



Otimização do agendamento de cirurgias em meio hospitalar

João Braga

UMinho | 2021

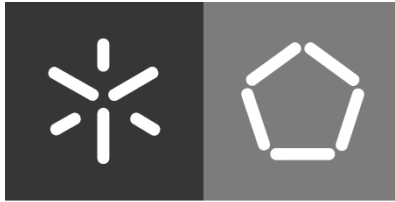


Universidade do Minho
Escola de Engenharia

João José Oliveira Braga

Otimização do agendamento de
cirurgias em meio hospitalar

Julho de 2021



Universidade do Minho

Escola de Engenharia

João José Oliveira Braga

Otimização do agendamento de cirurgias em meio hospitalar

Dissertação de Mestrado

Mestrado Integrado em Engenharia e Gestão de
Sistemas de Informação

Trabalho efetuado sob a orientação do

Professor Doutor Manuel Filipe Santos

Doutor Júlio Miguel Marques Duarte

Julho de 2021

DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença abaixo indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

Licença concedida aos utilizadores deste trabalho

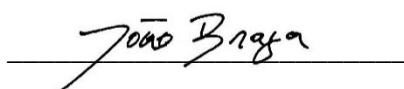


Atribuição

CC BY

<https://creativecommons.org/licenses/by/4.0/>

Universidade do Minho, 29 de julho de 2021

A handwritten signature in black ink, which appears to read 'João Braga', is written over a horizontal line.

REPOSITORY STATEMENT

Name: João José Oliveira Braga

E-mail: a77036@alunos.uminho.pt Phone: 925700185

ID card number/Passport number: 14348790

Type of academic work to archive:

- PhD: ☐ Thesis ☐ Other Specify

Master's: ☒ Dissertation ☐ Practicum Report ☐ Project ☐ Other Specify

Title of the academic work: Otimização do agendamento de cirurgias em meio hospitalar

Advisor(s): Manuel Filipe Santos

Conclusion date: 30/07/2021

Master's and specialization area/PhD and speciality: Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação

I hereby grant the University of Minho the non-exclusive and irrevocable right to archive, reproduce, communicate and/or distribute through its institutional repository, accordingly to the conditions mentioned below, the final version approved after the completion of the public defence trials and, when applicable, after confirmation by the advisor(s) of the insertion of requested amendments and corresponding homologation by the jury president¹, of my academic work referred to above in digital format.

I hereby authorize the University of Minho to archive more than one copy of my academic work referred to above and, without altering its content, to convert the delivered archive to any file format, media or support, for the purpose of preservation and access.

I declare that my academic work referred to above and now delivered is an original work, and if it contains material which I don't hold copyright, that I obtained previous permission from the copyright owner to grant the University of Minho the terms required by this license.

I also declare that the delivery of this document does not infringe, as far as I can know, the rights of any other person or entity.

If the submitted document is based on funded work or supported by other organization/funder rather than the University of Minho, I declare that I have fulfilled any obligations required by the respective contract or agreement.

I retain all copyrights pertaining to my academic work referred to above and the right to use it in future works (such as articles or books).

I agree that my academic work referred to above is placed in the University of Minho institutional repository with the following status (check only one of the three statutes):

1. ☒ Immediate availability of the work set for worldwide access

2. ☐ Availability of the work set for exclusive access within the University of Minho during:


☐ 1 year ☐ 2 years ☐ 3 years. After the indicated time I authorize worldwide access.

3. ☐ Other. Please, specify

If I selected the statutes 2 or 3, I assume the compromise of submitting the decision on my request to the University of Minho Rector, for my academic work under equation to be given an exceptional archive status, within five working days, after receiving it.

I declare that I was informed by the university of Minho that my personal data provided in this form will be used with the only purpose of managing the archive of my work in the RepositóriUM of the University of Minho. I was also informed that I can activate my rights relative to personal data protection in the Documentation Services that which are responsible for the RepositóriUM. I am aware that the University of Minho has a person responsible for data protection whose contacts are available from <http://www.uminho.pt/dataprotection>.

Braga/Guimarães, 30/07/2021

Signature: 

AGRADECIMENTOS

O desenvolvimento deste projeto, assim como a conceção deste documento, deve-se à cooperação e dedicação de várias pessoas, ao longo de mais de um ano de trabalho. Em primeiro lugar, quero agradecer ao meu orientador e Professor Manuel Filipe Santos, pela aceitação como seu orientando e pelos seus ensinamentos ao longo do trabalho.

Gostaria de expressar a minha gratidão ao meu co-orientador Júlio Miguel Marques Duarte e também ao Tiago Guimarães e João Lopes pela ajuda e prontidão que sempre me mostram quando enfrentava novos desafios.

Agradeço ainda aos profissionais de saúde e aos profissionais dos serviços de informação do Centro Hospitalar do Porto pela enorme colaboração facultada.

Aos meus amigos, que sempre estiveram presentes nos momentos altos e baixos do desenvolvimento desta dissertação, especialmente ao Gonçalo, Artur, Filipe, David, Francisco, Pratik e Vitor, agradeço a paciência, as palavras de motivação e os bons momentos que sempre me ofereceram.

Quero expressar um reconhecimento especial à Eva Correia pela presença, compreensão e apoio incondicional ao longo deste percurso académico, sem o seu apoio este percurso teria sido muito mais difícil.

Agradeço, ainda, aos meus amigos da Universidade do Minho, os elementos que fazem parte da “Família” deste curso, e que partilharam este percurso académico ao longo destes 5 longos anos. Relativamente a este projeto, um especial agradecimento ao João Abreu, Bruno Magalhães, Marco Cruz, André Amaral e Ricardo Cruz pelo apoio nas fases mais importantes.

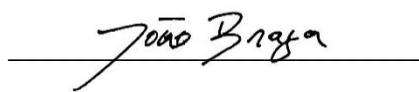
Deixo um agradecimento muito especial aos meus pais e irmão, com que pude, e posso, contar sempre em todos os momentos da minha vida, que me permitiu ter todas as condições para terminar esta fase académica.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.

Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 29 de julho de 2021

A handwritten signature in black ink, reading "João Braga", is written over a horizontal line.

RESUMO

Esta dissertação enquadra-se num projeto de dissertação de mestrado em Engenharia e Gestão de Sistemas de Informação na Universidade do Minho, sendo o tema da mesma “Otimização do agendamento de cirurgias em meio hospitalar”. A otimização é um tópico central do processo de um sistema *Adaptive Business Intelligence*, combinando ferramentas analíticas com algoritmos inteligentes que possibilitem a apresentação de informações complexas e competitivas para otimizar o agendamento de cirurgias e ajudar então no processo de tomada de decisão.

O Ministério da Saúde estabelece o tempo de resposta máximo garantido para as cirurgias, meios complementares de diagnóstico e consultas externas. Estes tempos são frequentemente ultrapassados, acarretando custos elevados para os hospitais e doentes. Posto isto, o principal objetivo desta dissertação consiste em recorrer a técnicas de otimização modernas para desenvolver modelos que permitam estabelecer agendamentos mais eficientes, o que permitirá a minimização de desperdícios do tempo de duração total dos turnos necessários, evitando os atrasos que são gerados pela falta de aproveitamento de turnos.

O trabalho desenvolvido seguirá não só a metodologia de investigação *Design Science Research*, como também a metodologia prática *Cross Industry Standard Process for Data Mining* (CRISP-DM). Neste documento é inicialmente realizado um enquadramento, respetivas motivações e a identificação dos objetivos e um estudo das abordagens metodológicas a serem utilizadas no projeto. De seguida é apresentado um estado de arte sobre os temas a abordar nesta dissertação. Posteriormente, uma descrição do trabalho desenvolvido e a discussão dos seus resultados. Por fim, uma síntese do trabalho realizado e uma reflexão sobre as conclusões relevantes a retirar do projeto.

Palavras-Chave: *Adaptive Business Intelligence*, Otimização, Sistemas de Apoio à Decisão

ABSTRACT

This dissertation is part of an MSc dissertation project in Engineering and Information Systems Management at the University of Minho, and its topic is "Optimization of surgery scheduling in a hospital environment". Optimization is a central topic in the process of an Adaptive Business Intelligence system, combining analytical tools with intelligent algorithms that enable the presentation of complex and competitive information to optimize surgery scheduling and then help in the decision-making process.

The Ministry of Health establishes maximum guaranteed response times for surgeries, complementary diagnostic procedures and outpatient medical appointments. These times are frequently exceeded, resulting in high costs for hospitals and patients. Therefore, the main objective of this dissertation is to be able to use modern optimization techniques to develop models to establish more efficient schedules, which will allow the minimization of waste of total duration time, avoiding the delays that are generated by the lack of utilization of shifts.

The work developed will follow not only the Design Science Research methodology, but also the practical Cross Industry Standard Process for Data Mining (CRISP-DM) methodology. This document initially contains a framework, respective motivations and the identification of objectives and a study of the methodological approaches to be used in the project. Next, a state-of-the-art on the themes to be addressed in this dissertation is presented. Afterwards, a description of the work developed and the discussion of its results. Finally, a summary of the work carried out and a reflection on the relevant conclusions to be drawn from the project.

Keywords: Adaptive Business Intelligence, Optimization, Decision Support Systems

ÍNDICE

1. INTRODUÇÃO.....	16
1.1 ENQUADRAMENTO E MOTIVAÇÃO	16
1.2 OBJETIVOS E RESULTADOS ESPERADOS	17
1.3 METODOLOGIAS DE INVESTIGAÇÃO.....	ERROR! BOOKMARK NOT DEFINED.
1.3.1 <i>Design Science Research</i>	18
1.3.2 <i>Cross Industry Standard Process for Data Mining</i>	21
1.4 ESTRUTURA DO DOCUMENTO	22
2. ESTADO DE ARTE.....	24
2.1 SISTEMAS DE INFORMAÇÃO.....	24
2.2 SISTEMAS DE INFORMAÇÃO NA ÁREA DA SAÚDE	26
2.3 BUSINESS INTELLIGENCE	27
2.3.1 <i>Introdução</i>	27
2.3.2 <i>Extração, Transformação e Carregamento de dados (ETL)</i>	30
2.3.3 <i>Data Warehouse</i>	32
2.4 ADAPTIVE BUSINESS INTELLIGENCE	34
2.4.1 <i>Data Mining</i>	35
2.4.2 <i>Previsão</i>	37
2.4.3 <i>Otimização</i>	38
2.5 ADAPTIVE BUSINESS INTELLIGENCE NA SAÚDE	42
2.5.1 <i>Problema das restrições na especialidade cirurgia</i>	42
2.5.2 <i>Caso de Estudo Relacionado</i>	44
3. PLATAFORMA DE OTIMIZAÇÃO DO AGENDAMENTO	47
3.1 CONCEÇÃO E DESENVOLVIMENTO.....	47
3.2 FASES CRISP-DM	51
3.2.1 <i>Compreensão do negócio</i>	51
3.2.2 <i>Compreensão dos dados</i>	52
3.2.3 <i>Preparação dos dados</i>	55
3.2.4 <i>Modelação</i>	57
3.2.5 <i>Avaliação</i>	57
4. DISCUSSÃO E APRESENTAÇÃO DE RESULTADOS OBTIDOS	61
4.1 OTIMIZAÇÃO.....	61
4.2 INTERFACE GRÁFICA	ERROR! BOOKMARK NOT DEFINED.
5. CONCLUSÃO.....	73
5.1 SÍNTESE	73
5.2 TRABALHO FUTURO	75
6. REFERÊNCIAS	77

LISTAS DE ABREVIATURAS E SIGLAS

ABI – Adaptive Business Intelligence

BI – Business Intelligence

CRISP-DM – Cross Industry Standard Process for Data Mining

CHP – Centro Hospitalar do Porto

CHUP – Centro Hospitalar e Universitário do Porto

DM – Data Mining

DW – Data Warehouse

DRG – Diagnosis Related Groups

DSRM – Design Science Research Methodology

EIS – Executive Information Systems

ETL – Extract Transform Load

IA – Inteligência Artificial

KPI – Indicadores de Performance

MCDT – Meios Complementares de Diagnóstico e Terapeuta

MOLAP – Multidimensional On-line Analytical Processing

ML – Machine Learning

OLAP – On-Line Analytical Processing

ROLAP – Relational On-line Analytical Processing

SAD – Sistemas de Apoio à Decisão

SI – Sistemas de Informação

TMRG – Tempos de Resposta Máxima Garantida

TI – Tecnologias de Informação

Índice de Figuras

FIGURA 1- FASES DA DSRM. ADAPTADO DE (PEFFERS, K., TUUNANEN, T., ROTHENBERGER, M. A., & CHATTERJEE, 2007).	18
FIGURA 2- MODELO DE PROCESSOS DO CRISP-DM. TRADUZIDO E ADAPTADO DE (CHAPMAN ET AL., 2000). .	21
FIGURA 3- ARQUITETURA DE UM SISTEMA DE BI. ADAPTADO DE (CHAUDHURI, DAYAL E NARASAYYA, 2011). .	28
FIGURA 4- PROCESSO DE ETL. ADAPTADO POR (JUN ET AL., 2009).....	31
FIGURA 5- ARQUITETURA DE DW. RETIRADO DE (MOODY, 2000).....	32
FIGURA 6- PROCESSOS DE ADAPTIVE BUSINESS INTELLIGENCE TRADUZIDO E ADAPTADO DE (MICHALEWICZ, 2007).....	34
FIGURA 7- PROCESSO DE DESCOBERTA DE CONHECIMENTO. TRADUZIDO E ADAPTADO DE (HAN, KAMBER, & PEI, 2012).....	36
FIGURA 8- HIERARQUIA DE ESCOLHAS DE DATA MINING ADAPTADO DE (EDELSTEIN, 1999)	37
FIGURA 9- CLASSIFICAÇÃO DOS MÉTODOS DE OTIMIZAÇÃO. RETIRADO DE (CORTEZ, 2014).	39
FIGURA 10- PROCURA POR NOVAS SOLUÇÕES EM MODELOS DE PROCURA LOCAL. RETIRADO (CORTEZ, 2014).	40
FIGURA 11- PROCURA BASEADA NA POPULAÇÃO. RETIRADO DE (CORTEZ, 2014)	41
FIGURA 12- MODELO DE RESTRIÇÕES NAS SALAS OPERATÓRIAS. RETIRADO DE (MESKENS, DUVIVIER, AND HANSET, 2013).....	43
FIGURA 13- AMOSTRA DA SOLUÇÃO INICIAL GERADA PELO SISTEMA.....	62
FIGURA 14- RESULTADO DO ALGORITMO HILL CLIMBING COM 10 ITERAÇÕES.	62
FIGURA 15- EVOLUÇÃO DO ALGORITMO DE HILL CLIMBING EM RELAÇÃO AO NÚMERO DE ITERAÇÕES.....	63
FIGURA 16- RESULTADO DO ALGORITMO HILL CLIMBING COM 10 ITERAÇÕES DA SEGUNDA SOLUÇÃO INICIAL. .	64
FIGURA 17- EVOLUÇÃO DO HILL CLIMBING EM RELAÇÃO AO NÚMERO DE ITERAÇÕES DA SEGUNDA SOLUÇÃO INICIAL.	64
FIGURA 18- RESULTADO DO ALGORITMO SIMULATED ANNEALING COM 10 ITERAÇÕES PARA A SEGUNDA SOLUÇÃO.....	67
FIGURA 19- EVOLUÇÃO DO ALGORITMO DE SIMULATED ANNEALING EM RELAÇÃO AO NÚMERO DE ITERAÇÕES.	68
FIGURA 20- RESULTADO DO EVOLUTIONARY ALGORITHM COM 10 ITERAÇÕES.....	69
FIGURA 21- EVOLUÇÃO DO EVOLUTIONARY ALGORITHM EM RELAÇÃO AO NÚMERO DE ITERAÇÕES	70
FIGURA 22- APLICAÇÃO PARA O AGENDAMENTO DE CIRURGIAS.....	ERROR! BOOKMARK NOT DEFINED.
FIGURA 23- INDICADORES DA APLICAÇÃO.....	60

ÍNDICE DE TABELAS

TABELA 1- PRINCIPAIS BENEFÍCIOS DE BUSINESS INTELLIGENCE. RETIRADO DE (SINGH & SAMALIA, 2014).	28
TABELA 2 - SUMARIO DOS MODELOS DE REGRESSÃO COM OS 14 ATRIBUTOS. ADAPTADO DE (FERREIRA ET AL., 2019).....	45
TABELA 3- OTIMIZAÇÃO SEQUENCIAL MENSAL DE DESPERDÍCIOS EM CIRURGIA VASCULAR. ADAPTADO DE (FERREIRA ET AL., 2019)	46
TABELA 4- ATRIBUTOS SELECIONADOS REFERENTE AOS TURNOS.	52
TABELA 5- ATRIBUTOS SELECIONADOS REFERENTE ÀS CIRURGIAS.	53
TABELA 6- CONJUNTO DE DADOS EXCLUÍDOS.	55
TABELA 7- RESULTADOS DE PENALIZAÇÃO NAS VÁRIAS ITERAÇÕES.....	63
TABELA 8- RESULTADOS DE PENALIZAÇÃO NAS VÁRIAS ITERAÇÕES DA SEGUNDA SOLUÇÃO INICIAL.	65
TABELA 9- RESULTADOS DE PENALIZAÇÃO NAS VÁRIAS ITERAÇÕES DO SIMULATED ANNEALING.	67
TABELA 10- RESULTADOS DE PENALIZAÇÃO NAS VÁRIAS ITERAÇÕES DO SIMULATED ANNEALING PARA SEGUNDA SOLUÇÃO.....	68
TABELA 11- RESULTADOS DE PENALIZAÇÃO NAS VÁRIAS ITERAÇÕES DO EVOLUTIONARY ALGORITHM.	70
TABELA 12- RESULTADOS DOS DIFERENTES ALGORITMOS EM COMPARAÇÃO COM O RESULTADO OBTIDO PELO CENTRO HOSPITALAR.	71

1. Introdução

Neste capítulo encontra-se a contextualização do projeto desenvolvido, apresentando um enquadramento do tema e identificando a motivação para a realização da dissertação. Por fim, explicita quais os objetivos pretendidos para o projeto e como será a organização do documento, através de uma breve descrição de cada capítulo.

1.1 Enquadramento e Motivação

Atualmente, a prestação de serviços hospitalares depende da execução de um elevado número de processos, que inclui um conjunto de atividades clínicas ou administrativas, bem como diferentes recursos, como médicos ou enfermeiros. No contexto atual, como resposta à modernização de vários setores na área da Saúde, pretende-se minimizar ou maximizar determinados objetivos através do apoio à decisão, e com a evolução dos sistemas computacionais, é possível resolver esses mesmos problemas de forma mais simples, permitindo uma melhor eficiência e eficácia da decisão tomada, poupando recursos humanos e financeiros.

