



Mestrado em Engenharia e Gestão Industrial

Business Intelligence no Suporte à Decisão Estratégica

Trabalho de Projeto
apresentado para a obtenção do grau de Mestre em
Engenharia e Gestão Industrial

Autor

Bruno Daniel Lopes Fernandes

Orientador

Prof. Jorge Alexandre Almeida

Professor Adjunto
Instituto Politécnico de Coimbra

Co-Orientador

Prof. Mateus Mendes

Professor Adjunto
Instituto Politécnico de Coimbra

Coimbra, março 2020



Mestrado em Engenharia e Gestão Industrial

Business Intelligence no Suporte à Decisão Estratégica

Trabalho de Projeto
apresentado para a obtenção do grau de Mestre em
Engenharia e Gestão Industrial

Autor

Bruno Daniel Lopes Fernandes

Orientador

Prof. Jorge Alexandre Almeida

Professor Adjunto
Instituto Politécnico de Coimbra

Co-Orientador

Prof. Mateus Mendes

Professor Adjunto
Instituto Politécnico de Coimbra

Coimbra, março 2020

AGRADECIMENTOS

Agradeço aos meus pais que me apoiam sempre, desde o primeiro dia em todas as minhas lutas, ao meu irmão, à minha família, amigos e aos colegas do mestrado.

À Sónia pelo seu carinho e apoio, principalmente, nas horas mais difíceis.

À Maria Leonor pelo meu tempo que era dela, e a priver da minha companhia, para a elaboração deste projeto de dissertação.

Um agradecimento especial, à empresa onde trabalho por me disponibilizar toda a informação necessária para a elaboração deste projeto, principalmente à diretora financeira.

Um agradecimento muito especial aos professores Eng.º Jorge Alexandre Almeida e Dr. Mateus Mendes, que acreditaram em mim, e me incentivaram na realização deste projeto.

A Deus, pelo dom da vida e por tudo o que tenho.

A todos que mesmo não mencionados não estão esquecidos.

Bem Hajam!

RESUMO

Em pleno século XXI, com um mundo em constante mudança, as organizações procuram constantemente prever os seus acontecimentos futuros, por forma a que tomem decisões por vezes com algum grau de complexidade, de uma forma ágil e expedita, baseada na informação que geram e dispõem. Os sistemas de apoio à decisão têm tido uma importância fundamental no meio organizacional, desde a automatização de processos das atividades diárias, até apoiar os agentes decisores a formular as melhores diretrizes ao nível estratégico, tático ou operacional. Com o novo paradigma da indústria 4.0, é levada a cabo uma revolução nestes sistemas nas organizações, que hoje se denomina de *Business Intelligence*. Neste contexto, são criados sistemas avançados de recolha de dados e informação relevante para a tomada de decisão, combinando diversas ferramentas que irão permitir aumentar a inteligência organizacional, auxiliando a estruturar conhecimento fundamental para delinear e executar as melhores resoluções para problemas internos e externos. Os sistemas de *Business Intelligence* trazem consigo tecnologias associadas, nomeadamente *Data Warehousing*, *Online Analytical Processing* e *Data Mining*, que serão abordadas no presente projeto de dissertação. Presente este novo paradigma, foi projetado para uma empresa de desenvolvimento de produtos de software uma arquitetura de um modelo de *Business Intelligence* para avaliar determinados parâmetros financeiros. Partindo dos dados do ERP da empresa, foi projetado um processo de *Data Mining* usando o software *RStudio*, caracterizado por ser *open source* e oferecendo grandes possibilidades de exploração de recursos de BI. Procedeu-se à construção de uma ferramenta *front-end* para visualização da informação usando o *RShiny*, com possibilidade de inúmeras personalizações. Vai permitir em tempo real a monitorização dos indicadores financeiros e de gestão, permitindo também fazer projeções para o futuro, que vão servir de suporte à tomada de decisão dos agentes decisores. Os avanços obtidos até ao momento possibilitam confirmar a utilidade destas ferramentas, havendo margem para grandes progressos.

Palavras Chave: *Business Intelligence*; Informação Financeira; *Data Mining*; Redes Neurais.

ABSTRACT

In the 21st century, with an ever-changing world, organizations are constantly seeking to predict their future events, in order to make decisions sometimes with some degree of complexity, in an agile and expeditious manner, based on the information they generate and provide. Decision support systems have been of fundamental importance in the organizational environment, from the automation of processes of daily activities, To support decision-makers to formulate the best guidelines at the strategic, tactical or operational level. With the new paradigm of the 4.0 industry, a revolution is carried out in these systems in organizations, which today is called Business Intelligence. In this context, advanced data collection systems and relevant information for decision-making are created, combining several tools that will allow to increase organizational intelligence, helping to structure fundamental knowledge to delineate and execute the best resolutions for internal and external problems. Business Intelligence Systems bring with it associated technologies, namely Data Warehousing, Online Analytical Processing and Data Mining, That will be addressed in this dissertation project. Present this new paradigm, was designed for a software product development company an architecture of a Business Intelligence model to assess certain financial parameters. Based on the company's ERP data, a Data Mining process was designed using the RStudio software, characterized by being open source and offering great possibilities for exploitation of BI resources. We proceeded to construct a front-end tool for viewing information using RShiny, with the possibility of numerous customizations. It will allow you to monitor financial and management indicators in real time - allowing you to make projections for the future - that will support decision-making by decision-makers. Advances so far have made it possible to confirm the usefulness of these tools, with room for great progress.

Keywords: Business Intelligence; Financial Information; Data Mining; Neuronal Network.

INDICE

CAPÍTULO 1 – INTRODUÇÃO	21
1.1 Enquadramento do Tema.....	21
1.2 Objetivos.....	21
1.3 Metodologia.....	22
1.4 Estrutura da Dissertação.....	22
CAPÍTULO 2 – ENQUADRAMENTO TEÓRICO	23
2.1. <i>Data Warehouse</i>	24
2.2 OLAP.....	26
2.3. <i>Data Mining</i>	27
2.3.1 Aplicações e Procedimentos de <i>Data Mining</i>	28
2.3.2 Métodos de <i>Data Mining</i>	29
2.3.3 Técnicas de <i>Data Mining</i>	30
2.4. Redes Neurais Artificiais.....	31
2.4.1 Redes Neurais Artificiais – Conceitos e Definições.....	32
2.4.2 Características de uma Rede Neuronal Artificial.....	33
2.4.3 Arquitetura de uma Rede Neuronal Artificial.....	34
2.4.4 Tipos de Aprendizagem de uma Rede Neuronal Artificial.....	37
2.4.5 Algoritmos Aplicados a Redes Neurais Artificiais.....	37
2.5. Evolução dos Sistemas de Apoio à Decisão.....	38
2.6. <i>Business Intelligence</i>	40
2.7. Arquitetura / Sistemas de <i>Business Intelligence</i>	41
2.8. Implementações de Sistemas de <i>Business Intelligence</i>	45
CAPÍTULO 3 – DESENVOLVIMENTO DO MODELO DE BI NUMA EMPRESA DE SOFTWARE	57
3.1. Caso de estudo numa empresa de software.....	57
3.2. Âmbito do Projeto.....	57
3.3. Descrição do Modelo Proposto.....	58
3.4. Descrição e Aplicação do Software RStudio.....	59
3.5. Arquitetura e Desenvolvimento do Modelo de BI.....	61
3.6. Arquitetura e Desenvolvimento do Modelo de BI.....	63
CAPÍTULO 4 – RESULTADOS DO MODELO DE BI	77
CONCLUSÃO	81
REFERÊNCIAS BIBLIOGRÁFICAS	83
ANEXOS	89
ANEXO 1 – ARTIGO “NEAR REAL TIME BUSINESS INTELLIGENCE FRAMEWORK USING R SHINY”	91

ÍNDICE DE FIGURAS

Figura 1 - Representação de um Datawarehouse. (Fonte: Elaboração própria)	25
Figura 2 - Visualizações parciais em cubos. (Fonte: Adaptado de Turban et al., 2009).....	27
Figura 3 - Estrutura de um neurónio artificial. (Fonte Nwufo, 2017).....	33
Figura 4 - Rede Feedforward de camada única. (Fonte: Lazzarin et al. 2018)	35
Figura 5 - Rede Feedforward multicamada. (Fonte: Lazzarin et al. 2018)	35
Figura 6 - Modelo de rede recorrente. (Fonte: Vieira e Silva, 2018)	36
Figura 7 - Evolução da terminologia analítica. (Fonte: Delen e Ram, 2018)	40
Figura 8 - Generic BI Technical Framework. (Fonte: Zeng et al., 2006)	42
Figura 9 - Arquitetura típica de BI. (Fonte: Negash, 2004)	43
Figura 10 - Estrutura típica de um sistema de BI. (Fonte: Sezões et al., 2006)	43
Figura 11 - Arquitetura Standard de Business Intelligence. (Fonte: Linden, 2019)	44
Figura 12 – Quadrante Mágico de plataformas de Business Intelligence e Analytics (Fonte: Howson et. al, 2019)	46
Figura 13 - Apresentação do software SQL Server com as tabelas da base de dados. (Fonte: Mussa et al., 2018)	47
Figura 14 - Tabela de Fatos Aluno. (Fonte: Mussa et al., 2018)	48
Figura 15 - Apresentação do ecrã da consola Pentaho. (Fonte: Mussa et al., 2018).....	48
Figura 16 - Configuração do cubo OLAP. (Fonte: Mussa et al., 2018)	49
Figura 17 - O Cubo Alunos e as suas dimensões. (Fonte: Mussa et al., 2018)	50
Figura 18 - Aplicação Cognos Report Studio. (Fonte: Sezões et al., 2006)	51
Figura 19 - Arquitetura de modelo de Business Intelligence. (Fonte: Costa e Santos, 2012)	52
Figura 20 - Dashboard de análise de vendas no Pentaho. (Fonte: Costa e Santos, 2012)..	52
Figura 21 - Os principais componentes da arquitetura de BI. Ali, et al., 2013)	54
Figura 22 - Detalhe de vendas por região (Fonte: Rao et al., 2019)	55
Figura 23 - Detalhes das vendas por cliente e por região. (Fonte: Rao et al., 2019)	56
Figura 24 - Interface RStudio (Fonte: RStudio, 2019).....	60
Figura 25 - Arquitetura do sistema de Business Intelligence desenvolvido. (Fonte: Elaboração Própria)	62
Figura 26 - Detalhes do Processo de Data Mining. (Fonte: Elaboração Própria).....	63
Figura 27 - Código em SQL para extração dos dados necessários.....	64
Figura 28 - Informação gerada pelo código SQL.....	65
Figura 29 - Código em R para normalização dos dados.....	66
Figura 30 - Apresentação dos dados normalizados	66
Figura 31 - Rede Neuronal Modelo A.....	70
Figura 32 - Rede Neuronal Modelo B.....	70
Figura 33 - Rede Neuronal Modelo C.....	70
Figura 34 - Página principal do dashboard.....	78
Figura 35 - Distribuição dos colaboradores pelo território nacional	79
Figura 36 - Faturação real e prevista	79

ÍNDICE DE QUADROS

Quadro 1 - Estrutura da UI. (Fonte: Konrath et al., 2018).....	61
Quadro 2 - Estrutura do Server. (Fonte: Konrath et al., 2018)	61
Quadro 3 - Parametrização dos Modelos de RNA.....	69
Quadro 4 - Resultados obtidos dos três modelos neurais, A, B e C, para prever a faturação da empresa. A tabela mostra o valor atual, o valor previsto e o erro entre os dois (desvio), em porcentagem, para cada modelo N	72
Quadro 5 – Precisão e desvio gerados pelos modelos	73
Quadro 6 - Resultado do MSE para os três modelos	73
Quadro 7 - Valores gerados pelo modelo C nos últimos trimestres.....	75
Quadro 8 – Previsões geradas pelo Modelo C com amostras de diferentes dimensões	76

ÍNDICE DE GRÁFICOS

Gráfico 1 - Valores previstos e reais do modelo A	74
Gráfico 2 -Valores previstos e reais do modelo B.....	74
Gráfico 3 - Valores previstos e reais do modelo C.....	74

LISTA DE ABREVIATURAS, ACRÓNIMOS E SIGLAS

BI - *Business Intelligence*

BIRT - *Business Intelligence Reporting Tool*

CASE - *Computer Aided Software Engineering*

CRM - *Customer Relationship Management*

DM - *Data Mining*

DSS - *Decision Support System*

DW - *Data Warehouse*

ERP - *Enterprise Resource Planning*

ETL - *Extract, Transform and Load*

KPI - *Key Performance Indicators*

MSE - *Mean Squared Error*

ODBC - *Open Database Connectivity*

OLAP - *Online Analytical Processing*

OLTP - *On-line Transaction Processing*

RNA - *Rede Neuronal Artificial*

SAD - *Sistemas de Apoio à Decisão*

SIE - *Sistema de Informação para Executivos*

SSAS - *SQL Server Analysis Services*

SSD - *Sistema de Suporte à Decisão*

SSIS - *SQL Server Integration Services*

SSRS - *SQL Server Reporting Services*

UI - *User Interface*

CAPÍTULO 1 – INTRODUÇÃO

1.1 Enquadramento do Tema

Hoje em dia, as empresas deparam-se com enormes quantidades de informação, interna e externamente, seja derivado dos seus processos internos, como produção, contabilidade, logística, comercial, área financeira ou do seu relacionamento externo com clientes, fornecedores e entidades bancárias. Ao produzir um conjunto muito vasto de dados derivado destas relações, as empresas sabem que possuem informações vitais para a sua continuidade no mercado e que pode trazer vantagens competitivas. Toda esta informação é suportada pelos sistemas de gestão das empresas, designado por *Enterprise Resource Planning* (ERP) que têm na sua principal tarefa, consolidar as informações da empresa numa única base de dados, agregando as funções de controlo da produção/projetos, gestão de stocks, compras, recursos humanos, faturação, tesouraria entre outros, eliminado deste modo a dificuldade em obter dados armazenados em vários sistemas. No entanto, o seu processamento através de folhas de cálculo, torna-se impraticável e limita as análises e resultados que se pretendem efetuar. Para superar esta dificuldade e extrair conhecimento sobre o que levou a factos passados, tenha o resultado sido positivo ou negativo, surgiram tecnologias e ferramentas que forma a base dos sistemas de *Business Intelligence* (BI), como os *Data Warehouse* (DW), *Online Analytical Processing* (OLAP), e *Data Mining* (DM) (Shim et al, 2002).

1.2 Objetivos

O presente projeto de Mestrado surge no âmbito da frequência do Mestrado em Engenharia e Gestão Industrial, visando a obtenção do grau de Mestre pelo Instituto Superior de Engenharia de Coimbra. A temática escolhida para este projeto de Mestrado tem como finalidade o estudo, conceção e implementação de um sistema de BI que suporte ao nível estratégico, uma organização na área de desenvolvimento de software para empresas de telecomunicações e de internet móvel, tendo como objetivo auxiliar nos procedimentos de tomada de decisão por parte dos responsáveis da empresa.

Com a elaboração deste projeto espera-se a realização de um sistema de BI que proporcione aos quadros da organização uma tomada de decisão mais eficiente e eficaz, através da disponibilização de vários mecanismos de recolha, exploração e análise de informação relevante, mais concretamente sobre a área financeira e recursos humanos.

É importante referir que o modelo proposto, é um modelo genérico podendo ser aplicado a diferentes organizações de diferentes áreas.

1.3 Metodologia

O presente projeto foi elaborado essencialmente em duas etapas: uma etapa de pesquisa e investigação e uma de desenvolvimento e implementação. A primeira etapa incidiu inicialmente na pesquisa de artigos científicos publicados em bases de dados e na consulta de obras de referência na área da temática do projeto. Foi efetuada uma revisão da literatura da área em estudo para poder clarificar e enquadrar conceitos associados ao BI e demonstrar ferramentas desenvolvidas, no que respeita às suas arquiteturas e implementações. A etapa seguinte, consistiu no desenho da arquitetura e desenvolvimento de um modelo de BI, recorrendo a diversas tecnologias, nomeadamente a uma ferramenta *opensource* a ser aplicada a uma empresa de produção de software, bem como posteriormente a sua implementação localmente.

1.4 Estrutura da Dissertação

O projeto de dissertação encontra-se dividido em três capítulos, sendo o capítulo um, inteiramente dedicado ao enquadramento teórico de conceitos relacionados com BI, com indicação das principais ferramentas usadas neste âmbito, tendo sido dado especial enfoque às redes neuronais artificiais (RNA), pelo facto, da construção do presente modelo ser estruturado e delineado para ser implementada uma análise que recorre às RNA. Na parte final deste capítulo é tratada a evolução dos sistemas de apoio à decisão, bem como, a apresentação de diversos conceitos relacionados com o termo BI, abordando diversas arquiteturas e implementações deste sistema.

No capítulo dois é caracterizada a organização onde se pretende implementar o modelo proposto, definindo o âmbito do projeto. É descrito o modelo e a sua aplicação, mencionando a sua arquitetura, com uma descrição sucinta das ferramentas a serem usadas e aplicadas. É desenvolvido uma RNA, sendo dado a conhecer os dados a serem trabalhados e analisados, a sua estrutura e parâmetros envolvidos, bem como a sua performance.

O capítulo 3 é dedicado aos resultados da aplicação desenvolvida, apresentando diversos *dashboards* com informação e análises anteriormente solicitadas pelos órgãos de gestão da empresa.

CAPÍTULO 2 – ENQUADRAMENTO TEÓRICO

O conceito de BI é um conceito lato e generalista sendo a sua principal meta recolher dados e transformá-los em informação válida, através da descoberta de tendências e padrões, e posteriormente em conhecimento proveitoso e adequado para a tomada de decisão. Este conceito “nascido” há cerca de 60 anos como sendo uma ferramenta automática para tratar informação de vários setores de qualquer empresa. Desde então a designação tem sido relacionada à propriedade dos Sistemas de Apoio à Decisão (SAD), sendo até empregue por diferentes autores para substituir designações como Sistema de Suporte à Decisão (SSD), Sistema de Informação para Executivos (SIE) (Negash, 2004) e (Zeng *et al*, 2006).

Existem numerosas definições e descrições do conceito de BI. O autor Luhn (1958) descreve-o como um sistema de processamento de dados através de computadores de modo a criar relações, verificando pontos de interesse e difundindo-os eficientemente para que seja possível levar uma determinada ação a um objetivo desejado.

Os autores Gangadharan e Swamy (2004) definem BI como o resultado de uma análise detalhada de dados do negócio, incluindo base de dados e aplicação de tecnologias, bem como práticas de análise. Ampliam a definição de BI como tecnicamente muito mais ampla como sendo uma ferramenta, que inclui a gestão de conhecimento potencialmente abrangente, planeamento de recursos empresariais para o suporte à decisão.

Segundo Zeng *et al.* (2006) define BI como um processo de recolha, tratamento e difusão de informações tendo como único objetivo, a redução de incerteza na tomada de decisões estratégicas. Afirma ainda, que é um meio usado para reunir e analisar dados e informações sobre uma empresa, a fim de ajudar a tomar decisões de negócios tornando a tomada de decisões bem mais fidedigna.

Sezões *et al.* (2006) afirmam que as linhas orientadoras de BI são, essencialmente, recolher e transformar os dados em informação através da perceção de padrões e tendências e, posteriormente, trabalhar a informação em conhecimento útil e apropriado para a tomada de decisão.

A evolução das ferramentas que possibilitam efetuar análises para a tomada de decisão nas empresas, tem vindo a ser identificado como fundamental para uma melhor qualidade e quantidade de informação disponível, para traçar um rumo estratégico com base em tomadas de decisões bem fundamentadas (Turban *et al.* 2010). Os sistemas de BI a um nível estratégico permitem definir objetivos com precisão e monitorizar a realização de tais objetivos, possibilitando a realização de relatórios comparativos sobre a rentabilidade de uma organização ou a eficácia dos canais de distribuição realizando simulações ou previsões de

resultados futuros. Num nível tático permite fornecer uma base para a tomada de decisões dentro do marketing, vendas, gestão do capital, a fim de ajudar a alcançar os seus objetivos estratégicos de uma forma mais eficaz. Ao nível operacional, os sistemas de BI são usados para realizar análises *ad-hoc* e responder a perguntas relacionadas com as operações correntes da atividade da organização, incidindo normalmente nos fornecedores e clientes (Olszak *et al*, 2007).

Num determinado período de tempo, uma empresa gera uma grande quantidade de dados relacionados com a sua produção, faturação, stocks, funcionários, materiais, produtos etc. Contudo, grande parte não sabe como utilizar esta grande quantidade de dados, desconhecendo a sua importante utilidade para obter informações úteis à gestão do negócio. Os dados estando no seu estado bruto, possuem pouca utilidade para retirar qualquer informação e conhecimento, pelo que, necessitam de ser tratados e trabalhados.

Os sistemas de BI permitem extrair informação útil com o recurso a diversas ferramentas de análise a partir da recolha e armazenamento dos dados. Segundo Santos e Ramos (2006) as tarefas habitualmente relacionadas ao BI são:

- Produzir previsões sustentadas em dados históricos dos desempenhos passados e atuais de uma organização;
- Gerar cenários que demonstrem o impacto da alteração de diversas variáveis;
- Analisar pormenorizadamente a organização, alcançando um conhecimento mais aprofundado da mesma.

Segundo Fortulan e Filho (2005) existem ferramentas adequadas no mercado comercializadas por empresas de tecnologia de informação como a Oracle, IBM, Seagate, SAP e Microsoft que parametrizam *softwares* às necessidades de cada empresa, sendo uma atividade que vem sendo tratada como SAD ou BI, termo mais usado ultimamente. Estas aplicações são encontradas em áreas muito vastas como o comércio, *marketing*, finanças, bancos e seguradoras.

