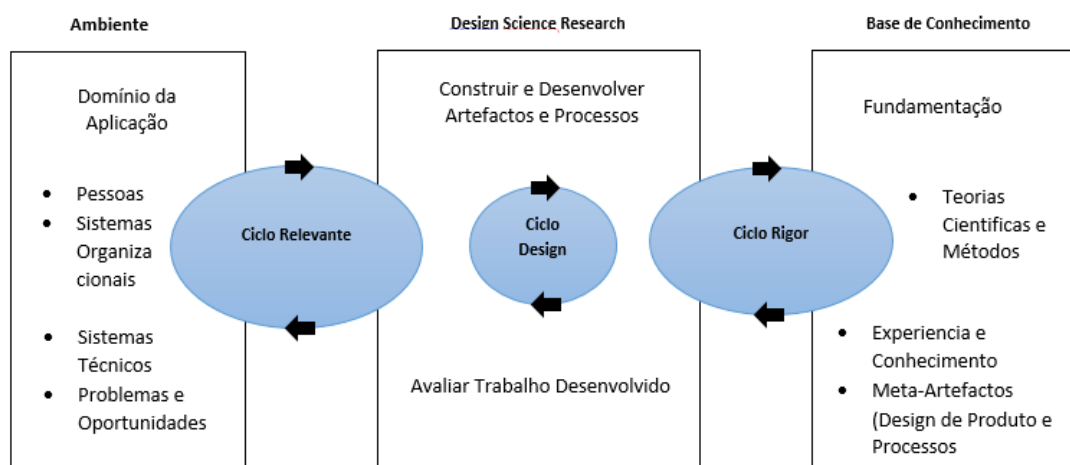


Figura 1 - Ciclos do DSR (adaptado de (Herter & Chatterjee, 2010))

Como podemos observar a metodologia é composta por 3 diferentes ciclos de pesquisa. O “Ciclo Relevante” enquadra o projeto e seus requisitos com o ambiente em que este se insere, juntamente com a atividade científica. O “Ciclo de Rigor” trata da interligação do projeto com o conhecimento científico existente na área. O ciclo central é o “Ciclo de *Design*”, trata da conciliação de toda a informação recolhida com a construção e desenho do protótipo, contribuindo também para a existência de vários pontos de controlo e de avaliação (Herver & Chatterjee, 2010).

Inicialmente o processo *DSR* começa com a identificação de um problema e com a procura de uma solução para o mesmo. Esta fase chama-se Consciencialização do Problema. Mason (Mason, 2006) defendia que neste momento inicia formal ou informalmente todo o processo de investigação. Após este ponto parte-se para uma definição de objetivos e consequente procura de uma solução, tratando-se de uma fase criativa da investigação denominada por Sugestão. Posteriormente ocorre a etapa de Desenvolvimento, onde os artefactos começam a ser construídos pelos investigadores perspetivando uma solução viável ao problema (Peffer, 2008). Após desenvolvimentos e implementação do artefacto dá-se a fase de Avaliação, onde ocorre uma análise e respetivos testes à implementação do artefacto de acordo com as condições estabelecidas para validação. Para finalizar há a fase de Conclusão onde os resultados da investigação são analisados e interpretados (Peffer, 2008) (Mattiello da Silva & Da Costa, 2013).

Como referido anteriormente, a metodologia *Design Science Research* foi utilizada como metodologia de investigação, ou seja, os contributos adquiridos desta metodologia foram aplicados sobretudo na fase de revisão literária e na procura e análise de estudos e projetos desenvolvidos dentro do mesmo âmbito que o projeto descrito nesta dissertação, sendo que a metodologia referida é utilizada durante todo o decurso do trabalho intervindo em todas as etapas estipuladas concentrando o seu contributo no desenvolvimento de artigos e artefactos descrevendo o conjunto de tarefas realizadas.

### 2.1.2 Cross Industry Standard Process for Data Mining

Para este projeto a metodologia prática que melhor se adequa é a CRISP-DM, devido ao facto de ser especialmente dedicada a projetos de Data Mining. Originalmente, esta metodologia é composta por 6 fases principais, tendo como principal vantagem, a transversalidade de aplicação em relação ao tipo de organização sobre estudo e quanto ao tipo de tecnologias e ferramentas a utilizar. Em seguida temos a figura 2, elucidativa das fases do CRISP-DM (CRISP-DM, 2014).

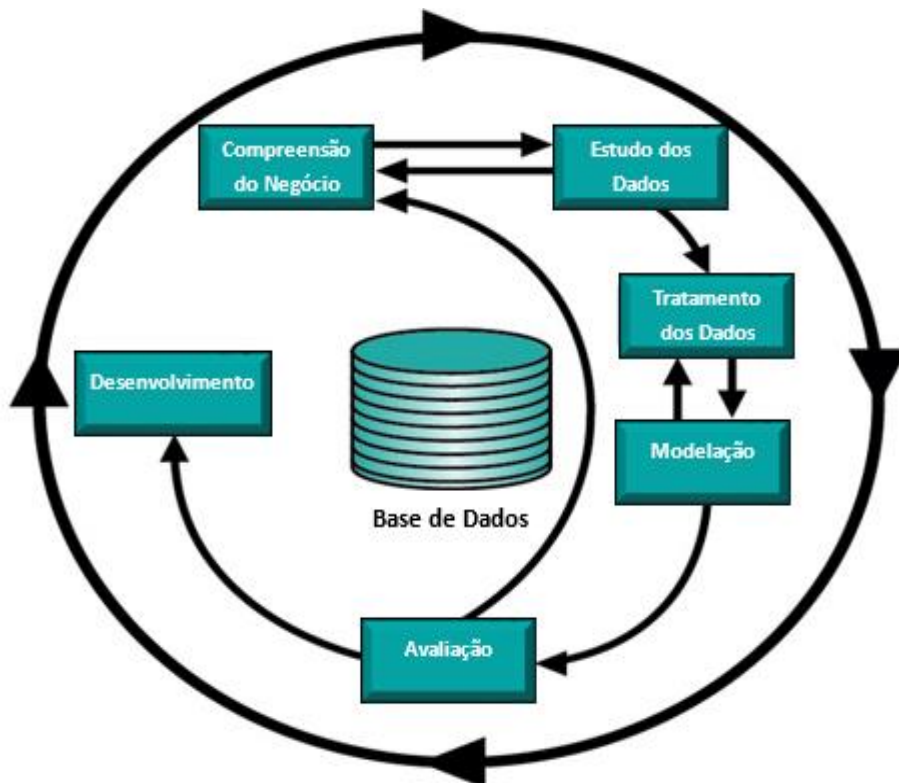


Figura 2 - Fases do CRISP-DM (adaptado (CRISP-DM, 2014))

Como este trabalho se caracteriza por ser também um projeto de *Data* e *Text Mining* serão seguidas as etapas da metodologia *CRISP-DM*, respeitando todas as indicações deste modelo. Como foi referido anteriormente, o *CRISP-DM* é constituído por 6 fases sendo a primeira denominada por “Compreensão do Negócio”, aqui será realizado um estudo conceptual do negócio, contendo também informações sobre o projeto a realizar. Posteriormente entra a fase “Estudo dos Dados”, onde deverá ser realizado um estudo intensivo dos dados que serão utilizados, deste ponto seguir-se-á a fase de “Tratamento de Dados” onde irá ser realizado o tratamento dos dados com base nas informações recolhidas anteriormente. A fase número 4 do *CRISP-DM* é a “Modelação”, esta fase é caracterizada por alcançar as respostas aos objetivos propostos e realizar os testes aos modelos. A fase seguinte denomina-se “Avaliação” e como o

próprio nome indica aqui são avaliados os resultados atingidos. Nesta fase poderá ser necessário rever todo o processo. Por fim a última fase desta metodologia é a “Implementação” onde todos os resultados obtidos serão apresentados ao público de uma forma simplificada e atrativa, ou seja, ocorre uma implementação de todo o trabalho desenvolvido em contexto real.

Através dos ensinamentos contidos na metodologia *CRISP-DM*, é possível compreender que a mesma irá contribuir para o correto desenvolvimento de cada etapa de trabalho, tendo maior influência no estudo e análise dos dados relacionados com o projeto, sendo possível que preste um grande contributo na fase de modelação e elaboração dos modelos de análise.

### 2.1.3 Modelo de Tomada de Decisão

O processo de tomada de decisão foi um modelo desenvolvido por uma das pessoas que mais contribuiu para o desenvolvimento da área da inteligência artificial, Herbert Simon (Marreiros, 2007). Este modelo segue o princípio da racionalidade limitada, e baseia-se sobretudo em três fases principais sendo estas a fase de inteligência, conceção e escolha, podendo ser aumentado pela fase de implementação e monitorização (Simon, 1960).

A fase de inteligência envolve uma identificação e consequente definição do problema. Ocorre aqui uma pesquisa e clarificação de informação sobre o problema em questão.

A fase de conceção diz respeito ao desenho de soluções para o problema encontrado, existe aqui portanto uma análise do panorama geral das várias soluções que podem dar uma resposta concreta ao problema encontrado.

A fase de escolha está relacionado com a seleção do caminho mais apropriado e mais eficaz para a resolução do problema encontrado. Neste ponto dá-se a seleção do melhor modelo que solucionará o nosso problema, sendo que esta decisão implica um estudo intensivo dos efeitos que irão resultar da sua implementação.

A fase de implementação está relacionada com a aplicação do modelo anteriormente estudado sobre o problema encontrado, caso seja detetada deficiências ou ineficácia da solução proposta, o utilizador desta metodologia deverá retornar à fase de conceção.

A fase de monitorização é a última desta metodologia e está relacionada com o controlo e avaliação dos resultados da fase de implementação.

O recurso aos ensinamentos contidos na metodologia de Tomada de Decisão de Simon são facilmente observáveis em cada uma das etapas de desenvolvimento deste projeto, sendo notória a sua presença na estruturação deste documento.

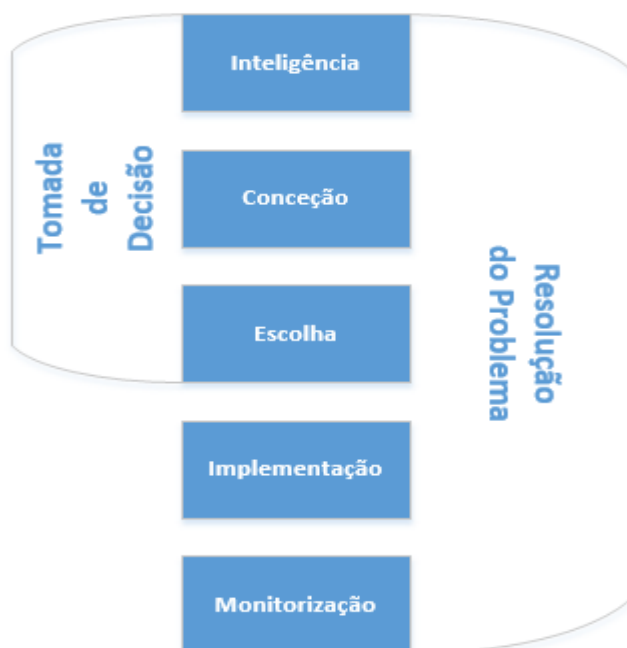


Figura 3 - Fases da Tomada de Decisão (adaptado de (Simon, 1960))

#### 2.1.4 Metodologia Utilizada

No desenvolvimento deste projeto foi decidido que a metodologia ideal para o problema em questão seria uma combinação das três metodologias apresentadas anteriormente. Desta forma a seguinte tabela (tabela 1) representa a utilização dos ensinamentos de cada metodologia em cada etapa de trabalho.

Tabela 1 - Metodologia Combinada

Metodologias		Etapa 1	Etapa 2	Etapa 3	Etapa 4	Etapa 5
<b>DSR</b>	Consciencialização	X				
	Sugestão		X			
	Desenvolvimento			X		
	Avaliação				X	
	Conclusão					X

<b>CRISP-DM</b>	Compreensão do Negócio	X				
	Estudo dos Dados		X			
	Tratamento dos Dados		X			
	Modelação			X		
	Implementação					
	Avaliação					X
<b>Tomada de Decisão</b>	Inteligência	X				
	Conceção		X			
	Escolha			X		
	Implementação					
	Avaliação					X

## 2.2 Ferramentas Utilizadas

No desenvolvimento deste projeto foi necessário recorrer à utilização de diversas ferramentas para as mais diversas etapas do trabalho. Por esta razão foi agregado na tabela 2 o conjunto de ferramentas que suportaram este projeto juntamente com a respetiva descrição de cada uma delas.

*Tabela 2 - Ferramentas Utilizadas*

<b>Oracle SQL Developer</b>	O <i>software Oracle SQL Developer</i> é uma ferramenta dedicada à simplificação da criação e gestão de base de dados. Foi utilizada neste projeto para gerir e armazenar os dados recolhidos na ERS.
<b>Microsoft Excel 2013</b>	A ferramenta <i>Microsoft Excel</i> é disponibilizada pela <i>Microsoft</i> no pacote de ferramentas utilitárias Office, esta é dedicada à exploração de dados através de folhas de cálculo, oferecendo diversos recursos aos utilizadores. Neste projeto este <i>software</i> foi utilizado sobretudo na exploração e tratamento de coleções de dados.



<p><b><i>Knime Analytics Platform</i></b></p>	<p>O <i>knime</i> é uma plataforma dedicada à exploração de dados, de cariz <i>open-source</i> esta é uma plataforma dedicada à análise de dados com a componente de estes obterem características textuais. Neste projeto o <i>knime</i> foi utilizado para exploração de dados.</p>
<p><b><i>Qlikview 11</i></b></p>	<p>O <i>QlikView</i> é uma ferramenta desenvolvida para o campo da visualização de dados, ele permite a pesquisa e exploração de grandes volumes de dados permitindo a sua visualização através de gráficos e estruturas de análise. Neste projeto o <i>QlikView</i> foi utilizado na análise exploratória de dados através do desenvolvimento de <i>Dashboards</i></p>
<p><b><i>Protégé</i></b></p>	<p>O <i>Protégé</i> é um <i>software</i> de distribuição livre e <i>open source</i> dedicado a apoiar o desenvolvimento de ontologias através de uma estrutura baseada na aquisição e estruturação de conhecimento. Desta forma o <i>Protégé</i> foi utilizado neste projeto de acordo com o efeito para que foi desenvolvido, ou seja, apoiou o desenvolvimento de uma ontologia sobre reclamações.</p>



## **3. ESTADO DE ARTE**

### **3.1 ERS - Entidade Reguladora da Saúde**

A Entidade Reguladora da Saúde (ERS) foi criada legalmente no nosso país no ano de 2003. Esta é uma entidade pública independente que adotou como missão regular toda a atividade desenvolvida em estabelecimentos prestadores de cuidados de saúde (Lobo, 2004). No setor da saúde em Portugal esta instituição é a única dotada de independência no exercício das suas funções, sendo que a necessidade que originou a sua criação prende-se com a necessidade de separar o Estado da função de regulador e supervisor, em relação às funções de operador e de financiador (Sancho & Quintas, 2009).

A ERS tem como competências cobrir a rede nacional do serviço de saúde (SNS) e a medicina privada, preocupando-se sobretudo com a defesa dos direitos dos utentes e os princípios basilares da prestação de serviço público, sendo eles a universalidade e igualdade no acesso, segurança e qualidade da prestação (Simões, 2004).

Nos dois primeiros anos após a criação desta entidade, a sua atividade foi quase inexistente devido à regulação da sua atividade e atribuição de recursos mínimos para cumprimento de funções (Entidade Reguladora da Saúde, 2011). Nos anos seguintes procederam-se a grandes mudanças estruturais provenientes de alterações legislativas, sendo que a ERS passou a deter quatro departamentos:

- Departamento de Gestão Interna, relacionada com gestão administrativa e de recursos humanos;
- Departamento de Proteção da Qualidade e Direitos dos Cidadãos, relacionada com a garantia de preservação dos direitos fundamentais dos utentes e qualificação de unidades de cuidados de saúde;
- Departamento de Acompanhamento do Sistema de Saúde e Defesa do Acesso e da Concorrência, relacionado com a proteção do acesso universal ao SNS e supervisão do mercado concorrencial de administração no setor da saúde;
- Departamento de Supervisão e Intervenção Jurídica, relacionado com cumprimento e aplicação de sanções previstas na legislação;

No âmbito das suas funções, durante os últimos anos, esta entidade tem visado o seu exercício de regulação sobretudo em duas dimensões: económica, no sentido de fixar preços, controlar a produção

nas unidades de saúde, coordenar a forma de distribuição do mercado, e a dimensão social, no sentido de humanizar serviços e controlar o cumprimento de direitos de utentes (Entidade Reguladora da Saúde, 2011).

A intervenção nestas duas dimensões tem sido efetuada através de um vasto leque de iniciativas, das quais se destacam as seguintes: diagnóstico da qualidade dos serviços públicos de saúde, avaliação dos cuidados de saúde primários, análise de queixas e reclamações dos utentes, carta dos direitos do utente dos serviços de saúde, sistema de registo das entidades reguladas, sistema de avaliação em saúde, deteção de práticas de indução artificial de procura, deteção de praticas de seleção de doentes, avaliação de praticas de transferência e referenciação de doentes, regime de licenciamento dos estabelecimentos prestadores de cuidados de saúde, regime de convenções celebrados pelos SNS, caracterização dos centros de nascimento não públicos, análise da concorrência no sector do transporte de doentes, análise da concorrência no setor da hemodiálise, informatização da informação através da criação e gestão do sitio na Internet da ERS (Entidade Reguladora da Saúde, 2013) (Almeida, 2010).

Uma dessas iniciativas executadas pela Entidade Reguladora da Saúde, nomeadamente a análise de queixas e reclamações dos utentes, deu o mote à criação deste projeto de Dissertação de mestrado.

### **3.2 Reclamações em Estabelecimentos de Saúde**

De acordo com a legislação vigente em Portugal, todos os utentes de estabelecimentos de saúde possuem o direito constitucional de apresentar queixa e reclamar formalmente sobre os serviços prestados, no momento em estes não são efetuados de forma adequada (Entidade Reguladora da Saúde, 2011). Este Livro simultaneamente é capaz de agrupar críticas construtivas, não se revendo unicamente como um livro de queixas, deve sobretudo agrupar opiniões pessoais dos utentes quanto aos serviços prestados, pode inclusivamente ser utilizado como forma de louvor ou partilha de agradecimento a uma instituição ou profissional, sendo assim utilizado com uma conotação mais positiva. Desta forma, todos os estabelecimentos de saúde sujeitos a regulação por parte da Entidade Reguladora da Saúde estão ao abrigo das disposições previstas na legislação publicada em Diário da República<sup>1</sup> (Entidade Reguladora da Saúde, 2009). Nestas diretivas governamentais está descrito que todas as entidades prestadoras de

---

<sup>1</sup> Decreto-Lei nº 156/2005, de 15 de Setembro, com as alterações introduzidas pelo Decreto-Lei nº118/2007, de 6 de Novembro e Decreto-Lei nº 371/2007, de 19 de Maio

cuidados de saúde sujeitas a regulação da ERS estão obrigadas a disponibilizar Livro de Reclamações em todos os seus estabelecimentos (Rodrigues, 2007). É também uma obrigação da entidade prestadora de cuidados de saúde facultar de forma livre e imediatamente ao utente o livro de reclamações, assim que este for solicitado. A entidade deve também afixar em todos os estabelecimentos um letreiro informando os utentes da existência de livro de Reclamações naquele local contendo também a identificação e morada da ERS, e manter por um período mínimo de três anos um arquivo de livros de Reclamações (Entidade Reguladora da Saúde, 2009).

Todos os cidadãos que enquanto utentes recorrem a estabelecimentos de saúde têm o direito constitucional fundamental a reclamar, ou seja fazer queixa sobre a forma como são tratados. Efetivamente este direito, e o de obter uma resposta clara e perceptível em matéria de saúde é uma das maiores preocupações da ERS, pelo que se revela importantíssimo o processamento adequado de todas as reclamações e exposições de desagrado por parte dos utentes. O principal objetivo é o de permitir uma melhor aferição dos pontos fracos do sistema nacional de saúde e identificação das áreas que necessitam de uma intervenção mais aprofundada.

Os estabelecimentos de saúde são obrigados a comunicar e partilhar à Entidade Reguladora uma cópia das reclamações e queixas dos utentes, constantes dos seus livros de Reclamações. Desta forma, a ERS efetua a sua análise e tratamento de todas as reclamações que lhe chegam provenientes dos livros de Reclamações. Sendo que este não é o único meio de efetuar queixas à ERS, também existem outros canais de comunicação que poderão ser utilizados, tal como via postal, *fax* e correio eletrónico (Entidade Reguladora da Saúde, 2011).

Mais recentemente, a ERS disponibilizou a todos os interessados um serviço inovador para efetuar reclamações, o Livro de Reclamações *online*. Este serviço, adequado às mais recentes normas de acessibilidade, conjugando a facilidade de consulta e de navegação com um *design* simplista, está disponível online a todos os utentes dos serviços de saúde que pretendam apresentar as suas queixas com rapidez e eficácia (Jornal de Notícias, 2008). Em 2009, a ERS recebeu um total de 7848 reclamações relacionadas com serviços de saúde, sendo que grande parte das queixas estavam relacionadas com tempos de espera e qualidade de assistência administrativa e dos cuidados de saúde. Desta forma, as reclamações são um bom indicador da resposta que o cliente espera receber da instituição de saúde. O processo de tratamento de reclamações deverá refletir as necessidades e expectativas dos utentes assim como de igual forma deverá estar de acordo com os objetivos das organizações (Firmino, 2011).

Uma gestão correta de reclamações afeta claramente o sentido de justiça e satisfação dos utentes, fortalecendo os laços de lealdade dos utentes para com a instituição em causa. Uma gestão eficaz e eficiente das reclamações poderá traduzir-se em sucesso organizacional, num mercado altamente competitivo de concorrência direta, como é observável no setor da saúde.