Com o aumento da importância dos sistemas de apoio à tomada de decisão, surge também a utilização dos sistemas *ABI* enquanto apoio à decisão, uma vez que permite prever, otimizar e adaptar a mudanças externas. Este tipo de sistema mostra-se mais eficiente que os sistemas *BI*, uma vez que estes apenas garantem o apoio à decisão, sendo incapaz de obter soluções com melhor qualidade e com algoritmos inteligentes que adotem as tomadas de decisão ao contexto real.

Este projeto será realizado em parceria com o *CHUP*, com o objetivo de introduzir novas soluções no que diz respeito a alocação de cirurgias em meio hospitalar. O objetivo centra-se na possibilidade de prever a execução dos vários eventos, assim como determinar os recursos necessários. Complementarmente, espera-se recorrer a técnicas de otimização modernas para desenvolver modelos que permitam estabelecer agendamentos mais eficientes, procurando a minimização de desperdícios do tempo de duração total dos turnos necessários, evitando os atrasos que são gerados pela falta de aproveitamento de turnos e minimização de penalizações associadas a cirurgias que ultrapassaram o *TMRG*. As motivações associadas a esta dissertação incidem no interesse pela área dos *SAD*, *DM* e Otimização. A

possibilidade de trabalhar na área da saúde e ajudar os profissionais a tomar decisões mais informadas foi uma motivação adicional ao desenvolvimento deste estudo.

1.2 Objetivos e Resultados Esperados

Nesta fase são identificados os objetivos que ajudam à resolução de problemas. Estes foram deduzidos racionalmente a partir da especificação do problema. O principal objetivo desta dissertação é **o auxílio da tomada de decisão relativamente ao processo de agendamento do CHUP.**

Para atingir esse objetivo, utilizam-se dados reais fornecidos pela mesma organização, sendo apenas utilizados no decorrer do projeto e conforme os critérios de confidencialidade. Assim, formularam-se as seguintes questões de investigação desta dissertação:

- **Qual o contributo de um modelo inteligente de otimização sobre o processo de agendamento de cirurgias no CHUP?**

Associados aos objetivos principais, existem objetivos intermédios que devem ser atingidos, tais como:

- Levantamento de conceitos relacionados com *ABI*, especificando modelos de otimização e casos de estudo similares à solução desta dissertação;
- Compreensão dos dados clínicos;
- Definir os critérios de eficiência de agendamento e taxa de utilização de turnos;

Com o desenvolvimento desta dissertação pretendem-se alcançar os seguintes resultados:

- Aquisição de um conhecimento mais aprofundado sobre os sistemas *ABI*;
- Obter conhecimentos claros sobre práticas clínicas e de gestão por parte das organizações que atuam na área da saúde;
- Obter os modelos de otimização que mais benefícios trazem para o *CHUP*.

1.3 Abordagens Metodológicas

Este capítulo abrange a descrição das metodologias utilizadas no desenvolvimento deste projeto. Para a fase de investigação utilizámos a metodologia Design Science Research Methodology (DSRM). Esta metodologia é uma *framework*, cujo principal objetivo é o desenvolvimento de artefactos que sejam utilizados para aplicação por profissionais da área de Sistemas de Informação (Peffers et al., 2007). Enquanto para a fase de desenvolvimento escolhemos utilizar a metodologia de Cross Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al., 2000) é uma estrutura criada para problemas de *DM* com vista à descoberta de conhecimento através da aplicação de algoritmos de *ML*.

1.3.1 Design Science Research

Para o desenvolvimento deste trabalho, será utilizada como metodologia de investigação a *DSRM*. Esta envolve um vasto conjunto de técnicas e perspetivas que conduzem a pesquisa científica nos sistemas de informação. Concentra-se no desenvolvimento e no desempenho de artefactos, com a intenção de melhorar o desempenho funcional destes na área de Sistemas de Informação. A Figura 1 representa as diferentes fases da metodologia.

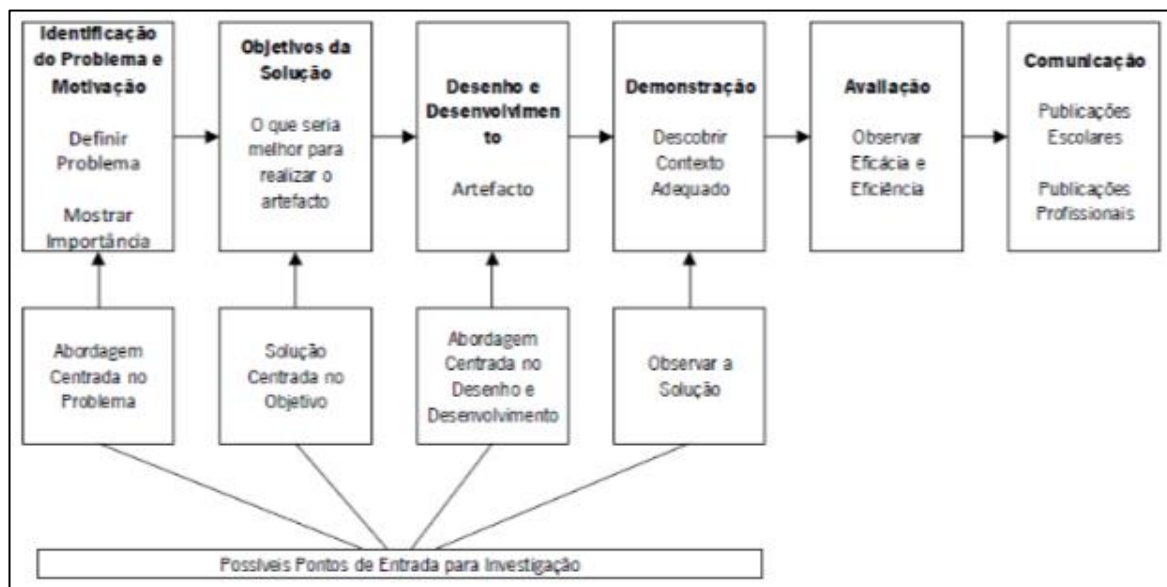


Figura 1- Fases da DSRM. Adaptado de (Peffers et al., 2007).

As fases da *DSRM* apresentam as seguintes características:

- **Identificação do Problema e Motivação:** Consiste em elaborar uma procura do máximo de informações possíveis e a solução do problema. Neste projeto, reconhece-se a não utilização, por parte das instituições hospitalares e outras instituições de saúde, de um sistema ABI. A motivação centra-se no desenvolvimento de um projeto que poderá potenciar o uso de este tipo de sistemas com a otimização de cirurgias em meio hospitalar;
- **Objetivos da solução:** Identificação dos objetivos que ajudam na resolução do problema. É espectável a criação de um artefacto para suportar a solução de problemas que até ao momento não tenham sido abordados. O conhecimento do estado do problema e as suas soluções atuais e a sua eficácia são os recursos requeridos para esta fase;
- **Desenho e desenvolvimento:** Permite determinar a funcionalidade desejada do artefacto e a sua arquitetura, para posteriormente criar o artefacto real. De modo a transitar para esta fase é necessário incluir o conhecimento da teoria que será utilizada na solução. Realiza-se o levantamento de técnicas, algoritmos e metodologias pertinentes para a execução do projeto;
- **Demonstração:** Consiste na demonstração da eficácia do artefacto na resolução de problemas. É feita a validação da solução desenvolvida. Os recursos necessários para a demonstração incluem o conhecimento efetivo de como usar o artefacto na resolução do problema.
- **Avaliação:** Esta fase tem como objetivo observar e medir corretamente a forma como o artefacto suporta a solução do problema. Após a plataforma desenvolvida, será feita uma avaliação de acordo com determinados critérios, seguindo um conjunto de métricas definidas, ou desenvolvendo-as através dos dados utilizados, por forma a verificar se esta plataforma vai ao encontro dos requisitos exigidos na proposta inicial.
- **Comunicação:** Última fase, dedicada à divulgação dos resultados que foram obtidos, bem como a importância e utilidade do trabalho. Esta fase encontra-se extremamente dependente dos resultados da fase anterior.

Primeiramente, o problema identificado está relacionado com as penalizações sofridas pelo hospital consequentes das cirurgias agendadas e realizadas para além do TMRG definido. O objetivo da conceção deste artefacto, passa exatamente por minimizar o desperdício de cada turno. Esta otimização é feita através da escolha, por parte do sistema, da combinação que melhor aproveita cada turno, diminuindo-se, assim, os desperdícios.

Esta solução evitará que as futuras cirurgias sejam atrasadas e sofram penalizações em consequência desse desperdício, precisamente pelo facto da disponibilidade de cada turno ser aproveitada ao máximo possível. Este resultado é obtido através do uso do método de otimização moderna, que visa gerar soluções capazes de auxiliar os profissionais na tomada de decisão do agendamento.

Assim, neste trabalho, foram implementados três métodos de otimização moderna no contexto do estudo descrito, sendo que dos vários testes realizados, foram escolhidos o Hill Climbing, o Simulated Annealing e *Evolutionary Algorithm*, por minimizarem o valor pretendido.

Uma vez identificado o problema e os objetivos do projeto, inicia-se a fase de desenvolvimento do artefacto, através dos três métodos escolhidos para solucionar o problema de agendamento de cirurgias. Para esta fase, foi necessário solucionar a representação da solução e a solução inicial, a função objetivo, a função de avaliação e as suas restrições ao problema em questão.

Neste projeto serão analisados as várias funções criadas e heurísticas desenvolvidas para validar a eficácia do artefacto, depois será procedido a uma fase de testes para verificar se é possível responder aos requisitos propostos, para isso serão analisados os resultados dos vários modelos gerados e comparado com os resultados CHUP, desta forma poderemos verificar se nos foi possível minimizar os desperdícios dos turnos, caso exista uma redução real dos mesmos.

Por fim, em relação à fase de comunicação para este projeto será desenvolvido o relatório de dissertação, de modo a mostrar todo o processo bem como todo o conhecimento adquirido.

1.3.2 Cross Industry Standard Process for Data Mining

CRISP-DM é uma metodologia que facilita os projetos de *DM* e a condução de todo o processo de extração de conhecimento. Esta metodologia tem como finalidade melhorar o desenvolvimento da plataforma, através de um aumento de segurança, exequibilidade e viabilidade dos projetos, assim como através da uma diminuição dos custos de execução (Santos & Azevedo, 2005). Assim, é possível inserir esta mesma metodologia dentro da fase de desenho e desenvolvimento da metodologia de investigação utilizada para a realização desta dissertação, *DSRM*. O *CRISP-DM* é constituído por seis fases, organizadas de maneira cíclica, conforme apresentado na Figura 2 (Chapman et al., 2000). O fluxo de execução destas fases não é unidirecional.

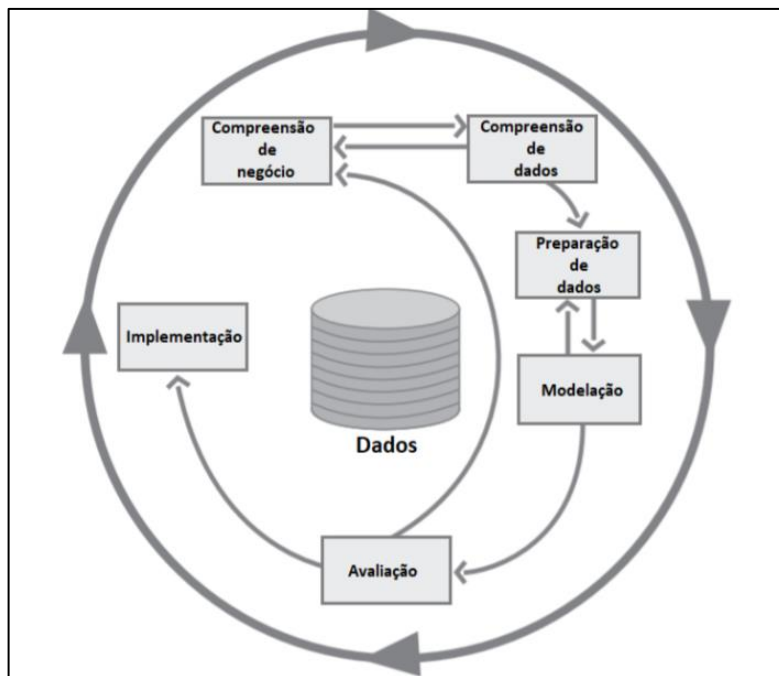


Figura 2- Modelo de processos do CRISP-DM. Traduzido e adaptado de (Chapman et al., 2000).

As fases da *CRISP-DM* apresentam as seguintes características:

- **Compreensão de negócio:** Consiste em perceber os objetivos e requisitos do negócio, de modo a compreender a informação com base no conhecimento que existe na definição do problema de *DM* existente;

- **Compreensão dos dados:** Abrange a recolha de dados e determina a exploração e verificação da qualidade dos mesmos. Esta fase tem como principal objetivo obter as informações ocultas através da perceção dos dados;
- **Preparação dos dados:** Compreende todas as atividades necessárias para construir o conjunto de dados final, que será usado na etapa de modelação. Este conjunto de dados será sujeito a várias tarefas de limpeza, seleção e transformação dos dados, até que estes apresentem a qualidade necessária;
- **Modelação:** Consiste na seleção e aplicação de técnicas de *DM*, bem como a configuração de parâmetros para obter a otimização de dados. Nesta fase, há que considerar uma adequação das técnicas ao problema existente;
- **Avaliação:** Nesta fase será avaliada a utilidade do método, de forma a verificar se os objetivos de negócio foram atingidos;
- **Implementação:** Consiste em avaliar e organizar os resultados obtidos, de maneira a possibilitar que o conhecimento extraído seja disponibilizado ao utilizador. Dependendo dos requisitos do projeto, poderá apenas ser necessário a elaboração de um relatório ou proceder-se à implementação, monitorização e manutenção do modelo desenvolvido.

1.4 Estrutura do documento

Este documento encontra-se dividido em 5 capítulos:

- **Capítulo 1:** Enquadramento do tema e identificada a motivação para a realização da dissertação, de acordo com as abordagens metodológicas estabelecidas;
- **Capítulo 2:** Revisão de literatura e estado de arte nos termos que foram abordados ao longo do desenvolvimento desta dissertação;
- **Capítulo 3:** Trabalho desenvolvido no âmbito do objetivo central deste estudo, com a descrição de todas as fases de implementação;
- **Capítulo 4:** Resultados obtidos e posterior discussão sobre os mesmos;

- **Capítulo 5:** Síntese geral do trabalho realizado, apresentando as conclusões mais relevantes do trabalho desenvolvido.

2. Estado de Arte

Neste capítulo pretende-se apresentar um enquadramento sobre os temas, envolvendo conceitos associados a sistemas *BI* e *ABI*, a relação dos mesmos na área da Saúde, com enquadramento da aplicabilidade dos modelos de previsão e otimização no estudo em questão. O estado de arte tem como principal objetivo clarificar a área de aplicação desta mesma dissertação, bem como os conceitos e metodologias que são úteis para a fundamentar. Este capítulo será também usado como diretriz para o desenvolvimento do projeto, melhorando o desempenho e a qualidade da dissertação.

Desta forma, é utilizada literatura proveniente da comunidade científica, com publicações relacionadas a esta temática, de modo a estabelecer um enquadramento sobre as diferentes perspetivas atualmente existentes.

2.1 Sistemas de Informação

Um sistema de informação pode ser perspetivado como um conjunto organizado de elementos, incluindo pessoas, dados, atividades e tecnologias da informação. Os sistemas de informação têm vindo a ganhar um papel progressivamente maior em diversos tipos de organizações com a informatização das infraestruturas e à digitalização da informação (O'brien, 2006).

De modo a compreender os Sistemas de Informação, é importante definir os conceitos básicos, tanto de sistema como de informação. Informação pode ser definida como o processamento ou organização dos dados de tal forma que represente uma modificação no conhecimento das pessoas que os vão receber. Cada conjunto de dados pode ser apresentado de maneira útil para uso no processo de tomada de decisão. Para proporcionar um melhor entendimento, Goldschmidt and Passos (2005) destacam as diferenças e a hierarquia dados, informação e conhecimento. Os dados, que se encontram na posição inferior da hierarquia, podem ser interpretados como factos elementares, recolhidos e armazenados através de tecnologias de informação. A informação caracteriza-se por ser entendida como objetos simbólicos construídos deliberadamente para possibilitar a comunicação e a formação de conhecimento. Numa posição superior da hierarquia encontra-se o conhecimento, podendo ser também designado com padrão ou conjunto de padrões, cuja formulação pode envolver

e relacionar dados e informações. Em geral, o conhecimento não pode ser extraído das bases de dados por recursos tradicionais da tecnologia da informação (Buckland, 1991).

Sistema é o resultado da visualização do mundo ativo a partir de um certo ponto de vista. Os sistemas podem dividir-se em dois subgrupos, os sistemas simples e os sistemas mais complexos. Por um lado, os sistemas simples são apenas objetos ativos sem autocontrole e sem capacidade de cognição. Adicionalmente, os sistemas mais complexos têm algum tipo de controlo sobre o que fazem. Eles também são capazes de aprender e tomar decisões e podem alcançar um alto nível de autonomia, ou seja, podem ter existência independente. No caso dos sistemas mais complexos, estes necessitam de conhecimento e realizam várias operações com ele (Carvalho, 2000).

Com base nas definições de informação e sistema fornecidas, é possível concluir que existem três interpretações possíveis. A primeira interpretação é focalizada a natureza dos objetos processados. Sistemas de informação são sistemas que processam apenas informações, isto é, objetos ou representações simbólicas. Objetos materiais e energéticos não são considerados relevantes. Objetos processados podem ser afetados na sua forma, espaço e tempo. No caso de informação, as operações possíveis sobre ele podem incluir: calcular, processar, criar, codificar, recolher, apresentar, transmitir, armazenar, memorizar. Portanto, qualquer objeto ativo que realize qualquer uma dessas operações poderá ter o direito de ser considerado uma informação. A segunda interpretação concentra-se no objetivo do sistema. Sistema de informação é um sistema cujo objetivo é informar, isto é, contribuir para a aquisição de conhecimento de alguém. Esse conhecimento é necessário para a execução de alguma ação em algum contexto. Também é possível considerar que um sistema de informação é um objeto ativo que lida com “processos” e cujo objetivo é informar. A terceira interpretação é uma combinação das outras duas. A existência de tal sistema é muito plausível como um sistema que informa os seus clientes/utilizadores tem que lidar com informações. No entanto, deve-se notar que nem todos os sistemas que lidam com informações têm a finalidade de informar. As companhias de seguros e os bancos são dois exemplos de tais sistemas (Carvalho, 2000).

2.2 Sistemas de Informação na área da Saúde

A importância atribuída à área da Saúde e os custos acrescidos na gestão desta área contribuíram para o surgimento do interesse neste estudo como uma importante área num futuro próximo. Os *SI* têm muito a oferecer na gestão de custos de assistência médica e na melhoria da qualidade dos cuidados (Kolodner et al., 2008). Além do papel incorporado das *TI* em equipamentos clínicos e de diagnóstico, os *SI* estão posicionados de maneira única para capturar, armazenar, processar e comunicar informações oportunas para a tomada de decisão, para uma melhor coordenação da assistência médica, tanto individual quanto profissional a níveis populacionais (Fichman, Kohli, & Krishnan, 2011).