Estas aplicações têm na sua principal função, melhorar a qualidade da informação. Para isso estas ferramentas incluem DW que funcionam como depósito dos dados da organização alimentados por um ou vários ERP's da organização sendo analisados com técnicas de sistemas analíticos de processamento de dados OLAP e ferramentas de DM.

2.1. Data Warehouse

Nos últimos anos, as bases de dados foram aperfeiçoadas para o tratamento de dados a nível operacional e analítico, havendo uma tendência para o nível operacional por parte dos

utilizadores, embora tivessem necessidade de ambos. Segundo Gray e Watson (1996) foram criados bancos de dados isolados, apenas para uma análise predominantemente analítica, designados de DW. Desta forma, deixa de ser necessário a recolha de dados de diversas fontes, que originava uma perda de produtividade, passando com esta ferramenta a ser possível uma consolidação e integração da informação. Para Scaff *et al.* (2005) um DW é uma base de dados integrada, variável em relação ao tempo, possuindo referências com um espaço temporal de 5 a 10 anos, o que não acontece com base dados operacionais. É não volátil, pelo que permite apenas a possibilidade de carregar informação para a base de dados e a sua consulta.

As organizações hoje em dia, possuem diversos sistemas operacionais como ERP's, que preenchem os DW que posteriormente são explorados por soluções de BI, fazendo destes sistemas uma base unificada e integrada da informação, conforme representado na figura 1.

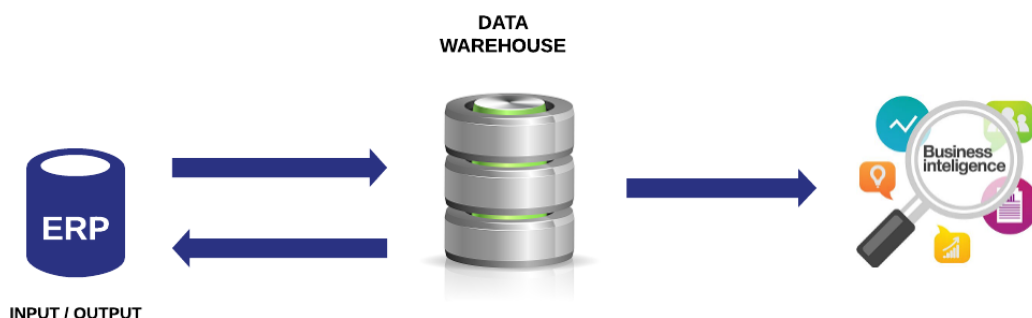


Figura 1 - Representação de um Datawarehouse. (Fonte: Elaboração própria)

Os DW são criados para armazenar *gigabytes* e até *terabytes* de informação e podem por vezes, ser divididos em unidades lógicas menores que são chamadas de *Data Marts*.

Segundo Kimball e Ross (2013), *Data Marts* são DW mas que contêm dados de uma determinada área ou departamento da organização. Considera-se sendo um subconjunto dos dados empresariais que engloba dados úteis apenas para uma determinada unidade de negócio específica, sendo mais viável para certas organizações o desenvolvimento de vários *Data Marts* orientados por assunto ou departamento para no final, integrá-los para chegar a um DW.

Importa referir, que num processo de armazenamento de dados para um DW ou *Data Marts*, a forma como os dados brutos são transportados é muito importante. Este procedimento é conhecido por *Extract, Transform and Load* (ETL). Segundo Rasmussen *et al.* (2002), uma ferramenta ETL efetua determinadas tarefas como:

- Extrai e lê dados de uma base dados, nomeadamente um ERP;

- Modifica, melhora e elimina elementos de dados distintos;
- Após a transformação e modificação são carregados na estrutura de um DW.

2.2 OLAP

No ponto anterior, foi descrita uma solução para o problema de recolha de dados dispersos por vários sistemas e posterior tratamento e armazenamento dos mesmos para auxiliar nos processos decisórios. No entanto, cálculos mais elaborados e análises mais complexas, como prever perspectivas futuras baseadas em tendências e padrões, não estão disponíveis em ferramentas como é o caso do DW. Para colmatar estas necessidades, uma outra ferramenta de BI, o OLAP, permite através de dados históricos, simular cenários para apoiar o decisor no processo de tomada de decisão.

Segundo Zeng *et al.* (2006), OLAP, refere-se às técnicas de execução análise complexa sobre as informações armazenadas num DW ou não, apto a transformar dados em bruto em informações, mais facilmente perceptíveis, analisando as variadas dimensões e combinações, reconhecendo tendências.

De acordo com Bispo e Cazarini (1998), a ferramenta OLAP apresenta-se apta a realizar análises de uma forma multidimensional, fazendo comparações de diversos ângulos, sendo caracterizada por:

- Realizar cálculos complexos;
- Fazer previsões e análises de tendências;
- Construir cenários a partir de fórmulas aplicadas pelos decisores através dos dados históricos disponíveis;
- Efetuar análises de regressão linear e não linear, probabilidades, correlação, desvio padrão e variância.

Os autores Santos e Ramos (2006), afirmam que os sistemas OLAP são usados para analisar DW, sendo que, possibilitam a criação de cubos analisando a informação gerada de vários ângulos. Após a criação de cubos, estes representam dados integrados com alguma medida de interesse, pelo que podem ter várias dimensões para analisar e representar um determinado negócio. A propensão atual é para recursos que prossigam uma visão multidimensional dos dados contabilísticos atribuindo custos ou proveitos financeiros não apenas a um segmento, mas sim a variados segmentos de modo a analisar a informação através das diferentes “faces do cubo”.

Tradicionalmente estes sistemas são usados para análises relacionadas com gamas de produtos, mas também a múltiplos segmentos, de forma a possibilitar a análise da

informação através das diversas “faces do cubo” conforme verificado na figura 2 (Jordan *et al*, 2008).

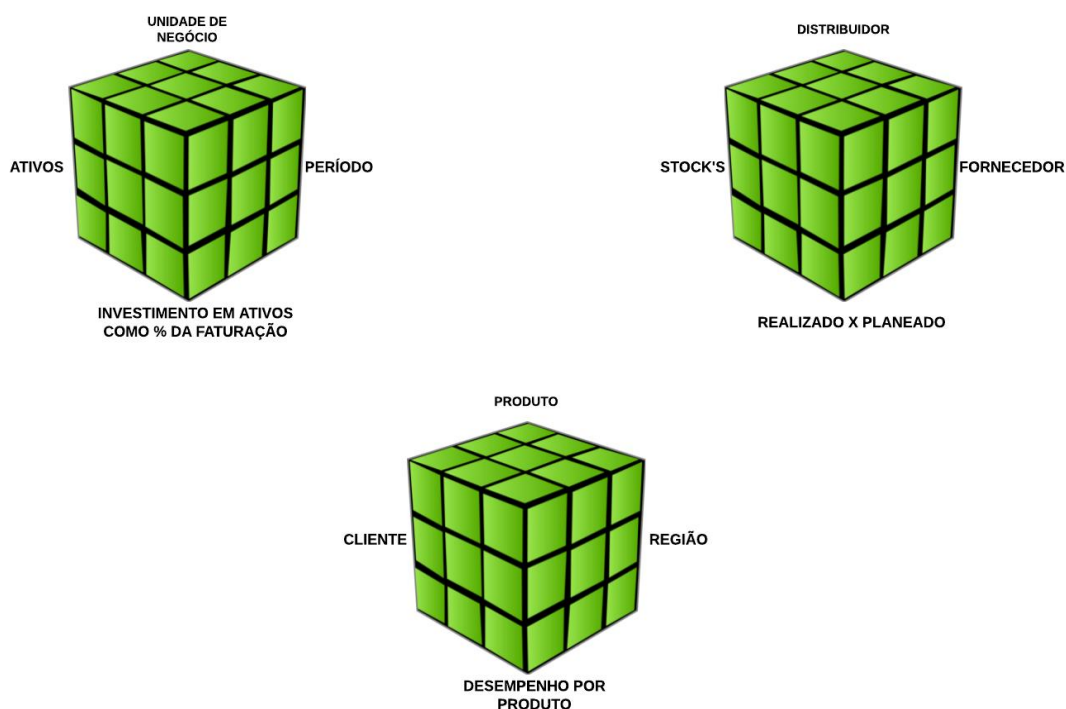


Figura 2 - Visualizações parciais em cubos. (Fonte: Adaptado de Turban *et al.*, 2009)

Utilizando o sistema OLAP o utilizador é capaz de consultar os dados através de várias dimensões, calcular os mais variados indicadores financeiros e de gestão, proceder à realização de previsões sobre a performance da organização, baseando-se em dados históricos e reais e a observação de desvios, que possam influenciar as decisões anteriormente planeadas.

2.3. Data Mining

Num mundo onde o tamanho de informação é cada vez maior e com a grande evolução das novas tecnologias, tem se assistido à acumulação e recolha de dados em massa, gerados através de alguns exemplos presentes na economia atual que só por si, produzem um conjunto de dados de enormes dimensões, nomeadamente: transações comerciais, perfis de clientes e empresas, descrição de produtos e promoções, informação contabilística e financeira entre outros. Com este volume de dados, surge a necessidade de criar ferramentas poderosas e polivalentes, para se proceder ao seu tratamento e organização, com a finalidade

de descobrir informação útil e algum conhecimento que esteja oculto nestes grandes volumes de dados, transformando-os em conhecimento organizado. A aplicação de DM tem por princípio a procura de relacionamentos e padrões que se encontram ocultos nos dados armazenados. Estes relacionamentos representam conhecimento do conjunto de informação gerada e das entidades nelas contidas.

As definições pesquisadas para o termo DM, têm diversos entendimentos, variando de acordo com o autor, área de especialização e a abordagem. De acordo com Newton e Reader (2013), DM automatiza a detecção de padrões e tendências relevantes numa base de dados, usando algoritmos para trabalhar dados atuais e históricos, que podem ser analisados para prever tendências futuras. Este método permite estender a possibilidade de descobrir padrões usando modelos mais fiáveis do que os métodos estatísticos tradicionais, e são, portanto, adequados para tornar os resultados mais compreensíveis para os agentes responsáveis pela tomada de decisão nas organizações.

Para Ali *et al.* (2013), DM permite atingir informação valiosa em dados previamente estruturados através de um processo automatizado de descoberta de padrões úteis que anteriormente eram desconhecidos. Os autores afirmam ainda que, um ambiente perfeito para um correto processo de DM, é usar um DW para servir de base ao processamento analítico, onde é possível dividir, classificar e filtrar dados para descobrir padrões. Sem a aplicação de técnicas de DM, é difícil perceber o potencial dos dados obtidos, em organizações com quantidades massivas de dados.

2.3.1 Aplicações e Procedimentos de *Data Mining*

As técnicas de DM têm sido amplamente usadas para resolver muitos problemas do mundo real. Algumas aplicações específicas e bem-sucedidas de diferentes áreas são apresentadas abaixo, de acordo com Newton e Reader (2013):

- **Marketing** – Vários profissionais de marketing são da opinião que, um dos aspetos mais importantes dos negócios é entender as necessidades individuais de seus clientes. Usando técnicas de DM, permite encontrar padrões de preferência e compra dos clientes e a sua segmentação demográfica para alcançar melhores resultados financeiros;
- **Setor Financeiro** – Neste setor, as técnicas de DM são usadas essencialmente para identificar padrões em possíveis transações fraudulentas e na análise de modelos de concessão de créditos;

- **Saúde** – Na identificação de padrões de sintomas associados a determinadas doenças, auxiliando os profissionais de saúde a obterem um diagnóstico mais fiável, contribuindo também para a otimização de recursos;
- **Produção e Fabrico** – As técnicas de DM, têm sido usadas com sucesso no setor fabril, na programação e planeamento da produção. Permite otimizar o processo de fabrico, com aplicação na previsão de manutenção e avarias de máquinas, identificando possíveis fatores relevantes para determinados acontecimentos.

No presente projeto, a etapa de DM abrange o processo de observação dos dados, com a finalidade de depreender automaticamente modelos e regras que revelam conhecimento implícito acerca dos dados observados e o processo de geração de modelos de previsão.

2.3.2 Métodos de *Data Mining*

Segundo Guohua e Francis (2000), para analisar uma grande quantidade de dados, o DM tem sido o processo mais adequado, transformando dados de baixo nível em dados de alto nível. Na sua área de atuação, DM é o campo no qual o resultado útil está sendo previsto a partir de um grande banco de dados, usando ferramentas já construídas para obter padrões e tendências que de certa forma estariam ocultas. Este processo envolve vários métodos que são constituídos por vários componentes, sendo os três principais:

- **Classificação,**
- **Associação,**
- **Clustering.**

A classificação, é considerada como uma das atividades de DM mais utilizada, consistindo gerar de forma automática um modelo que possa prever comportamentos futuros através de um processo de treino e análise do histórico de dados que se encontram em bases de dados. Os mesmos autores defendem que a previsão está na forma de prever a classe à qual os dados podem pertencer, isto é, refere-se à determinação de um objeto no qual se ajustará ou não ao perfil predefinido de um grupo. Após a classificação ter sido feita, é necessário descobrir o perfil de cada grupo em termos de atributos de um objeto, de modo a prever o grupo de um novo objeto. A maior parte das técnicas utilizadas na classificação consideram conjuntos de treino com exemplos pré-classificados, ou seja, aprendizagem supervisionada.

A associação, está presente numa base de dados referente a transações de um dado conjunto de artigos. É possível encontrar todas as associações onde a presença de um artigo numa transação, implica a presença de um outro artigo. Na associação, existem duas divisões principais de classificação: aprendizagem ou treino supervisionado e aprendizagem não supervisionada. O treino supervisionado exige que as amostras de treino sejam classificadas numa categoria conhecida. A aprendizagem não supervisionada, também conhecida como agrupamento, refere-se a metodologias projetadas para encontrar agrupamentos naturais sem o benefício de um conjunto de treino. O objetivo é descobrir novos relacionamentos dentro de um conjunto de dados. No ponto 4.4 serão abordados com mais pormenor, estes dois tipos de aprendizagem.

Clustering, refere-se à partição de determinados objetos em determinadas categorias com base nos recursos comuns intrínsecos a cada um, sendo uma forma eficiente de condensar grandes quantidades de dados brutos. O armazenamento em cluster tem a capacidade de poder obter informações importantes ou conhecimento essencial desses dados. Tomando como exemplo, os clientes de um determinado conjunto de produtos, podem ser agrupados em várias categorias, de acordo com suas particularidades pessoais, os produtos que adquiriram, a quantidade de dinheiro que gastaram e outras informações idênticas. Verificando as diferentes características entre os grupos, as empresas podem descobrir as particularidades de compra de cada grupo e, assim, alcançar o objetivo do marketing pretendido, com o intuito de poderem aumentar os seus lucros.

2.3.3 Técnicas de *Data Mining*

Até o momento, muitas técnicas têm sido usadas para realizar processos convencionais de DM, como associação, agrupamento, classificação, modelagem e padrões sequenciais. Entende-se como técnica, a integração de processos sustentados em conhecimentos científicos, de experimentação ou de cálculo, com o intuito de atingir um resultado. Estas técnicas podem tomar a forma de conjuntos estatísticos a conjuntos aproximados e até redes neuronais artificiais (RNA). Pode dizer-se que DM é uma combinação elegante destes conjuntos, RNA, e técnicas de *machine learning*, que abrangem quase todos os algoritmos de inteligência artificial (Guohua e Francis, 2000).

Ao delinear a estrutura de um modelo, é necessário definir as particularidades do sistema, procedendo-se à recolha dos dados essenciais para a sua construção, sendo que é necessário a divisão dos dados para gerar o modelo, chamado conjunto de treino, e para validar, chamado de conjunto de teste. Para o reconhecimento de padrões, relacionamentos e previsões são aplicados algoritmos ao modelo.

Segundo Newton e Reader (2013) a construção de modelos de DM podem abranger inúmeras técnicas. No entanto, é possível identificar as técnicas mais usadas, ou seja, árvores de decisão, aproximação de vizinhanças e RNA. Pela estrutura deste projeto, será dada uma maior ênfase às RNA.

- **Árvores de Decisão** – A árvore de decisão é uma estrutura semelhante a uma árvore na qual cada caminho, desde a raiz até uma folha corresponde a uma regra de classificação. Representam um conjunto de regras que seguem uma classificação de classes ou valores, utilizando uma lógica simples. A sua estrutura consiste na interligação de um conjunto de nós através de ramos.
- **Aproximação de Vizinhanças** – Permite classificar um conjunto de dados registados, baseados em dados semelhantes a partir de um histórico. Exemplificando, esta abordagem permite a partir de um determinado documento, solicitar ao sistema para procurar outros semelhantes. Esta técnica permite identificar classes, que são caracterizadas por terem informação idêntica para os registos que representam. Numa amostra onde o número de registos cresce consideravelmente, a complexidade na utilização desta técnica aumenta.
- **Redes Neurais Artificiais** – As RNA, são modelos preditivos não lineares que aprendem através do seu treino. Os autores afirmam que esta técnica é bastante usada para reconhecer fraudes nos sistemas bancários, sendo que por vezes são difíceis de implementar, exigindo um planeamento bem estruturado na forma como os dados são selecionados, de modo a que todas as entradas e saídas da rede sejam expressas numericamente.

2.4. Redes Neurais Artificiais

Na procura de tecnologias que possam aperfeiçoar procedimentos e que resolvam problemas com elevado nível de complexidade, tem sido cada vez mais notório, a procura de sistemas de inteligência artificial, que têm evoluído ao longo dos anos no que toca aos seus conceitos e técnicas. Em *machine learning* as RNA são modelos estatísticos que aprendem, inspirados no modo de funcionamento das redes neuronais biológicas como o sistema nervoso central, neste caso o cérebro, sendo usadas para prever ou aproximar funções que podem apresentar um elevado número de entradas na rede.

2.4.1 Redes Neurais Artificiais – Conceitos e Definições

O conceito de RNA é bastante amplo, tendo ao longo dos anos conceitos mais abrangentes e novas aplicações nos mais variados setores de atividades. Procedendo-se a uma revisão da literatura, diversos autores abordam a temática de diferentes perspectivas.

Segundo Abad e Chaparro (2016) as RNA são modelos de regressão que são ajustados para fazer aproximações. São caracterizadas por um conjunto de variáveis de entrada e de saída, onde cada conexão tem um peso. Obtendo informação oculta de dados complexos ou vagos, mostrar padrões ou tendências revelou-se a grande vantagem no seu uso.

Luckow *et al.* (2017), afirmam que as RNA são modeladas usando múltiplas camadas de neurônios. Na primeira camada, cada neurônio absorve uma entrada. Os neurônios da primeira camada geram saídas que alimentam a segunda camada, e assim sucessivamente, até à última camada em que cada neurônio corresponde a uma variável de saída. Um modelo que esteja estruturado com múltiplas camadas de neurônios, permite que a rede aprenda funções não lineares, que se adaptam e são utilizadas para dados não estruturados.

Os autores Nwufo *et al.* (2017), consideram que as RNA são uma poderosa ferramenta de modelagem que tem a capacidade de identificar relações complexas entre uma entrada e uma saída de dados. Têm a capacidade de reaprender e adaptar-se para aperfeiçoar o seu desempenho com a disponibilidade de dados atualizados. Uma RNA bem treinada pode servir como uma ferramenta de previsão para resolver problemas com alguma especificidade que envolva o processamento de informação idêntica à de um sistema neuronal biológico.

Segundo Vieira e Silva (2018) uma RNA é estruturada por diversos neurônios, formando uma rede inteligente quando interligados e que adquire conhecimento através da experiência resultante do seu treino. As redes geradas pela combinação de diversos neurônios artificiais, ao que normalmente se chama de rede neuronal, são nada mais, do que sistemas lógicos e matemáticos que se assemelham aos neurônios biológicos e procuram simular o comportamento destes. As RNA são utilizadas em setores de atividade que geram grande volume de dados, nomeadamente a área financeira, permitindo previsões com alto valor acrescentado, que possibilitam tomar decisões e ações mais fiáveis em tempo real sem a interferência humana.

2.4.2 Características de uma Rede Neuronal Artificial

Uma RNA é composta por neurónios, inspirada em redes neuronais biológicas. Os neurónios estão conectados para trocar informação entre si. As conexões, que representam as sinapses, têm pesos numéricos que podem ser organizados de acordo com o seu treino, o que leva a que as redes se adaptem e sejam capazes de aprender (Celik e Basarir, 2017).

De acordo com Nwufo *et al.* (2017), um neurónio artificial consiste principalmente na agregação de todos os pesos interligados com função de ativação. Cada neurónio recebe entradas x_1, x_2, \dots, x_n , cada entrada ponderada por um peso w_i mostrando a força da conexão para uma entrada específica como mostrado na figura 3. Cada entrada é então multiplicada pelo peso correspondente da conexão do neurónio. Em algumas situações pode ainda existir uma conexão extra, denominada de bias. Bias é um tipo de peso da conexão com um valor constante diferente de zero adicionado à soma dos produtos de entradas e pesos correspondentes, conforme equação 1.