### **3.3 Sistemas de Informação**

Ao longo dos anos as organizações têm vindo a sofrer diversas alterações no seu ambiente, e com isto, tiveram de se adaptar da melhor forma através de novas ferramentas e técnicas para melhorar a sua posição face à concorrência. O aparecimento dos sistemas de informação (SI) nas organizações contribuiu para a melhoria do grau de competitividade, registando simultaneamente um aumento constante de transformações através do aproveitamento eficiente das tecnologias.

O conceito de Sistemas de Informação sempre reuniu grande complexidade na sua explicitação, pois ainda não existe uma ideia homogénea, sendo que cada autor defende uma perspetiva diferente. Por forma a explicar este conceito será necessário primariamente definir o que se entende por sistema e informação (Fidalgo, 2013).

Informação está relacionada com o aumento de conhecimento provocado pela retenção de ideias aquando a existência de troca de mensagens. Falkenberg (Falkenberg, et al., 1998) defendia que o conceito de informação resultava da diferença entre a interpretação das mensagens e o conhecimento prévio à receção das mesmas. A informação poderá ser interpretada de 2 formas que se completam, pode ser visto como objetos simbólicos construídos deliberadamente a fim de permitir a comunicação e formação de conhecimento, mas noutra perspetiva é também o ato de fornecer ou partilhar conhecimento, ato de informar (Carvalho, 2000).

Sistema traduz-se num conjunto de elementos interdependentes que colaboram e trabalham em conjunto de forma a alcançar um objetivo comum através de um processo organizado (O'brien, 1993).

Aliando-se a estas duas definições Carvalho (Carvalho, 2000) conclui que um sistema de informação pode ser tanto um objeto ativo que lida com (processos) informações, como também um objeto ativo cujo seu propósito é informar.

Uma das opiniões mais antigas sobre a matéria foi dada por Buckingham em 1987 (Buckingham, et al., 1987) que definia este conceito como um sistema que recolhe, processa, armazena e distribui informação no seio de uma organização, garantindo que a informação é acessível a todos os que dela

necessitem. Compreende-se aqui que um sistema de informação tem como propósito apoiar a atividade humano, podendo ser suportado por computadores (Oliveira J. , 2014).

A nível científico este conceito foi sofrendo várias alterações, registando-se um crescimento de maturidade, ao mesmo nível que o espaço temporal foi-se desenvolvendo. Em 1993, Ein-Dor e Segev (Ein-Dor & Segev, 1993) propunham a ideia que se podia entender como sistema de informação um sistema informático que possuía uma interface de interação com o utilizador. Para outros autores seria impossível dissociar o tema em questão da sua natureza social, em que viam os SI como sistemas sociais cujo comportamento é fortemente influenciado pelos objetivos, crenças e valores de indivíduos e grupos, assim como pelo desempenho da tecnologia (Oliveira J. , 2014).

Após muitas transformações ao conceito, Sá-Soares (Soares, 2006) propôs que um Sistema de Informação pode ser visto como um “sistema social que tem por finalidade apoiar a significação e ação organizacional através da síntese organizada de informação”, referindo ainda que “os colaboradores da organização são parte integrante do seu sistema de informação”.

Devido ao papel desempenhado nas atividades das organizações e na sociedade, devemos encarar os sistemas de informação como parte integrante da própria organização, determinando aquilo que a organização deve e é capaz de produzir. Desta forma, poderemos verificar a total importância dos sistemas de informação nas organizações, sentindo-se a necessidade de existir uma gestão cuidada e atenta do próprio sistema, tendo em conta as suas características e as da empresa, por forma a atingir o sucesso organizacional (Fidalgo, 2013).

### **3.4 Sistemas de Informação na Saúde**

Desde cedo, os sistemas de informação tornaram-se num ferramenta essencial para qualquer organização atingir o sucesso. Na área da saúde despertaram grande interesse devido ao seu grande potencial aplicacional.

Conceptualmente os sistemas de informação na saúde podem ser encarados como um instrumento para adquirir, organizar e analisar dados necessários à definição de problemas e riscos para a saúde pública, avaliar a eficácia, eficiência e influência da prestação de serviços na qualidade de vida das pessoas, além de contribuírem para a formação de conhecimento nesta área e outras diretamente relacionadas (Branco, 1996).

Em 1984, Peter Reichertz (Haux, 2006) traçou uma perspetiva temporal da aplicação desta tecnologia a instituições hospitalares, e desde então esta área teve um progresso tremendo derivado das constantes inovações informáticas.

O objetivo inicial da introdução dos sistemas de informação na saúde continua inalterável desde a sua primeira aplicação, sendo este muito simples e direto: contribuir para uma prestação de serviços aos utentes de uma forma bastante eficiente e com grande qualidade (Haux, 2006).

Morris Collen no seu histórico estudo (Berg, et al., 1999) afirmava categoricamente que o desenvolvimento de um sistema de informação direcionado aos hospitais não era tarefa fácil, tratando-se de um processo muito mais complexo do que o Homem aterrar na lua. Esta citação deve-se ao facto de na altura existir um grande dificuldade de adaptar os sistemas ao contexto real e devido à imprecisão de dados, indubitavelmente necessários ao desenvolvimento do sistema (Berg, et al., 1999).

Nos primeiros trabalhos desenvolvidos nesta área, os sistemas de informação hospitalares eram sobretudo dirigidos e desenvolvidos para auxiliarem as tarefas diárias dos profissionais de saúde e administrativos. Hoje em dia, reconhecemos a necessidade de incluir todos os atores presentes no ambiente hospitalar no desenvolvimento da tecnologia, incluindo as expectativas dos doentes (Haux, 2006). Devido a esta questão, as primeiras implementações informáticas em meio hospitalar obtiveram resultados desastrosos, não produzindo o efeito esperado (Heeks, 2006). Sendo que outras questões potenciaram estes resultados tal como, a estrutura altamente complexa da rede hospitalar e a natureza imprevisível da capacidade de reação humana, assim como adesões inesperadas a instituições hospitalares potenciadas por epidemias ou ocorrência de catástrofes naturais (Berg, et al., 1999).

Nas últimas duas décadas, esta temática deparou-se com grandes transformações, com efeito a experiência recolhida permite compreender que hoje em dia o desenvolvimento de um sistema de informação para a saúde deve contar com uma visão mais dinâmica quanto aos aspetos de *design* e integração do sistema, pois estes dois aspetos estão constantemente em mudança (Heeks, 2006). Aqui surge a necessidade de reunir tanto administradores, como consultores, como representantes de todos os beneficiários do sistema para uma melhor integração do sistema, tendo sempre em atenção as condições organizacionais. Neste processo deve também ter-se em atenção o desenvolvimento de conteúdo, de ferramentas, estruturação de papéis e competências de forma a sustentar todas as necessidades dos utilizadores (Heeks, 2006).

A implementação de sistemas de informação no sector da saúde trouxe diversas oportunidades tanto através da redução de erros clínicos, como na facilitação do desempenho de funções dos profissionais



da área, e através da melhoria da eficiência e qualidade dos serviços prestados, mas esta fase do processo deve ser continuamente avaliada pois caso exista erros ou o sistema falhe isso poderá causar efeitos muito negativos tanto para utentes como para os profissionais médicos (Ammenwerth, et al., 2003).

O sucesso de uma implementação informática pode ser medido através de diversos fatores tais como: eficácia, eficiência, compromisso e empenho organizacional, satisfação dos trabalhadores e dos utentes (Berg, 2001).

Uma das grandes transformações que foi observada e ajudou a impulsionar o contributo dos sistemas na saúde foi a passagem de processamento de dados em papel para processamento e armazenamento através de computadores. Hoje em dia, o poder de produção de dados e informações em centros hospitalares é algo que não deve ser subestimado (Haux, 2006). Até porque a informação é a essência da profissão nesta área. Os profissionais de saúde necessitam de informação para exercer o correto cuidado, gestão e avaliação dos seus utentes, ou seja, todas as atividades dentro de um centro hospitalar estão diretamente relacionadas com a procura e utilização de conhecimentos e informações. Desta forma, quanto melhor esse conhecimento e essas informações forem registadas, armazenadas e disponibilizadas pelos sistemas de informação melhor será a atuação médica (Marin, 2010). Neste sentido, a informatização demonstra toda a sua importância para a organização pois encurta fluxos favorecendo a comunicação entre setores, departamentos e unidades, representando uma base concreta para o processo de gestão (Benito & Licheski, 2009).

Um estudo efetuado por investigadores da Universidade do Minho, mostrou que os sistemas de informação em hospitais portugueses detêm grande importância na gestão organizacional, sendo fundamentais em toda a cadeia de valor, tendo sido constatado que, atualmente a utilização intensiva de tecnologia otimiza todo o funcionamento da organização (Martinho, et al., 2014).

Poderemos observar alguns exemplos aplicativos de tecnologia nos cuidados de saúde observando a utilização de sistemas de suporte à decisão, observando a existência de servidores que colecionam a informação e conhecimento mais recente sobre medicina, e olhando para as estações de trabalho congregadas com as mais diversas funcionalidades de apoio a profissionais de saúde (Ammenwerth, et al., 2003). Outra área que tem usufruído das vantagens da introdução tecnológica na saúde é nomeadamente o ensino, esta parceria demonstra interesse pela melhoria do processo de ensino-aprendizagem e favorece a partilha de conhecimentos, estudos, análises e materiais educacionais entre alunos e docentes (Perez & Zwicker, 2006). Um exemplo tecnológico concreto de uma ferramenta

direcionada a esta área é o *EClass*, um sistema de apoio ao ensino que permite a partilha de materiais entre professores e alunos e permite a criação de um fórum de partilha e discussão de ideias e casos médicos entre utilizadores (Perez & Zwicker, 2006). Um exemplo diferente mas de sucesso da introdução de sistemas de informação na saúde, foi o desenvolvimento do “Prontuário Electrónico”, este é um sistema de recolha de dados de doentes, que regista toda a interação do paciente com os estabelecimentos de saúde, podendo conter diagnósticos médicos, exames realizados, bem como toda a documentação para consulta acerca de um determinado utente do hospital. (Perez & Zwicker, 2006)

### **3.5 Descoberta de Conhecimento em Bases de Dados**

Atualmente, as organizações estão sujeitas a um ritmo elevado de competição e constantemente são estabelecidas novas mudanças no seu ambiente organizacional potenciado pela quebra das barreiras comerciais e o potencial oferecido pelo uso da internet e das mais recentes inovações tecnológicas. Visto isto, é claro que as organizações necessitam de obter um processo de decisão seguro, rápido e estratégico de modo a alcançar vantagens competitivas sobre os seus concorrentes mais diretos. Para adquirir vantagens competitivas é necessário ter a perfeita noção de que a informação e o conhecimento são os bens mais importantes e necessários para qualquer instituição (Maimon & Rokach, 2010).

Hoje em dia, a produção de informação já ultrapassa a capacidade de processamento e análise humana sendo necessário recorrer a ferramentas tecnológicas. Desta forma, surgiu o grande interesse pela descoberta de conhecimento em bases de dados, também provocado pela contínua descida do custo de armazenamento de dados (Maimon & Rokach, 2010).

A descoberta de conhecimento pode ser definida como um processo de extração de conhecimento e informação, previamente desconhecida mas com potencial útil, a partir da leitura de dados (Frawley, et al., 1992). Esta área associada à exploração de bases de dados surge primeiramente em 1989, tendo o seu estudo e aplicação crescido de forma significativa nas últimas duas décadas. Vários autores consagrados definiram que, através do recurso a modelos e técnicas de *Data Mining*, o principal objetivo desta área depara-se com a extração de conhecimento útil a partir de grandes coleções de dados, através da identificação e reconhecimento de padrões e tendências. Desta forma, revela-se necessário no momento da utilização destas técnicas o uso de dados e informações concretas, minimizando a ocorrência de erros, não prejudicando o processo de tomada de decisão por parte de gestores e

consultores (Piateski-Shapiro & Frawley, 1991) (Fayyad, et al., 1996) (Santos & Azevedo, 2005) (Santos & Ramos, 2006) (Portela, 2009).

Esta área, cada vez mais, tem vindo a despertar atenção nos mais diversos contextos, patrocinado largamente pelo seu potencial organizacional, ressaltando vários estudos e aplicações tanto em contexto académico, científico e social.

O processo de Descoberta de Conhecimento em Bases de Dados é composto por cinco fases (Maimon & Rokach, 2010):

- **Seleção** – Nesta primeira fase deve ser definido claramente os objetivos do projeto, e depois selecionar os dados a utilizar. Esta seleção incluiu a descoberta do tipo de dados que estão disponíveis e integra-los num único *dataset*.
- **Pré-Processamento** – Na fase de pré-processamento a fiabilidade dos dados deve ser sempre uma preocupação. Esta fase é composta por tarefas como limpar dados e gerir ou remover dados em falta ou incoerentes, existe aqui uma espécie de filtragem à base de dados, para que esta contenha apenas dados interessantes para o projeto.
- **Transformação** – Nesta fase o expectável é reduzir a quantidade e projeção dos dados através da aplicação de técnicas uteis para a representação dos dados.
- **Data Mining** – Esta fase consiste na aplicação da análise de dados através de algoritmos de *Data Mining* para tentar encontrar padrões e relações globais nos dados.
- **Avaliação e Interpretação** – Para finalizar existe uma avaliação e interpretação dos resultados obtidos no processo de *Data Mining*, tendo em conta os objetivos definidos inicialmente. Nesta fase deve ser avaliada a compreensibilidade e utilidade dos modelos induzidos, tal como documentar a descoberta de conhecimento realizado.

Neste processo de descoberta de conhecimento em bases de dados (figura 4) a fase que mais se destaca é o *Data Mining*, devido ao facto de estar no núcleo de todo o processo confundindo-se por vezes com ele próprio. O termo *Data Mining* é mais utilizado pelos estatísticos, analistas de dados e pela comunidade dos gestores de Sistemas de Informação, enquanto o termo DCBD é mais utilizado pelos investigadores da área de Inteligência Artificial e de *Machine Learning* (Ramos & Lobo, 2003).

Além de iterativo, devido ao facto de poder existir um retrocesso para as etapas anteriores durante toda a execução do processo, a descoberta de conhecimento é também considerado um processo interativo, já que requer a participação do utilizador sempre que é necessário tomar uma decisão.

Normalmente, a fase de *Data Mining* representa cerca de 20% do tempo e esforço gasto durante todo o processo, sendo a única que é melhor interpretada e suportada automaticamente através de *software*. Todas as restantes fases constituem-se mais como uma questão de “arte”, do que por uma rotina que possa ser automatizada (Ramos & Santos, 2003) (Andrienko & Andrienko, 1998).

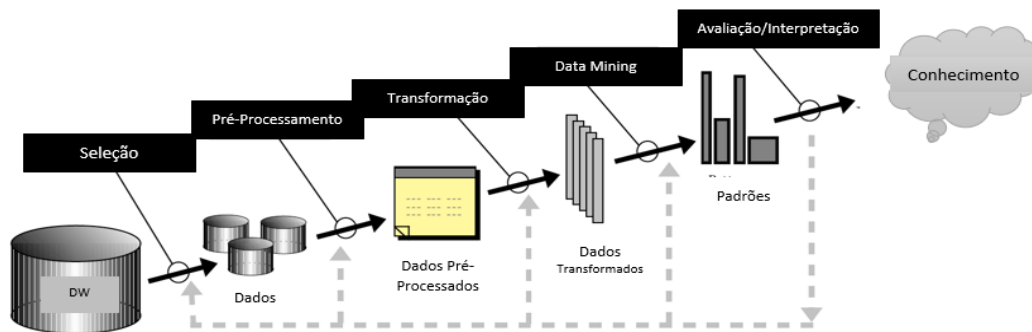


Figura 4 - Processo de DCBD (adaptado (Fayyad, et al., 1996))

### 3.6 Data Mining

Atualmente, o *Data Mining* é uma área de investigação que regista grande crescimento e expansão, causada pelo enorme volume de dados observado hoje em dia, sendo esta cada vez maior duplica a cada 20 meses que passam (Santos & Azevedo, 2005).

A arte de descobrir padrões nos dados de uma base de dados, ao longo da história da computação sempre foi designada por diversas denominações sendo a mais consensual e utilizada de todas, *Data Mining*. Este termo é frequentemente associado às comunidades estatísticas, de analistas de dados e de gestores de sistemas de informação.

A técnica de *DM*, como visto anteriormente integra o processo de DCBD e consiste numa aplicação de técnicas e métodos, sobre grandes coleções de dados, de modo a descobrir tendências ou padrões entre os mesmos (Fayyad, et al., 1996).

O DM é visto como o núcleo de todo o processo de aquisição de conhecimento e envolve a inferência de algoritmos que exploram os dados desenvolvem os modelos e descobrem padrões anteriormente invisíveis e por descobrir.

Estima-se que o conceito tenha surgido nos anos 80 do século XX, mas os primeiros passos nesta área foram dados por John Graunt, em 1662, que desenvolveu um modelo de previsão sobre a praga bubónica

em Londres, cujo objetivo era determinar a próxima praga através da análise do número de mortes em anos anteriores (Chemchem & Drias, 2015).

Esta temática tem-se revelado bastante útil no apoio ao negócio de diversas áreas, mas a sua principal utilização está relacionada com a identificação de oportunidades ou ameaças para a organização ou para tentar solucionar problemas encontrados.

Os modelos de *Data Mining* necessitam de ser cuidadosamente utilizados e construídos pois são suscetíveis de sofrerem de problemas de “subajustamento” ou “sobreajustamento”. No primeiro caso, o modelo gerado é demasiado generalizado, não atentando às particularidades diferenciais do conjunto de dados, por outro lado o segundo caso deve-se à utilização exagerada de demasiadas características específicas dos dados usados (Ramos L. F., 2014).

Tipicamente, as tarefas de *DM* são divididas em duas grandes classes (Tabela 3) relacionadas de acordo com o efeito do estudo a realizar, podendo ser agrupadas em descritivas ou preditivas. A classe descritiva tem o propósito de caracterizar as propriedades gerais dos dados, ou seja, permite estabelecer padrões entre dados para posterior análise humana aumentando o conhecimento e a informação que temos dos mesmos, sendo que as preditivas recorrem a algoritmos específicos para traçarem previsões, no sentido de apoiar o processo de tomada de decisão (Han, et al., 2000).

De acordo com o que é observável no estudo de Santos e Ramos (Santos & Ramos, 2006) podemos deduzir que nas tarefas de previsão devemos escolher o modelo que nos ofereça uma taxa de precisão mais elevada, mesmo que este seja de difícil acesso e interpretação. Já no caso da tarefa de descrição o modelo mais aconselhável é aquele que oferece um conhecimento mais alargado sobre os dados analisados, descartando em parte a precisão do estudo.

Como referido anteriormente e visível na tabela 3, a área de *Data Mining* é composta por diversas técnicas sendo cada uma delas associada a um tipo específico de projeto e de resultados esperados. Tendo em conta o cariz de dados utilizados e o trabalho desenvolvido será interessante destacar e especificar a técnica de classificação.

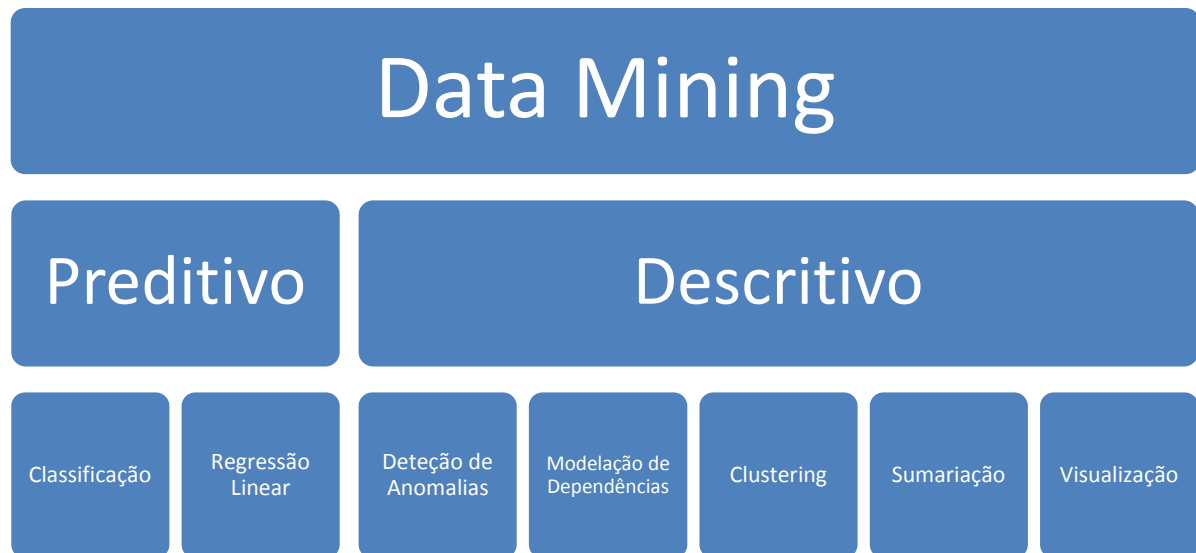
Desta forma poderemos referir que os problemas de classificação, como o próprio nome indica permite categorizar o conjunto de dados, por classes predefinidas. Este tipo de problema tem como objetivo principal o de prever valores discretos ou nominais.

O modelo divide-se em duas etapas, numa primeira fase ocorre uma divisão dos dados por classes predefinidas, sendo esta atribuição realizada a partir da análise de um conjunto de dados treino. A associação dos dados a classes é feita pela identificação de características comuns entre os mesmos.

Na fase posterior, em que já existe um modelo de classificação, este é aplicado ao conjunto de dados teste para verificar a sua fiabilidade de classificação (Santos & Azevedo, 2005).

Esta é a tarefa mais comum do processo de *Data Mining* utilizando técnicas, geralmente aplicada a casos de deteção de fraudes, aplicações de risco e tendências de mercados financeiros (Maimon & Rokach, 2010).

Tabela 3 - Modelos de DM (adaptado de (Santos & Ramos, 2006))



### 3.7 Text Mining

*Text Mining* é visto como uma técnica de descoberta de conhecimento, através de um processo automático de extração de informação a partir de documentos e fontes escritas, com recurso ao poder de processamento de computadores (Hearst, Marti, 2003).