A área da saúde gera grandes volumes de dados de hospitais, cirurgias, clínicas e laboratórios. Ainda assim, alguns processos são feitos manualmente, apesar de décadas de experiência de outras aplicações a usar *TI* dentro de indústrias com grandes quantidades de informação. Este atraso na digitalização da área deveu-se a um fraco investimento na área de *TI* numa fase inicial, falta de vontade política, mercados fragmentados com fontes de receita inadequados para apoiar o desenvolvimento de novos sistemas e falta de padrões. Além disso, existem alguns desafios específicos relacionados ao uso da *TI* em saúde, como a complexidade de dados médicos, preocupações de segurança e confidencialidade, bem como a ausência em muitos países de um identificador único do paciente e a falta de conhecimento geral dos benefícios e riscos da *TI* (Grimson, Grimson & Hasselbring, 2000).

No século XXI, cada organização de Saúde construiu a sua própria infraestrutura de dados de modo a suportar as suas próprias necessidades, mas os processos desta área estão sujeitos a mudanças, tanto devido a causas internas (e.g. novos procedimentos diagnósticos ou terapêuticos, mudança da estrutura departamental) quanto a causas externas (e.g. introdução de *DRG*, pressão econômica em relação a fusões hospitalares e redes integradas de assistência médica) que impõem mudanças no processo e criam necessidade de adaptar rapidamente as informações das infraestruturas criadas. Assim, é necessária uma infraestrutura de *TI* responsiva para poder rapidamente e adequadamente reagir às novas necessidades. Isso inclui a capacidade de uma infraestrutura de *TI* para adicionar novas funcionalidade com risco mínimo (Lenz & Kuhn 2004).

A importância que a área da saúde apresenta no dia-a-dia das pessoas proporciona um aumento do número de erros cometidos. Os erros médicos aumentam o tempo de

permanência no hospital e custam vidas humanas (Classen et al., 1997). A nível populacional, uma falha a controlar doenças infecciosas pode ameaçar fortemente a saúde pública. Um *SI* pode facilitar a busca, destacando e monitorizando os erros em vários estágios ao longo do atendimento contínuo (Fichman, Kohli, & Krishnan, 2011). Os hospitais tendem a desenvolver programas inovadores e tomar decisões estratégicas sobre o seu futuro, a qualidade e a disponibilidade de informações, que continuam a ganhar uma maior importância. As influências ambientais bastante mutáveis que acompanharam a introdução abundante de múltiplas TSI tornam o setor de saúde uma área importante para examinar o uso estratégico dos *SI* (Enthoven, 1988).

2.3 Business Intelligence

2.3.1 Introdução

Nos inícios de 1980, o conceito de *EIS* emergiu (Efraim Turban et al., 1995). Os *EIS* são definidos como um sistema que fornece aos executivos acessos fáceis a informações internas e externas e que sejam relevantes para os fatores críticos de sucesso (Watson, Rainer, & Koh, 1991). No início dos anos 90, Howard Dressner (1989) definiu o termo *BI*, agora amplamente utilizado para descrever aplicações analíticas. Em 2005, alguns sistemas *BI* começaram a incluir capacidades, integrando competências de *IA* com capacidades analíticas bastante eficazes (Efraim Turban et al., 1995). Hoje em dia, um *SI* empresarial baseado em *BI* contém toda a informação relevante para as necessidades dos gestores (Stackowiak, 2007). Santos (2005) entende que os sistemas *BI* asseguram um conjunto de técnicas e ferramentas que são propostas para auxiliar na administração de um negócio e na tomada de decisões. É proposto uma transformação dos dados em informações utilizáveis para maior efetividade estratégica, insights operacionais e benefícios reais para o processo de tomada de decisão nos negócios. O *BI* inclui atualmente vários componentes para execução de *ETL*, *DW*, análise de dados e processamento *OLAP* e *DM* (Ranjan, 2009). Estes sistemas vão permitir que as organizações tirem o melhor partido dos dados armazenados, tanto dados atuais como históricos, facilitando a gestão e o planeamento do negócio.

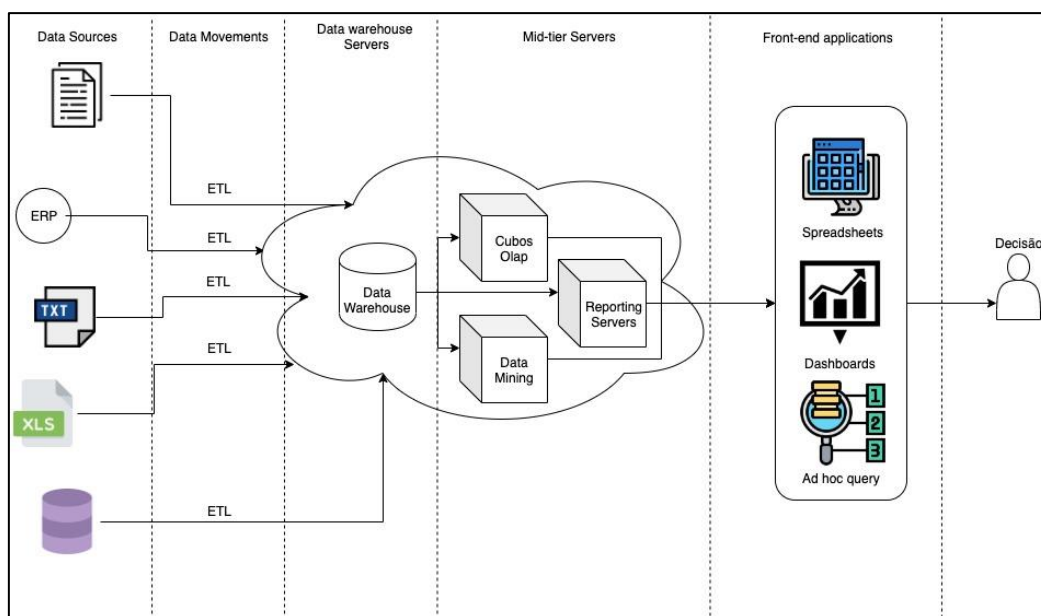


Figura 3- Arquitetura de um sistema de BI. Adaptado de (Chaudhuri, Dayal e Narasayya, 2011).

A Figura 3 fornece a estrutura principal de um sistema BI, de modo que seja possível fornecer o apoio necessário à tomada de decisão. Esta estrutura apresenta o processo inicial de extração de dados, onde posteriormente serão efetuados os tratamentos necessários aos dados utilizados. Os dados são estruturados para que facilmente sejam atualizados e acedidos. No entanto, é necessário ter outras estruturas para efetuar as análises necessárias para os gestores do negócio, como servidores *OLAP*, servidores de relatórios e mecanismos de *DM*. O acesso pode ser efetuado de diferentes formas possibilitando aos gestores do negócio visualizar os dados sob diversas formas. Todo este processo vai permitir uma tomada decisão muito mais informada e com muito menos risco por parte dos gestores.

Em suma, após analisada a aplicação de um sistema *BI*, é possível concluir que estes sistemas possuem uma aplicabilidade extensa em contexto organizacional, nos ambientes internos e externos de negócios (Wu, 2000). A Tabela 1 apresenta alguns dos principais benefícios dos sistemas *BI*, relacionando estes com as suas principais referências.

Tabela 1- Principais benefícios de Business Intelligence. Retirado de (Singh & Samalia, 2014).

Principais categorias de benefícios	Autores
Business reporting mais rápidos e precisos	Pellissier & Kruger (2011); Đekić & MladenovićRanisavljević (2010).

Processo de tomada de decisão melhorado	Hočevar and Jaklič (2010); Carver and Ritacco (2006)
Satisfação do cliente melhorada	Hočevar and Jaklič (2010); Carver and Ritacco (2006); Atre & Moss (2003)
Redução em custos	Carver and Ritacco (2006); Atre & Moss (2003)
Aumento de receita	Carver and Ritacco (2006); Atre & Moss (2003)
Qualidade de informação e de comunicação melhorada	Popovič, Turk and Jaklič (2010); Yeoh, Koronios & Gao (2008); Carver and Ritacco (2006);
Aumento na quota de mercado	Atre & Moss (2003)

OLAP é um método poderoso integrado nos sistemas de *BI* que analisa grandes quantidades de dados de um *DW*. Caracteriza as operações de consolidar, exibir, aplicar fórmulas e sintetizar dados em várias dimensões. Ajuda analistas e gestores a compreender o desempenho da organização, mediante uma ampla variedade de visualizações de dados organizadas para refletir a natureza multidimensional dos dados organizacionais. Os cubos *OLAP* são pré-calculados, que resultam num desempenho de *query* rápido, essencial para o uso executivo (Kimball & Ross, 2011). Segundo (Sarawagi, Agrawal, & Megiddo 1998), um cubo *OLAP* consiste em dois tipos de atributos: medidas e dimensões. O conjunto de dimensões consiste em atributos como nomes de produtos e nomes de lojas que juntos formam uma chave. As medidas são geralmente atributos numéricos, como volumes de vendas e lucro. As dimensões geralmente têm associadas hierarquias que especificam níveis de agregação. Todos os modelos multidimensionais são constituídos por tabelas de facto, que possuem então uma chave primaria composta pelas chaves estrangeiras de dimensões relacionadas com a tabela, e pelas medidas que cada entrada regista.

Relativamente ao modelo físico, existem duas principais visões: a arquitetura *ROLAP* e *MOLAP*. A *MOLAP* apresenta a vantagem de apresentar uma visão multidimensional direta dos dados, ao passo que a *ROLAP* é apenas uma *interface* multidimensional para os dados relacionais (Vassiliadis & Sellis 1999). Posteriormente, surgiu a arquitetura *HOLAP*, que é um sistema híbrido de *OLAP*, combinando *ROLAP* e *MOLAP*. Este tipo de arquiteturas pode conter dados em tempo real de um passado recente, em volumes reduzidos, e dados históricos que podem ser guardados em grandes volumes. A principal desvantagem é a redundância, uma vez que podem existir várias fontes para a mesma pesquisa. Existem várias operações analíticas de cubos *OLAP*. Estas permitem que perspectivas e vários níveis de detalhes sejam materializados explorando dimensões e suas hierarquias. Permite ainda fornecer um ambiente de análise de dados interativo. Cada operação efetuada no cubo altera a composição original do mesmo, sendo os mais comuns: *Roll-up*, *Drill-down*, *Slice and Dice*, *Pivot* e *Drill-across*.

Os *Dashboards* associam-se aos restantes componentes cruciais num sistema *BI*. Na década de 1990, Kaplan e Norton (1997) introduziram o *Balanced Scorecard*, uma metodologia de medição e gestão de desempenho, com ênfase nos principais *KPI's* como forma de fazer medições do estado do negócio. Definem-se como uma exibição visual das informações mais importantes, necessárias para alcançar um ou mais objetivos, consolidados e organizados, para que as informações possam ser monitorizadas rapidamente, facilitando a tomada de decisão. Os resultados desta ferramenta podem exibir métricas únicas, análises gráfica de tendências, medidores de capacidades, mapas geográficos, comparações de variações, entre outros (Few, 2006). O seu valor real reside na capacidade de substituir as técnicas de extração de dados *hunt-and-peck* por mecanismos adaptáveis, transformando repositórios de dados em informações de consumo (Hovis, 2002). Estas ferramentas ajudam as pessoas a identificar visualmente tendências, padrões e anomalias, de forma a perceber sobre o que veem e ajudar a orientá-las em direção a decisões efetivas (Brath & Peters, 2004). Os *KPI's* adquirem medidas de desempenho, úteis para realizar uma avaliação do que se passa na organização e definir ou redefinir metas.

2.3.2 Extração, Transformação e Carregamento de dados (ETL)

O processo de *ETL* é a base dos sistemas *BI*. Neste processo, as diversas fontes externas são acedidas, sendo efetuada a leitura e cópia dos dados para o sistema *ETL*, onde são

efetuadas todas as transformações necessárias. Como os dados provêm de diferentes fontes, é necessário efetuar várias transformações, entre elas, a limpeza de dados (e.g. correção de erros ortográficos, falta de elementos), a combinação dos vários dados e a verificação de erros estruturais. O carregamento dos dados é a fase em que a informação é colocada disponível para o utilizador. Em resumo, o processo *ETL* engloba a transferência de dados para um *DW*, de acordo com os requisitos estruturais previamente definidos, de forma a uniformizar a interpretação e implementação do *DW* (Chaudhuri et al. 2011). Segundo (Jun et al. 2009) a forma mais precisa e eficiente de desenhar o processo *ETL* contempla a divisão do processo em cinco módulos: Extração, Validação, Limpeza, Conversão e Carregamento. A Figura 4 representa a integração destas etapas no processo *ETL*.

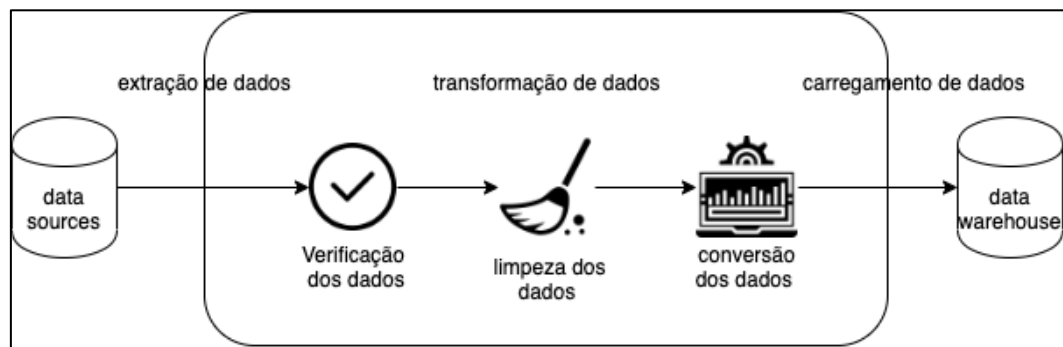


Figura 4- Processo de ETL. Adaptado por (Jun et al., 2009).

Estas fases representam um conjunto de características (Jun et al., 2009):

1. **Extração de dados:** Requer uma análise do negócio, do sistema e dos dados. É necessário compreender o tipo de tabelas que a base de dados possui e o correspondente significado de cada estrutura implementada;
2. **Validação de dados:** Envolve essencialmente tarefas de verificação da qualidade dos dados. Requer uma análise precisa quanto ao valor efetivo que os dados oferecem, bem como nas relações que posteriormente vão desenvolver;
3. **Limpeza de dados:** Esta tarefa resume-se em filtrar os dados indesejáveis. Estes incluem: dados incompletos, errados e duplicados;
4. **Conversão de dados:** Envolve tarefas finais que representem o desenvolvimento de mapeamentos, operações e conversões de dados;

5. **Carregamento de dados:** Última etapa do processo *ETL*, do qual os dados são transferidos para o *DW*. A implementação depende do tipo de operação e da quantidade de dados.

2.3.3 Data Warehouse

A definição de *DW* universalmente aceite foi desenvolvida na década de 1980 por Inmon (2002) como “*a subject-oriented, integrated, time variant and nonvolatile collection of data used in strategic decision making*”. Segundo Imhoff, Galemme, and Geiger (2003), o papel principal de um *DW* é servir como um repositório de dados que armazena dados de fontes diferentes, tornando-os acessíveis a outro conjunto de armazenamentos de dados, denominados *Data Mart*. A Figura 5 apresenta uma arquitetura genérica de um *DW*.

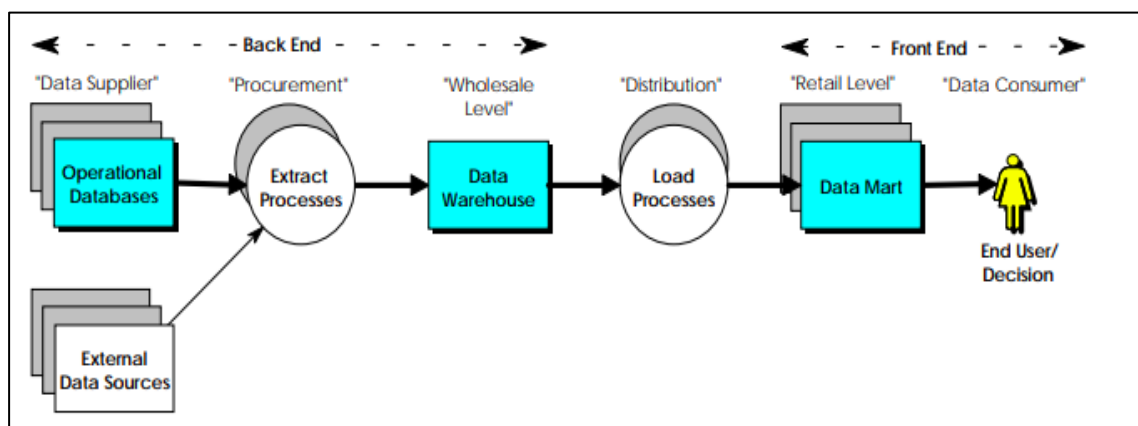


Figura 5- Arquitetura de DW. Retirado de (Moody, 2000).

Esta arquitetura consiste nos seguintes componentes (Moody, 2000):

1. **Sistemas operacionais**, que dão suporte às atividades diárias da empresa e que guardam os dados relativos ao negócio;
2. **Fontes externas de dados**, que apoiam a análise com *DW* incorporados, com dados de várias fontes externas;
3. **Processos de extração**, com processos que armazenam os dados no *DW* de uma forma regular. Os dados, extraídos de várias fontes, são depois armazenados num formato consistente;
4. **DW central**, integrado como um componente principal, utilizado para fornecer *Data Marts*;

5. **Processo de carregamento**, que contempla o processo de distribuição dos dados do *DW* central para os *Data Marts*. A infraestrutura de entrega de dados permanece razoavelmente estática, no entanto, estes necessitam de acompanhar constantemente as mudanças nas necessidades de informações comerciais;
6. **Data Marts**: fornecem dados com que os utilizadores finais consigam trabalhar. São implementados através de bases de dados tradicionais ou ferramentas *OLAP*. *Data marts* são um subconjunto de dados do *DW* e são onde a maioria das atividades analíticas ocorrem no ambiente de *BI* (Imhoff, Gallemmo, & Geiger, 2003)

À luz das observações anteriores, podemos definir um *DW* como um principal repositório, contendo dados disponíveis para o desenvolvimento de arquiteturas de *BI* e *SAD* (Vercellis, 2009). Estes propõem as seguintes características:

- **Orientado por assunto**, do qual o *DW* objetiva assuntos, ou seja, os dados são organizados por determinados tópicos de interesse de uma organização;
- **Integrado**, do qual o *DW* recebe os mais variados dados provenientes de diversas fontes. A integração é realizada visando padronizar os dados recolhidos das várias fontes numa representação única;
- **Variáveis temporais**, permitindo manter o histórico de dados durante um período de tempo superior aos sistemas transacionais. Os dados armazenados devem possuir informações sobre uma perspetiva histórica de forma a poder permitir a identificação de variáveis temporais para determinado tópico;
- **Não volátil**, do qual os dados são carregados para o *DW*, disponíveis para operações de consulta e exclusão.

