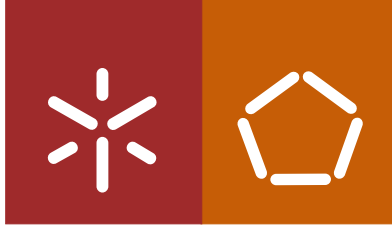


Universidade do Minho
Escola de Engenharia

Andreia Manuela Couto Brandão

**Business Intelligence: Indicadores da
Interrupção Voluntária da Gravidez**



Universidade do Minho
Escola de Engenharia

Andreia Manuela Couto Brandão

Business Intelligence: Indicadores da Interrupção Voluntária da Gravidez

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica
Ramo de Informática Médica

Trabalho efetuado sob orientação do
Professor Doutor António Carlos da Silva Abelha
e coorientação do
Doutor Carlos Filipe Portela

outubro de 2014

Declaração

Nome: Andreia Manuela Couto Brandão

Endereço eletrónico: andreiacoutobrandao@gmail.com

Cartão de Cidadão: 13859094

Título da Dissertação: Business Intelligence: Indicadores da Interrupção Voluntária da Gravidez

Orientador: Professor Doutor António Carlos da Silva Abelha

Co-orientador: Doutor Carlos Filipe Portela

Ano de conclusão: 2014

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica

Área de Especialização: Ramo de Informática Médica

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

Universidade do Minho, __ / __ / __

Assinatura: _____

Agradecimentos

Depois de terminada mais uma etapa crucial na minha vida acadêmica e pessoal, tenho que deixar um agradecimento a todas as pessoas que possibilitaram e que contribuíram para a concretização deste projeto. Assim sendo, deixo aqui o meu obrigado à minha família, por estarem sempre presentes e por me apoiarem em tudo.

Ao meu coorientador, o Doutor Filipe Portela, que me auxiliou na realização desta dissertação e que esteve sempre presente para disponibilizar conhecimentos e sugestões bastante úteis e muito proveitosas.

Ao meu orientador, o Professor Doutor António Abelha, e também ao Professor Doutor José Machado pela disponibilidade, pela presença, pela motivação e pelos conhecimentos transmitidos ao longo deste ano.

A todos os profissionais que me prestaram auxílio na realização deste projeto, nomeadamente à engenheira Sara Araújo e à enfermeira Alcina Borges, por toda a informação disponibilizada.

À minha colega de trabalho, Eliana Pereira, por ter sido uma excelente companheira, pela motivação e pelo grande apoio que me prestou durante a realização deste projeto.

Por último, mas não menos importante, a todas as minhas amigas que convivem comigo diariamente, em especial à Ana Novo, por todo o apoio, por todo o companheirismo e pelos incentivos prestados.

Resumo

Nos últimos anos, tem surgido um grande interesse na aplicabilidade das tecnologias de [Business Intelligence \(BI\)](#) na área da saúde. A grande satisfação resultante da sua implementação em outras áreas fez com que os profissionais de saúde juntamente com os profissionais de [Tecnologias de Informação \(TI\)](#) cooperassem para a conceção e o desenvolvimento de plataformas de BI em ambiente clínico.

A grande motivação para a sua implementação adveio da possibilidade de conceber um [Sistema de Apoio à Decisão \(SAD\)](#) médica, que suportasse o processo de tomada de decisão e que permitisse que este fosse realizado de forma mais rápida e eficaz. Além disso, com a implementação do [Processo Clínico Eletrónico](#), surgiu a necessidade de dar usabilidade aos dados dos utentes armazenados nas bases de dados, tornando-se isso possível com a implementação da plataforma de BI.

Neste projeto, incidiu-se numa área médica específica dentro da especialidade de Ginecologia e Obstetrícia, relacionada com a [Interrupção Voluntária da Gravidez \(IVG\)](#), por ser uma unidade relativamente recente, onde não existe nenhuma tecnologia semelhante implementada e, ainda, por ser um foco de interesse para os profissionais de saúde. O local de estudo foi o [Centro Materno Infantil do Norte \(CMIN\)](#), pertencente ao [Centro Hospitalar do Porto \(CHP\)](#), onde se teve acesso às fontes de dados necessárias para o desenvolvimento da plataforma de BI.

Além disso, foram também desenvolvidos modelos de [Data Mining \(DM\)](#), igualmente integrados na plataforma de BI, que permitem prever quais as utentes que recorrem à [IVG](#) que constituem um grupo de risco e quais as utentes que necessitam de acompanhamento da equipa de enfermagem du-

rante o processo da **IVG**. Os resultados obtidos foram bastante satisfatórios, uma vez que foram registados valores de 93% para a métrica da sensibilidade na questão relacionada com a probabilidade das utentes pertencerem ao grupo de risco, e valores de 91% e de 87% para a sensibilidade e acuidade, respetivamente, no problema relacionado com a previsão do local de realização de uma das etapas do processo de **IVG**.

Na seleção da tecnologia a utilizar para o desenvolvimento da plataforma de **BI**, optou-se pelo Pentaho BI Suite, depois de realizada uma pesquisa aprofundada sobre ferramentas *open source*.

Após a implementação da plataforma de **BI**, pode-se afirmar que o resultado foi satisfatório, uma vez que todos os indicadores de desempenho requisitados foram apresentados. Nestes indicadores, estão incluídos a distribuição das utentes por idades, por profissão, por localidade, pela presença na última consulta de avaliação, entre outros. Além disso, a informação representada é credível, pois esta foi submetida a um processo de validação por parte dos profissionais de saúde.

Abstract

In recent years, there has been great interest in the applicability of **Business Intelligence (BI)** technologies in healthcare. The great satisfaction resulting from its implementation in other areas meant that health professionals along with **Information Technologies (IT)** professionals cooperate for the design and development of **BI** platforms in a clinical environment.

The major motivation for its implementation came from the possibility of designing a medical support decision system, that supports the decision making process and allows this to be done more quickly and efficiently. Furthermore, with the implementation of electronic medical process, the need arose to give the usability of users data stored in databases, making it possible with the implementation of the **BI** platform.

This project is focused on a specific medical area, within the Gynaecology and Obstetrics specialty, related to **Voluntary Interruption of Pregnancy (VIP)**, being a relatively new unit, where there is no similar technology implemented and also for being a focus of interest for health professionals. The study site was the **Centro Materno Infantil do Norte (CMIN)** belonging to the **Centro Hospitalar do Porto (CHP)**, where it was possible to access to the data sources necessary for the development of the **BI** platform.

In addition, **Data Mining (DM)** models were developed, also integrated into the **BI** platform, that can predict which patients that resort to the **VIP** belong to the risk group and which patients require monitoring of the nursing team during the **VIP** process. The results were quite good, since it was reached a value of 93% of sensitivity in the question related to the probability of patients belong to the risk group, and values of 91% and 87% for sensitivity and accuracy, respectively, in the problem related to the probability of

patients require monitoring of the nursing team during the [VIP](#) process .

In selecting the software to be used for the development of [BI](#) platform, the choice was the Pentaho BI Suite, after conducting a thorough research on open source tools.

After the implementation of the [BI](#) platform, it can be stated that the result was satisfactory, since all the required performance indicators were presented. In these indicators are included the distribution of patients by age, by profession, by location, by the presence in the latter medical appointment for evaluation, among others. Furthermore, the information represented is credible because it has undergone a process of validation by the health professionals.

Conteúdo

Acrónimos	xviii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	2
1.3 Estrutura	4
1.4 Contextualização	5
1.4.1 As Tecnologias de Informação na Saúde	5
1.4.2 Sistemas de Apoio à Decisão	6
1.4.3 Business Intelligence na Área Médica	7
2 Caso de Estudo	9
2.1 Centro Materno Infantil do Norte	9
2.2 Interoperabilidade do CHP	10
2.3 Sistema de Apoio à Prática de Enfermagem	10
2.4 Interrupção Voluntária da Gravidez	11
2.4.1 Consulta Prévia	12
2.4.2 Período de Reflexão	12
2.4.3 Interrupção da Gravidez	12
2.4.4 Consulta de Revisão	13
3 Abordagem Metodológica	15
3.1 Metodologia de Investigação: Design Science Research	15
3.1.1 Atividade 1: Identificação do problema e motivação (Identify problem & motivate)	17

3.1.2	Atividade 2: Definição dos objetivos para a solução (Define objectives of a solution)	18
3.1.3	Atividade 3: Projeção e desenvolvimento (Design & development)	18
3.1.4	Atividade 4: Demonstração (Demonstration)	18
3.1.5	Atividade 5: Avaliação (Evaluation)	19
3.1.6	Atividade 6: Comunicação (Communication)	19
3.2	Metodologias aplicadas ao Sistema de BI	19
4	Descoberta do Conhecimento em BD	23
4.1	Introdução	23
4.1.1	Descoberta do Conhecimento em Bases de Dados	24
4.1.2	Cross Industry Standard Process para Data Mining (CRISP-DM)	25
4.1.3	Data Mining	27
4.1.4	Medidas Estatísticas	29
4.2	Estado da Arte	30
4.3	Gestão das Utentes da IVG através do DM	32
4.3.1	Objetivos	32
4.3.2	Resultados	33
4.3.3	Discussão dos Resultados	41
4.4	Previsão do Grupo de Risco na IVG através do DM	42
4.4.1	Objetivos	42
4.4.2	Resultados	43
4.4.3	Discussão dos Resultados	50
4.5	Conclusão	51
5	Aplicação do Business Intelligence na Saúde	53
5.1	Introdução	53
5.2	Estado da Arte	55
5.3	Objectivos	56
5.4	Aplicações de BI <i>open source</i> na Saúde	56
5.5	Requisitos das Aplicações em Ambientes Clínicos	59

<i>CONTEÚDO</i>	xi
5.6 Ferramentas de BI	63
5.6.1 QlikView	63
5.6.2 Palo BI Suite	65
5.6.3 Jaspersoft BI	67
5.6.4 Tableau Public	68
5.6.5 Spago BI	70
5.6.6 Pentaho BI Suite	71
5.6.7 Resultados - Análise comparativa entre as ferramentas	73
5.6.8 Discussão	77
5.6.9 Conclusão	79
6 Business Intelligence	81
6.1 Introdução	81
6.1.1 Data Warehouse	82
6.1.2 Metodologia de Kimball	83
6.2 Estado da Arte	87
6.3 Objetivos	89
6.4 Resultados	89
6.4.1 Planeamento do Projeto	90
6.4.2 Definição dos Requisitos da Organização	91
6.4.3 Seleção do Produto e Instalação	91
6.4.4 Modelação Dimensional	92
6.4.5 Projeção Física	93
6.4.6 Projeção e Desenvolvimento do ETL	93
6.4.7 Projeção da Aplicação de BI	95
6.4.8 Desenvolvimento da Aplicação BI	96
6.4.9 Plataforma de BI	103
6.5 Discussão dos Resultados	104
6.6 Conclusão	105
7 Conclusão	107
7.1 Considerações Finais	107
7.2 Trabalho Futuro	108

Bibliografia	117
Apêndices	117
A Publicações	119
A.1 Managing Voluntary Interruption of Pregnancy using Data Mining	119
A.2 Real-Time Business Intelligence Platform to Maternity Care	120
A.3 Predicting Risk Pregnancy in Maternity Care using Data Mining	121
A.4 Business Intelligence in Maternity Care	122
A.5 Pre-Triage Decision Support Improvement in Maternity Care by means of Data Mining	123
A.6 Improving Quality of Services in Maternity Care Triage System	124
A.7 Simulating a Multi-Level Priority Triage System for Maternity Emergency	125
A.8 Application of Business Intelligence Tools in Healthcare Environment	126
A.9 Step Towards to improve the Voluntary Interruption of Pregnancy Process by means of Business Intelligence	127
B Glossário	129

Lista de Figuras

3.1	Modelo do processo da metodologia Design Science Research. . .	17
4.1	Representação do processo do CRISP-DM.	26
4.2	Distribuição dos valores da variável alvo <i>Local de Administração da Segunda Dose</i> , que pode tomar os valores Domicilio(0) ou CMIN(1)	35
4.3	Distribuição de valores da variável alvo <i>Consulta de Revisão</i> , que pode apresentar os valores presente (1) ou não presente (0).	45
5.1	Fontes de dados e plataformas compatíveis com o QlikView. . .	65
5.2	Acesso aos dados multidimensionais, recorrendo ao Excel Add-in ou à interface <i>Web</i>	66
5.3	Arquitetura da ferramenta Jaspersoft BI.	69
5.4	Funcionalidades da ferramenta Spago BI.	71
5.5	Arquitetura da ferramenta Pentaho BI Suite.	73
6.1	Diagrama da metodologia de Kimball.	84
6.2	Modelo multidimensional do DW.	92
6.3	Arquitetura da plataforma de BI.	95
6.4	Caracterização do grupo de utentes por antecedentes obstétricos e por data.	96
6.5	Caracterização do grupo de utentes por número de gestações e por data.	97
6.6	Caracterização do grupo de utentes por número de experiências de IVG anteriores e por data.	98

6.7	Caracterização do grupo de utentes relativamente ao número de falhas do processo de IVG, por ano e por mês.	99
6.8	Caracterização do grupo de utentes com base na contracepção no início do processo por data.	99
6.9	Caracterização do grupo de utentes com base na idade gestacional no momento da toma do <i>mifepristone</i> , por ano, mês e dia.	100
6.10	Medidas estatísticas dos tempos médios entre a consulta inicial e a consulta final do processo de IVG, por ano.	101
6.11	Caracterização do grupo de utentes com base na localidade, por ano.	102
6.12	Caracterização do grupo de utentes por profissão e por data.	102
6.13	Caracterização do grupo de utentes por idades e por ano.	103

Lista de Tabelas

4.1	Expressões que definem as medidas da sensibilidade, especificidade e acuidade.	30
4.2	Medidas estatísticas das variáveis N_IVG, SG, Idade, Gesta e Para.	36
4.3	Percentagem de ocorrências das variáveis selecionadas.	36
4.4	% de ocorrências associada a cada um dos intervalos das classes CI_1 e CI_2.	38
4.5	% de ocorrências associada a cada intervalo da classe CSG.	38
4.6	Representação das variáveis utilizadas em cada um dos cenários.	39
4.7	Top 3 dos modelos para cada um dos algoritmos aplicados.	40
4.8	Número de casos que o modelo acertou e falhou na previsão, para cada um dos algoritmos.	42
4.9	Medidas estatísticas das variáveis N_IVG, SG, Idade, Gesta e Para.	45
4.10	Percentagem das ocorrências de algumas variáveis selecionadas.	46
4.11	Representação das variáveis utilizadas em cada um dos cenários.	47
4.12	Valores da sensibilidade, especificidade e acuidade para os três melhores modelos de cada um dos algoritmos aplicados, na primeira abordagem.	49
4.13	Valores da sensibilidade, especificidade e acuidade para os três melhores modelos de cada um dos algoritmos aplicados aos dados resultantes do <i>oversampling</i>	49
4.14	Os melhores modelos de DM obtidos e as respectivas técnicas utilizadas.	50

5.1	Comparação das ferramentas de BI em relação aos requisitos selecionados e atribuição de um grupo a cada uma das características.	74
5.2	Agrupamento das características em grupos com as respectivas percentagens.	76
5.3	Classificação final atribuída a cada um das ferramentas de BI.	78

Acrónimos

AD Árvores de Decisão. 28, 40, 41, 48, 52

AIDA Agência de Integração, Difusão e Armazenamento. 10, 34, 43, 64, 92, 95, 97

BI Business Intelligence. v–viii, 1–8, 13, 19–21, 41, 50, 55–58, 60–65, 67, 68, 73, 75, 77, 78, 80–83, 85–93, 96, 98, 105, 107, 108, 112

CHP Centro Hospitalar do Porto. v, vii, 9–11, 64, 81, 82

CIPE Classificação Internacional das Práticas de Enfermagem. 10

CMIN Centro Materno Infantil do Norte. v, vii, 1–4, 9–11, 13, 20, 22, 32–35, 37, 39–41, 44, 50, 52, 53, 57, 77, 85, 87, 88, 90–97, 102–104, 107–109, 112

CRISP-DM CRoss-Industry Standard Process for DM. 21, 22, 25, 26, 33, 43, 50

DCBD Descoberta de Conhecimento em Base de Dados. 7, 8, 22–27, 30–32

DM Data Mining. v, vii, 2–4, 6, 8, 11, 13, 20–35, 37–44, 46, 48–50, 52, 53, 57, 72–74, 81, 83, 107, 108, 111, 112

DSR Design Science Research. 15, 16

DW Data Warehouse. 6, 21, 63, 70, 83–88, 93–97, 108

EC Extração do Conhecimento. 21

ETL Extract, Transform and Load. 57, 62, 63, 66, 70, 72–74, 78, 84, 87, 93, 95, 97

FN Falso Negativo. 29, 30

FP Falsos Positivos. 29, 30

- HJU** Hospital Joaquim Urbano. 9
- HSA** Hospital de Santo António. 9
- IA** Inteligência Artificial. 31
- IGO** Interrupção da Gravidez Opcional. 11
- IT** Information Technologies. vii
- IVG** Interrupção Voluntária da Gravidez. v, vi, 1–4, 8, 9, 11, 12, 20, 22, 32–35, 41–45, 51, 52, 57, 61, 77, 85, 90–93, 99, 100, 102–104, 106–108, 111, 112
- MLG** Modelos Lineares Generalizados. 29, 39, 40, 48
- NB** Naive Bayes. 29, 40
- OLAP** Online Analytical Processing. 6, 57, 59, 61–63, 65, 67, 68, 70, 72–74, 78, 83, 89, 93, 96, 97, 102, 106
- OMS** Organização Mundial da Saúde. 11
- QV** QlikView. 65, 66
- SAD** Sistema de Apoio à Decisão. v, 4–6
- SAPE** Sistema de Apoio à Prática de Enfermagem. 6, 10, 11, 34, 43, 95, 96
- SGBD** Sistemas de Gestão de Bases de Dados. 65
- SI** Sistemas de Informação. 5, 7, 10, 15, 16
- SNS** Sistema Nacional de Saúde. 11
- SONHO** Sistema Integrado de Informação Hospitalar. 10, 95, 97
- SVM** Support Vector Machine. 29, 40, 42, 48, 52
- TI** Tecnologias de Informação. v, 1, 4, 5, 7, 15, 20, 56, 77, 78, 82, 93, 98
- TIC** Tecnologias da Informação e da Comunicação. 5
- VIP** Voluntary Interruption of Pregnancy. vii
- VN** Verdadeiro Negativo. 29, 30
- VP** Verdadeiro Positivo. 29, 30

Capítulo 1

Introdução

Neste capítulo é apresentada a principal motivação deste projeto, sendo também definidos os principais objetivos a realizar ao longo da dissertação. Numa segunda secção é apresentada a estrutura deste documento e, ainda, uma breve contextualização ao tema.

1.1 Motivação

Com a contínua evolução das [Tecnologias de Informação \(TI\)](#), com o desenvolvimento das tecnologias de [Business Intelligence \(BI\)](#) e a com a sua implementação bem sucedida nas mais variadas áreas, a ideia de expandir a aplicação dessas tecnologias à área da saúde tornou-se num grande objetivo. Assim, a grande motivação deste projeto foi a criação e a implementação de uma plataforma de [BI](#) na unidade de [Interrupção Voluntária da Gravidez \(IVG\)](#) do [Centro Materno Infantil do Norte \(CMIN\)](#) com o intuito de fornecer indicadores de desempenho desta unidade e de integrar modelos de previsão de procedimentos realizados na mesma. Esta motivação advém também do fato de se tratar de uma ideia inovadora, pelo que ainda não existe nenhum projeto semelhante implementado no setor, e ainda por dar usabilidade à grande quantidade de dados, existentes nas bases de dados, que contêm informações importantes capazes de originar conhecimento útil no momento da tomada de decisão.

Além disso, a realização deste projeto permitiu simplificar o trabalho que a equipa de enfermagem da unidade de **IVG** executa na conceção do relatório anual com os respetivos indicadores, que é disponibilizado à Direção Geral da Saúde. Portanto, outra das principais motivações deste trabalho resultou da utilidade e do interesse que esta plataforma tem para os profissionais de saúde.

1.2 Objetivos

O principal objetivo deste projeto é a conceção e implementação de uma plataforma de **BI** na unidade de **IVG** do **CMIN**, que disponibilize indicadores de desempenho desta unidade e que inclua modelos de previsão de algumas atividades realizadas na mesma. Deste modo, a questão de investigação que se coloca neste projeto é a seguinte: "Será possível a implementação de uma plataforma de **BI** na unidade de **IVG** capaz de suportar o processo de tomada de decisão e de prever eventos futuros?". De forma a dar uma resposta a esta questão, foram definidos alguns objetivos mais globais, cuja concretização origina uma resposta positiva à mesma. Assim, de uma forma geral, podem ser identificados três objetivos, como:

- obtenção de modelos de **Data Mining (DM)** capazes de prever a ocorrência de procedimentos relacionados com a **IVG**;
- identificação da ferramenta mais apropriada para a conceção da plataforma de **BI**;
- construção da plataforma de **BI** para unidade de **IVG** do **CMIN**.

Dentro de cada um destes objetivos podem ser definidos outros mais específicos e mais detalhados. Deste modo, relativamente aos modelos de **DM** a desenvolver, pretendeu-se :

- investigar possíveis modelos de **DM** na unidade de **IVG**;
- avaliar se era possível prever a necessidade de acompanhamento das utentes pela equipa de enfermagem no processo de **IVG**, recorrendo a métricas de avaliação e utilizando a técnica de classificação;

- avaliar se era possível prever quais as utentes que têm propensão para pertencer ao grupo de risco, recorrendo a métricas de avaliação e utilizando a técnica de classificação;
- analisar os resultados para cada um dos problemas de DM;
- registar as principais conclusões.

Quanto à identificação da ferramenta de BI a utilizar, os principais objetivos foram:

- seleção de um conjunto de ferramentas *open source* mais utilizadas e mais eficientes;
- instalação e compreensão do modo de funcionamento dessas ferramentas;
- análise e comparação das vantagens e das desvantagens de cada uma dessas ferramentas;
- seleção da ferramenta mais apropriada e mais vantajosa no contexto deste projeto.

Por último, no que diz respeito à criação da plataforma de BI a implementar no CMIN, pretendeu-se:

- compreender o modo de funcionamento da unidade de IVG, no CMIN;
- pesquisar quais os indicadores que seriam interessantes analisar;
- recolher as amostras de dados necessárias que possibilitassem a construção da plataforma de BI;
- proceder ao tratamento e à transformação desses dados e também à construção da estrutura responsável pelo armazenamento dos dados;
- desenvolver a plataforma de BI, recorrendo à ferramenta selecionada;
- implementar a plataforma de BI no CMIN.

1.3 Estrutura

A presente dissertação está dividida em sete capítulos, os quais abordam cada uma das etapas constituintes deste projeto. A revisão de literatura e o estado de arte desta dissertação encontram-se divididos entre os capítulos quatro, cinco e seis, uma vez que são abordados mais especificamente em cada um dos temas presentes nestes capítulos.

No presente capítulo são apresentadas as motivações e os principais objetivos que estão na base deste projeto, bem como a estrutura que este documento apresenta. Além disso, é realizada uma abordagem em termos de contextualização, onde são introduzidos temas como as **TI** na saúde, os **Sistema de Apoio à Decisão (SAD)** e o **BI** na área da saúde.

No segundo capítulo é introduzido o caso de estudo sobre o qual se incidiu nesta dissertação, onde é apresentada a instituição de saúde que possibilitou a realização deste projeto e são descritas algumas das particularidades que a caracterizam, como é o caso dos procedimentos realizados na unidade de **IVG**.

O terceiro capítulo incide sobre as metodologias aplicadas na realização deste projeto, quer seja em termos de metodologia de investigação, como em metodologias utilizadas no **DM** ou no **BI**.

No quarto capítulo é feita uma introdução teórica à temática do **DM**, seguindo-se a apresentação das questões envolvidas nos problemas de **DM**. Por conseguinte, são apresentados os resultados e a respetiva discussão dos mesmos para cada um dos problemas de **DM** descritos, finalizando-se este capítulo com as principais conclusões obtidas.

O quinto capítulo descreve algumas ferramentas *open source*, comparando-as entre si e analisando quais as características apropriadas e indispensáveis que uma aplicação *Web* deve ter para ser implementada na área da saúde. Este capítulo é finalizado com a conclusão de qual a ferramenta mais eficaz para se utilizar na conceção da plataforma desenvolvida neste projeto.

No capítulo seis, o tema de **BI** é abordado com mais detalhe, sendo explicitados os procedimentos realizados para a implementação da plataforma de **BI** no **CMIN**. Além disso, são apresentados os resultados finais decorrentes

do desenvolvimento da aplicação *Web*.

No último capítulo são apresentadas as principais conclusões obtidas no decorrer deste projeto, sendo também expostas algumas sugestões a se ter em consideração em projetos futuros.

1.4 Contextualização

Nesta secção é feito um enquadramento ao tema da dissertação, começando por abordar as **TI** na saúde, seguindo-se os **SAD** e terminando com a aplicação do **BI** na área médica.

1.4.1 As Tecnologias de Informação na Saúde

Atualmente, com a Era da Informática, as organizações estão cada vez mais aptas a dar resposta a novos desafios, novos requisitos, bem como a alterações constantes. Esta capacidade de adaptação surgiu com a necessidade de estas organizações se ajustarem a novas realidades, de modo a serem capazes de manter a competitividade no mercado e de fornecerem serviços de qualidade. A versatilidade destas organizações deve-se em parte às **Tecnologias da Informação e da Comunicação (TIC)** que apoiam os **Sistemas de Informação (SI)**, que são um conjunto de procedimentos que quando executados, produzem informação relevante para o processo de tomada de decisão e para a gestão da organização.

No setor da saúde, as **TI** influenciam os procedimentos médicos, na medida em que fornecem informação completa e credível aos profissionais de saúde e apoiam o processo de tomada de decisão, quer ao nível clínico, como administrativo, reduzindo conseqüentemente os erros médicos associados a essas decisões [1].

Com as **TIC**, há um grande potencial para aumentar a qualidade e a eficiência do sistema de cuidados de saúde, permitindo a partilha de informação, a comunicação, a colaboração e a coordenação entre os profissionais de saúde. [2–5]

Um exemplo da implementação de um **SI** na área da saúde é o caso focado

no auxílio aos enfermeiros, como o **Sistema de Apoio à Prática de Enfermagem (SAPE)**. O **SAPE** permitiu a redefinição das atividades de prestação de cuidados de enfermagem aos utentes e foi criado com o intuito de dar visibilidade ao trabalho realizado pelos enfermeiros, uma vez que estes são os utilizadores que mais produzem, processam e fornecem informação clínica. Com o armazenamento dos dados relacionados com os processos dos utentes no **SAPE**, torna-se possível utilizar essa informação para disponibilizar conhecimento útil nas práticas de enfermagem [2].

1.4.2 Sistemas de Apoio à Decisão

Os **SAD** são sistemas que funcionam como um suporte aos utilizadores no processo de tomada de decisão [6].

Os principais objetivos dos **SAD** são a melhoria da eficácia das decisões e a eficiência do processo de tomada de decisão [7].

Quanto à relação existente entre os **SAD** e o **BI**, o que se pode constatar é que o **BI** é utilizado como um suporte ao desenvolvimento dos **SAD**. Enquanto que os **SAD** integram um conjunto de ferramentas que possibilitam a consulta e a obtenção de relatórios capazes de extrair informação de suporte à tomada de decisão, os sistemas de **BI** integram um conjunto de ferramentas que permitem a recolha, o armazenamento e a análise de dados e, ainda a extração de conhecimento útil a partir dos mesmos [8].

Assim, pode-se afirmar que os sistemas de **BI** são o resultado de uma série de evoluções tecnológicas ocorridas nas últimas décadas. Alguns dos seus antecedentes são os primeiros sistemas de apoio à decisão, as linguagens de programação, os sistemas de informação executivos, o **Data Warehouse (DW)**, a tecnologia **Online Analytical Processing (OLAP)** e o **DM**. Esta evolução tecnológica permitiu que atualmente existam ferramentas que integram todos estes componentes num único sistema [9].

Além disso, esta evolução deu origem a novas formas de armazenamento de grandes quantidades de dados, novas ferramentas de extração e de limpeza dos dados, novas técnicas de análises de dados, bem como novas tecnologias relacionadas com a *Web*, que no global promoveram a existência de

novos ambientes de desenvolvimento, potenciando também o aparecimento das plataformas de BI [9].

1.4.3 Business Intelligence na Área Médica

O conceito de BI tem sido alvo de interesse por parte dos profissionais de saúde e dos profissionais de TI, dada a sua aplicabilidade no setor da saúde com a implementação do *Processo Clínico Eletrónico* [1].

O BI é um processo que compreende várias metodologias, aplicações e tecnologias para recolha, armazenamento, transformação e análise de dados, permitindo o acesso a informação que pode ser útil para apoiar os utilizadores a tomar decisões melhores e mais rápidas. Assim, o BI tem a capacidade de operacionalizar o conteúdo de repositórios do *Processo Clínico Eletrónico*, dando o suporte às práticas baseadas em evidências e melhorando a qualidade dos cuidados de saúde prestados [1, 5, 10].

Business Intelligence e o Data Mining

Nas organizações de saúde, a maioria dos dados clínicos que documentam toda a atividade diária médica são armazenados em *Sistemas de Gestão de Base de Dados Relacionais*. Devido à extensa quantidade de informação, esta é armazenada de diferentes formas, sendo, por isso, bastante heterogênea. Por outro lado, no processo de tomada de decisão, onde é necessário integrar todos ou quase todos os sistemas clínicos, médicos, financeiros e administrativos, as fontes de dados são de grandes dimensões e complexas, pelo que em casos deste calibre, torna-se extremamente importante aceder a dados de qualidade, que interferem diretamente no sucesso do processo de *Descoberta de Conhecimento em Base de Dados (DCBD)* [10, 11].

Deste modo, pode-se afirmar que a interoperabilidade dos SI numa instituição é um dos fatores chave para o processo de tomada de decisão, pois é esta que garante a uniformização dos sistemas e que permite ultrapassar toda a complexidade inerente às diferentes fontes de dados, garantido assim a qualidade dos dados.

Por outro lado, com o aumento da quantidade de dados registados, existe

também uma correspondente necessidade de aplicação de tecnologias de **DM** para a extração de conhecimento útil a partir da informação armazenada em bases dados, a fim de fornecer um suporte no processo de tomada de decisão em tempo real.