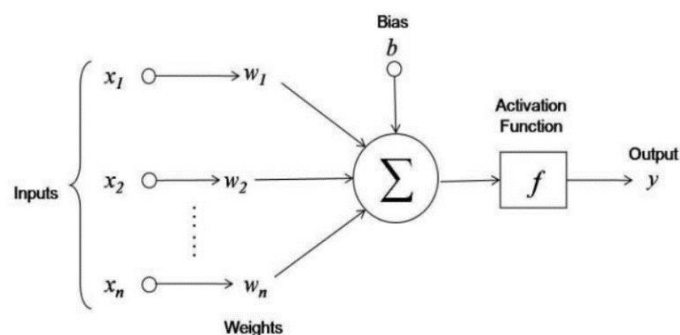


Figura 3 - Estrutura de um neurónio artificial. (Fonte Nwufo, 2017)

$$Y = \sum (weights \times input) + bias \quad (1)$$

Segundo Vieira e Silva (2018), os neurónios são formados por componentes que desempenham papéis essenciais durante o processamento da informação, sendo os seus elementos constituídos por:

- entradas (x_1, x_2, \dots, x_n);
- um conjunto de conexões, cada uma caracterizada por um peso (w_1, w_2, \dots, w_n);
- a função adição (Σ), pesando todas as entradas numa combinação linear;
- a função de ativação (f) que pode condicionar o sinal de saída, introduzindo uma componente de não linearidade no processo computacional;

- uma conexão extra (bias), sendo considerada uma espécie de neurónio na rede;
- uma saída relativa ao resultado gerado (y).

2.4.3 Arquitetura de uma Rede Neuronal Artificial

As RNA são estruturadas em camadas, com nós relacionados, em que cada neurónio tem uma função de ativação. A arquitetura é desenhada como uma rede, por meio de uma camada de entrada que se interliga como uma ou mais camadas ocultas, sendo este método executado através de conexões ponderadas. Os diferentes tipos de arquiteturas são voltados para cenários específicos e com utilizações bastante distintas (Celik e Basarir, 2017).

Segundo Agrawal e Agrawal (2015) uma RNA funciona como um sistema adaptativo, que muda a sua estrutura na fase de aprendizagem, em que, com arquiteturas simples ou complexas, podem ser facilmente modeladas. Uma RNA pode ser projetada para uma aplicação específica, como classificação de informação e categorização de padrões por meio de um processo de aprendizagem. Existem vários tipos de arquiteturas ou topologias de RNA, organizando-se em três categorias principais:

- Redes *Feedforward* de uma só camada;
- Redes *Feedforward* de multicamada;
- Redes Recorrentes.

Os autores Lazzarin *et al.* (2018), consideram que uma RNA é dita *feedforward* de uma só camada, se a corrente de informação seguir sempre em uma única direção, tendo origem na camada de entrada e seguindo para a camada de saída (figura 4). A camada de entrada não é contabilizada como camada numa RNA devido ao facto de nesta não serem efetuados cálculos. Normalmente é aplicada na resolução de problemas linearmente separáveis.

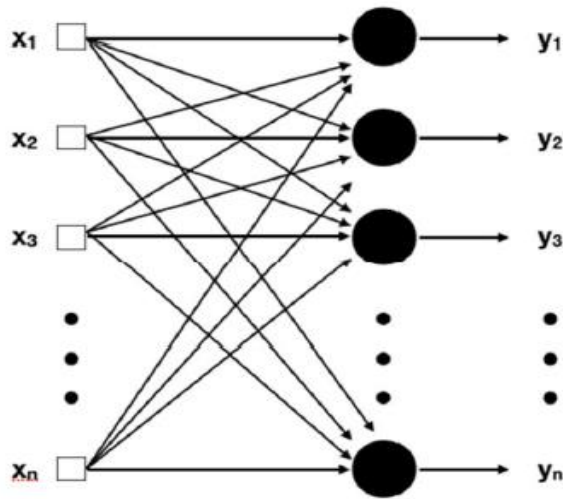


Figura 4 - Rede Feedforward de camada única. (Fonte: Lazzarin et al. 2018)

As RNA *feedforward* de multicamada, possuem um maior poder de aproximação devido às camadas adicionais conferidas à sua arquitetura. Aumentando o número de camadas intermédias, a capacidade de a rede trabalhar funções de maior complexidade aumenta, originando um acréscimo do tempo necessário para a aprendizagem (figura 5).

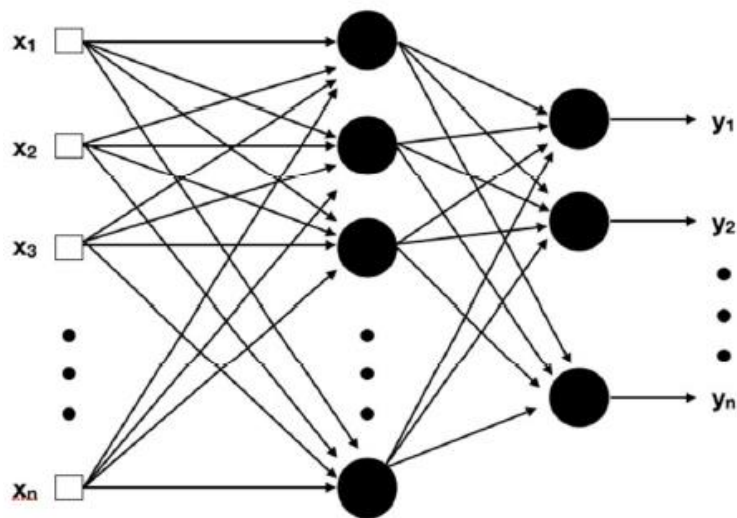


Figura 5 - Rede Feedforward multicamada. (Fonte: Lazzarin et al. 2018)

Os autores Hamdi *et al* (2016) referem que uma rede *feedforward* de uma só camada oferece vantagens a nível computacional, por se tornar uma arquitetura mais simples e menos morosa no processamento dos dados, enquanto que uma rede *feedforward* multicamada é a mais popular, sendo usada para resolver com maior precisão, problemas mais complexos. É necessário um conhecimento especializado e também variadas aplicações para determinar a arquitetura ótima de uma RNA multicamada, pelo que, segundo os autores, não existem

técnicas científicas para definir um projeto de uma RNA ideal, portanto, vários parâmetros devem ser ajustados para selecionar a melhor tipologia de rede, sendo os dois seguintes, os mais importantes:

- Número de camadas ocultas;

O número ideal de camadas ocultas usada numa rede com retropropagação tradicional, é geralmente uma ou duas camadas;

- Número de neurónios ocultos;

Não existem padrões universais para definir o número de neurónios ocultos. O melhor procedimento é usar a menor quantidade de unidades que permitam obter resultados significativos, pois muitos nós poderiam induzir um problema de *overfitting* e muito poucos poderiam causar um problema de *underfitting*. O número deve estar entre o tamanho da camada de entrada e o tamanho da camada de saída, sendo que a melhor forma de determinar o ideal é experimentalmente.

Os autores Vieira e Silva (2018), consideram as redes recorrentes diferente das redes *feedforward*, porque implementam uma camada de estado que tem papel relevante na camada de entrada, podendo ser aplicada em dados que variam com o tempo. Este processo existe em arquiteturas dinâmicas quando uma saída de um elemento influencia de algum modo a entrada para esse mesmo elemento, originando diversos circuitos fechados (figura 6).

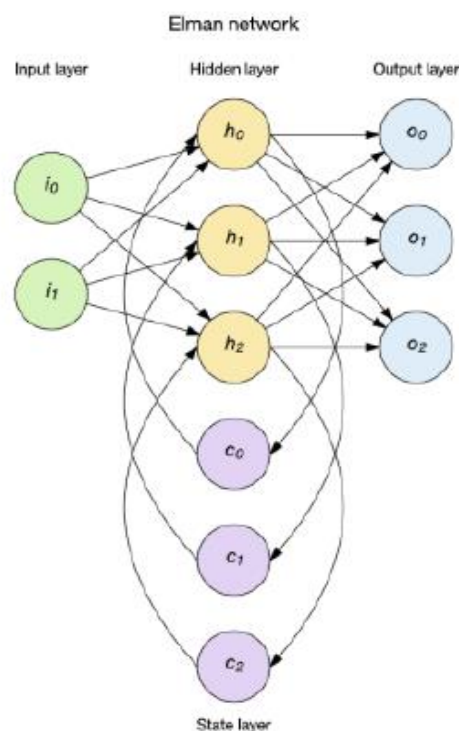


Figura 6 - Modelo de rede recorrente. (Fonte: Vieira e Silva, 2018)

2.4.4 Tipos de Aprendizagem de uma Rede Neuronal Artificial

De acordo com Agrawal e Agrawal (2015), a grande vantagem de uma RNA é a sua capacidade de aprender a partir de um algoritmo de aprendizagem. O processo de aprendizagem tem um comportamento de ajuste no qual uma rede neuronal se adapta a um impulso, produzindo de seguida a saída desejada. Segundo os autores Goodfellow e Courville (2016), os dois paradigmas fundamentais mais importantes de aprendizagem são:

- Supervisionada;
- Não Supervisionada.

O termo aprendizagem supervisionada tem origem numa resposta, dada pela configuração que é dada à rede, que é comparada com a resposta correta, sendo provida de um instrutor ou professor que mostra ao sistema o que fazer. No decorrer deste processo, efetuada a comparação entre o valor pretendido com o valor de saída, irá originar um erro. Este erro é usado para ajustar os pesos das conexões para que o erro seja diminuído.

A aprendizagem não supervisionada trata este paradigma de outra forma, ou seja, não é dado à rede uma indicação externa acerca da resposta correta. A aprendizagem é feita nos dados de entrada, reconhecendo as suas características, ajustando-se aos padrões de exemplos de treino. Segundo Falez *et al.* (2019), a aprendizagem de representação não supervisionada é reconhecida como um dos maiores desafios em *machine learning* estando a receber um crescente interesse em aplicações computacionais.

2.4.5 Algoritmos Aplicados a Redes Neurais Artificiais

Existem diversos algoritmos de aprendizagem supervisionada e não supervisionada para treino de uma RNA indicados para reconhecimentos de padrões. De acordo com Vieira e Silva (2018), referindo-se à aprendizagem supervisionada, afirmam que no passado, foi demonstrado que uma rede *feedforward* com duas ou mais camadas poderia corrigir muitas das limitações encontradas nas redes do tipo perceção. A rede perceção é uma rede *feedforward* com apenas uma camada de nodos com várias entradas e saídas. As redes perceção, por serem mais simples, são também mais rápidas e fáceis de treinar. As redes multicamada, sendo mais complexas, exigem mais dados para treino e um método mais sistemático. O método mais comum é o algoritmo de retropropagação (*backpropagation*).

Segundo Bisi *et al.* (2015), o algoritmo *backpropagation* destaca-se por ser adequado a redes multicamada, com pesos sinápticos sujeitos a ajustamentos e bias. Possui um treino

supervisionado com um erro que vai sendo reduzido para obter os pesos de todas as camadas. Na fase de treino do algoritmo o erro é calculado para cada saída, ou seja, é a diferença entre a saída desejada e a saída obtida. De seguida é calculado o erro propagado em sentido contrário, subindo os pesos se o erro é negativo e descendo se o erro é positivo.

O procedimento continua até o erro ser menor do que o prefixado anteriormente, minimizando o erro quadrático da rede.

Citando Huang (2018), algoritmos padrão de *machine learning*, exigem um imenso número de exemplos de treino para descobrir informação oculta. No treino de redes neuronais mais complexas o treino não supervisionado pode contribuir para aperfeiçoar o desempenho final, porque fornece uma boa região inicial no espaço de parâmetros a partir do qual o ajuste se inicia. No entanto, a aprendizagem não supervisionada numa rede neuronal profunda é normalmente muito complicada.

Os pontos seguintes são compostos pela análise bibliográfica que orienta a temática da investigação, no sentido de poder analisar, clarificar e compreender, a evolução histórica dos vários sistemas de apoio à decisão, as várias ferramentas disponíveis para os implementar e o processo decisório nas organizações. É abordado os desenvolvimentos destes sistemas começando com a construção dos mesmos orientados para modelos no final da década de 60, desenvolvimentos teóricos e testes de sistemas no início da década de 70 e a implementação de sistemas de planeamento financeiro e de apoio à decisão no início e meados da década de 80 e 90. Nas últimas três décadas, os sistemas de apoio à decisão adotaram uma definição mais estreita ou mais ampla, surgindo outros sistemas para auxiliar tipos específicos de problemas, bem como, diferentes tipos de responsáveis pela tomada de decisão.

2.5. Evolução dos Sistemas de Apoio à Decisão

Segundo Power (2007), os *Decision Support Systems*, comumente designados por DSS, começaram a tomar forma no final da década de 60 com pesquisadores a estudar sistematicamente o uso de modelos quantitativos computadorizados para auxiliar no planeamento e na tomada de decisões. O primeiro desenvolvimento teórico teve início com um estudo de Scott Morton em 1966, que envolveu a construção, implementação e teste de um sistema de decisão interativo aumentando a qualidade das decisões. Estudou como os computadores e modelos analíticos poderiam ajudar os agentes a tomar uma decisão no planeamento dos seus negócios. Na prática, os decisores das áreas do marketing e produção usaram um *Management Decision System* para coordenar o planeamento da produção de equipamentos de lavandaria. O mesmo autor, identificou quatro critérios para projetar modelos

e sistemas para apoiar a tomada de decisões. Estes critérios incluíam a robustez, facilidade e controlo, simplicidade e integridade de toda a informação, sendo que, todos eles permanecem relevantes na avaliação dos modernos sistemas de apoio à decisão. Power (2007) refere ainda, que o primeiro uso do termo DSS foi no artigo publicado por Gorry e Scott-Morton em 1971, na *Sloan Management Review*, em que os autores argumentaram que os *Management Information Systems*, se concentravam principalmente em decisões estruturadas e sugeriram que os sistemas de informação de suporte às decisões semiestruturadas e não estruturadas deveriam ser denominadas de DSS. Em 1975, Little expandiu as fronteiras dos modelos de modelação suportados por computador. O seu próprio DSS intitulado de *Braindaid*, foi projetado para apoiar decisões sobre produtos, promoções, preços e publicidade. Little ajudou também a desenvolver a linguagem de modelação financeira e *marketing* conhecida como *Express*. É importante referir, que no final da década de 70, várias empresas e investigadores haviam desenvolvido sistemas de informação que usavam os dados históricos e modelos para apoiar a gestão de topo das organizações a analisar problemas semiestruturados, no entanto, foi reconhecido que estes sistemas poderiam ser projetados para apoiar os tomadores de decisão em qualquer nível. Segundo Scaff *et al.* (2005), os primeiros sistemas de bases de dados começaram a surgir no início dos anos 80, sendo que a sua modelação passou a ser baseada na estrutura dos processos e não na estrutura do negócio. Nesta década, o processamento da informação, dominava a forma como as organizações viam o uso de computadores para a sua gestão. De acordo com Carlsson e Walden (2019), o foco era exclusivamente na rentabilidade e produtividade, sendo os sistemas executados principalmente em computadores *mainframe* (a IBM era líder de mercado), sendo as linguagens de programação dominantes o Cobol e Fortran. Os sistemas de informação eram grandes, complexos e rígidos, sendo os custos de investimentos bastante altos, pelo que, em muitos casos as reações foram francamente hostis por parte das organizações. O paradigma, no entanto, alterou-se com os desenvolvimentos de DSS a concentrarem-se na qualidade de um sistema a partir do valor e das prioridades dos usuários em vez do nível de tecnologia avançada aplicada. Na década de 90, ferramentas CASE (*Computer Aided Software Engineering*) surgiram com a finalidade de auxiliar as atividades de engenharia de software, na análise de requisitos, modelagem e testes, possibilitando um avanço na navegação e manipulação dos dados, no entanto, ainda débil em análises mais sofisticadas. Os sistemas de apoio à decisão no final da década de 90, observaram um avanço em poder de armazenamento, consultas e análises, justificado pela qualidade e evolução dos hardwares e softwares, sendo que as principais ferramentas de apoio à decisão até à atualidade são os DW, OLAP e DM (Scaff *et al.* 2005).

Os DSS tiveram que lidar na década de 2000 com o rápido crescimento do *Big Data*, assumindo de que estes processos serão mais complexos e difíceis na década de 2020 do que na década de 1970 (Figura 7). Para a década de 2020, um conjunto de pesquisas

industriais propôs que os DSS se suportem em inteligência partilhada de sistemas e humanos, de modo a que se utilize algoritmos rápidos e automáticos para grandes conjuntos de dados estruturados combinando com conhecimento adquirido por especialistas em contextos experientes. Desta forma, para se atingir este objetivo, os utilizadores precisam de aconselhamento relevante em contexto prático, (em tempo real, com dados e informações reais) que seja adaptado às suas capacidades cognitivas e conhecimento prévio, ou seja, conceitos que possam compreender e usar (Eswaran, 2017).

O DSS do século XXI serão possivelmente sistemas de treino digital que guiarão os utilizadores na economia digital, mais concretamente sobre os *smartphones*, *tablets*, *laptops*, serviços em nuvem e novos dispositivos de suporte digital que aparecerão como parte deste mesmo ambiente (Carlsson e Walden, 2019).

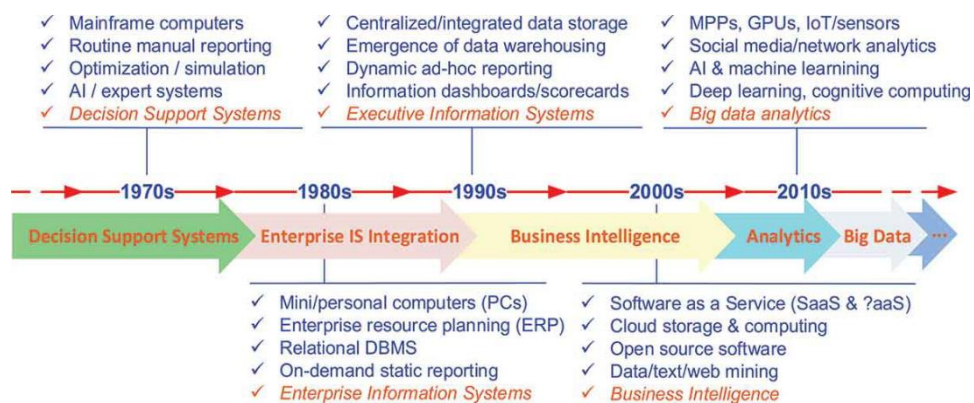


Figura 7 - Evolução da terminologia analítica. (Fonte: Delen e Ram, 2018)

2.6. Business Intelligence

Os sistemas de DSS e BI são reconhecidos como conceitos relacionados e são frequentemente associados. O autor Keenan (2019), refere que existem diferentes visões quanto ao relacionamento entre o DSS e o BI. Alguns autores caracterizam BI como uma forma de DSS orientada por dados, enquanto que o BI é um subcampo do DSS. Outros sugerem que os sistemas de BI evoluíram além do DSS ou até mesmo substituíram o DSS, ou simplesmente ignoram qualquer conexão com os sistemas de DSS.

Desde a Revolução Industrial, mais concretamente, sempre houve especialistas para analisar o desempenho e a otimização das atividades de comércio e indústria, servindo-se de métodos estatísticos e posteriormente DM, para suportar o processo de análise. Nas décadas de 80 e 90, presenciou-se a uma revolução enorme na informatização nas organizações,

sendo que as fontes de dados foram-se multiplicando, não só nos processos administrativos como nas cadeias e processos de produção. Sistemas e ferramentas específicas, surgiram nos anos 90 para dar um apoio eficaz e eficiente ao decisor, sobretudo plataformas de BI. Uma estrutura de BI, do ponto de vista do utilizador, aparece como uma ferramenta, composta por dashboards interativos, *reports*, alertas, ferramentas de consulta OLAP etc. A importância de uma plataforma de BI, é disponibilizar aos utilizadores, informação de toda a organização consultando apenas uma única fonte, evitando a consulta de múltiplas fontes, repletas de dados desnecessários e com problemas de qualidade (Linden, 2019).

De acordo com Al-Maaitah (2018), as organizações nas últimas décadas têm-se interessado pelo conceito de BI, porque representam uma mais valia na tomada de decisões oportunas perante a concorrência e a globalização. Adotaram também novos sistemas para identificar as necessidades dos clientes, bem como a análise de dados e extração de relatórios necessários à monitorização e desempenho do trabalho.

2.7. Arquitetura / Sistemas de *Business Intelligence*

O objetivo deste ponto, é pretender demonstrar algumas arquiteturas propostas por diversos autores para diferentes atividades, com objetivos distintos e com recurso a diversas ferramentas. Nos dias de hoje, os sistemas de BI têm sido vistos como um conjunto de ferramentas poderosas, de modo a criar valor para uma empresa. De acordo com Zeng *et al.* (2006), cada vez mais setores de atividade, implantaram soluções de BI avançadas para aumentar a sua competitividade, uma vez que, é importante para o funcionamento eficaz e eficiente da organização na gestão dos seus recursos e no desenvolvimento dos seus negócios. Um sistema de BI nunca é implementado por uma única tecnologia, produto ou fornecedor. Segundo Friedman (2002), a própria natureza destes sistemas, incentiva a que os gestores da informação passem a ter mais acesso e controlo sobre os dados, sendo que uma aplicação bem-sucedida numa empresa, deve considerar dois aspetos importantes. O primeiro deve considerar que os dados são corretos, válidos e íntegros, o outro, é o meio que transformará os dados em informação de decisão, contudo, nem os dados pretendidos, nem os meios eficazes, são facilmente adquiridos. A arquitetura/estrutura técnica visa deste modo resolver estes dois aspetos, conforme exemplificado na figura 8.