2.4 Adaptive Business Intelligence

ABI é a disciplina que usa técnicas de previsão e otimização para criar sistemas *self-learning* de decisões, ou seja, pode ser caracterizado como a combinação entre o sistema de *BI*, *DM*, métodos de previsão, técnicas de otimização e adaptabilidade. Esta nova abordagem permite elaborar uma melhor recomendação baseado em dados armazenados anteriormente (Michalewicz, 2007).

Os processos presentes no sistema *ABI* são similares aos presentes no *BI*, tal como mostra a Figura 6.

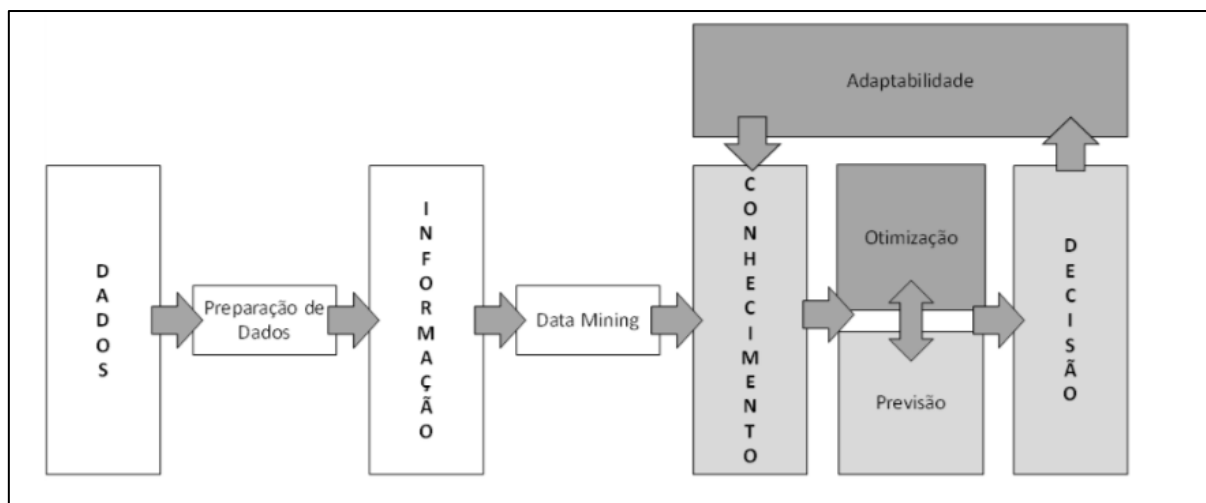


Figura 6- Processos de Adaptive Business Intelligence traduzido e adaptado de (Michalewicz, 2007)

Estes sistemas contemplam três módulos adicionais, comparativamente com os sistemas *BI* já apresentados: previsão, otimização e adaptabilidade. A previsão integra um conjunto de cenários que se tentam antecipar utilizando dados históricos. Com esta finalidade, os resultados obtidos com o processo analítico servem para construir um modelo capaz de prever alguma resposta através das relações e padrões que existem entre as variáveis nos dados. A otimização integra um conjunto de técnicas matemáticas, utilizadas no sentido de melhorar os resultados previstos no módulo anterior, minimizando erros preditivos. Por último, a adaptabilidade, que permite a utilização dos dados em tempo real para adaptar-se às mudanças organizacionais (Michalewicz, 2007).

De forma a ser possível um correto funcionamento do sistema é necessário (Michalewicz, 2007):

- Dados disponíveis devidamente preparados e analisados através de *DM*;
- Modelos de previsão desenvolvidos de acordo com critérios de sucesso previamente estabelecidos;
- Modelos de otimização integrados nos módulos de previsão, de forma a este recomendar a melhor solução;
- Adaptação dos modelos as mudanças organizacionais, utilizando dados em tempo real.

Os poderes dos sistemas ABI residem na sua capacidade de responder às duas questões fundamentais por trás de todas as decisões de negócios:

- 1. O que provavelmente acontecerá no futuro?**
- 2. Qual é a melhor decisão agora?**

As organizações que conseguirem responder com precisão a estas perguntas de maneira consistente terão uma vantagem competitiva sobre aquelas que não conseguem (Michalewicz, 2007).

2.4.1 Data Mining

DM surgiu na década de 1990 e pode ser vista como resultado da evolução natural das tecnologias de informação. O seu desenvolvimento está relacionado com o progresso do hardware nas ultimas três décadas, do qual os computadores ficaram mais baratos e aumentaram significativamente a performance, disponíveis para a sociedade em geral (Edelstein, 1999). Os avanços nos sistemas de gestão de base de dados aproveitaram esta evolução, recuperando o termo *DM* e como este pode beneficiar as organizações (Han, Kamber, and Pei, 2012).

DM é um processo que usa uma variedade de ferramentas de análise de dados para descobrir padrões e relacionamentos nos dados que podem ser usados para fazer previsões válidas. Pode também ser visto como a ciência que explora grandes conjuntos de dados, com o objetivo de extrair informações implícitas, anteriormente desconhecidas e potencialmente úteis (Gorunescu 2011). O processo de descoberta de conhecimento é constituído por uma sequência iterativa de etapas, também representadas na Figura 7 (Han, Kamber, & Pei, 2012):

- 1. Limpeza dos dados**, para remover ruído e dados inconsistentes;

2. **Integração dos dados**, onde múltiplas fontes de dados podem ser combinadas;
3. **Seleção dos dados**, do qual os dados relevantes são recolhidos da base de dados;
4. **Transformação dos dados**, onde os dados são transformados e consolidados em formas de acordo com os critérios definidos;
5. **Data Mining**, que constitui um processo essencial em que métodos inteligentes são aplicados para extrair padrões de dados;
6. **Avaliação de padrões**, do qual identifica os padrões verdadeiramente importantes baseado em critérios de avaliação;
7. **Apresentação do conhecimento**, onde a visualização e as técnicas de representação de conhecimento são usadas para apresentar os resultados aos utilizadores.

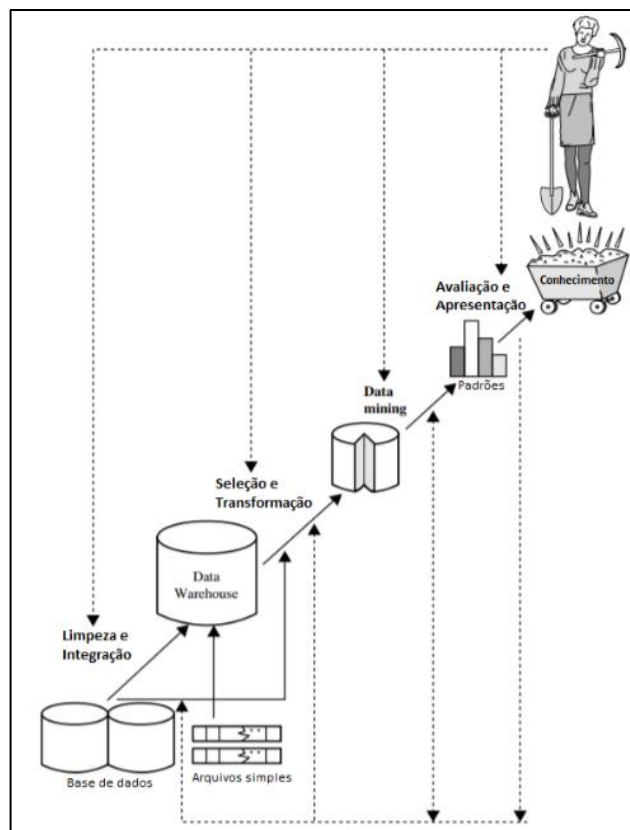


Figura 7- Processo de descoberta de conhecimento. Traduzido e adaptado de (Han, Kamber, & Pei, 2012).

DM permite a construção de um modelo para a descrição de padrões e relacionamentos nos dados utilizados, que podem ser usados com confiança para previsão.

Para tal, é importante ter em atenção a seguinte hierarquia de escolhas (Edelstein, 1999), presente na Figura 8.

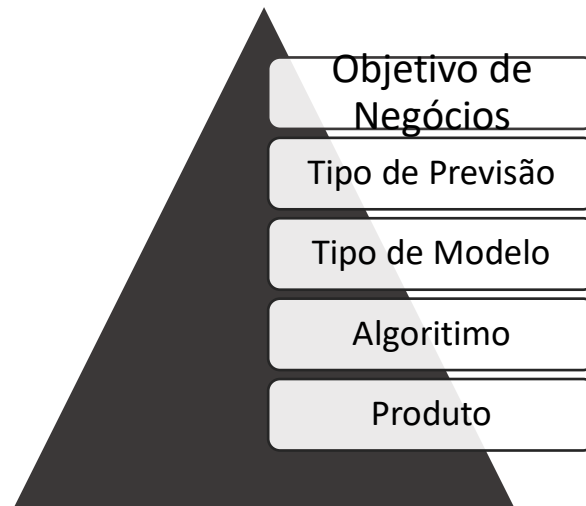


Figura 8- Hierarquia de escolhas de data mining adaptado de (Edelstein, 1999)

O objetivo de negócio é apresentado no nível mais alto, uma vez que o conhecimento das necessidades e objetivos da organização permite guiar na formulação do objetivo dos modelos. Em seguida, o próximo passo é decidir qual o tipo de previsão mais adequado. Por último, ao selecionar um produto de *DM*, é necessário ter em atenção que eles geralmente têm implementações diferentes de um algoritmo específico, mesmo quando o identificam com o mesmo nome. Essas diferenças de implementação podem afetar características operacionais, como uso de memória e armazenamento de dados, bem como características de desempenho, como velocidade e precisão.

2.4.2 Previsão

Os modelos de Previsão constituem técnicas e algoritmos para extrair conhecimento através dos dados existentes. Podem ser divididos em duas categorias: problemas de Classificação e problemas de Regressão, dependendo do tipo de variável que é esperado prever, discreta ou numérica. Aquando da introdução dos métodos de previsão numa destas três categorias é necessário ter em consideração o output esperado e o tempo. Os problemas de classificação normalmente apresentam dois possíveis outputs esperados, são problemas clássicos de classificação, pois eles atribuem novos casos a uma classe. Um problema de

classificação pode ter, no entanto, mais de dois outputs esperados. Nestes problemas, o tempo não existe. O “futuro” é entendido como a chegada de um novo caso ou é incluído como uma variável do caso. Os problemas de regressão têm como objetivo geral descobrir a relação entre várias variáveis independentes e uma variável dependente, com o output de saída a ser um número concreto. Um modelo de regressão também nos possibilita saber quais variáveis são melhores preditores, que são as variáveis independentes do modelo, e podemos facilmente identificar anomalias nas variáveis. Novamente, a questão do tempo é inexistente ou incluída como uma variável do caso (Michalewicz, 2007). O processo apenas em 3 passos (Michalewicz, 2007):

1. **Preparação dos dados:** Esta etapa inclui transformação de dados, normalização, criação de atributos derivados, seleção de variáveis, eliminação de dados ruidosos, fornecimento de valores ausentes e limpeza de dados. Esse estágio é frequentemente aumentado pela análise preliminar de dados para identificar as variáveis mais relevantes e determinar a complexidade do problema em causa;
2. **Criação do modelo:** Esta etapa inclui uma análise completa dos dados. A seleção do melhor método de previsão é com base na capacidade do modelo para explicar a variabilidade em questão, a produção de resultados consistentes e o desenvolvimento de um ou mais modelos de previsão;
3. **Avaliação e Deployment:** Esta etapa inclui a implementação do melhor modelo de previsão e a aplicação a novos dados para gerar previsões. Esta etapa é essencial para medir o desempenho do modelo de previsão e ajustá-lo ao de acordo com as mudanças no ambiente.

2.4.3 Otimização

Atualmente existe uma tendência crescente na adoção de ferramentas computacionais baseadas em métodos de otimização. Este é um tópico central do processo de *ABI*, que começou a ser desenvolvido na década de 1940, com introdução à programação linear. Ao longo dos anos, surgiram várias adaptações, desde otimização moderna (Michalewicz, 2007), metaheurísticas (Luke, 2012) ou métodos clássicos (Cortez, 2014). Qualquer que seja a técnica de otimização, existe sempre a necessidade de especificar três coisas:

1. A representação da solução;
2. O objetivo;
3. A função de avaliação.

A representação de uma solução irá determinar o espaço de procura por soluções, ou seja, o número de soluções possíveis para o problema, e o seu tamanho. Posteriormente, é necessário definir qual o objetivo do problema, essencialmente constituído por uma expressão matemática. Criar uma função de avaliação é a última etapa deste processo, permitindo comparar a qualidade de diferentes soluções (Michalewicz, 2007).

Segundo (Cortez, 2014), os métodos de otimização podem ser classificados de três formas que são:

- **Tipo de procura;**
- **Solução determinística ou estocástica**
- **Inspirado naturalmente.**

Para se perceber a classe de cada um dos métodos de otimização e ao grupo a que pertencem, a Figura 9 apresenta um conjunto de técnicas mais atuais, dentro das características já enunciadas:

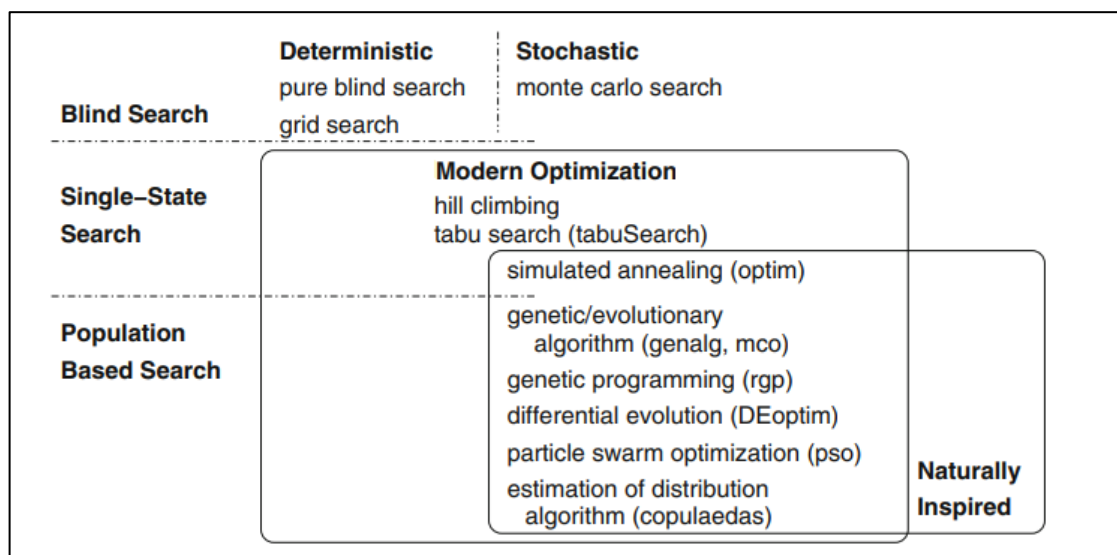


Figura 9- Classificação dos métodos de otimização. Retirado de (Cortez, 2014).

Em relação ao tipo de procura, este subdivide-se em três categorias, Cega, Local ou Baseada em População.

- 1. Procura Cega:** Assume o esgotamento de todas as alternativas, garantindo que todas as soluções são testadas. Só é admissível a espaços de procura discretos e possui uma fácil implementação. A maior desvantagem associada a esta técnica relaciona-se à viabilidade da mesma quando o espaço de busca é contínuo ou demasiado grande (Cortez, 2014). Tende a requerer mais esforço computacional, uma vez que a procura é efetuada através de um conjunto de soluções candidatas em vez de uma solução única (Sean, 2012).
- 2. Procura Local:** As técnicas mais modernas de otimização são baseadas em Procura Local. Tal como mostra a Figura 10, novas soluções são geradas a partir de soluções existentes, incluindo vários métodos que focam uma vizinhança local através de uma determinada solução inicial, ou seja, usam pesquisas anteriores para orientar a procura atual. (Cortez, 2014)

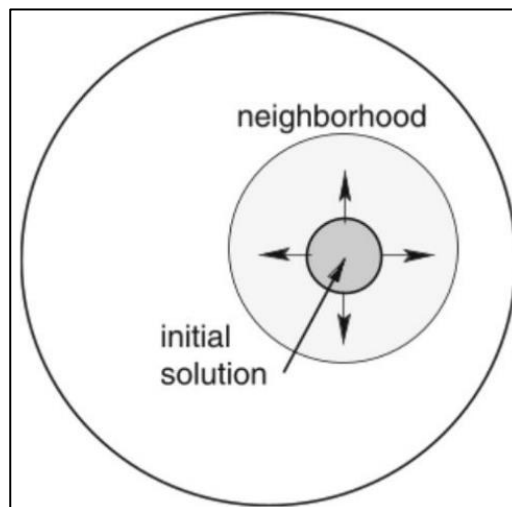


Figura 10- Procura por novas soluções em modelos de Procura Local. Retirado (Cortez, 2014)

- 3. Procura baseado na População:** Apresenta uma nova abordagem nos algoritmos de Otimização, em que utiliza um conjunto de soluções candidatas em vez de um único. Estes métodos por norma requerem um maior esforço computacional, quando comparados com modelos anteriores, em contrapartida a sua eficiência no reconhecimento de espaços de busca mais interessantes leva a que estes algoritmos sejam tomados em conta quando se procura melhores resultados. A Figura 11 explica

a exploração por regiões mais distintas no sentido de obter uma maior diversidade, com vista à definição de novas soluções. (Cortez, 2014).

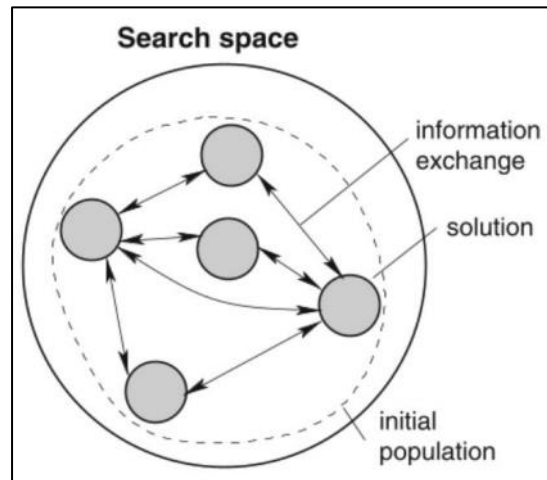


Figura 11- Procura baseada na População. Retirado de (Cortez, 2014)

Com estes aspetos resolvidos, as técnicas de otimização de pesquisa local e pesquisa baseada na população apresentam-se mais adequados para atingir o resultado pretendido. Ora, o Hill Climbing e Simulated Annealing são alguns dos vários métodos existentes de pesquisa local, dentro daqueles que são os métodos de otimização, sendo que os escolhemos. Em relação à pesquisa baseada na população escolhemos de entre os vários métodos, o *Evolutionary Algorithm*.