A maioria dos dados clínicos não são estruturados e as técnicas de **DM** são mais adequadas para dados estruturados. Assim, uma das vantagens da utilização do **BI** como tecnologia de apoio à decisão é permitir a combinação e transformação de dados não estruturados em dados estruturados, possibilitando assim a aplicação das técnicas de **DM**. Deste modo, o **BI** pode ser combinado com o **DM** e com a **DCBD** [1].

No próximo capítulo é feita uma breve introdução à instituição de saúde onde se pretende implementar a plataforma de **BI**, sendo abordados alguns dos sistemas informáticos implementados nesta e descritos quais os procedimentos adotados em específico no processo de **IVG**.

Capítulo 2

Caso de Estudo

Neste capítulo é feita uma breve apresentação da instituição de saúde que suportou este projeto, do sistema que garante a interoperabilidade da mesma, do sistema utilizado para os registos das práticas de enfermagem e, ainda, dos procedimentos realizados no processo para o qual se direccionou este trabalho, a [Interrupção Voluntária da Gravidez \(IVG\)](#).

2.1 Centro Materno Infantil do Norte

O [Centro Materno Infantil do Norte \(CMIN\)](#) é um dos três hospitais constituintes do [Centro Hospitalar do Porto \(CHP\)](#), sendo os restantes o [Hospital de Santo António \(HSA\)](#) e o [Hospital Joaquim Urbano \(HJU\)](#). A sua junção só se realizou em 2007, tendo estas sido instituições individuais anteriormente.

Com a criação do [CHP](#), as mulheres que necessitam de cuidados urgentes nas especialidades de Ginecologia e Obstetrícia são direccionadas para o [CMIN](#), o que não se verificava anteriormente.

Um dos módulos do [CMIN](#) é o da [IVG](#), para onde são direccionadas todas as mulheres que pretendem interromper a sua gravidez. O processo de interrupção da gravidez pode ser realizado através de medicamentos ou pode ser um procedimento cirúrgico. O recurso a medicamentos para pôr fim ao período de gestação origina a remoção do embrião/feto e da placenta do útero

da utente, procedimento este que foi adotado no CMIN [12].

2.2 Interoperabilidade do CHP

A comunicação entre os **Sistemas de Informação (SI)** dos diferentes hospitais do **CHP** é garantida por um sistema que se baseia no uso de agentes inteligentes denominado de **Agência de Integração, Difusão e Armazenamento (AIDA)** da informação médica. Este sistema de multiagentes permite a uniformização dos sistemas clínicos e permite ultrapassar a complexidade administrativa e médica das diferentes fontes de informação dos hospitais.

Além disso, a **AIDA** permite o envio e a receção de informação hospitalar, a gestão, o armazenamento, e a resposta a pedidos de informação por terceiros.

Deste modo, todos os sistemas de informação do **CHP** estão conectados através da **AIDA**, incluindo o **Sistema de Apoio à Prática de Enfermagem (SAPE)** e o **Sistema Integrado de Informação Hospitalar (SONHO)** [13, 14].

2.3 Sistema de Apoio à Prática de Enfermagem

O **SAPE** teve origem nos **Sistemas de Informação de Enfermagem** como uma alternativa ao registo de informação em papel e a sua arquitetura foi criada na Escola Superior do S. João pelo enfermeiro Abel Paiva. Este sistema foi desenvolvido tendo por base a **Classificação Internacional das Práticas de Enfermagem (CIPE)**, tendo sido construídos um número de diagnósticos e intervenções base a partir de um conjunto de eixos da **CIPE**, com os focos mais frequentes em cada um dos serviços.

O trabalho de parametrização do sistema é desenvolvido pelos serviços do hospital que utilizam a aplicação. Deste modo, para a sua implementação é necessário realizar uma pesquisa dos focos mais frequentes nesse mesmo serviço para, posteriormente, ser construído, para cada um desses focos, intervenções que possam ser utilizadas apropriadamente. Essas intervenções parametrizadas são apresentadas à enfermeira, no momento do registo da

informação da utente, e esta fica encarregada de selecionar das possíveis opções, a que considerar mais apropriada para o caso em questão. Além disso, a cada foco deve estar associado um conjunto de diagnósticos que retratem a evolução desse mesmo foco na utente [2, 15].

O SAPE é atualmente utilizado no CHP e conseqüentemente no CMIN para apoiar as práticas de enfermagem. No CMIN, o processo de IVG é um foco da enfermagem, pelo que os registos clínicos das utentes, bem como todos os procedimentos efetuados são suportados pelo SAPE, tornando possível o uso dessa informação para a extração de conhecimento no contexto da IVG [2]. Assim, o SAPE foi a principal fonte de informação para o desenvolvimento deste projeto, quer em termos dos modelos de Data Mining (DM) desenvolvidos, quer a nível dos indicadores da IVG criados.

2.4 Interrupção Voluntária da Gravidez

A IVG ou *Interrupção da Gravidez Opcional (IGO)* corresponde à interrupção da gravidez por desejo da mulher, sem a existência de problemas de saúde na mulher ou no embrião/feto que comprometam o normal decurso da gestação. A IVG pode dever-se também ao fato de a gravidez ter sido originada de um crime contra a liberdade e autodeterminação sexual. No entanto, a maioria dos casos dizem respeito a gravidezes não desejadas e não planeadas [16, 17].

Até 1984, a realização da IVG não era permitida em Portugal, independentemente das circunstâncias, o que originava soluções alternativas, como as práticas ilegais ou o recurso a outros países [18]. No entanto, desde essa altura que a legislação respeitante à IVG tem sofrido alterações, promovendo o direito à proteção da saúde da mulher e contribuindo para uma opção informada e livre [19]. Assim, a partir de 2007, a IVG passou a ser legal em Portugal, desde que realizada até às 10 semanas de gestação, sendo totalmente comparticipada pelo Sistema Nacional de Saúde (SNS) [20].

Um dos módulos do CMIN é o da IVG, onde são direcionadas todas as mulheres que pretendem interromper a gravidez. A metodologia adotada no CMIN para a realização do processo de IVG é a recomendada pela Orga-

nização Mundial da Saúde (OMS), que consiste na administração de dois medicamentos específicos, concretamente, o *Mifepristone* e o *Misoprostol*. O processo engloba várias etapas, nomeadamente, uma consulta prévia, um período de reflexão, o procedimento da interrupção da gravidez e, por fim, uma consulta de revisão [21].

2.4.1 Consulta Prévia

Esta consulta, realizada previamente ao procedimento de IVG, é uma consulta de carácter obrigatório com um médico, onde é identificada a idade gestacional da utente, são recolhidos dados clínicos, são fornecidas informações sobre os riscos que o procedimento acarreta na saúde da mulher e ainda são aconselhados alguns métodos contraceptivos a utilizar após o procedimento [19,20].

Tendo em consideração que os riscos associados ao processo de interrupção da gravidez são tanto menores quanto menor for a idade gestacional, o período entre a marcação e a concretização desta consulta não deve ser superior a 5 dias [22].

2.4.2 Período de Reflexão

Após a realização da consulta prévia, é obrigatório um período de reflexão de no mínimo três dias. Durante este período, a utente deverá procurar um aconselhamento psicológico e/ou social (recorrendo a um psicólogo ou a um assistente social) [20].

2.4.3 Interrupção da Gravidez

As circunstâncias que tornam a interrupção da gravidez um procedimento legal devem ser atestadas no *certificado de comprovação do tempo de gestação*, redigido e assinado antes da intervenção por um médico diferente do que dirige o procedimento [23]. Da mesma forma, o tempo de gestação deve ser comprovado ecograficamente ou através de outro meio adequado, no momento da interrupção [19,20].

Também a mulher deverá assinar um documento, após o período de reflexão, onde presta um *consentimento livre e esclarecido* [24].

Posteriormente, após a realização de todas as formalidades, é efetuada uma consulta com uma enfermeira, onde é dada a primeira dose de medicação e é realizada uma triagem para verificar se a utente está apta a fazer uma autoadministração da segunda dose da medicação em casa ou se necessita de monitorização da equipa de enfermagem no CMIN. Caso se verifique esta segunda situação, então numa outra fase a utente deve dirigir-se ao CMIN para a administração da medicação em ambulatório, como menciona o protocolo [21].

2.4.4 Consulta de Revisão

Por fim, após a interrupção da gravidez, deverá ser agendada uma consulta médica a realizar no prazo máximo de 15 dias, para confirmar a eliminação completa do conteúdo uterino e estabelecer um método contraceptivo a iniciar [25, 26].

Nesta consulta, na avaliação da utente podem-se verificar três situações distintas: o procedimento ter sido realizado com sucesso e, portanto, a interrupção foi conseguida; o procedimento ter sido incompleto, ou seja, terem sido observados restos ovulares na ecografia; ou por último, a interrupção não ter sido conseguida, sendo necessária a repetição de todo o processo [21].

No momento de optar por um método contraceptivo, de forma a que este seja eficaz, os profissionais devem tentar compreender o que falhou e transmiti-lo à utente, para evitar que essas falhas se repitam. Assim, nestes casos, devem-se considerar os métodos que sejam o mínimo dependentes das falhas humanas, como é o caso de esquecimentos [27].

O seguinte capítulo faz uma abordagem sobre as metodologias utilizadas no desenvolvimento deste projeto, nomeadamente, a metodologia de investigação, a metodologia aplicada na obtenção dos modelos de DM e a metodologia seguida na criação da plataforma de BI.

Capítulo 3

Abordagem Metodológica

3.1 Metodologia de Investigação: Design Science Research

Uma metodologia é "um sistema de princípios, práticas e procedimentos aplicados a uma determinada área de conhecimento". Esta tem como função apoiar os investigadores de [Sistemas de Informação \(SI\)](#) a produzirem e apresentarem uma investigação de grande qualidade, útil e rigorosa [28].

Para a metodologia [Design Science Research \(DSR\)](#), é necessário ter em consideração três elementos: princípios conceptuais que definam o que se pretende com a [DSR](#), regras práticas e um processo para a realização e apresentação da investigação [28].

Sendo uma metodologia relativamente recente, a metodologia [DSR](#) aplicada aos [SI](#) continua a evoluir, no entanto, já é possível definir qual a sua essência. "A [DSR](#) cria e avalia artefactos das [Tecnologias de Informação \(TI\)](#) com o intuito de solucionar problemas organizacionais identificados" [29]. Consiste num processo muito rigoroso que projeta artefactos para resolver problemas detetados, para realizar contribuições em pesquisas, para avaliar projetos e para comunicar os resultados a um público adequado. Tais artefactos incluem construções, modelos, métodos e instâncias, podendo ainda englobar inovações sociais ou novas propriedades de recursos técnicos, sociais e/ou informativos. Resumindo, esta definição abrange qualquer objeto

projetado com uma solução integrada para um problema de investigação compreendido [30].

As regras práticas para a realização de pesquisas em **SI** seguindo a metodologia **DSR**, com base em Hevner et al., estão representadas na forma de sete diretrizes, que descrevem as características que uma pesquisa deve ter para que seja bem realizada [29]. A mais importante destas diretrizes é que a investigação deve produzir "um artefacto criado para resolver um problema" [29]. Além disso, o artefacto deve ser relevante para a solução de um "problema até ao momento não solucionado e importante para o negócio". A sua "utilidade, qualidade e eficácia" devem ser avaliadas com rigor [29]. A investigação deverá representar uma contribuição fiável, pelo que o rigor deverá estar presente quer no desenvolvimento do artefacto, como na sua avaliação. O desenvolvimento do artefacto deverá incluir um processo de pesquisa que tem por base teorias e conhecimentos existentes, para a obtenção de uma solução de um problema definido. Por fim, a investigação deverá ser comunicada de forma eficaz ao público alvo [29].

O terceiro elemento nesta abordagem metodológica é o processo a utilizar para a **DSR**. Este processo deverá ter como finalidade três objetivos: fornecer um processo nominal para a execução da **DSR**, basear-se na literatura direcionada para a **DSR** em **SI** e oferecer aos investigadores um modelo mental ou um modelo para a estruturação dos resultados finais da pesquisa. Relativamente ao processo nominal, este é responsável por fornecer um trajeto a seguir pelos investigadores que pretendem utilizar a metodologia **DSR** na sua pesquisa [28]. Quanto à literatura, o processo deve ser suportado por uma pesquisa, quer na área de **SI**, quer na própria metodologia **DSR**, de modo a que todas as opções sejam facilmente justificáveis [28]. Por último, o modelo mental tem como finalidade conduzir com eficácia o investigador na realização e na apresentação da investigação. Na maioria dos casos, os resultados obtidos da pesquisa **DSR** são diferentes daqueles que eram esperados teoricamente e, portanto, o modelo mental vem nestes casos dar alguma orientação, quer ao investigador como ao utilizador, do que esperar em termos de resultados finais na investigação **DSR** [29, 31].

Depois de definidos os principais princípios desta metodologia, é impor-

3.1. METODOLOGIA DE INVESTIGAÇÃO: DESIGN SCIENCE RESEARCH¹⁷

tante definir as etapas que a constituem e que apresentam uma estrutura comumente aceite para a realização de pesquisas. Essas etapas estão representadas na Figura 3.1.

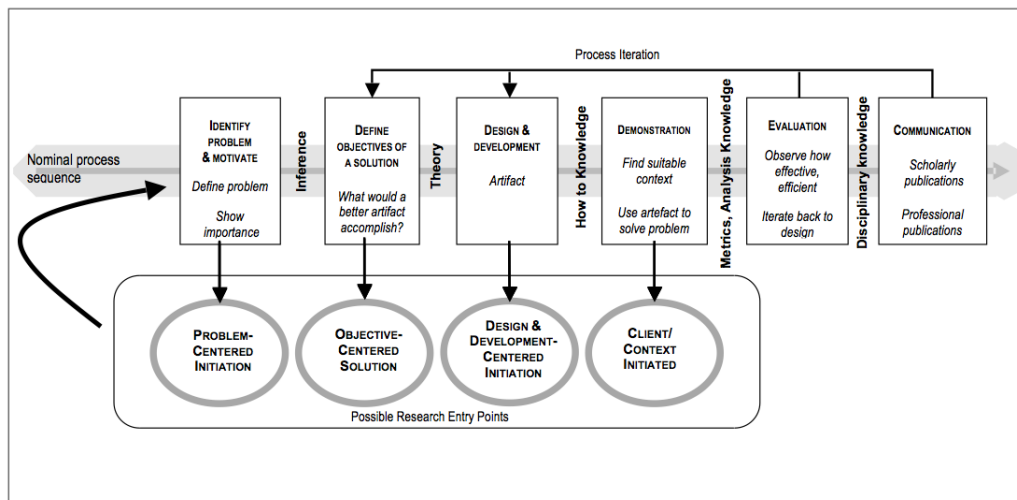


Figura 3.1: Modelo do processo da metodologia Design Science Research [28].

3.1.1 Atividade 1: Identificação do problema e motivação (Identify problem & motivate)

Nesta etapa, deve ser definido o problema específico em questão e deve ser justificado o valor da solução. Uma vez que a definição do problema é utilizada para desenvolver um artefacto que consiga de forma eficaz fornecer uma solução, é útil simplificar o problema conceptual, de tal modo que se possa reduzir a complexidade da solução. Com a justificação do valor da solução consegue-se motivar o investigador e o público alvo a optar por essa solução e a aceitar os seus resultados, e ainda, a facilitar a compreensão do raciocínio do investigador associado ao problema. Os recursos necessários para esta etapa incluem o conhecimento sobre o estado do problema e a importância da sua solução [29, 32–35].

3.1.2 Atividade 2: Definição dos objetivos para a solução (Define objectives of a solution)

Esta atividade compreende a inferência dos objetivos de uma solução a partir da definição do problema e a tomada de conhecimento do que é possível e fiável desenvolver. Os objetivos podem ser quantitativos, como é o caso em que a solução desejada será melhor do que as soluções existentes até ao momento; ou qualitativos, como por exemplo a descrição de como um novo artefacto poderá suportar soluções para problemas que ainda não foram abordados. Os objetivos devem ser definidos de forma racional, a partir da especificação do problema. Os recursos necessários nesta etapa incluem o conhecimento do estado do problema, das soluções já existentes e da sua eficácia [35, 36].

3.1.3 Atividade 3: Projeção e desenvolvimento (Design & development)

A terceira atividade envolve a criação dos artefactos. Estes podem ser principalmente construtores, modelos, métodos ou instâncias ou, ainda, "novas propriedades de recursos técnicos, sociais e/ou informativos". Em termos conceptuais, um artefacto pode ser qualquer objeto projetado, no qual está integrado uma contribuição para a investigação. Esta atividade inclui a determinação da funcionalidade desejada para o artefacto e a sua arquitetura, seguidos da criação do mesmo. Os recursos necessários nesta etapa são o conhecimento da teoria que pode ser aplicada na solução [29, 32, 36].

3.1.4 Atividade 4: Demonstração (Demonstration)

Nesta etapa, é demonstrada a utilização do artefacto para resolver uma ou mais instâncias do problema. Isto pode envolver a sua utilização na experiência, na simulação, no caso de estudo, na prova ou ainda noutra atividade apropriada. Os recursos necessários para a demonstração são o conhecimento efetivo de como utilizar o artefacto para a resolução do problema [32, 35, 36]

3.1.5 Atividade 5: Avaliação (Evaluation)

Ocorre a observação e a medição do quão eficiente é o artefacto a apoiar a solução do problema. Esta atividade envolve a comparação dos objetivos da solução com os resultados obtidos na utilização do artefacto na demonstração. Para tal, é necessário conhecimento de métricas relevantes e técnicas de análise. Dependendo da natureza do problema e do artefacto, a avaliação pode ser realizada de várias formas. Pode incluir a comparação da funcionalidade do artefacto com os objetivos da solução; a utilização de medidas quantitativas de desempenho objetivas, como orçamentos e itens produzidos; a realização de questionários de satisfação, análise do feedback dos utilizadores ou simulações. Além disso, pode englobar ainda medidas quantificáveis do desempenho do sistema, como o tempo de resposta e a disponibilidade. No final desta atividade, os investigadores podem decidir se é necessário voltar à terceira atividade para tentar melhorar a eficácia do artefacto ou se continuam na atividade da comunicação e deixam as questões de melhoria para os projetos subsequentes [29, 32, 33, 36].

3.1.6 Atividade 6: Comunicação (Communication)

Esta etapa inclui a comunicação do problema e da sua importância, do artefacto, da sua utilidade e da sua inovação, do rigor da sua conceção e da sua eficácia para os investigadores e para outros públicos relevantes. A comunicação exige um conhecimento da cultura disciplinar [28, 29].

3.2 Metodologias aplicadas ao caso de desenvolvimento da Plataforma de BI

Para o desenvolvimento deste projeto foi utilizada a metodologia abordada na secção 3.1. Para tal, foram seguidas todas as atividades que constituem esta metodologia. Assim, para cada uma das atividades mencionadas, foram descritas as etapas executadas no caso prático do desenvolvimento e implementação da plataforma de [Business Intelligence \(BI\)](#) na unidade de

Interrupção Voluntária da Gravidez (IVG) no Centro Materno Infantil do Norte (CMIN).

Deste modo, relativamente à identificação do problema e à motivação, pode-se afirmar que a ideia da criação de uma plataforma de BI para a unidade de IVG adveio de uma sugestão feita por profissionais de saúde, mais especificamente por uma equipa de enfermagem, que demonstrou um grande interesse na implementação de uma aplicação que disponibilizasse indicadores de desempenho específicos para a unidade de IVG. Esta solução tem como principal objetivo facilitar um trabalho adicional da equipa de enfermagem, que consiste na realização de relatórios submetidos anualmente para a Direção Geral de Saúde. Este trabalho até ao momento era realizado manualmente, com o recurso a folhas de cálculo. Com este projeto, pretendia-se automatizar todo o processo de criação de indicadores de desempenho, recorrendo aos dados armazenados, resultantes do *Processo Clínico Eletrónico*. Além disso, pretendia-se também com esta plataforma de BI a integração de modelos de *Data Mining* (DM), que permitam efetuar previsões de alguns dos procedimentos que são realizados no CMIN, na unidade de IVG.

Quanto à segunda atividade, verificou-se então que o artefacto a ser criado era a plataforma de BI que disponibiliza os indicadores de desempenho da unidade de IVG e que integra os modelos de DM. Até ao momento, verificou-se que não existe nenhuma plataforma de BI idêntica ao pretendido no mercado. Assim, pode-se assumir que este artefacto é uma inovação na área de IVG.

Na terceira atividade, a projeção da plataforma a desenvolver foi realizada com o apoio dos profissionais de saúde e dos profissionais de *Tecnologias de Informação* (TI). Os profissionais de saúde deram o seu contributo na definição dos indicadores a implementar e na interpretação e compreensão da informação armazenada nas bases de dados, enquanto que os profissionais de TI foram úteis na estruturação, no desenvolvimento técnico e na implementação da plataforma de BI. Ao longo do desenvolvimento deste projeto, foram realizadas várias reuniões com ambas as partes, de modo a assegurar que o resultado final fosse o pretendido e o planeado na fase inicial.

Na demonstração do artefacto, foi utilizada uma aplicação interativa com

os indicadores associados a uma amostra de dados relativa a um curto período de tempo. Esta aplicação teve como principal objetivo demonstrar aos profissionais de saúde o modo de funcionamento da plataforma de BI e o modo como a informação estaria representada na mesma.

Na avaliação, foi necessário recorrer aos profissionais que apresentavam maior interesse na plataforma, mais especificamente a equipa de enfermagem, para que fosse possível validar todo o conhecimento apresentado nesta e verificar se os indicadores de desempenho apresentados eram os pretendidos. Com o apoio destes profissionais tornou-se possível a criação de uma plataforma que, segundo eles, além de ter utilidade, tem ainda credibilidade. Depois desta etapa, concluiu-se que a plataforma de BI estava apta a ser utilizada, apesar de terem sido feitas sugestões para trabalho futuro, como por exemplo novos indicadores, que possivelmente serão realizados em projetos subsequentes.

Por último, a comunicação do desenvolvimento deste artefacto foi realizada através da redação de uma dissertação, de uma apresentação ao público alvo e ainda através da escrita e publicação de artigos científicos.

Relativamente à metodologia aplicada no desenvolvimento dos modelos de DM, recorreu-se ao [CRoss-Industry Standard Process for DM \(CRISP-DM\)](#). Esta metodologia é a mais utilizada no DM e foi desenvolvida em finais de 1996, por um grupo de empresas (Teradata, SPSS-ISL, Daimler-Chrysler e OHRA) que se dedicam a projetos de DM e que propuseram um guia de referência para o desenvolvimento de trabalhos no domínio do DM e da [Extração do Conhecimento \(EC\)](#). O [CRISP-DM](#) é independente de qualquer fornecedor, podendo ser utilizado com qualquer ferramenta de DM e ser aplicado a qualquer problema de DM. Esta metodologia além de definir as fases a serem implementadas num projeto de DM, define também as tarefas que devem ser realizadas em cada uma dessas fases. Assim, no geral, o [CRISP-DM](#) é constituído por seis fases que serão abordadas na secção 4.1.2 [37].

A metodologia utilizada para a criação da plataforma de BI foi a metodologia de Ralph Kimball. Esta foi concebida em meados da década de 1980 e tem sido implementada com sucesso em milhares de projetos de [Data](#)

Warehouse (DW) e de BI, em praticamente todos os setores, áreas de aplicação, negócios e plataformas técnicas. Esta metodologia será abordada mais detalhadamente na secção 6.1 [38].

O capítulo que se segue é um capítulo que incide na *Descoberta de Conhecimento em Base de Dados (DCBD)* e no *DM*, sendo descritos todos os procedimentos efetuados, através da metodologia *CRISP-DM*, para a obtenção dos modelos de *DM* associados à unidade de *IVG* no *CMIN*.

Capítulo 4

Descoberta do Conhecimento em Bases de Dados

No presente capítulo é feita uma introdução ao tema, que faz referência a conceitos pertinentes para a compreensão do capítulo e à metodologia utilizada durante o desenvolvimento do mesmo, seguida de uma breve revisão da literatura. Posteriormente, são apresentados os resultados da aplicação dessa metodologia aos casos de estudos, dando ênfase às técnicas de [Data Mining \(DM\)](#) aplicadas, juntamente com a discussão dos mesmos e as principais conclusões retiradas.

4.1 Introdução

O processo de [Descoberta de Conhecimento em Base de Dados \(DCBD\)](#) engloba todos os procedimentos necessários para obtenção de conhecimento útil a partir dos dados, enquanto que o [DM](#) apenas incide na aplicação de algoritmos nos dados para a extração de modelos. Até 1995, o [DCBD](#) e o [DM](#) eram considerados equivalentes. Atualmente, o [DM](#) representa uma das etapas do [DCBD](#) e consiste na procura de padrões ou relações que possam existir entre os dados armazenados nas bases de dados.

4.1.1 Descoberta do Conhecimento em Bases de Dados

O processo de **DCBD** consiste num conjunto de atividades que permitem partilhar o conhecimento obtido através da informação armazenada nas bases de dados. O principal objetivo do processo de **DCBD** é a descoberta de conhecimento útil, válido e relevante sobre uma atividade em particular, através de algoritmos, tendo em consideração dados de magnitude crescente [39]. De acordo com Fayyad et al. [40], este processo engloba cinco etapas, nomeadamente:

- *Seleção*, onde ocorre a seleção do conjunto de dados necessários para a realização do **DM**;
- *Pré-processamento*, que inclui a limpeza e o processamento dos dados, de modo a torná-los consistentes;
- *Transformação*, onde os dados são modificados, de acordo com a variável alvo selecionada;
- *Data Mining*, onde são definidos os objetivos que se pretendem atingir e o tipo de resultado que se pretende obter. De acordo com o resultado desejado, é definido o tipo de metodologia a seguir e identificado o tipo de técnica a utilizar. Posteriormente, é aplicada a técnica de **DM** selecionada ao conjunto de dados, de modo a originar a obtenção de padrões;
- *Interpretação/Avaliação*, que consiste na interpretação e avaliação dos padrões obtidos. A validade dos resultados obtidos pode ser avaliada aplicando esses padrões a novos conjuntos de dados [41].

De acordo com Freitas [42], o conhecimento a ser extraído deve apresentar três propriedades: deve ser correto, na medida do possível; deve ser facilmente compreendido pelos utilizadores e deve ser conhecimento novo, útil e interessante. Além disso, a metodologia de extração de conhecimento deve apresentar três características: deve ser eficaz e precisa; genérica, ou seja, deve ser aplicável a diferentes tipos de dados, e flexível, isto é, facilmente adaptável.

4.1.2 CRoss Industry Standard Process para Data Mining (CRISP-DM)

Existem várias metodologias que são utilizadas no processo de **DM**, porém no âmbito deste projeto optou-se pela metodologia mais conhecida, a de **CRoss-Industry Standard Process for DM (CRISP-DM)**, devido às características que apresenta:

- Independente da Indústria: o mesmo processo pode ser aplicado para a análise do negócio, dados financeiros, recursos humanos, produção, serviços, entre outros;
- É independente da ferramenta utilizada;
- Tem uma relação próxima com os modelos obtidos em procedimentos de **DCBD**, anteriormente descritos.

O **CRISP-DM** divide o processo de **DM** em seis etapas principais, que podem ser observadas na Figura 4.1. Estas etapas são:

- *Compreensão do Negócio (Business Understanding)*: esta fase inicial incide na compreensão do objetivo do projeto de uma perspectiva do negócio, definindo um plano preliminar para o alcance desse mesmo objetivo;
- *Compreensão dos Dados (Data Understanding)*: inclui a recolha dos dados e atividades de iniciação para a familiarização com os dados, identificando problemas e conjuntos de dados interessantes;
- *Preparação dos Dados (Data Preparation)*: esta fase inclui todas as questões relacionadas com a criação de cenários necessários para a construção da tabela de cenários. As atividades de preparação dos dados devem ser realizadas várias vezes e devem incluir a construção da tabela de cenários, a seleção dos atributos, a limpeza e a transformação dos dados. Além disso, nesta fase, é possível a criação de novos atributos obtidos a partir de atributos existentes. Esta etapa é muito

importante para a melhoria significativa da qualidade do conhecimento extraído através do DM;

- *Modelação (Modelling)*: nesta fase são aplicadas várias técnicas de modelação (algoritmos de árvores de decisão, regras de associação, regressão linear, redes neuronais, entre outros) aos dados. É possível que durante esta etapa seja realizada a calibração dos parâmetros para otimização, pelo que é comum o retorno à fase de preparação dos dados.
- *Avaliação (Evaluation)*: corresponde à seleção do modelo que parece ter maior qualidade de uma perspetiva de análise dos dados. É necessário durante esta fase verificar se o modelo está de acordo com os objetivos do negócio;
- *Implementação (Deployment)*: O conhecimento obtido através do modelo deve ser organizado e apresentado de forma a que o utilizador possa utilizá-lo.

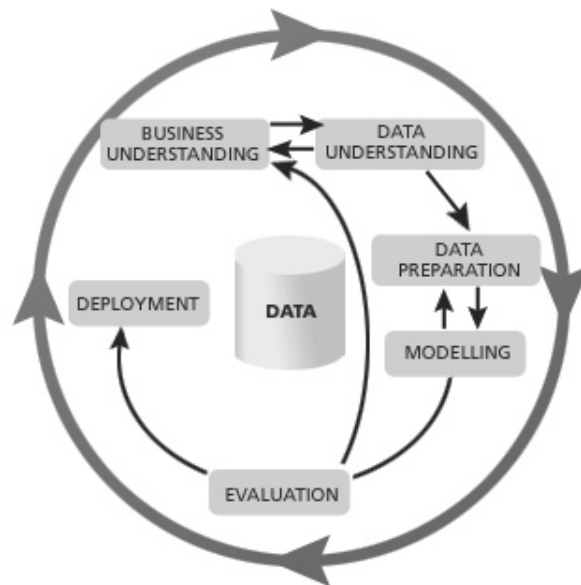


Figura 4.1: Representação do processo do CRISP-DM [37].

Como foi referido, comparando as etapas do processo de DCBD com as etapas do CRISP-DM, verifica-se que esta última metodologia incorpora

passos que devem preceder e acompanhar o processo de DCBD. Assim, a compreensão do domínio da aplicação, a aquisição de conhecimento prévio relevante e a definição dos objetivos do utilizador final são etapas que deverão preceder o processo de DCBD, enquanto que a consolidação do conhecimento deverá ser uma fase que acompanhe o processo de DCBD [43–45].