Esta é uma tecnologia criada a partir do *Data Mining*, e é já considerado um processo muito mais maduro, tendo sido já realizados diversos estudos acerca do problema de previsão baseado no reconhecimento de padrões em bases de dados. A principal diferença entre estas duas tecnologias é que enquanto o *DM* trabalha sobre conjuntos de dados estruturados e bem organizados, geralmente do tipo numérico, o *Text Mining* trabalha sobre grandes coleções de texto (Weiss, et al., 2005).

Nomeadamente, na desconstrução do conceito, a palavra *text* (texto, em português) refere-se a conhecimento escrito, ou seja, qualquer publicação que detenha palavras (Weiss, et al. 2005). Devido à interligação entre computadores ter-se tornado na espinha dorsal, tanto de áreas científicas como económicas, enormes coleções de documentos são diariamente criados. É estimado que 85% da

informação de uma organização está estruturada em texto através dos mais diversos documentos, ressaltando a importância do *TM*, que através de métodos de análise de informação, oferece a capacidade de tratamento e estudo de documentos com um grande volume de palavras, permitindo lidar com a resolução de más interpretações, de incerteza e imprecisão na informação (Hotho, et al., 2005). No desempenho do seu papel, o *Text Mining* recolhe ensinamentos de diversas áreas, tais como: a matemática, a estatística, a probabilidade, inteligência artificial e gestão de conhecimento (Weiss, et al.2005).

Hoje em dia, após muitas pesquisas terem sido realizadas na área, torna-se evidente que o *Text Mining* está intimamente ligado ao *Data Mining*, devido à adaptação dos seus algoritmos como forma de evitar problemas e alcançar o objetivo a que destina esta tecnologia de análise textual. Esta afirmação está relacionada com as diversas pesquisas e estudos realizados sobre as áreas que compõe o processo de *TM*, sendo elas o *Information Retrieval*, *Natural Language Processing* e *Information Extraction*, descritas em seguida (Hotho, et al., 2005):

- *Information Retrieval (IR)* – área que trata da procura por documentos que estejam relacionados com a resposta à pergunta colocada pelo utilizador através de uma *query*, não oferecendo a resposta em si.
- *Natural Language Processing (NLP)* – esta área tem como objetivo desenvolver um entendimento adequado da linguagem natural através de computadores, sendo usado também para melhorar a velocidade de processamento de textos.
- *Information Extraction (IE)* – área com o objetivo de extrair informação específica a partir de documentos.

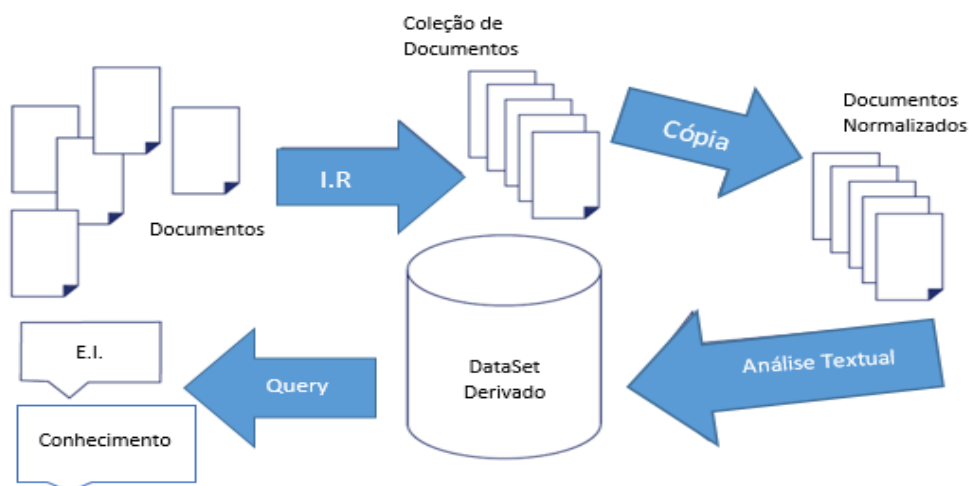


Figura 5 - Processo de TM (adaptado de (McDonald & Kelly, 2012))

A figura 5 mostra graficamente o processo de *TM*. Aqui é possível observar que quando existe uma necessidade de informação o utilizador cria uma *query*, que pode ser entendido como uma pergunta ao sistema. Através do *IR*, o sistema faz uma pesquisa no sistema sobre os documentos que estão relacionados com a *query*. Seguidamente organiza-os e apresenta-os ao utilizador sobre uma ordem de relevância. De seguida essa coleção de documentos é copiada e normalizada de forma a obter uma estrutura interpretável pelo sistema através do *NLP*, sendo simultaneamente realizada uma análise textual. Posteriormente, é formado uma base de dados textual já interpretável pelo sistema, sendo esta submetida ao processo de extração de informação, onde os dados contidos no *dataset* são virtualmente comparados com a *query* introduzida pelo utilizador. Desta forma através de técnicas de descoberta de conhecimento, onde serão analisados e reconhecidos padrões de dados na informação extraída estará a ser criado conhecimento (McDonald & Kelly, 2012).

Atualmente já existe um reconhecimento generalizado do potencial do *TM* e cada vez mais a sua aplicação nas mais diversas áreas tem vindo a ser constatada (Fan, et al., 2006). Ferramentas e aplicações desta tecnologia tem vindo a ser introduzidas na biomedicina, onde o volume de informação é imenso e torna-se complicado para qualquer investigador ou profissional encontrar informação credível e atualizada sobre um determinado tema específico (Spasic, et al., 2005). No Governo o *TM* tem vindo a ser aplicado em agências de inteligência e segurança nacional, como ferramenta de deteção de indícios terroristas. Na educação mostrou ser muito benéfico para estudantes e professores sempre que estes necessitam de se informar sobre algo, afastando-os de técnicas mais rudimentares e tradicionais. No mundo empresarial, as ferramentas de *TM*, têm-se mostrado grandes aliados dos gestores e



administrados, devido a fornecer informações sobre a concorrência, histórico de fornecedores e no desenvolvimento de estratégias de Marketing (Fan, et al., 2006).

Nesta área de *TM* apesar de constantemente estudadas para maior grau de maturação, já existe um conjunto principal de técnicas consolidadas, sendo interessante para o projeto destacar sobretudo a técnica de extração e de análise de distribuição.

### 3.7.1 Extração

Esta é a técnica de extrair informações específicas dos documentos, através da pesquisa de atributos, sendo que nem sempre os atributos estão presentes nos documentos. Esta técnica é utilizada no pré-processamento dos dados, e o resultado da sua aplicação geralmente é transformado num formato estruturado de dados (Leal & Rodrigues, 2003).

### 3.7.2 Análise de Distribuição

O objetivo desta técnica é verificar a existência de determinados termos numa coleção de documentos, analisando também a frequência de aparição desses termos. Os resultados são compactados numa lista, permitindo a análise de quais as palavras mais utilizadas (Leal & Rodrigues, 2003).

## **3.8 Information Retrieval**

Um estudo realizado nos anos 90 (Manning, et al., 2008), mostrou que a grande maioria das pessoas preferiam procurar informações e adquirir conhecimento através de comunicação com outras pessoas do que através das tecnologias, nomeadamente motores de busca. Isto devia-se a uma grande descredibilização da tecnologia existente, cujo seu potencial ainda estava bastante escondido para o público geral. Contudo durante a última década assistimos a grandes inovações e desenvolvimentos tecnológicos, e em simultâneo o tema *Information Retrieval* cresceu drasticamente. A otimização destes sistemas está por de trás da arquitetura dos motores de busca utilizados diariamente por milhões de pessoas, e para surpresa de todos este tema passou de apenas ser falado a nível académico para se tornar na principal base informacional do público geral (Manning, et al., 2008).

Hoje em dia, o volume de informação eletrónico é gigantesco, e as coleções de documentos armazenados, seja em computadores pessoais ou em grandes servidores empresariais, estão a

aumentar significativamente (Rodrigues D. F., 2010). À medida que esta “explosão” de informação se vai dando surge a necessidade de encontrar e extrair informação necessária dessas coleções. Como resposta a esta necessidade, surgiu o processo de *Information Retrieval* (Manning, et al., 2008).

Como Manning referiu “*Information Retrieval* trata de encontrar material (geralmente documentos) de natureza não estruturada (geralmente texto) que satisfaça as necessidades de informação do utilizador, dentro de grandes coleções (geralmente armazenados em computadores)” (Rodrigues D. F., 2010).

Por outras palavras, *Information Retrieval* é a ciência de extrair informação a partir de documentos, de acordo com o pedido que o utilizador efetuou. Esta técnica está diretamente relacionada com os motores de busca utilizados na internet, e com áreas como classificação textual e *clustering* textual. A partir da perspetiva da ciência de computadores esta temática trata de todos os aspetos relacionados com *design* e implementação de sistemas de busca, indexação e recolha de documentos (Manning, et al., 2008).

Com a proliferação da internet, hoje em dia vivemos na era da informação, e não existe tema que não tenha sido escrito, analisado, organizado e armazenado. O objetivo principal dos sistemas de *IR* é responder de forma rápida e útil às necessidades de informação e conhecimento de qualquer utilizador. O principal problema é que esta não é uma tarefa fácil pois os sistemas focam-se em encontrar itens na base de dados em vez de responderem diretamente à pergunta que utilizador quer ver resolvida (Mitra & Chaudhuri, 2000).

O processo de *IR* pode ser descrito em quatro fases (Bimbaum & Collins, 1991):

- **Indexação:** Documentos são convertidos, por métodos de representação textual, num formulário capaz de ser interpretado pelo sistema.
- **Formação da *Query*.** O utilizador terá de expressar a sua necessidade de informação através de uma *query*, cujo sistema seja capaz de entender.
- **Comparação:** O sistema compara a *query* introduzida pelo utilizador com os documentos armazenados, posteriormente deve decidir quais os documentos a retornar e por qual ordem. A *query* fornece a informação necessária para diferenciar a informação relevante da restante.
- **Feedback:** desde que o conteúdo do documento e a necessidade de informação do utilizador são representadas imperfeitamente os resultados iniciais nem sempre são os desejados. Algumas modificações sucessivas as *queries* terão de ser feitas até alcançar resultados aceitáveis.

### 3.9 Trabalhos Relacionados na Área

Hoje em dia a gestão da relação com os clientes por parte das organizações já é um assunto que lhe é dado a devida importância dentro e fora da comunidade científica, e como resultado disso os vários departamentos das organizações que interagem diretamente com os clientes são exaustivamente estudados por forma a descobrir mais e cada vez melhores técnicas de conquistar a satisfação dos clientes e consequente fidelização a uma marca, produto e serviço. Devido a esta ordem de ideias a gestão de reclamações tem vindo a despertar um interesse crescente por parte da comunidade científica e como resultado disso pode-se verificar o desenvolvimento de sistemas orientados a este processo de gestão. Existem alguns sistemas já desenvolvidos e inseridos no mercado comercial desenhados para auxiliarem os utilizadores a realizarem uma melhor e mais eficiente gestão de reclamações. Desta forma poderemos verificar que esta área em concreto, hoje em dia, já desperta a atenção de muitas pessoas tanto da parte de quem desenvolve os sistemas como da parte de quem os utiliza diariamente. Uma gestão eficaz e adequada das reclamações pode trazer benefícios para as empresas ao nível do fortalecimento de relações com os clientes, como na identificação de problemas e pontos fracos das organizações. Em seguida, serão apresentados alguns sistemas já implementados e em distribuição no mercado comercial.

#### 3.9.1 *FreshDesk*

Este *software* é desenvolvido pela empresa com o mesmo nome, este sistema é originário dos Estados Unidos da América tendo a empresa sido criada em 2010. O *software* é dedicado a pequenas e médias empresas que desejem servir-se de um sistema de gestão de reclamações ágil e rápido na sua manipulação sem a necessidade de uma formação extensiva nem grandes custos de manutenção. Este sistema tem um custo de aquisição de cerca de 15 dólares mensais por utilizador, oferecendo inicialmente uma versão experimental completamente grátis. O sistema dedica-se a controlar todo o tipo de reclamações dos clientes, sejam elas partilhadas com a organização através de telefone, correio eletrónico ou redes sócias.

#### 3.9.2 *Zoho Support*

Este sistema de gestão de reclamações é desenvolvido pela empresa *Zoho*, originária dos Estados Unidos da América com ano de fundação de 1996. O *Zoho Support* é um *software* de cariz *web-based*, que se

dedica a todo o tipo de indústrias e a empresas de qualquer dimensão. O *software* dedica-se a oferecer ferramentas de apoio à gestão de vendas e *marketing*, recursos humanos, *help-desk* e finanças. Com um custo fixo de 12 dólares mensais, esta ferramenta centra-se sobretudo no apoio à gestão de clientes e a área de *help desk*.

### 3.9.3 *NABD*

Esta é uma plataforma desenvolvida pela *NABD*, de origem egípcia dedicada ao auxílio das organizações na gestão da área de apoio ao cliente. Este sistema oferece suporte às áreas de gestão de reclamações e gestão de clientes através de diversas ferramentas como bases de conhecimentos e gestão de casos. O *NABD* permite a qualquer organização gerir a relação com os seus clientes em qualquer momento através de diversas plataformas de comunicação, com um custo mensal de 30 dólares.

### 3.9.4 *UsefulFeedback*

O *UsefulFeedback* é uma plataforma desenvolvida pela *iCasework* e distribuída a partir do Reino Unido. Esta é uma aplicação dedicada a médias e grandes empresas, tal como, instituições públicas oferecendo uma série de vantagens para ajuntamento de pontos de controlo e melhoria de serviços. Este *software* oferece serviços e ferramentas para a gestão e elaboração de relatórios sobre as reclamações e opiniões de clientes sobre toda a rede de negócios de uma organização. Através de plataforma *Web-Based* o *software* compromete-se a melhorar os níveis de eficiência de serviços e a melhorar a relação das organizações com os seus clientes.

### 3.9.5 *Complaints Management System*

*Software* desenvolvido pela empresa *QIT Consulting*, com sede nos Estados Unidos da América tendo sido fundada em 2002. Este é um *software* dedicado à gestão de reclamações através de uma plataforma flexível e dinâmica que promete aumentar o nível de satisfação dos clientes com as organizações, através da análise de dados de reclamações e identificando problemas dentro da organização.

## 4. AQUISIÇÃO DE CONHECIMENTO

### 4.1 Recolha e Estudo dos Dados

A matéria-prima utilizada para o desenvolvimento deste projeto foi indubitavelmente as reclamações criadas por utilizadores comuns de unidades de cuidados de saúde que de alguma forma se sentiram insatisfeitos ou sentiram a necessidade de formalmente partilhar a sua opinião (positiva ou negativa) com a entidade que rege e controla as unidades de prestação de serviços de saúde em Portugal. Desta forma, e como referido anteriormente a ERS, Entidade Reguladora da Saúde em Portugal, entidade determinante e parceira durante todo o projeto, concatena informação sobre todas as reclamações formalmente apresentadas em Livro de Reclamações ou Livro de Reclamações Online, em estabelecimentos de cuidados de saúde. É importantíssimo ressaltar que a informação e documentos utilizados para o desenvolvimento deste projeto, ou seja, as reclamações nem sempre obedecem à conotação negativa que lhes está intrinsecamente implícita, podendo ser também apresentado em Livro de Reclamações críticas construtivas aos serviços prestados aos utentes, testemunhos positivos sobre o trabalho desenvolvido em unidades de saúde e declarações de agradecimento a instituições ou profissionais médicos com cariz de louvor.

Sobre esta primeira fase de desenvolvimento de trabalho, estudo conceptual de negócio e recolha de dados, foram realizadas várias reuniões, no centro de operações da instituição reguladora e após ultrapassar alguns entraves a ERS facultou os dados principais para o desenvolvimento do projeto. Durante este processo foi necessário analisar o sistema de reclamações e após dialogar com membros da entidade foram definidos quais os dados mais importantes que deveriam ser disponibilizados. Todos os dados disponibilizados foram sujeitos a processos de anonimização de modo a manter a confidencialidade dos mesmos.

Posteriormente e após a recolha do *dataset* da ERS, este foi armazenado através de uma base de dados *Oracle SQL*.

O conjunto de dados recolhido é assim composto por 6 tabelas relacionais assumindo cada uma delas os nomes dados pela instituição, sendo realizada assim uma simples exportação de dados da instituição reguladora para a universidade por razões de comodidade e facilidade de acesso. Da exportação de

dados realizada facilmente são deduzíveis a existência de relações entre dados, resultando daí o modelo relacional observável na figura 6.

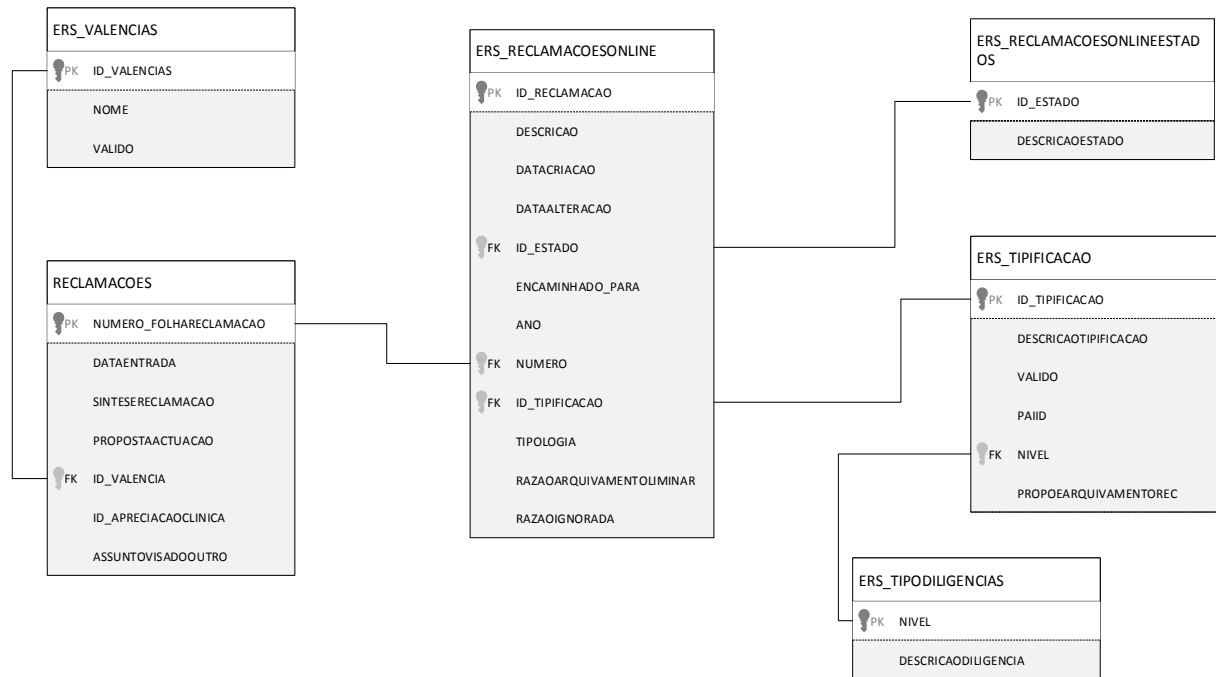


Figura 6 - Modelo Relacional de Base de Dados

A primeira tabela denominada por “ers\_reclamacoesonline”, como o nome indica diz respeito às reclamações apresentadas através de meio eletrônico, um serviço disponibilizado pela ERS no próprio *site* da instituição. Esta tabela é composta por 12 campos informacionais, sendo alguns destes a descrição, ou seja o campo que contém a reclamação transcrita na íntegra, dados relativos a datas tanto de criação como de alteração da reclamação, a tipologia e o id de tipificação da reclamação, entre outros campos informacionais.

De salientar que esta é a tabela principal de todo o conjunto de dados. As restantes tabelas apesar de também serem informacionais apresentam um cariz diferente e têm em comum o objetivo de ajudar a completar as informações existentes na tabela de reclamações *online*.

A seguinte tabela é denominada por “ers\_reclamacoesonlineestados”, sendo composta meramente por 2 campos, o id do estado e a descrição de estado. Com a análise ao conjunto de dados desta tabela é possível verificar a existência de 6 estados. Dentro dos estados que poderão ser atribuídos a uma reclamação *online*, esta poderá ser classificada como “anexada ao processo”, “gerou Rec”, “arquivada liminarmente”, “inserida”, “reencaminhada” ou “ignorada”.

Posteriormente poderá ser observado a existência de uma tabela denominada por “ers\_reclamacoes\_tipos\_diligencias” e a informação contida nesta tabela retrata a atenção que é dada

à reclamação em si. Esta tabela conta com formulação de dois campos, sendo o primeiro o identificador de diligência ou “id” e a segunda a descrição correspondente, podendo esta ser “abertura de processo interno”, “implementação de medidas preventivas”, “implementação de medidas corretivas”, “instauração de processo disciplinar” ou “outras”.

A tabela denominada por “ers\_tipificacao” contém a tipificação, ou seja, a categorização atribuída à reclamação após a sua análise por parte de um funcionário qualificado da ERS. Esta tabela é composta por um identificador de tipificação e a sua respetiva descrição, além disso apresenta ainda um campo denominado por “valido” que retrata o nível de arquivamento onde se situa a reclamação.

O conjunto de dados recolhidos da Entidade Reguladora é composto também por uma tabela que retrata as reclamações produzidas no livro de reclamações, sendo estas produzidas em modo tradicional no livro e posteriormente traduzidas para meio tecnológico por um agente qualificado aquando da análise na Entidade Reguladora da Saúde. Esta tabela denomina-se por “ers\_reclamacoes” e é composta por 7 campos informacionais. O primeiro campo retrata o código identificador da folha de reclamação, contendo também um campo para identificar a data de entrada da reclamação na instituição, sendo que esta concatena dia e hora no mesmo campo. Esta tabela inclui também com uma síntese da reclamação onde é realizado uma apreciação e resumo do conteúdo da reclamação, posteriormente existe um campo para propostas de atuação em que o agente da entidade reguladora faz uma apreciação do problema encontrado na reclamação e traça uma série de perspetivas para resolução do mesmo. Esta tabela contém também um campo denominado por valência que transcreve o id da valência correspondente à reclamação, um código de apreciação clínica e um campo para outros assuntos que possam estar implícitos à reclamação.

