

Otimização do Agendamento de Cirurgias em Meio Hospitalar

Índice

INTRODUÇÃO	3
ENQUADRAMENTO E MOTIVAÇÃO	3
OBJETIVOS E RESULTADOS ESPERADOS	3
ESTRUTURA DO DOCUMENTO	4
ESTADO DE ARTE	5
SISTEMAS DE INFORMAÇÃO.....	5
SISTEMAS DE INFORMAÇÃO NA ÁREA DA SAÚDE.....	6
BUSINESS INTELLIGENCE.....	8
<i>Introdução.....</i>	<i>8</i>
<i>Extração, Transformação e Carregamento de dados (ETL).....</i>	<i>10</i>
<i>Data Warehouse</i>	<i>11</i>
<i>Cubo OLAP.....</i>	<i>14</i>
<i>Dashboards.....</i>	<i>15</i>
ADAPTIVE BUSINESS INTELLIGENCE	16
<i>Data Mining.....</i>	<i>18</i>
<i>Otimização.....</i>	<i>22</i>
<i>Previsão</i>	<i>24</i>
ADAPTIVE BUSINESS INTELLIGENCE NA SAÚDE.....	26
<i>Caso prático.....</i>	<i>26</i>
METODOLOGIAS DE INVESTIGAÇÃO.....	30
DESIGN SCIENCE RESEARCH	30
CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING	32
PLANEAMENTO DO PROJETO E CRONOGRAMA	33
CONCLUSÃO	33

Introdução

Neste capítulo é feita a contextualização do projeto desenvolvido, é apresentado um enquadramento do tema e identificada 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.

Enquadramento e Motivação

Atualmente pretende-se minimizar ou maximizar um determinado objetivo através do apoio à decisão, e com a evolução dos sistemas computacionais, podemos resolver esses mesmos problemas e de forma mais simples, permitindo uma melhor eficiência e eficácia da decisão tomada, poupando recursos humanos e financeiros.

Com o aumento de importância dos sistemas de apoio à tomada de decisão, surge também com mais importância a utilização do sistema Adaptive Business Intelligence enquanto apoio à decisão, uma vez que permite prever, otimizar e adaptar a mudanças externas. Este tipo de sistema mostra-se mais eficiente que o Business Intelligence, uma vez que este apenas garante o apoio à decisão, sendo incapaz de obter soluções com melhor qualidade e disponibilidade em tempo-real.

O Ministério da Saúde estabelece tempos de resposta máximo garantido (TMRG) para as cirurgias, meios complementares de diagnóstico e consultas externas. Estes tempos são frequentemente ultrapassados acarretando custos elevados para os hospitais e doentes.

Este projeto será realizado em parceria com o Centro Hospitalar do Porto (CHP), mais especificamente com o Hospital Santo António (HSA). Esta instituição foca-se na prestação de serviços de saúde que melhorem a qualidade de vida dos utentes e da população. O Hospital Santo António também fornece apoio e articulação com as restantes instituições de saúde.

O objetivo deste trabalho centra-se na possibilidade de prever com acuidade 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, 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.

A principal motivação para a realização desta dissertação, incide primeiramente pelo interesse na área dos Sistemas de Apoio à Decisão (SAD), em particular a área de data mining e otimização. Outra motivação foi a possibilidade de trabalhar na área da saúde e ajudar os profissionais da área a tomar decisões mais informadas por forma poupar recursos humanos e financeiros.

Objetivos e Resultados Esperados

Nesta fase são identificados os objetivos que ajudam na resolução do problema. Estes objetivos deveram ser deduzidos racionalmente a partir da especificação do problema. Para a realização desta atividade os recursos requeridos são o conhecimento do estado do problema e as suas soluções atuais.

O principal objetivo desta dissertação é melhorar a qualidade da informação que as entidades de saúde possuem e que seja necessária ao processo de decisão. Para atingir

esse objetivo será essencial a utilização de uma plataforma de Adaptive Business Intelligence, com processos padronizados e otimizados.

A área da saúde é uma das principais áreas da nossa sociedade e requer um sistema como o Adaptive Business Intelligence.

Assim, e como objetivos particulares desta dissertação, temos:

- Prever com grande acuidade os tempos associados aos eventos.
- Definir os critérios de eficiência de agendamento e taxa de utilização de recursos;
- Otimizar o agendamento, tendo em conta os critérios estabelecidos, recursos disponíveis.
- Permitir simular cenários e determinar necessidades extra ou em excesso;

Estrutura do documento

Este documento encontra-se dividido em 6 capítulos:

- **Capítulo 1-** Neste capítulo encontra-se a Introdução, o pretendido com este capítulo é fazer uma contextualização do projeto desenvolvido. É apresentado um enquadramento do tema e identificada a motivação para a realização da dissertação.
- **Capítulo 2-** Neste ponto do projeto é apresentada a revisão de literatura e o estado de arte dos termos que foram abordados ao longo do desenvolvimento desta dissertação. O estado de arte é importante para sustentar o trabalho desenvolvido;
- **Capítulo 3-** Capítulo onde será realizada a Abordagem Metodológica, sendo aqui que é descrita a metodologia de investigação - Design Science Research Methodology – e a metodologia de desenvolvimento - Cross Industry Standart Process for Data Mining.
- **Capítulo 4-** Este capítulo trata-se do Planeamento do Projeto e Cronograma, em que são especificadas as atividades desenvolvidas e a desenvolver, a lista dos riscos que poderão surgir na realização deste projeto e a calendarização de todas as tarefas a realizar;
- **Capítulo 5-** Neste capítulo temos as conclusões, onde é feito um balanço do trabalho realizado, onde são identificadas as limitações e dificuldades encontradas assim como os riscos envolvidos no projeto
- **Capítulo 6-** Por fim, temos o capítulo das referências, onde são descritas as fontes de informação utilizadas para a revisão de literatura.

Estado de Arte

Sistemas de Informação

Um sistema de informação pode ser perspectivado 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, isto deve-se também à informatização das infraestruturas e à digitalização da informação, que acelerou todo o processo.

De modo a compreender os Sistemas de Informação é importante primeiro definir os conceitos básicos, tanto de sistema como de informação. Embora exista um risco assumido por não existir um acordo quanto ao significado destes dois termos, a necessidade de os apresentar é inegável.

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. Conjunto de dados cuja forma e conteúdo são então apresentados de maneira útil para uso no processo de tomada de decisão. A informação caracteriza-se por ser entendida como objetos simbólicos construídos deliberadamente para possibilitar a comunicação e a formação de conhecimento. 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.

As Informações consistem em representações dos dados processados, com significado e com textos bem definidos. As tecnologias de informação são utilizadas para processar os dados e obter informações uteis.

Na 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 abstraído das bases de dados por recursos tradicionais da tecnologia da informação.

Sistema é o resultado da visualização do mundo ativo a partir de um certo ponto de vista. Qualquer coisa (e especialmente uma coisa ativa) pode ser vista como um sistema. Um sistema (em geral ou em abstrato) pode ser definido como ativo (faz alguma coisa), estável (tem uma estrutura...) e evolutivo (...que muda com o tempo) coisa ou objeto que opera num ambiente (ele interage com outras coisas) com algum objetivo (do ponto de vista do modelador, há uma razão para o sistema fazer o que faz) [Le Moigne, 1977]. Existem autores que apenas definem sistemas como um conjunto de componentes que se interligam entre si.

Os sistemas podem se dividir em dois subgrupos, os sistemas simples e os sistemas mais complexos. Por um lado, existem os sistemas simples são apenas objetos ativos sem autocontrolo e sem capacidade de cognição. Por outro lado, 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.

Um sistema autónomo, por exemplo uma organização, pode se considerar ser subdividido em três subsistemas: operacional, de gestão e informativo. O subsistema

operacional inclui as atividades que executam ações diretamente relacionadas ao objetivo ou missão do sistema. O subsistema de gestão inclui as atividades que gerem (organizam, planeiam, controlam, coordenam etc.) as atividades operacionais. O subsistema informativo estabelece comunicação entre os outros dois subsistemas. A necessidade do sistema informacional é particularmente clara quando essa comunicação é assíncrona e a existência de mecanismos que suportam a memorização e recuperação de informações é evidente.

A ação de um sistema altera outros objetos. O objeto alterado pode ser passivo (matéria, energia ou informação) ou ativo (outro sistema). No caso do objeto passivo o sistema pode alterar objeto na sua forma, no espaço ou em tempo. Enquanto que no caso objeto ativo a modificação altera a estrutura do objeto.

Com base nas definições de informação e sistema fornecidas, é possível concluir que existem três interpretações possíveis para a definição de um sistema de informação. A primeira, é a que um sistema de informação é um objeto ativo que lida com (processos) informações; a segunda, é a que um sistema de informação é um objeto ativo cujo objetivo é informar. Por último temos uma interpretação híbrida que é a junção das duas interpretações.

A primeira interpretação focaliza 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, decodificar (formulário); coletar, apresentar, transmitir (espaço); armazenar, memorizar (hora). 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) objetos simbólicos e cujo objetivo é informar. A terceira interpretação é uma combinação das outras duas como já foi referido. 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.

Sistemas de Informação na área da Saúde

A importância atribuída à área da Saúde e os custos acrescidos da economia, contribuíram para surgimento desta área, como uma importante área de estudo.

Longos tempos de espera, cancelamentos e sobrecarga de recursos ocorrem com frequência na área da saúde. Os sistemas de informação (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 da tecnologia da informação (TI) em equipamentos clínicos e de diagnóstico, os sistemas de informação (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 tecnologias de informação (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 inadequadas 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, problemas de dados de entrada, preocupações de segurança e confidencialidade, a ausência em muitos países de um identificador único do paciente e uma falta geral de conhecimento 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 (por exemplo, novos procedimentos diagnósticos ou terapêuticos, mudança da estrutura departamental) quanto a causas externas (por exemplo, introdução de DRGs, 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. Posto isto, 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).

Nos ambientes de saúde, a informatização dos registos de saúde é adotada desde o início dos anos 1960, quando os hospitais começaram a usar computadores (Khoumbati et al., 2005). No ambiente hospitalar, médicos, enfermeiros e até funcionários administrativos lidam com vários tipos de dados, incluindo dados de pacientes, dados de funcionários, inventário de medicamentos e assim por diante. Todos os dados devem ser mantidos e tratados para facilitar o acesso no futuro. Atualmente, profissionais de saúde têm acesso a computadores, que são usados com intuito de armazenar, aceder e atualizar dados.

À medida que o ambiente contemporâneo toma forma e os hospitais se tornam organizações de saúde por meio da integração vertical (Conrad, et al., 1988), a necessidade de desenvolver sistemas de informação em toda a instituição para ajudar a tomada de decisões tornou-se primordial. As pressões cada vez mais competitivas promoveram novos incentivos para o desenvolvimento de estratégias organizacionais que vão muito além da tradicional declaração de missão em que os hospitais confiam há muito tempo (Enthoven, 1988; Harrell e Fors, 1986). Os hospitais tentam 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 arena importante para examinar o uso estratégico dos sistemas de informação.