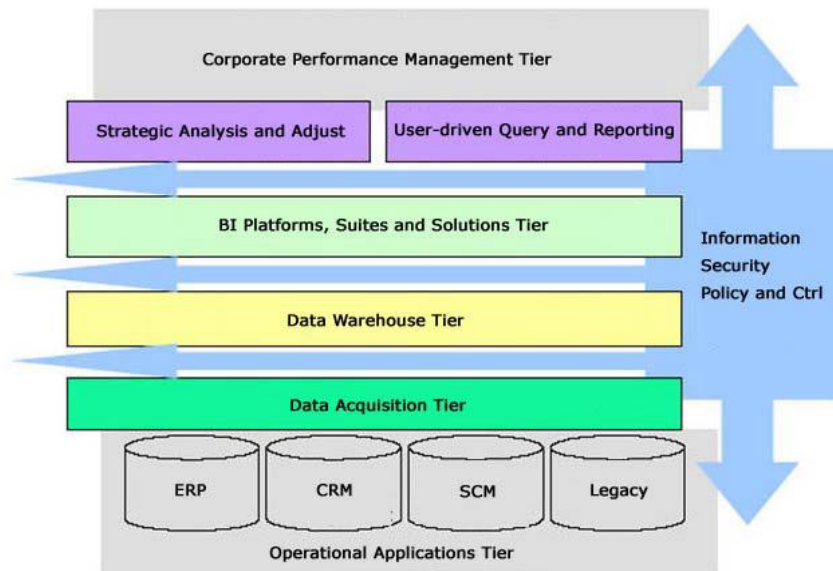


Figura 8 - Generic BI Technical Framework. (Fonte: Zeng et al., 2006)

A arquitetura de um sistema de BI, consiste em aplicativos operacionais, designadamente um ERP, aquisição de dados, consistindo na extração, transformação e carregamento e um DW que pode incluir *Data Marts*. Na área de BI, DW, OLAP e DM, são as três tecnologias mais significativas. Segundo Codd *et al.* (1998), DW pode ser entendido como um vasto repositório de dados históricos pertencentes à organização, sendo que a ferramenta OLAP refere-se às técnicas de execução e análise complexa sobre as informações armazenadas num DW. Os autores Han *et al.* (2000) consideram o DM como um processo de identificação e interpretação de padrões em dados para resolver um problema específico. Esta estratégia inclui uma classificação, estimativa, séries temporais, análise e previsão. De acordo com Negash (2004), numa arquitetura típica de BI, os dados são extraídos de sistemas operacionais e distribuídos, sendo os dados especificamente necessários para a análise de BI, carregados para um *Data Mart* usado por analistas. Posteriormente, a informação retirada do *Data Mart* será enviada para consulta dos analistas que interagem com a ferramenta OLAP. A informação final pode assumir várias formas, incluindo relatórios esporádicos, de rotina e repostas a solicitações específicas dos utilizadores. A informação é sempre enviada para os utilizadores quando se verifica que os parâmetros estão fora dos limites pré-estabelecidos. Na figura 9 é exemplificada a arquitetura defendida pelo autor.

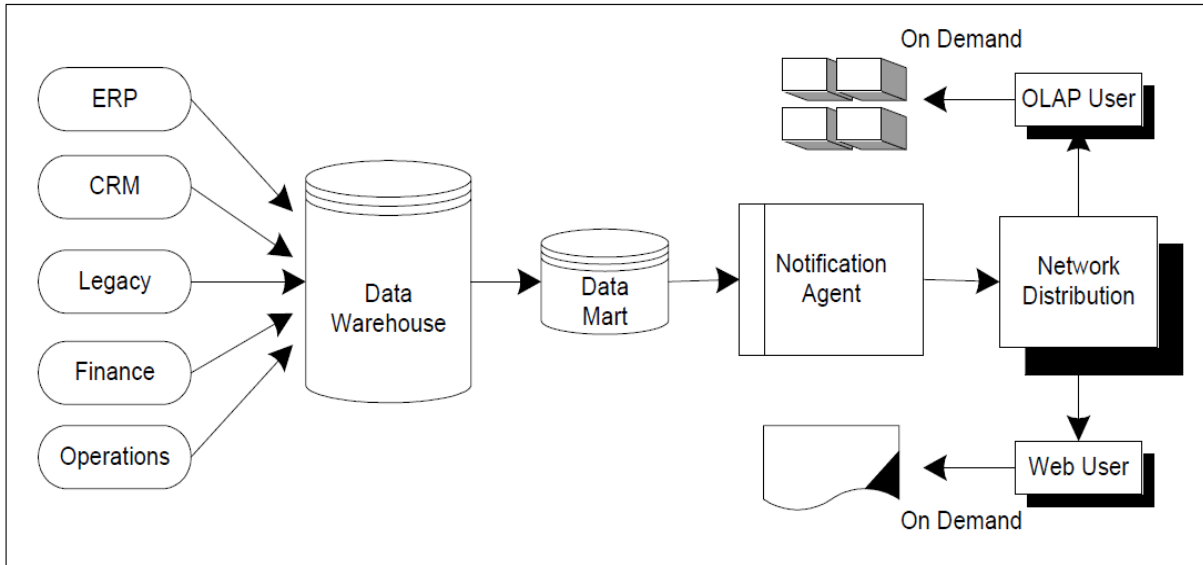


Figura 9 - Arquitetura típica de BI. (Fonte: Negash, 2004)

Um sistema de BI não funciona por si só, está inteiramente ligado às fontes de dados implícitas, sejam elas ficheiros de suporte ou sistemas internos ou externos, ou seja, tudo o que possa ser considerado informação relevante derivado do normal funcionamento de uma organização. Posteriormente é necessário transmitir a informação processada através de ferramentas de visualização, partindo do que foi processado e analisado.

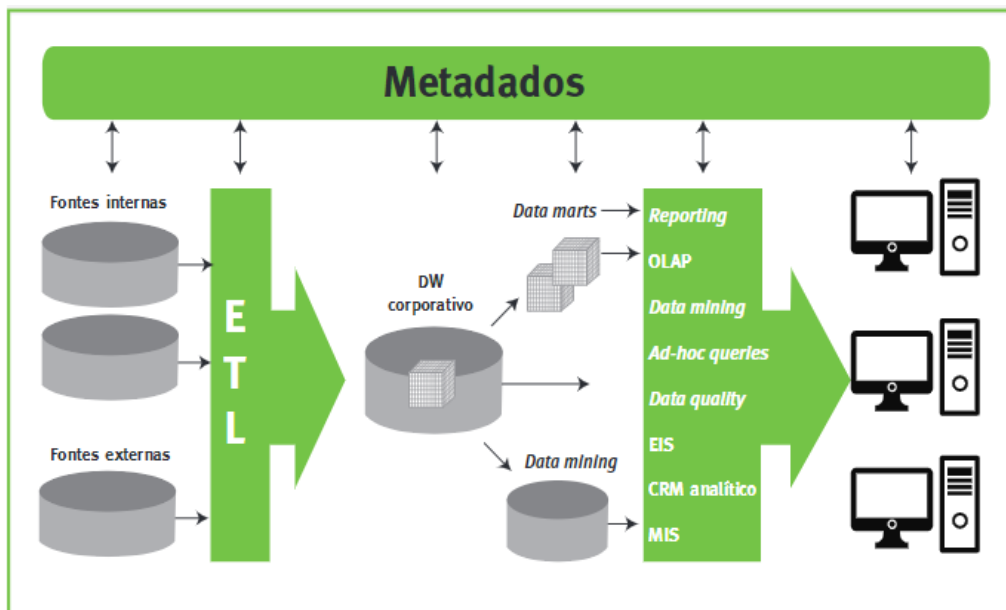


Figura 10 - Estrutura típica de um sistema de BI. (Fonte: Sezões et al., 2006)

Segundo Sezões et al. (2006), e de acordo com a figura 10, uma estrutura típica de um sistema de BI, é composto pelas fontes externas e internas de informação como ERP e

Customer Relationship Management (CRM); ferramentas ETL que se dedicam a carregar a informação vinda dos sistemas ERP ou ficheiros Excel; os DW ou Data Marts onde é agregada toda a informação dos sistemas operacionais de uma organização, funcionando como um repositório de dados históricos já filtrados. A estrutura termina com um *front-end*, visível aos utilizadores, sob a forma de relatórios, análises OLAP e técnicas de DM funcionando como cenários e projeções futuras.

O valor agregado de uma plataforma de BI, não reside tanto nas ferramentas disponibilizadas aos utilizadores, mas sim nas informações a que dá acesso. O objetivo é agregar todos os dados, evitando a consulta de múltiplas fontes. Para obter uma estrutura de referência, o autor Linden (2019) exemplifica, conforme ilustrado na figura 11. À direita encontram-se os utilizadores, que acedem às ferramentas de acordo com a sua função, nos diferentes níveis da organização. Do lado esquerdo, as mais variadas fontes de dados internas e externas. Como aspetos essenciais de tal implementação, o autor faz referência ao fluxo de dados, que parte de um sistema operacional para um sistema analítico. Nesta divisão, está implícito que, independentemente da carga de trabalho de análise, o sistema operacional não deve ser interrompido. No final do processo, todos os aplicativos oferecidos aos utilizadores, são consultados apenas numa única fonte: o DW. Apesar de serem considerados como bancos de dados tradicionais, os DW possuem um conjunto de especificidades que por vezes podem afetar a sua performance:

- Não suportam operações de atualização e exclusão, apenas carregamento de dados;
- As consultas enviam um amplo conjunto de dados apenas em linhas e colunas;
- Para grandes consultas, o tempo de resposta é maior.

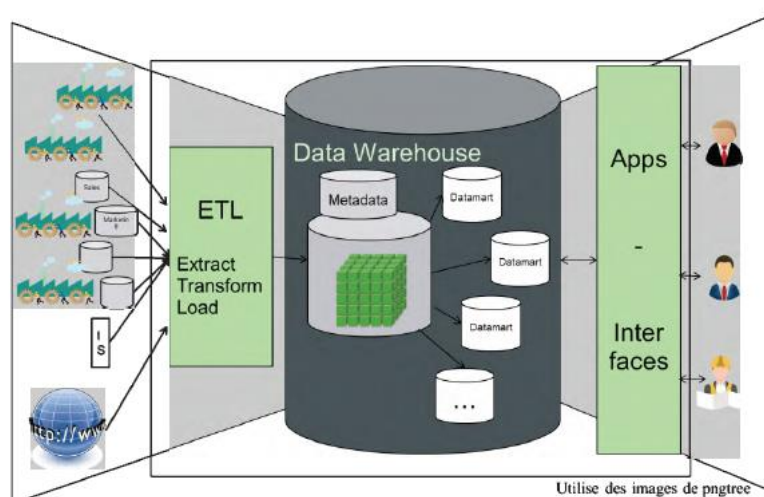


Figura 11 - Arquitetura Standard de Business Intelligence. (Fonte: Linden, 2019)

2.8. Implementações de Sistemas de *Business Intelligence*

Num determinado período de tempo, uma organização gera uma grande quantidade de dados brutos que terão pouca utilidade para retirar qualquer conclusão ou extrair qualquer conhecimento se não forem trabalhados. Segundo Mussa *et al.* (2018), uma ferramenta de BI tem como principal objetivo fornecer informações de maneira correta e no tempo certo para que a empresa possa aplicar as melhores decisões de maneira mais rápida e confiável.

A seleção e a implementação de uma ferramenta de BI, pode tornar-se uma tarefa morosa e difícil. Atualmente há empresas que oferecem um vasto conjunto de soluções, desde produtos *open source*, com baixa tecnologia como a apresentação de simples relatórios, até sofisticadas plataformas de BI. Conforme a funcionalidade e complexidade de cada organização, esta deverá optar pela ferramenta mais adequada. Tendo em consideração a evolução das organizações, estas deverão ter ferramentas facilmente adaptadas e responder às suas expectativas nos próximos anos. Segundo Olszak *et al.* (2007) os produtos de BI podem ser encontrados em diferentes segmentos de mercados. Os fornecedores de sistemas ERP, atualmente estão preocupados em equipar os seus produtos com módulos de BI, como é o caso da SAP, Oracle ou Microsoft. Os autores referem ainda a existência de fornecedores altamente especializados em soluções customizadas, adaptando às necessidades de cada organização. No sentido de sistematizar a análise das ferramentas de BI existentes, segundo os autores Howson *et al.* (2019), as plataformas modernas de BI são caracterizadas por ferramentas fáceis de usar, que derivam de um fluxo de trabalho analítico completo, da preparação de dados à aparência visual. Destacam-se das plataformas de BI tradicionais, não exigindo envolvimento significativo de equipes de TI para predefinir modelos de armazenamento de dados. Uma comparação da classificação das plataformas de BI é apresentada na figura 12. Plataformas como Tableau, Microsoft e outros fornecedores emergentes oferecem um conjunto restrito de recursos, mas são amplamente utilizados para uma variedade de funções de BI e relatórios personalizados. Em geral, eles são fáceis de usar e implantar. A líder de mercado, a Microsoft, oferece preparação de dados, descoberta de dados com base em dashboards, painéis interativos e uma análise aperfeiçoada por meio de um único produto, o Power BI. O Power BI Desktop pode ser usado como uma ferramenta de análise autônoma e gratuita, podendo ser usado por pessoas com poucas competências em tratamento de dados.



Figura 12 – Quadrante Mágico de plataformas de Business Intelligence e Analytics (Fonte: Howson et. al, 2019)

Segundo Fernandes et al (2019), os modelos de BI estão presentes em todos os setores de atividade, sendo que um sistema de BI nunca é implementado por uma única tecnologia, produto ou fornecedor.

Existido diversas ferramentas no mercado, a plataforma Pentaho, assume-se como uma ferramenta polivalente. Uma das variadas aplicações desta plataforma, pode verificar-se no sistema educativo. Segunda Mussa et al. (2018), através da inserção da informação para um DW, a utilização desta ferramenta numa instituição de ensino, permite apresentar os dados de uma forma multidimensional, permitindo análises específicas para os responsáveis educacionais. Inicialmente foi estudado a base de dados da Instituição de Ensino, utilizando o *software* SQL Server 2008 por forma a visualizar as tabelas e relacionamentos entre as mesmas para melhor entender as relações existentes. A figura 13 mostra o diagrama elaborado com a seleção de algumas tabelas.

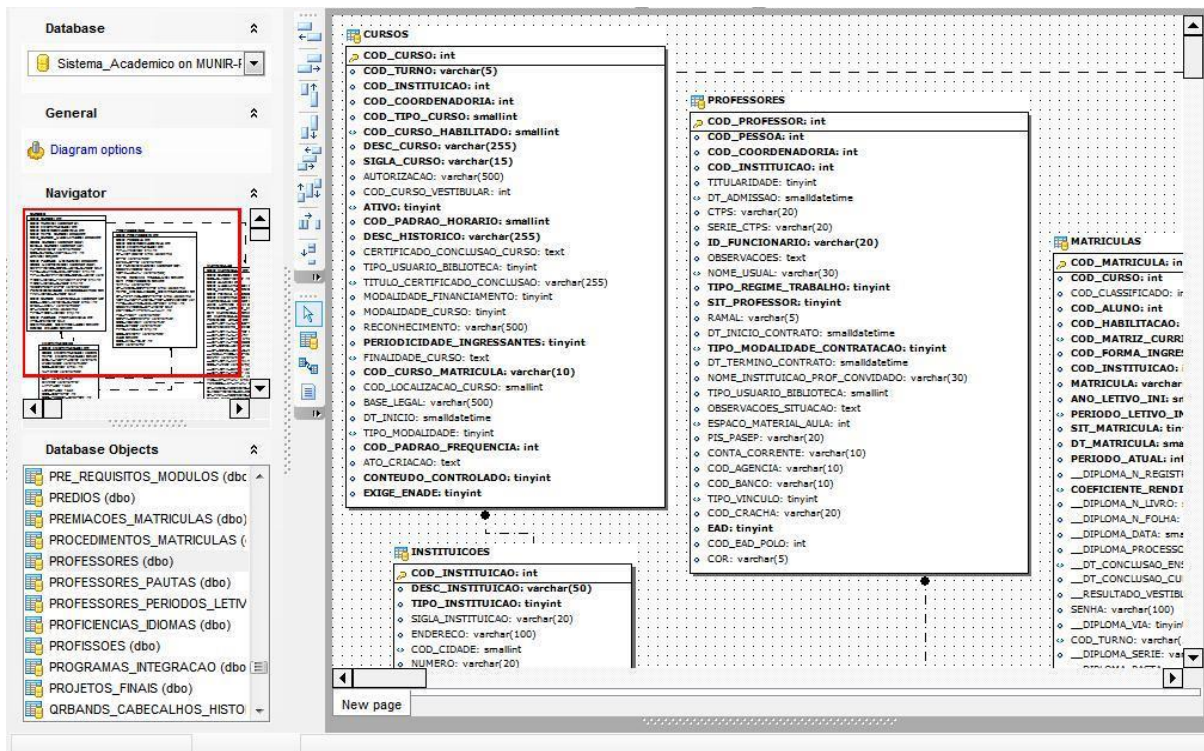


Figura 13 - Apresentação do software SQL Server com as tabelas da base de dados.
(Fonte: Mussa et al., 2018)

Após verificar os relacionamentos entre tabelas, foi feita a definição dos fatos, dimensões e medidas que iriam servir para a construção do cubo OLAP. Neste estudo de caso, foram criadas duas tabelas de fatos, Aluno (figura 14) e Professor, contendo cada uma as chaves primárias das respectivas tabelas de dimensões. A medida a ser analisada para a tabela de fato Aluno é o coeficiente de rendimento, sendo que, para a tabela Professor será a carga horária presencial.

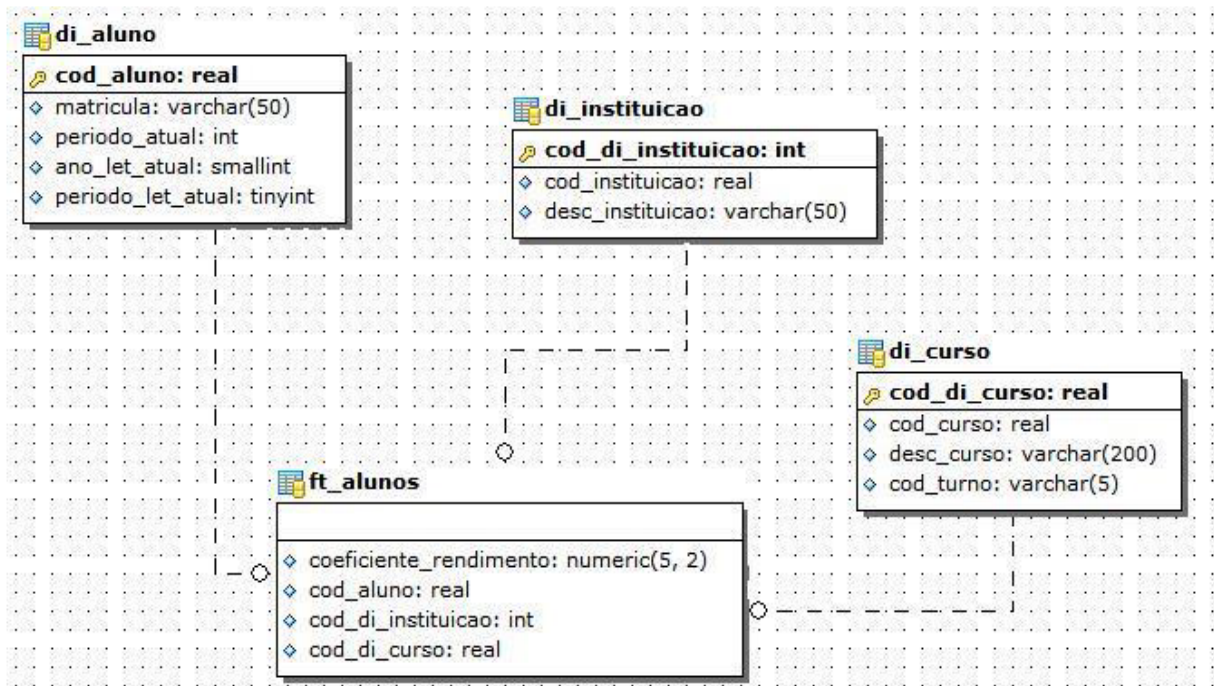


Figura 14 - Tabela de Fatos Aluno. (Fonte: Mussa et al., 2018)

O processo de extração, transformação e carregamento foi realizado com o recurso à ferramenta Pentaho que extrai os dados a partir da base de dados disponibilizada. Posteriormente, após criado o banco de dados multidimensional, é utilizado o *User Dashboard* Pentaho para fornecer informações aos utilizadores, havendo a possibilidade de criar relatórios ou análises conforme figura 15.