Para finalizar, o *Dataset* utilizado conta com uma tabela de valências denominado por “ers\_ac\_valencias” composta por três campos informacionais. Sendo eles o identificador da valência, o nome da valência que poderemos traduzir pela sua descrição onde é apresentado o campo da medicina ou departamento médico onde poderemos inserir a reclamação.

## **4.2 Caracterização dos Intervenientes**

A partir do estudo efetuado aos dados fornecidos pela Entidade Reguladora da Saúde e da análise da interação desta instituição com o meio social e comunitário em que ela se insere é facilmente compreensível a existência de três grupos maioritários de intervenientes em determinados pontos da cadeia de processo da gestão de reclamações em unidades de cuidados de saúde. Consequentemente os intervenientes deste processo de gestão poderão ser caracterizados através de duas dimensões pela sua atividade e pelas funções que desempenham nesta cadeia de processo. Obtemos assim uma dimensão que retrata os intervenientes de cariz ativo no processo e os intervenientes de cariz passivo durante todo o processo.

Os intervenientes de cariz ativo é denominação atribuída às pessoas que desempenham um papel ativo no processo de gestão de reclamação, ou seja, são as pessoas cujas suas atividades podem causar influência no decurso do processo. De acordo com esta caracterização poderemos verificar a existência de três intervenientes, sendo estes o Reclamante, Agente Supervisor das Reclamações e Reclamado.

### **Reclamante:**

- Papel - Produzir uma reclamação sobre um serviço que lhe foi prestado ou sobre a instituição que lhe prestou determinado serviço, sempre que este interveniente sentir a necessidade de o fazer ou se sentir desagradado com algo.
- Expectativa - Obter uma resposta adequada ao problema relatado formalmente em reclamação, de forma a encontrar uma solução para o seu caso concreto ou para que o problema não volte a acontecer a outras pessoas em casos semelhantes.

### **Agente Supervisor das Reclamações:**

- Papel - Analisar e estudar todas as reclamações recebidas pela ERS. Desta forma, caso o problema sugerido pelo reclamante se constitua realmente um defeito na prestação de serviços nas unidades de cuidados de saúde este deverá tecer noções preventivas para atenuar e eliminar o problema encontrado.
- Expectativa - Continuamente melhorar e aumentar a fiabilidade e qualidade da rede nacional de saúde.



**Reclamado:**

- Papel - Prestar serviços de cuidados de saúde com qualidade e excelência. Após a definição de um problema o reclamado deve atenuar ao máximo a sua ocorrência e aplicar as noções de prevenção tecidas pela ERS.
- Expectativa - Prestar serviços de saúde aos utentes com a melhor qualidade e segurança, de forma a garantir a satisfação dos seus utentes e a qualidade nos serviços prestados.

Numa segunda perspetiva, a caracterização dos intervenientes poderá contemplar outra dimensão, sendo esta agrupada pelos intervenientes de cariz passivo no processo de gestão. Sobre esta dimensão poderemos caracterizar as pessoas cuja sua influencia não é tão grande e tão forte no processo de gestão, mas que ainda assim as suas ações asseguram a estabilidade e normal funcionamento do processo.

**Agente do Reclamado:**

- Papel - Fornecer o Livro de Reclamações aos utentes sempre que este for solicitado.
- Expectativa - Executar as suas funções com a máxima qualidade possível e tentar minimizar conflitos de modo a não formularem reclamações que não se encontrem com uma incorreta disponibilização dos serviços.

**Agente da Entidade Reguladora da Saúde:**

- Papel - Inserir todas as reclamações redigidas em Livro de Reclamações no sistema informático da ERS e efetuar uma primeira triagem da informação.
- Expectativa - Executar as funções a que se destina com a máxima qualidade e fiabilidade possível.

Tendo como base a informação fornecida resta salvaguardar que os intervenientes de cariz passivo apenas intervêm no processo de gestão de reclamações quando este é feito de uma forma tradicional, ou seja, quando a reclamação é formulada em Livro de Reclamações. Caso a reclamação seja formulada eletronicamente este dois intervenientes de cariz passivo não têm qualquer relevância ou intervenção em qualquer ponto do processo.

### **4.3 Classificação do Problema**

O problema central de todo este estudo retrata o processo de gestão de reclamações em unidades de cuidados de saúde. Como foi explicado anteriormente a Entidade Reguladora da Saúde, recebe diariamente um grande volume de dados relativos a reclamações produzidas, sendo que grande parte delas indicam e identificam problemas reais na prestação de serviços aos utentes.

Após identificarmos os intervenientes diretos deste processo de gestão poderemos facilmente perceber que este assunto da gestão de reclamações tem um grande impacto na vida de cada um dos intervenientes, sobretudo na vida do reclamante pois este pode reclamar de um serviço médico que lhe foi prestado ou o caso de não ser dada a devida atenção à doença do mesmo.

No caso concreto do reclamado, o estudo do processo de gestão de reclamações poderá oferecer diversas oportunidades de melhoria institucional pois quando uma instituição sofre um grande conjunto de reclamações do mesmo departamento ou área clínica ou do mesmo serviço médico ou prestador do serviço, esse conjunto total de reclamações poderá indicar para além de um problema concreto ou uma ameaça à qualidade da instituição, poderá indicar que um agente está a desempenhar mal o seu papel dentro da organização e que não está a tratar devidamente os utentes a seu encargo. Da mesma forma é necessário lembrar que o Livro de Reclamações não agrupa unicamente reclamações ou críticas negativas, este Livro poderá conter também sugestões de melhoria por parte dos utentes e louvores e declarações de agradecimento, e a análise de toda esta distribuição poderá resultar em mudanças institucionais, pois quando um determinado profissional médico é consecutivamente destacado em declarações de agradecimento isto poderá indicar que ele em primeiro lugar está a realizar um bom trabalho mas também que é uma pessoa que poderá liderar o seu departamento ou unidade e poderá servir de modelo para outros profissionais.

Devido então a estas duas perspetivas e contando também com o facto de a entidade reguladora querer prestar o melhor serviço possível e exercer corretamente a missão e as tarefas a que se propõe poderemos facilmente concluir que este assunto é de altíssima importância e uma correta simplificação do problema poderá ter um impacto enorme na vida das pessoas.

## **5. DESENHO DA SOLUÇÃO**

### **5.1 Orientação de Tarefas**

Esta fase do trabalho é caracterizada pelo Modelo de Tomada de Decisão de Simon como a fase de conceção do projeto, é caracterizada pela delineação ou estruturação de uma solução ou várias para o problema encontrado. Desta forma, os resultados expectáveis neste ponto são análises estruturadas e fundamentadas do panorama geral envolvente do problema e das respetivas soluções delineadas.

Em sequência do que é traçado pela metodologia de Simon, este capítulo da dissertação oferece uma imagem geral do fluxo de informação envolvente à Entidade Reguladora da Saúde, incidindo-se também sobre o processo de gestão de reclamações empregado pela mesma entidade.

De forma a oferecer uma imagem interessante e realista do panorama geral e do meio envolvente ao processo de gestão de reclamações foi desenvolvido uma ontologia, que retrata toda a cadeia informacional deste processo, os agentes envolvidos e as etapas que compõem a gestão de reclamações.

### **5.2 Desenvolvimento da Ontologia**

Como é facilmente dedutível o desenvolvimento desta ontologia teve em conta informações recolhidas na Entidade Reguladora da Saúde, instituição portuguesa dedicada à regularização e controle dos serviços de saúde assumindo inclusivamente a responsabilidade de gerir o processo de análise de reclamações em unidades prestadoras de cuidados de saúde (Relatório sobre "A carta dos Direitos dos Utentes", 2011).

Desta forma, e tendo em atenção as noções e regras assimiladas a partir da análise de diversos estudos realizados no campo do desenvolvimento de ontologias, foi necessário aquando do processo de criação ter consciência de duas perspetivas principais: a primeira está relacionada com o enquadramento deste trabalho no desenvolvimento de um sistema para análise de reclamações, refletindo o ponto de vista de quem fornece dados para a cadeia informacional da ontologia (tanto os reclamantes como os reclamados como a instituição reguladora) e por último a perspetiva de futuro aproveitamento da ontologia em projetos de cariz semelhante e o seu contributo para a comunidade científica (Oliveira, et al., 2015).

Os utentes de uma unidade prestadora de cuidados de saúde têm como direito constitucional apresentar uma reclamação ou crítica ao serviço que lhes foi prestado sempre que sentirem necessidade de o fazerem. Por esta razão todos os estabelecimentos públicos ou privados são obrigados a apresentarem o livro de reclamações e a fornecerem-no aos clientes livremente sempre que este for solicitado (Oliveira, et al., 2015).

Este livro é dedicado unicamente a receber reclamações, tendo desta forma uma estrutura única sendo exatamente igual para todos os estabelecimentos. Desta forma quando algum cliente começa a preencher este livro com a sua reclamação, este pode ser visto como o ponto iniciador de todo o processo de gestão de reclamações. O utente que faz a reclamação, designado nesta fase como reclamante, inicia a realização da sua reclamação através do preenchimento dos campos relativos a informações pessoais, identificando a pessoa que reclama, através do nome, número de identificação civil, morada e número de telefone. Seguidamente e ainda no preenchimento da reclamação, existe um fornecimento de dados relativos à pessoa ou instituição sobre a qual a reclamação se incide, ou seja o reclamado, finalizando a formulação da reclamação com a escrita do assunto tratado, escolha de uma categoria ou tipologia para designar o documento e a formulação do texto onde é explicada toda a razão da insatisfação do reclamante que deu o mote para a escrita da reclamação. Concluindo esta fase, resta dizer que o preenchimento da livro de reclamações não é o único meio para o reclamante partilhar a sua insatisfação, este também o poderá fazer através do portal *online* da entidade que regula as reclamações em Portugal. Posteriormente as reclamações são recolhidas e enviadas para análise na instituição responsável. A análise é feita por um agente, destacado pela instituição, que supervisiona e analisa a reclamação, determina se esta está bem categorizada e processa a mesma no sistema através da indicação de informações do reclamante e do reclamado e através da escrita de uma síntese de informação que a reclamação contém. Após o processamento da reclamação ocorre uma fase de estudo em que se analisa o problema central do documento e define-se uma medida corretiva para solucionar o mesmo, sendo esta medida partilhada com o reclamado (Oliveira, et al., 2015).

Caso a medida traçada pela instituição seja bem aplicada pelo reclamado, o problema é solucionado e a reclamação em questão passa o seu estado para arquivada. Quando a medida traçada não surte o efeito esperado, ocorre uma comunicação entre a instituição reguladora e o reclamado para estudar as razões pelas quais o problema não foi solucionado ou se a medida traçada inicialmente está realmente bem adequada (Oliveira, et al., 2015).

Resta dizer que assim que se inicia a fase de processamento e análise de reclamações, é sempre atribuído um estado de reclamação a cada uma delas baseando-se nas conclusões preliminares da análise dos documentos, sendo este estado atribuído pelo agente supervisor responsável designado pela entidade reguladora.

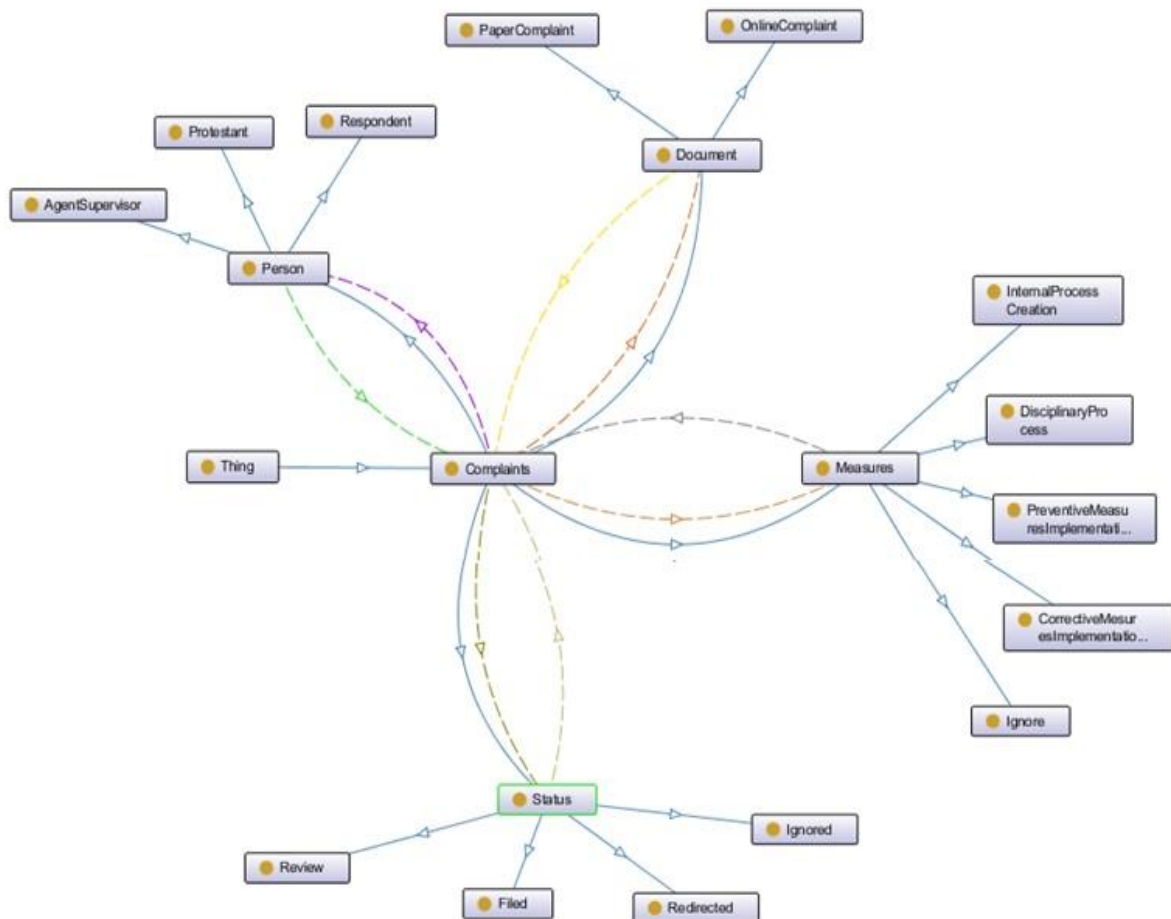


Figura 7 - Diagrama das Classes da Ontologia Reclamações (Oliveira, et al., 2015)

Na figura 7 e 8 é possível observar todas as classes principais que compõem a ontologia, envolvidas no processo de gestão de reclamações em estabelecimentos de cuidados de saúde e simultaneamente as relações existentes entre essas mesmas classes. A enunciação das classes em Inglês reside no fator da universalidade da ontologia e a mesma ter sido já alvo de publicação científica (Oliveira, et al., 2015).

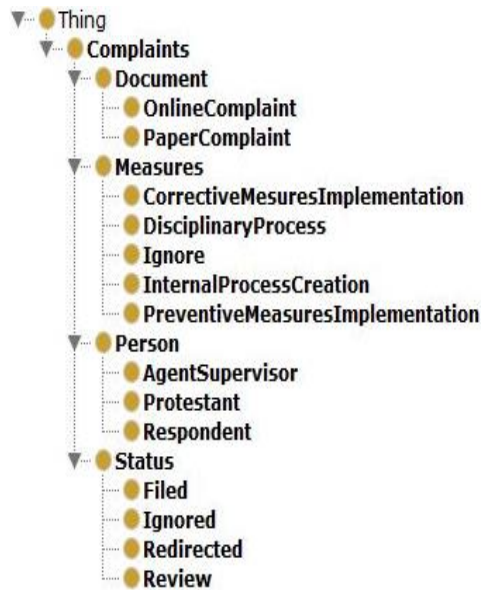


Figura 8 - Taxonomia da Ontologia Reclamações (Oliveira, et al., 2015)

### 5.3 Descrição da Ontologia

Uma ontologia é a transcrição de todo o conhecimento de um determinado domínio, ou seja, é um modelo de dados que agrega os vários conceitos e termos de um tema e especifica os vários relacionamentos existentes entre os mesmos.

Para o desenvolvimento desta ontologia foi utilizado uma abordagem “*Top-down*” em que inicialmente definiu os conceitos mais gerais e posteriormente aumentou o grau de especialização de ideias. Esta ontologia reflete todo o conhecimento adquirido sobre a área da gestão de reclamações de estabelecimentos prestadores de cuidados de saúde, desta forma o conceito principal de todo o estudo são as reclamações. Como é possível observar na figura 9 esta é a classe principal da ontologia, sendo que todas as outras classes criadas são efetivamente subclasses da classe reclamações.

Posteriormente é possível verificar a existência de dois níveis de classes, sendo o primeiro nível formado pela classe documento, medidas, estado e pessoas. No segundo nível de classes poderemos verificar um aumento do grau de especialização da ontologia, relacionada com a apresentação de classes ainda mais restritivas e especializadas quanto ao número de indivíduos envolvidos na questão.

Este diagrama ontológico (figura 9) é composto por classes, propriedades de objetos e propriedades do tipo de dados, de forma a representar exatamente o fluxo de dados e os principais conceitos envolvidos no processo de gestão de reclamações de unidades prestadoras de cuidados de saúde.



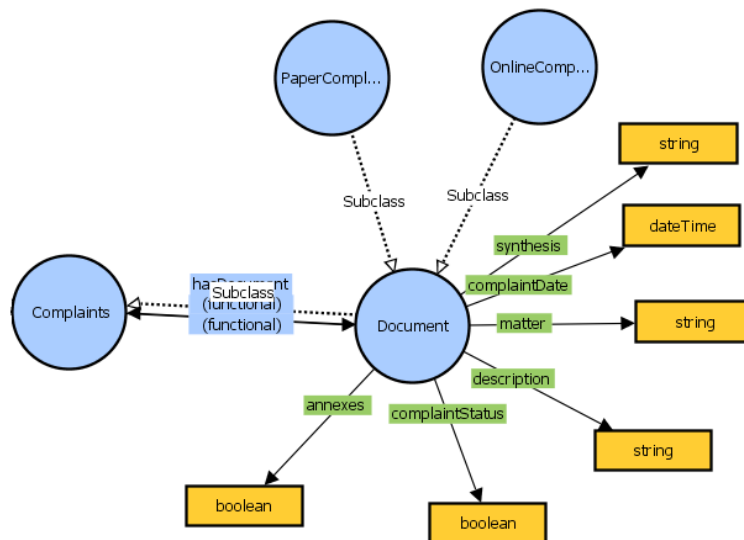


Figura 10 - Classe Documento da Ontologia

A classe Pessoas foi criada pelo facto de nesta ontologia existirem três classes que caracterizam recursos humanos na cadeia de processo e simultaneamente usufruírem de propriedades semelhantes. Desta forma a classe Pessoas apresenta como propriedades informacionais dados relativos ao nome e à morada das pessoas envolvidas no processo de gestão de reclamações. No segundo nível desta classe temos a classe Agente (*Agent Supervisor*), que diz respeito à pessoa que faz a análise das reclamações representando a instituição reguladora, esta classe tem como propriedades a atividade desenvolvida pelo agente e qual o seu cargo dentro da instituição. Outra classe aqui envolvida é o Reclamado (*Respondent*), que diz respeito à pessoa ou entidade sobre a qual a reclamação formulada incide. A última subclasse de Pessoas é a classe Reclamante (*Protestant*), ou seja, a pessoa que se sentiu insatisfeita com o serviço prestado e exerceu o seu direito constitucional de reclamar, esta classe apresenta como propriedades dados informacionais relativos ao número telefónico do reclamante e a sua nacionalidade.

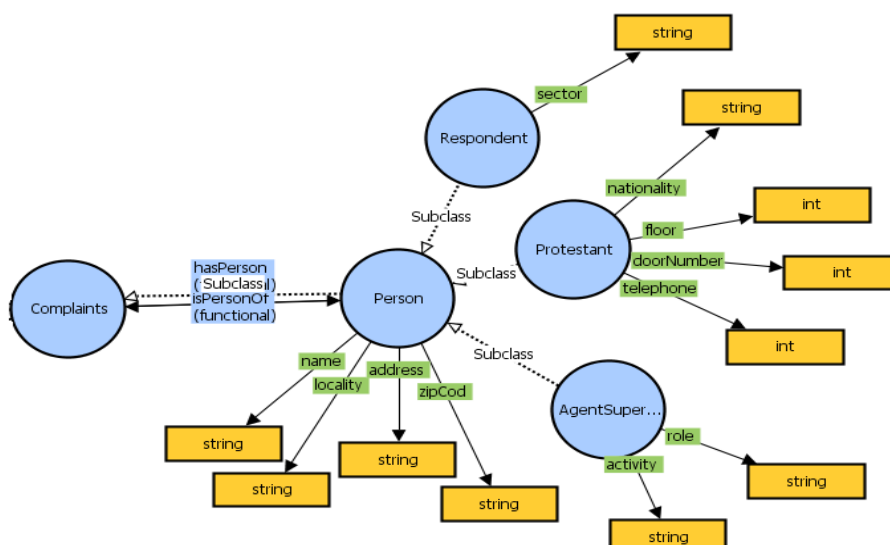


Figura 11 - Classe Pessoas da Ontologia



A classe seguinte é a classe Medidas (*Measures*) que está diretamente ligada à proposta de solução, fornecida pela entidade reguladora, ao problema que justificou a formulação da reclamação. Esta classe tem as suas subclasses diretamente relacionadas e com as medidas mais comuns e principais.

Nesta ordem de ideias a classe *Measures* é composta por Criação de Processo Interno (*Internal Process Creation*), Processo Disciplinar (*Disciplinary Process*), Implementação de Medidas Preventivas (*Preventive Measures Implementation*), Implementação de Medidas Corretivas (*Corrective Measures Implementation*) e Ignorar (*Ignore*).