A importância que a área da saúde apresenta para o dia-a-dia das pessoas, faz com os erros cometidos ganhem uma proporção maior. Os erros médicos, de uma forma geral, são caros, 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. Portanto, a qualidade da assistência médica é diligentemente perseguida e vigilantemente executada, e o sistema de informação pode

facilitar a busca, destacando e monitorizando os erros em vários estágios ao longo do atendimento contínuo.(Fichman et al., 2011)

Business Intelligence

Introdução

No início da década de 1970, os sistemas de apoio à decisão foram as primeiras aplicações desenhadas com intuito de apoiar a tomada de decisão. Estas novas aplicações contrastavam com as aplicações operacionais ou processamento de transações até lá usados, como exemplo entrada de pedidos, controle de stock e sistemas de folha de pagamento. No entanto, os sistemas de reporting eram estáticos e não tinham qualquer capacidade analítica.

Nos inícios de 1980, o conceito de Executive Information Systems (EIS) emergiu (Efraim Turban et al., 2011). O conceito de EIS é definido como um sistema computadorizado que fornece aos executivos acesso fácil a informações internas e externas e que sejam relevantes para os fatores críticos de sucesso (Watson, Rainer, & Koh, 1991). O conceito começou a expandir e aumentar o suporte computadorizado aos níveis de topo das organizações, os relatórios começaram a ser mais dinâmicos.

No início dos anos 90, Howard Dressner, então analista do Gartner Group, definiu o termo Business Intelligence. BI é agora amplamente utilizado para descrever aplicações analíticas. O conceito original de EIS foi então transformado em Business Intelligence.

Em 2005, alguns sistemas de BI começaram a incluir capacidades de inteligência artificial bem como capacidades analíticas bastante eficazes (Efraim Turban et al., 2011).

Hoje em dia, um bom sistema de informação empresarial baseado em Business Intelligence contém toda a informação relevante para as necessidades dos gestores. As organizações fazem depender a sua atividade da utilização de uma grande quantidade de dados, pelo que dependem do Business Intelligence para uma gestão mais eficaz.

Stackowiak (2007) define Business intelligence como o processo de armazenar grandes quantidades de dados, analisando esses dados e apresentando um conjunto de relatórios de alto nível que condensam a essência desses dados na base de ações de negócios, permitindo que o gestor tome decisões fundamentais de negócio diariamente. Por sua vez, Santos (2009) entende que BI é a aplicação de um conjunto de técnicas e ferramentas que são propostas para auxiliar na administração de um negócio e na tomada de decisões. Com o decorrer dos anos e com o aumento de conhecimento na área, novas propostas para definir Business Intelligence começaram a surgir. É então que DUAN & XU (2012), definem o termo como sendo o processo de transformação de dados brutos em informações utilizáveis para maior efetividade estratégica, insights operacionais e benefícios reais para o processo de tomada de decisão nos negócios.

Como já foi mencionado, os sistemas de BI tiveram um grande crescimento ao longo dos anos, chegando a ser a principal prioridade de muitos diretores de informação. Foi realizada uma pesquisa com 1.400 CIOs, o grupo Gartner constatou que os projetos de BI eram a prioridade número um da tecnologia em 2007. O BI “tornou-se uma iniciativa estratégica e agora é reconhecido pelos CIOs e líderes de negócios como um instrumento para impulsionar a eficácia e a inovação dos negócios”, de acordo com o vice-presidente de pesquisa da empresa, Andreas Bitterer.

O BI inclui vários softwares para extração, transformação e carregamento (ETL), data warehousing, consulta e relatório de base de dados (Berson et al., 2002; Curt Hall, 1999) análise de dados de processamento analítico multidimensional / on-line (OLAP), data mining e visualização.

A junção das várias ferramentas leva assim a um sistema de BI, 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.

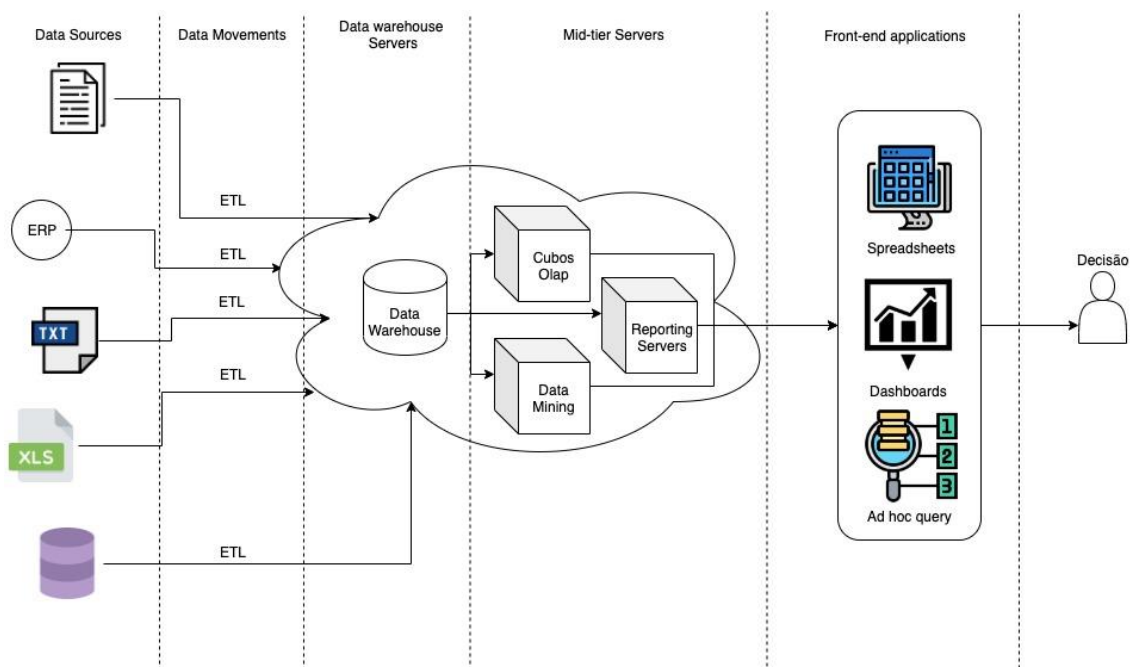


Figura 1 - Arquitetura de um sistema de BI adaptado de (Chaudhuri, Dayal e Narasayya, 2011).

Para que o sistema forneça o apoio necessário à tomada de decisão é importante que possua uma estrutura que permita extrair dados de várias fontes, posteriormente serão efetuados os tratamentos necessários, depois dos dados já terem sido tratados serão então armazenados. 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. É o caso de servidores Online Analytical Processing (OLAP), servidores de relatórios e mecanismos de data mining (estes temas serão todos abordados posteriormente).

O acesso pode ser efetuado de diferentes formas possibilitando aos gestores do negócio visualizar os dados sob diversas formas, das quais se destacam os relatórios, dashboards, gráficos. 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 do sistema de Business Intelligence, é possível concluir que implantação de sistemas de business Intelligence consome uma quantidade considerável de recursos e, ao mesmo tempo, sua ampla aplicabilidade nos ambientes internos e externos de negócios apresentam às organizações inúmeros benefícios, torna imperativo avaliar estes benefícios tangíveis e intangíveis Wu (2000). Em seguida vai ser apresentada a tabela 1 com alguns dos principais benefícios observados por uma empresa após a implantação de sistemas de business Intelligence:

Principais categorias de benefícios	Autores
Business reporting mais rápidos e precisos	Pellissier & Kruger. (2011); Đekić & MladenovićRanisavljević (2010).
Processo de tomada de decisão melhorado	Hočevvar and Jaklič (2010); Đekić & MladenovićRanisavljević (2010); Carver and Ritacco (2006)
Satisfação do cliente melhorada	Hočevvar 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); Hari Misra (2007); Carver and Ritacco (2006)
Aumento na quota de mercado	Atre & Moss (2003)

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

As principais categorias de benefícios de inteligência de negócios deve estar vinculado à estratégia de negócios, a fim de atingir os objetivos estratégicos da empresa (Hočevvar e Jaklič, 2010).

Extração, Transformação e Carregamento de dados (ETL)

O processo de ETL é a base dos sistemas de business Intelligence, e normalmente está relacionado com o sucesso ou falha dos mesmos, posto isto, é importante ter em atenção este processo e tentar melhorar o seu design e implementação.

Neste processo as diversas fontes externas são acedidas, sendo efetuada a leitura dos dados e de seguida copiados para o sistema de ETL, onde são efetuadas todas as manipulações necessárias. Como os dados provêm de diferentes fontes é necessário efetuar várias transformações, entre elas, a limpeza de dados (correção de erros ortográficos, falta de elementos, etc), a combinação dos vários dados e a verificação de duplicados. É importante nesta fase garantir a qualidade dos dados que serão usados posteriormente pelo DW. O carregamento dos dados é a fase em que a informação é colocada disponível para o utilizador.

Em resumo, o ETL é um processo de transferência da fonte de dados para um data warehouse e uma etapa importante da transformação dos dados das várias fontes, estes dados que vão ser então uniformizados de modo a facilitar a interpretação e implementação do data warehouse.

Segundo (Jun, Kai, Yu, & Gang, 2009) a maneira precisa de desenhar o processo de ETL, de modo com que este seja eficiente, flexível e com manutenção, é com a divisão do processo em cinco módulos: : extração de dados, validação de dados, limpeza dos dados, conversão de dados e carregamento de dados. Em seguida será representado o processo de ETL na figura 2.

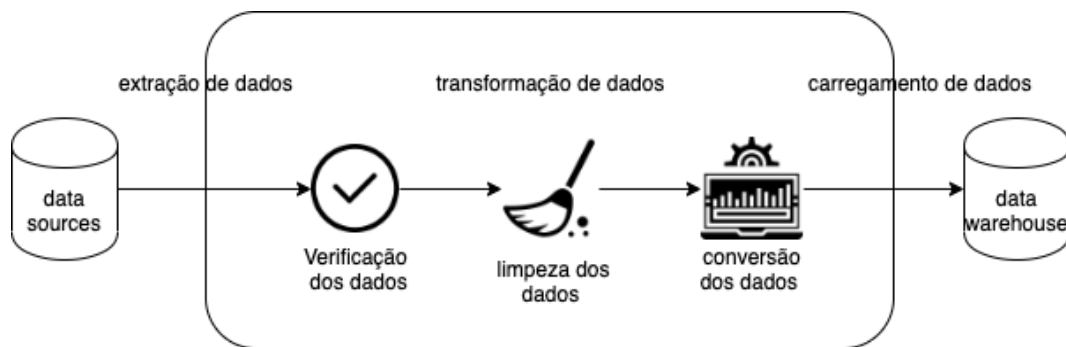


Figura 2 - Processo de ETL adaptado e traduzido (Jun et al., 2009)

- Extração de dados- Esta é a primeira etapa do processo, e requer uma análise do negócio e do sistema de onde os dados vieram, de modo a perceber qual o sistema de origem dos dados, que tipo de sistema de gestão de base de dados o servidor executa. Em segundo lugar, precisamos saber que tipo de estrutura da tabela que a base de dados possui e o correspondente significado de cada estrutura da tabela. Em terceiro lugar, precisamos verificar a existência de dados manuais e a quantidade dos mesmos. Em quarto e último lugar, devemos definir se existem os dados não estruturados.
- Validação de dados- A validação de dados envolve muito trabalho de verificação, incluindo o valor efetivo da propriedade, verificação da chave estrangeira e assim sucessivamente. Quanto aos dados de baixa quantidade, podemos recusá-los inicialmente, e em seguida, esses dados serão armazenados em ordem e corrigidos no campo da alteração;
- Limpeza de dados- A tarefa da limpeza de dados é filtrar os dados indesejáveis. Esses dados indesejáveis incluem: dados incompletos, dados errados, dados duplicados e assim por diante;
- Conversão de dados- Esta tarefa envolve os seguintes tipos: mapeamento direto, operações de campo, processamento de cadeia de caracteres, determinação do valor nulo, conversão dos dados, operação dos dados, montar operações e assim por diante;
- Carregamento de dados- Esta é normalmente a última etapa do processo de ETL, os dados são movidos para o data warehouse. Quanto à melhor maneira de carregar dados, a implementação depende do tipo de operação e da quantidade de dados. Existem duas maneiras de inserir ou atualizar dados na tabela do base de dados: SQL inserir / atualizar / excluir ou uma aplicação de carregamento de grandes quantidades de dados.