4.1.3 Data Mining

Como já foi mencionado, o DM é uma das etapas constituintes do processo de DCBD, cuja principal finalidade é a deteção de padrões em dados pré-processados.

Existem duas possíveis abordagens no DM que podem ser utilizadas para a obtenção de informação relevante [46]:

- **Abordagem direta ou modelo de verificação**, que é seguida quando o utilizador pretende testar a validade de uma hipótese formulada por si, através dos dados, de modo comprová-la ou a negá-la.
- **Abordagem indireta ou modelo de descoberta**, onde o ênfase está no sistema, que descobre automaticamente a informação relevante nos dados.

Na abordagem indireta, podem ainda ser identificados dois objetivos distintos [46]:

- **Descrição**: este objetivo pretende aumentar o conhecimento sobre os dados analisados, permitindo a identificação de regras que os caracterizam.
- **Previsão**: implica o desenvolvimento de modelos que sejam capazes de prever o valor de uma variável alvo, recorrendo para isso a um conjunto de atributos selecionados das bases de dados.

Associado a cada um destes objetivos, existe um conjunto de diferentes tarefas, às quais estão inerentes diferentes algoritmos de aplicação aos dados. As principais tarefas de DM são [47–50] :

- **Classificação:** processo que envolve a criação de um modelo ou função capaz de distinguir classes de dados, de forma a que, quando aplicado a novos dados, seja possível a previsão da classe a que esses novos dados pertencem. Dentro desta tarefa, existe uma grande variedade de técnicas que podem ser utilizadas, tais como árvores de decisão, análise de regressão, redes neuronais artificiais, algoritmo Naïve Bayes (NB), entre outros.
- **Associação:** tem como finalidade a descoberta de dependências significativas entre variáveis. Esta tarefa permite identificar grupos de dados ou fatos que possam ser diretamente ou indiretamente associados, sendo analisada a correlação que existe entre eles. Essas associações apresentam um determinado nível de suporte e de confiança que permitem a avaliação das relações encontradas.
- **Regressão:** consiste na descoberta de uma função capaz de representar, de forma aproximada, a maneira como os dados estão distribuídos. A principal finalidade desta tarefa é então a procura de uma função que represente comportamentos de variáveis de uma forma aproximada. A função de regressão mais conhecida é a regressão linear.
- **Clustering:** processo de partição de um conjunto de dados em grupos, denominados de *clusters*, de acordo com o grau de similaridade que apresentam entre si.

Destas tarefas apresentadas, pode-se afirmar que a classificação e a regressão apresentam um caráter de previsão, enquanto que a associação e o *clustering* estão para a descrição.

Neste projeto, a tarefa utilizada foi a de classificação, pelo fato de se pretender, neste caso, prever qual a classe a que um utente teria mais probabilidade de pertencer, com base numa classificação previamente feita do conjunto de dados. Nesta tarefa, foram utilizadas as técnicas que melhor se aplicavam aos problemas de DM abordados [37]:

- *Árvores de Decisão (AD)*: esta técnica gera automaticamente regras,

que são condições que revelam a lógica utilizada na construção da árvore.

- *NB*: recorre ao teorema de Bayes, uma fórmula que calcula a probabilidade através da contagem da frequência de determinados valores e de combinações de valores no histórico dos dados.
- *Modelos Lineares Generalizados (MLG)*: é uma técnica estatística bastante popular para modelação linear. Em Oracle DM, esta técnica é aplicada para classificação binária.
- *Support Vector Machine (SVM)*: é um poderoso algoritmo baseado em regressão linear e não-linear. No Oracle DM, o SVM é implementado para classificação binária e multiclasse.

4.1.4 Medidas Estatísticas

Com a utilização das técnicas de DM descritas na secção anterior, é possível obter quatro tipos de resultados, representados na *Matriz de Confusão*. O primeiro resultado designa-se de *Verdadeiro Positivo (VP)* e corresponde ao número de registos positivos classificados corretamente. Tem-se ainda os *Falsos Positivos (FP)* que estão relacionados com o número de registos negativos que foram classificados como sendo positivos. Além disso, há também os *Verdadeiro Negativo (VN)*, que correspondem ao número de registos classificados corretamente como sendo negativos, e por último, os *Falso Negativo (FN)*, que correspondem ao número de exemplos positivos classificados como sendo negativos.

A partir destes valores resultantes dos modelos, podem ser estimadas métricas estatísticas para a avaliação da qualidade dos dados, cujas expressões estão representadas na Tabela 4.1, tais como [37]:

- *Sensibilidade*: é a capacidade de detetar corretamente a ocorrência de casos positivos. Consiste no resultado da razão entre os valores VP e todos os valores correspondentes a positivos (VP + FN);

- *Especificidade*: por outro lado, corresponde à capacidade de identificar corretamente num modelo a não ocorrência de casos positivos. É medida através da razão entre os valores corretamente identificados como negativos (VN) e todos os valores correspondentes a negativos ($VN + FP$);
- *Acuidade*: corresponde à percentagem total de concordância entre as previsões corretamente efetuadas e todas as previsões feitas. Mede-se através da proporção de todos os resultados medidos corretamente ($VP + VN$) a partir dos modelos em todos os casos possíveis de serem obtidos ($VP + VN + FP + FN$).

Tabela 4.1: Expressões que definem as medidas da sensibilidade, especificidade e acuidade.

	Resultados Positivos	Resultados Negativos	Sensibilidade	Especificidade	Acuidade
Valores Positivos	VP	FP	$\frac{VP}{VP+FN}$	$\frac{VN}{VN+FP}$	$\frac{VP+VN}{VP+FP+FN+VN}$
Valores Negativos	FN	VN			

4.2 Estado da Arte

Historicamente, já foram associados uma grande variedade de nomes à necessidade de tentar encontrar padrões vantajosos nos dados, nomeadamente, o de *DM*, *Extração do Conhecimento (EC)*, descoberta de informação, aquisição de informação, arqueologia de dados e, ainda, processamento de padrões dos dados. O termo *DM* era essencialmente utilizado por estatísticos, analistas de dados e pelas comunidades de Sistemas de Gestão da Informação. Além disso, foi um termo que ganhou muita popularidade no setor das bases de dados.

Relativamente ao termo *DCBD*, este foi criado no primeiro workshop de *DCBD* em 1989 para enfatizar que o conhecimento é o produto final de uma descoberta orientada para os dados [51]. Este termo foi popularizado essencialmente nas áreas de *Inteligência Artificial (IA)* e de aprendizagem automática.

Atualmente, a **DCBD** corresponde ao processo global de descoberta de conhecimento útil através dos dados e o **DM** diz respeito a uma etapa em particular desse processo.

Com o grande crescimento das bases de dados, implementadas em grande parte das organizações, e com o acesso generalizado a grandes repositórios de dados, surgiu a necessidade de desenvolver técnicas que permitissem consolidar e extrair o conhecimento de toda a informação armazenada. A área de pesquisa das técnicas de **DM** surgiu em 1960 como um dos ramos da **IA** [48].

Assim, o **DCBD** é uma área interdisciplinar que tem evoluído bastante nos últimos anos através da interseção de técnicas de gestão de bases de dados, visualização de dados, reconhecimento de padrões, aprendizagem automática, algoritmos matemáticos, sistemas inteligentes e estatística para a extração de conhecimento útil, subentendido em grandes volumes de dados [52].

A criação da **DCBD** e do **DM** originou o desenvolvimento de aplicações em diversas áreas. A aplicação bem sucedida de técnicas de **DM** pode ser observada em diferentes setores, como é o caso das áreas de bioinformática, da ecologia e sustentabilidade, das finanças, da indústria, do marketing, da investigação científica, entre outras. Também na área da saúde, atualmente, tem sido desenvolvidas diversas aplicações direcionadas para a previsão e identificação de tendências [53]. Uma das principais motivações da implementação das técnicas de **DM** na área da saúde é ajudar, por exemplo, os médicos a identificar os tratamentos mais eficazes e as melhores práticas, fornecendo aos utentes melhores serviços de saúde e mais acessíveis. Alguns modelos de **DM** já implementados no setor da saúde são:

- Modelo de previsão para a gravidade da exacerbação da asma em utentes pediátricos no departamento de emergência [54];
- Sistema inteligente baseado na *Web* para a previsão de doenças cardíacas, que recorre a perfis de médicos para prever a propensão dos utentes para as doenças de coração [41];
- Modelo que prevê a recorrência do cancro da mama [55];
- Ferramenta que prevê precocemente o nascimento prematuro [56];

- Técnicas de processamento dos dados para avaliação do risco de suicídio [57];
- Entre outras.

No Centro Materno Infantil do Norte (CMIN) e, em particular, na unidade de Interrupção Voluntária da Gravidez (IVG), o DM nunca tinha sido utilizado, pelo que os modelos descritos neste capítulo podem ser tidos como uma inovação nesta área.

Nas próximas secções são abordados dois casos em particular da implementação do processo de DCBD aplicados à unidade de IVG do CMIN. A primeira questão diz respeito à possibilidade de prever se uma utente da IVG necessita de acompanhamento da equipa de enfermagem para a administração de uma das doses de medicamentos, enquanto que a segunda está relacionada com a possibilidade de prever se uma utente tem aptidão para pertencer ao grupo de risco.

4.3 Gestão das Utentes da Interrupção Voluntária da Gravidez através do Data Mining

A presente secção corresponde ao primeiro problema de DM abordado neste projeto, pelo que se segue uma descrição geral dos objetivos deste problema com uma posterior apresentação dos resultados obtidos [58].

4.3.1 Objetivos

Quando uma mulher tem a intenção de não dar continuidade a uma gravidez não planeada, esta tem que se dirigir a uma unidade de saúde especializada, como é o caso do CMIN. Os procedimentos utilizados na unidade de IVG dividem-se em duas etapas, onde são administrados dois tipos de medicamentos diferentes. A primeira dosagem de medicamentos é administrada com o auxílio da equipa de enfermagem, enquanto que a segunda dosagem poderá ser administrada pela utente no domicílio ou ainda requerer novamente o acompanhamento da equipa de enfermagem. Esta decisão é tomada

pela enfermeira que está a acompanhar o processo da utente e é tomada com base nas características da mesma.

Neste subcapítulo, o principal objetivo é prever qual o local mais apropriado para a administração da segunda dose de medicamentos, tendo por base algumas características cruciais da utente para essa tomada de decisão.

4.3.2 Resultados

De modo a se conseguir cumprir com os objetivos estabelecidos, foi utilizada a metodologia do **CRISP-DM** descrita anteriormente. Assim, recorrendo à ferramenta Data Miner do Oracle, foram realizados todos os passos que constituem esta metodologia, aplicados ao caso de estudo.

Compreensão do Negócio

Como já foi referido, o processo de **IVG** consiste na administração de duas doses de medicamentos, *mifepristone* e *misoprostol*. A primeira dose é sempre administrada no **CMIN** com a monitorização da equipa de enfermagem, enquanto que a segunda dose, pode ser aplicada de forma equivalente à da primeira ou então em casa ao cuidado da utente.

Neste caso, o problema em questão pode ser formulado como "*É possível prever se uma mulher grávida realizará a segunda etapa da administração de medicamentos em casa ou com o acompanhamento da equipa de enfermagem no **CMIN**?*". Este problema pode ser traduzido num problema de **DM** como "*Qual a probabilidade da administração da segunda dose de medicamentos numa mulher grávida ser realizada com o acompanhamento da equipa de enfermagem?*".

Pretende-se com este problema de **DM** construir um modelo capaz de prever onde a utente deverá realizar a segunda fase de medicamentos, tendo como base todos os dados que fazem parte do historial do processo de **IVG** até ao momento. Antes da construção do modelo, foi necessário selecionar os dados que aparentemente melhor descrevem as relações entre a mulher grávida e o local onde é realizada a segunda fase do processo de **IVG** (no **CMIN** ou no domicílio).

Para a indução dos modelos de DM foram selecionados os atributos idade, gesta (número de gestações), para (número de partos), número de IVGs realizadas anteriormente, profissão, contraceção e semanas de gestação.

Compreensão dos dados

Nesta etapa, foram extraídos dados do Sistema de Apoio à Prática de Enfermagem (SAPE) e da Agência de Integração, Difusão e Armazenamento (AIDA) e foi analisada a qualidade das possíveis variáveis a serem utilizadas no processo. A amostra compreende o período entre 1 de Janeiro de 2012 e 31 de Janeiro de 2012, correspondendo a 1124 registos de casos de IVG. Cada registo contém os seguintes campos de preenchimento:

- *Idade*: corresponde à idade da utente que se submeteu ao procedimento de IVG;
- *Número de IVGs realizadas anteriormente (N_IVG)*: esta variável define o número de vezes que a utente se submeteu a um procedimento de IVG anteriormente;
- *Gesta*: corresponde ao número de gestações anteriores da utente;
- *Para*: corresponde ao número de partos que a utente já teve;
- *Estado profissional (EP)*: esta variável informa se a utente em questão está empregada ou desempregada;
- *Falhas da Contraceção (FC)*: corresponde à falha associada ao método contraceptivo utilizado, responsável pela ocorrência da gravidez;
- *Local de Administração da Segunda Dose (LASD)*: corresponde ao local onde é administrada a segunda dose de medicamentos (Domicílio ou CMIN);
- *Método Contraceptivo (MC)*: esta variável informa se a utente utilizou método contraceptivo ou não;

- *Semanas de Gestação (SG)*: corresponde ao número de semanas de gestação da utente, quando deu entrada no CMIN.

Após uma análise superficial dos dados, verificou-se que estes tinham qualidade. No entanto tiveram que ser trabalhados de modo a poderem ser utilizados na formação do problema de DM, uma vez que nem todos os campos estavam preenchidos no registo das utentes. Na Figura 4.2 está representada a distribuição de valores da variável alvo *Local de Administração da Segunda Dose*. Nesta figura pode-se observar que, por exemplo, cerca de 47% das utentes realizaram a segunda etapa do processo de IVG no CMIN (1).

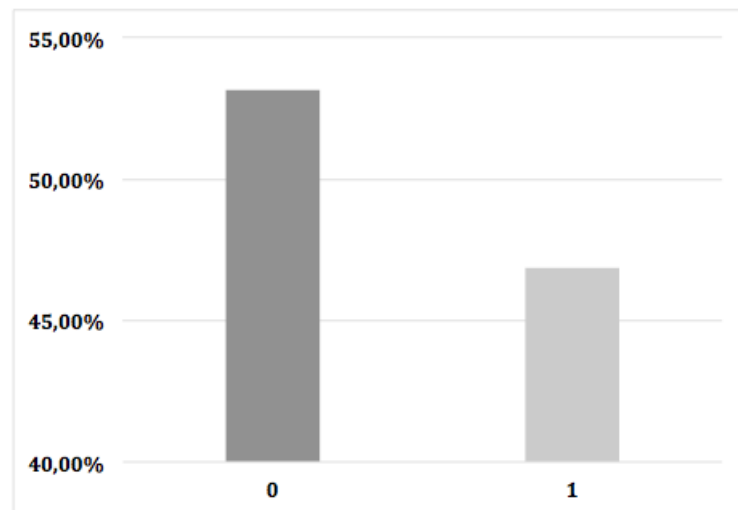


Figura 4.2: Distribuição dos valores da variável alvo *Local de Administração da Segunda Dose*, que pode tomar os valores Domicilio(0) ou CMIN(1) .

Relativamente à Tabela 4.2 estão representadas medidas estatísticas de algumas variáveis numéricas. Assim, para a variável N_{IVG} , correspondente ao número de interrupções da gravidez realizadas anteriormente, verifica-se que esta varia entre 0 e 4, que a média é de 0.22 e que o desvio padrão tem o valor de 0.53.

Tabela 4.2: Medidas estatísticas das variáveis N_IVG , SG , $Idade$, $Gesta$ e $Para$.

	Mínimo	Máximo	Média	Desvio Padrão
N_IVG	0	4	0.22	0.53
SG	3	10.4	7.14	1.46
$Idade$	13	46	27.4	6.95
$Gesta$	1	9	2.1	1.26
$Para$	0	8	0.8	0.98

A Tabela 4.3 apresenta a percentagem de ocorrências associada a cada uma das variáveis selecionadas. Por exemplo, a variável MC , correspondente à utilização de métodos contraceptivos, tem cerca de 39% de ocorrências para a opção não (0) e aproximadamente 61% para a opção sim (1).

Tabela 4.3: Percentagem de ocorrências das variáveis selecionadas.

Número	MC	EP	N_IVG	$Gesta$	$Para$
0	38.60	53.00	82.10	42.60	49.70
1	61.40	26.20	14.70	24.80	27.40
2		20.80	2.50	19.60	17.60
3		0.09	0.37	8.95	4.40
4				2.22	0.65
5				1.02	0.09
6				0.46	0.09
8				0.28	0.09
9				0.09	

Preparação dos dados

Nesta fase, tendo por base as variáveis mencionadas na subsecção anterior, foram selecionadas apenas as variáveis consideradas mais importantes na modelação do problema de DM , nomeadamente, a *idade*, o *gesta*, o *para*, o *número de IVGs anteriores*, o *estado profissional* e o *local de adminis-*

tração da segunda dose de medicamentos. Após a seleção dos dados, estes foram submetidos a uma fase de pré-processamento, onde todos os dados com valores nulos ou que apresentavam ruído (valores fora dos intervalos de normalidade aceites pelo universo em estudo) foram eliminados. Assim, foram eliminados aproximadamente 0.08% de nulos na variável *estado profissional*, cerca de 0.08% de nulos na variável *métodos contraceptivos*, 0.08% de nulos na variável *local de administração da segunda dose de medicamentos* e cerca de 0.7% de valores nulos na variável *número de IVGs anteriores*. Além disso, foram ainda eliminados cerca de 3% de dados com ruído, restando apenas 1082 registos no final.

Um exemplo da existência de ruído nos dados é o caso das semanas de gestação, onde os valores com casas decimais eram separados umas vezes por um ponto e outras por uma vírgula.

Além disso, foram ainda definidas classes para algumas variáveis de modo a permitir uma modelação ao problema com maior precisão, tendo sido analisada a distribuição de valores para cada classe. Foram então criadas as seguintes classes, presentes nas Tabelas 4.4 e 4.5:

- *Classe de Idade 1 (CI_1)*: esta divisão baseia-se em classes definidas em relatórios realizados pela equipa de enfermagem do CMIN, o que originou uma divisão dos dados nos mesmos intervalos.
- *Classe de Idade 2 (CI_2)*: esta segunda divisão das idades foi feita de forma a avaliar o impacto que a idade pode ter nos resultados.
- *Classe de Semanas de Gestação (CSG)*: esta divisão foi realizada de modo a agrupar as mulheres grávidas que recorrem ao CMIN, de acordo com o tempo de gestação, com o intuito de analisar o seu impacto nos resultados finais.

Na Tabela 4.4, pode-se observar que utentes com idades compreendidas entre os 13 e os 15 anos têm uma percentagem de cerca de 1% de ocorrências, enquanto a classe que engloba as utentes com idades entre os 0 e 18 anos apresenta uma percentagem de cerca de 6% de ocorrências.

Tabela 4.4: % de ocorrências associada a cada um dos intervalos das classes CI_1 e CI_2.

<i>CI_1</i>	% Ocorrências	<i>CI_2</i>	% Ocorrências
<i>13-15</i>	1.20	<i>0-18</i>	5.50
<i>16-18</i>	7.80	<i>19-23</i>	29.50
<i>19-21</i>	13.50	<i>24-29</i>	28.50
<i>22-24</i>	17.50	<i>30-35</i>	20.50
<i>25-27</i>	14.20	<i>36-41</i>	14.00
<i>28-30</i>	14.00	<i>42-46</i>	2.00
<i>31-35</i>	15.80		
<i>36-40</i>	12.90		
<i>41-46</i>	3.10		

Tabela 4.5: % de ocorrências associada a cada intervalo da classe CSG.

<i>CSG</i>	% Ocorrências
<i>0-4</i>	2.20
<i>5-8</i>	67.70
<i>9-12</i>	30.10

Com todas as transformações previamente referidas, tornou-se possível a construção da tabela de cenários, onde as técnicas de DM foram aplicadas.

Modelação

Na Tabela 4.6, pode-se observar a combinação das variáveis selecionadas em cada um dos 10 cenários que originaram os melhores resultados. As únicas variáveis que estão presentes em todos os cenários são *LASD*, que é a variável que se pretende prever, o *Gesta* e o *Para*, que são as variáveis que têm mais influência na variável alvo.

Tabela 4.6: Representação das variáveis utilizadas em cada um dos cenários.

	<i>LASD</i>	<i>Idade</i>	<i>N_IVG</i>	<i>Gesta</i>	<i>Para</i>	<i>EP</i>	<i>MC</i>	<i>SG</i>	<i>CI_1</i>	<i>CI_2</i>	<i>CSG</i>
Cenário 1	X	X	X	X	X	X	-	-	-	-	-
Cenário 2	X	X	-	X	X	-	-	-	-	-	-
Cenário 3	X	-	-	X	X	-	-	-	-	-	-
Cenário 4	X	X	X	X	X	-	-	-	-	-	-
Cenário 5	X	-	X	X	X	X	-	-	X	-	-
Cenário 6	X	-	X	X	X	X	-	-	-	X	-
Cenário 7	X	-	X	X	X	X	X	-	-	X	-
Cenário 8	X	-	X	X	X	X	-	X	-	X	-
Cenário 9	X	-	X	X	X	X	-	-	-	X	X
Cenário 10	X	-	X	X	X	X	X	-	-	X	X

Após a construção da tabela de cenários, foram aplicadas as técnicas de DM de modo a se obter o melhor modelo para o problema de DM em questão. Para a seleção das técnicas de DM teve-se em consideração dois aspetos: a interpretação dos modelos e a eficiência do mecanismo.

Para a utilização da técnica MLG, que realiza uma classificação binária, foi necessário efetuar inicialmente uma transformação da variável alvo *LASD* em binário, uma vez que esta variável era registada com os seguintes valores:

- **CMIN** : se a utente realizava a etapa de administração da segunda dose de medicação no **CMIN**, com a monitorização da equipa de enfermagem;
- **Dom**: Se a utente tinha capacidade para autoadministrar a medicação em casa.

No caso do valor da variável corresponder ao primeiro caso, **CMIN**, foi atribuído o valor "1"; no caso de corresponder a **Dom**, atribuiu-se o valor "0". Os modelos desenvolvidos podem ser representados pela seguinte expressão:

$$M_n = T_f + C_i + TDM_Y + A$$

A expressão significa que o modelo M_n pertence à tarefa (T) do tipo classificação e é caracterizado pelo cenário (C), pela técnica de DM (TDM) e pela abordagem (A):

$$T_f = \{ \textit{Classificação} \}$$

$$C_i = \{ \textit{Cenário 1} \dots \textit{Cenário 10} \}$$

$$TDM_y = \{ \textit{SVM}, \textit{NB}, \textit{MLG}, \textit{AD} \}$$

Avaliação

Esta etapa tem o intuito de avaliar os modelos. Para a avaliação dos resultados obtidos pelos modelos de *DM*, as métricas estatísticas descritas na secção 4.1.4 foram aplicadas. Neste caso em particular foram avaliados três parâmetros: a especificidade, a sensibilidade e a acuidade.

Durante a fase de desenvolvimento, foram criados 40 modelos (10 cenários x 4 técnicas de *DM* x 1 abordagem). A Tabela 4.7 apresenta o top 3 dos cenários, para cada um dos algoritmos aplicados.

Tabela 4.7: Top 3 dos modelos para cada um dos algoritmos aplicados.

Support Vector Machine				Naïve Bayes			
	Sensibilidade	Especificidade	Acuidade		Sensibilidade	Especificidade	Acuidade
Cenário 1	0.902	0.870	0.886	Cenário 2	0.887	0.880	0.884
Cenário 5	0.902	0.870	0.886	Cenário 3	0.900	0.855	0.877
Cenário 7	0.902	0.870	0.886	Cenário 4	0.887	0.880	0.884
Modelação Linear Generalizada				Árvores de Decisão			
	Sensibilidade	Especificidade	Acuidade		Sensibilidade	Especificidade	Acuidade
Cenário 2	0.904	0.815	0.856	Cenário 1	0.911	0.841	0.874
Cenário 3	0.900	0.825	0.860	Cenário 3	0.911	0.841	0.874
Cenário 4	0.889	0.864	0.877	Cenário 4	0.911	0.841	0.874

Os melhores modelos foram selecionados com base no valor da sensibilidade, uma vez que o interesse do *CMIN* é tomar conhecimento de todas as mulheres grávidas que necessitam de acompanhamento, apesar de existirem casos correspondentes a falsos positivos. Nestes casos, é preferível que ocorra a monitorização da utente sem que seja estritamente necessário do que a não ocorrência do acompanhamento de uma utente que até necessitava do apoio da equipa de enfermagem e que, posteriormente, resulte no insucesso do processo de *IVG*.

Implementação

Como já foi referido, após a validação do modelo, o conhecimento adquirido é organizado e apresentado aos profissionais de saúde, que poderão fazer uso dele. Dependendo dos requisitos, a fase de implementação pode ser simples, dando origem a relatórios; ou complexa, originando a implementação de um processo repetitivo de DM.

Neste caso em particular, estes processos de DM foram integrados na plataforma de Business Intelligence (BI) implementada e que se encontra em fase de testes no CMIN.

4.3.3 Discussão dos Resultados

Os resultados obtidos podem ser considerados bastante aceitáveis, tendo em consideração a avaliação realizada nos modelos de previsão. Nos modelos de classificação, as melhores previsões originaram um valor máximo da sensibilidade de aproximadamente 91%. Os três melhores modelos que atingiram este valor de sensibilidade foram os modelos 1, 3 e 4. Curiosamente, os melhores modelos foram obtidos com a aplicação da mesma técnica: AD.

O melhor cenário, resultante do valor máximo de sensibilidade com o melhor valor de acuidade, foi o cenário 3, que resulta da combinação das variáveis *Gesta* e *Para*. Este cenário foi considerado o melhor porque está presente no top 3 dos melhores modelos obtidos em três das técnicas de DM utilizadas, apresentando o maior valor de sensibilidade em duas dessas técnicas. Assim, pode-se concluir que com a combinação de todas as variáveis utilizadas nos diferentes modelos e com a aplicação das diferentes técnicas, o modelo 3 foi o mais eficaz na previsão da variável *LASD*.

A Tabela 4.8 apresenta os resultados em termos de precisão do cenário 3, identificando o número de casos que o modelo acertou e o número de vezes que o modelo falhou em cada um dos algoritmos utilizados.

Tabela 4.8: Número de casos que o modelo acertou e falhou na previsão, para cada um dos algoritmos.

	Errado	Correto
<i>Support Vector Machine</i>	48	382
<i>Naïve Bayes</i>	53	377
<i>Modelação Linear Generalizada</i>	60	370
<i>Árvores de Decisão</i>	54	376

De acordo com estes resultados, verifica-se que o algoritmo com maior acuidade aplicado ao modelo 3 foi o algoritmo **SVM**, uma vez que acertou num maior número de casos, nomeadamente em cerca de 89% dos casos.

Concluindo, pode-se afirmar que os resultados obtidos foram satisfatórios, podendo ser utilizados para apoiar o processo de tomada de decisão da equipa de enfermagem sobre o local da administração da segunda dose de medicamentos, no processo de **IVG**.

4.4 Previsão do Grupo de Risco das Grávidas na IVG recorrendo ao DM

Nesta secção é abordado o segundo problema de **DM** sobre o qual se incidiu durante a realização desta dissertação, seguindo-se a mesma sequência descritiva da secção anterior [59].

4.4.1 Objetivos

No processo de **IVG**, a última etapa consiste na consulta de avaliação, onde se determina se o procedimento foi conseguido, incompleto ou não conseguido, segundo as enfermeiras responsáveis pelo processo de **IVG**. As utentes que falham esta etapa são consideradas utentes de risco, pelo facto de se desconhecer o resultado final do procedimento.

O objetivo desta secção é avaliar o processo de **IVG** e consequentemente identificar se a utente irá pertencer a um grupo de risco, através da construção

4.4. PREVISÃO DO GRUPO DE RISCO NA IVG ATRAVÉS DO DM 43

de modelos de DM.

4.4.2 Resultados

Para se conseguir atingir os objetivos definidos, mais uma vez foi utilizada a metodologia do CRISP-DM descrita na secção 4.1.2. Assim, recorrendo novamente à ferramenta Data Miner do Oracle, foram realizados todos os passos que constituem esta metodologia.

Compreensão do Negócio

O processo de IVG compreende duas fases de administração de medicamentos, que são seguidas por uma examinação do médico. Esta examinação avalia se o processo foi bem sucedido ou não. No entanto, existe um grupo de mulheres que falham esta consulta médica e que, por isso, fazem parte do grupo de risco. Isto deve-se ao facto de não existir informação acerca destas utentes e não ser conhecido o seu estado atual de saúde nem o do feto.

Assim, este problema pode ser formulado da seguinte forma: "É possível prever se uma mulher pertence ao grupo das utentes de risco?". Esta questão pode ser traduzida num problema de DM: "Qual a probabilidade de uma mulher pertencer ao grupo de utentes de risco?".

Um modelo que prevê se uma mulher é considerada uma utente de risco foi construído, baseando-se numa amostra de dados registados no processo de IVG e nas informações clínicas recolhidas pelas enfermeiras responsáveis pela IVG.

Antes da construção desse modelo, foram selecionados os dados associados aos atributos que estabelecem a relação entre a utente e o grupo de mulheres que não se apresentam na consulta de avaliação.

Compreensão dos Dados

Esta fase inclui a extração de dados do SAPE e da AIDA, bem como a análise de possíveis variáveis a incorporar nos modelos de DM. Os dados extraídos compreendem o período de 1 de Janeiro de 2012 a 31 de Dezembro

de 2012. A quantidade de dados engloba um total de 1124 registos, sendo que cada registo requer o preenchimento de campos como:

- *Idade*: corresponde à idade da utente que se submeteu ao procedimento de IVG;
- *Número de IVGs realizadas anteriormente (N_IVG)*: esta variável define o número de vezes que a utente se submeteu a um procedimento de IVG anteriormente;
- *Gesta*: corresponde ao número de gestações anteriores da utente;
- *Para*: corresponde ao número de partos que a utente já teve;
- *Estado profissional (EP)*: esta variável informa se a utente em questão está empregada ou desempregada;
- *IVG Conseguida (IC)*: esta variável informa se o procedimento de IVG foi conseguido ou não;
- *IVG Incompleta (II)*: esta variável informa se o procedimento de IVG foi incompleto ou não;
- *IVG Não Conseguida (INC)*: esta variável informa se o procedimento de IVG falhou ou não;
- *Consulta de Revisão (CR)*: variável que informa se a utente esteve ou não presente na consulta de revisão
- *Método Contracetivo (MC)*: esta variável informa se a utente utilizou método contracetivo (1) ou não (0);
- *Semanas de Gestação (SG)*: corresponde ao número de semanas de gestação da utente, quando deu entrada no CMIN.