O método *Hill Climbing* é um método de pesquisa local, em que a solução satisfatória é alcançada através das soluções vizinhas, por comparação individual, de forma a verificar se a nova solução é a melhor. Este método tem um limite de execução através do critério de paragem da implementação (Michalewicz, 2007). Um método semelhante a este é o *Simulated Annealing*, inspirado no arrefecimento de metais, que envolve o arrefecimento controlado de determinado metal aquecido, através da refrigeração do mesmo. Este método difere da pesquisa do *Hill Climbing* por adotar um parâmetro que é usado para calcular a probabilidade de aceitar soluções inferiores (Cortez, 2014). Por último, o *Evolutionary Algorithm* aplica o fenómeno da seleção natural, ou seja, as soluções mais aptas sobrevivem, enquanto as soluções inadequadas morrem e não contribuem para o *pool* genético de gerações futuras (Cortez, 2014).

2.5 Adaptive Business Intelligence na saúde

Por forma a compreender melhor o funcionamento e a aplicabilidade que uma plataforma *ABI* pode ter na saúde, foram feitas várias pesquisas, nas quais podemos observar as técnicas e ferramentas utilizadas. Os seguintes temas tendem a relacionar a aplicabilidade destes estudos com o tema desta Dissertação.

2.5.1 Problema das restrições na especialidade cirurgia

Na área da saúde é frequente ocorrerem longos tempos de espera, cancelamentos e má gestão de recursos. Normalmente, para resolver este tipo de problemas, eram usadas técnicas tradicionais, como contratar mais profissionais ou comprar mais equipamentos, mas com a diminuição de orçamentos, esta abordagem tornar cada vez mais difícil. Em relação às cirurgias, o seu ambiente é uma das áreas mais caras de um hospital. Segundo (Macario et al., 1995), as despesas do setor cirúrgico representam quase 33% do orçamento hospitalar projetado, já que esta especialidade envolve custos dispendiosos (e.g. cirurgiões, anestesistas, enfermeiras) e custos de material, como camas e equipamentos cirúrgicos.

A utilização de um sistema capaz de otimizar o agendamento de cirurgias teria de permitir que os administradores hospitalares usem as salas de cirurgia da maneira mais eficiente possível, minimizando os custos operacionais, satisfazendo as necessidades dos pacientes e as preferências dos cirurgiões e levando em consideração as restrições de recursos humanos e materiais. A gestão de uma sala de operações consiste em duas fases: planeamento e agendamento de casos cirúrgicos. Inicialmente, o planeamento fornece a data da cirurgia para cada paciente, levando em consideração a disponibilidade de salas de cirurgia e cirurgiões. O agendamento diário determina a sequência de operações em cada sala de operação, levando em consideração a disponibilidade de recursos materiais e humanos (Meskens, Duvivier, and Hanset, 2013). A Figura 12 constitui um modelo composto pelas restrições que são normalmente encontradas em todas as salas de operações. De realçar o facto de uma operação ocorrer apenas uma vez, que duas operações não se sobreponem na mesma sala de operações e que um cirurgião não pode operar simultaneamente em várias salas e apenas um paciente é permitido numa sala de cirurgia ou num quarto de recuperação (Meskens, Duvivier, and Hanset, 2013). Assim, vários módulos estão conectados, cada um focado num aspeto muito específico:

- Programação do bloco cirúrgico;

- Operações;
- Recursos humanos, como cirurgiões, anestesistas e enfermeiros;
- Recursos materiais, como material cirúrgico e camas;
- Afinidades entre os funcionários.

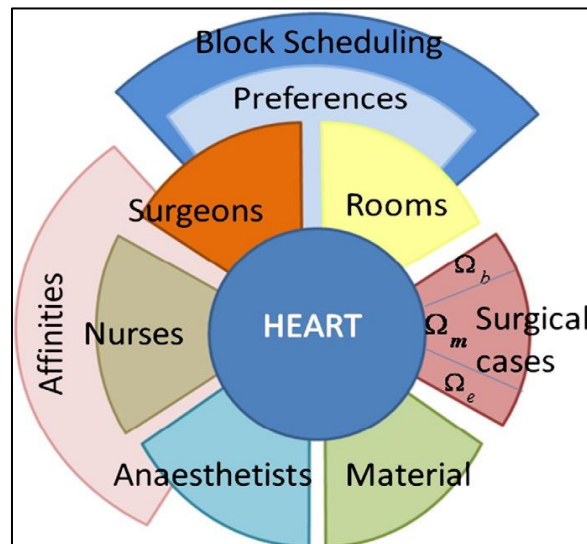


Figura 12- Modelo de restrições nas salas operatórias. Retirado de (Meskens, Duvivier, and Hanset, 2013)

O agendamento do bloco é uma estratégia usada para atribuir intervalos de tempo a todas as especialidades, de forma que os cirurgiões possam operar apenas nos horários atribuídos à sua especialidade. Além disso, na programação dos blocos, as salas de cirurgia são dedicadas a uma especialidade. Em relação às operações, pode-se considerar a existência de três níveis de prioridade: Alto, médio e baixo. Operações com alta prioridade são realizadas no início do dia. As operações com baixa prioridade são realizadas no final do dia e são principalmente casos de infecções que podem contaminar a sala, o que exige mais limpeza após a operação. Operações com prioridade média são todos os procedimentos que não têm requisitos especiais. As restrições relacionadas aos recursos humanos são divididas em vários módulos e variam de acordo com a categoria profissional (e.g. cirurgiões, enfermeiros, anestesistas). O módulo dos recursos materiais expressa a disponibilidade do material necessário para as operações. Incluem material específico para uma especialidade cirúrgica, bandejas cirúrgicas, carro cirúrgico, microscópios cirúrgicos, entre outros. Por último, existe o módulo das “afinidades” em que se tem em consideração, como o próprio nome indica,

afinidades entre os profissionais. Este módulo também tem em consideração incompatibilidades entre certas pessoas, bem como as preferências dos cirurgiões em trabalhar com um anestesista ou enfermeiro específico.

2.5.2 Caso de Estudo Relacionado

Neste ponto, apresento um estudo realizado por (Ferreira et al., 2019), intitulado como *“Adaptive Business Intelligence in Healthcare - A Platform for Optimising Surgeries”*. A Plataforma de otimização de cirurgias é um projeto que consiste na criação de uma plataforma *ABI*, que resulta de uma parceria com uma instituição hospitalar portuguesa. A ideia que move este caso prático é a melhoria da qualidade do serviço por meio da programação eficiente de cirurgias no processo de decisão. O principal objetivo deste estudo foi fornecer uma plataforma *ABI* capaz de estimar o tempo das cirurgias e, em seguida, otimizar o agendamento.

O agendamento da cirurgia nas salas de cirurgia do hospital é uma tarefa altamente complexa devido ao enorme volume de cirurgias e à variedade de combinações e restrições. A solução proposta consiste em dois blocos principais: previsão, que foi desenvolvido com a ferramenta *Weka*; o processo de otimização foi desenvolvido em *R*. O bloco de previsão prevê a duração de cada cirurgia, e integra posteriormente o bloco de otimização com o objetivo de minimizar o desperdício de turnos, evitando atrasos e aumentando a eficiência e a eficácia do cronograma.

A previsão contemplou a determinação do tempo padrão de cada cirurgia, utilizando várias modelos: Regressão Linear, *Multilayer Perception*, Árvores de Regressão (*IBk*) e *Support Vector Machine (SMOreg)*. O bloco de otimização irá utilizar o resultado do bloco de previsão como o seu novo conjunto de dados que contém eventos com o tempo de duração previsto no bloco anterior. Os métodos usados foram *Hill climbing* e *Simulated annealing*. A solução inicial foi obtida em turnos aleatórios ou sequenciais através da lista de cirurgias de especialidades correspondentes a cada turno, atribuindo um turno a cada índice de entrada. Os parâmetros inferior e superior representam os valores mais alto e mais baixo para cada dimensão, variando entre um e o número de turnos. A função de avaliação calcula os desperdícios de cada turno, subtraindo o tempo padrão total do turno atribuído e retorna a soma total do desperdício de cada turno. É introduzida uma função de mudança, que visa gerar a próxima solução, induzindo pequenas alterações na solução gerada anteriormente, de

forma a efetuar avaliações iterativamente. A Tabela 2 inclui um sumário dos resultados obtidos pelos modelos desenvolvidos com os 14 atributos utilizados.

Tabela 2 - Sumario dos modelos de regressão com os 14 atributos. Adaptado de (Ferreira et al., 2019).

REPTree	IBk	SMOreg	Regressão Linear	Multilayer Perceptron
Correlation coefficient	0.9958	0.813	0.9972	0.9955
Mean absolute error	269.7897	2107.5949	214.699	313.8986
Root mean squared error	405.5077	2850.6335	332.11	425.2081
Relative absolute error	7.8189%	61.0813%	6.222 3%	9.0973%
Root relative squared error	9.2617%	65.1079%	7.5853%	9.7117%
Total number of instances	237	237	237	237

O algoritmo *SMOreg* foi o mais eficiente, uma vez que apresentou a menor percentagem de erro absoluto. Na otimização, foram utilizadas 32 especialidades cirúrgicas, permitindo otimização anual e mensal, para encontrar a melhor solução. No final, com vista à discussão de resultados, são discutidos os resultados de duas especialidades, a que obteve menor e maior número de cirurgias. Dois tipos de otimização foram aplicados para resolver esse problema: procura cega num ano, e otimização distribuída para cada mês do ano. Os resultados para cada mês são apresentados na Tabela 3.

Tabela 3- Otimização sequencial mensal de desperdícios em cirurgia vascular. Adaptado de (Ferreira et al., 2019)

Mês do ano	Total desperdiçado
Janeiro	127089.205125
Fevereiro	134182.871451
Março	171413.244737
Abril	95961.169928
Mai	5401.727598
Junho	23751.340097
Julho	9194.226731
Agosto	9194.226731
Outubro	29114.226731
Total	605302.200000

Após análise dos resultados, conclui-se que o objetivo foi alcançado através da otimização mensal, uma vez que este apresentou-se como o mais rentável no agendamento de cirurgias. No entanto, para um número menor de cirurgias, o método de otimização que oferece um melhor resultado é o método de otimização anual, precisamente devido à capacidade de gerir melhor todas as combinações possíveis ao longo de um ano. Os resultados foram promissores e tornou viável o aprofundar este tema com uma abordagem *ABI*, eficaz no agendamento de eventos e adaptável a qualquer organização de Saúde.

3. Otimização do Agendamento de Cirurgias

O objetivo da presente dissertação incide no auxílio aos clínicos e gestores do *CHUP* no processo de gestão do agendamento de cirurgias. Este é um estudo em soluções *ABI* com vista a obter recomendações ideais numa área tão delicada como a Saúde. Assim, este trabalho visa criar um modelo inteligente capaz de responder aos requisitos delineados para o agendamento de cirurgias no *CHUP*. Este capítulo aborda a conceção do subsistema da plataforma e todo o processo de desenvolvimento do mesmo, utilizando a metodologia *CRISP-DM*.

O trabalho desenvolvido para esta dissertação encontra-se dentro de um projeto, logo por uma questão de confidencialidade não poderei mostrar ou mencionar qualquer tipo de informação relativa ao código utilizado.

3.1 Conceção e Desenvolvimento

Atualmente, os administradores de hospitais são pressionados a cumprir vários objetivos, tais como proporcionar melhor assistência médica, aumentar o número de camas disponíveis para novas entradas, e reduzir as listas de espera para cirurgia. Assim, a criação de artefacto capaz de otimizar o agendamento das cirurgias irá trazer grandes benefícios a estes mesmos administradores, dado que podem tomar uma decisão muito mais informada e orientada para atingir os objetivos definidos. Esta solução foi constituída com dois blocos: o bloco de previsão, que neste trabalho oferece um modelo simplista para uma previsão das datas-limite das cirurgias; e um bloco de otimização, que tem o desígnio de minimizar a penalização obtida com as cirurgias agendadas fora das datas-limite definidas. Estes blocos têm o intuito de impedir atrasos nas realizações de cirurgias e aumentar a eficiência e eficácia do agendamento.

O modelo inteligente realizado para o bloco de otimização foi elaborado na linguagem R, por meio da ferramenta RStudio, que facilita a utilização da mesma. No que diz respeito ao bloco de otimização, foi necessário proceder-se a um estudo, tanto de modelos e métodos a utilizar, como de heurísticas previamente desenvolvidas. Esta abordagem permitiu-nos adquirir um maior conhecimento em relação ao problema em causa e, conseqüentemente, ajudou a definir a melhor abordagem a ter por forma a resolver o mesmo. Os *datasets*

fornecidos são utilizados para testar o artefacto desenvolvida e os dois blocos desenvolvidos. O primeiro é responsável pela previsão das datas-limite para cada cirurgia, enquanto o segundo é responsável pela minimização das penalizações associadas a cirurgias que ultrapassem a data-limite. Os métodos de otimização moderna escolhidos foram o *Hill Climbing*, *Simulated Annealing* e *Evolutionary Algorithms*.

O problema de otimização consiste, no âmbito desta dissertação, na minimização da penalização de cada cirurgia alocada depois da data-limite. Para esse fim, foi utilizada uma função objetivo sem restrições. As restrições do projeto foram tidas em conta aquando da realização da solução inicial, que através da função *change*, faz pequenas perturbações na solução. Além das heurísticas utilizadas na construção da solução inicial, foram também utilizadas heurísticas de pesquisa local, que se concentram na melhoria da solução através da troca de duas diferentes cirurgias entre dias de cirurgia.

Os métodos implementados foram o *Hill Climbing* e *Simulated Annealing*. Estes são caracterizados por elaborarem uma pesquisa em soluções vizinhas, são capazes de alcançar soluções satisfatórias. Como se trata de modelos de otimização moderna, é necessário especificar os seguintes aspetos:

- **Representação da solução:** apresentada no formato lista, que contém *slots* diários de cirurgias e as cirurgias alocadas a cada espaço;
- **Função objetivo:** minimiza a penalização associada a cirurgias alocadas a *slots* depois da data-limite;
- **Função de avaliação:** calcula a penalização total do agendamento das cirurgias.

Posto isto, iremos agora explicar a função do algoritmo utilizada. Esta função foi retirada e adaptada do livro (Cortez, 2014).

`hclimbing (par, fn, change, control,`

A função acima apresenta cinco parâmetros de entrada, sendo eles:

- **A solução inicial (par):** é obtida através da alocação das cirurgias em *slots* disponíveis para a especialidade por um método *first fit*. Nesta solução inicial existe a preocupação

de respeitar as limitações temporais associadas a cada cirurgia e os *turnovers* existentes com a adição de múltiplas cirurgias a um *slot*;

- **A função de avaliação (fn):** esta função tem o objetivo de avaliar a penalização total das cirurgias alocadas;
- **A função *change* (change):** função responsável por gerar o próximo candidato, cria pequenas perturbações na solução inicial através de trocas de cirurgias;
- **O *control*:** lista com o método de paragem, sendo o número máximo de iterações (*\$maxit*) e a frequência de informações de controlo (*\$REPORT*) em que envia um relatório do resultado obtido da solução;
- **O tipo (type):** tipo do método de otimização, variável entre maximização e minimização. Neste caso, uma vez que o objetivo é minimizar a penalização, foi selecionado o tipo “min”.

Concluída a função, e descoberta a solução satisfatória, é apresentada a minimização da penalização das cirurgias alocadas. A solução permite uma redução considerável de gastos e uma marcação de cirurgias mais eficiente.

Em relação ao *Simulated_Annealing*, este método de estado único difere do *Hill Climbing*, uma vez que adota um parâmetro de temperatura de controlo que é usado para calcular a probabilidade de aceitar soluções inferiores. Em contraste com *Hill Climbing* estocástico, que adota um valor fixo para este parâmetro de controlo, o *Simulated Annealing* utiliza um valor de temperatura variável durante a pesquisa. O método começa com uma temperatura alta, que vai gradualmente diminuindo o parâmetro de controlo até atingir o valor mínimo ou até que o número de iterações definido seja alcançado (Cortez, 2014).

`simulated_annealing (func, s0, niter, step)`

A função acima descrita apresenta quatro parâmetros de entrada, sendo eles:

- **A função de avaliação (func):** Avalia a penalização total das cirurgias alocadas;

- **A solução inicial (s_0):** é obtida da mesma forma que métodos *hill climbing*, através da alocação das cirurgias em *slots* disponíveis para a especialidade por um método *first fit*. Nesta solução inicial existe a preocupação de respeitar as limitações temporais associadas a cada cirurgia e os turnovers existentes com a adição de múltiplas cirurgias a um *slot*;
- **Número de iterações (*niter*):** Utilizado como método de paragem, sendo este o número máximo de iterações;
- **Parâmetro de arrefecimento (*step*):** parâmetro de controlo da velocidade de arrefecimento do modelo.

A função *change* neste modelo, ao contrário do *Hill Climbing*, é inserida dentro da função em vez de inserida inicialmente. Por outro lado, em comparação com o modelo *Hill Climbing*, para obter a solução satisfatória, é apresentada a minimização da penalização das cirurgias alocadas. De forma a alcançar o modelo com melhores resultados para utilizar como método de otimização, dever-se-á comparar as soluções obtidas, escolhendo a que apresenta menor penalização, e em caso de valores semelhantes comparar o número de cirurgias penalizadas.

Por último, o método de pesquisa de população *Evolutionary Algorithm*, inspirado no fenómeno de seleção. Segundo (Luke, 2012), uma solução candidata é frequentemente denominada indivíduo. O genótipo, genoma, ou cromossoma indica a representação individual da estrutura de dados. Um gene é uma posição de valor nessa mesma representação e um alelo é um valor particular para um gene. A criação de novas soluções é chamada reprodução e ocorre devido à aplicação de operadores genéticos, como *crossover* e mutação. O *crossover* envolve a seleção de duas ou mais soluções, pai e mãe, por forma a gerar filhos, enquanto a mutação geralmente realiza uma ligeira mudança num único

rbga(stringMin, stringMax, popsize, iters, monitorfunc,

indivíduo.

Em contraste com os modelos anteriormente descritos, os valores iniciais do *Evolutionary Algorithm* são gerados aleatoriamente dentro do limite inferior (*stringMin*) e superior (*stringMax*). A função descrita apresenta cinco parâmetros de entrada, sendo eles:

- **Limite inferior (*stringMin*):** este parâmetro serve de limite inferior para os cromossomas gerados pelo modelo;
- **Limite superior (*stringMax*):** este parâmetro serve de limite superior para os cromossomas gerados pelo modelo;
- **Tamanho da população(*popsiz*e):** este parâmetro representa o número ideal de indivíduos para este problema;
- **Número de iterações (*iters*):** este parâmetro serve como método de paragem, sendo este o número máximo de iterações;
- **Função de monitorização (*monitorfunc*):** este parâmetro é usado para calcular estatísticas uteis, como medidas de diversidade populacional;
- **Função de avaliação (*evalfunc*):** Esta função tem o objetivo de avaliar a penalização total das cirurgias alocadas.