Data Warehouse

A definição de data warehouse universalmente aceite foi desenvolvida na década de 1980 por Bill Inmon como “*a subject-oriented, integrated, time variant and nonvolatile collection of data used in strategic decision making*” (Inmon, 2002).

O data warehouse faz parte da arquitetura de BI desde o começo e por isso várias metodologias e pessoas tentaram dar nome a este componente, um exemplo disso, The information warehouse desenvolvido pelos investigadores, Devlin e Murphy, e utilizado pela IBM, abrange não apenas o repositório de dados históricos, mas também os vários data marts na sua definição. A expressão pode ser considerada antecessora do termo data warehouse.

Segundo (Imhoff, Galemme, & Geiger, 2003), o papel principal do data warehouse é servir como um repositório de dados que armazena dados de fontes diferentes, tornando-os acessíveis a outro conjunto de armazenamentos de dados, data marts.

O data warehouse pode ser considerado uma metáfora de supply chain, uma vez que podemos considerar os dados “o produto” que é obtido através de “fornecedores” de dados, sendo eles sistemas operacionais ou fontes externas, estes dados que temporariamente serão guardados num “armazém” de dados. Os dados serão depois entregues através de data marts para os “consumidores finais”. Em seguida, será apresentada uma arquitetura genérica de um data warehouse.

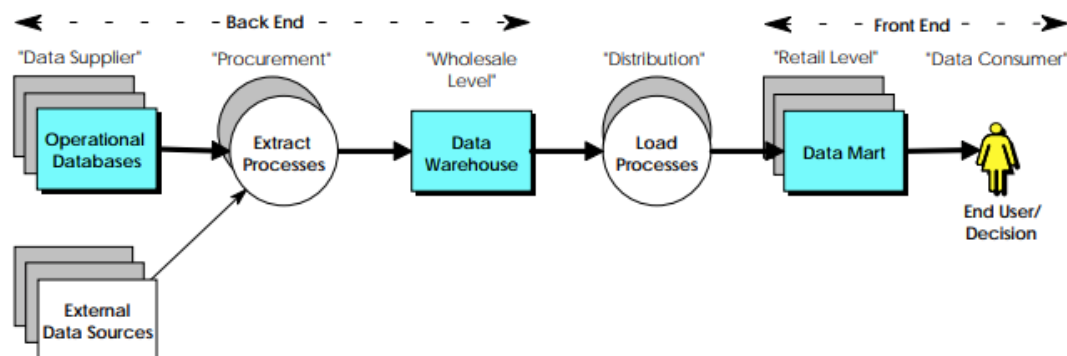


Figura 3 - Arquitetura de data warehouse retirado de (Moody, n.d.)

A arquitetura descrita na figura 3 consiste nos seguintes componentes:

- **Sistemas operacionais:** Sistemas operacionais são os que dão suporte às atividades diárias da empresa, e que guardam os dados relativos ao negócio. São nestes sistemas que maior parte dos dados utilizados para tomadas de decisão são produzidos.
- **Fontes externas de dados:** De modo a apoiar a análise, os data warehouse incorporam dados de várias fontes externas.
- **Processos de extração:** estes processos armazenam os dados no data warehouse de uma forma regular (diariamente, semanalmente, mensalmente). Os dados, extraídos de várias fontes, são depois armazenados num formato consistente. Sem o conjunto único de transformações de regras de negócios que a camada de extração de dados contém, as empresas podem acabar a criar data marts isolados, específicos para utilizadores ou departamentos, estes data marts que geralmente não podem ser combinados para produzir informações válidas.
- **Data warehouse central:** este componente é utilizado para é usado para fornecer data marts. O data warehouse central é geralmente implementado usando um DBMS relacional tradicional.
- **Processo de carregamento:** este é o processo de distribuição dos dados do data warehouse central para os data marts. A infraestrutura de entrega de dados permanece razoavelmente estática, no entanto os data marts necessitam de acompanhar constantemente as mudanças nas necessidades de informações comerciais, isto significa que este processo também deve ser flexível para acompanhar os novos requisitos.
- **Data Marts:** este componente fornece dados com que os utilizadores finais consigam trabalhar. Data marts são geralmente adaptados às necessidades de um grupo específico de utilizadores ou tarefa de tomada de decisão. Data marts

podem ser implementados através de DBMS relacional tradicional ou Ferramentas OLAP.

- Utilizadores finais: escrever consultas e análises de dados armazenados em data marts através da utilização de ferramentas de consulta “user friendly”.

DBMS- data base management systems

Data marts são um subconjunto de dados do data warehouse e são onde a maioria das atividades analíticas ocorrem no ambiente de BI (Imhoff et al., 2003).

Como já foi referido nos componentes da arquitetura do data warehouse, os data marts podem ser definidos como uma coleção de áreas de estudo organizadas para suporte à decisão com base nas necessidades de um determinado departamento (Inmon, 1999). Posto isto, dentro de uma empresa podem existir vários data marts, um data mart pode suportar uma função comercial específica, processo de negócios ou unidade de negócios.

A existência de Data Warehouses e Data Marts pressupõe que as organizações devem ser capazes de identificar qual a arquitetura que melhor satisfaz as suas necessidades (Santos & Ramos, 2009).

À luz das observações anteriores, podemos definir um Data Warehouse como um principal repositório, contendo dados disponíveis para o desenvolvimento de arquiteturas de Business Intelligence e sistemas de suporte à decisão (Vercellis, 2009).

Conforme foi definido pelo, já mencionado, Bill Inmon, este propôs as seguintes características próprias que compõem um Data Warehouse:

- Orientado por assunto- O DW objetiva assuntos, ou seja, os dados são organizados por determinados tópicos de interesse de uma organização. Contrariamente às bases de dados operacionais que apenas são direcionadas à produção e otimização, o DW é direcionado por assunto permitindo uma compreensão mais profunda da organização e os objetos em análise.
- Integrado- 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- O DW varia em relação ao tempo, permitindo manter o histórico de dados durante um período de tempo superior aos sistemas transacionais. Enquanto que por um lado os sistemas transacionais têm como finalidade fornecer as informações no momento exato, os DW por outro lado têm como finalidade analisar o comportamento destas informações no contexto da organização num período de tempo maior. 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- Após os dados serem carregados para o DW, estes não poderão sofrer qualquer alteração, somente sofrendo operações de consulta e exclusão.

Apresentadas as características anteriores relativas a um Data Warehouse, é necessário reforçar que este integra a informação relativa a um determinado assunto da organização. Isto significa que caracteriza a organização como um todo e não subdividida em várias partes, mesmo assim algumas organizações preferem implementar e desenvolver soluções de BI de forma incremental, através de uma série de DM de dados

integrados, em vez de um DW central, de maneira a reduzirem o tempo de implementação e diminuir o número de incertezas dos resultados (Vercellis, 2009).

Cubo OLAP

On-Line Analytical Processing (OLAP) é um método poderoso de suporte à decisão, que analisa grandes quantidades de dados do data warehouse. OLAP caracteriza as operações de consolidar, exibir, aplicar fórmulas e sintetizar dados em várias dimensões.

O software OLAP 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. Um modelo de dados cada vez mais popular para o método OLAP é a base de dados multidimensional, também referido como cubo OLAP. O cubo OLAP é uma estrutura de dados otimizada para análise de dados muito rápida.

Os cubos OLAP são pré-calculados, que resulta num desempenho de *query* rápido, essencial para o uso executivo (Kimball & Ross, n.d.).

Segundo (Sarawagi, Agrawal, & Megiddo, 1998) um cubo de dados 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. Por exemplo, nome da loja → cidade → distrito, é uma hierarquia na dimensão da loja. O facto de estes sistemas estarem organizados em tabelas de facto, com métricas bem definidas e dimensões estruturadas segundo uma hierarquia, faz com que estejam presentes propriedades simples e bastante flexíveis.

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. Como já foi mencionado em cima, estas medidas são obrigatoriamente numéricas, pois apenas desta forma é possível realizar operações matemáticas. Ao conjunto de dimensões e medidas numa entrada pode-se denominar facto.

Em seguida, será representada na figura 4 um esquema multidimensional, que apresenta uma tabela de factos com 3 dimensões (loja, produtos, tempo) e 2 medidas (lucro_euros, unidades_vendas).

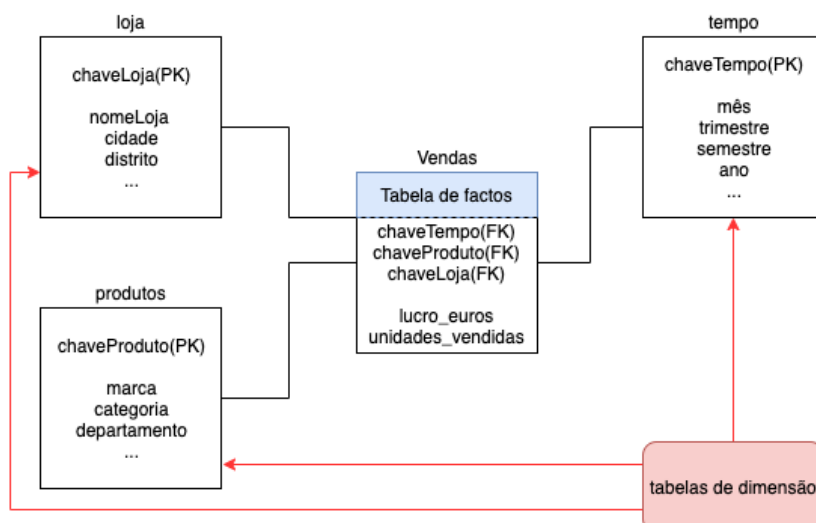


Figura 4- Esquema Multidimensional

Relativamente ao modelo físico, de apoio ao OLAP, existe um debate que está centrado em duas visões principais, a arquitetura ROLAP (Relational On-line Analytical Processing) e a arquitetura MOLAP (Multidimensional On-line Analytical Processing).