Uma análise estatística demonstrou que os dados têm qualidade, mas que, no entanto, precisam de algumas transformações para poderem ser incorporados nos modelos de DM. A Figura 4.3 ilustra a distribuição de valores (em

4.4. PREVISÃO DO GRUPO DE RISCO NA IVG ATRAVÉS DO DM 45

percentagem) da variável alvo *Consulta de Revisão*. Assim, verifica-se que cerca de 10% das utentes que recorrem à IVG não vão à consulta de revisão (0) e, portanto, pertencem ao grupo de risco.

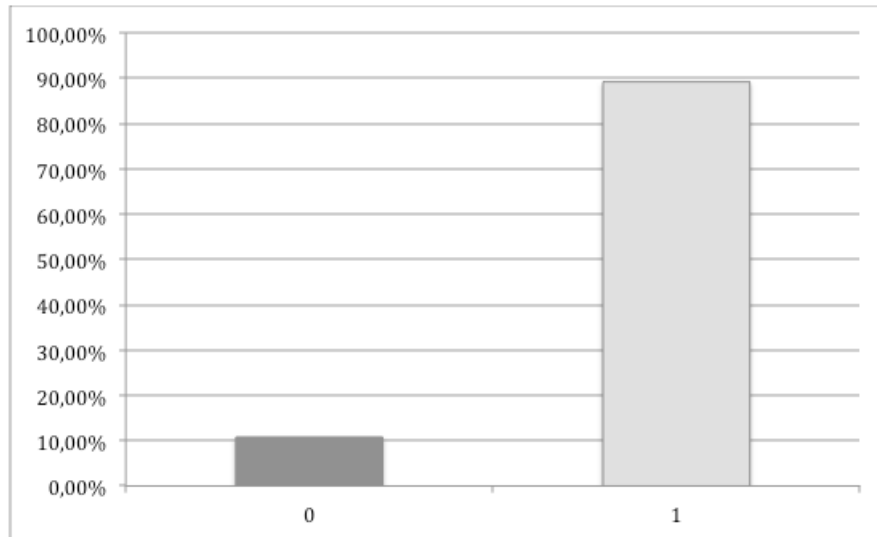


Figura 4.3: Distribuição de valores da variável alvo *Consulta de Revisão*, que pode apresentar os valores presente (1) ou não presente (0).

Na Tabela 4.9, algumas medidas estatísticas aplicadas às variáveis numéricas são apresentadas. Por exemplo, pode-se observar que a variável *idade*, correspondente à idade das utentes que recorrem à IVG, varia entre os 13 e 46 anos, que a média é de aproximadamente 27 anos e o desvio padrão é de 6.95.

Tabela 4.9: Medidas estatísticas das variáveis *N_IVG*, *SG*, *Idade*, *Gesta* e *Para*.

	Mínimo	Máximo	Média	Desvio Padrão
<i>N_IVG</i>	0	4	0.22	0.52
<i>SG</i>	3	10.4	7.16	1.46
<i>Idade</i>	13	46	27.43	6.95
<i>Gesta</i>	1	9	2.14	1.30
<i>Para</i>	0	8	0.82	0.98

Quanto à Tabela 4.10, nesta é demonstrada a percentagem de ocorrências para cada uma das variáveis selecionadas. Analisando esta tabela, tendo em atenção a variável *MC*, associada ao uso de métodos contraceptivos, verifica-se que cerca de 39% não usa contraceção, enquanto que 61% das utentes recorre aos métodos contraceptivos.

Tabela 4.10: Percentagem das ocorrências de algumas variáveis selecionadas.

Número	<i>MC</i>	<i>EP</i>	<i>N_IVG</i>	<i>Para</i>	<i>Gesta</i>
0	38.79	53.11	82.36	48.78	41.58
1	61.21	26.55	14.58	27.90	24.75
2		20.34	2.43	17.82	19.80
3			0.36	4.59	9.18
4			0.27	0.63	2.61
5				0.09	1.08
6				0.09	0.45
8				0.09	0.36
9					0.18

Preparação dos Dados

Nesta fase, foram selecionadas as variáveis mais apropriadas ao problema de *DM*, tendo como base as variáveis definidas na subsecção anterior. Assim, as variáveis utilizadas neste problema foram *Consulta de revisão*, *Gesta*, *Para*, *Estado Profissional*, *Número de IVGs anteriores*, *Idade*, *Métodos Contraceptivos* e *Semanas de Gestação*. Posteriormente, os dados selecionados foram submetidos a uma fase de pré-processamento, onde todos os registos que apresentassem campos sem preenchimento ou com ruído (valores fora do intervalo) foram eliminados. Após esse processamento, foram utilizados apenas 1119 registos dos iniciais.

Alguns dos procedimentos realizados para eliminar o ruído dos dados foram a substituição da vírgula pelo ponto, na separação das casas decimais na variável *Semanas de Gestação*; a eliminação de texto associado a variáveis numéricas; bem como a atribuição de valores numéricos à variável *Estado*

4.4. PREVISÃO DO GRUPO DE RISCO NA IVG ATRAVÉS DO DM 47

Professional, onde o valor 0 foi associado ao desemprego, o valor 1 foi associado às utentes com emprego e, por último, o valor 2 às estudantes.

Depois de efetuadas todas as transformações nos dados, tornou-se possível construir a tabela de cenários.

Numa segunda etapa da preparação dos dados, recorreu-se a uma técnica de *oversampling*, numa tentativa de obtenção de melhores resultados. Esta técnica consiste em manter o conjunto de dados de um dos resultados da variável alvo e proceder à replicação do conjunto de dados do resultado contrário da mesma variável, de forma a se encontrar uma proximidade na distribuição dos resultados dessa variável. Assim, o número de ocorrências da variável alvo com o valor "0" manteve-se, enquanto que o número de ocorrências com o valor "1" foi replicado até ser atingida essa tal proximidade na distribuição. No final desta etapa, o total do número de registos passou a ser 1959, atribuindo uma percentagem de 49% de ocorrências ao valor "0" da variável e de 51 % ao valor "1".

Modelação

Na etapa da modelação, foi construída a tabela de cenários apresentada na Tabela 4.11, onde estão representados 10 cenários, resultantes de diferentes combinações das variáveis. Em cada cenário, a variável alvo *CR* está representada, bem como outras variáveis consideradas cruciais para a construção dos modelos de previsão.

Tabela 4.11: Representação das variáveis utilizadas em cada um dos cenários.

	<i>CR</i>	<i>Idade</i>	<i>N_IVG</i>	<i>Gesta</i>	<i>Para</i>	<i>EP</i>	<i>MC</i>	<i>SG</i>
Cenário 1	X	X	X	X	X	X	X	X
Cenário 2	X	X	X	X	X	X	X	-
Cenário 3	X	X	X	X	X	X	-	-
Cenário 4	X	X	X	X	X	-	-	-
Cenário 5	X	X	X	-	-	-	-	-
Cenário 6	X	X	X	X	X	X	-	X
Cenário 7	X	X	X	X	X	-	-	X
Cenário 8	X	X	X	X	X	-	X	X
Cenário 9	X	X	X	X	X	-	X	-
Cenário 10	X	X	X	-	-	-	X	X

Após a construção da tabela de cenários, os dados foram submetidos a técnicas de DM, técnicas essas selecionadas a fim de ser possível identificar o melhor modelo de previsão para o problema de DM em questão. Neste caso, as técnicas de DM utilizadas foram o de AD, SVM e o de MLG.

Cada modelo, resultante da aplicação de uma técnica em particular num dado cenário de DM, pode ser definido através da expressão:

$$M_n = T_f + C_i + TDM_y + A$$

Nesta expressão, o modelo M_n pertence à tarefa (T) do tipo classificação e é caracterizado pelo cenário (C), pela técnica de DM (TDM) e pela abordagem (A):

$$T_f = \{ \textit{Classificação} \}$$

$$C_i = \{ \textit{Cenário 1} \dots \textit{Cenário 10} \}$$

$$TDM_y = \{ \textit{SVM}, \textit{MLG}, \textit{AD} \}$$

Para este problema de DM, foram gerados 60 modelos de previsão, tendo em consideração os modelos criados com base nos dados resultantes da técnica de *oversampling*. Este número total de modelos resulta de 10 cenários x 3 TDM x 2 abordagens.

Avaliação

Para a avaliação dos resultados obtidos nos modelos de DM, foi considerada uma métrica estatística, denominada de sensibilidade. Esta medida avalia a capacidade do modelo detetar corretamente se a utente pertence ao grupo de risco. É, portanto, a métrica mais apropriada para a avaliação dos modelos, uma vez que o principal objetivo é detetar a ocorrência das utentes com forte probabilidade de pertencerem ao grupo de risco.

Nesta fase, os três melhores modelos de cada técnica de DM utilizada foram selecionados e estão representados na Tabela 4.12.

A partir dos dados resultantes do *oversampling* também foi possível a aplicação das técnicas de DM utilizadas na primeira abordagem. Para estes resultados, foram calculadas métricas estatísticas, incluindo a sensibilidade, a especificidade e a acuidade para os melhores modelos obtidos na abordagem anterior (presentes na Tabela 4.12), como se pode observar na Tabela 4.13.

4.4. PREVISÃO DO GRUPO DE RISCO NA IVG ATRAVÉS DO DM 49

Tabela 4.12: Valores da sensibilidade, especificidade e acuidade para os três melhores modelos de cada um dos algoritmos aplicados, na primeira abordagem.

Support Vector Machine		Modelação Linear Generalizada		Árvores de Decisão	
Sensibilidade		Sensibilidade		Sensibilidade	
Cenário 4	0.929	Cenário 4	0.925	Cenário 4	0.926
Cenário 7	0.924	Cenário 5	0.929	Cenário 5	0.926
Cenário 10	0.926	Cenário 9	0.923	Cenário 9	0.926
Especificidade		Especificidade		Especificidade	
Cenário 4	0.100	Cenário 4	0.093	Cenário 4	0.090
Cenário 7	0.093	Cenário 5	0.092	Cenário 5	0.090
Cenário 10	0.101	Cenário 9	0.092	Cenário 9	0.090
Acuidade		Acuidade		Acuidade	
Cenário 4	0.574	Cenário 4	0.543	Cenário 4	0.446
Cenário 7	0.583	Cenário 5	0.437	Cenário 5	0.446
Cenário 10	0.631	Cenário 9	0.579	Cenário 9	0.446

Tabela 4.13: Valores da sensibilidade, especificidade e acuidade para os três melhores modelos de cada um dos algoritmos aplicados aos dados resultantes do *oversampling*.

Support Vector Machine		Modelação Linear Generalizada		Árvores de Decisão	
Sensibilidade		Sensibilidade		Sensibilidade	
Cenário 4	0.644	Cenário 4	0.594	Cenário 4	0.873
Cenário 7	0.693	Cenário 5	0.645	Cenário 5	0.524
Cenário 10	0.608	Cenário 9	0.595	Cenário 9	0.873
Especificidade		Especificidade		Especificidade	
Cenário 4	0.753	Cenário 4	0.613	Cenário 4	0.524
Cenário 7	0.650	Cenário 5	0.579	Cenário 5	0.524
Cenário 10	0.822	Cenário 9	0.614	Cenário 9	0.524
Acuidade		Acuidade		Acuidade	
Cenário 4	0.680	Cenário 4	0.602	Cenário 4	0.555
Cenário 7	0.670	Cenário 5	0.605	Cenário 5	0.555
Cenário 10	0.657	Cenário 9	0.603	Cenário 9	0.555

Analisando as duas tabelas, é possível verificar que os valores da sensibilidade diminuíram da Tabela 4.12 para a Tabela 4.13. No entanto, na Tabela 4.13 os valores de especificidade e de acuidade são melhores, o que se deve ao facto de nesta segunda abordagem existir um número de casos mais equilibrado na variável alvo.

Implementação

Após a validação do modelo, foi apresentado e disponibilizado o conhecimento extraído aos profissionais de saúde. O processo de DM descrito neste subcapítulo foi também integrado na plataforma de BI implementada e em fase de testes no CMIN.

4.4.3 Discussão dos Resultados

Após a aplicação da metodologia do CRISP-DM, foi possível verificar que os resultados obtidos são satisfatórios, tendo em consideração a avaliação realizada na subsecção 4.4.2. Nos modelos de classificação, as melhores previsões obtidas, baseadas na métrica da sensibilidade, foram de aproximadamente 93%. Na Tabela 4.14 são apresentados os três melhores modelos.

Tabela 4.14: Os melhores modelos de DM obtidos e as respetivas técnicas utilizadas.

Modelos	Técnica de DM	Sensibilidade
Modelo 4	Support Vector Machine	0.929
Modelo 5	Modelação Linear Generalizada	0.929
Modelo 9	Árvores de Decisão	0.926

Como se pode observar na Tabela 4.14, os três melhores modelos, tendo em consideração os valores da sensibilidade, são os modelos 4, 5 e 9. Assim, pode-se verificar que os atributos que melhor caracterizam as possíveis utentes pertencentes ao grupo de risco são a idade, o número de IVGs anteriores, o gesta e o para. Além disso, os resultados obtidos através da segunda fase de preparação dos dados demonstraram que, existindo um equilíbrio no número de ocorrências dos valores "0" e "1" da variável alvo, os modelos são

mais uniformes e a capacidade de detetar corretamente as utentes que pertencem ao grupo de risco é idêntica à capacidade de detetar as utentes que não pertencem ao mesmo grupo.

Concluindo, os resultados obtidos para a previsão das utentes pertencentes ao grupo de utentes de risco, tendo em consideração a presença ou ausência da utente na consulta de avaliação, foram satisfatórios. Consequentemente, os modelos de previsão obtidos podem ser utilizados para apoiar o processo de tomada de decisão da equipa de enfermagem responsável pelo processo de *IVG*. Assim, o profissional de saúde pode ter em consideração *a priori* as características que identificam uma mulher pertencente ao grupo de risco e ficar com a ideia se a utente é uma possível candidata a este grupo, tomando medidas preventivas neste tipo de utentes.

4.5 Conclusão

Com este estudo demonstrou-se que é possível a obtenção de modelos de classificação de *DM* quer para a previsão do local onde a utente da *IVG* deve proceder à administração da segunda dose de medicamentos, como para a previsão de se uma utente pertence ao grupo de risco no processo de *IVG*. Este estudo foi realizado utilizando dados reais de processos de *IVG*, recolhidos no *CMIN*, correspondentes a um ano de atividade, nomeadamente ao ano de 2012.

Neste capítulo foram obtidos bons resultados em ambos os problemas de *DM*. Relativamente ao primeiro problema, obtiveram-se valores de sensibilidade e acuidade de, aproximadamente, 91% e 87%, respetivamente, no modelo 3, resultante da aplicação da técnica *AD* ao cenário 3. Assim, pode-se considerar que os fatores mais importantes para a determinação do local da administração da segunda dose de medicação são o número de gestações (*Gesta*) e o número de partos (*Para*) que a utente teve.

Quanto à segunda questão de *DM*, obteve-se um valor máximo de aproximadamente 93% no modelo 4, em termos da métrica de sensibilidade, utilizando a técnica *SVM*. Deste modo, verifica-se que os fatores mais relevantes na determinação das utentes pertencentes ao grupo de risco são a idade, o

número de IVGs realizadas anteriormente, o número de gestações e o número de partos.

Pode-se concluir então que, recorrendo a técnicas de classificação de dados e ao historial de dados das utentes da unidade de IVG do CMIN, é possível prever qual a localização mais adequada para a realização da segunda etapa do processo de IVG, bem como as utentes com maior aptidão para pertencer ao grupo de risco no processo de IVG.

Este trabalho demonstra que é possível o desenvolvimento de modelos de DM com boa capacidade preditiva, a fim de fornecer à mulher grávida um melhor tratamento nas condições apropriadas. No entanto, estes problemas de DM descritos neste capítulo estão direcionados apenas para a IVG, em específico no CMIN. Os modelos podem ser aplicados em outros hospitais, desde que estes apresentem os mesmos problemas e utilizem os mesmos procedimentos no processo de IVG.

Por último, os modelos de DM criados estão aptos para serem induzidos em tempo real, sempre que uma utente dê entrada no CMIN. Porém, os melhores modelos obtidos podem variar de acordo com o conjunto de dados utilizados.

Pesquisas futuras podem considerar alguns aspetos como a exploração de diferentes tipos e configurações de técnicas de DM (como o *Clustering*, que é capaz de caracterizar utentes de risco), a incorporação de novas variáveis nos modelos de previsão e a repetição da experiência com novos dados.

Capítulo 5

Aplicação do Business Intelligence na Saúde

Neste capítulo são mencionadas algumas das características que uma plataforma de [Business Intelligence \(BI\)](#) deverá apresentar, sendo abordadas algumas ferramentas disponíveis para o desenvolvimento da mesma. Além disso, é realizada ainda uma comparação entre as respectivas ferramentas, que permite concluir qual a mais adequada para a construção da plataforma de [BI](#) a implementar neste projeto.

5.1 Introdução

Cada vez mais, no setor da saúde, são procuradas aplicações capazes de apoiar as práticas diárias dos profissionais de saúde. Neste contexto, a integração de ferramentas *open source* no setor da saúde tornou-se numa realidade cada vez mais frequente, originando um progresso viável para este tipo de aplicações. As ferramentas *open source* são um modelo de desenvolvimento que além de ser grátis, tem o seu código fonte totalmente disponível para visualização, leitura, modificação e redistribuição dos utilizadores, sem qualquer tipo de restrição de propriedade do produto, ao contrário das ferramentas grátis, que apenas permitem a sua utilização sem a necessidade de pagamento.

As ferramentas *open source* diferem essencialmente do proprietário, no processo de desenvolvimento e nas licenças do produto. Todas as aplicações *open source* são licenciadas por uma licença de *open source*, que dá ao utilizador o direito de usar a ferramenta, aceder e modificar o código fonte ¹ e de a redistribuir sem quaisquer custos. Este tipo de ferramentas é bastante popular devido às inúmeras vantagens que apresentam, na medida em que prometem acelerar a difusão das soluções das **Tecnologias de Informação (TI)** na área da saúde, reduzir os custos do desenvolvimento, estimular a inovação a indivíduos que possam estar geograficamente distribuídos e aumentar a usabilidade das aplicações.

A opção de uma ferramenta com uma licença *open source* para a informática médica e para a saúde é importante na medida em que determina os direitos dos utilizadores, constitui um estímulo que influencia os programadores a participarem no projeto, determina a qualidade do produto e tem influência na vontade dos utilizadores para adotar determinada aplicação.

Em termos de custos, as organizações de saúde têm algumas vantagens com as ferramentas *open source*, uma vez que conseguem economizar em taxas de licenciamento e reduzir os custos ao realizarem a compra do *hardware* específico para um determinado produto ao proprietário deste [60, 61].

A aquisição de uma ferramenta de **BI** e a sua implementação é bastante vantajosa para a organização de saúde na medida em que a sua utilização poderá ter um impacto positivo global. Por exemplo, a plataforma de **BI** poderá originar a otimização de recursos nos mais variados departamentos; poderá conduzir o planeamento, o orçamento e a previsão de forma mais eficiente e mais precisa; poderá originar uma gestão dos suprimentos e da logística da organização para redução de custos e para garantir um fornecimento consistente; e possibilitará a melhoria da condição clínica do utente através de diagnósticos eficientes e da identificação e aplicação dos protocolos das melhores práticas de tratamento [62].

No presente capítulo são avaliadas algumas ferramentas *open source*/grátis, nomeadamente, o Qlikview, o Palo BI Suite, o Jaspersoft BI, o Tableau Pu-

¹Conjunto de palavras ou símbolos, correspondentes a instruções de uma das linguagens de programação existentes, estruturadas de forma lógica.

blic, o Spago BI e o Pentaho BI Suite. Após uma aprofundada revisão da literatura e de averiguar quais as ferramentas utilizadas nos sistemas já implementados, foram selecionadas estas ferramentas com base no seu bom desempenho nas mais diversas áreas [63–65].

A análise destas ferramentas de BI emergiu durante o desenvolvimento deste projeto cujo objetivo é a criação de uma plataforma para a visualização de indicadores da unidade de Interrupção Voluntária da Gravidez (IVG) do Centro Materno Infantil do Norte (CMIN) e, ainda, a integração de modelos de Data Mining (DM).

5.2 Estado da Arte

Atualmente, as organizações procuram cada vez mais flexibilidade, redução de custos e eficiência das plataformas de BI que implementam. Independentemente da dimensão da organização, estas têm vindo a dar cada vez mais valor à informação, pois é a partir desta que se torna possível a obtenção de conhecimento útil, indispensável para os utilizadores responsáveis pelos processos de tomada de decisão.

O BI veio então contribuir para o melhor desempenho das organizações, na medida em que engloba aplicações e tecnologias que permitem o armazenamento, a análise e o acesso à informação, apoiando os utilizadores no processo de tomada de decisão [66].

Historicamente, as pequenas e médias empresas eram excluídas do grupo de organizações que aderiam à implementação do BI, pelo fato de os custos serem excessivamente elevados e pela falta de acessibilidade. Com a tradicional arquitetura de BI, que requeria uma base de dados, ferramentas de Extract, Transform and Load (ETL), ferramentas para a criação de relatórios e para análises Online Analytical Processing (OLAP), as taxas iniciais das licenças desses produtos eram muito elevadas, pelo que esta não era uma solução viável para a maioria das organizações. Hoje em dia, qualquer ferramenta *open source* BI está disponível para *download* gratuito, facultando todas as ferramentas necessárias para a implementação do BI [67].

A variedade de características das plataformas de BI *open source* evoluiu

bastante nos últimos cinco anos. A quantidade e a qualidade dos produtos *open source* disponíveis têm vindo a aumentar, tornando-se neste momento alternativas viáveis para qualquer solução empresarial de média dimensão [68].

5.3 Objectivos

Nos últimos anos, têm sido aplicadas tecnologias de BI a várias áreas a fim de apoiar o processo de tomada de decisão dos utilizadores e permitir a extração de conhecimento a partir dos dados armazenados. As organizações de saúde não são exceção e, portanto, têm sido desenvolvidas plataformas de BI para múltiplas unidades de diferentes organizações de saúde. O objetivo deste capítulo é analisar algumas ferramentas *open source* e/ou grátis de BI existentes no mercado e a sua aplicabilidade na área clínica. A principal finalidade é a seleção de uma ferramenta que preferencialmente não tenha custos associados para a organização e que cumpra um conjunto de requisitos definidos.

5.4 Aplicações de BI *open source* na Saúde

Cada vez mais, as organizações de saúde produzem e armazenam grandes quantidades de dados. A falta de tecnologias que permitam recolher, analisar e distribuir os dados mais relevantes faz com que estas organizações continuem a ser ricas em dados, mas pobres em informação [62]. As organizações de saúde com uma visão de futuro encontram-se cientes de que os dados e o seu tratamento através das tecnologias de BI são essenciais para um processo de tomada de decisão fundamentado e preciso, indispensável para melhorar os serviços e assegurar o futuro dessas organizações [60, 61].

Os casos de organizações de saúde que adotaram ferramentas de BI *open source* em aplicações médicas são escassos. No entanto, os grandes benefícios que resultam da sua implementação em outras áreas (ex. retalho, finanças, gestão, entre outras) originou a sua introdução nesta área.

Assim, com a urgência de adquirir aplicações de informática médica na área da saúde, as ferramentas *open source* tem sido alvo de grande atenção por parte da indústria médica. Por exemplo, tem sido desenvolvido na Europa o projeto *open source* denominado de *Care2X*, composto por quatro componentes: o sistema de informação do hospital, a gestão prática, o servidor central de dados e o protocolo de permuta da saúde. O *Care2X* foi desenvolvido com o intuito de ultrapassar problemas de integração numa rede de múltiplos programas incompatíveis. É possível a integração de quase qualquer tipo de serviço, sistema, departamento, processo, dados ou comunicação de um hospital. O *Care2X* oferece suporte para o fluxo de trabalho clínico, incorporando grupos relacionados com o diagnóstico, bem como módulos de agendamento e de prescrição eletrónica.

Outro caso, o *openEHR*, que é patrocinado pela Fundação openEHR, promove o "desenvolvimento de uma plataforma de computação na saúde, interoperável e livre, onde o principal componente é o registo clínico eletrónico interoperável e clinicamente eficaz".

A Canada Health Infoway, criada com fundos federais e provinciais, iniciou uma iniciativa *open source* em 2005 para o desenvolvimento de uma ferramenta que os hospitais e os programadores pudessem utilizar para garantir a segurança na partilha de registos médicos de utentes entre várias entidades.

Estas iniciativas sugerem que desenvolvimentos em *open source* são uma forma viável de criação de aplicações na área da saúde [69–72].

Além disso, atualmente existem já algumas aplicações, desenvolvidas em ferramentas *open source*, implementadas em organizações de saúde, como por exemplo, as aplicações:

- **Turin ASL3** é uma plataforma desenvolvida em conjunto com a ferramenta *open source* Spago BI, que permite atribuir permissões de utilização conforme os diferentes tipos de utilizadores, disponibiliza documentos analíticos, possibilita a visualização de dados e permite ainda a utilização da tecnologia OLAP. Esta solução está implementada nas instituições de cuidados de saúde locais, no Serviço Nacional de Saúde

Italiano. O Turin ASL 3 nasceu em 1995 e está implementado em, pelo menos, dois hospitais da cidade de Turin (Amadeo di Savoia e Maria Vittoria), em Itália [73].

- **St antonius** é uma aplicação desenvolvida através do Pentaho BI Suite, com a finalidade de analisar os tempos de espera dos utentes. As vantagens desta aplicação são a melhoria da eficiência operacional, a eliminação de custos resultantes da criação de relatórios manuais, a análise comportamental, a identificação de padrões e a análise de riscos. Esta aplicação está implementada no hospital St Antonius, localizado em Nieuwegein, na Holanda [74].
- **NHS** é uma aplicação desenvolvida para o tratamento médico preventivo para utentes de alto risco. Esta aplicação recorre também ao Pentaho BI Suite e permite uma melhoria dos serviços clínicos prestados, através do acesso à informação que apoia o processo de tomada de decisão; permite uma redução dos tempos de espera do utente e ainda a partilha da informação do mesmo ao longo do processo de tratamento. Um dos locais onde foi realizada a sua implementação é em Islington, um bairro londrino, sendo esta uma das 150 organizações locais do NHS (National Health Service) no Reino Unido [74].
- **ShareableInk** é uma aplicação destinada a melhorar os cuidados prestados ao utente e a analisar métricas de custos e de cumprimento através da análise de grandes fluxos de dados clínicos. É também uma aplicação desenvolvida através do Pentaho BI Suite e permite que a organização de saúde evite multas, maximize o financiamento público disponível e permita a prevenção e a deteção de fraudes. Recentemente, a empresa responsável pela aplicação, desenvolveu uma parceria estratégica com a T-System, que é uma empresa especializada em documentação de departamentos de emergência com mais de 1700 clientes, nomeadamente, hospitais nos Estados Unidos. [74].

Estas aplicações são alguns exemplos de soluções BI implementadas através de ferramentas *open source* e que, como se pôde verificar, apresentam

muitas vantagens para as organizações de saúde.

5.5 Requisitos das Aplicações em Ambientes Clínicos

Os requisitos das organizações de saúde para a implementação de uma plataforma de BI são principalmente a capacidade de fornecer informação que apoie o processo de tomada de decisão a um nível estratégico, com implicações a um nível operacional, recorrendo aos dados atuais e históricos (dos utentes), permitindo a sua visualização através de relatórios, gráficos e indicadores de desempenho por meio de ferramentas de processamento analítico. Assim, em termos práticos, as aplicações médicas, mais concretamente as de BI para a área da IVG, devem apresentar um conjunto de requisitos para satisfazerem as necessidades dos profissionais de saúde. Algumas dessas características são:

- **Desempenho:** avalia se a ferramenta tem um bom desempenho no processamento de consultas com um elevado volume de dados. Na área da saúde é importante que o desempenho seja bom, uma vez que é uma área onde as decisões tomadas tem um grande impacto na vida dos seres humanos.
- **Consultas *ad-hoc* OLAP:** avalia se a ferramenta permite ao utilizador ter a liberdade de definir consultas que considera apropriadas num determinado contexto. O OLAP é uma técnica que permite aos utilizadores realizar análises *ad-hoc* nos dados, considerando múltiplas dimensões, oferecendo a informação necessária para o processo de tomada de decisão ainda mais eficiente. Isso permite a análise do histórico dos documentos, a utilização de operações como o *roll-up*, o *drill-down*, o *slice e dice* e o *pivot*. No sector da saúde, a análise do histórico dos dados clínicos é muito importante, uma vez que permite a visualização da evolução do utente ao longo do tempo. Assim, é muito importante a existência desta técnica na plataforma de BI.

- **Arquitetura:** avalia se a ferramenta implementa arquiteturas OLAP que possuem alta escalabilidade, isto é, que seja capaz de processar de forma uniforme a informação, mesmo que a sua carga seja crescente.
- **Exibição de Key Performance Indicators:** avalia se a ferramenta disponibiliza a visualização de indicadores de desempenho da organização.
- **Plug-ins:** avalia se a ferramenta possibilita o desenvolvimento e a utilização de *plug-ins* que acrescentem funcionalidades à mesma.
- **Visualização dos dados interativa:** avalia se a ferramenta permite a ocorrência de interatividade entre o utilizador e as *dashboards*, os relatórios e os gráficos. Esta é uma característica que apesar de parecer pouco importante na área clínica, torna-se bastante relevante pelo fato de a interatividade ser bastante apelativa ao utilizador e facilitar a compreensão da informação demonstrada.
- **Documentação:** avalia o nível de qualidade da documentação disponibilizada pela ferramenta. Esta característica é bastante importante para o programador que desenvolve a aplicação, pois o processo de instalação por vezes é um procedimento complicado que requer documentação.
- **Dashboards:** avalia se a ferramenta disponibiliza a criação de painéis, que possibilitem a integração de gráficos, tabelas e outras análises, como por exemplo a de OLAP.
- **Recursos de navegação:** avalia se a ferramenta possibilita a criação de relatórios ou *dashboards*, recorrendo a operações de *roll-up*, *drill-down*, *slice e dice* e *pivot*.
- **Extract, Transform and Load (ETL):** característica que, no caso do BI, se aplica à opção de se integrar o processo de extração, transformação e carregamento dos dados, através da criação de procedimentos incorporados na ferramenta.