Terminada a função, é apresentada a melhor solução encontrada, ou seja, com menor penalização. Esta solução será apresentada na estrutura de dados vetor, em que cada posição do mesmo representa uma cirurgia do *dataset* utilizado e cada número associado a essa posição o turno ao qual essa mesma cirurgia foi atribuída. Os números de cada turno também são referentes às linhas do *dataset* turnos.

3.2 Fases CRISP-DM

Neste capítulo são apresentadas todas as fases de desenvolvimento da metodologia *CRISP-DM*, referencial adotado para esta dissertação.

3.2.1 Compreensão do negócio

A otimização do agendamento de cirurgias enquadra-se dentro da problemática do tempo de resposta máximo garantido (*TMRG*) estabelecido pelo Ministério da Saúde para as cirurgias, meios complementares de diagnóstico, e consultas externas. A ultrapassagem

destes mesmos limites acarreta custos elevados para os hospitais e doentes. O objetivo foi definido em termos de minimizar a penalização associada às cirurgias que ultrapassem o tempo de resposta estabelecido. Para tal, será necessário realizar uma previsão do tempo limite de cada cirurgia, dado que a solução satisfatória será alcançada através da utilização dessa previsão no método de otimização moderna.

Os métodos implementados encontram a solução através das soluções vizinhas; por outras palavras, percorre várias iterações, calculando a penalização total, até encontrar o valor menor para este problema. Em relação aos recursos, o *CHUP* possui uma solução tecnológica responsável pelo registo e armazenamento das informações clínicas, e por esse facto não foi necessário obter nenhuma tecnologia focada na extração de dados.

3.2.2 Compreensão dos dados

Os dados extraídos e utilizados neste estudo referem-se a atos médicos realizados entre fevereiro de 2019 e fevereiro de 2020. Durante este período, um total de 893 cirurgias foram armazenadas, relacionados com 338 turnos, emergindo os dois *datasets* utilizados. Estes contêm os dados alusivos às cirurgias e aos turnos da especialidade em questão. Cada instância é correspondente a 146 atributos, enquanto uma instância dos turnos corresponde a 19 atributos. De seguida, deu-se início à análise desse conjunto de dados, visto que essa análise permite entender melhor os atributos que serão analisados e quais as vertentes a tomar para atingir mais agilmente os resultados. Uma vez que se trata de imensos atributos, nas tabelas seguintes serão descritos os atributos selecionados que possam vir influenciar a realização do modelo.

Tabela 4- Atributos selecionados referente aos turnos.

<i>Atributos de turnos</i>	<i>Descrição</i>	<i>Tipo de dados</i>
<i>ANO</i>	Ano do turno	Integer
<i>MÊS</i>	Mês do turno	Integer
<i>DIA</i>	Dia do turno	Integer
<i>HORA_INICIO</i>	Hora de início de turno em s	Integer
<i>HORA_FIM</i>	Hora de fim de turno em s	Integer

TURNO	Especificação do tipo de turno, se é de manhã, tarde ou todo o dia	String
COD_SALA	Identificador da sala de cirurgias	Integer
COD_ESPECIALIDADE	Identificador da especialidade	Integer

Tabela 5- Atributos selecionados referente às cirurgias.

Atributos de cirurgias	Descrição	Tipo de dados
COD_GRUPO	Código do grupo de especialidade	Integer
DES_GRUPO	Designação da especialidade da cirurgia	String
DTA_MARCACAO	Data de marcação da cirurgia	String
DTA_REGISTO	Data de registo	String
DTA_OPERADO	Data da realização da cirurgia	String
NUM_LISTA_ESPERA	Identificador de cirurgia	String
TEMPO5	Tempo despendido em cada cirurgia	Integer
TEMPO7	Duração da cirurgia	Integer
LICMAX	Data máxima limite de alocação da cirurgia	String
PRIORIDADE	Atributo com o valor da urgência de cada cirurgia	String
LICMIN	Data mínima limite de alocação da cirurgia	String
NUM_MEDICO	Identificador do medico interveniente	Integer
COD_PATOLOGIA	Código da patologia	Integer
DES_PATOLOGIA	Descrição da patologia	String
DTA_PREVISTA	Data prevista da cirurgia	String
COD_INTERV_CIRURGICA	Código da intervenção cirúrgica	Integer
DES_INTERV_CIRURGICA	Designação da intervenção cirúrgica	String
DISTRITO	Distrito de habitação do utente	String
SEXO	Sexo do utente	Integer

SERVICO	Código do serviço que realiza a cirurgia	Integer
COD_SALA	Código da sala da cirurgia	Integer
CONVENCIONAL	Indicação se a cirurgia foi convencional	Integer
PROGRAMADA	Indicação se a cirurgia foi programada	Integer
AMBULATORIO	Indicação se a cirurgia foi em regime de ambulatório	Integer
URGENTE	Indicação se a cirurgia foi urgente	Integer
NORMAL	Indicação se a cirurgia foi normal	Integer
SIGIC	Indicação se a cirurgia esteve em lista de espera	Integer
BASE	Indicação se a cirurgia foi realizada em horário de trabalho normal	Integer
ADICIONAL	Indicação se a cirurgia foi realizada em horário de trabalho extraordinário	Integer
ESPECIALIDADE	Designação da Especialidade	String
DUR_TURNO	Duração do turno	Integer
TURNOVER	Tempo despendido na limpeza da sala de operações	Integer
INTERVENCAO1	Código da principal intervenção	Integer
INTERVENCAO2	Código da intervenção secundária	Integer
INTERVENCAO3	Código da intervenção terciária	Integer
D1	Designação da principal intervenção	String
D2	Designação da intervenção secundária	String
D3	Designação da intervenção terciária	String
INIS	Início da cirurgia	Integer
COD_ASA	Código do risco de anestesia	Integer
DES_ASA	Designação do risco de anestesia	String
MEDICOS_PRINC	Código, nome e função do médico principal	String
MEDICOS_AJUD	Código, nome e função do médico ajudante	String

ENFERMEIROS	Código, nome e função o dos enfermeiros	String
COD_ANESTESIA	Código de anestesia	Integer
DES_ANESTESIA	Designação da anestesia	String

Analizados os dados, foi necessário elaborar um relatório da qualidade dos mesmos. Foi possível retirar algumas conclusões:

- **Várias colunas com todos os dados a *null*;**
- **Colunas com informação complementar;**
- **Colunas com poucos registos;**
- **Colunas com dados similares;**
- **Colunas de intervenientes numa cirurgia que não vamos ter em conta numa fase inicial;**

Para todos os problemas encontrados, a solução focou-se em métodos de resolução dos dados incompletos. Em último caso, eliminou-se as colunas, se estas não apresentassem resolução possível.

3.2.3 Preparação dos dados

Esta fase permitiu a obtenção de um conjunto de dados final para o desenvolvimento de técnicas de *DM*. O *dataset*, com os dados inicialmente fornecidos pelo *CHUP*, era composto por 1231 combinações entre cirurgias e turnos. No entanto, estes mesmos eram compostos por linhas que apresentavam bastantes duplicações nos vários atributos que o constituíam, ou por atributos que apenas continham valores *null*. A tabela seguinte apresenta os atributos excluídos.

Tabela 6- Conjunto de dados excluídos.

Dados excluídos	Justificação
------------------------	---------------------

DES_GRUPO DES_PATOLOGIA DES_ANESTESIA DES_ASA ESPECIALIDADE DES_INTERV_CIRURGICA D1 D2 D3	<p>Apresenta inúmeros atributos com duplicados.</p> <p>Deste modo, e visto que se torna repetitiva a informação, será mais adequado excluir a designação dos vários atributos.</p>
MEDICOS_PRINC MEDICOS_AJUD ENFERMEIROS	<p>Foi necessário excluir estes atributos, uma vez que cada coluna continha o código, nome e trabalho desempenhado pelos profissionais, informação considerada irrelevante no contexto do estudo.</p>
CONVENCIONAL NORMAL BASE	<p>Estes dados foram excluídos, uma vez que a informação contida era complementar à informação contida nas colunas AMBULATORIO, URGENTE e ADICIONAL, respetivamente.</p>
Atributos com valores null	<p>Não possuem qualquer tipo de informação relevante para o projeto</p>
SEXO DISTRITO	<p>Atributos que apenas acrescentam informação sobre o paciente também acabaram por ser excluídos uma vez que não seriam uma mais-valia para o modelo</p>

Em relação à limpeza e formatação dos dados, não foi necessário qualquer tipo de transformação, visto que não foram identificadas anomalias significativas.

Posteriormente, foi identificada a necessidade de construir dois novos atributos referente aos turnos, o tempo e o tempo ocupado. O atributo tempo consiste no tempo total do turno. O tempo ocupado é utilizado para guardar o histórico de tempo ocupado em cada turno.

3.2.4 Modelação

Nesta fase, foram selecionadas e aplicadas as diferentes técnicas para obter os modelos de otimização. Na previsão dos valores de penalização de cada cirurgia, foi utilizada a média interquartil, tendo em conta que é uma medida que vai ao encontro da maior densidade de valores, excluindo de certa forma os *outliers*. Na otimização da penalização inerente às cirurgias que ultrapassem a data-limite, os métodos de otimização moderno implementados são o *Hill Climbing*, *Simulated Annealing* e *Evolutionary Algorithem*.

Os modelos implementados são compostos por 3 atributos cruciais: a solução inicial, função de avaliação e função objetivo. O primeiro passo consistiu na elaboração da solução inicial. Uma vez que o modelo tinha um grau de complexidade elevado, determinou-se que a primeira solução deveria já conter as restrições do modelo e optou-se pela utilização do método *first fit*. Deste modo, as cirurgias eram inseridas consoante a dimensão das cirurgias e dos turnos. De seguida, foi elaborada a função de avaliação, que apresenta a finalidade de somar as penalizações obtidas com as cirurgias na sua totalidade. Esta penalização é obtida através da multiplicação do número de dias em que a cirurgia se encontra em atraso em relação à data-limite pela prioridade associada a essa mesma cirurgia.

Por fim, foi necessário elaborar a função objetivo, capaz de minimizar as penalizações através de perturbações na solução inicial. Para cada iteração é dada uma penalização total das cirurgias fazendo-se comparações com as restantes soluções. As iterações terminam no limite atribuído no método de otimização, sendo que este irá retornar o melhor valor das soluções obtidas. No final, é escolhida a solução satisfatória de entre os métodos de otimização testados.

3.2.5 Avaliação

Nesta fase foram apresentados e avaliados os resultados obtidos dos diferentes modelos de otimização construídos para resolução do problema. A análise destes resultados foi conduzida tendo em conta a seguinte questão:

- **A utilização destes modelos apresenta uma vantagem significativa para o agendamento do hospital?**

Por forma a responder a esta questão foi necessário avaliar os resultados que o hospital apresentou no mesmo espaço temporal e comparar com os resultados obtidos através dos modelos. Os desenvolvimentos foram satisfatórios, refletindo-se numa redução significativa da penalização, permitindo a colocação de todas as cirurgias sem a ocorrência de problemas, desde a preparação dos dados até a modelação, conseguindo-se minimizar a penalização das cirurgias de forma eficaz e eficiente. Posteriormente, é necessário comparar os métodos, escolhendo-se sempre a melhor solução entre as encontradas. Os resultados serão discutidos ao pormenor no Capítulo 4.

3.3 Protótipo final

O desenvolvimento dos algoritmos até aqui referenciados mostram um conjunto de resultados que identifica qual modelo permite responder ao problema existente. Com vista a dar suporte ao *CHUP*, na otimização do agendamento cirúrgico, foi desenvolvida uma *web app* que disponibiliza todo o trabalho desenvolvido com os dados fornecidos, permitindo a aplicação do conhecimento e o auxílio nas decisões clínicas e na organização dos serviços.

Após um conjunto de reuniões com a equipa administrativa do *CHUP*, definiu-se um conjunto de métricas e informações a serem disponibilizados na mesma aplicação desenvolvida. Ficou assim estabelecido a disponibilização de:

1. Indicadores estatísticos que comprovem a eficácia da aplicação dos algoritmos em dados reais;
2. Visualização do agendamento, permitindo perceber as cirurgias que foram agendadas e quais é que se mantiveram em lista de espera;
3. Controlo sobre os cenários de otimização, de forma a permitir os profissionais qual a melhor solução em determinado contexto.

A Figura 13 apresenta a solução desenvolvida. Constituí uma barra lateral para controlo dos cenários, variando os resultados de acordo com as variáveis inseridas. Estas correspondem:

- **Turnover:** Tempo associado à limpeza do bloco cirúrgico;

- **Eficiência:** Percentagem de eficiência, permitindo alterar os valores do tempo padrão das cirurgias;
- **Folga:** Tempo extra. Esta variável permite alocar tempo adicional a cada turno já definido.

As variáveis dão uma maior flexibilidade, tanto ao algoritmo implementado, como na tomada de decisão, permitindo realizar pequenas alterações que se ajustem à situação atual da organização. A proposta de agendamento é apresentada num conjunto de separadores que permitem consultar:

- Lista de cirurgias agendadas;
- Lista de cirurgias que continuam em lista de espera;
- Métricas de avaliação à proposta executada;
- Lista dos pacientes relativos a cada cirurgia.

	data_turno	tempo_disponivel	tempoocu	TURNO	slot1.NUM_LISTA_ESPERA	slot1.Prioridade	slot1.t
11	2021-01-06	22500	0	M	0	0	0
23	2021-01-06	15300	0	T	0	0	0
2	2021-01-08	15300	0	T	0	0	0
12	2021-01-13	22500	0	M	0	0	0
13	2021-01-13	15300	0	T	0	0	0
14	2021-01-20	15300	0	T	0	0	0
79	2021-02-03	15300	0	T	0	0	0
80	2021-02-10	15300	0	T	0	0	0
15	2021-02-12	15300	0	T	0	0	0
81	2021-02-17	15300	0	T	0	0	0

Figura 13- Aplicação para o agendamento de cirurgias.

Os modelos, após serem gerados, desenvolvem um conjunto de resultados, sendo necessário resumir os resultados alcançados. Para isso recorreu-se a métricas como:

- Ocupação de salas;

- Número de cirurgias agendadas;
- Número de cirurgias não agendadas;
- Penalização total;
- Cirurgias para além do *TMRG*.

A Figura 14 comprova a exibição das métricas que permite avaliar a eficiência da proposta de agendamento.

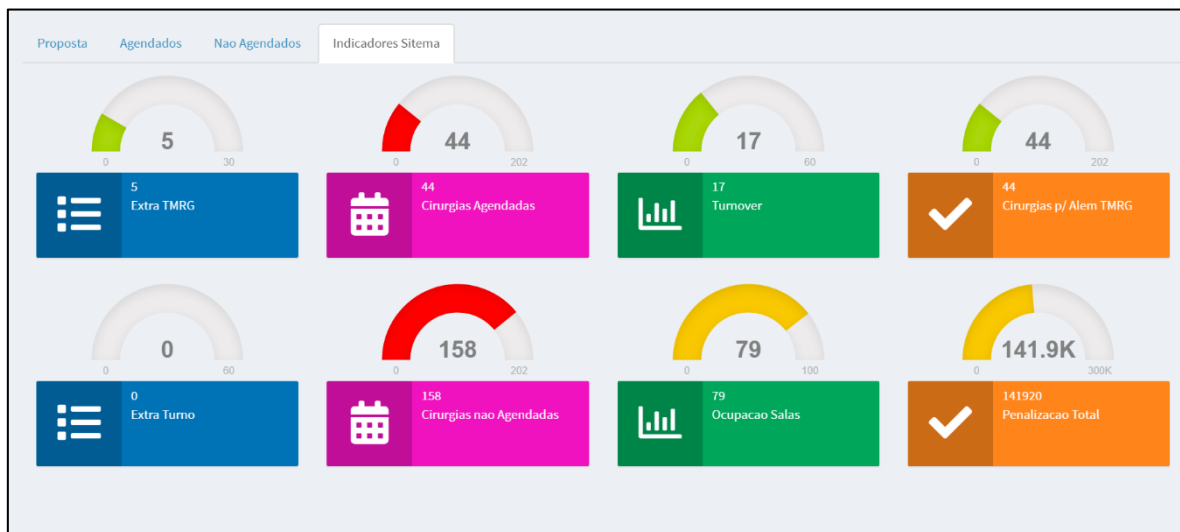


Figura 14- Indicadores da Aplicação.

4. Resultados obtidos

Neste capítulo é promovida a discussão dos resultados obtidos do bloco de otimização elaborado. Serão então analisados os vários modelos inteligentes desenvolvidos, de forma a desvendar qual deles apresenta a melhor solução de distribuição de cirurgias, indicando as penalizações associadas a cada solução. O valor desta penalização é calculado para cada cirurgia inserida fora do prazo estipulado pelo Ministério da Saúde.

4.1 Otimização

De forma a comparar todos os modelos desenvolvidos, foi necessário testar todas as fases de execução, uma vez que só assim seria possível perceber qual das soluções seria capaz de gerar um maior rendimento e eficiência aquando do agendamento de cirurgias.

Para a resolução deste problema foi utilizada uma otimização anual. Com o intuito de definir quais os valores de penalização que seriam aceitáveis, foi utilizada a mesma fórmula matemática nos resultados obtidos pelo centro hospitalar. Assim, foi possível analisar se as propostas de otimização conseguiriam obter resultados melhores do que a situação atual. O resultado de penalizações anual obtido pela fórmula matemática apresentou um valor de 91370.

Havendo então a necessidade de alcançar um valor inferior ao do centro hospitalar, foi implementado o modelo de *Hill Climbing*. Para este primeiro modelo, foi necessário efetuar dois testes, variáveis entre si na solução inicial gerada. Tendo em conta que um dos testes apresentava uma solução inicial apenas baseada na longevidade dos pacientes na lista de espera, isto é, pacientes mais antigos são inseridos no modelo por *first fit*, enquanto o outro teste já apresenta uma solução inicial que tem em consideração a longevidade, mas também as prioridades cirúrgicas.

data_turno	tempo_disponivel	tempooc	TURNO	paciente1.NUM_LISTA_ESPERA	paciente1.TempoPadrao	paciente1.Prioridade	paciente1.Penalização	paciente1.Controlo
2/1/2019	22500	0	M	0	0	0	0	0
3/1/2019	17353	5147	M	19	5147	1	215	AGEN
4/1/2019	19800	0	M	0	0	0	0	0
7/1/2019	12174	3126	T	13	3126	1	211	AGEN
8/1/2019	19979	2521	M	49	2521	1	215	AGEN
8/1/2019	13510	1790	T	213	1790	2	1	AGEN
9/1/2019	22500	0	M	0	0	0	0	0
10/1/2019	22500	0	M	0	0	0	0	0
10/1/2019	15300	0	T	0	0	0	0	0
11/1/2019	19800	0	M	0	0	0	0	0
14/1/2019	12779	2521	T	89	2521	1	173	AGEN
14/1/2019	12779	2521	T	92	2521	1	173	AGEN
15/1/2019	22500	0	M	0	0	0	0	0
15/1/2019	15300	0	T	0	0	0	0	0
16/1/2019	22500	0	M	0	0	0	0	0
17/1/2019	9468	5832	T	228	1320	1	215	AGEN
17/1/2019	19979	2521	M	53	2521	1	215	AGEN
18/1/2019	19800	0	M	0	0	0	0	0
21/1/2019	8943	6357	T	119	2521	1	199	AGEN

Figura 15- Amostra da solução inicial gerada pelo sistema

A Figura 15 representa uma amostra da solução inicial desenvolvida pelo sistema. Esta solução inicial segue uma estrutura de dados em lista, sendo que os turnos representam as linhas da lista, tendo depois seis slots disponíveis para cada turno. A lista é composta por vários atributos respetivos aos turnos e a cada slot específico, sendo que a esta mesma estrutura foi projetada através de uma comunicação direta com administradores e médicos do *CHUP*.