A arquitetura MOLAP apresenta a vantagem de apresentar uma visão multidimensional direta dos dados, ao passo que a arquitetura ROLAP é apenas uma *interface* multidimensional para os dados relacionais. Em contrapartida, a arquitetura ROLAP pode ser facilmente integrada noutros sistemas de base de dados relacionais e os dados relacionais podem ser armazenados com mais eficiência do que dados multidimensionais. (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 operações permitem que perspetivas 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 (Golfarelli & Rizzi, 2009):

Operação	Definição
drill-down	analisa os dados em níveis de agregação progressivamente mais detalhados, ou de menor granularidade
roll-up	analisa os dados em níveis de agregação progressivamente menos detalhados, ou de maior granularidade
slice and dice	restringe os dados sendo analisados um subconjunto destes dados slice: corte para um valor fixo dice: seleção de faixas de valores
pivot	reorienta a visão multidimensional dos dados, oferecendo diferentes perspetivas dos mesmos dados
drill-across	compara medidas numéricas distintas que são relacionadas entre si através de pelo menos uma dimensão em comum

Dashboards

Em muitos aspetos, o Dashboard corresponde ao renascimento dos Sistemas de Informação Executivos (EIS) da década de 1980, que acabaram por nunca descolar. Depois disso, já na década de 1990, Kaplan e Norton introduziram Balanced Scorecard, uma metodologia de medição e gestão de desempenho, com ênfase nos principais indicadores de desempenho (KPI's) como forma de fazer medições do estado do negócio, abrindo as portas para os Dashboards. O que realmente impulsionou a popularidade dos Dashboards foi, no entanto, o escândalo da Enron em 2001, um dos maiores casos de banca rota corporativa na história dos Estados Unidos, que deixou as empresas desesperadas por novas maneiras de monitorização e controlo do seu negócio.

As ferramentas de visualização, como é o caso dos dashboards facilitam a tomada de decisão, pois fornecem uma maneira útil de visualizar dados e informações. Os resultados desta ferramenta podem exibir métricas únicas, análise gráfica de tendências, medidores de capacidades, mapas geográficos, comparações de variações, etc. Um design de interface do tipo dashboard permite aos utilizadores facilmente compreender e “digerir” relacionamentos complexos e métricas de desempenho. Mais especificamente, este tipo de design de interface consegue diminuir significativamente a curva de aprendizagem e aumenta a probabilidade de utilização eficaz da ferramenta.

Um dashboard é uma exibição visual das informações mais importantes, necessárias para alcançar um ou mais objetivos, consolidados e organizados numa única tela para que as informações possam ser monitorizadas rapidamente. (Few, 2006)

A informação num dashboard é apresentada de uma forma visual, por norma, uma combinação de gráficos e texto, com ênfase nos gráficos, uma vez que é possível comunicar de uma forma mais eficaz e com significados mais ricos através de gráficos do que simplesmente com texto.

Um dashboard para ser eficaz precisa de equilíbrio entre o apelo visual dos ícones, estrutura dos ícones visuais, quantidade de informação, utilidade e relevância da informação.

O valor real dos produtos de dashboard reside na sua capacidade de substituir as técnicas de extração de dados *hunt-and-peck* por um mecanismo incansável, adaptável e de fluxo de informações. Os dashboards transformam repositórios de dados em informações de consumo. (Babu & Sastry, 2014)

Dashboards e visualização são ferramentas cognitivas que melhoram a "amplitude de controlo" sobre muitos dados de negócio. Essas ferramentas ajudam as pessoas a identificar visualmente tendências, padrões e anomalias, raciocinar sobre o que veem e ajudar a orientá-las em direção a decisões efetivas. (Brath & Peters, 2004)

Definir e entender KPIs é fundamental para um bom design de dashboards. Um KPI pode ser definido como uma medida que é usada para avaliar o desempenho face à meta futura. Os KPIs capturam medidas de desempenho, estas medidas são úteis para realizar uma autoavaliação do que se passa na organização e definir ou redefinir metas. Os KPIs criam a base e o contexto para as informações que serão apresentadas visualmente no dashboard. Como exemplo, o KPI pode ser custo por unidade, vendas mensais ou número de pacientes.

Adaptive Business Intelligence

Adaptive Business Intelligence é a disciplina que usa técnicas de previsão e otimização para criar sistemas self-learning de decisões (Michalewicz, 2007), ou seja, pode ser caracterizado como a combinação entre o sistema de Business Intelligence, data mining, métodos de previsão, técnicas de otimização e ainda adaptabilidade.

A utilização da plataforma de Adaptive Business Intelligence é uma tendência muito recente no mercado. Esta nova abordagem de business intelligence permite elaborar uma recomendação do melhor curso de ação, baseado em dados armazenados anteriormente. Um sistema Adaptive Business Intelligence incorpora módulos de previsão e otimização para recomendar decisões quase ideais e um "módulo de adaptabilidade" para melhorar as recomendações futuras (Michalewicz, 2007).

Os processos presentes no sistema de Adaptive Business Intelligence são similares aos presentes no Business Intelligence, tal como mostra a Figura 5, que se segue:

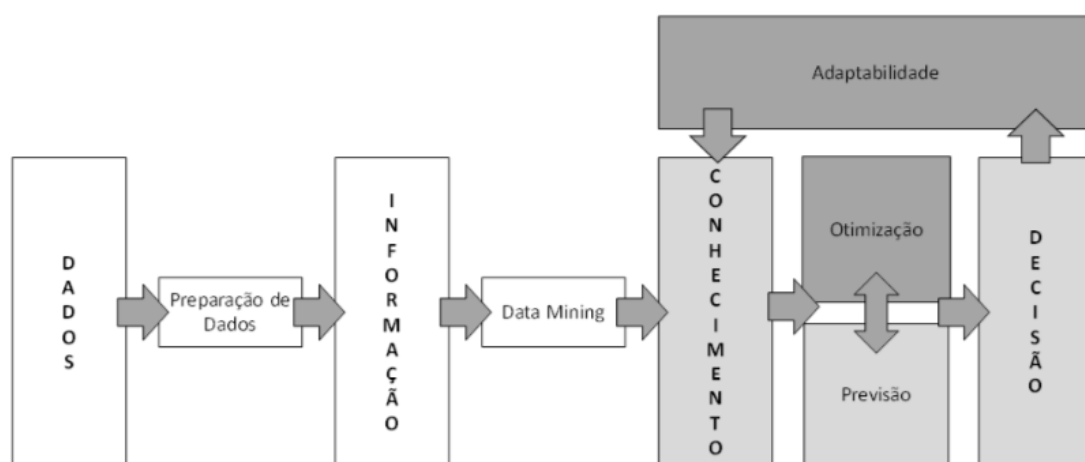


Figura 5- Processos de Adaptive Business Intelligence traduzido e adaptado de (Michalewicz, 2007).

Na primeira fase, os dados são extraídos e sofrem uma transformação para posteriormente serem armazenados. Esta transformação é composta por várias atividades, desde normalização dos dados, criação de atributos derivados, seleção de variáveis, eliminação de dados, suprimento de valores em falta e a consequente limpeza dos dados. Em seguida, segue-se a fase do data mining, cuja função é elaborar um processo analítico dos dados armazenados, de maneira a encontrar padrões entre variáveis e identificar quais as variáveis mais relevantes. Na fase da previsão, 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 no conjunto de dados do sistema (Michalewicz, 2007).

A otimização é uma técnica de procura de melhores valores de parâmetros. O objetivo desta fase será pesquisar os valores dos parâmetros que minimizam o erro de previsão, baseado nos dados do modelo de previsão. Isto permite diminuir os desperdícios dos turnos, sendo que para tal é preciso prever os agendamentos dos mesmos (Michalewicz, 2007).

Por último, temos a fase da adaptabilidade, que apresenta grande importância e permite medir o erro da previsão. Compara os valores obtidos pelo modelo de previsão com os dados em tempo real, permitindo ao modelo adaptar-se às mudanças no ambiente exterior. Assim, o modelo de adaptabilidade será capaz de se melhorar ao longo do tempo, aprendendo com os seus próprios erros de previsão.

Como já foi mencionado, os processos envolvidos nos sistemas de Adaptive Business Intelligence e Business Intelligence têm muitas semelhanças, uma vez que apenas diferem na fase de aquisição de conhecimento. Posto isto, verificamos que o sistema de Business Intelligence terá como foco principal o apoio às decisões, enquanto que o sistema de Adaptive Business Intelligence terá também a capacidade de usar os dados/resultados da organização para obter previsões, otimizar essa mesma previsão e diminuir os seus erros, tendo por isso uma capacidade de se adaptar às mudanças externas. Tendo em consideração, que os problemas com os sistemas de Adaptive Business Intelligence têm de lidar são mais complexos, é imperial oferecer um melhor processo de decisão, assim sendo, um sistema Adaptive Business Intelligence deve incluir três componentes principais: (1) um componente para fazer previsões, (2) um componente para tomar decisões quase ideais e (3) um componente para adaptar o módulo de previsão às mudanças no ambiente (Michalewicz, 2007).

Como todas as organizações operam em ambientes que mudam com o tempo, são forçadas a adaptar-se e ajustar constantemente. Os sistemas baseados em Adaptive Business Intelligence visam resolver problemas do mundo real, que apresentam:

- Restrições complexas, o problema é tão severamente restrito que é difícil encontrar até uma solução viável;
- Ambientes mutáveis com o tempo, um ambiente de mudança de tempo requer um conjunto de soluções (em vez de uma única solução);
- Multiobjectivos (possivelmente conflitantes);
- Grande quantidade de soluções, o número de soluções possíveis para o problema é tão grande que proíbe qualquer pesquisa sistemática da melhor solução.

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

- Que os dados disponíveis sejam preparados e analisados minuciosamente através de data mining;
- Que o modelo de previsão seja desenvolvido tendo por base os resultados da fase anterior, nomeadamente a fase de data mining;
- Que o modelo de otimização seja desenvolvido usando o modelo de previsão, por forma a este recomendar a melhor solução;
- Que o modelo de adaptabilidade seja desenvolvido, por forma a ser responsável na adaptação do modelo de previsão as mudanças externas.

Embora o principal objetivo do business Intelligence seja ajudar os gestores a tomar decisões mais rápidas e inteligentes, as organizações agora percebem que existe um vasto abismo entre ter o "conhecimento certo" e tomar a "decisão certa". As organizações precisam de sistemas que possam prever o futuro e recomendar o melhor curso de ação. Por esse motivo, o futuro do setor de business intelligence está em sistemas que possam prever, otimizar e adaptar (Michalewicz, 2007).

O poder e o apelo dos sistemas Adaptive Business Intelligence residem na sua capacidade de responder às duas questões fundamentais por trás de todas as decisões de negócios: "O que provavelmente acontecerá no futuro?" e "Qual é a melhor decisão agora?". Sem dúvida, as organizações que conseguirem responder com precisão a estas perguntas de maneira consistente terão uma vantagem competitiva sobre aquelas que não conseguem (Michalewicz, 2007).

Data Mining

A noção de data mining surgiu na década de 1990 e pode ser vista como resultado da evolução natural das tecnologias de informação. O setor de base de dados e gestão de dados evoluiu no desenvolvimento de várias funcionalidades críticas: recolha de dados e criação de base de dados, gestão de dados (incluindo armazenamento, recuperação de dados e processamento de transações de base de dados) e análise avançada de dados (envolvendo data warehousing e data mining). Atualmente, vários sistemas de base de dados oferecem processamento de query e transações como prática comum. A análise avançada de dados tornou-se naturalmente o próximo passo (Han, Kamber, & Pei, 2012).