- **Conexão à base de dados:** é muito importante que a ferramenta de BI permita que seja estabelecida uma conexão a diferentes bases de dados, de modo a que seja possível incorporar informação de diferentes fontes de dados. Existem ferramentas em que a única possibilidade de ligação inerente é a visualização de dados, outras em que a conexão pode ser feita via ETL e via visualização de dados e, por último, outras em que a ligação é feita apenas via ETL. Este último tipo de ferramentas são integradas, o que não permite que estas estejam conectadas a várias fontes de dados. Num hospital, esta é também uma das principais características, porque normalmente estas organizações possuem sistemas interoperáveis, onde são comuns diferentes bases de dados que armazenam a informação clínica. Assim, de modo a facilitar a obtenção do Data Warehouse (DW), geralmente opta-se por uma ferramenta específica para a construção do DW e, posteriormente, utiliza-se uma ferramenta de BI para a criação dos cubos OLAP e para a visualização dos indicadores de desempenho.
- **Integração do modelo dimensional:** avalia se a ferramenta de BI permite a integração do modelo dimensional do DW.
- **Open source:** avalia se a ferramenta apresenta um modelo de desenvolvimento que, além de ser gratuito, tem o código fonte totalmente disponível para os utilizadores o visualizarem, modificarem, e redistribuírem sem restrições do proprietário do produto.
- **Exportação:** avalia se a ferramenta permite a exportação para outros formatos como PDF, HTML, entre outros.
- **Pervasive:** avalia se a versão *open source* da ferramenta fornece o próprio servidor, permitindo a visualização/abertura da aplicação num *browser*. Se a ferramenta possui esta característica, então não é necessária a sua instalação em todos os computadores da organização destinados para a utilização da aplicação, mas apenas num servidor, sendo que todos os computadores conectados à rede da organização conseguem aceder à aplicação. Num hospital, esta característica é bastante

importante na medida em que além de reduzir os custos nos processos de instalação em todos os computadores interessados na aplicação, permite também reduzir o tempo gasto, o que na área da saúde é muito importante. Especificamente, no Centro Hospitalar do Porto (CHP), este aspeto é também muito importante para o desenvolvimento das aplicações de BI, uma vez que a plataforma de BI estará integrada na Agência de Integração, Difusão e Armazenamento (AIDA), permitindo a interoperabilidade em todas as organizações constituintes do CHP. Assim, este atributo *pervasive* da aplicação é implementado neste contexto, ou seja, de forma a que a informação seja distribuída pelos utilizadores da organização e não focada apenas para o topo da pirâmide organizacional.

- **Aprendizagem *Online*:** avalia se a ferramenta disponibiliza o recurso de ajuda *online*.
- **Suporte a dispositivos móveis:** avalia se a ferramenta suporta o uso de dispositivos móveis, o que poderá ser bastante útil numa organização de saúde, na medida que os profissionais de saúde poderão aceder à informação por outros meios distintos do computador.
- **Data Mining:** avalia se a ferramenta disponibiliza o recurso de utilização dos dados para prever tendências futuras, situações clínicas e padrões de comportamento.
- **Facilidade de uso:** avalia a facilidade com que um utilizador pouco experiente conseguirá identificar as funcionalidades da ferramenta, onde encontrá-las e como executá-las.
- **Atratividade:** avalia o grau de atratividade da interface que a ferramenta possui.
- **Personalização da interface:** identifica se a ferramenta permite a customização da interface, pelo o administrador.

- **Perfil de utilizador:** verifica se a ferramenta permite que o administrador defina níveis hierárquicos, atribuindo diferentes permissões de acesso aos utilizadores do sistema.
- **Tempo Real:** é uma abordagem para a análise de dados, que permite que os utilizadores possam aceder à aplicação desenvolvida e à informação nesta presente em tempo real. Esta é uma característica importante nas organizações de saúde, pois é crucial que os profissionais de saúde possam aceder a dados atuais que suportem o processo de tomada de decisão.

Após a análise exaustiva de cada uma das ferramentas, incluindo a instalação e o teste das mesmas, estas características foram avaliadas e comparadas na secção 5.6.7 para cada uma das ferramentas abordadas na secção seguinte [63–65, 75–77].

5.6 Ferramentas de BI

5.6.1 QlikView

O QlikView (QV) é uma ferramenta de BI desenvolvida pela empresa sueca QlikTech. Apesar de ser um produto patenteado, a empresa oferece uma versão integral gratuita para o desenvolvimento da mesma. Assim, esta empresa disponibiliza várias opções de licenciamento, limitando a utilização da ferramenta de acordo com a licença adquirida pelo o utilizador.

Uma característica chave do QV é o facto de ele possibilitar uma extração bastante simples dos dados a partir de diferentes fontes, permitindo o estabelecimento de conexões através do *Open Database Connectivity* e do *Object Linking and Embedding Database*, que são interfaces de comunicação entre o sistema operacional e as várias bases de dados. Esta ferramenta permite ainda a utilização da tecnologia OLAP, disponibilizando vários tipos de operações que facilitam a navegação do utilizador entre diferentes dimensões através das consultas *ad-hoc*. Além disso, possibilita também a criação

de uma interface flexível e intuitiva com gráficos e tabelas *pivot*, análises estatísticas, entre outros.

Outro atributo muito importante é que ela não utiliza **Sistemas de Gestão de Bases de Dados (SGBD)** como ferramenta de armazenamento. O sistema conecta-se à base de *Online Transaction Processing*, utilizado apenas após a conclusão do processo de carregamento de dados. Assim, os dados são submetidos aos processos de **ETL** e comprimidos num ficheiro com a extensão *.qvw*, que são compreendidos pelo **QV**. Este ficheiro contém todos os pormenores necessários para a análise dos dados, incluindo os próprios dados, a informação necessária para atualizar o arquivo do *script* com novos dados a partir da base de dados, a informação do *layout* (pastas, listas, gráficos, etc.), alertas, favoritos, documentos e relatórios, informação sobre as restrições de acesso e as macros do módulo.

Além disso, a distribuição de informação torna-se facilitada e a análise dos dados pode ser realizada independentemente da localização dos dados originais ou das condições de rede. Assim, surge a possibilidade do utilizador visualizar a aplicação de diferentes formas, uma vez que o ficheiro *.qvw* gerado pode ser utilizado em qualquer máquina com os dados e relatórios produzidos. O acesso à aplicação pode também ser feito a partir de um *browser*, através de um servidor da aplicação, de acordo com as regras de segurança estipuladas pelo cliente. No entanto, isto só é possível na versão paga do **QV**. Na versão gratuita, o documento final só pode ser utilizado na máquina onde foi gerado, apenas através do utilizador que o desenvolveu.

As vantagens do **QV** resumem-se então à facilidade de criação de relatórios por parte dos utilizadores finais, uma vez que depois de realizado o processo de **ETL**, os relatórios podem ser produzidos com a inclusão de gráficos e conhecimentos básicos de funções agregadas, como por exemplo, somas e contagens [64, 78].

Na Figura 5.1 estão representadas as diversas fontes de dados às quais o **QV** se consegue conectar e as múltiplas plataformas que permitem o acesso ao **QV**.



Figura 5.1: Fontes de dados e plataformas compatíveis com o QlikView [79].

5.6.2 Palo BI Suite

Os produtos Jedox, empresa fundada na Alemanha, são totalmente *open source*, preparados para serem instalados e utilizados. O Palo BI Suite é um produto Jedox, que tem como objetivo minimizar a quantidade de sobrecarga, tempo, custos e de apoio informático na criação e manutenção de soluções BI. Assim, os utilizadores assumem um papel de liderança através da utilização e da gestão do Palo BI Suite.

Esta ferramenta na versão comunitária é constituído pelos seguintes componentes:

- Palo OLAP Server;
- Palo Excel Add-in;
- Palo Worksheet-Server;
- Palo ETL-Server

O Palo OLAP Server apresenta uma tecnologia OLAP multidimensional em memória (Molap). Os dados que são armazenados estão organizados em

cubos, dimensões, elementos e atributos de elementos. Comparativamente ao OLAP relacional (Rolap), o Molap tem o potencial de ser 100 vezes mais rápido. Um simples cubo Palo pode conter dados de múltiplas fontes de dados, o que simplifica a análise de dados de diferentes fontes, como por exemplo, no caso da comparação de um conjunto de dados atuais e um conjunto de dados previstos. Além disso, os utilizadores podem importar os dados que necessitam de forma incremental, a partir de qualquer base de dados que o Palo tenha acesso, e armazená-los numa estrutura definida por eles. O modo como as tabelas estão dispostas na base de dados não é relevante, uma vez que a arquitetura é definida novamente na própria ferramenta.

O Palo Excel Add-in permite o estabelecimento de uma ligação entre o Excel e uma base de dados multidimensional e centralizada como o Palo OLAP Server, tornando-se numa ferramenta de BI altamente sofisticada.

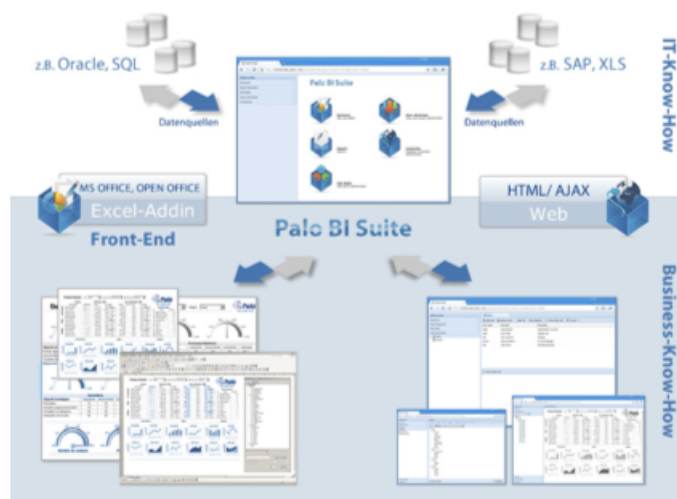


Figura 5.2: Acesso aos dados multidimensionais, recorrendo ao Excel Add-in ou à interface *Web* [80].

O Palo Worksheet-Server oferece um sistema de análise e de relatórios baseado na *Web*, que pode ser gerido pelos utilizadores. Os relatórios podem ser criados e publicados na *Web*, sendo fornecida uma gestão do conteúdo flexível e segura. Sendo este componente bastante semelhante ao Excel, os utilizadores facilmente se adaptam ao método de criação de relatórios. Além

disso, tem a vantagem de que as folhas de cálculo resultantes são atualizadas automaticamente sempre que surgem novos dados nas bases de dados.

Na Figura 5.2 está representado o acesso aos dados multidimensionais, recorrendo ao Excel Add-in ou à interface *Web*.

Por último, o Palo ETL-Server apoia a extração, a transformação e o carregamento dos dados. O Palo tem a capacidade de permitir a extração de grandes quantidades de dados originárias de uma ampla gama de fontes de dados. O ETL-Server não se limita a carregar o Palo Olap Server, mas adapta-se às necessidades específicas de importação/exportação dos dados para/de os modelos Palo [63, 65, 75].

5.6.3 Jaspersoft BI

O Jaspersoft BI é uma ferramenta que foi desenvolvida em 2001 nas linguagens de programação Java e Perl. Esta é uma ferramenta *open source*, que está disponível em duas versões, a versão comunitária e a versão empresarial. A versão comunitária do Jaspersoft BI é constituída por vários componentes, representados na Figura 5.3:

- Jaspersoft iReport Designer;
- Jaspersoft Studio;
- JasperReports Library;
- JasperReports Server;
- Jaspersoft OLAP;
- Jaspersoft ETL;

O Jaspersoft iReport Designer é responsável pelo *design* dos relatórios do JasperReports Library e do JasperReports Server. Este componente permite a criação de *layouts* sofisticados com a presença de imagens, gráficos, tabelas, entre outros. Além disso, permite o acesso aos dados via JDBC, TableModels, JavaBeans, XML, Hibernate, CSV e fontes de dados personalizadas. Os

relatórios podem ainda ser exportados nos formatos PDF, RTF, XML, XLS, CSV, HTML, DOCX ou em formatos OpenOffice.

O Jaspersoft Studio apresenta as mesmas funcionalidades que o Jaspersoft iReport Designer, distinguindo-se apenas pelo fato de utilizar um ambiente integrado de desenvolvimento baseado no Eclipse.

O JasperReports Library é um mecanismo de criação de relatórios *open source*, desenvolvido totalmente em Java, que permite a utilização de dados de qualquer fonte de dados e a produção de documentos que podem ser visualizados, impressos e exportados numa grande variedade de formatos.

O JasperReports Server é um servidor que permite o acesso aos relatórios e a realização das respectivas análises, que podem ser incorporadas em páginas *Web* ou em aplicações móveis. Estes relatórios podem fornecer assim informação em tempo real e serem programados para *browsers*, dispositivos móveis, impressoras ou correio eletrónico em diferentes tipos de formatos de ficheiros. Este componente está otimizado para partilhar, proteger e gerir centralmente os relatórios Jaspersoft e as respectivas visões analíticas.

O Jaspersoft OLAP é um ambiente para análises de dados, que pode ser acedido pelo utilizador através de uma interface bastante intuitiva. Este componente foi criado para a análise de grandes conjuntos de dados e para a realização de consultas analíticas complexas. Esta ferramenta, que consiste num mecanismo **OLAP**, oferece um ambiente interativo para os utilizadores realizarem operações de *slice e dice*, *pivot*, filtragem e de sintetização dos dados em tempo real através de uma interface baseada na *Web* ou no Microsoft Excel.

Por último, o Jaspersoft ETL consiste na implementação e na execução dos procedimentos de **ETL** nos vários sistemas externos. É utilizado para extrair dados de um sistema transacional para a criação de um **DW** ou de um *Data Mart*, para a posterior conceção dos relatórios e análises [63, 65, 75, 76].

5.6.4 Tableau Public

O Tableau Public é uma ferramenta baseada na *Web* correspondente à versão *open source* do Tableau Desktop. Esta ferramenta permite a criação

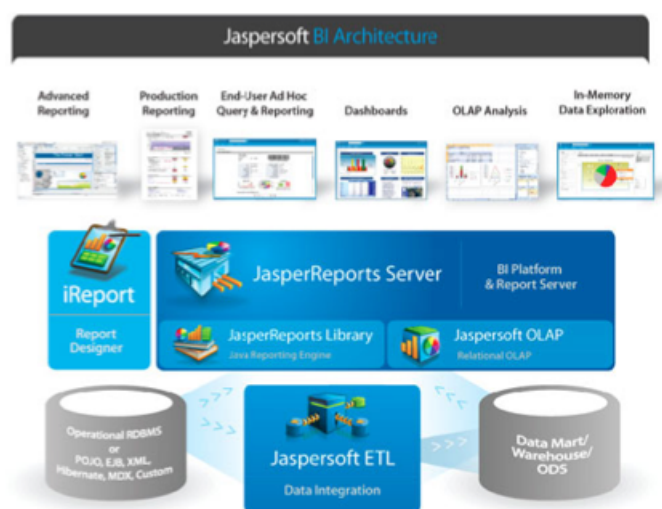


Figura 5.3: Arquitetura da ferramenta Jaspersoft BI [75].

de visualizações dos dados de forma interativa e ainda a sua integração num *website*, publicado na galeria do Tableau Public ou partilhado na comunidade Tableau Public. Esta versão *open source* difere essencialmente da versão paga, pelo fato de na versão *open source* não ser possível a gravação local das visualizações, possibilitando apenas a geração de um código que pode ser integrado em qualquer página *Web*. No entanto, o Tableau Public oferece também vários tipos de representação dos dados, como gráficos, tabelas e mapas.

Esta ferramenta permite a utilização de vários tipos de formatos de ficheiros, como ficheiros de texto, folhas de cálculo, bases de dados e ainda arquivos *csv*. Quando se estabelece uma conexão a uma base de dados, esta ferramenta identifica o papel de cada campo, de forma a verificar se o respetivo contém dimensões e factos, isto é, tem a capacidade de reconhecer o modelo dimensional da estrutura onde os dados estão armazenados. Assim, quando os dados são selecionados para serem visualizados com o Tableau Public, são identificados os títulos, as colunas e as linhas dos dados estruturados. O processo de construção de gráficos fica simplificado a apenas movimentos de arrastar e de largar os valores nos sítios apropriados [81].

5.6.5 Spago BI

O Spago BI é uma ferramenta completamente *open source*, existindo apenas uma única versão, a versão comunitária gratuita. É uma aplicação desenvolvida pela Spago World, cujas funcionalidades estão representadas na Figura 5.4, e é constituída por vários módulos:

- Spago BI Server;
- Spago BI Studio
- Spago BI Meta;
- Spago BI SDK;
- Spago BI Applications;

O Spago BI server corresponde ao módulo principal. Este é o núcleo da ferramenta e oferece todas as suas capacidades analíticas.

O Spago BI Studio é o ambiente de desenvolvimento utilizado, que se baseia no Eclipse. Este componente permite ao utilizador criar e modificar todos os documentos de análise, como é o caso dos relatórios, dos cubos **OLAP**, das *dashboards* e ainda do **DM**. A interação entre este módulo e o Spago BI Server é possível graças ao módulo Spago BI SDK.

O Spago BI Meta é um componente orientado para a gestão de metadados e para a pesquisa. A plataforma gere os metadados, permitindo ao utilizador editá-los e importá-los a partir de ferramentas externas específicas para o **ETL**. Este módulo enriquece a base de conhecimento dos metadados do Spago BI Server, de modo a que eles possam ser facilmente consultados através de ferramentas disponíveis, como o **OLAP**.

O Spago BI SDK é uma ferramenta utilizada para integrar os serviços fornecidos pelo servidor. Especificamente, é utilizado pelo Spago BI Studio, para que os utilizadores possam fazer o *download* ou o *upload* de documentos de análise de/para o servidor. Este módulo permite a integração dos documentos, graças à ampla gama de serviços disponíveis através de um serviço *Web*, e a publicação de documentos Spago BI num portal externo.

Por último, o Spago BI Applications é uma coleção de modelos analíticos desenvolvidos, utilizando o Spago BI. Estes modelos são criados tendo em consideração o setor do mercado, a finalidade do produto final, bem como a componente analítica específica [63, 65, 75].



Figura 5.4: Funcionalidades da ferramenta Spago BI [65].

5.6.6 Pentaho BI Suite

O Pentaho BI Suite foi desenvolvido pela Pentaho Corporation em 2001, na linguagem Java, sendo a primeira ferramenta de BI a ser lançada no mercado como uma alternativa *open source*. O Pentaho oferece dois tipos de licenças, a Community Edition (CE), que corresponde à versão *open source*, e a Enterprise Edition (EE).

O projecto Pentaho BI Suite engloba um conjunto de produtos: a plataforma BI (servidor), criação de relatórios, análises OLAP, integração dos dados (ETL), dashboards e DM.

A plataforma que integra o Pentaho é 100% *Java 2 Platform, Enterprise Edition*, assegurando a escalabilidade, a integração e a portabilidade deste. Além disso, permite o estabelecimento de conexões a bases de dados dos tipos JDBC, IBM, DB2, Microsoft SQL Server, MySQL, Oracle, PostgreSQL,

Firebird e NCR Teradata.

O Pentaho BI, cuja arquitetura está representada na Figura 5.5, está dividido em vários módulos:

- Pentaho BI Platform: fornece vários serviços aos utilizadores finais como a criação de relatórios, a integração de ferramentas, uma segurança centralizada e integrada dos dados, entre outros.
- Pentaho Reporting: permite o desenvolvimento de relatórios com facilidade, possibilitando a organização da forma de aceder, formatar e distribuir a informação. Este módulo contém todas as características gráficas para a construção de relatórios, bem como a execução de consultas *ad-hoc*.
- Pentaho Analysis: fornece uma análise OLAP, apoiando os utilizadores no processo de tomada de decisão. Este módulo possibilita uma exploração interativa da informação através da interseção dos dados, permitindo ainda uma integração completa com outros serviços disponíveis no Pentaho BI Suite, através de *plug-ins*.
- Pentaho Data Integration: ferramenta bastante poderosa para a realização do processo de ETL, utilizando uma abordagem inovadora, orientada por metadados.
- Community Dashboard Editor: oferece um ambiente gráfico que permite aos utilizadores ter acesso a informação essencial para a compreensão e para a otimização do desempenho da organização. Este componente pode ser completamente integrado com o Pentaho Reporting e com o Pentaho Analysis.
- Weka Pentaho Data Mining: o Weka é uma ferramenta de DM *open source*, que permite a realização de análises de previsão, fornecendo informação sobre padrões ocultos e relações entre dados, bem como sobre indicadores de desempenho. Este módulo oferece uma interface gráfica para o pré-processamento dos dados, para a utilização de técnicas de classificação, regressão, clustering e de regras de associação nos

dados e ainda para a visualização dos mesmos. Como se pode verificar, o Pentaho permite integrar esta ferramenta, juntamente com as suas restantes funcionalidades [63–65, 75].

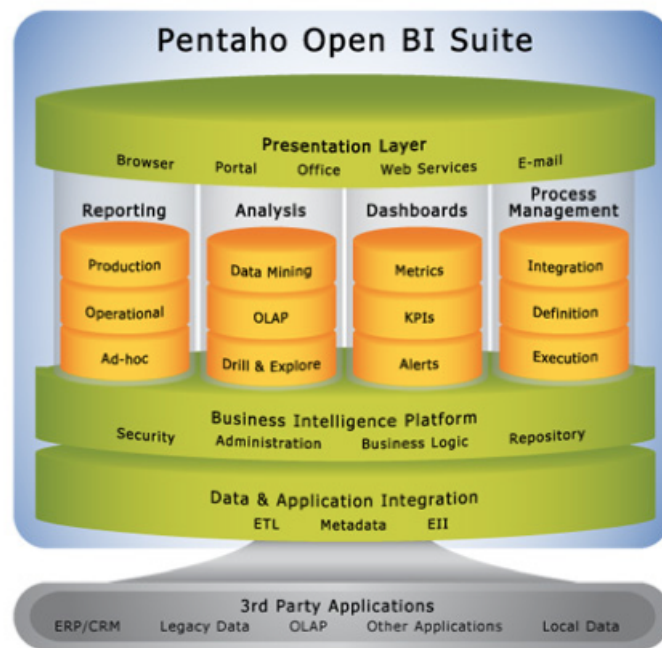


Figura 5.5: Arquitetura da ferramenta Pentaho BI Suite [65].

5.6.7 Resultados - Análise comparativa entre as ferramentas

Nesta secção, é apresentada a relação existente entre cada uma das características referenciadas como requisitos para a plataforma de BI no ambiente clínico e todas as ferramentas anteriormente analisadas. Assim, a Tabela 5.1 consiste numa tabela comparativa das ferramentas de BI relativamente aos requisitos pretendidos.

74CAPÍTULO 5. APLICAÇÃO DO BUSINESS INTELLIGENCE NA SAÚDE

Tabela 5.1: Comparação das ferramentas de BI em relação aos requisitos selecionados e atribuição de um grupo a cada uma das características.

	Ferramentas de BI						Grupo
	Jaspersoft BI	Palo BI Suite	Pentaho BI Suite	QlikView	Spago BI	Tableau Public	
Desempenho	4	3	4	3	4	4	Outros Importantes
Consultas <i>ad-hoc</i> OLAP	1	5	5	3	5	4	Tecnologias
Arquitetura	4	4	5	4	5	4	Outros Importantes
Exibição de <i>KPIs</i>	1	1	5	4	4	4	<i>Must have</i>
<i>Plug-ins</i>	3	0	5	0	0	3	Outros Importantes
Visualização de dados interativa	5	4	5	5	4	4	Utilizador final
Documentação	4	4	2	2	2	3	Administrador
<i>Dashboards</i>	1	1	4	4	5	4	Tecnologias
Recursos de navegação	5	4	4	1	2	4	Utilizador final
ETL	4	5	5	3	4	1	Tratamento de dados
Conexão à base de dados	5	4	4	5	5	3	<i>Must have</i>
Integração do modelo dimensional	1	1	1	2	4	1	Tratamento de dados
<i>Open source</i>	5	5	5	1	5	5	Outros Importantes
Exportação	5	2	5	2	5	4	Utilizador final
<i>Pervasive</i>	5	5	5	1	5	4	<i>Must have</i>
Aprendizagem <i>online</i>	4	2	3	4	3	4	Administrador
Suporte a dispositivos móveis	4	1	0	5	5	3	Utilizador final
Data Mining	1	1	3	2	4	1	Tecnologias
Facilidade de uso	4	4	4	4	4	5	Administrador
Atratividade	4	3	4	5	5	4	Utilizador final
Personalização da interface	4	0	5	5	5	5	Administrador
Perfil do utilizador	5	4	5	1	4	0	Outros Importantes
Tempo real	5	4	5	1	5	1	<i>Must have</i>

Para a análise comparativa das ferramentas selecionadas, foi adaptado um critério de classificação com os seguintes parâmetros:

- 0 - desconhecido;
- 1 - inexistente;
- 2 - insuficiente;
- 3 - suficiente;
- 4 - bom;
- 5 - excelente;

Assim, para cada uma das características foi atribuída uma classificação, conforme o nível de satisfação que a respectiva ferramenta apresenta. Esta atribuição foi realizada com base na experimentação e na opinião crítica dos utilizadores.

Depois de analisada cada uma destas características para cada uma das ferramentas, procedeu-se ao agrupamento desses requisitos por grau de afinidade entre eles. Assim, foram considerados seis grupos, onde se encontram distribuídos todos os requisitos, aos quais foram atribuídos uma percentagem de acordo com a importância que têm para a plataforma de BI a implementar no módulo de IVG no CMIN. Estas ponderações atribuídas foram definidas com base na opinião crítica dos profissionais de saúde e dos profissionais das TI.

Deste modo, inicialmente, definiu-se o grupo das características indispensáveis (*Must have*), cuja ausência impede o correto funcionamento da plataforma de BI. A este grupo não foi associada nenhuma percentagem, pois inclui características cuja importância não pode ser quantificada, isto é, são estritamente necessárias.

Posteriormente, foi criado um grupo em que os requisitos estão direcionados para o proveito do administrador da ferramenta responsável pela construção da plataforma de BI. A este grupo foi atribuída uma percentagem de 5%, pelo fato de não ter grande importância para a organização de saúde em

questão. Poderá ter algum valor apenas para o programador que administra a ferramenta.

Tabela 5.2: Agrupamento das características em grupos com as respetivas percentagens.

Grupo	Caraterísticas	Percentagem
<i>Must have</i>	<i>Pervasive</i>	-
	Tempo real	
	Exibição de <i>KPIs</i>	
	Conexão às bases de dados	
Administrador	Documentação	5%
	Aprendizagem <i>online</i>	
	Facilidade de uso	
	Personalização da interface	
Utilizador final	Visualização dos dados interativa	25%
	Recursos de navegação	
	Exportação	
	Suporte a dispositivos móveis	
	Atratividade	
Tecnologias	Consultas <i>ad-hoc</i> OLAP	30%
	<i>Dashboards</i>	
	Data Mining	
Outras Importantes	Desempenho	25%
	Arquitetura	
	<i>Plug-ins</i>	
	<i>Open source</i>	
	Perfil de utilizador	
Tratamento de dados	ETL	15%
	Integração do modelo dimensional	

Outro grupo definido foi o grupo cujas características são vantajosas para o utilizador final da plataforma de BI. A este grupo foi atribuída uma percentagem de 25%. Este tem uma percentagem mais elevada, pois é importante

que os profissionais de saúde encarregues de utilizar a aplicação de BI estejam satisfeitos com a mesma e não tenham dificuldades em utilizá-la; caso contrário, esta poderá ser evitada e poderá originar o insucesso da sua implementação.

O grupo com maior percentagem associada (30%) corresponde às tecnologias que a ferramenta incorpora, pelo fato de ser o foco de estudo e da análise das várias ferramentas. Sem características como tecnologias OLAP ou *dashboards*, a plataforma de BI perde todo o interesse que suscita nos profissionais de TI e nos profissionais de saúde.

Existe ainda o grupo das características que não são obrigatórias na ferramenta a utilizar, mas que são muito importantes para a organização de saúde. Deste modo, a percentagem atribuída foi de 25%.

Por fim, o último grupo criado está associado ao tratamento dos dados e tem uma percentagem de 15%. Neste grupo, as características incluídas seriam vantajosas se a ferramenta as incorporasse, porém, não tem muita influência no sucesso da construção da plataforma de BI. Por exemplo, no caso do ETL, se a ferramenta pudesse incorporar os procedimentos de ETL, teria muitos benefícios. No entanto, o ETL pode ser realizado utilizando outras ferramentas distintas destas.

Os grupos, juntamente com as características associadas e as respetivas percentagens estão representadas na Tabela 5.2.

5.6.8 Discussão

Depois de construída a Tabela 5.1 e a Tabela 5.2, pode-se obter uma classificação final para cada uma das ferramentas de BI, somando cada uma das classificações atribuídas aos requisitos em cada um dos grupos definidos e multiplicando pelas respetivas percentagens. Por exemplo, para o caso do Pentaho BI Suite tem-se as seguintes classificações para cada um dos grupos:

- **Must have:** $5+5+5+4=19$
- **Administrador:** $(2+3+4+5) \times 0.05=0.7$
- **Utilizador final:** $(5+4+5+0+4) \times 0.25=4.5$

- **Tecnologias:** $(5+4+3) \times 0.3=3.6$
- **Outras importantes:** $(4+5+5+5+5) \times 0.25=6$
- **Tratamento de dados:** $(5+1) \times 0.15=0.9$

Pode-se verificar então que a classificação final do Pentaho BI Suite é de $19+0.7+4.5+3.6+6+0.9=34.7$. Este procedimento foi repetido para cada uma das restantes ferramentas, estando representados na Tabela 5.3 as respectivas classificações finais.

Tabela 5.3: Classificação final atribuída a cada um das ferramentas de BI.

	Classificação Final
Spago BI	34.85
Pentaho BI Suite	34.7
Jaspersoft BI	29.45
Palo BI Suite	25
Tableau Public	24.6
QlikView	21.95

Analisando os valores registados na Tabela 5.3, verifica-se que à primeira vista, a ferramenta mais apropriada para a implementação da plataforma de BI é o Spago BI pelo fato de apresentar a melhor classificação global. No entanto, esta não foi a ferramenta selecionada, uma vez que o processo de instalação desta é bastante complexo, não existe grande documentação como suporte de apoio e, além disso, esta ferramenta ocupa grande parte da memória *Random Access Memory*.