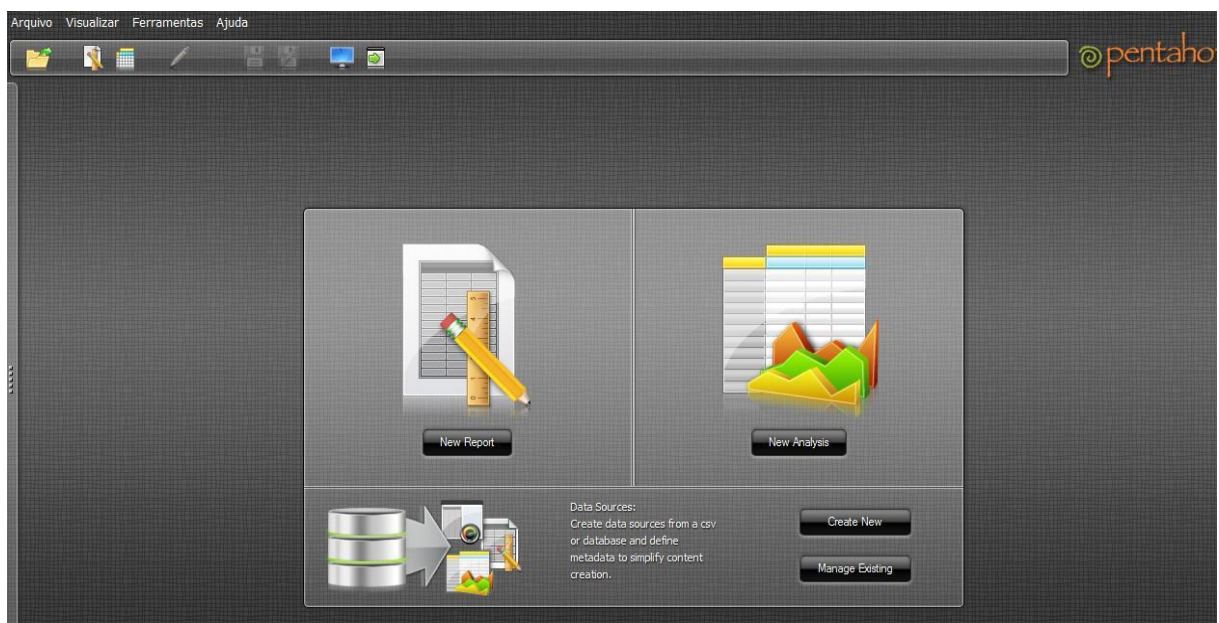


Figura 15 - Apresentação do ecrã da consola Pentaho. (Fonte: Mussa et al., 2018)

Para estar disponível aos utilizadores, foi configurada a conexão com a base de dados da consola para que o Pentaho BI Server pudesse realizar a conexão. Para recolher a informação da base de dados, foi necessário inserir consultas através do SQL para definir os campos e tabelas que se deseja obter. Após a inserção da *query* é fundamental configurar o cubo definindo as dimensões, medidas e hierarquias de cada dimensão conforme é exemplificado na figura 16, onde é possível verificar a configuração feita para a tabela de fatos Alunos.

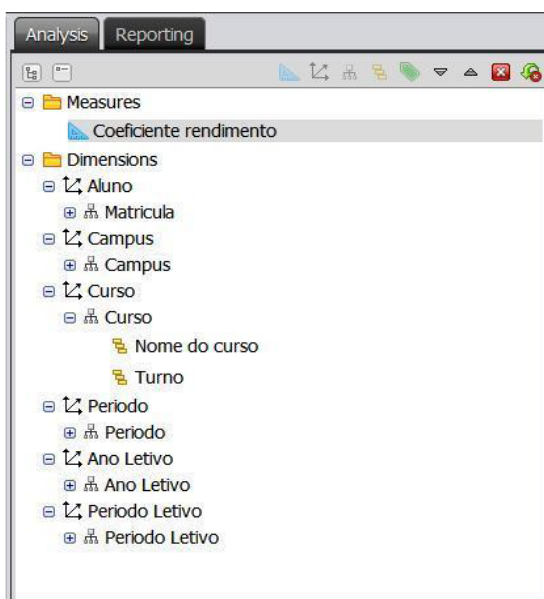


Figura 16 - Configuração do cubo OLAP. (Fonte: Mussa et al., 2018)

A figura 17 mostra o cubo com todas as dimensões previamente definidas e a medida, o coeficiente de eficiência. Para esta etapa, basta iniciar uma nova visualização analítica e seleccionar o cubo desejado.

Nova Visão Analítica

Matricula	Campus	Curso	Período	Ano Letivo	Período Letivo	Measures
+ All Aluno.Matriculas	- All Campuss	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	● Coeficiente rendimento 6,52
	Campus Bom Jesus	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	6,22
	Campus Cabo Frio	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	6,32
	Campus Campos-centro	+ All Cursos	- All Periodos	+ All Ano Letivos	+ All Período Letivos	6,59
			1	+ All Ano Letivos	+ All Período Letivos	5,98
			2	+ All Ano Letivos	+ All Período Letivos	6,16
			3	+ All Ano Letivos	+ All Período Letivos	7,01
			4	+ All Ano Letivos	+ All Período Letivos	7,49
			5	+ All Ano Letivos	+ All Período Letivos	6,66
			6	+ All Ano Letivos	+ All Período Letivos	7,02
			7	+ All Ano Letivos	+ All Período Letivos	7,33
			8	+ All Ano Letivos	+ All Período Letivos	7,61
			9	+ All Ano Letivos	+ All Período Letivos	7,55
			10	+ All Ano Letivos	+ All Período Letivos	7,73
	Campus Itaperuna	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	6,12
	Campus Macaé	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	6,53
	Campus Quissamã	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	6,02
	Câmpus Campos - Guarus	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	5,86
	Reitoria	+ All Cursos	+ All Periodos	+ All Ano Letivos	+ All Período Letivos	8,18

Figura 17 - O Cubo Alunos e as suas dimensões. (Fonte: Mussa et al., 2018)

Neste ponto, o utilizador pode alargar as dimensões de acordo com a informação que pretende consultar. A pesquisa das várias informações dá aos analistas a possibilidade de verificar dados que por vezes se encontram dispersos em diferentes *softwares*. A facilidade na análise dessas informações é um fator decisivo para o apoio à tomada de decisão dos responsáveis no momento certo, constituindo um importante diferencial para a instituição.

Embora uma ferramenta de BI seja por norma associada ao setor comercial e financeiro, conclui-se, pelo caso de estudo cima descrito, que pode ser usado para o apoio à tomada de decisão em qualquer ambiente, estando reunidas as condições necessárias. Esta aplicação permite um grande número de configurações e alterações ao nível das tabelas nos cubos OLAP, gerando deste modo mais informações para os responsáveis prosseguirem decisões estratégicas. O autor Sezões *et al.* (2006), identifica algumas ferramentas mais relevantes e de utilização mais difundida. A aplicação da ferramenta Cognos Report Studio, permite a monitorização das operações de uma organização numa base diária, transmitindo dados importantes como valores, quantidades e alertas para desvios de valores. As organizações que terão mais necessidade destes *reportings* constantes são normalmente as indústrias transformadoras e as cadeias de retalho. Na figura 18, a imagem da ferramenta.

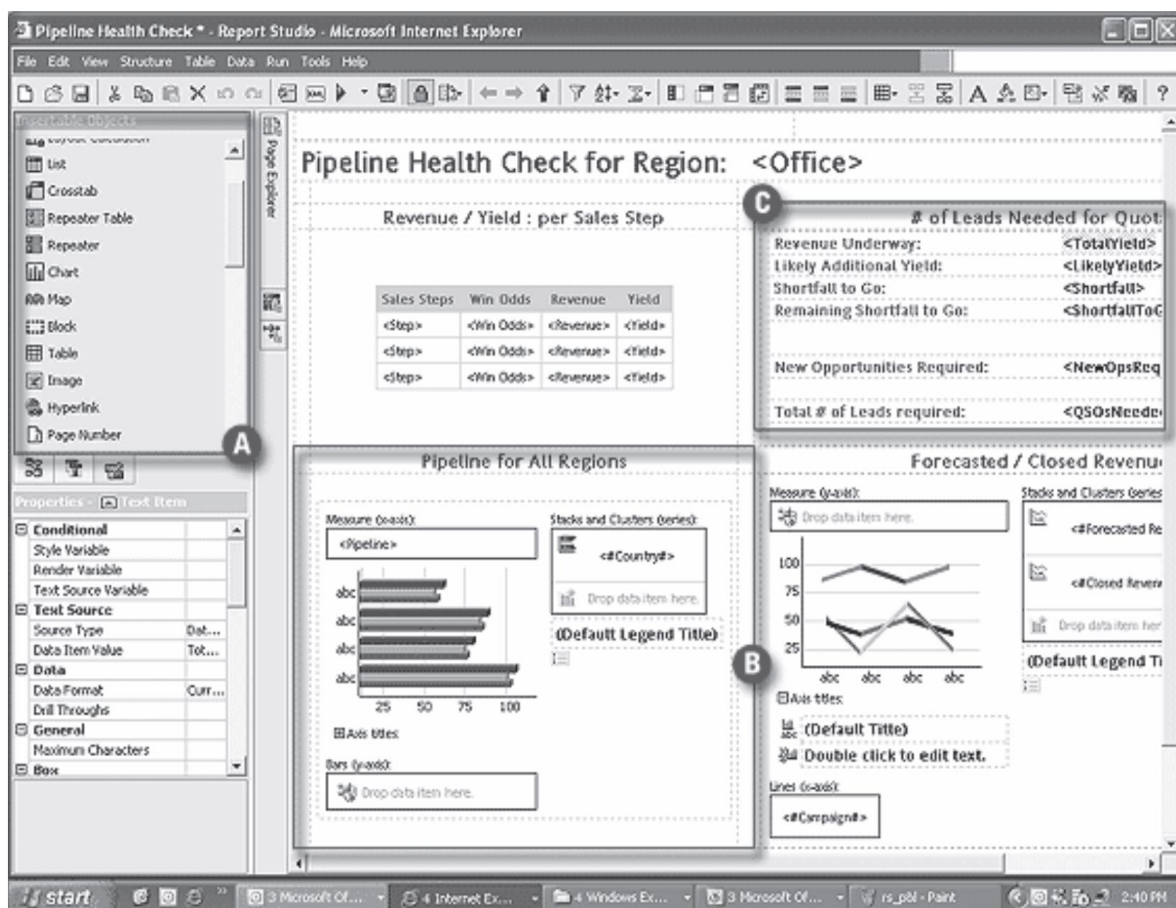


Figura 18 - Aplicação Cognos Report Studio. (Fonte: Sezões et al., 2006)

Existindo vários modelos e implementações de sistemas de BI, os autores Costa e Santos (2012) propõem um modelo que possibilite a uma organização obter informação que apoie os responsáveis no processo decisório em todos os níveis de planeamento, sendo eles estratégicos, táticos ou operacionais. Usufruindo de todo o histórico de informações da sua base de dados, a solução implementada permite à organização o acesso à informação das compras e vendas de produtos, informação contabilística e financeira, bem como, contas correntes de fornecedores. Utilizando como *software* de gestão, o ERP PHC Advanced CS, este é suportado pelo sistema de gestão de base de dados Microsoft SQL Server 2005. A arquitetura definida pelos autores, é composta por um primeiro servidor com a respetiva base de dados do ERP, um segundo servidor com o Data Warehouse, com o processo ETL e os servidores OLAP e de relatórios. A análise de relatórios é suportada por um ambiente *Web* com diversos painéis informativos e também do Microsoft Office Excel. A descrição do modelo implementado pelos autores, é caracterizado conforme figura 19.

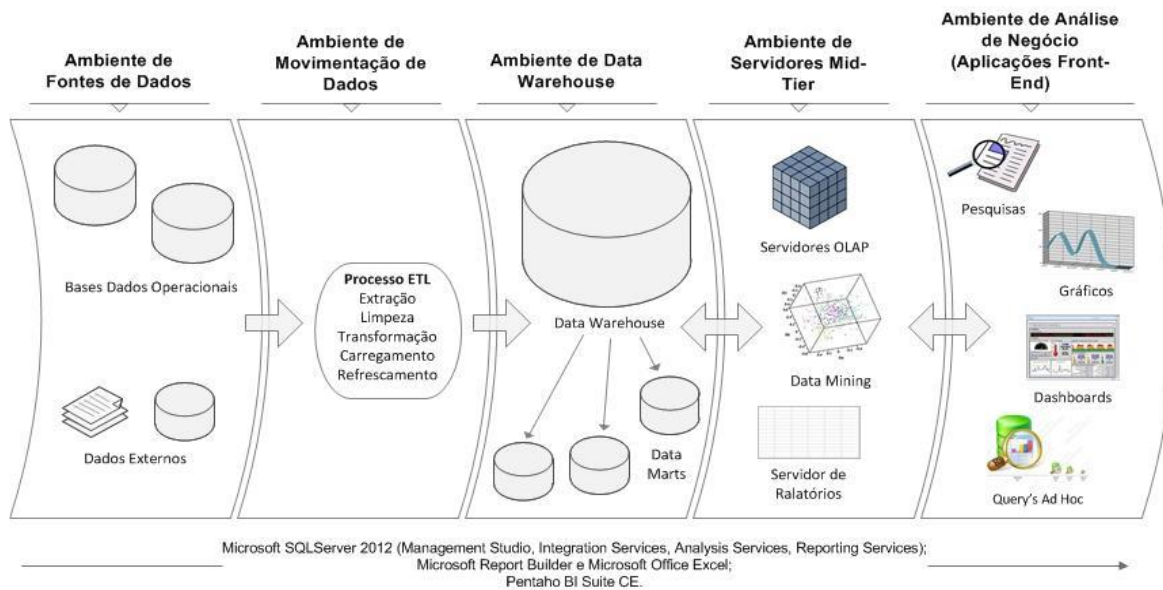


Figura 19 - Arquitetura de modelo de Business Intelligence. (Fonte: Costa e Santos, 2012)

Devido tratar-se de uma base de dados complexa, os autores optaram por criarem views das tabelas consideradas essenciais de modo a não sobrecarregar o sistema de BI. O ambiente de DW além de permitir a modelação multidimensional dos dados, poderá permite a introdução de técnicas de DM. Na criação da interface, os autores implementaram três cenários, i) utilizando uma solução de *front-end* como o Reporting Services; ii) utilizando o Microsoft Excel; iii) recorrendo à ferramenta Business Intelligence Pentaho conforme figura 20.

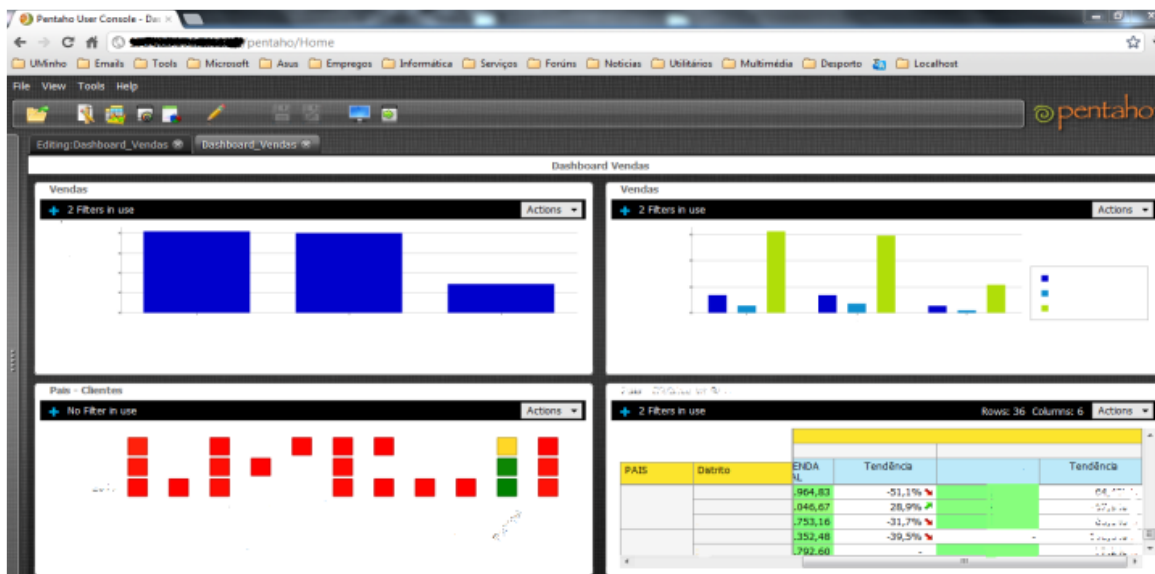


Figura 20 - Dashboard de análise de vendas no Pentaho. (Fonte: Costa e Santos, 2012)

Um outro modelo de implementação de BI, é proposto por Ali, *et al.* (2013), mais concretamente na área da saúde. Com o aumento significativo de taxas de infeção, que acarreta aos sistemas de saúde custos elevadíssimos, a intenção deste projeto é definir e recomendar uma estrutura apropriada de gestão de dados e relatórios para apoiar a estratégia de tomada de decisão. Os autores afirmam que implementar uma sólida estrutura para obter, gerir e relatar os dados e a informação, é um fator chave para o sucesso nestes projetos, com o objetivo, neste caso em particular, de reduzir o número de infeções, e os custos inerentes.

Este projeto de implementação de BI, recorre a ferramentas On-line Transaction Processing (OLTP) responsáveis pelo tratamento de dados de rotina que são gerados diariamente através dos sistemas informacionais da empresa, recorrendo também ao uso do SQL Server 2008 Analysis Services, que se afigura como um mecanismo de dados analíticos usado na análise de negócios e suporte de decisão, que permite resolver problemas de uma forma simples e económica através de uma visão simples e integrada dos dados. Para implementação de um Data Mart foi usado o Microsoft SQL Server 2008 R2 Management Studio, para a extração e transformação dos dados (processo ETL) foi utilizado a package Business Intelligence Development Studio com Visual Studio 2008 do Microsoft SQL Server Integration Services (SSIS). Para o reporting da informação, os autores beneficiaram do potencial da ferramenta Microsoft SQL Server Reporting Services (SSRS) ou do Microsoft Excel 2010. É apresentado na figura 21, a arquitetura de BI, desenvolvida pelos autores:

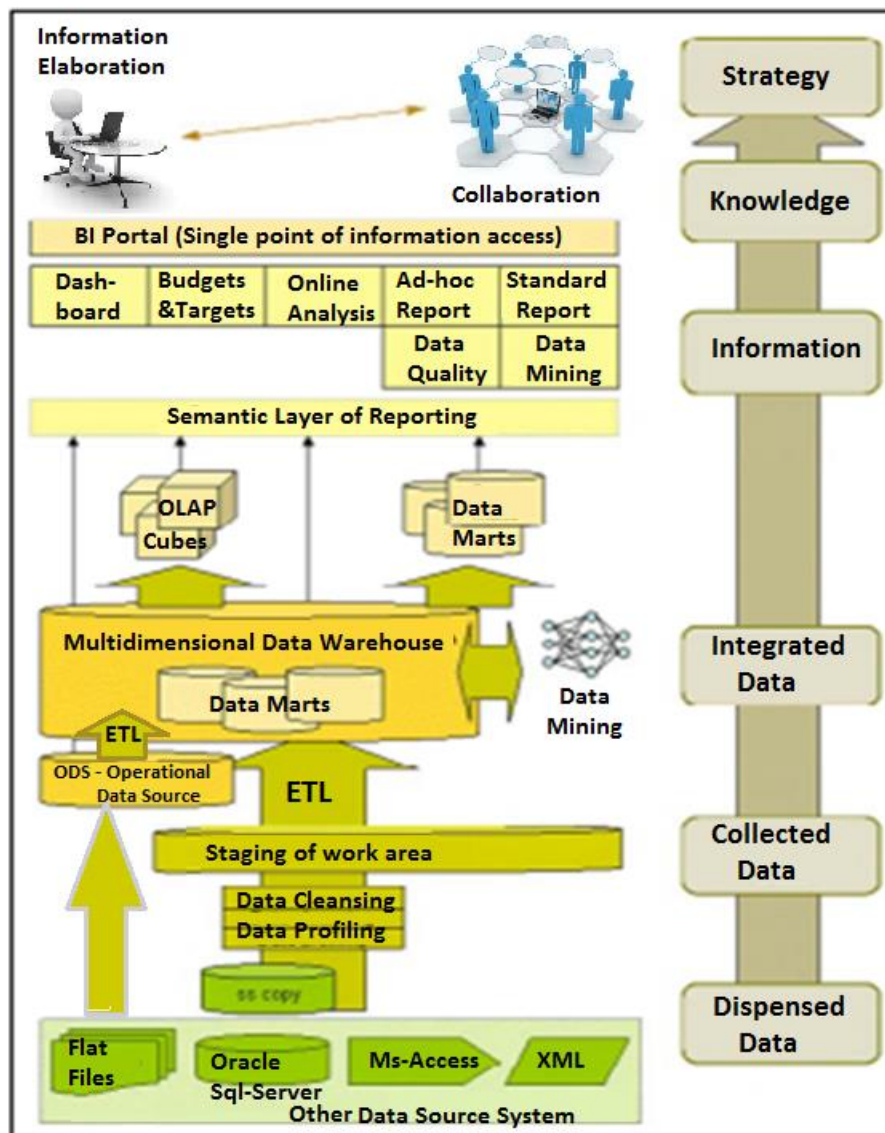


Figura 21 - Os principais componentes da arquitetura de BI.
(Fonte: Ali, et al., 2013)

Neste projeto os autores definiram os seguintes etapas de desenvolvimento: 1) Projeção do *Data Mart* usando o SQL Server Management Studio para desenhar as tabelas de dimensões e fatos; 2) Definir o processo ETL usando a ferramenta SSIS; 3) Com os dados pretendidos, desenhar a estrutura dos cubos OLAP recorrendo a ferramenta SQL Server Analysis Services (SSAS), criando e exibindo a fonte do dados, criar as dimensões associadas, criar o cubo multidimensional e elaborar os *Key Performance Indicators* (KPI) para exibir as métricas mensais; 4) Definir e construir os *reports* de informação usando o SSRS ou o Excel usando tabelas dinâmicas; 5) Adicionar componentes de DM. Com as organizações de saúde a sentirem dificuldades em retirarem informação valiosa dos seus dados, esta ferramenta permite incrementar o pensamento estratégico, através de modificações às ferramentas clássicas de informação para soluções de BI.