As propriedades desta classe prendem-se com a adequação das medidas corretivas traçadas pelo agente que supervisionou a reclamação, se a mesma surtiu o efeito esperado ou não e em que estado se encontra a medida. Neste caso específico podemos observar que a classe Medidas tem a propriedade data de análise estando esta estruturada em formato *dateTime* e diz respeito à data em que a medida foi analisada.

Posteriormente temos a propriedade estado da medida, esta propriedade como o nome indica contém o estado em que se encontra a medida, ou seja se a medida corretiva traçada foi aprovada, se foi aplicada ou se foi rejeitada. Podemos observar que a seguinte propriedade Resultado da Medida, diz respeito ao resultado obtido após aplicação da medida corretiva. Temos também como propriedades a data de aplicação, quando foi aplicada a medida corretiva e por último a Descrição da Medida, onde é detalhada a proposta de resolução para o problema identificado na reclamação.

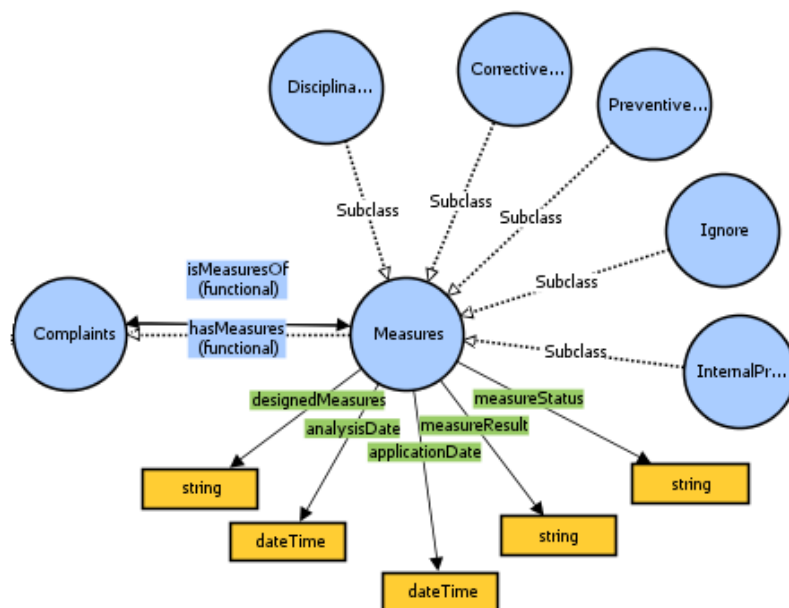


Figura 12 - Classe Medidas da Ontologia

A última classe Estado (*Status*) é uma classe da ontologia unicamente dedicada à análise do estado em que se encontra a reclamação, caracteriza-se através de 4 subclasses sendo que todas representam um estado particular da reclamação sendo elas: Revisto (*Review*), Arquivado (*Filed*), Reencaminhado (*Redirect*) e Ignorado (*Ignored*). A propriedade principal da classe *Status* é designada por *StatusDate* referindo-se à data em que o agente regulador atribui um estado à reclamação sendo do tipo de dados data.

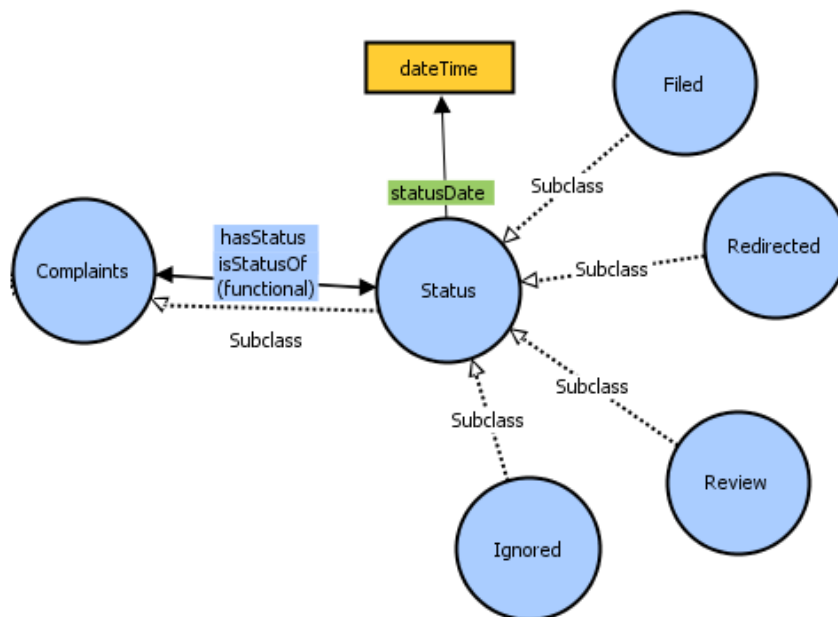


Figura 13 - Classe Estado da Ontologia

## 5.4 Seleção de Dados para Análise

Como referido anteriormente, foram fornecidos pela Entidade Reguladora da Saúde diversos dados relativos às reclamações. Desse grande conjunto de dados recebidos, foi decidido junto da equipa de trabalho que apenas o conjunto de dados relativos às reclamações *online* presentes na tabela "ers\_reclamacoesonline" é que seriam tidos em atenção. Esta decisão é fundamentada pela maior consistência dos dados presentes nesta tabela, pelo facto de as reclamações executadas de forma tradicional e portanto redigidas em Livro de Reclamações terem de ser transcritas ou digitalizadas para o sistema informático da ERS através de um agente designado pela instituição e devido a esta razão a reclamação pode sofrer alterações.

Desta forma foram desenvolvidas algumas *views* sobre os dados, criando assim um *dataset* a partir da base de dados da instituição reguladora constituída unicamente pelos principais e mais interessantes campos que foram definidos para o estudo. Sendo assim a *view* referida (figura 14) é constituída por 5 campos informacionais, contendo dados sobre a descrição da reclamação, ou seja, todo o corpo textual da reclamação criada, o seu ano de criação, o número identificativo da reclamação e dados sobre a sua tipificação e tipologia.

	COLUMN_NAME	DATA_TYPE
1	DESCRICAO	LONG ()
2	ANO	NUMBER (38)
3	NUMERO	NUMBER (38)
4	TIPIFICACAOID	NUMBER (38)
5	TIPOLOGIA	VARCHAR2 (1000)

Figura 14 - Campos do Dataset Desenvolvido

Resta salvaguardar que para facilitar a posterior interpretação dos dados, foi executada uma *query SQL* capaz de transformar os dados contidos no campo “TIPIFICACAOID” nos dados contidos na descrição da tipificação, desta forma os dados contidos neste campo em vez de serem o código identificador de cada tipificação, ou seja, dados do tipo numérico seriam então o nome relativo a cada tipificação atribuída às reclamações.



## 6. DESENVOLVIMENTO DA SOLUÇÃO

### 6.1 Desenvolvimento do Estudo no Knime

Na realização de análises sobre os dados de reclamações *online*, foi definido a aplicação de técnicas de *Text Mining* para avaliação dos textos que compõem as reclamações e para tal foi utilizado o *software open-source Knime*.

Antes de iniciar os trabalhos nesta ferramenta foi decidido que apenas seria interessante trabalhar com três campos informacionais do *dataset* criado a partir da base de dados da ERS, sendo estes a descrição da reclamação, a sua respetiva tipologia e tipificação.

	COLUMN_NAME	DATA_TYPE
1	RECLAMACAO	LONG ()
2	TIPIFICACAO	VARCHAR2 (1000)
3	TIPOLOGIA	VARCHAR2 (1000)

Figura 15 - Dataset Utilizado no Knime

Dentro do *Knime*, a informação contida nestes três campos informacionais iria ser carregada e analisada, através de modelos e técnicas de *text mining*. A análise definida para este programa seria o estudo das palavras mais usadas em cada tipologia identificada e para tal foi utilizado o modelo apresentado em baixo.

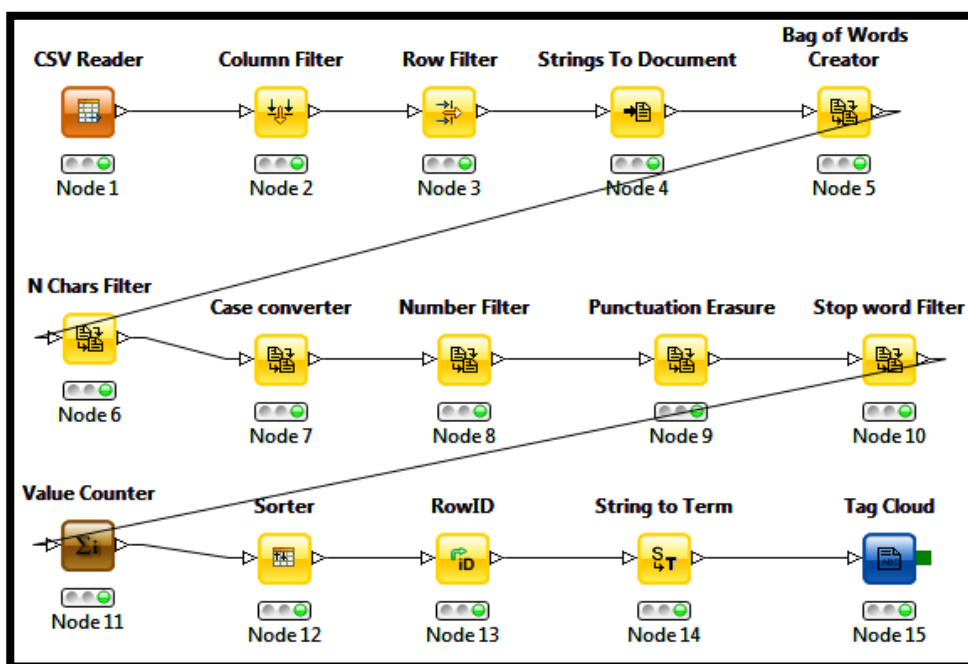


Figura 16 - Modelo Utilizado no Knime

Da análise do modelo de *Knime* poderemos verificar que no *Node 1* temos um leitor de *CSV*, ou seja, temos aqui um carregamento de dados para o *software*. Posteriormente poderemos verificar a existência de dois tipos de filtros, em que um executa um filtro das colunas que vão ser analisadas e outros das linhas que serão trabalhadas. No filtro por colunas serão filtradas a coluna de descrição e tipificação, isto indica que a coluna de tipologia será retirada deste primeiro estudo e no filtro por linhas será selecionada uma tipologia específica das várias tipologias existentes.

Posteriormente poderemos verificar a existência de um nodo designado de “*strings to document*”, cuja sua função é transformar os dados do tipo *string* como os existentes no campo “descrição da reclamação” para o tipo *document*, a necessidade de realizar esta ação prende-se com o facto de os restantes filtros do modelo não se aplicarem sobre tipos de dados *string*, aplicam-se unicamente sobre dados do tipo *document*.

Na continuação do modelo poderemos verificar o uso de uma série de algoritmos relacionadas com o tratamento dos dados para que a contagem de palavras dedica-se unicamente a palavras que possam ser interessantes para o estudo e que tenham peso na análise final, revelando resultados com valor científico, dentro desta ferramentas para tratamento de dados poderemos destacar a “*bag of words*” que individualiza as reclamações por termos, ou seja, ocorre um isolamento das palavras que compõem as reclamações, existe também a ferramenta “*punctuation erasure*” cujo seu objetivo é eliminar todos os símbolos relacionados com pontuação dentro de cada reclamação e a ferramenta “*Stop Word Filter*” que elimina termos especiais de conclusão de frases existentes no respetivo dicionário de “*stop words*” portuguesas oferecido pelo *knime*.

Posteriormente à fase de tratamento dos dados poderemos verificar a utilização da ferramenta “*value counter*” dedicada ao cálculo da ocorrência dos termos identificados nas reclamações.

O modelo encerra com a utilização da ferramenta “*Tag cloud*” que oferece ao utilizador um modo de visualização alternativo e interessante sobre o estudo executado, sendo esta experiência composta por uma nuvem das palavras com maior frequência nas reclamações.

## 6.2 Análise de Palavras com o Knime

Com o modelo demonstrado no ponto anterior foi realizado uma análise das palavras contidas nas reclamações, sendo esta análise realizada, num ponto inicial do estudo, sobre o total de reclamações sem aplicar qualquer restrição sobre a tipificação e posteriormente foram selecionadas as 10 tipificações que agregam mais reclamações, dentro do *dataset* utilizado.

Em todas as análises realizadas foram recolhidas as 20 palavras que apontam maior frequência, e com isto queremos dizer as 20 palavras mais usadas em cada conjunto de reclamações. Posteriormente essa frequência é submetida a dois cálculos de percentagens para obtermos uma ideia mais precisa sobre a sua incidência no conjunto de dados. Os dois cálculos efetuados sobre a frequência de palavras são assim formulados:

$$\% \text{ Frequencia por Reclamações} = \left( \frac{\text{Frequência de Palavras}}{\text{Número de Reclamações}} \right) \times 100$$

$$\% \text{ Frequência por Palavras} = \left( \frac{\text{Frequência de Palavras}}{\text{Número de Palavras}} \right) \times 100$$

### 6.2.1 Aplicação ao Total de Palavras

Nesta primeira aplicação do modelo do *Knime* sobre a totalidade das reclamações *online*, visível na tabela 4, que compõe o *dataset* utilizado é necessário informar que o mesmo é composto por 889 reclamações nas quais verifica-se a utilização de 15201 palavras.

Tabela 4 - Top 20 Palavras mais Frequentes

Palavras	Frequência	% de Frequência por Reclamações	% de Frequência por Total de Palavras
saúde[]	404	45%	3%
hospital[]	398	45%	3%
consulta[]	366	41%	2%
situação[]	350	39%	2%
tinha[]	325	37%	2%
serviço[]	291	33%	2%
médico[]	289	33%	2%
centro[]	273	31%	2%

<b>reclamação</b>	272	31%	2%
<b>resposta</b>	272	31%	2%
<b>mesmo</b>	271	30%	2%
<b>médica</b>	237	27%	2%
<b>qualquer</b>	235	26%	2%
<b>espera</b>	194	22%	1%
<b>tempo</b>	187	21%	1%
<b>utente</b>	182	20%	1%
<b>urgência</b>	181	20%	1%
<b>mesma</b>	180	20%	1%
<b>venho</b>	174	20%	1%
<b>saber</b>	159	18%	1%

Nesta primeira análise realizada, poderemos verificar que a palavra mais usada em reclamações de unidades prestadoras de cuidados de saúde é efetivamente a palavra saúde, sendo que esta surge 404 vezes e integra 45 % das reclamações criada, esta palavra contém também uma margem de 3% na totalidade de palavras, o que à primeira vista pode dar a sensação que é um número pequeno mas tendo em atenção de que o conjunto de palavras é de cerca de 15000 palavras esta percentagem torna-se bastante razoável. No seguimento deste estudo poderemos referir que as palavras hospital e consulta compõem o pódio das mais utilizadas apresentando ambas uma percentagem de frequência acima dos 40 %, em que este resultado não é tão surpreendente pois ambas estão relacionadas com saúde e intimamente ligadas ao meio em que foram obtidos os dados para a realização do estudo. Perto da margem dos 40% estão também as palavras situação e tinha, sendo que a palavra “tinha” é um verbo auxiliar e por esta razão é uma palavra com importância mais reduzida do que as restantes. Em seguida poderemos visualizar a nuvem de palavras oferecida pelo *software*.

Por forma a concluir a análise resta dizer que o conjunto de palavras figurado na tabela número 4, cuja soma de frequência de palavras total é de 5240, representa 34% das palavras normalmente utilizadas na formulação de reclamações.







### 6.2.3 Aplicação à Tipificação “Qualidade nos Cuidados de Saúde”

A tipificação “qualidade nos cuidados de saúde” conta com 124 reclamações, sendo estas compostas por um total de 4668 palavras (tabela 6).

*Tabela 6 - Top 20 Palavras da Tipificação "Qualidade nos Cuidados de Saúde"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Hospital</b>	75	60%	2%
<b>situação</b>	62	50%	1%
<b>tinha</b>	61	49%	1%
<b>saúde</b>	54	44%	1%
<b>serviço</b>	54	44%	1%
<b>mesmo</b>	50	40%	1%
<b>urgência</b>	48	39%	1%
<b>médico</b>	47	38%	1%
<b>reclamação</b>	46	37%	1%
<b>resposta</b>	46	37%	1%
<b>médica</b>	45	36%	1%
<b>consulta</b>	40	32%	1%
<b>dores</b>	37	30%	1%
<b>pessoas</b>	31	25%	1%
<b>deveria</b>	28	23%	1%
<b>sendo</b>	27	22%	1%
<b>tempo</b>	27	22%	1%
<b>venho</b>	27	22%	1%
<b>estado</b>	26	21%	1%
<b>utente</b>	26	21%	1%



#### 6.2.4 Aplicação à Tipificação “Direitos dos Utentes”

A próxima análise incide sobre a tipificação “direitos dos utentes”, sendo que esta concatena um grupo de 103 reclamações (tabela 7).

*Tabela 7 - Top 20 Palavras da Tipificação "Direitos dos Utentes"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Saúde</b>	53	51%	1%
<b>situação</b>	49	48%	1%
<b>hospital</b>	46	45%	1%
<b>consulta</b>	42	41%	1%
<b>serviço</b>	40	39%	1%
<b>centro</b>	37	36%	1%
<b>médico</b>	37	36%	1%
<b>direito</b>	34	33%	1%
<b>mesmo</b>	33	32%	1%
<b>reclamação</b>	33	32%	1%
<b>resposta</b>	33	32%	1%
<b>qualquer</b>	31	30%	1%
<b>médica</b>	30	29%	1%
<b>mesma</b>	29	28%	1%
<b>tinha</b>	27	26%	1%
<b>família</b>	25	24%	1%
<b>tempo</b>	24	23%	1%
<b>espera</b>	24	23%	1%
<b>utente</b>	24	23%	1%
<b>utentes</b>	23	22%	1%

Através do estudo da tabela de frequências poderemos verificar uma pequena alteração no pequeno grupo das 5 primeiras palavras relativamente às análises realizadas nas tipificações anteriores e na geral. Esta alteração é relativa à exclusão da palavra “tinha” deste pequeno grupo de palavras, no entanto ela



### 6.2.5 Aplicação à Tipificação “Tempos de Espera”

Esta tipificação é composta por 102 reclamações e um total de 2958 palavras. Como poderemos observar na seguinte tabela 8 as palavras mais usadas são consulta e espera, palavras estas que compõem mais de metade das reclamações desta tipificação. Este conjunto de palavras compõe cerca de 21% dos documentos tipificados com “Tempos de Espera”.

*Tabela 8 - Top 20 de Palavras da Tipificação "Tempos de Espera"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Consulta</b>	61	60%	2%
<b>espera</b>	54	53%	2%
<b>hospital</b>	50	49%	2%
<b>tempo</b>	41	40%	1%
<b>médico</b>	39	38%	1%
<b>saúde</b>	37	36%	1%
<b>serviço</b>	37	36%	1%
<b>situação</b>	34	33%	1%
<b>qualquer</b>	30	29%	1%
<b>resposta</b>	26	25%	1%
<b>tendo</b>	26	25%	1%
<b>tinha</b>	24	24%	1%
<b>aguardar</b>	23	23%	1%
<b>centro</b>	22	22%	1%
<b>mesmo</b>	22	22%	1%
<b>reclamação</b>	22	22%	1%
<b>urgência</b>	22	22%	1%
<b>marcada</b>	21	21%	1%
<b>marcação</b>	20	20%	1%
<b>venho</b>	20	20%	1%

### 6.2.6 Aplicação à Tipificação “Livro de Reclamações”

A tipificação “Livro de Reclamações” é composta por 60 reclamações e 2720 palavras, e o grupo de palavras representado na tabela agrega cerca de 14% destas reclamações. Nesta tipificação as palavras que mais se destacam são a palavra “reclamação” e “consulta” (tabela 9), sendo que a palavra “tinha” apesar de apresentar grande frequência não deverá ser tida em conta na análise por ser um verbo auxiliar.

*Tabela 9 - Top 20 de Palavras da Tipificação “Livro de Reclamações”*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Tinha</b> □	29	48%	1%
<b>reclamação</b> □	28	47%	1%
<b>consulta</b> □	25	42%	1%
<b>médico</b> □	22	37%	1%
<b>resposta</b> □	22	37%	1%
<b>hospital</b> □	21	35%	1%
<b>livro</b> □	20	33%	1%
<b>médica</b> □	19	32%	1%
<b>saúde</b> □	19	32%	1%
<b>mesmo</b> □	18	30%	1%
<b>médicos</b> □	18	30%	1%
<b>qualquer</b> □	18	30%	1%
<b>serviço</b> □	18	30%	1%
<b>centro</b> □	17	28%	1%
<b>reclamações</b> □	17	28%	1%
<b>mesma</b> □	16	27%	1%
<b>situação</b> □	16	27%	1%
<b>entidade</b> □	14	23%	1%
<b>ficar</b> □	14	23%	1%
<b>utente</b> □	14	23%	1%



### 6.2.7 Aplicação à Tipificação “Taxas Moderadoras”

Do estudo da tipificação “Taxas Moderadoras” poderemos verificar que a mesma é composta por 43 reclamações que agregam cerca de 1312 palavras. Como poderemos verificar na tabela 10 as duas palavras com mais frequência são as mesmas que intitulam esta tipificação surgindo também a palavra saúde em quase 50 % das reclamações. Este grupo de palavras constitui cerca de 20 % das reclamações assim tipificadas.

*Tabela 10 - Top 20 de Palavras da Tipificação "Taxas Moderadoras"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Taxas</b>	24	56%	2%
<b>moderadoras</b>	21	49%	2%
<b>saúde</b>	21	49%	2%
<b>isenção</b>	18	42%	1%
<b>centro</b>	15	35%	1%
<b>resposta</b>	15	35%	1%
<b>hospital</b>	14	33%	1%
<b>reclamação</b>	13	30%	1%
<b>moderadora</b>	12	28%	1%
<b>situação</b>	12	28%	1%
<b>valor</b>	12	28%	1%
<b>mesmo</b>	11	26%	1%
<b>consultas</b>	10	23%	1%
<b>pagamento</b>	10	23%	1%
<b>consulta</b>	9	21%	1%
<b>cumprimentos</b>	9	21%	1%
<b>filha</b>	9	21%	1%
<b>declaração</b>	8	19%	1%
<b>Direito</b>	8	19%	1%
<b>esclarecimento</b>	8	19%	1%

### 6.2.8 Aplicação à Tipificação “Acesso”

A aplicação do modelo sobre a tipificação “acesso” mostra que a mesma é composta por 36 reclamações sendo estas constituídas por 1618 palavras das quais 15% são formadas pelas 20 palavras que integram a tabela de frequências número 11.