Assim, a ferramenta selecionada para a plataforma de BI desenvolvida para o CHP foi o Pentaho BI Suite. Esta ferramenta além de apresentar a segunda melhor classificação, é uma ferramenta que apresenta uma agradável interface, uma grande variedade de formas de representação da informação e possui uma grande lista de *plug-ins* que permitem ao utilizador personalizar a aplicação.

A opção por esta ferramenta pode ser justificada por uma comparação entre o Pentaho e cada uma das restantes ferramentas. Assim, considerando

o Jaspersoft BI, o Pentaho fornece mais capacidades em termos de *dashboards*. Além disso, o Pentaho possui muitos mais *plug-ins* desenvolvidos, como o CDE, o Saiku e o OpenI, disponíveis no *marketplace*, ao contrário do Jaspersoft BI, que apresenta um número muito limitado de *plug-ins* até ao momento.

Relativamente ao Palo BI Suite, este não permite a apresentação de indicadores de desempenho e, ainda, não permite a integração de processos de DM, ao contrário do Pentaho.

A grande desvantagem do QV relativamente ao Pentaho é o facto de não permitir o acesso à aplicação através de *browsers*. Esta é uma característica indispensável, visto que é essencial que a aplicação possa ser acedida em qualquer lugar através de uma página *Web* e que possa ser acedida por múltiplos utilizadores em simultâneo. Além disso, o QlikView não é uma ferramenta *open source*, é apenas gratuito.

Por último, no Tableau Public salienta-se a ausência de uma das principais características que é a ligação a uma base de dados Oracle, visto que todos os dados utilizados neste projeto estavam armazenados numa base de dados Oracle.

5.6.9 Conclusão

Neste capítulo, foi inicialmente realizado um levantamento dos requisitos considerados fundamentais para a conceção e a implementação de uma plataforma de BI na área clínica. Depois de analisadas essas características, foi selecionado um conjunto de ferramentas de BI, com base na revisão da literatura.

Depois de estabelecido o critério de avaliação a utilizar para diferenciar as ferramentas selecionadas, procedeu-se ao processo de instalação de cada uma das ferramentas e avaliou-se cada um dos requisitos definidos com uma escala estabelecida. Convém salientar que muitas das características definidas tem um carácter subjetivo, pelo que a sua avaliação pode variar de utilizador para utilizador. Posteriormente, essas características foram agrupadas em seis conjuntos de acordo com a sua proximidade e foram atribuídas percentagens

a cada um desses grupos, com base no conhecimento dos profissionais de **TI** e de saúde. Essas percentagens traduzem o nível de importância que o conjunto desses requisitos têm para o **CHP**, pelo que poderão tomar diferentes valores no caso de aplicados a outras organizações de saúde.

Assim, com base nesta investigação realizada, pode-se concluir que a ferramenta que melhor se adequa para a finalidade pretendida, isto é, para o ambiente clínico, é o Pentaho BI Suite, não deixando de salientar que esta opção é bastante influenciada pela organização de saúde onde se pretende aplicar.

Por último, é então importante salientar que a seleção de uma ferramenta de **BI** implica uma análise prévia dos requisitos que são considerados cruciais e indispensáveis na ferramenta, bem como os fatores que caracterizam o ambiente onde se pretende aplicar a ferramenta.

Capítulo 6

Business Intelligence

Neste capítulo é realizada uma introdução teórica sobre a metodologia utilizada no processo de **Business Intelligence** (BI), seguindo-se uma breve revisão literária do tema em questão. Posteriormente, são estabelecidos os objetivos pretendidos com a realização deste capítulo e descritas todas as etapas realizadas ao longo deste projeto para a conceção e implementação da plataforma de BI. Por fim, são discutidos os resultados obtidos e mencionadas as principais conclusões.

6.1 Introdução

Como se poderá constatar na próxima secção, a definição de BI varia entre os investigadores. Atualmente, uma plataforma de BI pode ser definida como "um conjunto de ferramentas, tecnologias e produtos programados e integrados que são utilizados para a recolha, a análise e apresentação dos dados". O BI tem a capacidade de transformar a informação armazenada em conhecimento útil e de o fornecer ao utilizador certo, em tempo útil, para apoiar o processo de tomada de decisão [1, 82].

Das ferramentas e tecnologias que integram o BI, podem-se destacar o **Data Warehouse** (DW), o **Data Mining** (DM), as ferramentas **Online Analytical Processing** (OLAP), entre outras.

O DW é a base para qualquer plataforma de BI, podendo apresentar

diferentes arquiteturas no seu desenvolvimento.

6.1.1 Data Warehouse

Segundo Bill Inmon [83], o criador do DW, este é definido como:

"a subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision-making"

Isto é, um DW é orientado por temas, uma vez que os dados são divididos por tabelas de acordo com o assunto a que estão associados; é integrado, pois os dados são uniformizados antes de serem carregados nesta estrutura; é variável no tempo, uma vez que cada entrada no DW corresponde a um particular momento no tempo; e é não volátil, pelo fato de os dados aí armazenados não serem removidos.

Assim, um DW é um sistema de computação que permite armazenar informação relacionada com a atividade de uma organização e proveniente de bases de dados existentes nesta. Com o DW é possível a análise de grandes volumes de dados, recolhidos dos sistemas transacionais da organização.

Este repositório de dados é construído com o intuito de apoiar o processo de tomada de decisão de uma organização. Os dados armazenados neste sistema deverão ser estruturados de modo a serem disponibilizados num formato válido e consistente para uma posterior utilização em processamento analítico, como consultas *ad-hoc*, elaboração de relatórios com *Key Performance Indicators*, entre outros [84].

O processo de conceção e implementação de um DW passa por um carregamento dos dados, provenientes de diversas fontes, nesta estrutura. Antes desse carregamento, são necessários procedimentos de extração dos dados das várias bases de dados e de transformação dos dados, de forma a torná-los consistentes e uniformes. Todo este processo é denominado de *Extract, Transform and Load (ETL)*.

Os DW devem ainda ser atualizados, sempre que existam novos dados nas fontes de dados, definindo-se uma periodicidade ou regularidade para tal.

Um DW pode apresentar diferentes abordagens na sua arquitetura, nomeadamente:

- *Abordagem Top-Down*: o DW é desenvolvido com base no modelo de dados da empresa, em que um único repositório de dados alimenta vários *Data Marts*, que são repositórios de menores dimensões. Esta arquitetura requer muito tempo para ser implementada. Um exemplo desta arquitetura é a de "The Inmon's approach".
- *Abordagem Bottom-Up*: esta é uma arquitetura cuja a implementação requer menos tempo e é realizada por etapas. Neste caso, o fluxo de dados é feito da fonte de dados para *Data Marts* e, posteriormente, dos *Data Marts* para um DW. Um exemplo desta arquitetura é a abordagem de Kimball, que é uma das metodologias mais utilizadas e é abordada com mais detalhe na seguinte secção [11, 85].

6.1.2 Metodologia de Kimball

No desenvolvimento da plataforma de BI para a unidade de Interrupção Voluntária da Gravidez (IVG) no Centro Materno Infantil do Norte (CMIN) foi seguida a abordagem de Kimball [38, 86]. Esta metodologia identifica uma sequência de tarefas e destaca atividades que devem ser executadas em paralelo, que pressupõem as seguintes etapas, representadas na Figura 6.1:

- *Planeamento do Projeto/Programa (Program / Project Planning)*: Esta fase inclui o planeamento do projeto, a compreensão dos requisitos da organização e a identificação das tarefas.
- *Gestão do Projeto/Programa (Program / Project Management)*: Nesta fase é monitorizado todo o projeto desde a fase inicial até ao fim, sendo responsável pelo reforço do plano do projeto inicialmente estabelecido.
- *Definição dos requisitos da organização (Business Requirements Definition)*: O sucesso do projeto depende de uma compreensão sólida dos requisitos da organização.

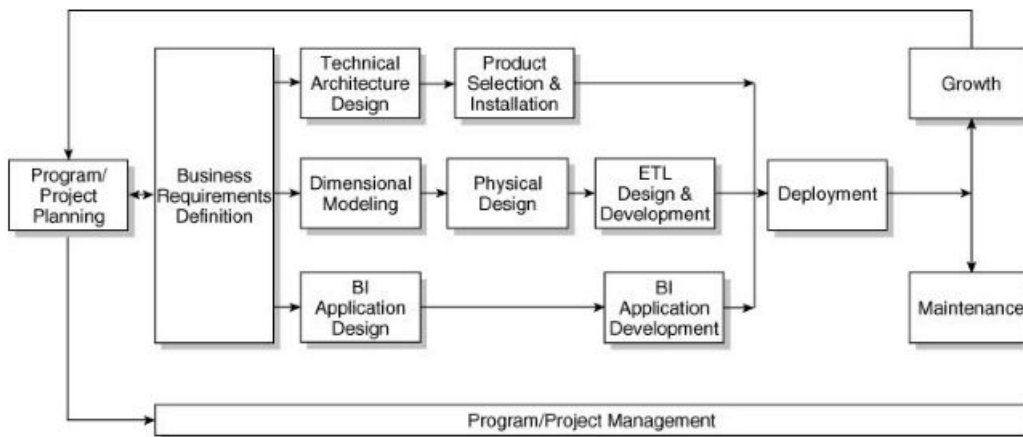


Figura 6.1: Diagrama da metodologia de Kimball [38].

No diagrama da Figura 6.1, pode-se observar que existem três conjuntos de atividades do projeto que ocorrem em simultâneo. Estes conjuntos estão direcionados para a vertente tecnológica, para os dados e para a aplicação de BI. As setas no diagrama indicam o fluxo das atividades dentro de cada um dos percursos paralelos. As dependências entre as tarefas são ilustradas pelo alinhamento vertical das caixas das respetivas tarefas [38]. Estes percursos são:

- Tarefas Tecnológicas, que incluem:
 1. *Projeto da arquitetura tecnológica (Technical Architecture Design)*: é realizado um esquema técnico da arquitetura global.
 2. *Seleção do produto e instalação (Product Selection & Instalation)*: nesta etapa, devem ser seleccionadas e instaladas todas as ferramentas de BI mais apropriadas de maneira a ser possível a construção da solução de BI.
- Tarefas direcionadas para os dados, que incluem:
 1. *Modelação dimensional (Dimensional Modeling)*: O modelo dimensional é uma técnica de projeção do DW, que tem como finalidade fornecer um apoio facilitado aos utilizadores finais das

consultas. O modelo dimensional é constituído pela tabela de fatos e pelas tabelas das dimensões. Os fatos são tipicamente valores numéricos que podem ser agregados, enquanto que as dimensões são grupos de hierarquias e de valores que definem os fatos. Para a projeção do modelo dimensional, podem ser utilizados vários tipos de arquiteturas. As mais comuns são o esquema em estrela e o esquema em floco de neve. O esquema em estrela integra uma única tabela de fatos (o centro da estrela) e múltiplas tabelas de dimensões ligadas à tabela de fatos. O esquema em floco de neve é semelhante ao esquema em estrela, mas apresenta algumas (ou todas) das suas dimensões normalizadas. Assim, este último é uma estrutura irregular, uma vez que as suas dimensões podem ter um número diferente de sub-hierarquias ou dimensões.

2. *Projeção Física (Physical Design)*: Nesta fase é definida a estrutura física dos dados, que envolvem tarefas como a definição do ambiente da base de dados e configuração da segurança apropriada.
3. *Projeção e Desenvolvimento do ETL (ETL Design & Development)*: É a fase mais importante deste processo, que ocupa cerca de 70% do tempo [86] da projeção do DW. Esta fase inclui a extração, a transformação e o carregamento dos dados. A extração consiste em duas etapas, a extração inicial e a extração dos dados alterados. Na primeira etapa, a extração dos dados iniciais ocorre antes do carregamento do DW, enquanto que a segunda etapa é uma atualização contínua do DW, como é o caso da extração contínua dos dados intrínseca às bases de dados do CMIN. A transformação corresponde à limpeza, à integração e otimização dos dados, a fim de se obter dados uniformes, consistentes e precisos. Esta etapa não tem qualquer efeito nos dados originais, apenas nos dados extraídos para o DW. Um exemplo desta etapa é a eliminação de valores nulos ou de valores duplicados. Por último, tem-se a fase do carregamento, em que os dados submetidos

às etapas anteriores são inseridos numa estrutura multidimensional (DW) direcionada para um tipo específico de utilizadores finais e sistemas de aplicações.

- A Aplicação Prática do BI, inclui:
 1. *Projeção da Aplicação de BI (BI Application Design)*: o objetivo é fornecer capacidades para melhorar o apoio ao processo de tomada de decisão no CMIN, que sejam aceites por esta organização. Para tal, é necessário identificar as aplicações de BI candidatas e as interfaces de navegação apropriadas para satisfazer as necessidades dos profissionais do CMIN.
 2. *Desenvolvimento da Aplicação de BI (BI Application Development)*: nesta fase, algumas tarefas são realizadas, como a construção e a validação das aplicações de BI operacionais específicas e do portal de navegação.
- *Implementação (Deployment)*: esta fase inicia o desenvolvimento da aplicação de BI, pelo que é crucial que um planeamento adequado seja realizado, de modo a garantir que os percursos da tecnologia, dos dados e da aplicação de BI sejam testados e funcionem corretamente em conjunto .
- *Manutenção (Maintenance)*: ocorre quando o sistema está em produção. Inclui tarefas operacionais técnicas que são necessárias para manter o sistema com um desempenho ideal (monitorização da utilização, ajuste de desempenho, manutenção do índice, *back-up* do sistema).
- *Expansão (Growth)*: os sistemas de DW têm tendência a expandir, se forem bem sucedidos. Neste caso, deve ser reiniciado novamente o ciclo de Kimball e desenvolver as questões já estabelecidas, focando nos novos requisitos. No caso do CMIN, a plataforma de BI pode ser expandida com novos módulos de interesse na organização para o processo de apoio à tomada de decisão [38, 86].

6.2 Estado da Arte

BI é um termo que foi introduzido por Howard Dresner pertencente ao Gartner Group, em 1989. Com este termo, Dresner pretendia descrever um conjunto de conceitos e métodos utilizados para melhorar o processo de tomada de decisão, através de um sistema de apoio computadorizado, baseado em fatos [87].

A primeira definição científica foi estabelecida por Ghoshal e Kim [88] e associou o BI a uma filosofia de gestão e a uma ferramenta capaz de ajudar as organizações a gerirem e a aperfeiçoarem a informação empresarial com o objetivo de tomarem decisões mais eficientes.

Mais tarde, Eckerson [89] compreendeu que o BI deveria ser capaz de fornecer ferramentas como relatórios de produção, consultas do utilizador final, análises OLAP, *dashboards* e ferramentas de planeamento e modelação.

Posteriormente, Lonnqvist e Pirttimaki [90] constataram que o termo BI pode ser utilizado para se referir a dois conceitos distintos. O primeiro está associado à informação e ao conhecimento relacionado com uma organização, que descreva o ambiente empresarial em que se encontra, as condições do mercado, os clientes, os concorrentes e ainda questões económicas. O segundo conceito engloba processos sistemáticos através dos quais a organização obtém, analisa e distribui informação para a tomada de decisões sobre operações empresariais.

A revisão literária em torno do tema BI mostra uma divisão dos pontos de vista técnico e administrativo. A abordagem administrativa defende que o BI é um processo onde os dados reunidos a partir do interior e do exterior da organização são integrados de modo a gerar informação relevante para o processo de tomada de decisão. A abordagem técnica considera o BI um conjunto de ferramentas que apoiam esse mesmo processo. Neste caso, o foco não é no processo em si, mas nas tecnologias, nos algoritmos e nas ferramentas que permitem o armazenamento, a manipulação e a análise dos dados e da informação [91].

Bose [92] descreveu também uma abordagem administrativa do BI como sendo o processo de obtenção da informação correta para os utilizadores

apropriados, no tempo certo, de forma a que estes pudessem tomar decisões que melhorassem o desempenho da organização.

Outra definição é a de Ranjan [93], que considera o BI como sendo a transformação consciente dos dados provenientes de qualquer fonte de dados em novas formas de conhecimento, direcionado para a organização e orientado para os resultados. Esta definição engloba uma variedade de ferramentas, bases de dados e fornecedores, a fim de originarem uma infraestrutura que não incorpora apenas uma solução inicial, mas também tem a capacidade de mudança de acordo com as alterações verificadas no mercado.

Wu et al. [94] definiram BI como um termo de gestão de negócios, utilizado para descrever aplicações e tecnologias que são utilizadas para recolher, dar acesso e analisar dados e informação relacionada com a organização para apoiar o processo de tomada de decisão.

Elbashir et al. [95] referem as plataformas de BI como um grupo importante de sistemas utilizados para a análises de dados, que apoiam os gestores em diferentes níveis da organização, de forma relevante e oportuna, permitindo-lhes tomar decisões mais acertadas.

Recentemente, Jalonen e Lonnqvist [96] escreveram que o BI gera análises e relatórios de tendências no ambiente organizacional, explicando ainda que as análises podem ser geradas sistematicamente e regularmente ou então consultadas em contextos específicos.

De todas estas definições podem ser deduzidas duas importantes proposições:

- A maioria das abordagens ao BI são limitadas por ferramentas de apoio, sistemas ou tipo de sistemas.
- O BI tem como principal objetivo fornecer informação analítica relevante para a gestão de uma organização, apoiando a administração das suas atividades.

6.3 Objetivos

Com o aumento dos registos eletrónicos e com o aumento da pressão para a tomada de decisões mais rápidas e eficazes, que muitas vezes são responsáveis pela ocorrência de erros médicos, tornou-se oportuna a construção de um sistema de apoio ao processo de tomada de decisão específico para a unidade de IVG, no CMIN, recorrendo à tecnologia de BI.

Por outro lado, são realizados relatórios anuais por uma equipa de enfermagem, que contém indicadores relacionados com a informação registada durante os procedimentos de IVG. O trabalho que era realizado pelas enfermeiras para a criação dos indicadores deixa de ser necessário com o recurso às tecnologias de BI, uma vez que todo o processo passa a ser automatizado.

Portanto, o principal objetivo deste capítulo é a construção de uma plataforma de BI capaz de auxiliar os profissionais de saúde do CMIN no processo de decisão no setor da IVG [97].

6.4 Resultados

Antes de se proceder ao processo realizado para a implementação da solução de BI proposta, é de salientar que esta é bastante vantajosa para a instituição de saúde em estudo. Alguns desses benefícios são:

- Organização da informação dispersa, extraindo apenas aquela que poderá ser relevante para a organização, e a construção de cenários que apoiem os profissionais de saúde na tomada de decisão;
- Maior autonomia e flexibilidade dos utilizadores no que diz respeito à criação de relatórios;
- Análises dos dados clínicos mais rápidas e simples;
- Maior apoio no processo de tomada de decisão;
- Reações positivas das utentes devido à melhoria nos cuidados de saúde prestados;

- A partir dos gráficos, por exemplo, os fluxos de trabalho podem ser monitorizados, as dificuldades podem ser identificadas e os processos melhorados;
- A comparação dos resultados obtidos a partir de decisões realizadas no passado com resultados presentes, a fim de se registrar a evolução;
- É também possível uma maior proteção dos dados da utente, restringindo o acesso a apenas utilizadores com níveis de acesso apropriados.
- Melhor gestão dos dados, o que leva a uma melhor assistência da utente em termos do tempo de admissão, do diagnóstico e do risco de acidentes.

Como se pode constatar, é elevado o número de vantagens que a organização em causa, o CMIN, pode usufruir com a implementação desta plataforma de BI.

Assim, seguindo a metodologia de Kimball, descrita na secção 6.1.2, foi projetada a plataforma de BI para a unidade de IVG no CMIN.

6.4.1 Planeamento do Projeto

Como já foi referido, a ideia de implementação desta solução de BI surgiu como uma hipótese de apoio ao trabalho realizado por uma equipa de enfermagem do CMIN. Estas profissionais são responsáveis pela elaboração de relatórios anuais com indicadores de desempenho da unidade de IVG, com a finalidade de serem entregues à Direção Geral de Saúde. A questão é que todos os indicadores eram obtidos manualmente, o que pode ser um processo moroso e trabalhoso. O principal objetivo deste projeto era portanto a automatização do processo de obtenção de indicadores, recorrendo aos dados que eram registados e armazenados nas bases de dados, associados à unidade de IVG.

Assim, nesta etapa, foram realizadas tarefas como: pesquisas de indicadores importantes para a unidade de IVG do CMIN, um planeamento inicial de todas as etapas da abordagem de Kimball e um estabelecimento de prazos para essas mesmas etapas definidas.

6.4.2 Definição dos Requisitos da Organização

No CMIN, os registos das utentes são armazenados na *Agência de Integração, Difusão e Armazenamento (AIDA)*, que suporta o *Processo Clínico Eletrónico*. No caso particular da unidade de *IVG*, os dados são também armazenados no *Processo Clínico Eletrónico*, porém a informação armazenada nas bases de dados raramente é utilizada para extração do conhecimento. Assim, quando há a necessidade de extrair conhecimento a partir dos dados, têm que ser aplicadas técnicas apropriadas, que requerem bastante tempo e que podem ser bastante caras. Nestes casos, o desenvolvimento de indicadores clínicos é realizado através de ferramentas rudimentares, como as folhas de cálculo para a obtenção de gráficos ou relatórios, método que tem sido utilizado nos últimos anos na unidade de *IVG*. A utilização destas técnicas geralmente não é apropriada, pois não apresentam a qualidade que os dados clínicos requerem, o que condiciona diretamente um bom processo de tomada de decisão.

Assim, uma das principais motivações para o desenvolvimento desta plataforma de *BI* é dar usabilidade à grande quantidade de dados registados, permitindo a criação de conhecimento e possibilitando a otimização e automatização de todo o processo de obtenção de indicadores, para que esse conhecimento possa estar disponível no momento de decisão.

Nesta fase, foi então realizado um planeamento/pesquisa de todos os indicadores necessários na unidade de *IVG* com o auxílio dos profissionais de saúde e dos profissionais das *Tecnologias de Informação (TI)* do CMIN, a fim de se obter uma ideia mais precisa dos requisitos finais da plataforma de *BI*.

6.4.3 Seleção do Produto e Instalação

Nesta fase, optou-se por algumas ferramentas, que depois de instaladas, foram também testadas, nomeadamente:

- **Oracle Database**, como aplicação para a base de dados;
- **Oracle SQL Developer**, como ferramenta para extração dos dados e para a implementação de procedimentos de *ETL*;

- **Pentaho BI Suite**, como ferramenta para análises OLAP e para a criação das *dashboards*.

6.4.4 Modelação Dimensional

Na estruturação do modelo dimensional implementado no DW, foi utilizado o esquema em estrela, por ser o que mais se adequava à estrutura pretendida para a disposição dos dados. Além disso, no esquema em estrela, apesar de a redundância ser superior, a complexidade é bastante inferior, sendo por isso, mais vantajosa a sua utilização com as ferramentas de BI. Neste esquema, existe uma única tabela de factos, onde estão representados os atributos que estão associados aos indicadores a serem analisados e os atributos que fazem a ligação às tabelas das dimensões (através de conexões do tipo chave primária/chave estrangeira). Na Figura 6.2 está representada o modelo multidimensional projetado para o DW implementado no CMIN.

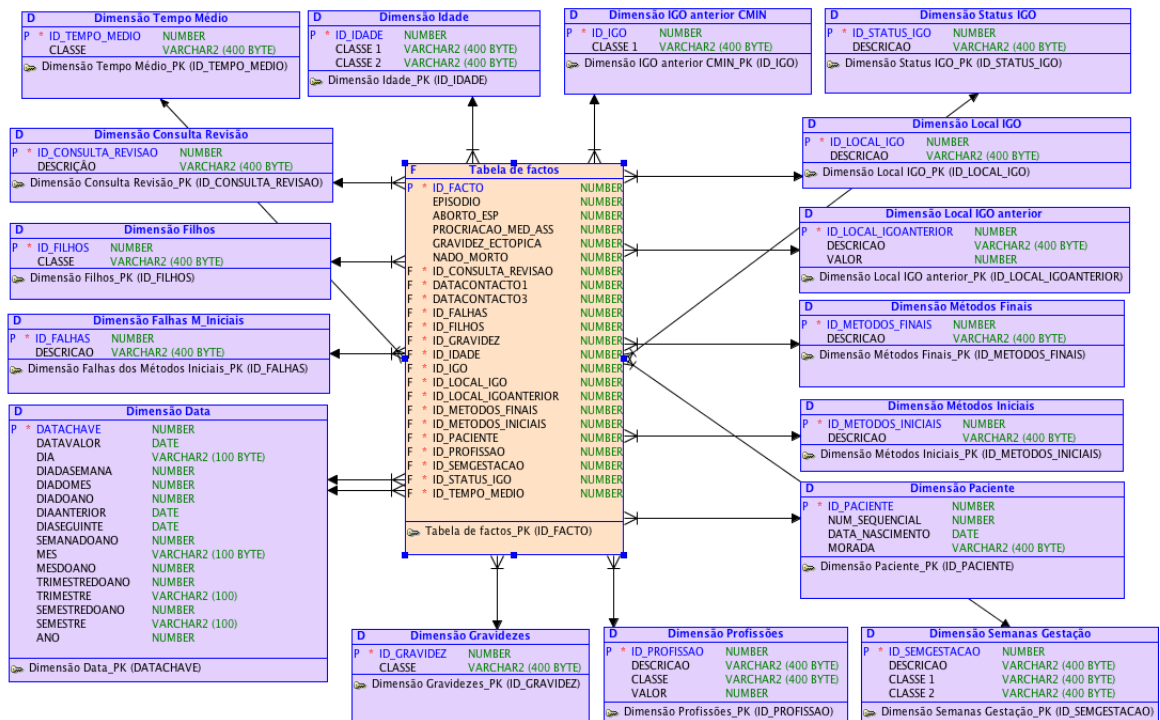


Figura 6.2: Modelo multidimensional do DW.

Como se pode observar na Figura 6.2, na tabela de factos estão presentes atributos que vão dar origem a um determinado indicador, como é o caso do atributo *ABORTO_ESP* responsável pelo indicador que caracteriza o grupo utentes com base no número de abortos espontâneos existentes anteriormente, e atributos que fazem a ligação às tabelas das dimensões. Neste último caso, tem-se o exemplo do atributo *ID_GRAVIDEZ* que faz ligação à *Dimensão Gravidezes* e é através desta ligação que é possível caracterizar o grupo de utentes com base no número de gestações anteriores e ainda agrupá-las em classes.

6.4.5 Projeção Física

Nesta etapa, recorreu-se ao Oracle Database para se criarem as ligações às bases de dados do *CMIN* e para a criação da estrutura apresentada na Figura 6.2. Assim, foram criadas todas as tabelas, quer as correspondentes às dimensões, quer à tabela de fatos, e foram feitas as ligações que estabelecem as relações entre elas.

6.4.6 Projeção e Desenvolvimento do ETL

Esta etapa inicia-se com a seleção da informação necessária para a obtenção dos indicadores, informação essa presente nas tabelas das bases de dados: *Processo Clínico Eletrónico*, *SAPE* e *SONHO*. Como se pode verificar, os dados utilizados são armazenados em bases de dados diferentes que estão interligadas pela *AIDA*.

Dependendo das variáveis recolhidas e dos alvos definidos, os dados foram automaticamente validados e preparados de acordo com regras de transformação e de processamento definidas. Posteriormente, os dados foram carregados para o *DW*, no modelo multidimensional criado, através do esquema em estrela. A este processo designa-se de *ETL*. Para a realização deste processo, foram implementados procedimentos e funções através do SQL Developer.

Nos procedimentos de *ETL*, alguns dos objetivos a atingir foram a eliminação dos valores nulos, a eliminação dos dados repetidos e a transformação da informação. Além disso, foram também realizados procedimentos para

uniformizar a informação, como por exemplo, a correção de erros ortográficos, a substituição da vírgula pelo ponto nos valores com casas decimais, entre outros. Alguns exemplos concretos são, por exemplo:

- No caso dos métodos contraceptivos, foi atribuída uma única designação para cada um dos diferentes métodos, tendo sido transformados todos os dados que apresentassem letras minúsculas ou letras maiúsculas numa ordem diferente da definida e, ainda, todos os termos que tivessem erros de ortografia;
- No caso das profissões das utentes, foram selecionados todos os dados que estivessem na fonte de dados, associados a uma *label* igual a "Profissão".
- Relativamente ao número de interrupções realizadas anteriormente no **CMIN**, foi contabilizado o número de vezes que a expressão "Oportunidade de abortar" estava associada ao número sequencial de cada uma das utentes, nos registos utilizados.
- A idade das utentes foi calculada, recorrendo à data de nascimento das mesmas e à data de admissão no **CMIN**;
- Entre outros.

Para a fase de carregamento foram também criados procedimentos para carregar/atualizar a informação do **DW**. Estes procedimentos são pré-definidos e programados de forma a automatizar o processo.

Após o carregamento do **DW**, seguiu-se o desenvolvimento do cubo **OLAP**. O cubo **OLAP** consiste numa metodologia que oferece respostas rápidas a consultas *ad-hoc* às dimensões e às métricas armazenadas no **DW**. Isto é, pode-se afirmar que uma base de dados **OLAP** tira uma "foto" de um subconjunto do **DW**, reestrutura esses dados e armazena-os num cubo **OLAP**. As consultas *ad-hoc* são executadas nos resultados armazenados nesses cubos. Além disso, as estruturas dos cubos **OLAP** são altamente otimizadas para a realização das consultas *ad-hoc*, através do armazenamento de resultados pré-computorizados para diferentes combinações de dimensões e de métricas.

6.4.7 Projeção da Aplicação de BI

Nesta etapa, foram analisadas várias ferramentas de BI, tendo em consideração os principais requisitos de um ambiente na área clínica. Esta análise está detalhada no capítulo 5.

Em resultado desta análise, conclui-se que a ferramenta mais apropriada para o caso em estudo era o Pentaho BI Suite.