O desenvolvimento do data mining está relacionado com o progresso do hardware nas últimas três décadas, este progresso fez com que o mercado dos computadores ficasse mais acessível, tendo em consideração que os computadores ficaram mais baratos e aumentaram a performance. O aumento nas capacidades de hardware levou a grande

impulso do setor de base de dados e informações e permitiu que um grande número de base de dados e repositórios de informações esteja disponível para gestão de transações, recuperação de informações e análise de dados (Han et al., 2012). Os avanços nos sistemas de gestão de base de dados para aproveitar esse paralelismo de hardware também beneficiaram o data mining (Edelstein, 1999).

Data mining 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 (Edelstein, 1999).

Também pode ser visto como a ciência de explorar grandes conjuntos de dados para extrair informações implícitas, anteriormente desconhecidas e potencialmente úteis (Gorunescu, 2011)

O termo data mining é muitas vezes usado como sinônimo de outro termo popularmente usado, *knowledge discovery from data* ou KDD, pode, no entanto, ser apenas associado a uma etapa essencial no processo de descoberta de conhecimento. O processo de descoberta de conhecimento é mostrado em seguida, como uma sequência iterativa das seguintes etapas:

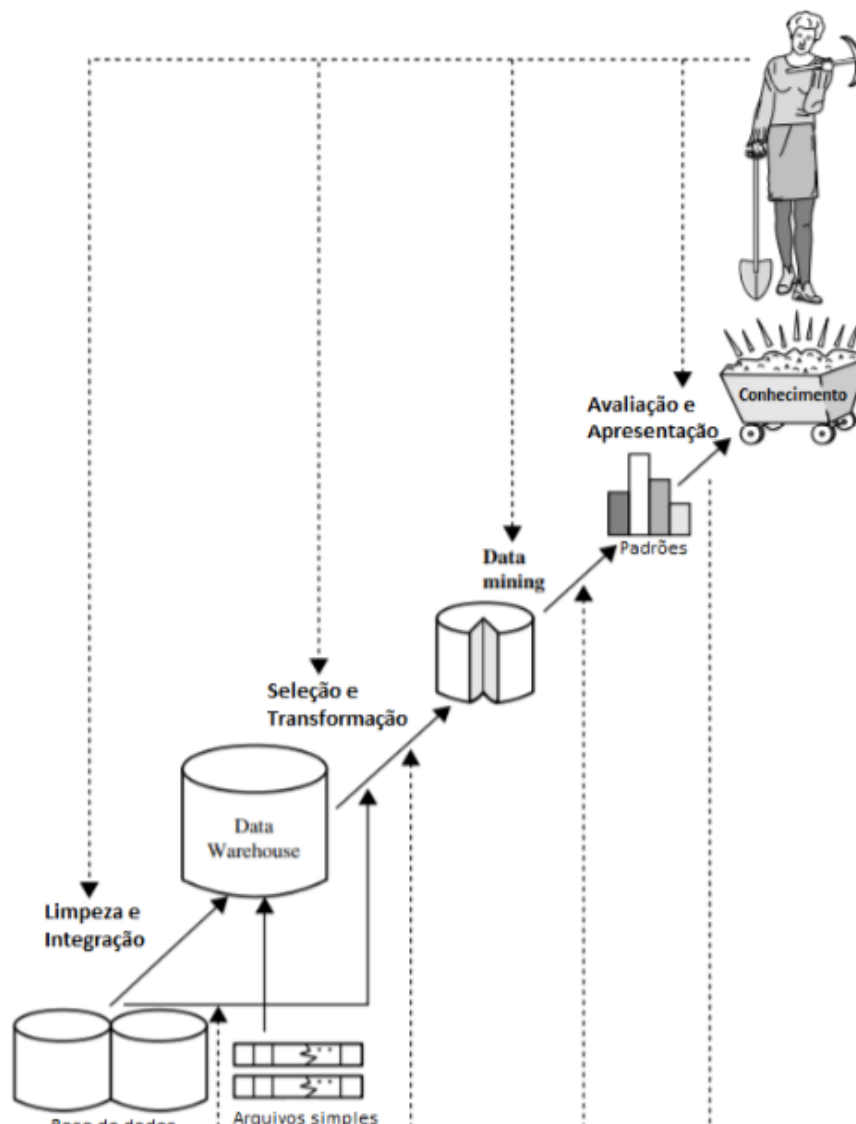


Figura 6- Processo de descoberta de conhecimento traduzido e adaptado de (Han, Kamber, & Pei, 2012).

- Limpeza dos dados (para remover ruído e dados inconsistentes);
- Integração dos dados (onde múltiplas fontes de dados podem ser combinadas);
- Seleção dos dados (onde os dados relevantes para análise são recolhidos da base de dados);
- Transformação dos dados (onde os dados são transformados e consolidados em formas apropriadas para data mining);
- Data Mining (um processo essencial em que métodos inteligentes são aplicados para extrair padrões de dados);
- Avaliação de padrões (identifica os padrões verdadeiramente interessantes que representem conhecimento baseado em medidas interessantes);
- Apresentação do conhecimento (onde a visualização e as técnicas de representação de conhecimento são usadas para apresentar os resultados do data mining aos utilizadores).

As 4 primeiras etapas do processo são diferentes formas de processamento de dados, onde os dados são preparados para a etapa 5, data mining. Como já foi mencionado, data mining tanto é considerado o processo de descoberta de conhecimento, como apenas uma etapa do mesmo. Posto isto, (Han et al., 2012) definiu o termo como o processo de descobrir padrões e conhecimentos interessantes de grandes quantidades de dados. As fontes de dados podem incluir bases de dados, data warehouses, a Web, outros repositórios de informações ou dados transmitidos dinamicamente para o sistema.

Os objetivos do processo de descoberta de conhecimento são definidos pelo uso pretendido do sistema. Podemos distinguir dois tipos de objetivos: **Verificação** e **Descoberta**. Com a verificação, o sistema limita-se a verificar as hipóteses do utilizador. Com a descoberta, o sistema autonomamente encontra novos padrões. Ainda se pode subdivir o objetivo de descoberta em **previsão**, onde o sistema encontra padrões para prever o futuro comportamento de algumas entidades e **descrição**, onde o sistema encontra padrões para apresentar a um utilizador num formato que seja possível o entendimento humano (Fayyad, Piatetsky-Shapiro, & Smyth, 1997).

Data mining envolve então uma grande análise de dados, por forma a descobrir padrões e ajustar modelos de previsão. Os modelos ajustados desempenham o papel do conhecimento inferido. Dois formalismos matemáticos primários são usados no ajuste do modelo: **estatístico** e **lógico**. A abordagem estatística permite efeitos não determinísticos no modelo, enquanto um modelo lógico é puramente determinístico (Fayyad et al., 1997). A abordagem estatística tende a ser mais amplamente usada para aplicações práticas de data mining, dada a presença típica de incerteza nos processos de geração de dados do mundo real.

A maioria dos métodos de data mining é baseada em técnicas experimentadas e testadas de machine learning, reconhecimento de padrões e estatística. Os objetivos da previsão e descrição podem ser obtidos com uso de uma variedade de métodos específicos de data mining. (Fayyad et al., 1997):

- **Classificação:** O modelo é derivado com base na análise de um conjunto de dados de treinamento (ou seja, objetos de dados pelos quais os rótulos de classe são conhecidos). O modelo é usado para prever o rótulo da classe de objetos para os quais o rótulo da classe é desconhecido (Han et al., 2012).

- **Regressão:** este método aprende uma função que mapeia um item de dados para uma variável de previsão com valor real. A análise de regressão é uma metodologia estatística mais frequentemente usada para previsão numérica.
- **Clustering:** é uma tarefa descritiva comum em que se procura identificar um conjunto finito de categorias ou clusters para descrever os dados (Jain e Dubes, 1988; Titterington, Smith, and Makov, 1985).
- **Summarization:** envolve métodos para encontrar uma descrição compacta para um subconjunto de dados. Um exemplo simples seria tabular a média e os desvios padrão para todos os campos (Agrawal, 1996).
- **Modelação de dependências:** consiste em encontrar um modelo que descreva dependências significativas entre variáveis. Os modelos de dependência existem em dois níveis: (1) o nível estrutural do modelo especifica (geralmente em forma gráfica) quais as variáveis são localmente dependentes uma da outra e (2) o nível quantitativo do modelo especifica os pontos fortes das dependências usando alguma escala numérica.

O resultado da construção do modelo é uma descrição de padrões e relacionamentos nos dados que podem ser usados com confiança para previsão. De modo a desenvolver um modelo preciso e de confiança é importante ter em atenção a seguinte hierarquia de escolhas (Edelstein, 1999):

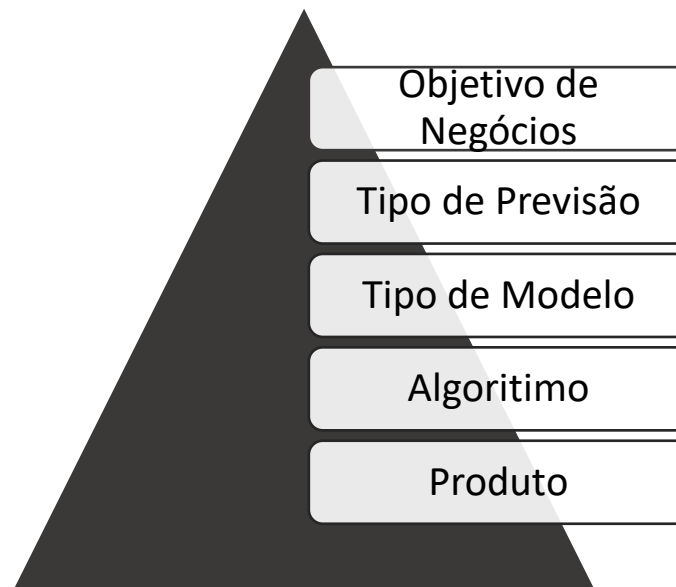


Figura 7- Hierarquia de escolhas de data mining adaptado de

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, sendo que pode ser um modelo de classificação, regressão, etc, como foi mencionado.

Agora, tendo em conta que o tipo de previsão já foi selecionado é possível escolher o tipo de modelo, como por exemplo, uma rede neural para executar a regressão.

Muitos algoritmos estão disponíveis para a construção de modelos, como por exemplo, construir a rede neural com uso de funções de retropropagação ou base radial.

Por último, ao selecionar um produto de data mining, 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.

Muitos objetivos de negócios são melhor atendidos através da criação de vários tipos de modelo usando uma variedade de algoritmos. Posto isto, pode ser importante uma fase experimentação de várias abordagens, de maneira a conseguir determinar qual é o melhor tipo de modelo.

Otimização

Atualmente existe uma tendência crescente de adoção de ferramentas computacionais baseadas em métodos de otimização, para analisar os dados de mundo real e tomar decisões mais informados. Em vez de tomar decisões com base na experiência ou intuição humana.