Desta tabela podemos destacar que as palavras “saúde” e “consulta” apresentam uma frequência maior relativamente às restantes palavras surgindo em mais de metade das reclamações deste grupo.

*Tabela 11 - Top 20 de Palavras da Tipificação "Acesso"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Saúde</b>	21	58%	1%
<b>consulta</b>	20	56%	1%
<b>hospital</b>	17	47%	1%
<b>centro</b>	16	44%	1%
<b>situação</b>	16	44%	1%
<b>espera</b>	12	33%	1%
<b>peessoas</b>	12	33%	1%
<b>serviço</b>	12	33%	1%
<b>acesso</b>	11	31%	1%
<b>familia</b>	11	31%	1%
<b>médica</b>	11	31%	1%
<b>médico</b>	11	31%	1%
<b>tempo</b>	11	31%	1%
<b>doente</b>	10	28%	1%
<b>reclamação</b>	10	28%	1%
<b>resposta</b>	10	28%	1%
<b>marcação</b>	9	25%	1%
<b>podéria</b>	9	25%	1%
<b>utente</b>	9	25%	1%
<b>consultas</b>	9	25%	1%

### 6.2.9 Aplicação à Tipificação “Rejeição Infundada”

Na análise à tipificação “rejeição infundada” poderemos verificar que a mesma é composta por 35 reclamações com 1450 palavras. É possível verificar na tabela 12, a palavra “saúde” apresenta uma elevada percentagem de frequência nas reclamações com cerca de 86%, o que teoricamente sugere a frequência desta palavra na maioria das reclamações que são tipificadas com “Rejeição Infundada”. Poderemos visualizar também que as palavras “centro” e “consulta” acompanham de uma forma mais atenuada esta tendência surgindo em mais de metade das reclamações criadas.

*Tabela 12 - Top 20 de Palavras da Tipificação "Rejeição Infundada"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Saúde</b>	30	86%	2%
<b>centro</b>	23	66%	2%
<b>consulta</b>	20	57%	1%
<b>médica</b>	17	49%	1%
<b>médico</b>	16	46%	1%
<b>situação</b>	16	46%	1%
<b>familia</b>	15	43%	1%
<b>hospital</b>	15	43%	1%
<b>tinha</b>	15	43%	1%
<b>consultas</b>	14	40%	1%
<b>mesmo</b>	14	40%	1%
<b>atender</b>	12	34%	1%
<b>dirigi-me</b>	12	34%	1%
<b>funcionária</b>	12	34%	1%
<b>resposta</b>	12	34%	1%
<b>médicos</b>	11	31%	1%
<b>reclamação</b>	11	31%	1%
<b>facto</b>	10	29%	1%
<b>foi-me</b>	10	29%	1%
<b>motivo</b>	10	29%	1%

### 6.2.10 Aplicação à Tipificação “Qualidade”

Da análise à Tipificação “qualidade” poderemos verificar que a mesma é composta por 34 reclamações e 1556 palavras das quais o grupo das 20 palavras mais frequente constitui 14 % das reclamações. Decorrendo a análise da tabela 13 de frequências poderemos destacar que a única palavra que compõe mais de 50% das reclamações é a palavra “saúde”.

*Tabela 13 - Top 20 de Palavras da Tipificação "Qualidade"*

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>Saúde</b>	20	59%	1%
<b>hospital</b>	15	44%	1%
<b>serviço</b>	15	44%	1%
<b>qualquer</b>	13	38%	1%
<b>serviços</b>	13	38%	1%
<b>fiquei</b>	12	35%	1%
<b>reclamação</b>	12	35%	1%
<b>médica</b>	11	32%	1%
<b>doentes</b>	10	29%	1%
<b>médico</b>	10	29%	1%
<b>peessoas</b>	10	29%	1%
<b>senhores</b>	10	29%	1%
<b>situação</b>	10	29%	1%
<b>tinha</b>	10	29%	1%
<b>outubro</b>	9	26%	1%
<b>problemas</b>	9	26%	1%
<b>quase</b>	9	26%	1%
<b>urgências</b>	9	26%	1%
<b>atendimento</b>	8	24%	1%
<b>consulta</b>	8	24%	1%

### 6.2.11 Aplicação à Tipificação “Faturação”

A última submissão ao modelo do *software knime* foi à tipificação “faturação”, contando com uma coleção de 34 reclamações e 1522 palavras. A tabela de frequências de palavras (tabela 14) indica-nos que a palavra mais usada nas reclamações desta forma tipificadas é a palavra “situação” com valores de cerca de 50%, visto que a palavra “tinha” não será tida em conta por ser um verbo auxiliar. Destaca-se também palavras como “valor” ou “pagamento” que estão intimamente ligadas ao tema qualificador destas reclamações, ou seja, são termos económicos (tabela 14).

Tabela 14 - Top 20 de Palavras da Tipificação “Faturação”

<b>Palavras</b>	<b>Frequência</b>	<b>% de Frequência por Reclamações</b>	<b>% de Frequência por Total de Palavras</b>
<b>tinha</b> □	18	53%	1%
<b>situação</b> □	17	50%	1%
<b>valor</b> □	16	47%	1%
<b>pagamento</b> □	15	44%	1%
<b>anexo</b> □	14	41%	1%
<b>mesmo</b> □	13	38%	1%
<b>qualquer</b> □	13	38%	1%
<b>consulta</b> □	12	35%	1%
<b>hospital</b> □	12	35%	1%
<b>mesma</b> □	12	35%	1%
<b>reclamação</b> □	12	35%	1%
<b>contacto</b> □	11	32%	1%
<b>resposta</b> □	11	32%	1%
<b>recebi</b> □	10	29%	1%
<b>tendo</b> □	10	29%	1%
<b>exames</b> □	9	26%	1%
<b>informação</b> □	9	26%	1%
<b>saúde</b> □	9	26%	1%
<b>euros</b> □	8	24%	1%
<b>junto</b> □	8	24%	1%

Resta salientar que todas as análises requerem um estudo por parte de especialistas no negócio a fim de retirar conclusões úteis e com influência no processo de decisão. Neste sentido todos os estudos e informações contidos nesta dissertação serão entregues na Entidade Reguladora da Saúde para futura análise por especialistas.

### **6.3 Processo de Criação de *Dashboards***

Atualmente, o desenvolvimento de *dashboards* sobre dados e indicadores de negócio tem vindo a oferecer aos utilizadores uma nova experiência de visualização da informação muito atrativo melhorando a análise e estudo dos dados e auxiliando os processos de tomada de decisão, definição de metas a atingir ou identificação de problemas dentro das organizações. A representação de informação através de *dashboards* pode ser efetuada de diversas formas seja através de gráficos, semáforos, mapas entre outras representações de forma a agilizar o processo de decisão dentro de uma organização.

A criação de *dashboards* neste trabalho contribui para um maior entendimento dos dados trabalhados. Como pontos de análise interessantes poderemos recorrer aos dados envolvidos em questões temporais das reclamações, mas também poderemos analisar as distribuições de reclamações por tipificação ou tipologia.

Resta dizer que os *dashboards* criados pertencem sempre a uma tabela dinâmica e o conjunto de dados trabalhados é sempre o mesmo para todas as análises podendo divergir nos campos utilizados para cada unidade gráfica. Devido à condição da construção sobre uma tabela dinâmica sempre que pretendemos selecionar um grupo de dados num dos gráficos a página assume essa seleção como um filtro de dados, podendo a informação contida noutros gráficos pertencentes à mesma página alterarem-se de forma a melhorar a experiência de visualização do utilizador.

### **6.4 Análise da Informação Contida nas *Dashboards***

Para a realização de análises aos dados através de *dashboards* foi submetido no *software Qlikview* uma base de dados composta pelo campo Reclamação contendo as reclamações submetidas através da plataforma da Entidade Reguladora da Saúde, pelos campos Tipificação e Tipologia das reclamações categorizando-as através destas duas dimensões, e pelos campos relacionados com o espaço temporal

das mesmas sendo estes, o Ano, o Mês e a Hora de Criação. Na figura 21 é possível visualizar o quadro geral de estudos realizados através de *dashboards*.

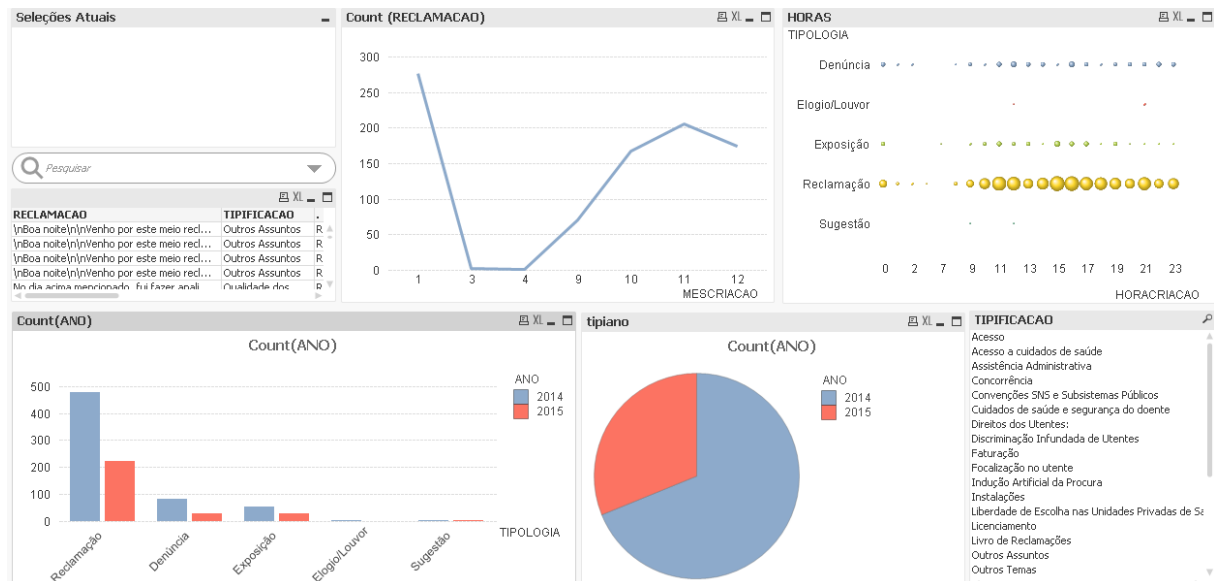


Figura 21 - Quadro Geral de Dashboards

#### 6.4.1 Análise do Número de Reclamações por Ano de Criação

A primeira análise realizada teve em conta sobretudo dois campos da base de dados, sendo estes o número de reclamações e o ano das mesmas, como pode ser visualizado na figura 22.

Desta forma obtemos um gráfico dinâmico do tipo *pie chart* que nos mostra a distribuição de reclamações por anos de criação. Da análise do gráfico podemos verificar que a quantidade de reclamações, no conjunto de dados em estudo, datada de 2014 é bastante superior do que a coleção de 2015, em números concretos o ano de 2014 totaliza 619 reclamações enquanto que o ano 2015 (até junho) totaliza 280 documentos. Como foi referido anteriormente, a parte interessante destes gráficos é que se

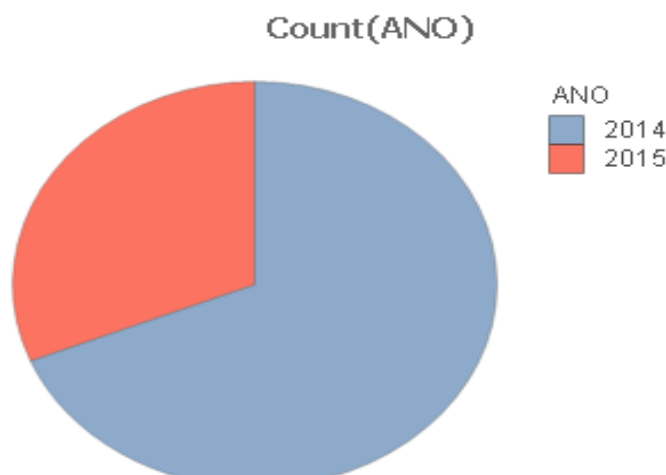


Figura 22 - Dashboards Distribuição de Reclamações por Ano

escolhermos um ano específico todos os gráficos seguintes irão apresentar as suas informações respetivamente para o ano escolhido.

#### 6.4.2 Análise de Numero de Reclamações por Ano e Tipologia

A próxima análise (figura 23) relaciona três dimensões da base de dados através de um gráfico de barras, sendo estas o número de reclamações quantificado pelo eixo vertical, a tipologia que a reclamação pode assumir representado pelo eixo horizontal e ano diferenciando os gráficos por cores.

Da análise do gráfico podemos facilmente observar que a tipologia que agrega mais reclamações é a denominada por “reclamação” obtendo assim um total de 699 reclamações sendo que a maioria delas foi redigida no ano de 2014. Podemos facilmente observar que a tipologia “Reclamação” domina a distribuição de documentos neste estudo, ultrapassando a quantidade de documentos desenvolvidos pelas outras tipologias (a soma de todos os anos). A tipologia classificada como “Denúncia” é a segunda a conter mais documentos sendo esta composta por cerca de 113 reclamações. No caso da tipologia “exposição” verificamos uma coleção de 81 documentos, na tipologia “elogio/louvor” temos um conjunto de 4 reclamações e no caso da “sugestão” apenas 2 reclamações. Simultaneamente poderemos acrescentar que no ano de 2015 a distribuição de reclamações entre “Denúncia” e “Exposição” é bastante semelhante, diferença que tinha sido maior no ano anterior. Para finalizar resta destacar a inexistência até ao momento de documentos sobre a tipologia “Elogio/Louvor” em 2015.

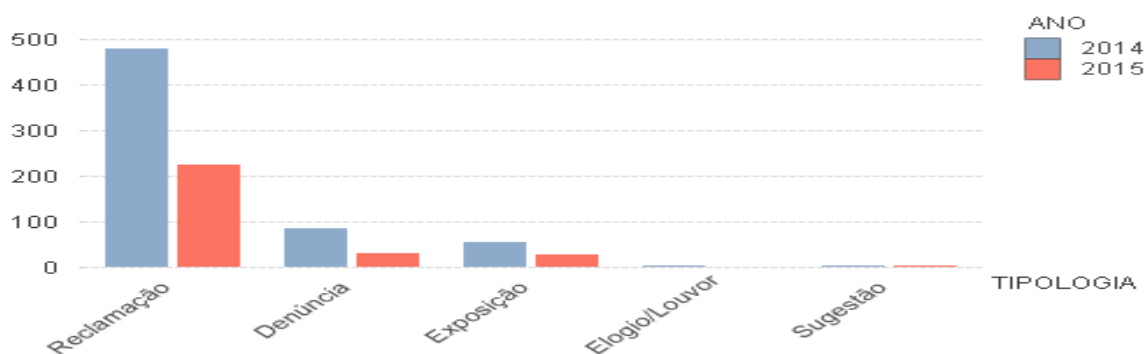


Figura 23 - Dashboard Distribuição de Reclamações por Ano e Tipologia



### 6.4.3 Análise da Criação de Reclamações por Meses

A análise seguinte procura mostrar o nível de desenvolvimento de reclamações por meses do ano. Desta forma, para oferecer uma experiência mais correta foi necessário dividir o gráfico por anos, obtendo assim um gráfico para 2014 (figura 24) e um para 2015 (figura 25).

Da análise dos gráficos, é possível facilmente deduzir que o período de criação de reclamações no qual o conjunto de dados se concentra é de Setembro de 2014 a Abril de 2015. Simultaneamente é possível verificar que o mês de 2014 com mais reclamações é Novembro e no ano de 2015 é indubitavelmente Janeiro. No mês de Fevereiro regista-se uma inexistência da criação de reclamações sendo que os dois meses seguintes apenas concentram juntos um total de 4 reclamações. Resta dizer que no ano de 2014 a distribuição de reclamações procede a um aumento exponencial até ao pico máximo de Novembro decrescendo de forma pouco acentuada até ao mês de Dezembro. Este estudo tem como base um conjunto de dados a nível temporal pouco expansivo, sendo que este facto não afeta de qualquer maneira o desenvolvimento dos gráficos apresentados, estando estes preparados para realizarem análises semelhantes para um conjunto de dados maior e com maior distribuição temporal.

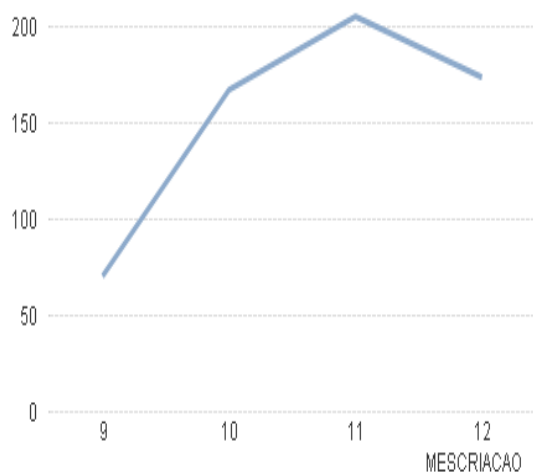


Figura 25 - Distribuição de Reclamações por Meses em 2014

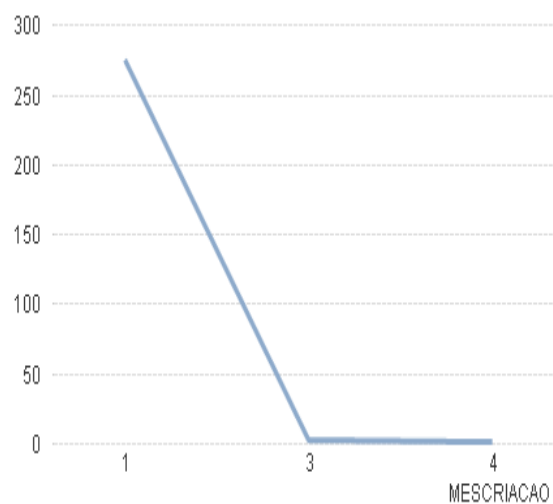


Figura 24 - Distribuição de Reclamações por Meses em 2015

#### 6.4.4 Análise do Número de Reclamações por Tipologia e Horas

A análise seguinte retrata a distribuição de reclamações, desenvolvendo uma relação de envolvimento entre os campos tipologia e horas de criação. Do estudo do gráfico, representado na figura 26, é facilmente perceptível que as *plots* de maior dimensão são as que agregam mais reclamações.

Através da visualização do gráfico facilmente perceptível que durante as horas noturnas, um intervalo entre a meia-noite e as 7 horas da manhã existe uma fraca produção de reclamações sendo este quase inexistente ou mesmo inexistente em algumas tipologias. Outro facto que poderemos constatar analisando o gráfico é que existe dois períodos do dia em que se constata uma extraordinária produção de reclamações, sendo estas na parte da manhã o intervalo das 10h00 até ao 12h00 e da parte da tarde das 15h00 até às 17h00. Poderemos também destacar as 21h00 como a hora da parte noturna que coleciona mais reclamações. Para concluir a análise resta destacar o facto de que os dados do gráfico apenas contemplam a hora de criação da reclamação dentro da Entidade Regulador, sendo que este dado pode não corresponder à hora exata do acontecimento ou situação exposta na reclamação.

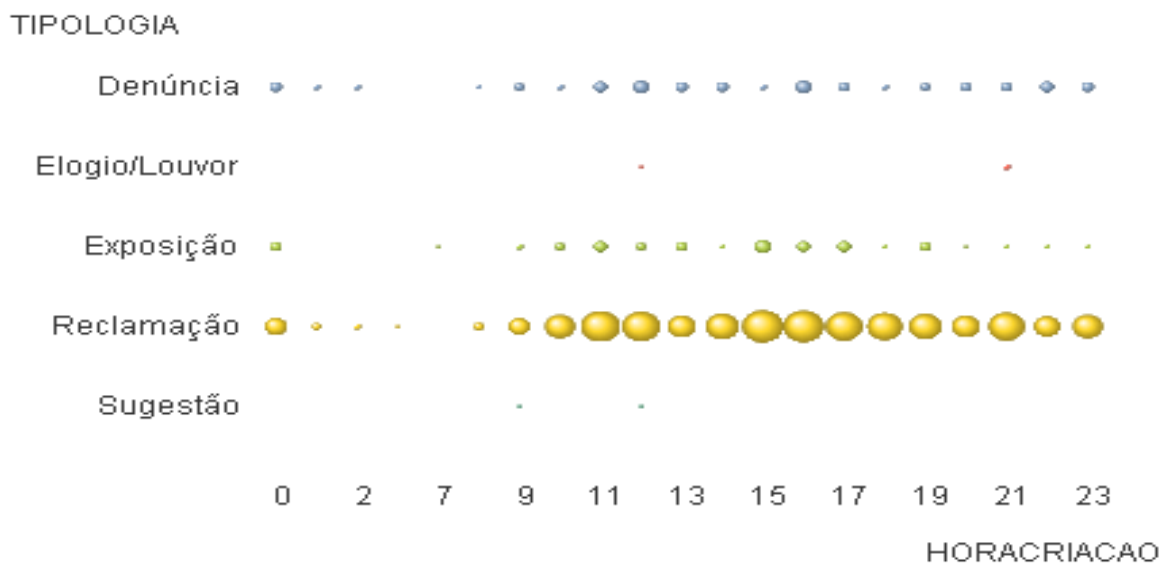


Figura 26 - Distribuição de Reclamações por Horas e Tipologia

## 6.5 Estudo das palavras mais usadas nas reclamações

De forma a obter uma visualização mais atrativa sobre a distribuição de palavras por tipologia foram criadas *dashboards* para o estudo realizado anteriormente. Para a correta realização desta análise foi criado um novo conjunto de dados tendo em conta três dimensões sendo assim foi desenvolvido um campo agregando as palavras mais frequentes recolhidas na análise de cada tipologia, um campo com a frequência de cada palavra e um campo com a respetiva tipologia associada à palavra.