Assim, até à presente etapa, o processo pode ser resumido pela ocorrência, numa primeira fase, da análise dos dados presentes nas tabelas do SAPE, Processo Clínico Eletrónico e SONHO, existentes na AIDA. Posteriormente, estes dados foram submetidos ao processo de ETL, onde foi realizada a extração dos dados da base de dados original, um tratamento desses mesmos dados para que estes passassem a ser uniformes e, por último, deu-se o carregamento para o DW. A estrutura do DW adotada neste caso foi a do esquema em estrela, por ser a mais adequada para os dados armazenados. Em seguida, ocorreu a construção dos cubos OLAP, terminando com a obtenção e a visualização da informação na ferramenta *open source* Pentaho BI Suite. Este processo está representado na Figura 6.3.

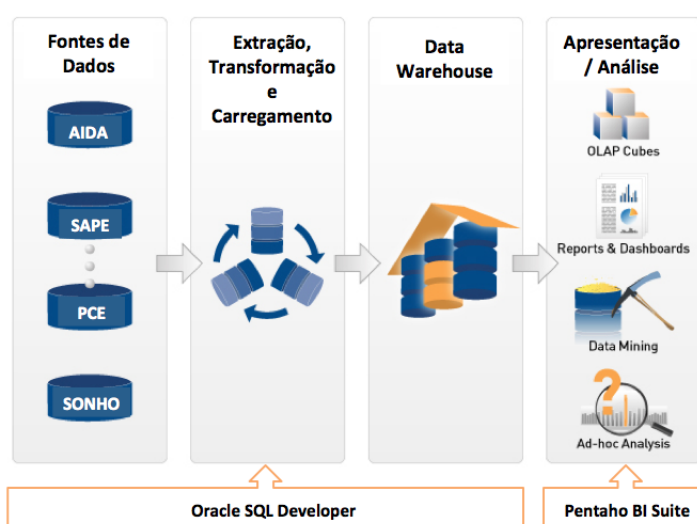


Figura 6.3: Arquitetura da plataforma de BI.

6.4.8 Desenvolvimento da Aplicação BI

Nesta fase, recorreu-se então ao Pentaho BI Suite para a construção dos cubos OLAP e para a visualização dos indicadores de desempenho requisitados pelo CMIN.

Devido a questões de confidencialidade do projeto, apenas serão apresentados alguns dos indicadores desenvolvidos. Assim, alguns dos indicadores selecionados com o apoio dos profissionais de saúde juntamente com os profissionais das TI para serem implementados na plataforma de BI foram os seguintes:

- *Caracterização do grupo de utentes por antecedentes obstétricos e por data:* permite identificar o número e a percentagem de utentes que apresentam antecedentes obstétricos, como a procriação médica assistida, a gravidez ectópica, o aborto espontâneo ou nado morto, num período de tempo selecionado. Este indicador foi apresentado na forma de gráficos circulares, como se pode observar na Figura 6.4



Figura 6.4: Caracterização do grupo de utentes por antecedentes obstétricos e por data.

- *Caracterização do grupo de utentes por número de gestações e por data:*

identifica o número e a percentagem de utentes por número de gestações tidas anteriormente, num período de tempo selecionado. Este indicador está representado sob a forma de um gráfico linear e de uma tabela de percentagens, como se pode observar na Figura 6.5.

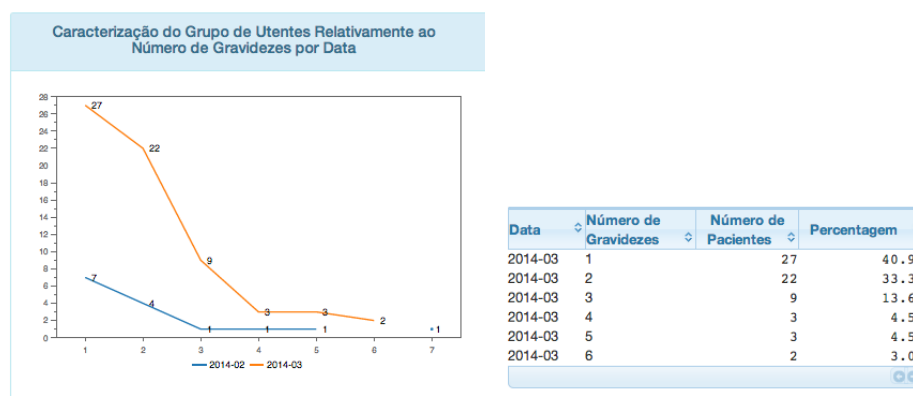


Figura 6.5: Caracterização do grupo de utentes por número de gestações e por data.

- *Caracterização do grupo de utentes por número de filhos e por data:* identifica o número e a percentagem de utentes por número de filhos que apresentam, num período de tempo selecionado. Este indicador está também representado sob a forma de um gráfico linear e de uma tabela de percentagens.
- *Caracterização do grupo de utentes por número de experiências de IVG anteriores e por data:* identifica o número e a percentagem de utentes que apresentam experiências anteriores de IVG, num período de tempo selecionado. Este indicador está representado sob a forma de um gráfico de barras e de uma tabela de percentagens, como se pode observar na Figura 6.6. No gráfico de barras, o número de utentes associado a 0 experiências de IVG anteriores toma um valor negativo, devido à falta de relevância deste valor e de modo a salientar mais os restantes.

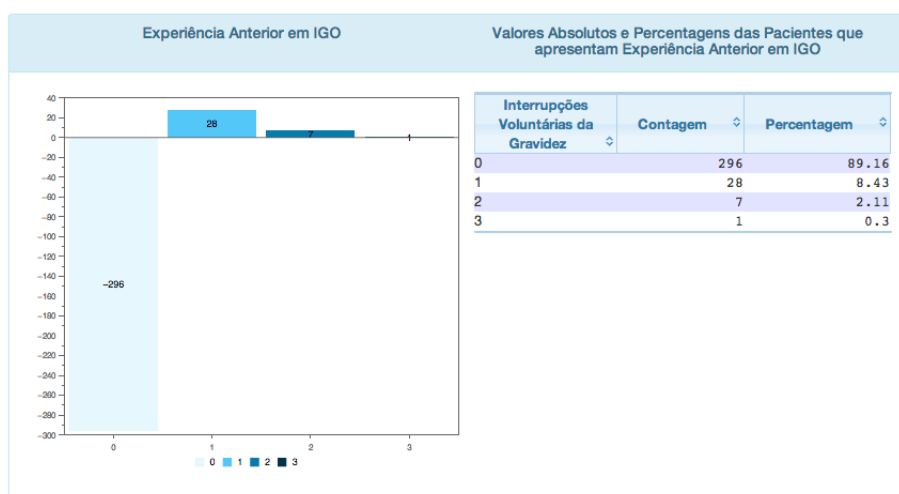


Figura 6.6: Caracterização do grupo de utentes por número de experiências de IVG anteriores e por data.

- *Caracterização do grupo de utentes relativamente à consulta de revisão por data*: apresenta o número e a percentagem de utentes que estiveram presentes na consulta de revisão, correspondente à consulta efetuada no final do processo de IVG, durante um determinado intervalo de tempo. Este indicador está representado sob a forma de um gráfico circular.
- *Caracterização do grupo de utentes com base no resultado final do processo de IVG e por data*: identifica o número e a percentagem de utentes pelos três tipos de resultados possíveis no final do processo de IVG (aborto conseguido, aborto incompleto e aborto não-conseguido), num período de tempo selecionado. Este indicador está também representado sob a forma de um gráfico circular.
- *Caracterização do grupo de utentes relativamente ao número de falhas do processo de IVG, por ano e por mês*: identifica o número de utentes cujo o processo de IVG não foi conseguido com sucesso. Este indicador está representado num gráfico de barras, onde o ano pode ser selecionado através de um *picker*, como se pode observar na Figura 6.7

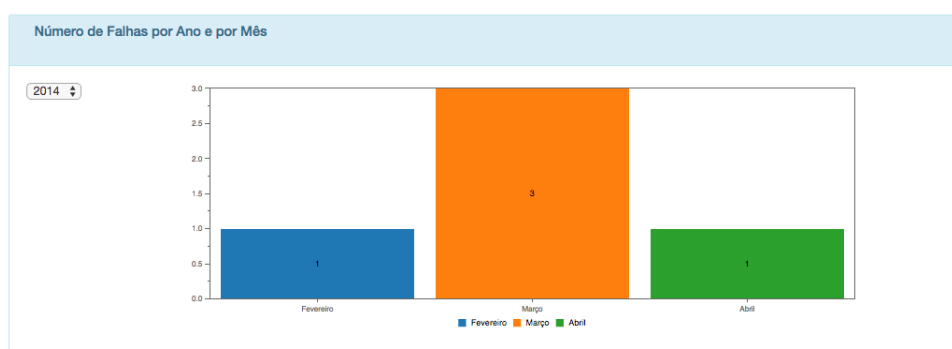


Figura 6.7: Caracterização do grupo de utentes relativamente ao número de falhas do processo de IVG, por ano e por mês.

- *Caracterização do grupo de utentes com base na contraceção no início do processo por data:* identifica o número de utentes que utilizavam métodos contraceptivos no início do processo, quais os métodos contraceptivos e quais as falhas associadas a cada um deles, num determinado período de tempo. Para tal, utilizou-se um gráfico circular e dois gráficos de barras, como se pode observar na Figura 6.8

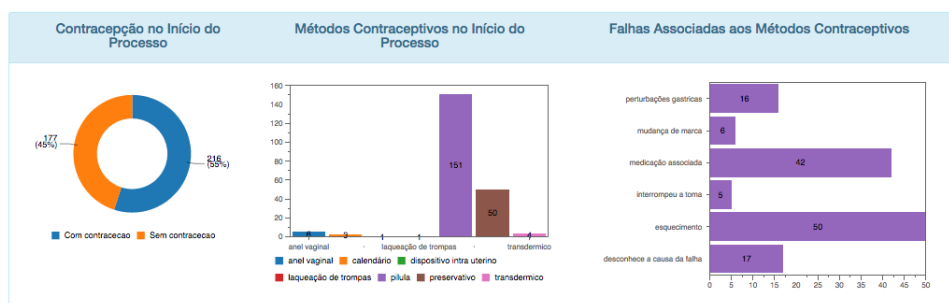


Figura 6.8: Caracterização do grupo de utentes com base na contraceção no início do processo por data.

- *Caracterização do grupo de utentes com base na contraceção no final do processo de IVG por data:* identifica o número de utentes que adotaram métodos contraceptivos no final do processo e quais os métodos

contracetivos escolhidos, num determinado período de tempo. Para tal, utilizou-se um gráfico circular e um gráficos de barras.

- *Caracterização do grupo de utentes com base na idade gestacional no momento da toma do mifepristone, por ano, mês e dia:* identifica o número de utentes para cada uma das possíveis idades gestacionais, por ano, mês e dia. Para tal, foi criado um cubo OLAP que permite fazer o *drill-down* e o *roll-up* na data, facultando a visualização dos meses e dos dias para um determinado ano, como se pode observar na Figura 6.9.

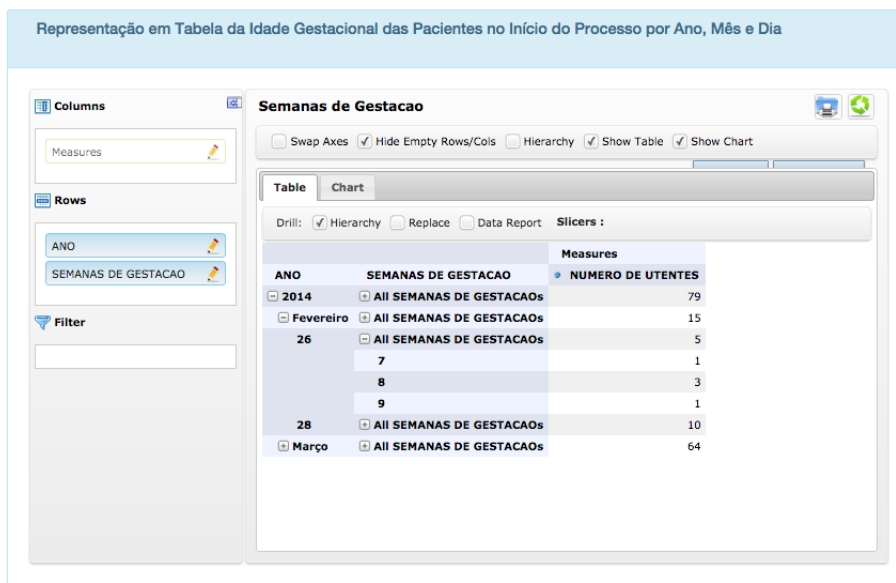


Figura 6.9: Caracterização do grupo de utentes com base na idade gestacional no momento da toma do *mifepristone*, por ano, mês e dia.

- *Caracterização do grupo de utentes com base no local de consecução do processo, por data:* identifica o número de utentes, tendo por base o local de consecução do processo (domicílio ou CMIN), por data. Este indicador está representado sob a forma de um gráfico de barras.
- *Caracterização do grupo de utentes com base no tempo médio entre a consulta inicial e a consulta final do processo de IVG, por ano:* identi-

fica o número de utentes, tendo por base o tempo médio entre a consulta inicial e a consulta final do processo de IVG. Foi também criado um outro indicador que apresenta algumas medidas estatísticas dos tempos entre consultas, por ano, como se pode observar na Figura 6.10.



Ano	Máximo	Mínimo	Média
2010	1342	153	864
2011	840	3	28
2012	614	3	35
2013	227	3	39
2014	89	2	24

Figura 6.10: Medidas estatísticas dos tempos médios entre a consulta inicial e a consulta final do processo de IVG, por ano.

- *Caracterização do grupo de utentes com base na localidade, por ano:* identifica o número de utentes, tendo por base a localidade que foi registada no momento da admissão no CMIN. Este indicador está representado sob a forma de gráficos circulares interativos, onde estão registadas as 10 localidades mais frequentes, por ano, como se pode observar na Figura 6.11.
- *Caracterização do grupo de utentes com base no local onde foi realizada a última experiência de IVG:* identifica o número de utentes que realizaram o último procedimento de IVG no CMIN ou no exterior.
- *Caracterização do grupo de utentes por profissão e por data:* identifica o número de utentes que são desempregadas, estudantes e empregadas, apresentando também as respetivas profissões, num determinado período de tempo. Este indicador está representado sob a forma de um gráfico circular e dois gráficos de barras, como se pode observar na Figura 6.12.

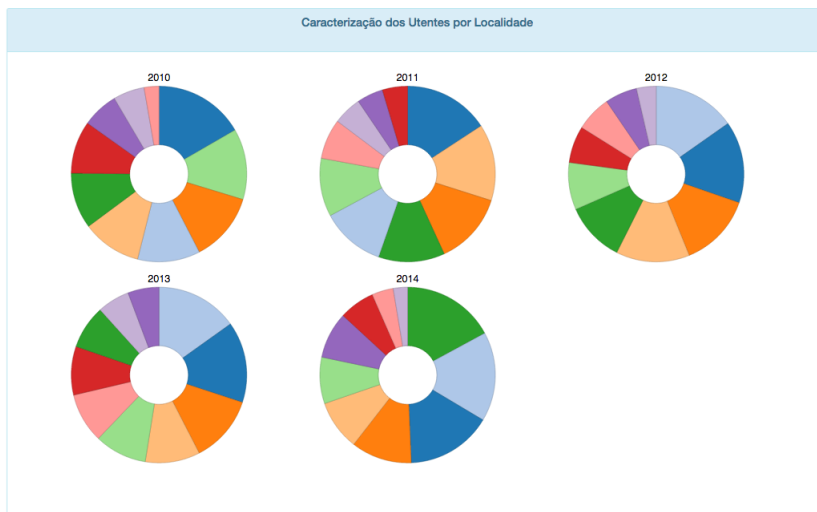


Figura 6.11: Caracterização do grupo de utentes com base na localidade, por ano.

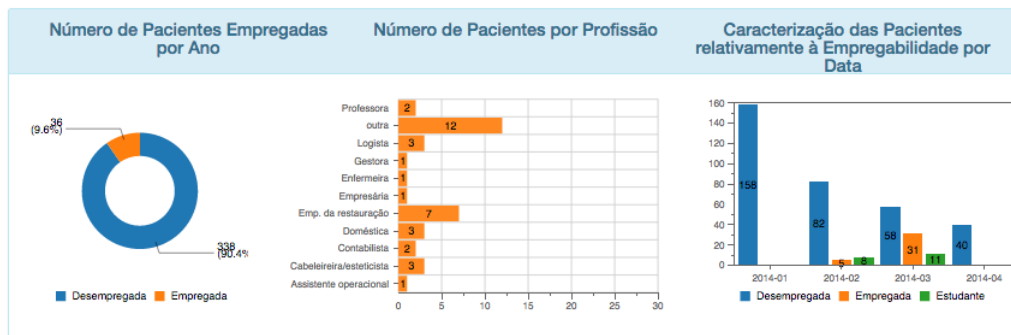


Figura 6.12: Caracterização do grupo de utentes por profissão e por data.

- *Caracterização do grupo de utentes por idades e por ano*: identifica o número de utentes por classes de idades, num determinado ano. Este indicador está representado sob a forma de um gráfico de linhas, como se pode observar na Figura 6.13.

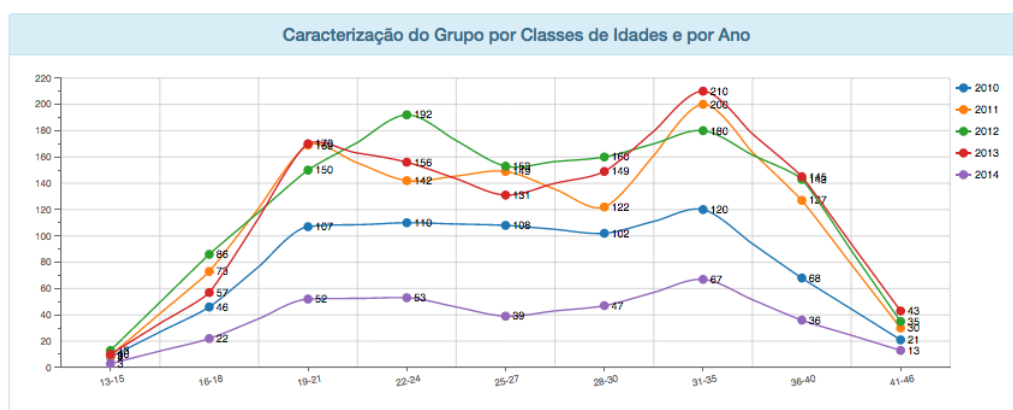


Figura 6.13: Caracterização do grupo de utentes por idades e por ano.

6.4.9 Plataforma de BI

A plataforma de BI é baseada na *Web*, que pode ser acedida a qualquer momento e em qualquer lugar, de acordo com os privilégios de acesso.

Nesta aplicação, estão representados todos os indicadores pretendidos, apresentados na secção anterior, e também outros indicadores que podem ser criados pelos utilizadores, através da tecnologia OLAP.

A tecnologia OLAP permite que os utilizadores finais façam consultas *ad-hoc* nos dados com múltiplas dimensões, fornecendo assim a informação e o conhecimento necessários para uma melhor tomada de decisão. Quanto maior o número de medidas analisadas, maior o número de perspectivas existentes para compreender e analisar a informação disponível. Com esta ferramenta podem ser realizadas várias operações, em particular, o *roll-up*, o *drill-down*, o *slice e dice* e o *pivot*.

A operação *roll-up* consiste na soma dos dados que constituem o cubo, o que resulta na ascensão a um nível superior da hierarquia de uma determinada medida ou na eliminação de uma dimensão do cubo em análise.

O *drill-down* permite a navegação em profundidade nos dados agregados a fim de permitir uma análise de dados mais detalhada. É a operação inversa do *roll-up*.

O *slice* consiste na seleção e a definição de um valor para um atributo de uma dimensão, permitindo a sua análise em relação a outras dimensões, enquanto que o *dice* permite definir um subcubo, selecionando atributos em duas ou mais dimensões.

Por último, a operação *pivot* permite a rotação dos dados em torno de eixos, de modo a oferecer uma representação alternativa do mesmo.

Nos indicadores referenciados na secção anterior, alguns deles foram representados também através de cubos OLAP, a fim de permitir que os dados pudessem ser analisados com mais pormenor a nível das datas, sendo possível a sua visualização ao nível do ano, dos meses e dos dias. Alguns desses indicadores foram, por exemplo:

- Caraterização do grupo de utentes por número de gestações e por ano, mês e dias;
- Caraterização do grupo de utentes por número de filhos e por ano, mês e dias;
- Caraterização do grupo de utentes por presença ou ausência de contração no início do processo de IVG e por ano, mês e dias;
- Caraterização do grupo de utentes por sucesso ou insucesso do procedimentos de IVG e por ano, mês e dias;
- Caraterização do grupo de utentes por situação profissional (empregada, desempregada ou estudante) e por ano, mês e dias;

Além disso, a plataforma de BI no CMIN incorpora também os modelos de previsão de DM, abordados no capítulo 4, para prever o local onde é administrada a segunda dose dos medicamentos no processo de IVG e para prever quais as utentes que tem propensão para pertencer ao grupo de risco.

6.5 Discussão dos Resultados

Existem muitos benefícios em integrar a tecnologia BI no CMIN, sendo os principais a autonomia e a flexibilidade que os utilizadores finais adqui-

rem, no que diz respeito à criação de relatórios. Além disso, permite que sejam realizadas análises aos dados de forma simples e rápida, resultando numa melhoria do processo de tomada de decisão e numa maior eficiência operacional, bem como uma maior variedade de novas funções analíticas.

Ao nível científico esta plataforma traz muitas vantagens, na medida em que se trata de um projeto pioneiro na área de **IVG** no **CMIN** e a nível mundial, caracterizada pela obtenção de indicadores específicos. Tendo em consideração os resultados obtidos e acreditação dada pelos profissionais de saúde, esta plataforma pode fornecer um ponto de partida para a implementação de medidas de melhoria no módulo de **IVG**.

Além disso, pode ainda ser útil para a investigação e o desenvolvimento de novos módulos de interesse e não apenas para o módulo de **IVG**, bem como para outras instituições de saúde. Um exemplo de outro módulo de interesse poderá ser o caso das gravidezes não evolutivas, uma vez que requerem apenas uma pequena adaptação do sistema atual a algumas novas condicionantes desse procedimento.

A implementação da plataforma de **BI** na unidade de **IVG** é muito vantajosa, na medida em que permite, por exemplo, a interpretação dos registos dos dados existentes, a obtenção de indicadores que traduzam o funcionamento global desta unidade, a aquisição de estratégias de melhoria contínua nos cuidados prestados com base nos resultados obtidos, e a divulgação da atividade desenvolvida nesta área.

O principal objetivo desta plataforma de **BI** é apoiar a prestação de melhores cuidados de saúde e ir ao encontro das necessidades das mulheres grávidas que pretendem se submeter ao processo de **IVG**. Com esta plataforma pretende-se prestar cuidados personalizados à utente, possibilitando um aumento das práticas clínicas baseadas em evidências.

6.6 Conclusão

Com a concretização deste projeto foi possível explorar a importância de integrar a tecnologia **BI** no **CMIN**. A implementação desta tecnologia garante uma análise dos dados extraídos, apesar da maioria dos dados clínicos serem

pouco estruturados.

A criação desta plataforma de BI envolveu a construção de um DW para a unidade de IVG, essencial para a obtenção dos indicadores pretendidos. Assim, a partir desta plataforma, podem ser gerados indicadores de IVG, como por exemplo a caracterização das utentes por idade ou por profissão. Os indicadores mencionados neste capítulo foram considerados pelos clínicos como os mais significativos, no entanto, podem ser adicionados outros indicadores à plataforma sempre que necessário. Além disso, esta plataforma integra também modelos de previsão de DM de modo a permitir a extração de conhecimento a partir dos dados resultantes dos registos eletrónicos e a oferecer um apoio em tempo real aos médicos e profissionais de saúde no processo de tomada de decisão.

A implementação da plataforma de BI no CMIN é uma ideia inovadora, na medida que é a primeira vez que as tecnologias de BI são aplicadas na área de IVG.

No futuro, pretende-se que seja realizado uma prestação de cuidados personalizados às utentes, através da criação de novos indicadores e de uma maior exploração das capacidades dos modelos de DM a serem desenvolvidos no CMIN. Pretende-se ainda uma melhoria gradual da plataforma, respondendo às necessidades dos profissionais de saúde do CMIN. Futuramente, a plataforma poderá também ser expandida para outras unidades de interesse do CMIN, bem como a outras organizações materno-infantil.

Capítulo 7

Conclusão

Neste capítulo, são realizadas algumas considerações finais que foram obtidas ao longo desta dissertação e ainda algumas sugestões futuras para projetos subsequentes.

7.1 Considerações Finais

Com a realização deste projeto, pode-se afirmar que todos os objetivos definidos inicialmente foram totalmente cumpridos.

Relativamente aos modelos de [Data Mining \(DM\)](#) desenvolvidos, quer para a previsão do local onde é administrada a segunda dose de medicamentos do processo de [Interrupção Voluntária da Gravidez \(IVG\)](#), quer para a previsão do grupo de risco das mulheres que recorrem à unidade de [IVG](#), foram obtidos bons resultados, permitindo a obtenção de conclusões significativas. Para o primeiro caso, verificou-se que as características que maior influência têm nesta questão são o número de gestações e o número de partos que a utente já teve. Neste caso, obtiveram-se valores de sensibilidade e acuidade de, aproximadamente, 91% e 87%. Relativamente ao segundo problema de [DM](#), é a idade, o número de [IVGs](#) anteriores, o número de gestações e o número de partos que maior influência têm na probabilidade da utente pertencer ao grupo de risco. Obteve-se um valor máximo de aproximadamente 93% no modelo 4, em termos da métrica de sensibilidade.

Assim, com esta parte do trabalho, demonstra-se que é possível o desenvolvimento de modelos de **DM** eficazes na área da saúde, que possibilitem aos profissionais de saúde ter uma perspectiva *a priori* das possíveis situações avaliadas.

Quanto à análise comparativa de diferentes ferramentas de **BI**, verificou-se que, com base nas características identificadas como as mais necessárias na área clínica, a ferramenta mais adequada para utilizar neste projeto foi o Pentaho BI Suite. Esta foi, portanto, a ferramenta que foi usada no processo de desenvolvimento da plataforma de **BI**. É de referir que esta opção foi escolhida com base nas condições que o **CMIN** dispõe e nas ponderações indicadas pelos profissionais, pelo que poderão existir ferramentas mais adequadas para outras organizações.

Por último, pode-se ainda dizer que foi concretizada a ideia de implementar uma plataforma de **BI** na unidade de **IVG** do **CMIN**, pelo que esta disponibiliza todos os indicadores que eram pretendidos pelos profissionais de saúde, tendo sido bastante positivo o *feedback* por eles apresentado, na altura em que a plataforma foi colocada em fase de testes.

7.2 Trabalho Futuro

Futuramente, existem alguns aspetos que poderão ser melhorados. Além disso, este é um projeto ao qual se pode dar continuação, pois exige uma manutenção contínua, devido às novas circunstâncias que vão surgindo.

Assim, relativamente aos modelos de **DM**, podem ser exploradas novas técnicas, podem ser selecionadas novas variáveis e ainda se podem repetir as experiências com novos conjuntos de dados.

Quanto à plataforma de **BI** pretende-se, no futuro, integrá-la no sistema de **BI** já existente no **CMIN** para outras unidades da instituição. Além disso, poderão ainda ser explorados outros indicadores de interesse na unidade de **IVG** ou ainda, direcionar esta plataforma de **BI** para outras unidades ainda não abordadas.

Bibliografia

- [1] W. Bonney, “Applicability of Business Intelligence in Electronic Health Record,” *Procedia - Social and Behavioral Sciences*, vol. 73, pp. 257–262, Feb. 2013.
- [2] L. F. B. Pinto, “Sistemas de informação e profissionais de enfermagem,” Ph.D. dissertation, Universidade de Trás-os-Montes e Alto Douro, 2009.
- [3] A. Abelha, C. Analide, J. Machado, J. Neves, M. Santos, and P. Novais, “Ambient intelligence and simulation in health care virtual scenarios,” in *Establishing the Foundation of Collaborative Networks*, L. M. Camarinha-Matos, H. Afsarmanesh, P. Novais, and C. Analide, Eds. Boston, MA: Springer US, 2007, pp. 461–468.
- [4] S. Khodambashi, “Business process re-engineering application in healthcare in a relation to health information systems,” *Procedia Technology*, vol. 9, pp. 949–957, Jan. 2013.
- [5] F. Portela, M. Vilas-Boas, M. F. Santos, A. Abelha, J. Machado, A. Cabral, and I. Aragao, “Electronic health records in the emergency room,” in *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*. IEEE, Aug. 2010, pp. 195–200.
- [6] E. Turban, J. E. Aronson, and T.-P. Liang, “Decision support systems and intelligent systems (7th edition),” *Prentice-Hall, Inc.*, Apr. 2004.
- [7] P. G. Keen, “Decision support systems: The next decade,” *Decision Support Systems*, vol. 3, no. 3, pp. 253–265, Sep. 1987.
- [8] S. Negash and P. Gray, “Business intelligence,” in *AMCIS 2003 Proceedings*. AIS Electronic Library, 2003.
- [9] S. Negash, “Business intelligence,” *Communications of the Association for Information Systems*, vol. 13, no. 1, 2004.
- [10] T. Mettler and V. Vimarlund, “Understanding business intelligence in the context of healthcare.” *Health informatics journal*, vol. 15, no. 3, pp. 254–64, Sep. 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19713399>

- [11] M. Viana, O. Oliveira, A. Abelha, and J. Machado, "Step towards Simulation and Monitoring of Hospital Waiting Lists," in *Proceedings of the ESM*, Essen, Germany, 2012. [Online]. Available: <http://pt.linkedin.com/in/jmmachado>
- [12] A. Cabral, A. Abelha, M. Salazar, C. Quintas, F. Portela, J. Machado, J. Neves, and M. F. Santos, *Knowledge acquisition process for intelligent decision support in critical health care*. IGI Global Book, 2013.
- [13] J. Machado, V. Alves, A. Abelha, and J. Neves, "Ambient intelligence via multi-agent systems in the medical arena," *Engineering intelligent systems for electrical engineering and communications (2002)*, vol. 15, no. 3, pp. 151–157, 2007.
- [14] H. Peixoto, M. Santos, A. Abelha, and J. Machado, "Intelligence in interoperability with aida," in *Lecture Notes in Computer Science, Foundations of Intelligent Systems - 31*, ser. Lecture Notes in Computer Science, L. Chen, A. Felfernig, J. Liu, and Z. Raś, Eds. Springer Berlin Heidelberg, 2012, vol. 7661, pp. 264–273.
- [15] L. Nogueira and B. Ferreira, "A informática e a sua aplicação na Área de enfermagem," *Enferm UNISA*, vol. 1, pp. 1114–1117, 2000.
- [16] C. Lomba, J. Cabral, C. Ribeiro, R. Aguiar, and M. Fontes, "Circular informativa nº 02/2011-planeamento familiar e interrupção voluntária da gravidez (ivg)," Administração Regional de Saúde do Norte, Ministério da Saúde, Tech. Rep., 2011. [Online]. Available: <http://portal.arsnorte.min-saude.pt/portal/page/portal/ver-STAGE/ARSNorte/Conteúdos/Notícias/PlaneamentoFamiliarInterrupç~{a}oVoluntáriadaGrav.pdf>
- [17] A. Faúndes, *O drama do aborto: em busca de um consenso*. Campinas: Editora Komedi, 2004.
- [18] V. Beja and I. Leal, "Abortion counselling according to healthcare providers: a qualitative study in the lisbon metropolitan area, portugal." *The European journal of contraception & reproductive health care : the official journal of the European Society of Contraception*, vol. 15, no. 5, pp. 326–35, Oct. 2010. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20863255>
- [19] C. Canário, B. Figueiredo, and M. Ricou, "Abortamento: enquadramento legal, deontológico e perspectiva ética," 2011.
- [20] "decreto lei n.º 16/2007 de 17 de abril," Diário da República nº 75 - I Série, Lisboa, Tech. Rep., 2007. [Online]. Available: http://www.acss.min-saude.pt/Portals/0/Circulares/Portarian781Ade200716deJulho_IVG.pdf
- [21] C. Valente, T. Cristina, F. Rosário, and B. Alcina, "Acompanhamento de enfermagem na interrupção da gravidez por opção da mulher (i.g.o.)," Unidade da Mulher e da Criança. Centro Hospitalar do Porto, Porto, Tech. Rep., 2012.