A otimização é um tópico central do processo de Adaptive Business Intelligence, que começou a ser desenvolvido na década de 1940, com a programação linear. Depois, na década de 1960 surgiu branch and bound (Schrijver 1998), estas que são algumas das técnicas clássicas de otimização. Mais recentemente, nas últimas décadas, surgiram novos algoritmos de otimização, frequentemente denominada “otimização moderna” (Michalewicz et al. 2006), “heurísticas modernas” (Michalewicz e Fogel, 2004) ou “metaheurísticas” (Luke, 2012).

Em contraste com os métodos clássicos, os métodos modernos de otimização são solucionadores com um propósito geral, ou seja, aplicáveis a uma ampla gama de problemas distintos, uma vez que é necessário pouco conhecimento de domínio (Cortez, 2014). Por exemplo, problemas com descontinuidades, mudanças dinâmicas, objetivos múltiplos ou restrições rígidas e flexíveis, mais difíceis de serem tratadas pelos métodos clássicos (Michalewicz et al. 2006). Também em contraste com os métodos clássicos, a otimização moderna não garante que a solução ideal seja sempre encontrada. No entanto, os métodos modernos geralmente alcançam soluções de alta qualidade com um uso muito mais razoável dos recursos computacionais (por exemplo, esforço de memória e processamento) (Michalewicz e Fogel 2004).

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 pesquisa, ou seja, o número de soluções possíveis para o problema, e o seu tamanho. Por consequência, a escolha acertada do espaço de pesquisa é de grande importância, pois se o domínio correto não for selecionado, pode impedir que a solução certa seja encontrada. O tamanho do espaço de pesquisa não é determinado pelo problema, mas sim pela sua representação. Depois de selecionado o espaço de pesquisa, é necessário definir qual o objetivo do problema. O objetivo é uma declaração matemática da tarefa que se pretende alcançar, não é uma função, mas uma expressão. Após o objetivo ser claramente definido, a próxima etapa é criar uma função de avaliação que permita comparar a qualidade de diferentes soluções. Algumas funções de avaliação produzem uma classificação para várias soluções, enquanto outras são numéricas e fornecem também uma classificação e uma pontuação de medida de qualidade. Ao comparar as soluções possíveis, podemos facilmente saber se uma solução é melhor que outra e por quanto. É necessário ter em consideração que a técnica de otimização não sabe qual é o problema a ser resolvido,

apenas sabe a representação da solução e a função de avaliação. Se a função de avaliação não corresponder ao objetivo, existe uma procura pela resposta certa para o problema errado (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 é apresentado o seguinte mapeamento:

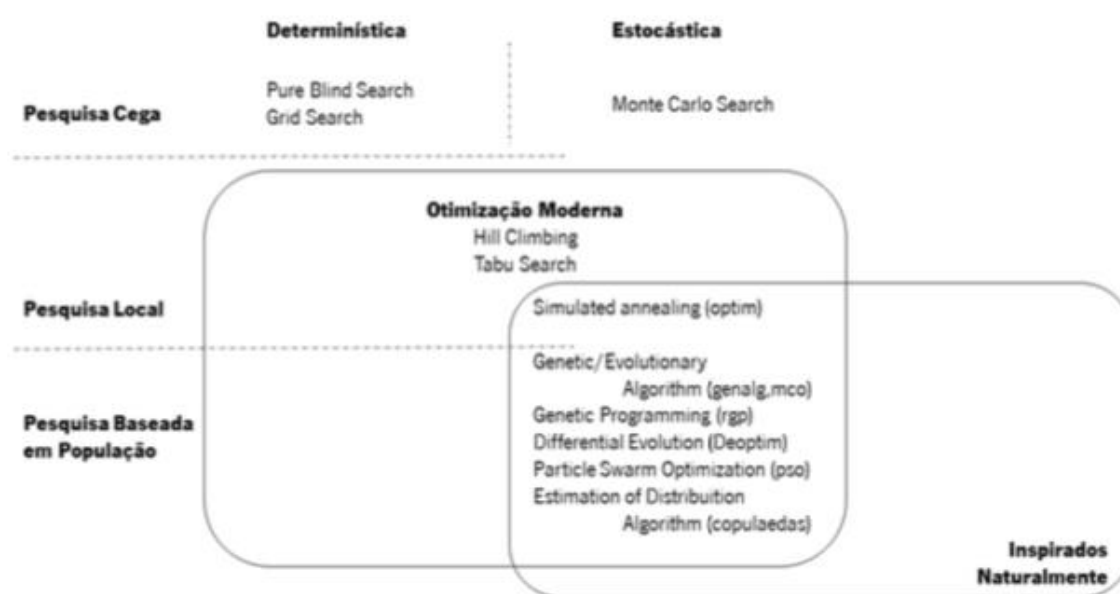


Figura 8- Classificação dos métodos de otimização, adaptado e traduzido de (Cortez, 2014).

Em relação ao tipo de procura, este tipo de classificação subdivide-se em duas categorias, sendo que pode ser as cegas ou guiada. O primeiro pressupõe o esgotamento de todas as alternativas para encontrar a solução ideal, enquanto o último usa pesquisas anteriores para orientar a pesquisa atual (Cortez, 2014). Através da observação da figura 8 é possível concluir que métodos de otimização moderno apenas utilizam o tipo de procura guiada. Na procura guiada esta pode ser local ou pode ser global. Na procura local é feita uma pesquisa na vizinhança de um espaço até se encontrar a solução ótima, enquanto que por sua vez, na procura global, o seu funcionamento é através de uma pesquisa envolvendo um conjunto de populações. A procura global 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. Grande parte dos algoritmos de otimização baseados em população são inspirados naturalmente (Sean, 2012).

Vários métodos modernos empregam algum grau de aleatoriedade, pertencendo à família de métodos de otimização estocástica (Luke 2012). Este tipo de métodos são, na sua maioria, naturalmente inspirados.

O processo de otimização moderna, devido á complexidade dos seus problemas, pode lidar com várias restrições para permitir ao utilizador obter o resultado final dentro das suas prioridades. Existem dois tipos de restrições (Michalewicz, 2007): *hard* e *soft*. Restrições *hard* não podem ser violadas e são devidas a fatores como leis ou restrições físicas. As restrições *soft* estão relacionadas a outros objetivos do utilizador (geralmente não prioritárias), como aumentar a eficiência da produção e reduzir os custos ambientais (Cortez, 2014).

Para evitar que as restrições *hard* possam gerar soluções pouco viáveis são aplicados procedimentos para tratar esse tipo de problemas, há vários métodos que podem ser aplicados (Michalewicz et al., 2007):

- “penalidade de morte”, é aplicado um valor demasiado elevado na pesquisa, para que essa solução seja descartada. No entanto não é muito eficiente, pode acontecer de se descartar soluções que podem conter informações uteis;
- penalidade de peso, é ideal para procedimentos com poucas restrições, o objetivo desta solução é aplicar um ou mais “pesos” à solução ótima até ser encontrada a solução ideal. No entanto esta solução que é fácil de implementar, pode tornar-se bastante difícil de encontrar os “pesos” ótimos.
- reparação, faz com que uma solução que não é viável se torne viável. Isto pode ser alcançado utilizando informação dependente do domínio ou através da procura de novas soluções na vizinhança;
- gerar soluções flexíveis, apenas se usam soluções viáveis ou flexíveis, descartando-se todas as soluções que não são viáveis para o procedimento da otimização (Michalewicz et al., 2007).

Previsão

Os algoritmos de previsão são basicamente métodos supervisionados, nos quais dados de entrada e saída válidos estão disponíveis para o processo de treinamento. Durante a fase de treinamento, os algoritmos tentam determinar os relacionamentos existentes nos dados para corresponder ao resultado "conhecido". Durante esse processo, as regras são geradas para prever o resultado para um novo conjunto de dados desconhecido (Babu & Sastry, 2014).

A maioria dos problemas de previsão podem ser categorizados como:

- Problemas de classificação;
- Problemas de regressão;
- Problemas de *time series*.

A quando 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, na verdade, o número de classes possíveis pode ser bastante significativo. Nesses problemas de classificação, o tempo não existe; o “futuro” é entendido como a chegada de um novo caso ou é incluído como uma variável do caso (Michalewicz, 2007).

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).

Os problemas de *time series*, em contraste com os problemas de classificação e regressão, dão grande importância ao tempo, neste tipo de problemas cada caso contém vários valores medidos ao longo de um período de tempo no passado. Em problemas de *time series*, o futuro é mencionado explicitamente nestes modelos, ou seja, permite prever o valor de uma variável no futuro (Michalewicz, 2007).

O processo de criar um modelo previsão, por norma, consiste 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.

A construção de modelos preditivos é um processo iterativo no qual um modelo é criado a partir de uma hipótese inicial e depois refinado até produzir um resultado comercial valioso ou descartado em favor de outro modelo com mais potencial (Babu & Sastry, 2014).

Muitos métodos de previsão foram desenvolvidos ao longo dos anos, alguns destes métodos apenas diferem entre si na representação de uma solução, bem como na capacidade de explicar a previsão ou facilidade em editar uma solução. Posto isto, selecionar o melhor método de previsão é um processo complicado que deve ter em conta várias considerações. Embora o erro de previsão seja possivelmente a medida mais importante, ele fornece apenas uma dimensão da qualidade de um modelo. Para problemas de negócios do mundo real, outros fatores devem ser considerados, como (Michalewicz, 2007):

- Tempo de resposta-Existem métodos de previsão que lidam com uma grande frequência de previsões, para este tipo de métodos o tempo de resposta é um fator crítico;
- Editar o modelo- Alguns modelos de previsão são difíceis de editar, enquanto outros são fáceis, a edição possibilita a alteração do modelo no futuro e pode ser um fato importante na escolha;
- Justificação da previsão- Para algumas aplicações dos métodos de previsão, é muito importante justificar a previsão, em alguns casos, isso pode até ser exigido por lei;
- Tolerância a ruído- Todos os métodos de previsão requerem alguma abordagem para a deteção de valores ausentes ou imprecisos. Alguns modelos podem lidar melhor com esses dados que outros;

- **Modelo compacto-** Um modelo de previsão não deve ser excessivamente grande e complexo, pois isso dificultaria o entendimento dos humanos e pode demorar mais tempo para fazer previsões.

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 usadas na área da Saúde.

Adaptive Business Intelligence é uma tendência muito recente e pouco utilizada no mercado, visto que a maioria das organizações já utilizam Business Intelligence. Por se tratar de um método ainda recente pode ser a causa de ser pouco utilizado na área da saúde.

Caso prático

Este caso pratico consiste na criação de uma plataforma de ABI, que resulta de uma parceria com uma instituição hospitalar portuguesa, que forneceu o dataset para a investigação.

A ideia que move este caso prático é a ausência de abordagens de Business Adaptive Business Intelligence (ABI) em hospitais e outras instituições de saúde para melhorar a qualidade do serviço por meio da programação eficiente de cirurgias no processo de decisão. (Rocha, Adeli, Reis, & Costanzo, 2019). A principal tarefa deste caso pratico é então fornecer uma plataforma baseada em ABI capaz de estimar o tempo das cirurgias e, em seguida, otimizar o agendamento (minimizando o desperdício de recursos).