Tendo em atenção ao conjunto de informações contidas no *dataset* utilizado foram desenvolvidos dois gráficos para análise. O conjunto de gráficos desenvolvidos para esta temática poderá ser visualizado na figura 27.

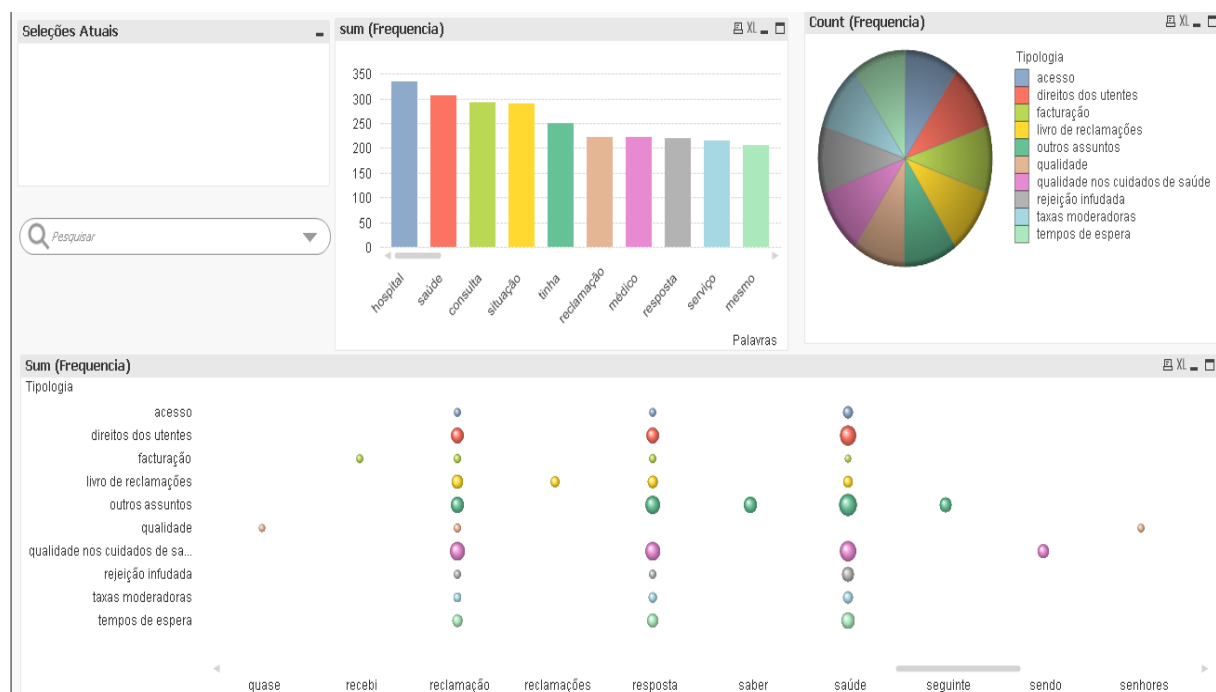


Figura 27 - Quadro Geral de Dashboards para Estudo de Palavras

### 6.5.1 Estudo de Palavras por Frequência

No primeiro estudo desenvolvido foram trabalhados os campos palavras e frequência de palavras. Neste estudo verificasse uma soma das frequências das palavras por cada palavra. O motivo que levou à tomada desta decisão foi o facto de que a mesma palavra poder apresentar frequências diferentes em várias tipologias, como podemos observar nas tabelas de frequências apresentadas anteriormente.

No gráfico de barras apresentado poderemos observar as palavras com maior frequência, mas dentro do *software QlikView* poderemos visualizar todas as palavras existentes na coleção através do deslizamento da barra horizontal visualizada na figura.

Da análise do gráfico presente na figura 28, podemos verificar que a palavra mais utilizada no desenvolvimento de reclamações de unidades prestadoras de cuidados de saúde é indubitavelmente a palavra “Hospital”, sendo que esta conta com uma frequência de 333 aparições. Logo depois podemos visualizar que o grupo formado pelas palavras “Saúde”, “Consulta” e “Situação” registam as maiores frequências sendo desta forma bastante utilizadas nas reclamações sendo a diferença de frequência entre as mesmas pouco acentuada. Por outro lado poderemos verificar que as palavras que registam menor frequência nesta análise são as palavras “esclarecimento”, “euros”, “junto”, “declaração” e “atendimento”, todas com uma frequência de valor 8.

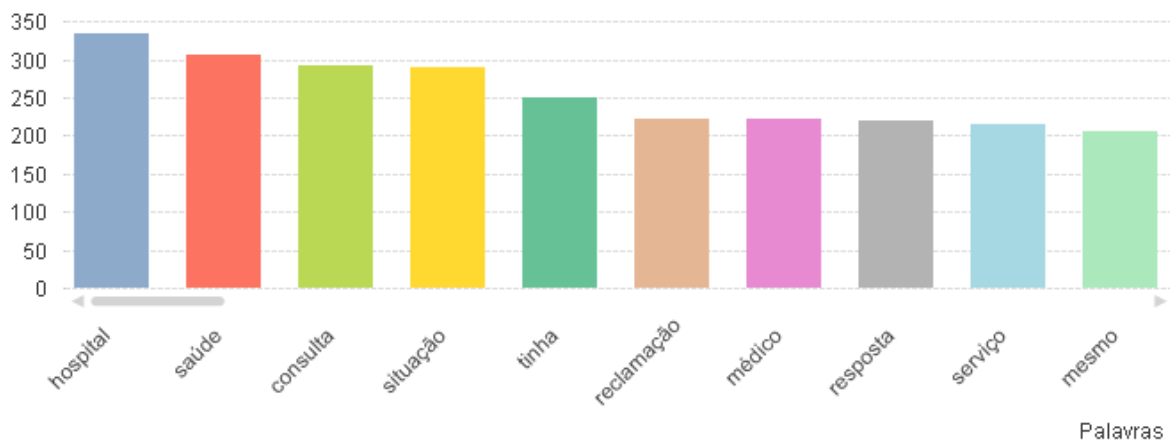


Figura 28 - Dashboard Frequência Total de Palavras

### 6.5.2 Estudo de Palavras por Tipologia

Esta análise contempla apenas dois campos de dados informacionais sendo estes a tipologia e o número de palavras.

Facilmente observamos que o intuito do desenvolvimento deste gráfico é meramente informativo pois todas as tipologias contam com 20 palavras cada e devido a esta razão que todos os quadrantes do gráfico (figura 29) têm a mesma dimensão.



Figura 29 - Dashboard de Palavras por Tipologia

### 6.5.3 Estudo da Frequência de Palavras por Tipologia

Este estudo foi desenvolvido tendo em conta todas as dimensões da base de dados. Desta forma poderemos visualizar através da figura 30 a relação de cada palavra com as tipologias existentes.

Neste gráfico de *plots* poderemos visualizar a ocorrência de cada palavra em cada tipologia. Através de um exemplo específico poderemos visualizar que a palavra consulta ocorre em todas as tipologias sendo que as tipologias onde esta palavra obtém uma maior frequência serão em “Direitos dos Utentes”, “Outros Assuntos” e “Tempo de Espera”, pois quanto maior for a dimensão dos *plots* maior é o real número de frequência de palavras. Vendo de outro prisma podemos verificar que algumas palavras são restritas a uma única tipologia como o caso da palavra “acesso” que surge apenas na tipologia “Acesso” ou o caso da palavra “aguardar” que surge apenas na tipologia “Tempo de Espera”.



Figura 30 - Dashboard Distribuição Geral de Palavras por Tipologia

## 6.6 Elaboração de modelos de Classificação de Reclamações

Após a realização de várias análises e estudos sobre os dados envolvidos neste projeto e as informações contidas nas reclamações verificou-se a necessidade realizar algumas experiências tecnológicas por forma a desenvolver uma resposta concreta para a questão central de toda esta dissertação.

Numa tentativa inicial de desenvolver um modelo de classificação automático de reclamações foi utilizado um conjunto de dados com informações relativas à tabela de reclamações *online* recolhida na Entidade Reguladora da Saúde. Seguidamente foram experimentadas diversas ferramentas disponíveis no mercado dedicadas à interpretação de dados em linguagem natural, pelo que todas as ferramentas utilizadas para o desenvolvimento da solução se revelaram inadequadas à resolução do problema em questão, apontando como principal fator desta tendência a inexistência de um dicionário de português técnico incorporado nos *softwares* utilizados, tornando-o assim incapaz de interpretar o texto relativo às reclamações e de identificar os principais termos contidos no mesmo. Desta forma, esta experiência revelou-se incapaz de oferecer uma resposta exata ao problema central da dissertação pondo em causa o desenvolvimento de um sistema capaz de classificar automaticamente as reclamações, sendo estritamente necessária a criação de um dicionário técnico a ser incorporado nos sistemas de exploração de dados de modo a facilitar a classificação de reclamações.

Após a realização do estudo de frequência de palavras contidas em reclamações por cada tipologia associada, foram delineadas novas experiências para tentar dar uma resposta positiva à questão da possibilidade de criação de um modelo de classificação.

Seguidamente à realização do estudo foi desenvolvido um *dataset* composto unicamente pelas 20 palavras mais frequentes em cada tipificação, contendo assim dois campos informacionais sendo o primeiro referente às palavras verificadas e o segundo referente à tipologia associada a cada uma das palavras. Para a realização desta experiência foi utilizado o *software weka*, escolhido pela sua larga oferta de algoritmos dedicados à análise e exploração de dados. Seguidamente foram utilizados sobre o conjunto de dados algoritmos de classificação, obtendo nas primeiras aplicações deste sistema resultados de instâncias corretamente classificadas na ordem dos 40%, ou seja resultados abaixo do expectável quando aplicado a um conjunto de dados previamente classificado. Por forma a tentar melhorar o valor de classificação o conjunto de dados foi retratado, através da remoção das colunas identificativas e da alteração da disposição de dados. Após o tratamento de dados realizaram-se novamente aplicações de algoritmos de classificação no *weka* sendo obtido um valor de instâncias corretamente classificadas de cerca de 80%, continuando a não ser um valor satisfatório para o conjunto de dados utilizado. Após a conclusão destas experiências perspetivou-se que o reduzido volume do conjunto de dados poderá ser o principal fator para o insucesso de classificação nestas últimas tentativas, juntamente com o facto de que as mesmas palavras apresentam percentagens de frequência semelhantes em tipologias diferentes sendo assim incapazes de categorizarem uma única tipologia.

Perspetivando a impossibilidade de oferecer uma resposta positiva à questão central deste projeto, ainda foi desenhada uma última experiência, que pressupunha o desenvolvimento de um conjunto de dados composto pelo grupo das 10 palavras mais frequentes em cada reclamação formulada sendo incluído também um campo para a tipologia associada à reclamação. Esta experiência tornou-se impossível de realizar pelo facto do sistema utilizado para análise de frequência de palavras, nomeadamente o *Knime*, revelar-se incapaz de oferecer a frequência de palavras individualizada de cada reclamação e pelo fator lógico de que na generalidade das reclamações analisadas é pouco frequente a utilização de mais do que duas vezes a mesma palavra numa única reclamação.

Por todas as razões apresentadas e todas as experiências realizadas, poder-se-á afirmar que atualmente ainda não é possível desenvolver um modelo de classificação automático de reclamações, sendo estritamente necessário para o caso o desenvolvimento de dicionário de português técnico capaz de ser integrado em *softwares* de análise e exploração de dados sobre forma textual.





## 7. MONITORIZAÇÃO DO PROJETO

### 7.1 Tarefas de Monitorização do Projeto

Este capítulo dedica-se a consagrar a última fase do Modelo de Tomada de Decisão de Simon, e desta forma pretende estabelecer métricas de avaliação e monitorização do projeto desenvolvido. Torna-se importante referir que todo o trabalho realizado, ao nível de implementação de modelos de *text mining* para averiguação e análise de frequência de palavras nas reclamações como *dashboards* de análise de informação serão implementados na Entidade Reguladora da Saúde para futura análise e validação da importância dos estudos e informação desenvolvidos e recolhidos nesta dissertação. Desta forma, será possível obter uma noção real e exata do impacto e dos contributos dos estudos desenvolvidos para a instituição em causa. Assim sendo, todo o trabalho e estudos desenvolvidos serão entregues aos agentes responsáveis e destacados pela Entidade Reguladora com intuito de implementar as análises desenvolvidas neste projeto em contexto real, para que estas sejam avaliadas, proporcionando o desenvolvimento de um conjunto de considerações mais especializado feito por responsáveis e especialistas da área.

Para realizar uma correta validação da informação é importante ter em atenção e monitorizar uma série de aspetos, sendo alguns destes em seguida apresentados:

- Categorização exata da reclamação quanto ao canal de comunicação em que a mesma foi processada;
- Número exato de reclamações em análise;
- Número exato de reclamações em análise para cada tipologia;
- Número exato de reclamações em análise para cada tipificação;
- Valor exato do ano mês e hora da reclamação;
- Nível de importância das palavras por cada reclamação;
- Identificação exata da frequência de palavras por cada tipologia;
- Nível de funcionalidade e utilidade da solução desenvolvida;
- Percentagem de reclamações recebidas que foram processadas pela solução.

## **7.2 Propostas de Melhoria de Dados e Condições do Estudo**

Após a realização deste estudo, foram traçadas algumas perspectivas que poderão melhorar a qualidade dos resultados obtidos, as condições de realização do estudo, e ideias para a realização de outras análises que não se viram concretizadas nesta dissertação devido a motivos externos. Em seguida serão apresentadas algumas propostas:

### **Eliminação da replicação de reclamações**

Desta forma, durante a realização deste projeto e nomeadamente em relação aos dados recolhidos na Entidade Reguladora da Saúde, foi denotado a existência de várias reclamações consecutivamente replicadas. Isto é um facto que não deveria acontecer pois para além de causar confusão na realização deste projeto e complicar a compreensão dos dados, esta situação poderá causar confusão na Entidade Reguladora, pois poderá a reclamação ser “sobre-analisada” sendo traçadas medidas corretivas ou punitivas sobre o reclamado mais severas do que seria normalmente empregues, devido ao fator de suposta reincidência. A situação contrária à referida anteriormente também poderá acontecer, isto é poderá a mesma reclamação ser entregue para análise a dois agentes supervisores diferentes e nenhum destes analisara reclamação devido a esta já deter o estatuto de analisada, ou até os dois agentes analisarem a mesma reclamação e ambos traçarem perspectivas diferentes sobre o mesmo documento.

### **Melhoria das condições de extração de reclamações**

Durante a realização deste projeto e mais uma vez na fase de análise e exploração de dados, foi verificado que o corpo da reclamação ou seja a componente descritiva da reclamação, nomeadamente no conjunto de dados recolhidos, contém diversos erros e falsos caracteres. Este facto poderá dificultar a leitura do corpo da reclamação e induzir o agente supervisor em erro. Visto que não é possível traçar uma solução exata para o problema de erros gramaticais por parte do Reclamante seria interessante pelo menos resolver o problema da inserção de falsos caracteres no corpo da reclamação, por parte dos sistemas de bases de dados ou dos sistemas de recolha de reclamações, nomeadamente as reclamações *online*.

### **Melhoria do Desenho da Base de Dados**

Na fase de estudo e recolha de dados foi verificado que por razões externas a base de dados utilizada não apresenta uma estrutura eficiente apresentando algumas deficiências que poderão constituir-se como dificuldades para futuros trabalhos ou estudos a desenvolver-se sobre a mesma. Desta forma é

importante referir a existência de chaves primárias de algumas tabelas que são enunciadas como chaves estrangeiras noutras tabelas mas com denominação diferente. Um exemplo deste caso poderá ser visível na figura 6 desta dissertação que representa o modelo relacional de base de dados desenvolvido em que é possível verificar que a chave primária da tabela “ers\_reclamacoes” designada como “numero\_folhadereclamacao” é a chave estrangeira na tabela “ers\_reclamacoesonline” sendo aqui designada como “numero”. Esta situação deveria ser solucionada com uma simples reestruturação da base de dados de modo a adequar corretamente os campos de dados.

### **Melhoria do Processo de Preenchimento de Reclamações Online**

As condições de formulação de reclamações *online* deveriam ser melhoradas na perspectiva de obrigar ou incentivar o reclamante a preencher todos os campos informacionais, pois desta forma poderia ser suprimida uma lacuna verificada na análise dos dados, sendo esta a existência de campos informacionais em branco ou sem valor atribuído.

### **Estudo relativo à informação da hora da reclamação**

Esta dissertação contém um estudo que foi pensado para outro tipo de dados, sendo o estudo visível na secção 6.4.4 desta dissertação. Neste estudo realizado é possível verificar a distribuição de reclamações por horas do dia. Desta perspectiva o estudo contempla com o campo “horadecriacao” a hora em que a reclamação entrou nos sistemas da Entidade Reguladora da Saúde. Seria interessante desenvolver um estudo semelhante com outro campo informacional, sendo este a hora em que se deu a situação descrita na reclamação, pois a hora em que é preenchida ou formulada a reclamação poderá não corresponder à mesma hora em que se deu a situação descrita no corpo da reclamação. Com este novo atributo é possível realizar um estudo semelhante ao da secção 6.4.4 mas que do qual seria obtido outro tipo de informações. Nomeadamente neste hipotético estudo seria criado certamente um conjunto de novas informações das quais poderia ser verificado quais os turnos médicos que produziam mais reclamações denotando assim uma das áreas que poderia sofrer questões de melhoria.



## 8. CONCLUSÃO

### 8.1 Considerações Finais

Para concluir a realização deste documento, poderemos visualizar que aqui se consagra o desenvolvimento de uma dissertação que procura responder à questão da possibilidade de categorizar a informação recolhida em estabelecimentos prestadores de cuidados de saúde através de técnicas dedicadas à análise e exploração de dados. Para a realização deste trabalho foi estabelecida uma parceria entre a Entidade Reguladora da Saúde e um grupo de investigação da Universidade do Minho. Neste projeto foram recolhidos dados sobre as reclamações obtidas pela Entidade Reguladora, sendo que ficou definido em reuniões com os responsáveis pela instituição que se deveria garantir a confidencialidade dos dados. Após a recolha dos dados, sobre a forma de uma base de dados informacional contendo todos os campos informacionais que compõe a formulação de uma reclamação, estes foram devidamente estudados e analisados com a finalidade de perceber quais os estudos que poderiam ser aplicados e desenvolvidos sobre reclamações. Para o desenvolvimento de todo o trabalho foi adotado uma metodologia combinada, sendo esta composta por três metodologias diferentes: a *Design Science Research*, a *CRISP-DM* e a metodologia dedicada ao processo de Tomada de Decisão consagrada por Herbert Simon. Seguindo os ensinamentos agrupados nesta metodologia combinada foi desenvolvida uma ontologia sobre reclamações. Desta forma poderemos visualizar todas as classes que compõem o processo de gestão de reclamações, contendo também informações sobre as relações entre as classes e o fluxo de informação implícito a este processo de gestão, sendo que a ontologia foi desenvolvida respeitando a característica de universalidade de modo a que este estudo ontológico poderá ser utilizado a nível global por outros grupos de investigação em projetos de cariz e área semelhante. Posteriormente ao desenho da ontologia foi realizado um estudo e compreensão dos dados submetidos a análise, na tentativa de perceber os dados que seriam utilizados e perspetivando os estudos que seriam desenvolvidos. Nesta fase do projeto, foi decidido pela equipa de trabalho que os dados que seriam utilizados para análise seriam unicamente os dados remetentes a reclamações *online*. Esta decisão deveu-se a uma maior integridade deste conjunto de dados e ao facto de que estes encontram-se praticamente inalterados desde a criação da reclamação por parte do reclamante. Posteriormente este conjunto de dados composto pelas reclamações *online* foi submetido a técnicas de *text mining* onde

foram estudadas sobretudo a frequência de palavras nas reclamações sendo realizado em seguida diversos estudos sobre a frequência de determinadas palavras para as várias tipologias atribuídas às reclamações. Simultaneamente à realização destes estudos foram estudadas através de técnicas de visualização, ou seja através da análise e visualização de informações, juntamente com técnicas de extração de informação sobre documentos textuais, as distribuições de reclamações por diversas dimensões, tendo em conta as tipificações e tipologias que são designadas a cada reclamações e dimensões relacionadas com espectro temporal, originando assim um conjunto de *dashboards* preparados para análise de reclamações em tempo real. Relativamente à questão de investigação foi possível verificar que atualmente e tendo em conta o conjunto de dados que é registado, à semelhança entre reclamações (reclamações de diferentes tipologias com um número elevado de palavras iguais ou semelhantes) e todas as experiências efetuadas ainda não é possível criar modelos capazes de categorizar automaticamente as reclamações. Este facto é sustentado pelas experiências e testes realizados para tentar desenvolver um sistema de classificação automática com base em informações e palavras que constituem as reclamações analisadas que obteve taxas de acuidade muito reduzidas e de erro muito elevadas

Na conclusão do projeto e tendo em conta as informações recolhidas nas diversas análises realizadas salienta-se que toda a informação recolhida e processada será entregue à Entidade Reguladora da Saúde para posterior análise por especialistas para que estes possam tirar conclusões mais adequadas sobre os estudos realizados, sendo que este projeto conta também com uma componente de questões de melhoria que poderão ser aplicadas para melhorar o desempenho e resultados das análises realizadas.

## **8.2 Limitações e Dificuldades**

Este foi um projeto de exploração submetido para realização como projeto de Dissertação de Mestrado. Desta forma foi observável desde o início da realização do projeto que algumas questões relacionadas com a definição dos requisitos e da viabilidade do projeto conforme o delineado inicialmente ainda não estavam completamente definidos. Este fator alinhado com a dificuldade da recolha de dados na entidade responsável, em relação ao que foi planeado inicialmente, fez atrasar um pouco o desenvolvimento inicial do projeto.