Os autores Rao *et al.* (2019), descrevem uma solução de BI onde são exibidas as principais métricas relativas à função financeira, produção, vendas e recursos humanos para monitorizar o desempenho de uma dada organização. A partir da ferramenta instalada SAP BI, que não oferece o processamento de *Big Data* desejado, devido às limitações da arquitetura desta ferramenta, os autores apresentaram uma nova metodologia a partir do repositório de dados SAP *Business* existente, utilizando a ferramenta *Business Intelligence Reporting Tool* (BIRT) de código aberto. Para a preparação de dados é usado o MySQL, para o processo ETL, no desenvolvimento de relatórios é usado a ferramenta Eclipse BIRT, sendo que, para fornecer acesso e implementar aplicativos *Web* é usado o aplicativo Tomcat.

O *dashboard* BIRT BI, foi projetado para permitir que os utilizadores, visualizem de uma forma instantânea as suas operações preferenciais. Para esta implementação, os autores usaram os dados do SAP *Business Warehouse* e a versão Eclipse Oxygen para o desenvolvimento de relatórios em BIRT BI apoiados pelo servidor de aplicativos Apache Tomcat que está ainda conectado ao servidor de base de dados MySQL. O painel relativo às vendas é projetado para fornecer o seu desempenho por região conforme ilustrado na figura 22.

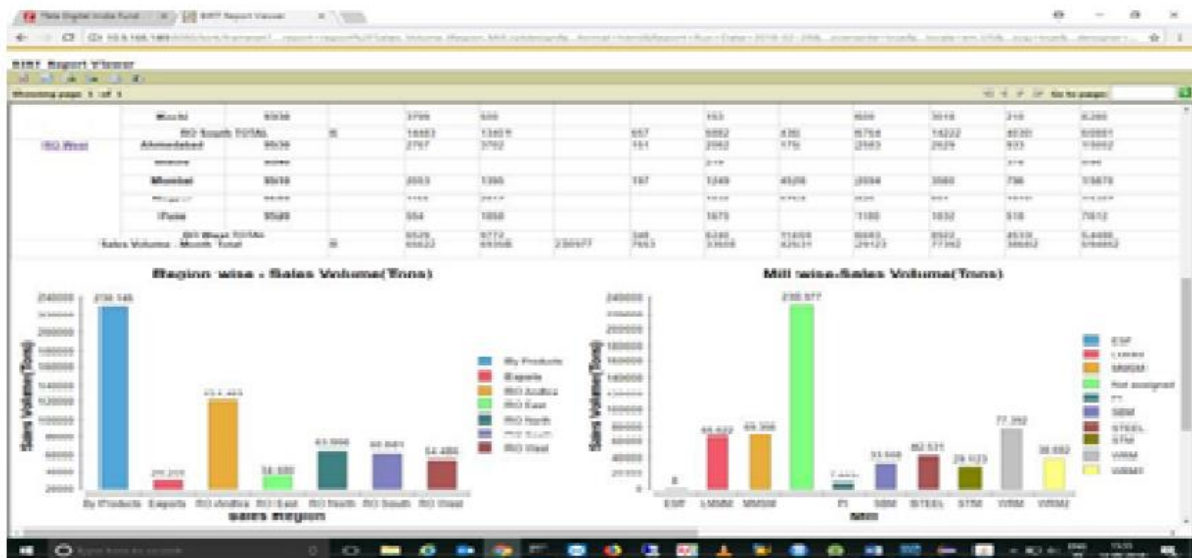


Figura 22 - Detalhe de vendas por região (Fonte: Rao *et al.*, 2019)

As vendas líquidas para cada segmento de clientes, é apresentado na figura 23.

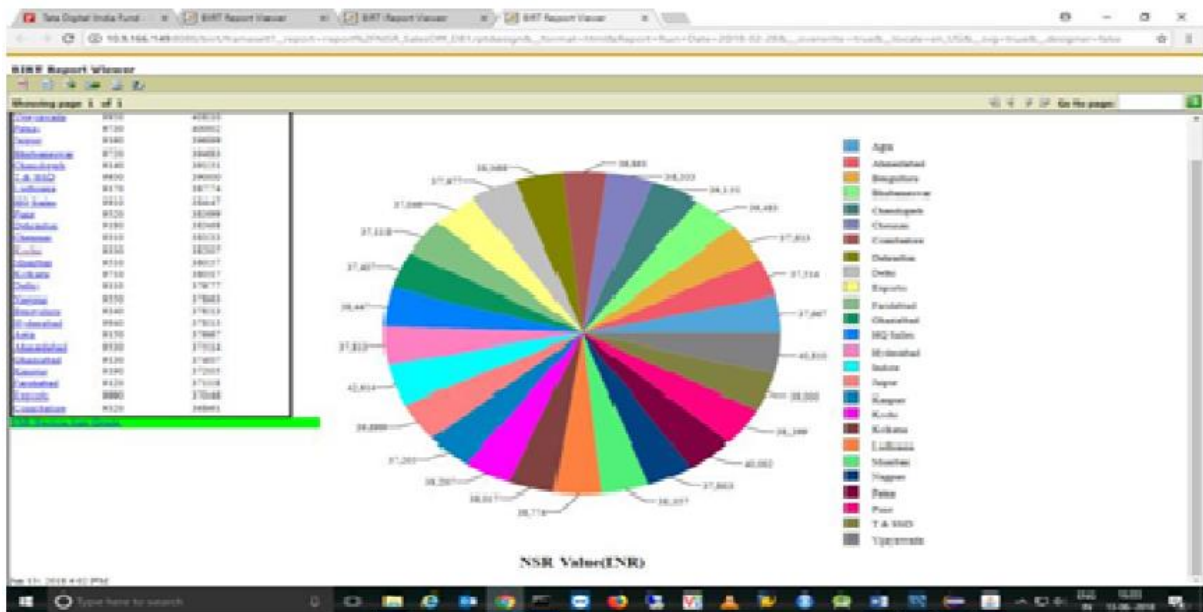


Figura 23 - Detalhes das vendas por cliente e por região. (Fonte: Rao et al., 2019)

A monitorização do desempenho de uma organização, através de qualquer solução de BI, é uma mais valia para aumentar a sua *performance*. Os autores propõem uma solução capaz de fornecer relatórios e KPI num interface visual, personalizado de acordo com o tipo e quantidade de dados exibidos, utilizando ferramentas de BI de código aberto.

CAPÍTULO 3 – DESENVOLVIMENTO DO MODELO DE BI NUMA EMPRESA DE SOFTWARE

3.1. Caso de estudo numa empresa de software

O presente projeto descreve uma solução de BI implementada numa empresa portuguesa, com capital exclusivamente privado, dedicando-se exclusivamente a desenvolver produtos de *software* para o setor das telecomunicações móveis, internet móvel e empresas de outros setores que pretendam fornecer serviços móveis aos seus clientes e colaboradores.

A empresa verificou ao longo dos anos um crescimento notável em receita, número de funcionários e instalações. É classificada como uma grande empresa, de acordo com as classes definidas no Decreto-Lei n. 158/2009 de 13 de julho, que aprova o sistema de normalização contabilístico português. Além da sua sede, onde o seu principal centro de desenvolvimento de *software* e estão instalados os seus serviços centrais, é detentora de mais quatro pontos de desenvolvimento no território nacional.

No início do século XXI, quando Portugal começa a sofrer um momento de estagnação, a sua economia definha, a empresa inicia a sua atividade com apenas oito profissionais produzindo *software* para as telecomunicações móveis essencialmente para o mercado nacional, contudo a empresa rapidamente se afirmou, e em 2005 já contava nos seus quadros com cerca de setenta colaboradores e em 2006 é já certificada pela norma ISO 9001 demonstrando a sua capacidade consistentemente de fornecer produtos de qualidade que atendam às expectativas do cliente. Assegurou continuamente novas parcerias e entrou em novos mercados, nomeadamente com empresas de telecomunicações com uma forte presença nos mercados nacional e europeu. O foco da empresa será manter essa tendência de evolução, conquistando mercados emergentes no setor de telecomunicações.

3.2. Âmbito do Projeto

Num determinado espaço temporal, uma empresa gera uma grande quantidade de dados, contudo, grande parte não sabe como utilizar esta grande quantidade de dados, desconhecendo a sua importante utilidade para obter informações úteis à gestão do negócio. Os dados estando no seu estado bruto, pouca utilidade têm para retirar qualquer informação e conhecimento, pelo que, necessitam de ser tratados e trabalhados.

Devido à estrutura considerável e ao crescimento projetado pela empresa em estudo, um sistema de suporte no nível de tomada de decisão, por meio da implementação de um sistema de BI, é de todo apropriado e necessário. Para que os agentes decisores simplifiquem os processos ao fazer a gestão estratégica, a empresa procura suprimir as folhas de cálculo usadas para analisar indicadores de custo e desempenho, com informações por vezes extensas, sendo um processo moroso e sujeito a falhas. Para atender a essas restrições, a empresa necessita de um sistema que forneça informações de dados históricos e atuais disponíveis na base de dados do seu ERP, e que permita fazer projeções para o futuro auxiliando a administração na tomada de decisões. O desenvolvimento da solução de BI proposta, iniciou-se com reuniões com os agentes decisores da empresa, assim como com o departamento técnico, para compreender os requisitos do modelo de BI e a sua integração na infraestrutura técnica já em funcionamento. Atualmente, a empresa utiliza um sistema ERP como um sistema de gestão integrado, incluindo todas as informações contábilísticas, financeiras e de recursos humanos. A administração mostrou interesse na solução de BI proposta, que apresentaria em tempo real a evolução e as tendências de receita e despesas, por mercado, localização geográfica dos colaboradores e projeções para o futuro ao nível da faturação.

3.3. Descrição do Modelo Proposto

Após a fase inicial na definição dos requisitos da empresa, verificou-se a existência do software de gestão ERP – PHC *Advanced CS*, servindo-se do *Microsoft SQL Server 2012* para a sua gestão de base de dados. Tendo presente todos os requisitos necessários e a estrutura do sistema de informação da empresa, foi efetuada uma análise sobre a melhor ferramenta a aplicar que pudesse interagir com os sistemas já existentes. Neste sentido, a preferência recaiu sobre o software RStudio. Sendo uma ferramenta *open source*, vai ao encontro de todos os requisitos pretendidos para a concretização desta implementação. Para a definição e construção da *user interface* recorreu-se novamente ao *software RStudio* que através do *package R Shiny* permite construir o *dashboard* com os *reports* pretendidos.

O objetivo será extrair as informações necessárias relativas à atividade da empresa, provenientes do ERP, sendo carregadas para o sistema de BI, para posteriormente serem transformadas de modo a poderem ser elaboradas as análises e cálculo dos indicadores económico e financeiros. O *software open source* proposto, está preparado para receber algoritmos de inteligência artificial, nomeadamente redes neuronais, para construir as

projeções de tendências futuras. Sang *et al.* (2016) e a Free Software Foundation¹, definem *software open source* como um *software* que pode ser usado livremente, modificado e compartilhado por qualquer pessoa. Portanto, esta ferramenta pode sempre ser melhorada, sendo por isso a escolha selecionada para o sistema atual.

A apresentação da informação permite a visualização dos dados em tempo real. É gerado através do mesmo *software* em que a interface do utilizador é criada através de painéis informativos, contendo tabelas, gráficos e ferramentas interativas de análise multidimensional que permitem avaliar o estado atual do negócio. Na elaboração desta solução, é esperado a realização de uma ferramenta de BI que proporcione a pesquisa, tratamento e análise da informação considerada pertinente para os responsáveis da gestão de topo, para levarem a cabo uma tomada de decisão mais eficaz e eficiente.

3.4. Descrição e Aplicação do Software RStudio

O *software* RStudio é uma ferramenta de código aberto, que atende a todos os requisitos do sistema R. É uma linguagem de programação e ambiente computacional, especializado em manipulação de dados, análise estatística e visualização gráfica, estando disponível para diferentes sistemas, como Windows, MacOS, GNU/Linux e Unix, sendo o resultado do esforço de vários engenheiros de *software* e investigadores estatísticos a nível mundial. Além de disponibilizar funções base na área da estatística, é possível utilizar diversos pacotes desenvolvidos pela comunidade R, sendo considerado uma plataforma multidisciplinar que pode contribuir em diversas áreas como Economia, Finanças, Engenharia, Saúde e Ciências Sociais (Konrath *et al.*, 2018). Este *software* possui uma extensa coleção de pacotes adicionais também gratuitos, podendo ser descarregado em <https://www.rstudio.com/products/rstudio/download/>.

A tela principal do RStudio é composta por quatro painéis de visualização conforme apresentado na figura 24. A janela Source, no canto superior esquerdo, é uma área onde são apresentados os códigos de programação, inseridos pelos utilizadores. A área do canto superior direito regista as ações e funções executadas, o histórico de todos os *scripts* e as conexões com outras bases de dados. No canto inferior esquerdo, está localizada a janela Console, que permite visualizar os resultados dos *scrips* executados na janela Source. Em baixo à direita, são disponibilizados vários separadores tendo funções como carregar ficheiros, visualizar gráficos, inserir pacotes disponíveis no R e o menu ajuda.

¹ <http://www.fsf.org> (last visited 2019-05-09).

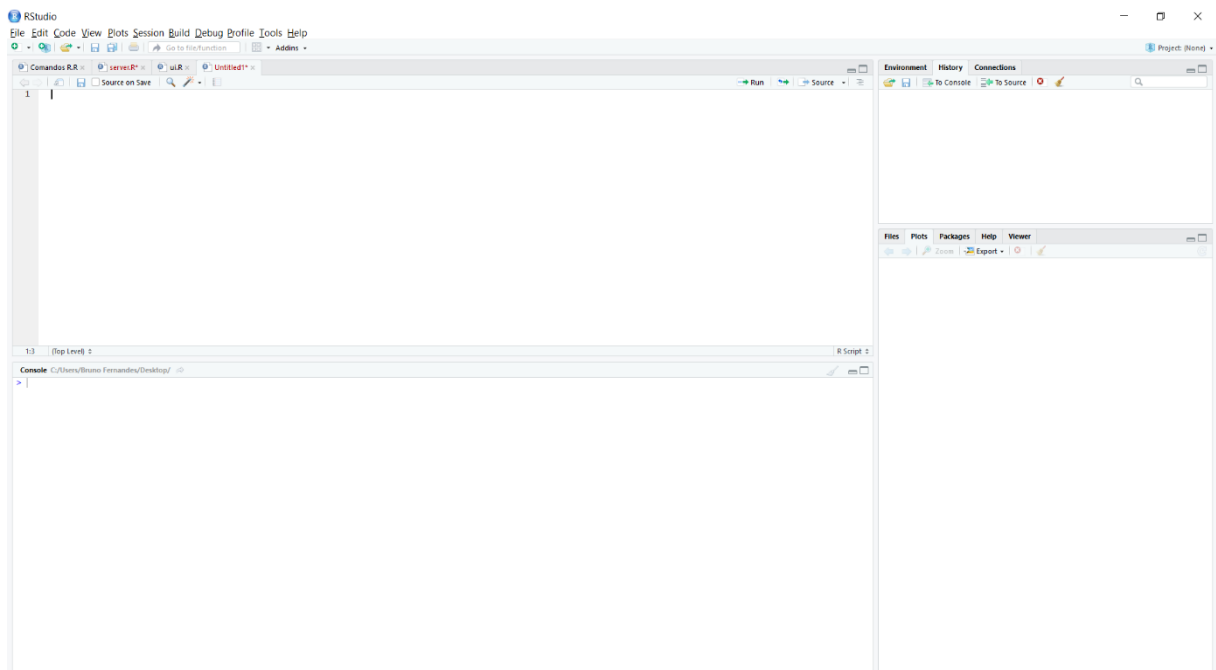


Figura 24 - Interface RStudio (Fonte: RStudio, 2019)

Para o desenho da interface, foi escolhido um pacote específico para o efeito pertencente ao software RStudio. Tem o nome de RShiny, que facilita a criação de aplicativos da *Web* interativos que podem ser usados para exibir dados de maneira interativa com recursos de visualização inovadores. Na construção de aplicativos *web* e desenho da interface gráfica, segundo Konrath *et al.* (2018), o Shiny apresenta dois *scripts* que são gravados num diretório do R. Conforme apresentado no quadro 1 as suas funções e finalidades, o script, *ui.R*, é referente à *user interface* do utilizador (UI). A UI tem a função de controlar o *layout* e o aspeto do *dashboard*.

Quadro 1 - Estrutura da UI. (Fonte: Konrath et al., 2018)

Função	Finalidade
Library (shiny)	Carregar o pacote Shiny.
Shiny UI (fluidpage)	Criar uma interface com o utilizador.
TitlePanel()	Criar um painel contendo um título do aplicativo.
SidebarLayout()	Criar um layout com uma barra lateral e área principal. A barra lateral é exibida com uma cor de fundo distinta e geralmente contém controlos de entrada. A área principal ocupa 2/3 da largura horizontal e geralmente contém saídas.
SidebarPanel()	Criar um painel com barra lateral, que contenha controlos de entrada que, por sua vez, possam ser passados para SidebarLayout.
MainPanel()	Criar um painel principal contendo elementos de saída que, por sua vez, pode ser passado para SidebarLayout.
Shiny Server()	Definir a lógica do servidor do aplicativo Shiny. Isso geralmente envolve a criação de funções que mapeiam entradas dos utilizadores para vários tipos de saída.
Function(input, output) {}	Funções render (funções do R)

O segundo *script* é o servidor e recebe o nome de server.R. A estrutura do *server* é elencada no quadro 2.

Quadro 2 - Estrutura do Server. (Fonte: Konrath et al., 2018)

Função	Finalidade
library(shiny)	Carregar o pacote Shiny.
shinyServer()	Definir a lógica do servidor do aplicativo Shiny. Isso geralmente envolve a criação de funções que mapeiam entradas de usuários para vários tipos de saída.
Function (input,output){}	Funções render (funções do R)

3.5. Arquitetura e Desenvolvimento do Modelo de BI

A figura 25 mostra a arquitetura do modelo, destacando os componentes e tecnologias utilizados para a sua construção. O respetivo processo foi desenhado recorrendo à ferramenta Lucidchart que se define como uma plataforma de fluxogramas e diagramas baseada na *web*, tendo à disposição uma panóplia de recursos para os mais variados fins.

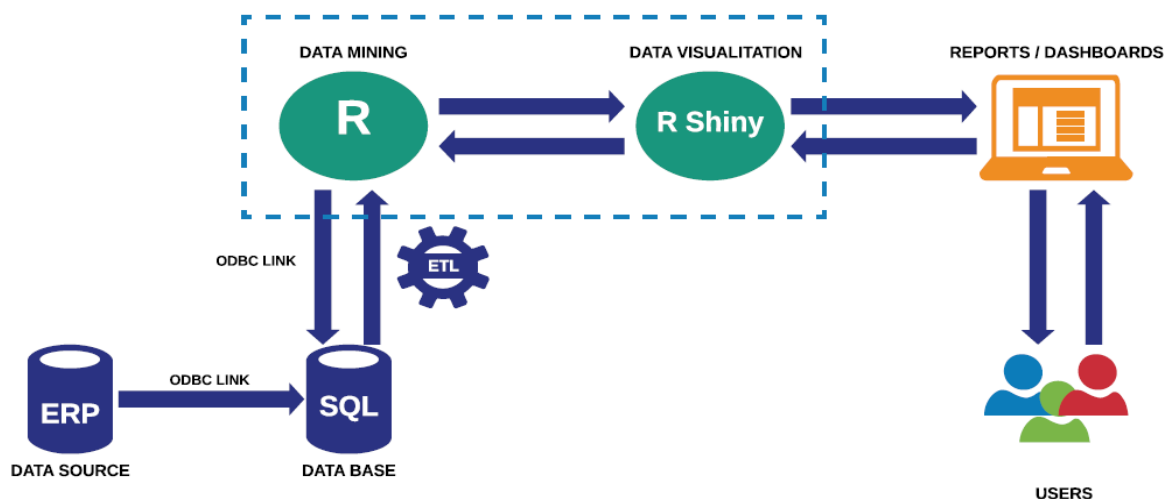


Figura 25 - Arquitetura do sistema de Business Intelligence desenvolvido.
(Fonte: Elaboração Própria)

A arquitetura acima mencionada compreende um servidor que armazena informações copiadas do ERP. As cópias dos dados do ERP para o servidor são feitas automaticamente a cada cinco minutos, através de uma ligação *Open Database Connectivity* (ODBC) do ERP para o Microsoft SQL Server 2012. Esse método garante que as informações do ERP estejam disponíveis quase em tempo real para o sistema de BI, enquanto os dados de BI e os dados do ERP são separados por razões de integridade. O sistema de BI tem acesso somente à leitura ao banco de dados do ERP, enquanto o ERP não tem acesso ao banco de dados da ferramenta de BI. O RStudio acede posteriormente ao banco de dados de BI por meio de outra conexão ODBC, que concede acesso de leitura e gravação, como ilustrado na figura 25. Depois de o conjunto de dados ser extraído da base de dados, os mesmos são processados em R elaborando *queries* em SQL para extrair a informação necessária. Dada a dimensão da base de dados, com este método, não é necessário implementar *views* das tabelas do ERP-PHC consideradas essenciais para o tratamento da informação. Este procedimento designado por processo ETL, permite extrair os dados dos campos das tabelas, que serão utilizados na aplicação em estudo.