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. Esse tipo de atividade é crítica e geralmente está associado a atrasos constantes e reprogramação significativa que podem levar a grandes perdas em recursos humanos e financeiros. Combinando dados operacionais com ferramentas analíticas, a plataforma ABI é capaz de apresentar informações complexas e competitivas para otimizar o agendamento de cirurgias.

O caso pratico foi realizado para estender os recursos de uma plataforma de BI hospitalar já existente com os recursos da ABI. Com a utilização desta nova plataforma o hospital beneficiará de uma melhoria na qualidade das informações que possui. A plataforma de ABI pode ajudar a prever o tempo da cirurgia e agendar cirurgias de otimização ao longo dos turnos disponíveis. Foram utilizados dados reais retirados do hospital.

A plataforma criada foi desenvolvida baseada na arquitetura de Adaptive Business Intelligence. A plataforma consiste em dois blocos, o bloco de previsão que foi desenvolvido com a ferramenta Weka e o bloco de otimização que foi desenvolvido com a ferramenta RStudio. O bloco de previsão prevê a duração de cada cirurgia ou mais geralmente um evento, enquanto que o bloco de otimização visa minimizar o desperdício de turnos, evitando atrasos e aumentando a eficiência e a eficácia do cronograma.

Bloco de Previsão

O primeiro passo é prever o tempo padrão de cada evento (cirurgia), por meio do treinamento de modelos de data mining, na ferramenta Weka. As técnicas utilizadas são as seguintes: Regressão Linear (Regressão Linear); k-Vizinhos mais próximos (IBk); Árvore de Decisão (REPTree); Regressão de vetores de suporte (SMOreg); Perceptron de várias camadas (MultilayerPerceptron). O *target* corresponde ao evento a ser previsto,

neste caso a duração das cirurgias. A seleção de recursos é obtida através do algoritmo "CorrelationAttributeEval", que faz um ranking das ocorrências dessas e, em seguida, utiliza as 10 a 15 características mais relevantes. Em seguida os modelos foram treinados e analisados de modo a escolher o modelo com a taxa de erro mais baixa.

Bloco de Otimização

O resultado do bloco de previsão é um novo conjunto de dados que contém eventos com o tempo de duração previsto no bloco anterior. Os métodos usados neste bloco são *hill climbing* e *simulated annealing*. A solução inicial (par) foi obtida em turnos aleatórios ou sequenciais através da lista de cirurgias de especialidades correspondentes ao turno, atribuindo um turno a cada índice de dimensão. Os parâmetros inferior e superior representam os valores mais alto e mais baixo para cada dimensão, variando de 1 ao número de turnos de uma dimensão. 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 turno. É também nesta parte da função que as restrições são definidas e os caminhos indesejados são penalizados, atribuindo um valor alto ao desperdício, para que essa solução não seja escolhida. As restrições encontradas para o bom funcionamento da otimização foram: Data do evento <Data do turno e prazo do evento> Data do turno e desperdício > = 0. O limite de iteração será o número que passará pela otimização e controle e permitirá a apresentação de um relatório contendo a solução de cada uma das iterações, onde cada solução a seguir é alterada pela função de mudança. A função de mudança visa gerar a próxima solução, que sofrerá pequenas alterações na solução gerada anteriormente, e assim por diante. Para entender o melhor modelo a ser utilizado como método de otimização, será necessário comparar as duas soluções obtidas, escolhendo o que para cada especialidade cirúrgica apresenta o menor desperdício.

Um conjunto de dados foi utilizado para prever o tempo de duração da cirurgia e

Especialidade	Nº de turnos	Quantidade de cirurgias	Tempo de turnos standard	Tempo de cirurgias standard	Desperdício total
Oftalmologia	1078	3162	34905600	5307459	29598141
Oftalmologia pediátrica	23	30	909900	116817,8	793082,2
Cirurgia vascular	569	796	19389000	3456938	15932062
Ortopedia	875	1440	26650800	9509328	17141472
Ortopedia pediátrica	94	163	3187800	848623,6	2339176,4

outro foi usado para a otimização. Essa divisão acontece porque a tabela de cirurgias disponíveis para serem utilizadas no bloco de otimização não possuem cirurgias suficientes para o processo de *forecasting*.

Para testar o bloco de previsão, foram utilizados dados das especialidades cirúrgicas que incluíram o número mais significativo de cirurgias em todo o conjunto de dados, a especialidade ortopedia. Em seguida, será apresentada uma tabela que identifica

as variáveis que parecem ter um poder mais descritivo pela ocorrência de apenas alguns valores, selecionando os 14 atributos mais relevantes para a tarefa de previsão.

<i>Atributos</i>	<i>Pontuação</i>
TEMPO6	0.969341
TEMPO7	0.942850
TEMPO2	0.677621
TEMPO8	0.376831
SERVICE	0.345452
GDH	0.299350
MCONBASE	0.296299
COD_ANESTESIA	0.233695
PROCS1	0.233320
INTERVENCAO1	0.232584
COD_INTERV_CIRURGICA	0.230257
COD_SALA	0.144060
PROCS8	0.117009
COD_PATOLOGIA	0.092911

Tabela 2- Atributos com maior poder descritivo

Depois de selecionados os atributos podemos ver o sumário dos resultados obtidos pelos modelos de previsão com os 14 atributos na seguinte tabela.

REPTree	IBk	SMOreg	LinearRegression	MultilayerPerceptron
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%	6.2223%	9.0973%
Root relative squared error	9.2617%	65.1079%	7.5853%	9.7117%
Total number of instances	237	237	237	237

Tabela 3- Sumario dos modelos de regressão com os 14 atributos

Analisando as percentagens de erro de cada modelo, conclui-se que o algoritmo de regressão de vetores de suporte (SMOreg) é mais eficiente (porque é o que apresenta a menor percentagem em torno de 6% de erro absoluto relativo). A Tabela 4 apresenta um exemplo contendo dez cirurgias da especialidade ortopedia, o tempo de duração esperado, a duração prevista e o erro.

<i>Cirurgias</i>	<i>Atual</i>	<i>Previsão</i>	<i>Erro</i>
1	8100	8049.027	-50.973
2	7020	7127.73	107.73

3	15300	15304.194	4.194
4	5100	5165.485	65.485
5	9000	9026.354	26.354
6	4200	4267.64	67.64
7	9300	9331.48	31.48
8	10560	10761.003	201.003
9	6960	7042.839	82.839
10	7860	7721.515	-138.485

Tabela 4- Amostra de previsão da duração das cirurgias ortopédicas (tempo em segundos)

Para a tarefa da otimização, como já foi mencionado em cima, os métodos usados são *hill climbing* e *simulated annealing*. Estes métodos foram utilizados em 32 especialidades cirúrgicas, permitindo otimização anual e mensal, para encontrar a melhor solução. O primeiro passo é obter a solução inicial, dada aleatoriamente ou sequencialmente, através dos turnos necessárias e do tamanho das cirurgias.

Em seguida, é aplicada a função de avaliação, que visa passar pelas restrições e, em seguida, somar o total de resíduos de cada turno. Cada valor de resíduo é obtido subtraindo o tempo padrão das cirurgias, considerando um tempo de rotatividade de 17 min (exceto a última cirurgia), no momento do turno designado e, no final, o total de resíduos é devolvido à solução obtida. Para cada uma das iterações, um conjunto completo de valores de resíduos é obtido para ser comparado entre soluções alternativas.

As iterações terminam no limite designado no método de otimização, que irá devolver o melhor valor das soluções obtidas. No final, a solução satisfatória é escolhida entre os dois métodos de otimização testados.

Para o conjunto de dados de cirurgias e turnos, apenas serão discutidos os resultados de duas especialidades, a com menor e maior número de cirurgias, pois basta entender a minimização dos resíduos. A especialidade com maior tempo de duração da cirurgia é a cirurgia vascular, que contém 796 cirurgias (3456938 s) a serem distribuídas por um total de 19389000 s, correspondendo a 569 turnos possíveis. Dois tipos de otimização foram aplicados para resolver esse problema: otimização global por um ano; e otimização distribuída para cada mês do ano. A otimização foi realizada para cada um dos 12 meses que tiveram cirurgias pendentes, utilizando como dimensão a capacidade de cirurgias de uma respetiva especialidade para o respetivo mês.

Os resultados para cada mês são como na Tabela 5.

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

Tabela 5- Otimização sequencial mensal de desperdícios em cirurgia vascular.

Para esta especialidade, pode-se concluir que o objetivo foi alcançado através da otimização mensal. Após análise dos resultados, o método de otimização mensal do processo seria o mais rentável para o agendamento de cirurgias na especialidade ortopédica. 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 (o que nunca poderia acontecer com grandes conjuntos de dados).

Os resultados são promissores e tornam essa abordagem uma plataforma de ABI eficiente e eficaz de agendamento de eventos, adaptável a qualquer organização ou instituição que precise agendar grandes listas de eventos. Isso pode resultar numa redução de atrasos imprevistos e um aumento na eficácia do serviço, minimizando o desperdício de turnos.

Metodologias de Investigação

Design Science Research

Para o desenvolvimento da dissertação será utilizada como metodologia de investigação a Design Science Research Methodology (DSRM). Esta metodologia envolve um vasto conjunto de técnicas e perspectivas que conduzem a pesquisa científica, para sistemas de informação, é uma metodologia que se concentra no desenvolvimento e no desempenho de artefactos, com a intenção de melhorar o desempenho funcional destes na área de Sistemas de Informação. Em seguida, na figura 9, estão representadas as diferentes fases da metodologia:

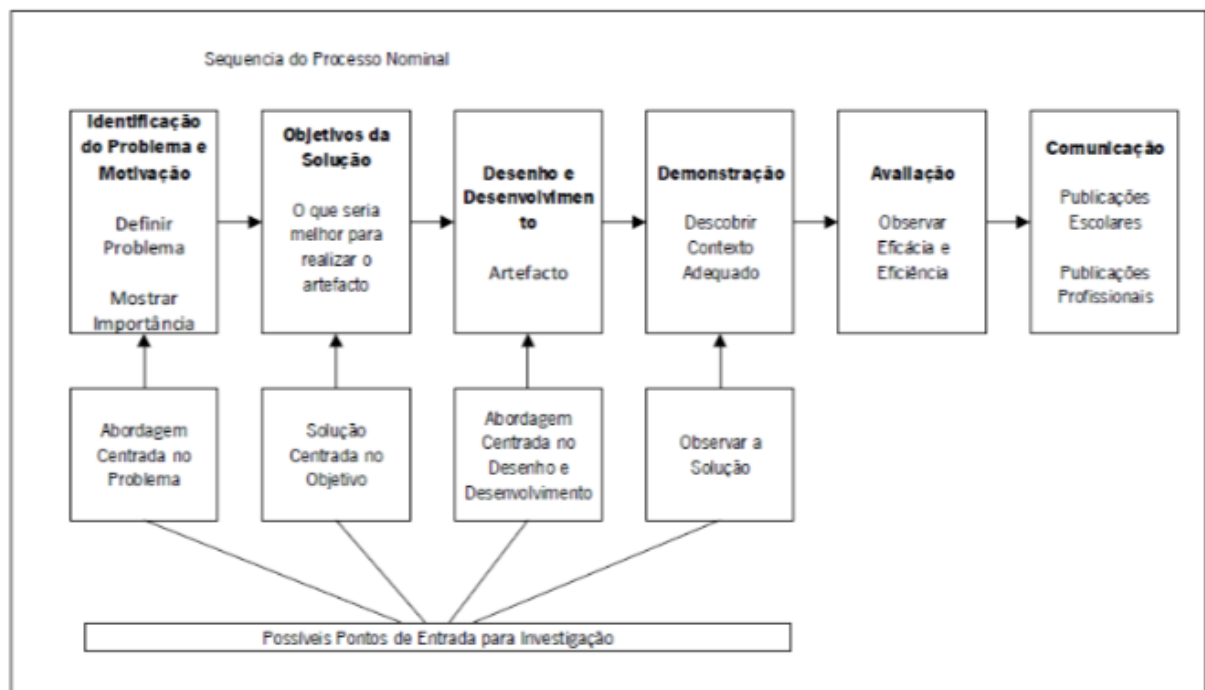


Figura 9- Fases Design Science Research (adaptado de (Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, 2007))

Identificação do problema e motivação

Esta primeira fase de identificação do problema a resolver, consiste em definir a especificidade do problema de investigação e 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 de Adaptive Business Intelligence. 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

Nesta fase são identificados os objetivos que ajudam na resolução do problema e 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.