[illegible]

Figura 16- Resultado do algoritmo Hill Climbing com 10 iterações.

A Figura 16 corresponde a uma amostra do primeiro teste do método de otimização de *Hill Climbing*, em que a solução inicial tem em consideração apenas a longevidade. Sendo que:

- i corresponde ao número de iterações do modelo, executado com 10 iterações;
- s representa o melhor agendamento de cirurgias;

- **f** corresponde à penalização das soluções calculadas;
- **s'** representa uma nova solução de agendamento para avaliação;
- **best** representa a melhor solução encontrada pelo modelo;
- **eval** corresponde à penalização total da melhor solução encontrada

A solução inicial apresenta um valor de penalização igual a 493 000, e com o decorrer do algoritmo, este baixa para 470 280. A Figura 17 contém a evolução dos resultados do algoritmo *Hill Climbing* em relação ao número de iterações do mesmo.

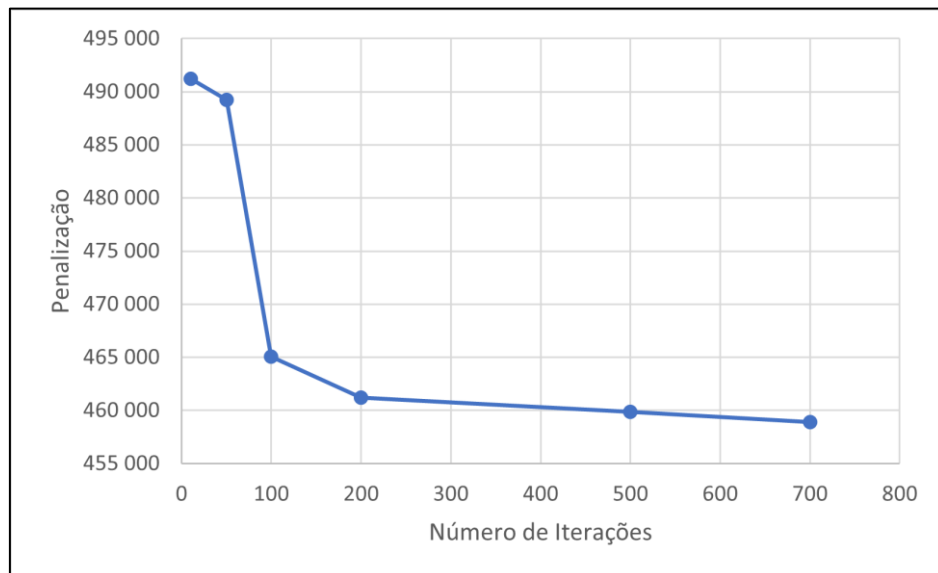


Figura 17- Evolução do Algoritmo de Hill Climbing em relação ao número de iterações.

A Tabela 7 apresenta o valor de penalizações obtido em cada teste do algoritmo correspondente à Figura 15.

Tabela 7- Resultados de penalização nas várias iterações.

Iterações	Valor de Penalizações ótimo obtido
10	491 200
50	489 210
100	465 080

A Tabela 8 apresenta o valor de penalizações obtido em cada teste do algoritmo correspondente à Figura 19.

Tabela 8- Resultados de penalização nas várias iterações da segunda solução inicial.

Iterações	Valor de Penalizações ótimo obtido
10	13 810
50	13 810
100	13 610
200	13 610
500	13 610
700	13 610

Mediante os resultados obtidos, é possível verificar que o algoritmo consegue obter a solução ótima mais eficiente com a solução inicial que tem em consideração a longevidade e prioridade dos pacientes, em comparação com a que apenas tem em consideração a longevidade dos pacientes. Esta diferença deve-se ao facto de a solução inicial inserir no modelo algumas cirurgias que, apesar de serem mais antigas, apresentam uma prioridade mais reduzida e, conseqüentemente, uma data de agendamento mais alargada. Este acontecimento faz com que cirurgias mais recentes e com maior urgência sejam forçosamente alocadas para *slots* mais afastados, acarretando uma maior penalização para o hospital e para os recursos humanos.

Em contrapartida, a solução inicial usada na Figura 18 permitiu colmatar estas mesmas lacunas, apresentando um resultado mais promissor. Com a otimização, foi possível minimizar ambas as penalizações das soluções iniciais, sendo que a primeira solução exibiu uma penalização final de 112990 e a segunda solução uma penalização de 13610. Uma vez que a penalização relacionada com o agendamento do hospital foi de 91370, é possível concluir que a solução que apenas tem em consideração a longevidade dos pacientes obteve um resultado insatisfatório.

Seguidamente, serão apresentados os resultados da otimização *Simulated Annealing* para ambas as soluções iniciais. A Figura 20 representa uma amostra da execução do algoritmo em que a solução inicial utilizada tem apenas em consideração a longevidade do paciente. Sendo que:

- **It:** é o número de iterações do sistema;
- **Best:** é o valor de penalização da melhor solução encontrada pelo algoritmo;
- **Current:** é o valor de penalização da solução atual;
- **Neigh:** é o valor de penalização da solução vizinha gerada;
- **Temp:** é o valor associado à temperatura de cada iteração.

```
> sol <- simulated_annealing(sumbin, par2 )
```

It	Best	Current	Neigh	Temp
0	493000.0000	493000.0000	493000.0000	1.0000
1	493000.0000	493000.0000	493000.0000	0.9900
2	491500.0000	491500.0000	491500.0000	0.9801
3	490200.0000	490200.0000	490200.0000	0.9703
4	490100.0000	490100.0000	490100.0000	0.9606
5	490100.0000	490100.0000	490100.0000	0.9510
6	490100.0000	490100.0000	490100.0000	0.9415
7	490100.0000	490100.0000	490100.0000	0.9321
8	490100.0000	490100.0000	490100.0000	0.9227
9	490100.0000	490100.0000	490100.0000	0.9135
10	490100.0000	490100.0000	490100.0000	0.9044

Figura 20- Resultado do algoritmo Simulated Annealing com 10 iterações.

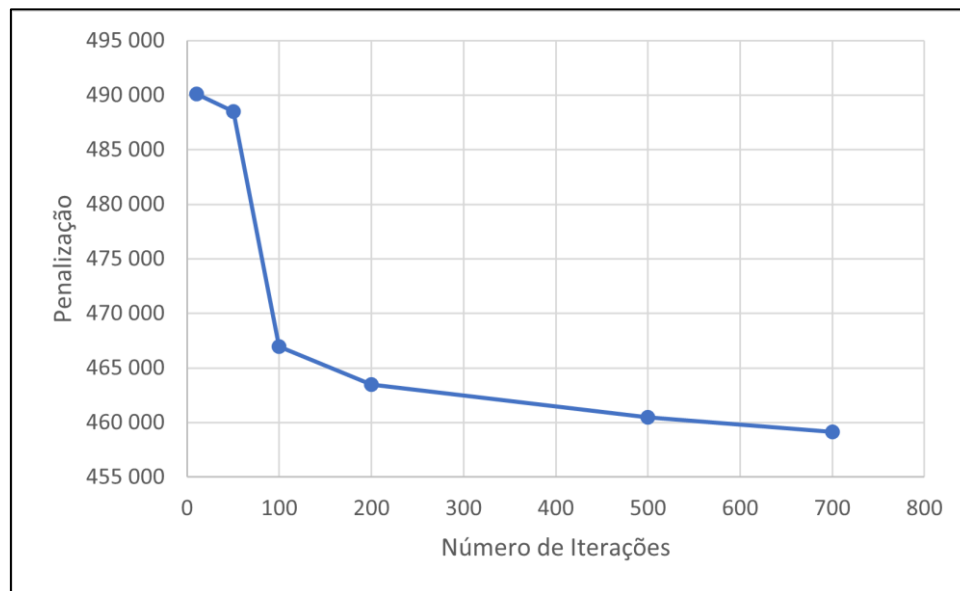


Figura 21- Evolução do Algoritmo de Simulated Annealing em relação ao número de iterações.

Após a análise, é possível averiguar que o modelo apresenta um valor de penalizações inicial de 493000 e com o decorrer do algoritmo este mesmo valor baixa para 459100.

A Tabela 9 apresenta o valor de penalizações obtido em cada iteração do algoritmo correspondente à Figura 21.

Tabela 9- Resultados de penalização nas várias iterações do Simulated Annealing.

Iterações	Valor de Penalizações ótimo obtido
10	490 100
50	488 510
100	466 980
200	463 480
500	460 480
700	459 110

A Figura 22 corresponde a uma amostra do processo do método de otimização do *Simulated Annealing*, desta vez através da solução inicial gerada, tendo em conta a prioridade do paciente e a longevidade na lista de espera:

> sol <- simulated_annealing(sumbin, par2)				
It	Best	Current	Neigh	Temp
0	13810.0000	13810.0000	13810.0000	1.0000
1	13810.0000	13810.0000	13810.0000	0.9900
2	13810.0000	13810.0000	13810.0000	0.9801
3	13810.0000	13810.0000	13810.0000	0.9703
4	13810.0000	13810.0000	13810.0000	0.9606
5	13810.0000	13810.0000	13810.0000	0.9510
6	13810.0000	13810.0000	13810.0000	0.9415
7	13810.0000	13810.0000	13810.0000	0.9321
8	13810.0000	13810.0000	13810.0000	0.9227
9	13810.0000	13810.0000	13810.0000	0.9135
10	13810.0000	13810.0000	13810.0000	0.9044

Figura 22- Resultado do algoritmo Simulated Annealing com 10 iterações para a segunda solução.

Neste caso, e em semelhança ao modelo *Hill Climbing*, é possível verificar que o algoritmo consegue obter a solução ótima mais eficiente com a solução inicial que tem em consideração a longevidade e prioridade dos pacientes. A solução inicial apresenta um valor de penalizações de 13 810 e com o decorrer do algoritmo este mesmo valor mantém-se inalterado. Seguidamente, será apresentado um gráfico da evolução do algoritmo relacionado com o número de iterações.

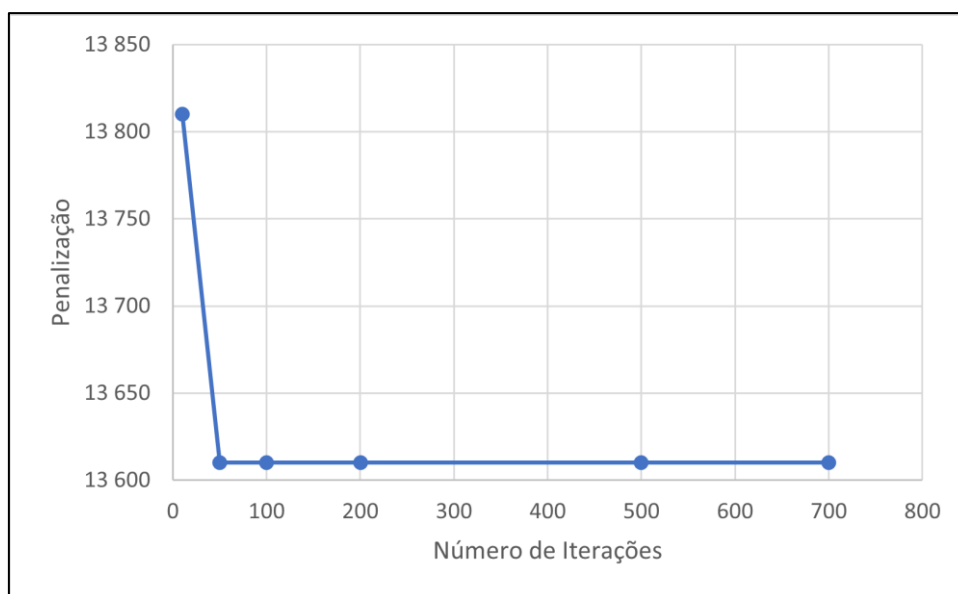


Figura 23- Evolução do Algoritmo de Simulated Annealing em relação ao número de iterações.

A Tabela 10 apresenta o valor de penalizações obtido em cada teste do algoritmo correspondente à Figura 23.

Tabela 10- Resultados de penalização nas várias iterações do Simulated Annealing para segunda solução.

<i>Iterações</i>	<i>Valor de Penalizações ótimo obtido</i>
10	13 810
50	13 610
100	13 610
200	13 610
500	13 610
700	13 610

Concluiu-se assim que ambos os modelos apresentaram valores muito semelhantes no que diz respeito ao número de penalização das soluções. Este resultado era expectável visto que são modelos de otimização de pesquisa local muito semelhantes. Assim, ainda é possível concluir que a utilização de uma solução inicial que tem apenas em conta a longevidade dos pacientes é incomportável para o hospital. A solução híbrida entre longevidade e prioridade apresentou valores promissores e conseguiu diminuir a penalização do hospital em 85%.

Posteriormente, de forma a tentar obter resultados distintos, foi implementado um modelo de procura baseado em população: *Evolutionary Algorithm*. Este necessita de um procedimento de classificação diferente, visto que não apresenta uma solução inicial e requer uma abordagem diferente em relação à apresentação da solução. Apresenta uma solução em forma de vetor, enquanto os modelos anteriores utilizam a estrutura de dados em lista. O vetor apresentado terá o tamanho da lista de cirurgias, sendo que cada posição do vetor representa uma cirurgia. Além disto, cada posição irá apresentar um número, que será o identificador do turno em que a cirurgia será inserida. O algoritmo não terá solução inicial, portanto é necessário adicionar mais penalizações ao modelo, por forma a que este consiga que a solução final encontrada satisfaça os requisitos do *CHUP*. Em seguida, na Figura 24, é apresentado o resultado do algoritmo. Sendo que:

- **best:** representa o melhor agendamento de cirurgias encontrado;
- **f:** é o valor de penalização da melhor solução encontrada pelo algoritmo.

```
best: 184 221 63 311 148 10 6 229 104 262 194 79 185 112 28 6 238 6 32 137 260 149 66 31 85 236 134 218 25 141 37 96 28 101 14 190 43 159 249 11 260 235 269 32 243 334 72 41 66 134 176 263 307 15
8 333 246 83 9 259 71 100 23 166 140 66 94 250 226 185 137 145 186 14 336 117 279 331 253 173 77 149 242 230 288 238 32 278 44 115 198 289 46 96 221 182 88 36 153 98 178 90 206 156 334 268 210 12
3 169 80 90 187 304 195 39 132 191 111 292 64 281 140 50 267 217 188 5 178 288 259 231 177 238 221 132 129 97 44 223 85 236 179 85 322 98 323 206 124 91 198 129 201 47 45 221 288 34 14 128 332 29
3 40 237 181 199 99 29 29 103 256 33 25 231 298 208 315 148 308 205 80 191 95 259 108 314 211 203 320 121 190 211 168 307 212 78 34 326 264 71 40 66 123 241 318 212 13 202 199 250 278 3 12 182 79
84 31 220 152 29 313 297 87 255 236 110 53 37 52 35 94 91 40 301 233 150 103 15 303 326 264 295 63 295 324 171 148 298 257 2 201 149 336 128 153 244 324 62 318 166 156 143 105 212 257 70 61 77 2
24 84 120 311 330 214 4 1 153 34 126 112 75 329 288 256 201 323 182 300 166 156 60 197 187 26 141 196 25 91 150 313 133 10 297 141 209 217 305 6 99 248 139 236 44 170 287 91 302 215 105 314 128 1
78 28 97 93 242 34 156 36 48 220 15 149 175 232 141 317 251 336 35 231 166 9 208 14 178 77 145 117 32 15 219 179 334 114 216 51 228 201 282 128 235 116 193 60 98 160 165 200 217 135 323 228 315 2
50 295 202 163 142 330 261 84 313 107 36 299 263 232 309 130 252 157 249 174 81 27 156 320 156 333 324 293 88 57 88 232 174 294 44 85 252 10 287 187 61 246 164 44 241 173 45 9 201 71 101 294 121
135 238 54 289 196 186 245 203 45 126 312 232 220 1 297 297 23 248 119 192 40 171 261 22 320 273 247 109 65 11 30 250 295 46 107 89 43 109 199 260 285 327 145 335 164 327 221 120 8 283 5 255 88
232 139 282 8 213 2 199 250 274 305 294 176 189 51 327 130 31 55 301 194 129 275 258 69 222 77 118 172 139 31 1 6 173 106 35 148 4 27 54 18 19 308 196 210 236 94 4 253 76 71 256 218 286 91 205 1
48 138 20 21 51 79 17 12 136 311 295 23 60 187 156 30 90 32 226 17 181 75 252 43 239 50 316 324 53 280 117 143 106 19 208 218 137 225 201 153 41 136 271 322 239 2 46 241 115 315 35 175 260 328 31
5 136 245 246 125 229 237 310 303 103 220 105 52 238 68 255 243 41 201 250 120 214 213 122 105 71 35 16 107 317 284 202 334 28 266 231 80 61 266 146 276 137 64 17 52 315 22 150 133 250 96 141 245
242 114 89 138 300 17 177 129 89 163 229 204 279 192 258 7 72 245 216 257 25 7 306 41 186 91 150 31 299 264 47 182 300 325 334 238 70 316 203 187 159 269 32 118 268 83 17 129 65 136 265 300 281
90 93 329 33 253 260 60 120 248 268 133 136 176 54 153 145 305 112 216 337 42 54 298 148 113 32 95 44 316 65 24 3 13 273 251 75 153 292 324 298 311 337 113 115 110 275 275 293 209 253 143 159 15
5 116 40 53 10 291 306 123 306 36 63 48 51 110 196 70 226 133 81 206 284 164 128 190 234 145 72 280 319 35 285 279 175 90 6 163 162 111 28 39 8 194 310 96 24 105 57 236 278 335 308 273 180 129 31
126 37 172 68 119 212 282 15 58 311 269 243 161 260 120 164 50 178 38 97 213 152 134 233 114 336 148 329 37 114 41 293 278 311 134 41 305 46 10 300 64 312 264 228 249 136 93 242 180 165 19 179 6
9 156 70 180 196 204 241 84 271 268 238 297 174 145 293 178 79 257 193 12 227 226 251 191 4 32 286 267 87 34 135 231 237 318 9 268 243 233 68 f: 6057130
```

Figura 24- Resultado do *Evolutionary Algorithm* com 10 iterações.