Após o processamento e tratamento da informação, é enviado para o R Shiny que utiliza os dados processados para visualização no *dashboards*, oferecendo diferentes opções de visualização para o utilizador, para que as comunicações entre R, R Shiny e o utilizador sejam bidirecionais. A figura 26 mostra o processo de tratamento de dados com mais detalhes, nomeadamente o processo de DM. Os dados selecionados são extraídos da base de dados, alimentando a consola do R responsável pela aplicação da inteligência estatística e artificial necessária, como os algoritmos. Este módulo permite o cálculo dos parâmetros financeiros, com base nas informações extraídas da base de dados, produzindo posteriormente os

gráficos. O módulo também está preparado para incluir uma rede neuronal avançada, nomeadamente modelos preditivos supervisionados, disponíveis em R. A rede neuronal planeada, é o modelo de três camadas. O número de neurônios será determinado experimentalmente nos pontos seguintes, sendo treinada e usada para projetar o crescimento futuro da faturação da empresa. Depois de elaborado o resultado da aplicação da rede neuronal, é preparada a saída da informação sendo visualizada graficamente para o próximo módulo alimentando a *interface*. A *interface* do utilizador foi projetada em R Shiny, como indicado anteriormente, recorrendo a diversos pacotes, para manipulação dos dados e trabalhar os gráficos.

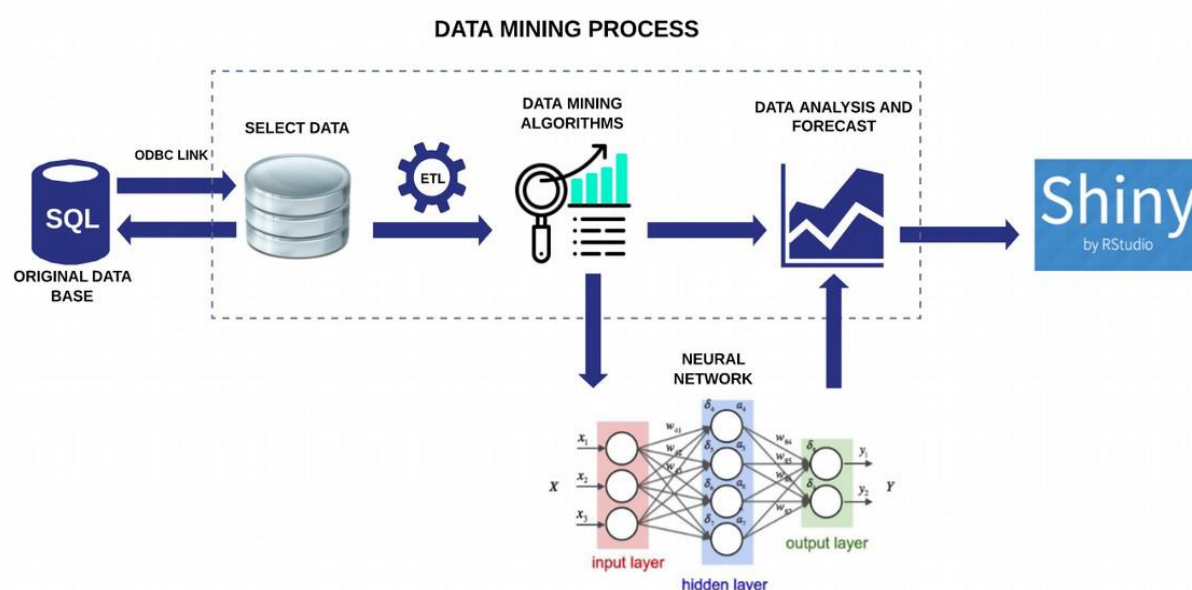


Figura 26 - Detalhes do Processo de Data Mining. (Fonte: Elaboração Própria)

3.6. Arquitetura e Desenvolvimento do Modelo de BI

Um dos objetivos da empresa, é prever a faturação dos próximos trimestres através de uma RNA com base nos dados históricos de faturação. O conjunto de dados usados tiveram por base a faturação líquida incluindo regularizações, os gastos com os colaboradores dos anos de 2010 a 2019. Foram testados diversos modelos de RNA, nomeadamente uma rede *feedforward* de camada única e de multicamada. Para o efeito, foi usado o pacote *neuralnet* do RStudio. Nos pontos seguintes, será detalhado as aplicações das RNA's.

- 1º Passo: Definir e Carregar o Dataset

Ao conjunto de dados, faturação e gastos com colaboradores, agrupou-se a informação por trimestre e por ano. Para a entrada na rede neuronal foram considerados trinta e nove trimestres, compreendidos entre os anos de 2010 a 2019. Os dados foram extraídos aplicando *queries* utilizando a linguagem SQL, conforme figura 27.

```
library(neuralnet)
softmax = custom <- function(x) {log(1+exp(x))}
options(max.print=999999)

importedDataOriginal <- dbGetQuery(con, 'select
  ft.FTANO AS ano,
  CASE WHEN DATEPART(quarter, ft.FDATA) = 1 THEN 1 ELSE 0 END AS Trimestre1,
  CASE WHEN DATEPART(quarter, ft.FDATA) = 2 THEN 1 ELSE 0 END AS Trimestre2,
  CASE WHEN DATEPART(quarter, ft.FDATA) = 3 THEN 1 ELSE 0 END AS Trimestre3,
  CASE WHEN DATEPART(quarter, ft.FDATA) = 4 THEN 1 ELSE 0 END AS Trimestre4,
  (SUM(ft.ettiliq)/1000000) AS TotalInvoice,

  (
  SELECT (sum(pcsa.EDEB-pcsa.ECRE)/1000000)
  FROM pcsa
  where pcsa.conta =\'63\'
  AND DATEPART(quarter, ft.FDATA) = DATEPART(quarter, DATEFROMPARTS ( 2010, pcsa.mes, 1 ) )
  AND ft.FTANO = pcsa.ano
  AND pcsa.mes between 1 and 12

  ) as Employees_Expenses

from ft
where ft.anulado = 0 and ft.fno >0 and ft.FTANO >= year(getdate())-9 and ft.tipodoc <= 3

GROUP BY ft.FTANO, DATEPART(quarter, ft.FDATA)
ORDER by ft.FTANO, DATEPART(quarter, ft.FDATA) ')|
```

Figura 27 - Código em SQL para extração dos dados necessários

Assim se apresentam os dados carregados na consola do RStudio:

```
Console C:/Users/Bruno Fernandes/Desktop/
> print(importedDataOriginal)
  ano Trimestre1 Trimestre2 Trimestre3 Trimestre4 TotalInvoice Employees_Expenses
1 2010          1          0          0          0      0.622770      0.444148
2 2010          0          1          0          0      0.970594      0.580923
3 2010          0          0          1          0      0.540602      0.427841
4 2010          0          0          0          1      2.163074      0.915010
5 2011          1          0          0          0      0.764351      0.537834
6 2011          0          1          0          0      0.878585      0.665344
7 2011          0          0          1          0      1.512774      0.555272
8 2011          0          0          0          1      2.848625      0.789626
9 2012          1          0          0          0      0.565167      0.681452
10 2012          0          1          0          0      2.008653      0.579892
11 2012          0          0          1          0      1.479866      0.777717
12 2012          0          0          0          1      3.772839      1.103740
13 2013          1          0          0          0      2.126514      0.956116
14 2013          0          1          0          0      2.299599      1.165864
15 2013          0          0          1          0      1.694107      1.066090
16 2013          0          0          0          1      4.085913      0.697364
17 2014          1          0          0          0      2.964165      1.227445
18 2014          0          1          0          0      2.876977      1.564341
19 2014          0          0          1          0      5.156611      1.370475
20 2014          0          0          0          1      4.967271      1.567170
21 2015          1          0          0          0      7.472494      1.552791
22 2015          0          1          0          0      5.324634      1.721432
23 2015          0          0          1          0      3.299802      1.619397
24 2015          0          0          0          1      5.216366      2.614950
25 2016          1          0          0          0      9.246789      2.157824
26 2016          0          1          0          0      5.775975      2.273319
27 2016          0          0          1          0      4.601202      2.244031
28 2016          0          0          0          1      5.973593      3.072913
29 2017          1          0          0          0     12.092951      2.525437
30 2017          0          1          0          0      3.438704      2.387622
31 2017          0          0          1          0      6.608662      2.341706
32 2017          0          0          0          1     10.696924      3.730569
33 2018          1          0          0          0     10.675696      3.050424
34 2018          0          1          0          0      6.432794      3.055873
35 2018          0          0          1          0      5.466885      2.905118
36 2018          0          0          0          1      9.702720      4.178666
37 2019          1          0          0          0      2.931822      3.724365
38 2019          0          1          0          0      8.769353      3.598970
39 2019          0          0          1          0      4.323612      2.108696
>
```

Figura 28 - Informação gerada pelo código SQL

- 2º Passo: Normalização dos dados

Um dos procedimentos mais importantes na formação de uma RNA é a normalização dos dados (figura 29). Pressupõe ajustar os dados para uma escala comum, de modo a comparar com precisão os valores previstos e reais. A falha em normalizar os dados normalmente resultará no valor de previsão que permanece o mesmo em todas as observações, independentemente dos valores de entrada.

```

normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}

importedData <- as.data.frame(lapply(importedDataOriginal, normalize)) #valores normalizados

print(importedData)

```

Figura 29 - Código em R para normalização dos dados

Os dados normalizados são apresentados da seguinte forma:

```

> returnHasAllData <- myprediction(importedDataOriginal, list())
[1] "normalized"

```

	ano	Trimestre1	Trimestre2	Trimestre3	Trimestre4	TotalInvoice	Employees_Expenses
1	0.0000000	1	0	0	0	0.007112666	0.004347577
2	0.0000000	0	1	0	0	0.037221175	0.040812888
3	0.0000000	0	0	1	0	0.000000000	0.000000000
4	0.0000000	0	0	0	1	0.140445203	0.129883159
5	0.1111111	1	0	0	0	0.019368269	0.029325015
6	0.1111111	0	1	0	0	0.029256647	0.063320203
7	0.1111111	0	0	1	0	0.084153621	0.033974126
8	0.1111111	0	0	0	1	0.199788199	0.096454780
9	0.2222222	1	0	0	0	0.002126407	0.067614725
10	0.2222222	0	1	0	0	0.127078138	0.040538015
11	0.2222222	0	0	1	0	0.081305023	0.093279745
12	0.2222222	0	0	0	1	0.279790456	0.180200089
13	0.3333333	1	0	0	0	0.137280479	0.140842348
14	0.3333333	0	1	0	0	0.152263146	0.196762846
15	0.3333333	0	0	1	0	0.099850256	0.170162298
16	0.3333333	0	0	0	1	0.306890919	0.071856991
17	0.4444444	1	0	0	0	0.209789628	0.213180834
18	0.4444444	0	1	0	0	0.202242418	0.303000007
19	0.4444444	0	0	0	0	0.399573195	0.251313778
20	0.4444444	0	0	0	1	0.383183455	0.303754241
21	0.5555556	1	0	0	0	0.600041775	0.299920684
22	0.5555556	0	1	0	0	0.414117683	0.344881726
23	0.5555556	0	0	1	0	0.238843200	0.317678377
24	0.5555556	0	0	0	1	0.404745736	0.583100784
25	0.6666667	1	0	0	0	0.753629154	0.461227330
26	0.6666667	0	1	0	0	0.453186880	0.492019222
27	0.6666667	0	0	1	0	0.351495614	0.484210807
28	0.6666667	0	0	0	1	0.470293185	0.705197390
29	0.7777778	1	0	0	0	1.000000000	0.559235901
30	0.7777778	0	1	0	0	0.250866902	0.522493318
31	0.7777778	0	0	1	0	0.525266333	0.510251745
32	0.7777778	0	0	0	1	0.879156438	0.880533749
33	0.8888889	1	0	0	0	0.877318890	0.699201642
34	0.8888889	0	1	0	0	0.510042763	0.700654389
35	0.8888889	0	0	1	0	0.426431283	0.660461898
36	0.8888889	0	0	0	1	0.793095673	1.000000000
37	1.0000000	1	0	0	0	0.206989938	0.878879713
38	1.0000000	0	1	0	0	0.712301109	0.845448401
39	1.0000000	0	0	1	0	0.327466734	0.448129412

Figura 30 - Apresentação dos dados normalizados

- 3º Passo: Preparação e ajuste de parâmetros da rede neuronal

A função *neuralnet* usada no RStudio para treinar uma RNA, permite definir entre outros parâmetros, o número de neurónios nas camadas ocultas de acordo com a complexidade necessária. A complexidade da função calculada aumenta com a adição de camadas escondidas ou neurónios escondidos. Segundo os autores Guenther e Fritsch (2019), os parâmetros padrão usados para o treino de uma rede neuronal e aplicados ao caso em estudo, foram os seguintes:

- **hidden:** Permite definir o número de neurónios nas camadas ocultas;
- **threshold:** O limite é definido como 0,01, o que significa que se a alteração no erro durante uma iteração for menor que 1%, nenhuma otimização adicional será realizada pelo modelo;
- **stepmax:** Número de passos máximos definidos para o treino da rede neuronal. Atingindo seu máximo leva a uma quebra do processo de treino da rede neuronal;
- **rep:** representa o número de repetições para o treino da rede neuronal;
- **startweights:** Um vetor contendo valores iniciais para os pesos. Por norma definido como NULL para inicialização aleatória;
- **learningrate.limit:** define um vetor ou uma lista contendo o limite mais baixo e mais alto para a taxa de aprendizagem, sendo usado apenas para os algoritmos “rprop” e “grprop”. Normalmente definido como NULL para inicialização aleatória;
- **learningrate.factor:** um vetor ou uma lista contendo os fatores de multiplicação para a taxa de aprendizagem mínima e máxima, sendo usado apenas para os algoritmos “rprop” e “grprop”;
- **learningrate:** um valor numérico especificando a taxa de aprendizagem usada pela retropropagação tradicional. Definido como NULL para inicialização aleatória;
- **lifesign:** uma *string* que especifica o número de mensagens que aparecem no ecrã durante o processo de aprendizagem. Poderão ser usados os parâmetros: 'None', 'minimal' ou 'full';
- **lifesign.step:** É uma especificação inteira do tamanho dos passos para imprimir o limite mínimo aplicado apenas quando é usado o *full lifesign mode*;
- **algorithm:** uma *string* contendo o tipo de algoritmo para calcular a rede neuronal. Podem ser usados os seguintes tipos: 'backprop', 'rprop +', 'rprop-', 'sag' ou 'slr'. 'backprop' refere-se à retropropagação, 'rprop +' e 'rprop-' referem-se à retropropagação resiliente com e sem retrocesso de peso, enquanto 'sag' e 'slr' induzem o uso do algoritmo globalmente convergente ('grprop')

- **err.fct:** uma função diferenciável que é usada para o cálculo do erro. Como alternativa, as strings "sse" e "ce", que representam a soma dos erros quadrados;
- **act.fct:** É a função de ativação e é usada para suavizar o resultado do produto vetorial da covariável ou dos neurónios e pesos. Além disso, as strings, 'logistic' e 'tanh' são possíveis para a função logística e tangente hiperbólica;
- **linear.output:** Se act.fct não for aplicado aos neurónios de saída, define-se a saída linear para TRUE, caso contrário, FALSE;
- **exclude:** vetor ou matriz especificando pesos que devem ser excluídos do treinamento. Uma matriz com n linhas e três colunas excluirá n pesos, onde a primeira coluna indica a camada, a segunda coluna o neurónio de entrada de um peso, e o terceiro neurónio a saída do neurónio do peso. Se dado como vetor, a numeração exata deve ser conhecida. A numeração pode ser verificada usando o lote fornecido ou os pesos iniciais salvos;
- **constant.weights:** um vetor que especifica os valores dos pesos que são excluídos do processo de treinamento e tratado como correção;
- **likelihood:** permite encontrar as estimativas de probabilidade máxima de modelos estatísticos usando um algoritmo de otimização global;

Segundo os mesmos autores, para os parâmetros acima descritos, foram usados os seguintes valores padrão e condições genéricas conforme previsto na coluna do modelo A do quadro 3. De modo a otimizar o modelo inicial, foram alterados alguns valores e condições, mantendo os restantes parâmetros gerais iniciais. No modelo B, alterou-se o número de neurónios na camada escondida para quatro bem como o número de repetições para o treino da RNA para 20. Nos modelos A e B é representada uma rede *feedforward* de camada única, sendo que no modelo C, com a alteração do número de neurónios e camadas, é apresentada uma rede *feedforward* multicamada com trinta repetições para o treino da rede. Com a conjugação dos parâmetros alterados em relação ao teste inicial, procedeu-se a uma análise comparativa da precisão, do desvio e do *Mean Squared Error* (MSE) para cada um dos modelos com os respetivos parâmetros modificados.

Quadro 3 - Parametrização dos Modelos de RNA

Parâmetros Aplicados	Valores Definidos Modelo A	Valores Definidos Modelo B	Valores Definidos Modelo C
<i>hidden</i>	2	4	(6,3)
<i>threshold</i>	0.01	0.01	0.01
<i>stepmax</i>	1e+06	1e+06	1e+06
<i>rep</i>	10	20	30
<i>startweights</i>	NULL	NULL	NULL
<i>learningrate.limit</i>	NULL	NULL	NULL
<i>learningrate.factor</i>	list(minus = 0.5, plus = 1.2)	list(minus = 0.5, plus = 1.2)	list(minus = 0.5, plus = 1.2)
<i>learningrate</i>	NULL	NULL	NULL
<i>lifesign</i>	"full"	"full"	"full"
<i>lifesign.step</i>	10000	10000	10000
<i>algorithm</i>	"rprop+"	"rprop+"	"rprop+"
<i>err.fct</i>	"sse"	"sse"	"sse"
<i>act.fct</i>	"logistic"	"logistic"	"logistic"
<i>linear.output</i>	TRUE	TRUE	TRUE
<i>exclude</i>	NULL	NULL	NULL
<i>constant.weights</i>	NULL	NULL	NULL
<i>likelihood</i>	FALSE	FALSE	FALSE

Definidos os parâmetros no quadro acima, são apresentadas nas figuras seguintes, as redes neurais geradas, de acordo com os modelos A, B e C. Para os dados de entrada são considerados o ano, employees expenses (gastos com os colaboradores), o trimestre 1, trimestre 2, trimestre 3 e o trimestre 4. Os dados de saída são o total de faturação para cada trimestre.

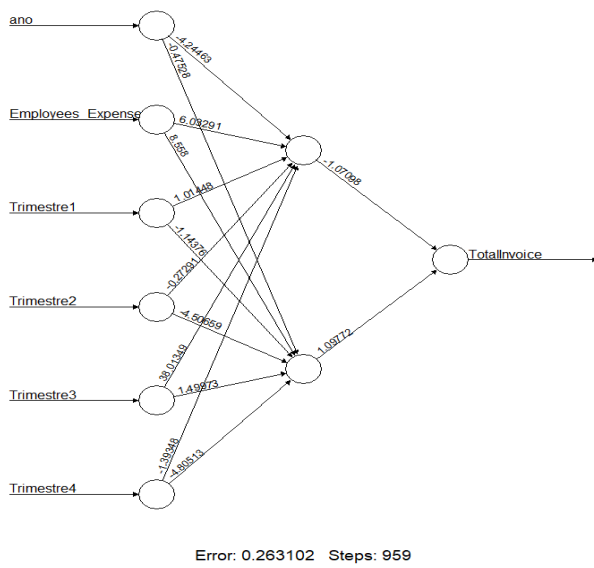


Figura 31 - Rede Neuronal Modelo A

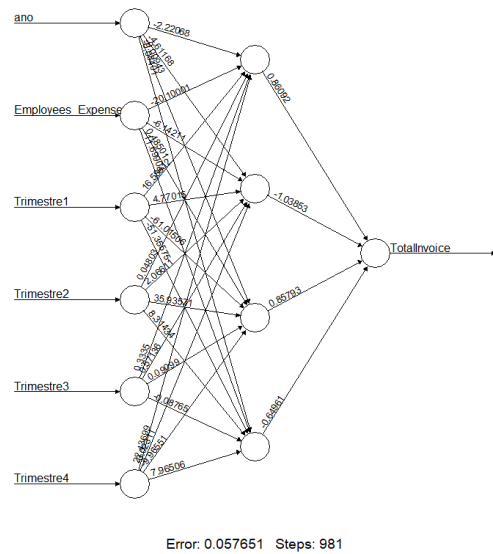


Figura 32 - Rede Neuronal Modelo B

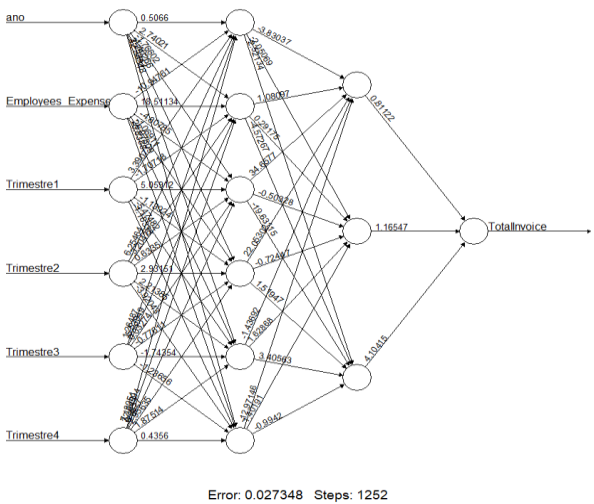


Figura 33 - Rede Neuronal Modelo C

- 4º Passo: Validação do Modelo

Posteriormente são validados e relatados os dados na saída do teste da RNA. Os resultados previstos são comparados com os resultados reais e mencionado o respetivo

desvio dos trinta e nove trimestres analisados. O desvio é calculado em R segundo a equação número 2 abaixo indicada, representando uma porcentagem, exemplificando, se “*Desvio*” igual a -0.05, significa -5%. Pode assim considerar-se, que os valores gerados para os desvios, são os erros entre os valores pretendidos e os valores obtidos.