Para este trabalho, o principal objetivo da criação desta plataforma, passa exatamente por minimizar o desperdício de cada turno, para cada uma das especialidades das cirurgias. Esta otimização é feita através da escolha, por parte do sistema, da combinação que melhor aproveita cada turno, diminuindo-se, assim, os desperdícios.

Desenho e desenvolvimento

Para esta fase de desenho e desenvolvimento será necessário o uso de software de desenvolvimento, onde o resultado será um artefacto. Esta atividade permite determinar a funcionalidade desejada do artefacto e a sua arquitetura, para posteriormente criar o artefacto real. Neste caso a utilização do Adaptive Business Intelligence na área da saúde. De modo a transitar para esta fase é necessário incluir o conhecimento da teoria que será utilizada na solução.

Aqui faz-se o levantamento de técnicas, algoritmos e metodologias pertinentes para a execução do projeto. Para este projeto foram seguidas as fases da metodologia CRISP-DM.

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, ou seja, depois de ter a plataforma desenvolvida será feita uma avaliação de acordo com determinados critérios, seguindo um conjunto de métricas definidas, ou desenvolvendo-as através do dataset, por forma a verificar se esta plataforma vai ao encontro dos requisitos exigidos na proposta inicial. Esta é uma atividade que requer conhecimento em métricas relevantes e técnicas de análise.

Comunicação

A última fase consiste em divulgar os resultados que foram obtidos bem como a importância e utilidade do trabalho. Esta requer conhecimento da cultura disciplinar.

Esta fase encontra-se extremamente dependente dos resultados da fase anterior, uma vez que se encontrada uma resposta satisfatória ao problema proposto, levar-nos-ão a concluir o projeto, mas se por sua vez não for encontrada uma solução eficiente, poderá levar-nos de volta à fase inicial.

Relativamente a este projeto foi desenvolvido o relatório de dissertação, onde o processo é descrito.

Cross Industry Standard Process for Data Mining

CRISP-DM para sistemas de informação é uma metodologia que facilita os projetos de Mining Data e a condução de todo o processo de KDD.

Esta metodologia tem como finalidade melhorar o desenvolvimento da plataforma, através de um aumento de celeridade, segurança e exequibilidade e viabilidade dos projetos, assim como através da uma diminuição dos custos de execução (M.F. Santos & Azevedo, 2005).

O CRISP-DM é constituído por seis fases, organizadas de maneira cíclica, conforme apresentado na figura 10 (Chapman et al., 2000). Apesar de ser composto por fases, o fluxo não é unidirecional, sendo que o círculo externo representa o ciclo natural da metodologia CRISP-DM.

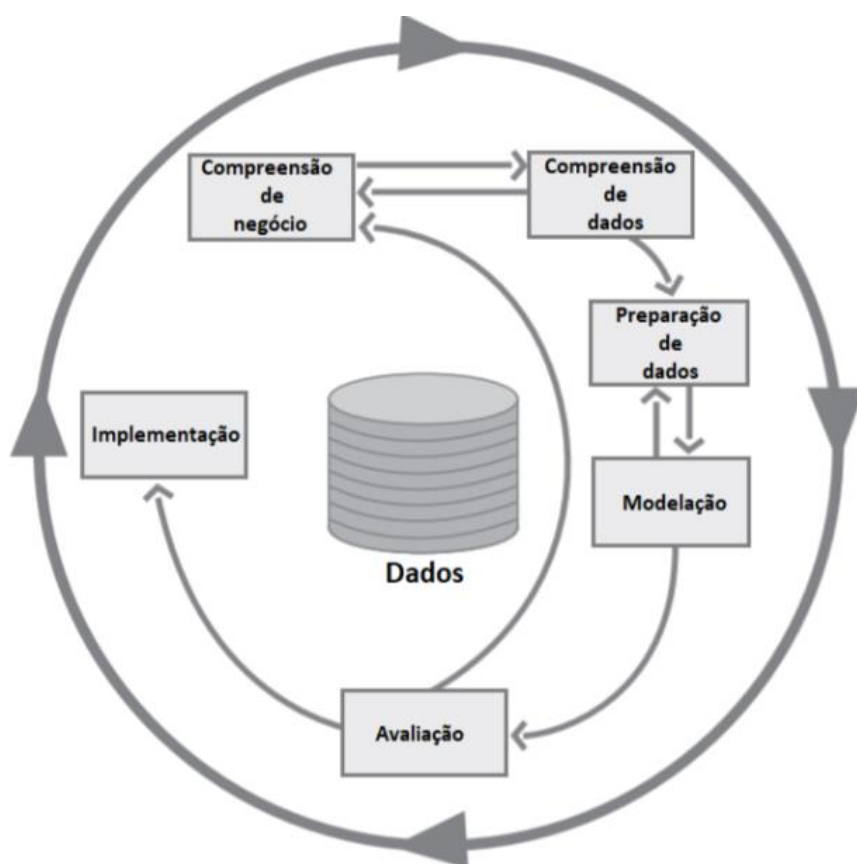


Figura 10- Modelo de processos do CRISP-DM traduzido e adaptado de (Chapman et al., 2000)

Compreensão de negócio: A primeira fase do modelo CRISP-DM, foca-se em perceber os objetivos e requisitos do negócio. Ao fim de ser estudada a informação, e com base nesse conhecimento existe a definição do problema de data mining, de forma a atingir os objetivos projetados por um plano preliminar.

Compreensão dos dados: Esta fase abrange a recolha de dados e segue-se com a exploração e verificação da qualidade dos mesmos. Esta fase tem como principal objetivo obter as informações ocultas através da percepção dos dados.

Preparação dos dados: A fase de preparação abrange todas as atividades necessárias para construir o conjunto de dados final, que será usado na ferramenta de modelação. Este conjunto de dados será sujeito a várias realizações, uma vez que neste processo de organização dos dados é efetuada uma limpeza, seleção e transformação dos mesmos, o número de vezes necessário até que estes apresentem qualidade, de forma a serem futuramente analisados.

Modelação: A fase de modelação consiste na seleção e aplicação de técnicas de data mining, bem como a configuração de parâmetros para obter a otimização de dados. Nesta fase, há que atender não só à adequação da técnica ao problema da plataforma, mas também poderá ser necessária uma submissão previa dos dados.

Avaliação: Nesta fase será avaliada a utilidade do método (ou métodos), bem como se deverá efetuar uma revisão dos passos executados, e forma a verificar se os objetivos de negócio foram atingidos.

Implementação: A fase da implementação consiste em avaliar e organizar os resultados obtidos, de maneira a possibilitar que o conhecimento extraído seja disponibilizado ao utilizador. Por fim, dependendo dos requisitos do projeto, poderá apenas ser necessário a elaboração de um relatório ou proceder-se à implementação, monitorização e manutenção do método.

Planeamento do projeto e Cronograma

Conclusão

Referências

- Babu, M. S. P., & Sastry, S. H. (2014). Big data and predictive analytics in ERP systems for automating decision making process. *2014 IEEE 5th International Conference on Software Engineering and Service Science*, 259–262. <https://doi.org/10.1109/ICSESS.2014.6933558>
- Brath, R., & Peters, M. (2004). *Dashboard Design: Why Design is Important*.
- Cortez, P. (2014). *Modern optimization with R*. Springer.
- Edelstein, H. A. (1999). *Introduction to data mining and knowledge discovery*. Two Crows Corp.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1997). *From Data Mining to Knowledge Discovery in Databases*. 18.
- Few, S. (2006). *Information Dashboard Design*. O'Reilly Media.
- Fichman, R. G., Kohli, R., & Krishnan, R. (Eds.). (2011). **Editorial Overview** —The Role of Information Systems in Healthcare: Current Research and Future Trends. *Information Systems Research*, 22(3), 419–428. <https://doi.org/10.1287/isre.1110.0382>
- Gorunescu, F. (2011). *Data Mining concepts, models and techniques*. Springer.
- Grimson, J., Grimson, W., & Hasselbring, W. (2000). The SI challenge in health care. *Communications of the ACM*, 43(6), 48–55. <https://doi.org/10.1145/336460.336474>
- Han, J., Kamber, M., & Pei, J. (2012). *DATA MINING Concepts and Techniques*.

Elsevier.

Imhoff, C., Gallemmo, N., & Geiger, J. G. (2003). *Mastering data warehouse design: Relational and dimensional techniques*. Wiley Pub.

Jun, T., Kai, C., Yu, F., & Gang, T. (2009). The Research Application of ETL Tool in Business Intelligence Project. *2009 International Forum on Information Technology and Applications*, 2, 620–623. <https://doi.org/10.1109/IFITA.2009.48>

Kimball, R., & Ross, M. (n.d.). *The Data Warehouse Toolkit*.

L. Hovis, G. (2002). *Stop Searching for Information Monitor it with Dashboard Technology*.

Lenz, R., & Kuhn, K. A. (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>

Michalewicz, Z. (2007). *Adaptive business intelligence*. Springer.

Rocha, Á., Adeli, H., Reis, L. P., & Costanzo, S. (Eds.). (2019). *New Knowledge in Information Systems and Technologies: Volume 3* (Vol. 932). Springer International Publishing. <https://doi.org/10.1007/978-3-030-16187-3>

Sarawagi, S., Agrawal, R., & Megiddo, N. (1998). Discovery-driven exploration of OLAP data cubes. In H.-J. Schek, G. Alonso, F. Saltor, & I. Ramos (Eds.), *Advances in Database Technology—EDBT'98* (Vol. 1377, pp. 168–182). Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0100984>

Singh, H., & Samalia, H. V. (2014). A Business Intelligence Perspective for Churn Management. *Procedia - Social and Behavioral Sciences*, 109, 51–56. <https://doi.org/10.1016/j.sbspro.2013.12.420>

Vassiliadis, P., & Sellis, T. (1999). A survey of logical models for OLAP databases. *ACM SIGMOD Record*, 28(4), 64–69. <https://doi.org/10.1145/344816.344869>