O resultado obtido demonstra-se bastante inferior quando comparado com os modelos de procura na vizinhança. O *Evolutionary Algorithm* gera soluções aleatórias, ou seja, para se obter um resultado que compreenda as *hard constraints* do problema, foi necessário adicionar penalizações elevadas às mesmas. O facto de ser feita uma programação anual também dificulta a chegada do modelo ao resultado ótimo devido ao elevado número de combinações possível. O resultado final do algoritmo compreende três tipos de penalizações:

- **C1:** Penalização de cirurgias, cujo tempo de realização previsto ultrapassa o tempo disponível no *slot*;
- **C2:** Penalização de cirurgias em que a data-limite mínima é superior à data em que cirurgia foi inserida;

- **C3:** Penalização usada nos algoritmos anteriores, caso a cirurgia ultrapasse a data definida pelo *TRMG* é então punida, tendo em consideração a sua prioridade. Quanto maior a prioridade maior será a penalização associada.

Seguidamente, a Figura 25 apresenta o gráfico que demonstra a evolução do algoritmo com o decorrer das iterações.

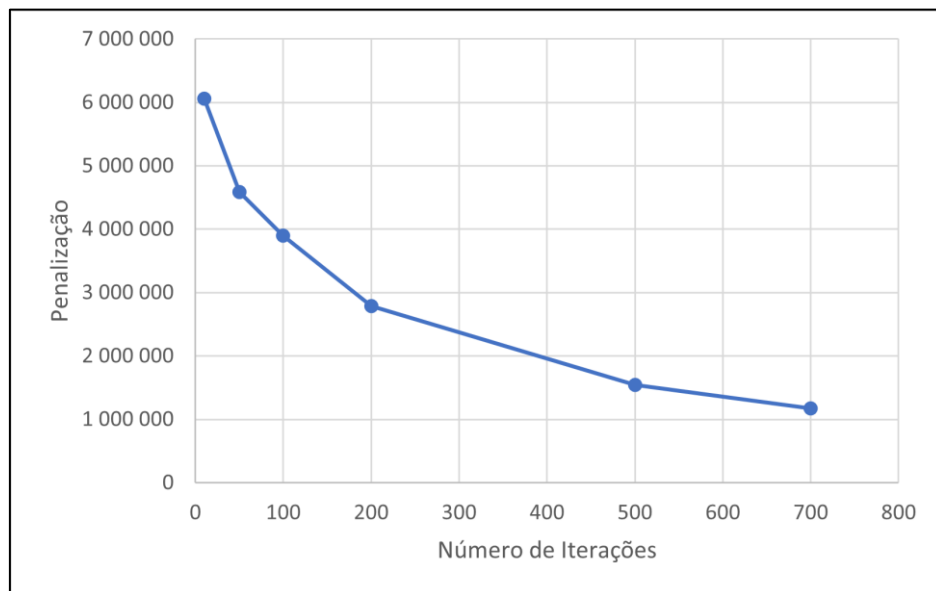


Figura 25- Evolução do *Evolutionary Algorithm* em relação ao número de iterações

Este modelo apresenta um poder computacional mais exigente comparativamente aos desenvolvidos anteriormente. Esta situação deve-se possivelmente à falta de *constraints* e de solução inicial, implicando um maior número de combinações possível. A Tabela 11 apresenta o valor de penalizações obtido em cada iteração do algoritmo correspondente à Figura 25.

Tabela 11- Resultados de penalização nas várias iterações do *Evolutionary Algorithm*.

<i>Iterações</i>	<i>Valor de Penalizações ótimo obtido</i>
10	6 057 130
50	4 586 600
100	3 900 140
200	2 787 480
500	1 544 670
700	1 169 310

Após a observação dos resultados obtidos, comprova-se que mesmo envolvendo múltiplas penalizações, este foi demasiado avultado, não atingindo os benefícios esperados. Posto isto, este algoritmo permitiu minimizar o desperdício total de cirurgias num ano, atingindo um valor de 1169310, comparativamente com os 6057130 obtidos inicialmente.

A Tabela 12 apresenta uma reflexão sumária de todos os estudos desenvolvidos, exibindo índices de avaliação relativamente a todos os modelos gerados.

Tabela 12- Resultados dos diferentes algoritmos em comparação com o resultado obtido pelo centro hospitalar.

	CHUP	Hill Climbing	Hill Climbing 2º solução	Simulated Annealing	Simulated Annealing 2º solução	Evolutionary Algorithm
Penalização total	91 370	458 890	13 600	459 110	13 600	1 169 310
Cirurgias efetuadas	891	889	889	889	889	891
Cirurgias não efetuadas	0	5	5	5	5	0
Nº de cirurgias penalizadas	893	108	19	107	19	-
Tempo de execução com 10 iterações	-	00:01:26	00:06:80	00:00:82	00:01:17	01:04:53

Os tempos de execução dos modelos de *Hill Climbing* e *Simulated Annealing* são apenas referentes ao algoritmo, sendo que o desenvolvimento da solução inicial não foi contabilizado. Contudo apresentam uma duração de execução de cerca de 00:37:42 para ambos os modelos.

Conclui-se assim que os métodos de otimização de procura local com a solução inicial a ter em conta a longevidade e prioridade dos pacientes seriam os mais apropriados para o agendamento das cirurgias, uma vez que apresentaram os melhores resultados. Não obstante, para um número menor de cirurgias, ou para um intervalo menor de tempo, o método baseado em população poderia oferecer uma maior capacidade de otimização, precisamente pelo potencial de gestão de todas as combinações possíveis durante um ano.

Consequentemente, e considerando que o objetivo é implementar algoritmos inteligentes para futura integração numa plataforma *ABI* de apoio à decisão no *CHUP*, constata-se que, embora ainda existam penalizações em múltiplas cirurgias depois da execução dos algoritmos, a avaliação do modelo gerado é bastante positiva, uma vez que revela melhores resultados que o atual sistema de agendamento, contribuindo para o aumento da eficiência desta área.

5. Conclusão

Este estudo começa por abordar a atual situação na área da Saúde, com destaque a um conjunto de tecnologias integrados em sistemas de *IA* que podem revolucionar uma organização.

É apresentado um enquadramento relativo a temas que englobam um sistema *BI* e o que representam no funcionamento de uma entidade hospitalar, introduzindo um conjunto de associados a *DM*, explicando a relação com modelos de previsão e otimização, e cruzando estas temáticas com as motivações e objetivos associados a este projeto de investigação.

Posteriormente, desenvolveu-se um conjunto de testes e implementações com vista a identificar quais algoritmos poderiam dar resposta ao problema do agendamento cirúrgico em ambiente hospitalar. De seguida, demonstra-se os resultados obtidos, identificando vantagens e possibilidades inerentes à utilização dos diferentes algoritmos.

Com vista à conclusão deste estudo, é possível identificar um conjunto de ideias, relacionando as mesmas com o trabalho a desenvolver no futuro.

5.1 Síntese

Este projeto pode ser descrito como a criação de artefactos para o apoio à decisão na área do agendamento de cirurgias. O desenvolvimento do mesmo partiu da necessidade de tratar grandes quantidades de dados clínicos produzidos diariamente pelos *SI* nesta unidade hospitalar, e da expectativa de adquirir um maior conhecimento com valor para apoiar os serviços prestados.

Atualmente, existe um problema transversal a todas as organizações e instituições que lidam diariamente com a marcação de cirurgias. Desde falhas na programação de salas, cancelamentos de cirurgias ou outro tipo de problemas, gerando assim atrasos não previstos e diminuindo a eficácia do serviço.

O processo de tomada de decisão no setor da saúde é bastante complexo. Cada decisão efetuada neste setor tem associado um impacto direto ou indireto na sociedade e na instituição de saúde. A obtenção de decisões fundamentadas e pensadas envolve o processamento de um grande volume de informações e conhecimentos, fora do alcance do ser humano. Assim, para a realização desta dissertação foi proposto um caso de estudo que

envolvia a criação de modelos inteligentes de otimização no agendamento do setor da saúde. Estes modelos poderão posteriormente fazer parte de uma plataforma de ABI, de forma a facilitar aos profissionais de saúde o acesso aos mesmos.

O módulo de otimização elabora uma solução apta a cumprir todos os requisitos impostos pelo hospital. Esta mesma solução alcançará o número mínimo de agendamentos com penalização encontrados, auxiliando dessa forma os administradores na tomada de decisão do agendamento de cirurgias. Considerou-se importante a necessidade de resolver este problema através da utilização de métodos de otimização moderna, de forma a minimizar o desperdício das penalizações de cirurgias que ultrapassem a data-limite de realização. Após a elaboração dos múltiplos algoritmos de otimização moderna e as respetivas avaliações, foi possível retirar algumas conclusões:

1. A realização da solução inicial para os modelos de procura local evidencia qual das abordagens terá um maior impacto no problema existente, acabando por se destacar a abordagem que insere as cirurgias tendo em conta a sua longevidade na lista de espera e a sua prioridade.
2. Foi igualmente possível averiguar qual o tipo de modelo mais vantajoso. Após a realização de testes, verificou-se que, entre os dois tipos de procura estudados, a procura local é a opção mais válida.
3. Por fim, os testes permitiram compreender que os dois modelos de procura local apresentaram resultados muito idênticos. Conclui-se que qualquer um dos algoritmos será suficientemente capaz de ser implementado numa plataforma ABI.

5.2 Objetivos Cumpridos

Este capítulo pretende responder aos objetivos que conduziram esta investigação, que foram apresentados da seguinte forma:

- **Qual o contributo de um modelo inteligente de otimização sobre o processo de agendamento de cirurgias no CHUP?**

O artefacto desenvolvido apresenta contributos positivos para o agendamento de cirurgias, uma vez que foi possível diminuir a penalização das mesmas em comparação ao

hospital em 85%. Adicionalmente, foi possível tornar alguns processos automáticos, simplificando a alocação de data cirúrgica no sistema hospitalar.

- **Levantamento de conceitos relacionados com *ABI*, especificando modelos de otimização e casos de estudo similares à solução desta dissertação:**

Este projeto permitiu compreender os vários modelos de otimização, resultando na seleção dos modelos e heurísticas mais apropriados para o problema em causa. Foi possível compreender a complexidade do estudo, identificando possíveis formas de resolução do problema, abrindo novas possibilidades de investigação na área em questão.

- **Definir os critérios de eficiência de agendamento e taxa de utilização de turnos**

Uma vez que se trata de um projeto com uma componente prática, é importante existir critérios de eficiência capazes de fornecer aos utilizadores soluções consideradas bem-sucedidas. Assim, os critérios selecionados foram: Ocupação de salas, Número de cirurgias agendadas, Número de cirurgias não agendadas, Penalização total e Cirurgias para além do *TMRG*. Estes critérios permitem aos administradores perceber se as soluções apresentadas vão de encontro as suas metas.

5.3 Trabalho futuro

Como já foi referido anteriormente, esta dissertação permitiu desenvolver o subsistema de otimização para uma plataforma de *ABI*, que tem como principal objetivo recorrer a técnicas de otimização modernas para desenvolver modelos que permitam estabelecer agendamentos mais eficientes.

Hoje em dia, com a crescente necessidade das mais diversas indústrias adquirirem novos conhecimentos para melhorar os seus serviços, faz cada vez mais sentido optar por uma plataforma de *ABI* capaz de prever e gerar possíveis soluções. Tendo em conta os resultados positivos desta dissertação, podemos contribuir que estes sejam referenciados, não só na gestão do agendamento de cirurgias, mas também como proposta de otimização em áreas similares no *CHUP*.

Relativamente ao sistema desenvolvido, surgem alguns aspetos que podem ser implementados ou melhorados para tornar o sistema mais eficaz e mais propício a ser implementado na instituição:

- Realização de um estudo mais aprofundado sobre modelos inteligentes a serem integrados nesta área, com otimização da proposta desenvolvida neste trabalho;
- Adição de novas funcionalidades, de acordo com requisitos solicitados por profissionais do *CHUP*;
- Integração deste modelo no sistema *ABI*, com suporte a *API's* que possibilitem a aplicabilidade desta solução em contexto real.

6. Referências

- Brath, Richard, and Michael Peters. 2004. Dashboard Design: Why Design Is Important.
- Buckland, M. K. (1991). Information and information systems (No. 25). ABC-CLIO.
- Carvalho, J. A. (2000). Information System? Which one do you mean?. In Information system concepts: an integrated discipline emerging (pp. 259-277). Springer, Boston, MA.
- Carver, A., & Ritacco, M. (2006). The Business Value of Business Intelligence. A Framework for Measuring the Benefits of Business Intelligence. *Business Objects*, 33.
- Chapman, Pete, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, and Rüdiger Wirth. 2000 "Step-by-Step Data Mining Guide," 76.
- Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88.
- Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., & Burke, J. P. (1997). Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality. *Jama*, 277(4), 301-306.
- Cortez, Paulo. 2014. *Modern Optimization with R*. New York: Springer.
- Dresner, H. (1989). *Business intelligence*. Gartner Inc.
- Edelstein, Herbert A. 1999. *Introduction to Data Mining and Knowledge Discovery*. Potomac, Maryland: Two Crows Corp.
- Enthoven, A. Managed competition of alternative delivery systems. *Jornal of Health, Politics, Policy and Law*, vol. 13, p. 305-321, 1988
- Ferreira, J., Portela, F., Machado, J., & Santos, M. F. (2019, April). Adaptive Business Intelligence in Healthcare-A Platform for Optimising Surgeries. In *World Conference on Information Systems and Technologies* (pp. 611-620). Springer, Cham.
- Few, Stephen. 2006. *Information Dashboard Design*. O'Reilly Media.
- Fichman, Robert G., Rajiv Kohli, and Ranjani Krishnan, eds. 2011. "Editorial Overview —The Role of Information Systems in Healthcare: Current Research and Future Trends." *Information Systems Research* 22 (3): 419–28. <https://doi.org/10.1287/isre.1110.0382>.

Goldschmidt, R., & Passos, E. (2005). Data mining: um guia prático. Gulf Professional Publishing.

Gorunescu, Florin. 2011. Data Mining Concepts, Models and Techniques. Springer.

Grimson, Jane, William Grimson, and Wilhelm Hasselbring. 2000. "The SI Challenge in Health Care." Communications of the ACM 43 (6): 48–55. <https://doi.org/10.1145/336460.336474>.

Han, Jiawei, Micheline Kamber, and Jian Pei. 2012. DATA MINING Concepts and Techniques. Waltham, USA: Elsevier.

Hočevár, B., & Jaklič, J. (2010). Assessing benefits of business intelligence systems—a case study. Management: journal of contemporary management issues, 15(1), 87-119.

Imhoff, Claudia, Nicholas Gallemmo, and Jonathan G. Geiger. 2003. Mastering Data Warehouse Design: Relational and Dimensional Techniques. Indianapolis, IN: Wiley Pub.

Jonathan, Wu. (2000), "Business Intelligence: What is Business Intelligence?", DM Review.

Jun, Tang, Cui Kai, Feng Yu, and Tong Gang. 2009. "The Research Application of ETL Tool in Business Intelligence Project." In 2009 International Forum on Information Technology and Applications, 2:620–23. <https://doi.org/10.1109/IFITA.2009.48>.

Kaplan, R. S., Norton, D. P., & Horváth, P. (1997). Balanced scorecard. Schäffer-Poeschel.

Kimball, R., & Ross, M. (2011). The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons.

Kolodner, R. M., Cohn, S. P., & Friedman, C. P. (2008). Health Information Technology: Strategic Initiatives, Real Progress: There is nothing "magical" about the strategic thinking behind health IT adoption in the United States. Health Affairs, 27(Suppl1), w391-w395.

L. Hovis, Gregory. 2002. Stop Searching for InformationMonitor It with Dashboard Technology.

Lenz, R., and K.A. Kuhn. 2004. "Towards a Continuous Evolution and Adaptation of Information Systems in Healthcare." *International Journal of Medical Informatics* 73 (1): 75–89. <https://doi.org/10.1016/j.ijmedinf.2003.09.003>

Luke, S. (2009). *Essentials of metaheuristics*. Lulu, Retrieved January 20th, 2012.

Macario, A., Vitez, T. S., Dunn, B., & McDonald, T. (1995). Where are the costs in perioperative care?: Analysis of hospital costs and charges for inpatient surgical care. *The Journal of the American Society of Anesthesiologists*, 83(6), 1138-1144.

Meskens, Nadine, David Duvivier, and Arnould Hanset. 2013. "Multi-Objective Operating Room Scheduling Considering Desiderata of the Surgical Team." *Decision Support Systems* 55 (2): 650–59. <https://doi.org/10.1016/j.dss.2012.10.019>.

Michalewicz, Zbigniew. 2007. *Adaptive Business Intelligence*. Berlin ; New York: Springer.

Moody, D. L., & Kortink, M. A. (2000, June). From enterprise models to dimensional models: a methodology for data warehouse and data mart design. In *DMDW* (p. 5).

Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional.

O'brien, J. A., & Marakas, G. M. (2006). *Management information systems* (Vol. 6). McGraw-Hill Irwin.

Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*. <http://doi.org/10.2753/MIS0742-1222240302>

Ranjan, J. (2009). Business intelligence: Concepts, components, techniques and benefits. *Journal of theoretical and applied information technology*, 9(1), 60-70.

Santos, M. F., & Azevedo, C. S. 2005. "Data Mining: Descoberta de Conhecimento em Bases de Dados". FCA Editores.

Sarawagi, Sunita, Rakesh Agrawal, and Nimrod Megiddo. 1998. "Discovery-Driven Exploration of OLAP Data Cubes." In *Advances in Database Technology — EDBT'98*, edited by

Hans-Jörg Schek, Gustavo Alonso, Felix Saltor, and Isidro Ramos, 1377:168–82. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0100984>.

Stackowiak, R., Rayman, J., & Greenwald, R. (2007). Oracle data warehousing & business intelligence SO. John Wiley & Sons.

Turban, E., & Walls, J. G. (1995). Executive information systems—a special issue. *Decision Support Systems*, 14(2), 85-88.

Vassiliadis, Panos, and Timos Sellis. 1999. “A Survey of Logical Models for OLAP Databases.” *ACM SIGMOD Record* 28 (4): 64–69. <https://doi.org/10.1145/344816.344869>.

Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making* (p. 417). United Kingdom: Wiley.

Watson, H. J., Rainer Jr, R. K., & Koh, C. E. (1991). Executive information systems: a framework for development and a survey of current practices. *MIS quarterly*, 13-30.

Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: a critical success factors framework. *International Journal of Enterprise Information Systems (IJEIS)*, 4(3), 79-94.