- [22] "circular normativa n.º11/sr - organização dos serviços para implementação da lei n.º 16/2007 de 17 de abril," Direcção-Geral da Saúde, Tech. Rep., 2007. [Online]. Available: http://www.spdc.pt/files/legix/11270_3.pdf
- [23] "circular normativa n.º14/dir de 12/07/2007-interrupção da gravidez - determinação do tempo de gravidez ", Direcção-Geral de Saúde, Tech. Rep., 2007. [Online]. Available: <http://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas/circular-normativa-n-14dir-de-12072007.aspx>
- [24] "circular informativa n.º 20/sr de 18/07/2007 - modelo normalizado do consentimento livre e esclarecido para a interrupção da gravidez", Direcção-Geral de Saúde, Tech. Rep., 2007. [Online]. Available: <http://www.dgs.pt/?cr=11609>
- [25] "portaria n.º 741-a/2007 de 21 de junho", Diário da República n.º118 - I Série, no. 2, Lisboa, Tech. Rep., 2007. [Online]. Available: http://www.chuc.min-saude.pt/media/c_gravidez/Portaria_741-A-2007.pdf
- [26] "circular normativa n.º 09/sr de 21/06/2007 - interrupção medicamentosa da gravidez", Direcção-Geral de Saúde, Tech. Rep., 2007. [Online]. Available: <http://www.dgs.pt/directrizes-da-dgs/normas-e-circulares-normativas/circular-normativa-n-09sr-de-21062007.aspx>
- [27] "circular informativa n.º 02/2011 - planeamento familiar e interrupção voluntária da gravidez (ivg)", Administração Regional de Saúde do Norte, Tech. Rep., 2011. [Online]. Available: <http://portal.arsnorte.min-saude.pt/portal/page/portal/ARSNorte/Conteúdos/Notícias/PlaneamentoFamiliarInterrupç~{a}oVoluntáriadaGravidez>
- [28] K. Peffers, T. Tuunanen, M. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, pp. 45–78, 2008.
- [29] A. Hevner, S. March, J. Park, and S. Ram, "Design science in information systems research," *Mis Quarterly*, vol. 28, pp. 75–105, 2004.
- [30] P. Järvinen, "Action research is similar to design science," *Quality & Quantity: International Journal of Methodology*, vol. 41, no. 1, pp. 37–54, 2007.
- [31] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decision Support Systems*, vol. 15, no. 4, pp. 251–266, Dec. 1995. [Online]. Available: <http://dl.acm.org/citation.cfm?id=222827.222832>
- [32] J. NunaMaker, M. Chen, and T. Purdin, "Systems development in information systems research," *Journal of Management Information Systems*, vol. 7, pp. 80–106, 1991.

- [33] M. Rossi and M. Sein, "Design research workshop: a proactive research approach," *Presentation delivered at IRIS*, vol. 26, 2003.
- [34] H. Takeda, P. Veerkamp, T. Tomiyama, and H. Yoshikawa, "Modeling design processes," *AI Magazine*, vol. 11, no. 4, pp. 37–48, Oct. 1990. [Online]. Available: <http://dl.acm.org/citation.cfm?id=95788.95795>
- [35] J. G. Walls, G. R. Widmeyer, and O. A. El Sawy, "Building an information system design theory for vigilant eis," *Information Systems Research*, vol. 3, no. 1, pp. 36–59, Mar. 1992. [Online]. Available: <http://pubsonline.informs.org/doi/abs/10.1287/isre.3.1.36>
- [36] J. Eekels and N. Roozenburg, "A methodological comparison of the structures of scientific research and engineering design their similarities and differences," *Design Studies*, vol. 4, pp. 197–203, 1991.
- [37] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, "Crisp-dm 1.0," Tech. Rep., 2000.
- [38] R. Kimball and M. Ross, "The Data Warehouse Toolkit - The Complete Guide to Dimensional Modeling ," 2002.
- [39] R. Goebel, J. Siekmann, and W. Wahlster, *Advances in Knowledge Discovery and Data Mining*. Springer, 2008.
- [40] U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *American Association for Artificial Intelligence*, pp. 37–54, 1996.
- [41] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," in *Proceedings of the 2008 IEEE/ACS International Conference on Computer Systems and Applications*, ser. AICCSA '08. Washington, DC, USA: IEEE Computer Society, pp. 108–115.
- [42] M. Steiner, N. Soma, and T. Shimizu, "Abordagem de um problema médico por meio do processo de kdd com ênfase à análise exploratória dos dados," *Gest Prod*, 2006.
- [43] J. a. Gonçalves, F. Portela, M. F. Santos, A. Silva, J. Machado, A. Abelha, and F. Rua, "Real-time Predictive Analytics for Sepsis Level and Therapeutic Plans in Intensive Care Medicine," *International Information Institute*, 2013.
- [44] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth, *The CRISP-DM User Guide*. Brussels: NCR Systems Engineering Copenhagen, 2000.
- [45] K. Cios, W. Pedrycz, R. Swiniarski, and L. Kurgan, *Data Mining. A knowledge Discovery Approach*. Springer, 2007.

- [46] G. Linoff and M. Berry, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd ed. Wiley, 2011.
- [47] I. Colak, S. Sagioglu, and M. Yesilbudak, "Data mining and wind power prediction: A literature review," *Renewable Energy*, vol. 46, no. C, pp. 241–247, 2012.
- [48] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications - a decade review from 2000 to 2011," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 303–11 311, Sep. 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2221970.2222049>
- [49] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. Morgan Kaufmann Publishers, 2006.
- [50] A. Pandey, "Data Mining Techniques using WEKA," Ph.D. dissertation, VGSOM IIT Kharagpur, 2014.
- [51] G. Piatetsky-Shapiro, "Knowledge discovery in real databases: A report on the ijcai-89 workshop," *AI Magazine*, 1990. [Online]. Available: <http://aaai.org/journals/ai/magazine/article/viewFile/873/791>
- [52] H.-H. Tsai, "Global data mining: An empirical study of current trends, future forecasts and technology diffusions," *Expert Systems with Applications*, vol. 39, no. 9, pp. 8172–8181, Jul. 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2169475.2169824>
- [53] A. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Information Systems*, vol. 37, no. 2, pp. 99–116, Apr. 2012. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2051373.2051552>
- [54] D. O'Sullivan, W. Elazmeh, S. Wilk, K. Farion, S. Matwin, W. Michalowski, and M. Sehatkar, "Using secondary knowledge to support decision tree classification of retrospective clinical data," in *Mining Complex Data SE - 19*, ser. Lecture Notes in Computer Science, Z. Raś, S. Tsumoto, and D. Zighed, Eds. Springer Berlin Heidelberg, 2008, vol. 4944, pp. 238–251.
- [55] A. Razavi, H. Gill, H. Ahlfeldt, and N. Shahsavar, "Predicting metastasis in breast cancer: Comparing a decision tree with domain experts," *Journal of Medical Systems*, vol. 31, no. 4, pp. 263–273, 2007.
- [56] C. Catley, M. Frize, C. R. Walker, and D. C. Petriu, "Predicting high-risk preterm birth using artificial neural networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 540–549, 2006.

- [57] S. Chattopadhyay, P. Ray, H. S. Chen, M. B. Lee, and H. C. Chiang, "Suicidal risk evaluation using a similarity-based classifier," in *Advanced Data Mining and Applications SE - 7*, ser. Lecture Notes in Computer Science, C. Tang, C. Ling, X. Zhou, N. Cercone, and X. Li, Eds. Springer Berlin Heidelberg, 2008, vol. 5139, pp. 51–61.
- [58] A. Brandão, E. Pereira, F. Portela, M. F. Santos, J. Machado, and A. Abelha, "Managing Voluntary Interruption of Pregnancy using Data Mining," *Procedia Technology. Elsevier*, 2014.
- [59] A. Brandão, E. Pereira, F. Portela, M. F. Santos, A. Abelha, and J. Machado, "Predicting Risk Pregnancy in Maternity Care using Data Mining," in *ICAART 2015 - 7th International Conference on Agents and Artificial Intelligence*, 2014.
- [60] B. Janamanchi, E. Katsamakos, W. Raghupathi, and W. Gao, "The State and Profile of Open Source Software Projects in health and medical informatics," *International Journal of Medical Informatics*, vol. 78, no. 7, pp. 457–472, Jul. 2014. [Online]. Available: [http://www.ijmijournal.com/article/S1386-5056\(09\)00031-8/abstract](http://www.ijmijournal.com/article/S1386-5056(09)00031-8/abstract)
- [61] T. Waring and P. Maddocks, "Open Source Software Implementation in the UK Public Sector: Evidence from the Field and Implications for the Future," *Int. J. Inf. Manag.*, vol. 25, no. 5, pp. 411–428, 2005.
- [62] I. D. Muraina and A. Ahmad, "Healthcare Business Intelligence: The Case of University's Health Center." [Online]. Available: http://www.academia.edu/4465749/Healthcare_Business_Intelligence_The_Case_of_Universitys_Health_Center
- [63] J. Bernardino, "Open Source Business Intelligence Platforms for Engineering Education," in *1st World Engineering Flash Week*, Lisbon, 2011.
- [64] T. Sá, A. Mateus, M. Iwanaga, A. Ribeiro, T. Escovedo, and R. N. Melo, "Uma análise Comparativa entre as Ferramentas OLAP como Apoio a Soluções de BI nas Empresas."
- [65] M. Tereso and J. Bernardino, "Open source business intelligence tools for SMEs," *IEEE*, pp. 1–4, 2011.
- [66] J. Bernardino and P. Ribeiro, "Open source Business Intelligence: an alternative to proprietary tools," in *International Journal of Electronic Business*, 2011, vol. 3.
- [67] E. Damiani, F. Frati, and C. Monteverdi, "Open Source BI adoption," in *Ow2 BI Initiative*, 2009. [Online]. Available: <http://www.ow2.org/xwiki/bin/download/BusinessIntelligence/Documents/OSBIadoption-v10.pdf>
- [68] C. Thomsen and T. B. Pedersen, "A Survey of Open Source Tools for Business Intelligence," Aalborg Universitet, Aalborg, Tech. Rep., 2008. [Online]. Available: <http://vbn.aau.dk/files/14833824/DBTR-23.pdf>

- [69] C. H. Foundation, *Open Source Software : A Primer for Health Care Leaders Open Source Software : A Primer for Health Care Leaders*, 2006, no. March.
- [70] C. L. Huntley, *Organizational learning in open-source software projects: an analysis of debugging data*, 2003, vol. 50, no. 4.
- [71] P. Nagy, “Open source in imaging informatics.” *Journal of digital imaging*, vol. 20 Suppl 1, no. August, pp. 1–10, Nov. 2007. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2039818&tool=pmcentrez&rendertype=abstract>
- [72] P. Welche, D. Bazarko, K. Ritten, Y. Burgess, R. Harmon, and L. Sandy, “Electronic Health Records in Four Community Physician Practices : Impact on Quality and Cost of Care,” pp. 320–328, 2007.
- [73] SpagoBI, “SpagoBI for Areas.” [Online]. Available: <http://www.spagoworld.org/xwiki/bin/view/SpagoBI/ApplicationsAreas>
- [74] Pentaho, “Pentaho Healthcare Solutions,” 2012. [Online]. Available: <http://pt.slideshare.net/Pentaho/pentaho-healthcare-solutions-presentation>
- [75] J. Lapa, J. Bernardino, and A. Figueiredo, “A comparative analysis of open source business intelligence platforms,” *Proceedings of the International Conference on Information Systems and Design of Communication - ISDOC '14*, pp. 86–92, 2014. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2618168.2618182>
- [76] J. Bernardino, “Open Business Intelligence for Better Decision-Making,” *International Journal of Information Communication Technologies and Human Development*, vol. 5, no. 2, pp. 20–36, 2013. [Online]. Available: <http://www.igi-global.com/article/open-business-intelligence-better-decision/77865>
- [77] A. Marinheiro and J. Bernardino, “OpenBRR evaluation of an open source BI suite,” in *Proceedings of the International Conference on Computer Science and Software Engineering - C3S2E '13*. New York, New York, USA: ACM Press, Jul. 2013, p. 134. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2494444.2494463>
- [78] Microstrategy, “Microstrategy 9 vs. qliktech (qlikview11). a comparison white paper,” Tech. Rep.
- [79] Pwi, “QlikView: análise de resultados com a melhor ferramenta de BI do mercado.” [Online]. Available: <http://www.pwi.com.br/qlikview.asp>
- [80] I. Güclü and S. Müller, “Business Intelligence mit Open Source: Vergleich der BI-Lösungen Jaspersoft, Jedox Palo und Pentaho,” *t3n*, 2010. [Online]. Available: <http://t3n.de/magazin/vergleich-bi-losungen-jaspersoft-jedox-palo-pentaho-224644/>

- [81] M. Rueter and E. Fields, “Tableau for the Enterprise : An Overview for IT,” Tech. Rep., May 2012.
- [82] J. Reinschmidt and A. Francoise, *Business Intelligence Certification Guide*. IBM, 2000.
- [83] W. H. Inmon, “Building the Data Warehouse,” *John Wiley & Sons, Inc.*, Jan. 1992. [Online]. Available: <http://dl.acm.org/citation.cfm?id=531510>
- [84] S. d. O. Alcântara, “Business Intelligence (BI) como auxílio à Gestão do Negócios,” Faculdade de Tecnologia da Zona Leste, São Paulo, Tech. Rep., 2010.
- [85] S. Chaudhuri, U. Dayal, and V. Narasayya, “An Overview of Business Intelligence Technology,” *Commun. ACM*, vol. 54, no. 8, pp. 88–98, 2011.
- [86] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. El Bastawissy, “A proposed model for data warehouse ETL processes,” *Journal of King Saud University - Computer and Information Sciences*, vol. 23, no. 2, pp. 91–104, Jul. 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S131915781100019X>
- [87] A. Nylund, “Tracing the BI Family Tree,” Knowledge Management, 1999.
- [88] S. Ghoshal and S. K. Kim, “Building Effective Intelligence Systems for Competitive Advantage,” *Sloan Management Review*, 1986.
- [89] W. W. Eckerson, *Performance Dashboards: Measuring, Monitoring, and Managing Your Business*. Wiley, 2005. [Online]. Available: <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471757659.html>
- [90] A. Lonnqvist and V. Pirttimaki, “The Measurement of Business Intelligence,” *Information Systems Management*, 2006. [Online]. Available: <http://mail.im.tku.edu.tw/~cjou/bi2008/MeasurementOfBI.pdf>
- [91] M. Petrini and M. Pozzebon, *Data Mining Applications for Empowering Knowledge Societies*, H. Rahman, Ed. IGI Global, Jul. 2008. [Online]. Available: <http://www.igi-global.com/chapter/role-business-intelligence-playing-developing/7555>
- [92] R. Bose, “Advanced analytics: opportunities and challenges,” *Industrial Management and Data Systems*, pp. 155–172, 2009.
- [93] J. Ranjan, “Business justification with business intelligence,” *Emerald Group Publishing Limited*, Apr. 2013.
- [94] L. Wu, G. Barash, and C. Bartolini, “A Service-oriented Architecture for Business Intelligence,” in *IEEE International Conference on Service-Oriented Computing and Applications (SOCA '07)*. IEEE, Jun. 2007, pp. 279–285. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1270402.1271807>

- [95] M. Z. Elbashir, P. A. Collier, and M. J. Davern, “Measuring the effects of business intelligence systems: The relationship between business process and organizational performance,” *International Journal of Accounting Information Systems*, vol. 9, no. 3, pp. 135–153, Sep. 2008.
- [96] H. Jalonen and A. Lönnqvist, “Predictive business – fresh initiative or old wine in a new bottle,” *Emerald Group Publishing Limited*, Nov. 2012.
- [97] E. Pereira, A. Brandão, C. F. Portela, M. F. Santos, J. Machado, and A. Abelha, “Business intelligence in maternity care,” in *Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS '14*. New York, New York, USA: ACM Press, Jul. 2014, pp. 352–355. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2628194.2628248>

Apêndice A

Publicações

A.1 Managing Voluntary Interruption of Pregnancy using Data Mining

Autores: Andreia Brandão, Eliana Pereira, Carlos Filipe Portela, Manuel Filipe Santos, António Abelha, José Machado

Conferência: ESM : CENTERIS 2014 - Conference on ENTERprise Information Systems / ProjMAN 2014 - International Conference on Project MANagement / HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies

Ano: 2014

Estado: Aceite para publicação

Abstract: When a woman aims to terminate an unplanned pregnancy, she must go to a specialized healthcare unit, such as Júlio Dinis Maternity Hospital. In this unit, the procedures of voluntary interruption of pregnancy are done by two kinds of drug administration: the first one is always done by a nursing team, the second one can be performed at home or by a nursing team, depending on patient features. It is important to give the best option to the pregnant. In this paper, it is proposed to predict whether the second drug phase is done at home

or at the hospital. The use of Data Mining (DM) helps in performing this step. Throughout this study, DM models capable to make predictions in a real environment using real data are induced. It was adopted the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. Four distinct techniques were considered: Decision Tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM) and Generalized Linear Models (GLM) to perform classification tasks. Using these techniques it was possible to obtain acceptable results for each model. A value greater than 89% of accuracy and 91% in sensibility was achieved in some models.

Keywords: Data Mining, Maternity Hospital, Crisp-DM, Voluntary Interruption of Pregnancy, Pregnancy

A.2 Real-Time Business Intelligence Platform to Maternity Care

Autores: Andreia Brandão, Eliana Pereira, Filipe Portela, Manuel Santos, António Abelha and José Machado

Conferência: IEEE Conference on Biomedical Engineering and Sciences

Ano: 2014

Estado: Aceite Para Publicação

Abstract: The motivation for the implementation of a decision support system in maternity care came from the fact that people are constantly making quick decisions based on incomplete information. There is a significant impact on the patient health, as well as in increasing medical errors. To implement this system, it was resorted to the technologies of Business Intelligence, which involved the construction of two data warehouses with a dimensional structure in a star shape, for

A.3. *PREDICTING RISK PREGNANCY IN MATERNITY CARE USING DATA MINING*

two distinct modules, in Gynaecology and Obstetrics cares. The feasibility of an evidence-based practice and medical decision making in real time with universal and interoperable features are some of the benefits resulting from the implementation of decision support system in maternity care. In this paper we present the architecture of BI solution, some clinical outcomes and some benefits of the BI solution in a real world context.

A.3 Predicting Risk Pregnancy in Maternity Care using Data Mining

Autores: Andreia Brandão, Eliana Pereira, Filipe Portela, Manuel Filipe Santos, António Abelha, José Machado

Conferência: ICAART 2015 - 7th International Conference on Agents and Artificial Intelligence.

Ano: 2014

Estado: Aceite Para Publicação

Abstract: When a woman has the intension to make an Interruption of Pregnancy she should generally use a health unit specialized in the process, as it is the case of Maternidade Júlio Dinis. The process is comprised of four stages, where the first three steps consist in the performance of the procedure itself and the last stage focuses on the evaluation process. So the purpose of this article is to evaluate the process of Voluntary Interruption of Pregnancy and consequently identify the risk patients group. To this work it was induced Data Mining (DM) models to make predictions in a real environment using real data. It was applied Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to support this work. Three different techniques were considered: Decision Tree (DT), Support Vector Machine (SVM) and Generalized Linear Models (GLM) to perform the classification task.

Using these techniques, it was possible to obtain good results from the models induced, achieving a sensitivity of approximately 93%.

Keywords: Data Mining, Intelligent Decision Support Systems, Voluntary Interruption of Pregnancy, Business Intelligence, Technology Acceptance.

A.4 Business Intelligence in Maternity Care

Autores: Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha

Conferência: IDEAS 2014 - International Database Engineering & Applications Symposium

Editora: ACM

Ano: 2014

Estado: Publicado

Abstract: The emergency services are usually pressured to make quick decisions with incomplete information on most cases, and this situation has a significant impact on healthcare as well on increasing medical errors. On the other hand, there has been an increase of the Electronic Health at Maternity Care. The combination of these two factors allows the construction of a Decision Support System specific for Maternity Care Unit using Business Intelligence technology. This solution is supported by a Data Warehouse that uses the dimensional structure snowake and makes the modeling of the maternity care database. With this solution it is intended to turn possible a clinical evidence-based practice, allowing for real time medical decision making with pervasive and interoperable characteristics. This paper presents the architecture, KPIs and benefits of Business Intelligence solution for the real context. This platform has several modules of clinical importance.

A.5. *PRE-TRIAGE DECISION SUPPORT IMPROVEMENT IN MATERNITY CARE BY*

The Obstetric Gynaecological Emergency and the Voluntary Interruption of Pregnancy modules are object of study. This solution has an innovative contribution to the medical and scientifically community studying the problem in Maternity area.

Keywords: Business Intelligence; Data Mining; Triage; Emergency Department; Intelligent Decision Support Systems; Manchester Triage System; Pervasive; Real-time; Interoperability.

A.5 Pre-Triage Decision Support Improvement in Maternity Care by means of Data Mining

Autores: Eliana Pereira, Andreia Brandão, Maria Salazar, Carlos Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha and Jorge Braga

Livro : Integration of Data Mining in Business Intelligence Systems

Editora: IGI Global Book

Ano: 2014

Estado: Publicado

Abstract: The conventional triage systems like Manchester Triage System (MTS) are not suitable for maternity care, a decision model for pre-triaging patients in emergency (URG) and consultation (ARGO) classes was built and incorporated into a Decision Support System (DSS) implemented in Centro Materno Infantil do Norte (CMIN). Complementarily, DSS produces several indicators to support clinical and management decisions. A recent data analysis revealed a bias in the classification of URG cases. Frequently, cases classified as URG correspond to ARGO. This misclassification has been studied by means of

Data Mining (DM) techniques in order to improve the pre-triage model and to discover knowledge for developing a new triage system based on waiting times and on a 5-scale of classes. This chapter presents a kind of sensitivity analysis combining input variables in six scenarios and considering four different DM techniques. CRISP-DM methodology was used to conduct the project.

- **Keywords:** AIDA; Data Mining; Decision Support System; Interoperability; Knowledge Discovery from Databases; Manchester Triage System; Maternity Care; Obstetric Triage Acuity Scale, SWOT analysis

A.6 Improving Quality of Services in Maternity Care Triage System

Autores: Eliana Pereira; Andreia Brandão; Carlos Filipe Portela; Manuel Filipe Santos; José Machado; António Abelha; Jorge Braga.

Journal: International Journal of E-Health and Medical Communications (IJEHMC)

Editora: IGI Global

Ano: 2014

Estado: Aceite para publicação

Abstract: The main objectives in triage are to improve the quality of care and reduce the risks associated to the waiting time in emergency care. Thus, an efficient triage is a good way to avoid some future problems and how much quicker it is, more the patient can benefit. The most common triage system is the Manchester Triage System that is a reliable system focused in the emergency department of a hospital. However, its use is more suitable for more widespread medical emergencies and not for specialized cases, like Gynecological and Obstetrics

A.7. *SIMULATING A MULTI-LEVEL PRIORITY TRIAGE SYSTEM FOR MATERNITY*

emergencies. To overcome these limitations, an alternative pre-triage system, integrated into an intelligent decision support system, was developed in order to better characterize the patient and correctly defined her as urgent or not. This system allows the increase of patient's safety, especially women who need immediate care. This paper includes the workflow that describes the decision process in real time in the emergency department, when women are submitted to triage and identify points of evolution.

Keywords: Triage; Emergency Department; Intelligent Decision Support System; Manchester Triage System; Triage System Gynecology Obstetrics; Real Time Decision Support

A.7 Simulating a Multi-Level Priority Triage System for Maternity Emergency

Autores: António Abelha, Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Santos and José Machado

Conferência: ESM - 28 th European Simulation and Modelling Conference

Ano: 2014

Estado: Aceite para publicação

Abstract: Nowadays Decision Support Systems are increasingly used in order to help health professionals. An example of this application is the implementation of a triage system in hospital emergency. In Centro Materno Infantil do Norte it was implemented an intelligent system of pre-triage which aims to prioritize the patients on two levels: Urgent and or outpatient service. However, although specific for obstetrics and gynaecology cases, the system does not meet all clinical requirements. Thus using a simulation algorithm developed within this framework, it was intended to simulate a specific priority triage system

for gynaecology and obstetrics but with five levels of acuity as suggested by the Portuguese general department of Health (Direção Geral de Saúde). For this study the repository of specific pre-triage system was used to test the algorithm. After application, it was found that the implementation of this system will reduce waiting time, allowing a uniform distribution according to the waiting time and the clinical features. The percentage of deviation between the waiting time and the actual time obtained by simulation algorithm is approximately 121.6%

Keywords: Maternity Care, Intelligent Decision Support Systems, Real-Time, Interoperability, Specific Priority Triage System, Gynecology and Obstetrics.

A.8 Application of Business Intelligence Tools in Healthcare Environment

Autores: Andreia Brandão, Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Santos, António Abelha and José Machado

Journal: Journal of Convergence Information Technology.

Ano: 2014

Estado: Submetido

Abstract: In recent years have been applied technologies of Business Intelligence (BI) to various areas in order to support the decision making process of users and enable the extraction of knowledge from data stored. The health care industry is no exception and, as such, has been under investigation for development of business intelligence applications across multiple units of different institutions. Thus, in this article, it is intended to analyze some Business Intelligence open source / free tools on the market and its applicability in the clinical area, taking into consideration the general characteristics of the clinical environment. For this purpose were selected, analyzed and tested

A.9. *STEP TOWARDS TO IMPROVE THE VOLUNTARY INTERRUPTION OF PREGNANCY*

in practical environment six BI tools. Then, it was created a metric of comparison and performed a ranking of the tested applications to choose the one that best applies to the extraction of useful knowledge of clinical data.

A.9 Step Towards to improve the Voluntary Interruption of Pregnancy Process by means of Business Intelligence

Autores: José Machado, António Abelha, Filipe Portela, Andreia Brandão

Livro : Applying Business Intelligence to Clinical and Healthcare Organizations

Editora: IGI Global Book

Ano: 2014

Estado: Submetido

Abstract: With the implementation of Information and Communication Technologies in the health sector, it became possible the existence of an electronic record of information for the patients, enabling the storage and the availability of their information in databases. However, without the implementation of a BI system, this information has no value. Thus, the major motivation of this paper is to create a decision support system that allows the transformation of information into knowledge, giving usability to the stored data. The particular case addressed in this paper is the Centro Materno Infantil do Norte, in particular the Voluntary Interruption of Pregnancy unit. With the creation of a BI system for this module, it is possible to design an interoperable, pervasive and real-time platform to support the decision making process of health professionals, based on cases that occurred.

Furthermore, this platform enables the automation of the process for obtaining key performance indicators that are presented annually by this health institution. In this paper, the BI system implemented in the Voluntary Interruption of Pregnancy unity in CMIN, some of the key performance indicators evaluated as well as the benefits of this implementation are presented.

Apêndice B

Glossário

Java 2 Platform, Enterprise Edition É uma plataforma de programação para servidores na linguagem de programação Java. A plataforma fornece uma interface e um ambiente de tempo de execução para o desenvolvimento e execução de softwares corporativos, incluindo serviços de rede e web.. 73

Key Performance Indicators São medidas quantificáveis para compreender se os objetivos de uma determinada organização estão a ser atingidos.. 62, 84

Matriz de Confusão A matriz de confusão de uma determinada hipótese oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas versus as classificações previstas para cada classe, sobre um conjunto de exemplos.. 29

Object Linking and Embedding Database É uma interface desenvolvida pela Microsoft que permite o acesso universal a diversas fontes de dados.. 65

Online Transaction Processing Sistemas que se encarregam de registrar todas as transações existentes numa determinada operação organizacional.. 66

Open Database Connectivity Padrão para acesso a sistemas de gestão de bases de dados, que define um conjunto de interfaces que permitem o uso de linguagens de programação como Visual Basic, Delphi, Visual C++, Java, entre outras.. 65

Processo Clínico Eletrónico É um registo que contém informação clínica da saúde e da doença de um paciente, após este ter procurado auxílio médico. Habitualmente as notas são feitas por médicos, enfermeiros e outros profissionais de saúde. Estes registos contêm considerações, achados, resultados de meios complementares de diagnóstico e informações sobre o tratamento do processo patológico.. [v](#), [7](#), [20](#), [92](#), [95](#), [97](#)

Random Access Memory É um tipo de memória que permite a leitura e a escrita, utilizada como memória primária em sistemas eletrónicos digitais.. [81](#)

Sistemas de Gestão de Base de Dados Relacionais São aplicações informáticas complexas, onde grandes quantidades de informação necessitam de ser combinadas ou exploradas, de diversas formas.. [7](#)