Das sucessivas reuniões com os responsáveis pela Entidade Reguladora da Saúde foi acordado que neste projeto não seriam trabalhados dados identificativos tanto do reclamante como do reclamado, assim garantindo confidencialidade dos dados e respeito pela privacidade das informações trabalhadas.

Por último, um dos fatores que mais dificultou o desenvolvimento do trabalho e diminui o espectro de resultados esperados prendeu-se com a tecnologia utilizada e existente para a resolução do projeto, sendo esta alinhada com uma inexperiência inicial ao projeto, no campo do *text mining*. Foi verificado que nas ferramentas utilizadas para o desenvolvimento dos estudos contidos nesta dissertação inexistia a capacidade de leitura de textos em linguagem natural portuguesa, este fator dificultou a resolução de determinados estudos perspetivados inicialmente pela incapacidade de compreensão de informação por parte dos *softwares* utilizados. No desenvolvimento do trabalho foram consecutivamente exploradas diversas ferramentas existentes no mercado, sendo que nenhuma delas contemplava os resultados esperados devido à inexistência de *softwares* de análise de dados capazes realizarem estudos sobre textos de língua portuguesa ou que na sua constituição incluam a capacidade de interpretação de palavras através da inclusão de um dicionário de português técnico. Posteriormente à consideração deste facto foram delineadas algumas experiências na tentativa de construir um modelo ou exemplo para a questão de classificação. Neste sentido, foi construído um *dataset* contendo unicamente as palavras mais frequentes de cada tipificação sobre o qual foram aplicados algoritmos de classificação sendo que os resultados foram completamente insatisfatórios. Da decorrente análise de resultados foi determinado que a razão principal de insucesso devia-se à reduzida amostra de dados, na qual tornar-se-ia impossível a determinação de um padrão entre os mesmos. Foi delineada ainda outra experiência, consistindo esta na construção de um conjunto de dados composto pela determinação das palavras mais utilizadas em cada reclamação juntamente com a tipificação associada, sendo que construção deste grupo de dados tornou-se insustentável incapacitando a realização da experiência.

Devido a todas estas razões torna-se essencial para o desenvolvimento de um sistema automático de classificação e tipificação de reclamações e a criação de um dicionário técnico capaz de ser integrado com os sistemas de análises existentes no mercado.

### 8.3 Análise de Riscos

A estratégia de identificar atempadamente os riscos para o projeto e definir estratégias de atenuação dos efeitos dos mesmos é essencial para um projeto desta dimensão, sendo esta análise geralmente produzida nos momentos iniciais do desenvolvimento do projeto este artefacto deverá ser tido em atenção durante toda a sua execução. Seguindo isto, a tabela 15 apresenta uma retrospectiva de todos os riscos que afetaram o desenvolvimento do projeto contendo uma descrição dos riscos identificados, contando com uma previsão do seu impacto para o projeto incluindo também a respetiva ação atenuante.

Tabela 15 - Análise de Riscos

<b>Descrição</b>	<b>Probabilidade</b>	<b>Impacto</b>	<b>Seriedade<sup>2</sup></b>	<b>Ação Atenuante</b>	<b>Risco Verificado</b>
Complexidade do Projeto	4	5	20	Desdobramento do projeto em tarefas e atividades. Delineação inicial dos requisitos e objetivos esperados.	Sim
Má Gestão do Projeto	4	5	20	Planeamento deve ser adequado. Alocação de tempo extra para resolução do projeto.	Não
Atraso na Disponibilização dos Dados do Projeto	4	5	20	Comunicação constante com a entidade fornecedora dos dados do projeto	Sim

---

<sup>2</sup> Seriedade = Probabilidade \* Impacto



				para viabilizar a disponibilização	
Má compreensão dos dados	4	4	16	Estudo orientado da informação contida no <i>dataset</i> . Colaboração com a instituição fornecedora dos respectivos dados.	Não
Perda dos Dados do Projeto	2	5	10	Realizar cópias de segurança dos dados envolvidos no projeto. Armazenamento na base de dados da Universidade.	Não
Fraca qualidade dos dados	4	5	20	Consulta dos orientadores para auxílio na compreensão dos dados.	Não
Incapacidade de responder ao problema central do projeto	4	5	20	Justificação fundamentada dos resultados obtidos e da razão da incapacidade	Sim
Falta de experiência na manipulação das ferramentas	4	5	20	Consulta de tutoriais para aquisição de experiência. Pedido de apoio aos	Sim

				orientadores do projeto	
Escolha de técnicas inadequadas	3	4	12	Estudo intensivo sobre as técnicas disponíveis para escolha da melhor	Não
Utilização de metodologias inadequadas	2	4	8	Desenvolvimento de uma metodologia combinando ensinamentos de várias	Não
Atraso nos prazos de entrega	2	5	10	Marcação de datas de controlo do projeto	Não

## 9. REFERÊNCIAS

- AKINúmeros. (2014). *AKINúmeros Portugal em números e em tempo real!* (OK-Sites) Obtido de <http://akinumeros.com/>
- Almeida, L. d. (2010). *A criação da ERS*. (Comunicação por conta e risco) Obtido em Dezembro de 2014, de <https://porcontaerisco.wordpress.com/2010/12/03/a-criacao-da-entidade-reguladora-da-saude-em-portugal/#comments>
- Ammenwerth, E., Graber, S., Hermmann, G., Burkle, T., & Konig, J. (2003). Evaluation of health information systems—problems and challenges. *International Journal of Medical Informatics*, 71(2-3), 125-135. doi:doi:10.1016/S1386-5056(03)00131-X
- Andrienko, G. L., & Andrienko, N. (1998). *Knowledge Extraction from Spatially Referenced Databases: a Project of an Integrated Environment*. California, USA: Varenus Workshop on Status and Trends in Spatial Analysis.
- Benito, G. A., & Licheski, A. P. (2009). *Sistemas de Informação apoiando a gestão do trabalho em saúde*. Brasil: Revista Brasileira de Enfermagem.
- Berg, M. (2001). Implementing information systems in health care organizations: myths and challenges. *International Journal of Medical Informatics*, 64(2-3), 143-156. doi:10.1016/S1386-5056(01)00200-3
- Berg, Marc. (1999). Patient care information systems and health care work: a sociotechnical approach. *International Journal of Medical Informatics*, 55(2), 87-101. doi:doi:10.1016/S1386-5056(99)00011-8
- Bimbaum, L. A., & Collins, G. C. (1991). Learning in Intelligent Information Retrieval. Em *Machine Learning: Proceedings of the International Workshop* (pp. 235-239). Chicago, USA: University of Chicago.
- Branco, M. A. (1996). *Sistemas de informação em saúde no nível local*. Fundação Oswaldo Cruz, Núcleo de Estudos em Saúde Coletiva, Centro de Pesquisas Aggeu Magalhães. Rio de Janeiro, Brasil: Cadernos de Saúde Pública.
- Buckingham, R. A., Hirschheim, R. A., Land, F. F., & Tully, C. J. (1987). Information Systems Curriculum: a Basis for Course Design. Em *Information systems education: recommendations and implementation* (pp. 14 - 133). New York, U.S.A.: Cambridge University Press.

- Carvalho, J. Á. (2000). Information System? Which One do You Mean? *Information Systems Concepts: An Integrated Discipline Emerging*. Leiden, Holanda.
- Chemchem, A., & Drias, H. (2015). *From data mining to knowledge mining: Application to intelligent agents*. Expert Systems with Applications.
- Chen, M.-S., Han, J., & Yu, P. (1996). Data mining: an overview from a database perspective. *Knowledge and Data Engineering, IEEE Transactions*, 8. doi:10.1109/69.553155
- CRISP-DM. (2014). *CRISP-DM*. Obtido de <http://crisp-dm.eu/reference-model/>
- CRISP-DM. (2014). *CRISP-DM*. Obtido em Dezembro de 2014, de <http://crisp-dm.eu/reference-model/>
- Ein-Dor, P., & Segev, E. (1993). *A Classification of Information Systems: Analysis and Interpretation*. Tel Aviv University, Faculty of Management. Information Systems Research. doi:<http://dx.doi.org/10.1287/isre.4.2.166>
- Entidade Reguladora da Saúde. (2009). *ERS Livro de Reclamações – Regras e Obrigações dos Prestadores de Cuidados de Saúde*. Obtido em Janeiro de 2015, de [https://www.ers.pt/pages/106?news\\_id=180](https://www.ers.pt/pages/106?news_id=180)
- Entidade Reguladora da Saúde. (2011). *Relatório sobre "A carta dos Direitos dos Utentes"*. Porto: Entidade Reguladora da Saúde.
- Entidade Reguladora da Saúde. (2013). *Relatório de Atividades de 2013*. Porto: Entidade Reguladora da Saúde.
- Falkenberg, E. D., Hesse, W., Lindgreen, P., Nilsson, B. E., Han Oei, J. L., Rolland, C., . . . Voss, K. (1998). *A Framework of Information System Concepts*. University of Leiden, Department of Computer Science. The Netherlands: FRISCO Report.
- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the Power of Text Mining. *Communications of the ACM*, 49(9), 77-88.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Widener, T. (1996). *The KDD process for extracting useful knowledge from volumes of data*. Communications of the ACM.
- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17. doi:<http://www.aaai.org/ojs/index.php/aimagazine/article/view/1230/1131>
- Feldman, R., & Sander, J. (2007). *The Text Mining Handbook*. Cambridge University Press. Obtido de <http://www.google.pt/books?hl=pt->

PT&lr=&id=U3EA\_zX3ZwEC&oi=fnd&pg=PR1&dq=text+mining+&ots=2MxJHjGzKB&sig=fJRZkio  
31KwAjue73mzaA7MCUNE&redir\_esc=y#v=onepage&q=text%20mining&f=true

- Fidalgo, P. M. (2013). *Sintomatologia do Desalinamento de Sistemas de Informação*. Dissertação de Mestrado em Engenharia e Gestão de Sistemas de Informação, Universidade do Minho, Departamento de Sistemas de Informação, Guimarães.
- Firmino, C. F. (2011). *Reclamações: A sua importância nas Unidades de Saúde Privadas*. Lisboa: Universidade Nova de Lisboa. Escola Nacional de Saúde Pública.
- Frawley, W. J., Piatetsky-Shapiro, G., & Matheus, C. J. (1992). Knowledge Discovery in Databases: An Overview. *AI Magazine*, 13(3).
- Han, J., & Kamber, M. (2006). So, what is data mining? Em *Data Mining concepts and techniques*. Elsevier, Inc. doi:[http://www.google.pt/books?hl=pt-PT&lr=&id=AfL0t-YzOrEC&oi=fnd&pg=PP2&dq=data+mining&ots=UwTYrV7tA4&sig=AjIMt-4IGJ8vqgR\\_SaZrIntsE8s&redir\\_esc=y#v=onepage&q&f=true](http://www.google.pt/books?hl=pt-PT&lr=&id=AfL0t-YzOrEC&oi=fnd&pg=PP2&dq=data+mining&ots=UwTYrV7tA4&sig=AjIMt-4IGJ8vqgR_SaZrIntsE8s&redir_esc=y#v=onepage&q&f=true)
- Han, Jiawei; Kamber, Micheline. (2000). *Data Mining: Concepts and Techniques*. California, USA: Elsevier.
- Hand, D., Mannila, H., & Smyth, P. (2001). Principles of Data Mining. Massachusetts Institute of Technology. doi:[http://www.google.pt/books?hl=pt-PT&lr=&id=SdZ-bhVhZGYC&oi=fnd&pg=PR17&dq=data+mining&ots=yxL8trpupZ&sig=QxisCfrpnRBSA1gYMtk9UJ01bFw&redir\\_esc=y#v=onepage&q=data%20mining](http://www.google.pt/books?hl=pt-PT&lr=&id=SdZ-bhVhZGYC&oi=fnd&pg=PR17&dq=data+mining&ots=yxL8trpupZ&sig=QxisCfrpnRBSA1gYMtk9UJ01bFw&redir_esc=y#v=onepage&q=data%20mining)
- Haux, R. (2006). *Health information systems – past, present, future*. Technical University of Braunschweig, Institute for Medical Informatics, Braunschweig, Alemanha. doi:<http://dx.doi.org/10.1016/j.ijmedinf.2005.08.002>
- Hearst, M. (2003). What Is Text Mining? Obtido de <http://www.jaist.ac.jp/~bao/MOT-Ishikawa/FurtherReadingNo1.pdf>
- Hearst, Marti. (2003). *What Is Text Mining?* Berkeley, USA: Universidade de Berkeley.
- Heeks, R. (2006). Health information systems: Failure, success and improvisation. (Elsevier, Ed.) *International Journal of Medical Informatics*, 75(2), 125-137. doi:[doi:10.1016/j.ijmedinf.2005.07.024](https://doi.org/10.1016/j.ijmedinf.2005.07.024)
- Herver, A., & Chatterjee, S. (2010). Information Systems Research. *Springer* (Design Science Research in Information Systems).

- Hotho, A., Nurnberger, A., & Paab, G. (2005). *A Brief Survey of Text Mining*. Alemanha. Obtido de <http://www.kde.cs.uni-kassel.de/hotho/pub/2005/hotho05TextMining.pdf>
- Jornal de Noticias. (2008). Saúde tem livro de reclamações online. *Jornal de Noticias*(Remodelação das urgências), Página Web. Obtido de  
[http://www.jn.pt/Dossies/dossie.aspx?content\\_id=886638&dossier=Remodela%E7%E3o%20das%20urg%Eancias](http://www.jn.pt/Dossies/dossie.aspx?content_id=886638&dossier=Remodela%E7%E3o%20das%20urg%Eancias)
- Leal, L. M., & Rodrigues, F. (2003). *Text Mining*. ISEP, Departamento de Engenharia Informatica, Porto. Obtido de <http://www.dei.isep.ipp.pt/~paf/proj/Set2003/Text%20Mining.pdf>
- Lobo, A. F. (2004). *Entre o direito a reclamar e o direito à saúde. Serviço social em gabinetes do cidadão, SNS*. Coimbra: Instituto Superior Miguel Torga, Escola Superior de Altos Estudos.
- Maimon, O., & Rokach, L. (2010). Introduction to Knowledge Discovery and Data Mining. Em *Data Mining and Knowledge Discovery Handbook* (pp. 1-15). Israel: Springer.
- Manning, C. D., Raghavan, O., & Schütze, H. (2008). Introduction to Information Retrieval. *Cambridge University Press*(Information Retrieval). Obtido de [http://www.langtoninfo.co.uk/web\\_content/9780521865715\\_frontmatter.pdf](http://www.langtoninfo.co.uk/web_content/9780521865715_frontmatter.pdf)
- Marin, H. d. (2010). Sistemas de informação em saúde: considerações gerais. *Journal of Health Informatics*, 2(1), 20-24.
- Marreiros, M. G. (2007). *Agentes de Apoio à Argumentação e Decisão em Grupo*. Braga, Portugal: Universidade do Minho.
- Martinho, R., Varajão, J., Cruz-Cunha, M. M., & Baloni, A. (2014). *Tecnologias e Sistemas de Informação em entidades hospitalares – Dois casos de hospitais portugueses*. Guimarães: Ministério de Saúde do Brasil.
- Mason, N. J. (2006). Is operation research really research? *Operations Research Society of South Africa*, 22(2), 155-180.
- Mattiello da Silva, J. V., & Da Costa, R. M. (2013). *DESIGN RESEARCH É UMA METODOLOGIA DE APLICAÇÃO PRÁTICA?* Brasil: UNISINOS.
- Mcdonald, D., & Kelly, U. (2012). *Value and Benefits of Text Mining*. (Jisc) Obtido em Fevereiro de 2015, de <http://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining>
- Mitra, M., & Chaudhuri, B. B. (2000). Information Retrieval from Documents: A Survey. *Information Retrieval*, 141-163.

- O'brien, J. A. (1993). *Management information systems: a managerial end user perspective* (Universidade de Michigan ed.). Irwin.
- Oliveira, A., Portela, C. F., Santos, M. F., Machado, J., Abelha, A., Neves, J. M., . . . Silva, Á. (2015). *Developing an Ontology for Health Complaints Management*. Braga, Portugal: KMIS Conference.
- Oliveira, J. (2014). *Percursos de Carreira dos Profissionais de Tecnologias e Sistemas de Informação*. Universidade do Minho, Departamento de Sistemas de Informação, Guimarães.
- Peffer, K. (2008). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 24(3), 45-77.
- Perez, G., & Zwicker, R. (2006). *Adoção de Inovações Tecnológicas: Um Estudo sobre o uso de Sistemas de Informação na Saúde*. São Paulo, Brasil: Universidade de São Paulo.
- Piatetski-Shapiro, G., & Frawley, W. (1991). *Knowledge Discovery in Databases*. MA, USA: MIT Press Cambridge.
- PORDATA. (2012). *PORDATA base de dados de Portugal contemporaneo*. (Pordata) Obtido de <http://www.pordata.pt/Portugal/Ambiente+de+Consulta/Tabela>
- Portela, C. F. (2009). *Sistemas de apoio à decisão na medicina intensiva baseados na descoberta de conhecimento em base de dados*. Dissertação de Mestrado, Universidade do Minho, Departamento de Sistemas de Informação, Guimarães.
- Prather, J. C., Lobach, D. F., Goodwin, L. K., Hales, J. W., Hage, M. L., & Hammond, W. E. (1997). Medical data mining: knowledge discovery in a clinical data warehouse. *AMIA Annu Fall Symp*. Obtido de <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2233405/>
- Ramos, C., & Lobo, F. (2003). Descoberta de conhecimento em bases de dados. Em *Descoberta de conhecimento em bases de dados*. Algarve: Universidade do Algarve.
- Ramos, I., & Santos, M. Y. (2003). *Data Mining no suporte à construção de Conhecimento Organizacional*. Guimarães: Associação Portuguesa de Sistemas de Informação.
- Ramos, L. F. (2014). *Deteção e Caracterização Geo-Espacial das Zonas de Acumulação de Acidentes Rodoviários*. Universidade do Minho, Departamento de Sistemas de Informação, Guimarães.
- Rodrigues, D. F. (2010). *Implementation of an Information Retrieval System within a Central Knowledge Management System*. Tese de Mestrado, Faculdade de Engenharia da Universidade do Porto, Departamento de Engenharia de Computadores, Porto.
- Rodrigues, F. S. (2007). *A Comunicação nos Serviços de Saúde. Um Estudo sobre as Reclamações do SNS Português*. Universidade Aberta.

- Sancho, P., & Quintas, P. (2009). *Ordem dos Médicos - ERS Entidade Reguladora da Saúde*. (Ordem dos Médicos) Obtido em Janeiro de 2015, de <https://www.ordemdosmedicos.pt/?lop=conteudo&op=ed3d2c21991e3bef5e069713af9fa6ca&id=e2c420d928d4bf8ce0ff2ec19b371514>
- Santos, M. F., & Azevedo, C. S. (2005). *Data Mining – Descoberta de conhecimento em bases de dados*. Guimarães: FCA - Editora Informática.
- Santos, M. Y., & Ramos, I. (2006). *Business Intelligence – Tecnologias da Informação na Gestão de Conhecimento*. Guimarães: FCA - Editora Informática.
- Saúde, E. -E. (2014). *Relatório de Atividades de 2013*. Porto: ERS. doi:[https://www.ers.pt/uploads/document/file/4192/Relat\\_rio\\_de\\_actividades\\_2013.pdf](https://www.ers.pt/uploads/document/file/4192/Relat_rio_de_actividades_2013.pdf)
- Simões, J. A. (2004). *As parcerias público-privadas no sector da saúde em Portugal*. Ministério da Saúde. Parcerias Público-Privadas.
- Simon, H. A. (1960). *The New Science of Management Decision Making*. New York-Evanston.
- Soares, F. d. (2006). *Interpretação da segurança de sistemas de informação segundo a teoria da acção*. Tese de Doutoramento, Universidade do Minho, Departamento de Sistemas de Informação, Guimarães.
- Spasic, I., Ananiadou, S., Mcnaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: Making sense of raw text. *Briefing in Bioinformatics*, 6(3), 239-251.
- Weiss, S. M., Indurkha, N., Zhang, T., & Damerau, F. J. (2005). *Text Mining Predictive Methods for Analyzing Unstructured Information*. Springer Science Business Media, Inc. Obtido de [http://www.google.pt/books?hl=pt-PT&lr=&id=NZteXd4qf9sC&oi=fnd&pg=PP5&dq=text+mining+&ots=1iA0aQai6M&sig=D8QOaa74k-HSP-CeoyDx2JPB\\_7Q&redir\\_esc=y#v=onepage&q=text%20mining&f=true](http://www.google.pt/books?hl=pt-PT&lr=&id=NZteXd4qf9sC&oi=fnd&pg=PP5&dq=text+mining+&ots=1iA0aQai6M&sig=D8QOaa74k-HSP-CeoyDx2JPB_7Q&redir_esc=y#v=onepage&q=text%20mining&f=true)
- Weiss, Sholom M.; Indurkha, Nitin; Zhang, Tong; Damerau, Fred J. (2005). *Text Mining Predictive Methods for Analyzing Unstructured Information*. USA: Springer.



## **ANEXO I – PUBLICAÇÃO CIENTÍFICA**

### **Developing an Ontology for Health Complaints Management**

**Autores:** André Oliveira, Carlos Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha, José Maria Neves, Suzana Vaz e Álvaro Silva

**Conferência:** KMIS 2015 - 7th International Conference on Knowledge Management and Information Sharing

**Ano:** 2015

**Abstract:** The dissatisfaction of healthcare institutions users has increased in Portugal in the recent years. This fact can be seen through the increase of complaints that the entity responsible in this country has been receiving lately. More and more technical efforts has been done to understand and analyse this tendency. In this paper the authors are proposing an ontology about the whole process of complaints management from healthcare institutions. All the work was developed after analysing the entire process and the data collected by the entity responsible with this matter in Portugal. The ontology developed can show the main concepts involved in the process and the relationship between them. As main ontology entities are person, document, measure and status.

**Keywords:** Ontology, Complaints Management, Healthcare, Information System