O quadro 4 apresenta os resultados já desnormalizados dos três modelos em análise, quando foram treinados e testados com todo o conjunto de dados, sendo que, se pode verificar, que o modelo C apresenta valores previstos mais próximos da realidade e conseqüentemente um reduzido desvio nas respectivas iterações.

$$Desvio = \frac{Atual - Previsto}{Atual} \quad (2)$$

Quadro 4 - Resultados obtidos dos três modelos neurais, A, B e C, para prever a faturação da empresa. A tabela mostra o valor atual, o valor previsto e o erro entre os dois (desvio), em percentagem, para cada modelo

Anos / Trimestres	Atual	Modelo A		Modelo B		Modelo C	
		Previsto	Desvio	Previsto	Desvio	Previsto	Desvio
2010 T1	0,6228	1,3120	-1,1067	-0,5755	1,9241	0,6843	-0,0988
2010 T2	0,9706	0,3159	0,6745	0,9446	0,0268	0,9699	0,0007
2010 T3	0,5406	1,5473	-1,8623	0,8332	-0,5412	0,6707	-0,2406
2010 T4	2,1631	3,4567	-0,5980	1,6843	0,2213	2,2277	-0,0299
2011 T1	0,7644	1,5838	-1,0721	0,5791	0,2423	0,7649	-0,0007
2011 T2	0,8786	0,5158	0,4129	1,3691	-0,5584	0,9088	-0,0344
2011 T3	1,5128	1,5834	-0,0467	1,1395	0,2468	1,0215	0,3247
2011 T4	2,8486	3,4581	-0,2139	2,3903	0,1609	2,7748	0,0259
2012 T1	0,5652	2,0502	-2,6276	1,8506	-2,2744	0,9337	-0,6521
2012 T2	2,0087	0,5931	0,7047	1,6291	0,1890	1,9813	0,0136
2012 T3	1,4799	1,6558	-0,1189	1,6949	-0,1453	1,7163	-0,1597
2012 T4	3,7728	3,4976	0,0729	3,7406	0,0086	3,8050	-0,0085
2013 T1	2,1265	3,0459	-0,4324	3,4616	-0,6278	1,5087	0,2905
2013 T2	2,2996	1,9143	0,1675	2,4879	-0,0819	2,1401	0,0694
2013 T3	1,6941	1,8230	-0,0761	2,4567	-0,4501	2,7131	-0,6015
2013 T4	4,0859	3,4720	0,1502	4,0700	0,0039	4,1559	-0,0171
2014 T1	2,9642	4,3564	-0,4697	4,8719	-0,6436	3,2046	-0,0811
2014 T2	2,8770	3,5838	-0,2457	3,2918	-0,1442	3,6032	-0,2524
2014 T3	5,1566	2,1835	0,5766	3,2820	0,3635	3,5773	0,3063
2014 T4	4,9673	3,6949	0,2562	5,7997	-0,1676	4,8934	0,0149
2015 T1	7,4725	5,8521	0,2168	6,2004	0,1702	7,4014	0,0095
2015 T2	5,3246	4,5790	0,1400	3,7456	0,2966	4,6120	0,1338
2015 T3	3,2998	2,7727	0,1597	3,9309	-0,1913	3,9247	-0,1894
2015 T4	5,2164	5,2071	0,0018	6,9661	-0,3354	5,1829	0,0064
2016 T1	9,2468	7,3702	0,2029	7,6363	0,1742	9,9631	-0,0775
2016 T2	5,7760	6,2453	-0,0813	5,0055	0,1334	4,7855	0,1715
2016 T3	4,6012	4,8285	-0,0494	5,4630	-0,1873	5,2502	-0,1410
2016 T4	5,9736	7,0065	-0,1729	7,5537	-0,2645	6,0382	-0,0108
2017 T1	12,0930	7,8171	0,3536	8,3597	0,3087	11,0891	0,0830
2017 T2	3,4387	6,6120	-0,9228	5,3815	-0,5650	4,7031	-0,3677
2017 T3	6,6087	5,6526	0,1447	5,6148	0,1504	5,7191	0,1346
2017 T4	10,6969	9,3409	0,1268	7,9706	0,2549	10,7216	-0,0023
2018 T1	10,6757	8,0317	0,2477	8,8987	0,1665	10,7151	-0,0037
2018 T2	6,4328	7,1263	-0,1078	6,8721	-0,0683	6,2076	0,0350
2018 T3	5,4669	7,5448	-0,3801	6,5426	-0,1968	5,5533	-0,0158
2018 T4	9,7027	10,2487	-0,0563	8,2527	0,1494	9,7523	-0,0051
2019 T1	2,9318	8,0860	-1,7580	9,1500	-2,1209	2,9975	-0,0224
2019 T2	8,7694	7,2415	0,1742	7,8089	0,1095	9,0264	-0,0293
2019 T3	5,7341	2,5895	0,5484	1,5076	0,7371	1,8464	0,6780

Segundo os autores Ciaburro e Venkateswaran (2017), a precisão é a medida do quão bom é o nosso modelo. Se um determinado modelo tiver um bom desempenho, tudo indica que está muito próximo do valor 1.

Determinar qual o número de camadas ocultas numa rede neuronal para um determinado conjunto de dados, não é uma ciência exata. Uma possibilidade, é comparar

como a precisão dos prognósticos muda à medida que modificamos o número de camadas ocultas. O quadro 5 apresenta os valores da precisão de cada um dos modelos testados.

Para o cálculo da precisão dos modelos, foi usada a seguinte equação número 3 em R:

$$Precisão = 1 - |(De\acute{s}vio)| \quad (3)$$

Quadro 5 – Precisão e desvio gerados pelos modelos

	Modelo A	Modelo B	Modelo C
Precisão	0.7926518	0.8943626	0.962269
Desvio	0.2073482	0.1056374	0.037731

Como é evidenciado no quadro acima, o modelo C é o mais preciso e o modelo A é o menos preciso. Isso mostra que o problema de prever rotatividade é melhor resolvido por uma rede com pelo menos duas camadas ocultas, no caso atual com 6 e 3 neurónios cada. Para demonstrar que as previsões geradas pelos três modelos são precisas, é avaliado o desempenho da RNA nos vários modelos, de acordo com os resultados do teste previsto e os resultados observados, com base na medição do erro quadrático médio MSE, que indica como até onde estão, as previsões dos dados reais. O MSE resulta da diferença entre a produção esperada (valor real) e a produção prevista pela RNA de acordo com a equação 4. Os erros MSE para cada modelo são mostrados no quadro 6.

$$MSE = \frac{\sum(Real - Previsão)^2}{Totallinhas} \quad (4)$$

Quadro 6 - Resultado do MSE para os três modelos

	Modelo A	Modelo B	Modelo C
MSE – Mean Squared Error	0.02007205	0.01941092	0.002183372

A partir dos resultados obtidos, os modelos A e B geraram um MSE maior. Esse é um resultado esperado, considerando os valores de precisão obtidos acima. Para uma melhor percepção do erro gerado pelos três modelos e para avaliar a precisão das estimativas geradas pela RNA, foi gerado um gráfico de regressão linear para avaliar o desempenho dos três modelos. Nos gráficos abaixo, é possível ver que as previsões feitas pelo modelo C estão

mais próximas da linha de destino, quando comparadas aos modelos A e B. Uma combinação perfeita com a linha indicaria um MSE de 0 e, portanto, uma previsão perfeita ideal.

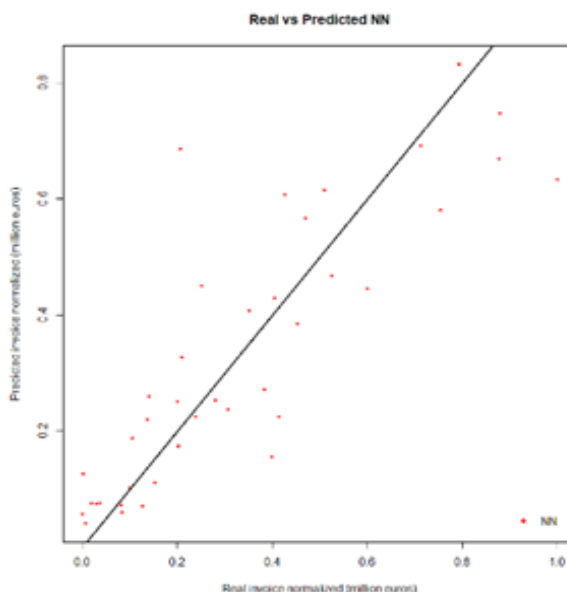


Gráfico 1 - Valores previstos e reais do modelo A

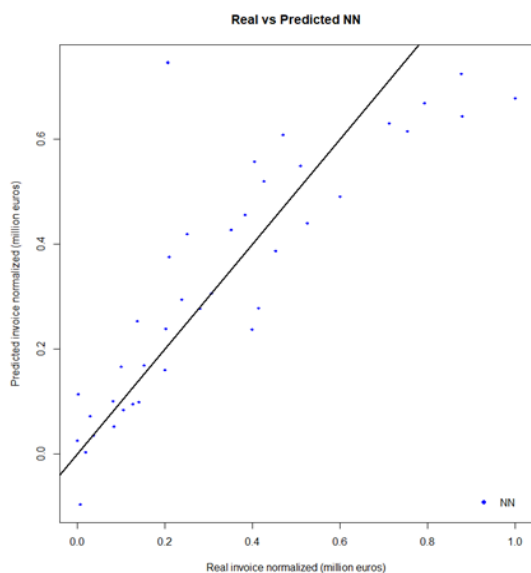


Gráfico 2 - Valores previstos e reais do modelo B

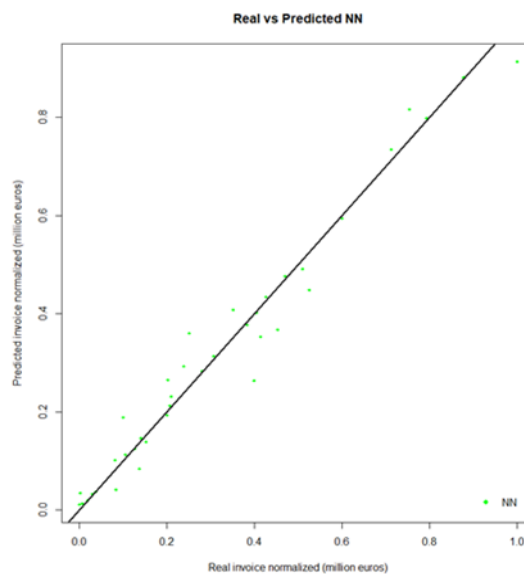


Gráfico 3 - Valores previstos e reais do modelo C

Para uma melhor percepção de valores e a título exemplificativo, é mencionado por trimestre no quadro 7, os valores reais de 2018 e 2019, a previsão que a rede gerou para os mesmos trimestres e o seu desvio, do modelo que melhor previsão efetuou, neste caso, o modelo C. O quadro abaixo é um resumo do quadro nº 4 e é treinado com os mesmos dados

definidos inicialmente. Dado que o ano 2019 foi um ano atípico para a empresa em termos de faturação, verifica-se um maior desvio a tender para valores negativos nos últimos trimestres, o que indica que a faturação ficou aquém do previsto.

Quadro 7 - Valores gerados pelo modelo C nos últimos trimestres

Trimestres	Faturação Real	Faturação Prevista	Desvio
2018 T1	10.675.696 Eur	10.715.07,66 Eur	1%
2018 T2	6.432.794 Eur	6.207.554,0 Eur	4%
2018 T3	5.466.885 Eur	5.553.257,1 Eur	-2%
2018 T4	9.702.720 Eur	9.752.262,1 Eur	-1%
2019 T1	2.931.822 Eur	2.997.455,2 Eur	-2%
2019 T2	8.769.353 Eur	9.026.425,0 Eur	-3%
2019 T3	5.734.065 Eur	6.037.150,8 Eur	-5%
2019 T4	--	9.752.262,1 Eur	--

Dado que a faturação prevista pela rede se aproxima bastante da realidade, verificado pelo baixo desvio, pressupõe-se que a rede está bem adaptada. Segundo Blundell *et al* (2015), as RNA *feedforward* são propensas a *overfitting*, quando aplicadas a modelos com aprendizagem supervisionada, sendo frequentemente incapazes de avaliar corretamente a incerteza nos dados de treino tomando decisões por vezes excessivamente confiantes sobre a previsão. Uma das soluções mais eficazes de reduzir o *overfitting* é aumentar o volume ou tamanho da amostra dos dados de treino. No caso de estudo acima, os dados são reduzidos e não foi criado um conjunto de treino e um de teste, pelo que não é possível comprovar a existência de *overfitting*.

Como verificado no quadro 8, para avaliar o comportamento do modelo e seu poder preditivo, o modelo C foi treinado com apenas pequenos subconjuntos de dados e testado para prever a faturação nos trimestres subsequentes. No primeiro teste, o conjunto de treino incluiu dados até 2016 e o modelo foi testado para prever a faturação até o terceiro trimestre de 2019. Em seguida, o conjunto de treino foi ampliado com dados de 2017 e 2018. Como mostra o quadro abaixo, os modelos treinados com mais dados continuam aprendendo e produzindo erros menores em geral nas amostras de teste.

Quadro 8 – Previsões geradas pelo Modelo C com amostras de diferentes dimensões

Anos/ Trimestres	Valores Reais €	Modelo C1		Modelo C2		Modelo C3		Modelo C4	
		Treino até 2016		Treino até 2017		Treino até 2018		Treino até 2019 T3	
		Valor Previsto €	Desvio %	Valor Previsto €	Desvio %	Valor Previsto €	Desvio %	Valor Previsto €	Desvio %
2016 T1	9 246 789	9 321 833	-1%	9 927 438	-7%	9 710 440	-5%	10 356 139	-12%
2016 T2	5 775 975	5 996 440	-4%	4 650 941	19%	4 852 647	16%	4 924 912	15%
2016 T3	4 601 202	4 642 592	-1%	5 313 559	-15%	5 126 983	-11%	5 270 449	-15%
2016 T4	5 973 593	5 959 229	0,2%	7 375 256	-23%	7 323 376	-23%	6 516 695	-9%
2017 T1	12 092 951	9 321 833	-23%	11 829 040	2%	10 751 133	11%	11 300 211	7%
2017 T2	3 438 704	5 996 440	-74%	4 991 816	-45%	5 120 572	-49%	4 315 380	-25%
2017 T3	6 608 662	4 642 592	30%	5 927 395	10%	5 351 830	19%	5 621 137	15%
2017 T4	10 696 924	5 959 229	44%	9 134 817	15%	9 088 547	15%	10 290 431	4%
2018 T1	10 675 696	9 321 833	13%	11 829 040	-11%	11 494 284	-8%	10 518 618	1%
2018 T2	6 432 794	5 996 440	7%	4 991 816	22%	6 072 696	6%	6 347 099	1%
2018 T3	5 466 885	4 642 592	15%	5 927 395	-8%	6 038 877	-10%	5 428 274	1%
2018 T4	9 702 720	5 959 229	39%	9 134 817	6%	10 223 594	-5%	9 777 454	-1%
2019 T1	2 931 822	9 321 833	-218%	11 829 040	-303%	11 494 284	-292%	3 157 808	-8%
2019 T2	8 769 353	5 996 440	32%	4 991 816	43%	6 072 696	31%	8 641 418	1%
2019 T3	5 734 065	4 642 592	19%	5 927 395	-3%	6 038 877	-5%	6 037 150	-5%
2019 T4	0	5 959 229	-	9 134 817	-	10 223 594	-	9 777 454	
Precisão		96%		95%		97%		98%	
Erro Quadrático Médio		0,003949948		0,005496366		0,005603120		0,002520385	

Após analisar os valores gerados pelo modelo C, é possível verificar que a rede neuronal se adapta continuamente aos dados e aprende, apesar do tamanho pequeno e da baixa qualidade do conjunto de dados. É difícil prever a faturação da empresa em particular, pois apresenta grandes flutuações de ano para ano e de trimestre para trimestre. No entanto, a rede neuronal aprende e ajusta a qualidade das previsões, quando mais amostras estão disponíveis. Portanto, o modelo pode melhorar continuamente ao longo do tempo, quando mais dados estão disponíveis para a mesma empresa.

CAPÍTULO 4 – RESULTADOS DO MODELO DE BI

Tipicamente um *dashboard* é uma área de visualização que pode conter vários elementos como tabelas, gráficos, indicadores ou mapas. É uma representação visual que possibilita exibir a performance de uma organização. Segundo Barros (2013), os *dashboards* assumem uma importância fundamental na supervisão da organização apoiando o processo de apoio à decisão. Os mesmos devem ser organizados graficamente, devendo ter vistas gráficas adequadas para a informação que se pretende exibir.

Com o projeto finalizado da implementação da ferramenta de BI e ao desenho e aspeto do *dashboard*, é objetivo principal, fornecer uma ferramenta de BI que possibilite a análise dos dados de maneira fácil e intuitiva.

A figura 34, mostra a página principal do *dashboard* produzido pelo sistema de BI proposto. Como pode ser evidenciado, o primeiro separador, “*Financial Reports*” exibe informação relativa à evolução da faturação por ano, por mercado, gastos com pessoal e reconhecimentos de trabalhos efetuados. O segundo separador na figura 35, mostra a distribuição dos colaboradores pelo território nacional. Na figura 36, o terceiro separador evidencia a evolução da faturação e as respetivas previsões, usando RNA anteriormente desenvolvidas no capítulo anterior.

O separador “*Financial Reports*” (figura 34) evidencia em tempo real a situação da empresa em relação aos seus gastos e rendimentos. As caixas a verde exibem o ano com mais faturação registada e rendimentos reconhecidos e os valores presentes à data respetivamente. As caixas informativas a laranja, mostram os gastos com os colaboradores e os gastos totais da empresa para o ano atual, em comparação com o ano com maiores gastos registados. A primeira linha de gráficos evidencia a faturação e os valores reconhecidos de 2010 até à data, bem como, os resultados gerados por mercado. É de ressaltar o facto, que a empresa tem vindo a obter bons resultados de ano para ano, no entanto, o ano corrente mostra-se atípico com um valor de faturação registada abaixo do normal, tendo em conta os sete meses já decorridos. É possível também destacar a posição da empresa perante o mercado, revelando-se claramente uma empresa exportadora, tendo um grande impacto o peso dos clientes da União Europeia e Países Terceiros, sendo o valor da faturação a nível nacional praticamente residual.

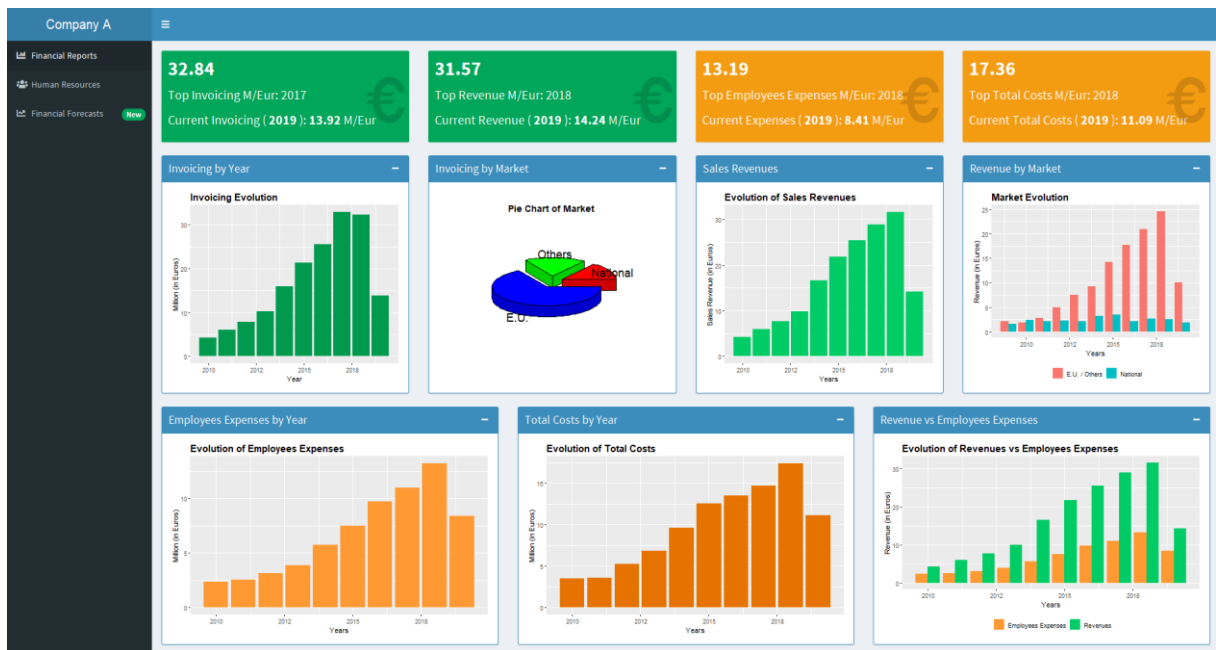


Figura 34 - Página principal do dashboard

A figura 35, mostra o segundo separador, “*Human Resources*” com a distribuição dos recursos humanos da empresa. Esta informação visual permite verificar onde se encontra a maior força de trabalho, sendo que atualmente, a maior concentração de colaboradores está no centro de Portugal, com a segunda maior no Norte e uma pequena representação no Centro Sul de Portugal. Caso os responsáveis da empresa decidam mudar de localização, visualmente, esta informação pode ajudar na procura de um novo escritório, podendo a empresa optar por se instalar numa zona onde se verifique uma maior concentração de colaboradores, originando uma redução dos tempos de deslocação e de gastos com transporte para os mesmos.

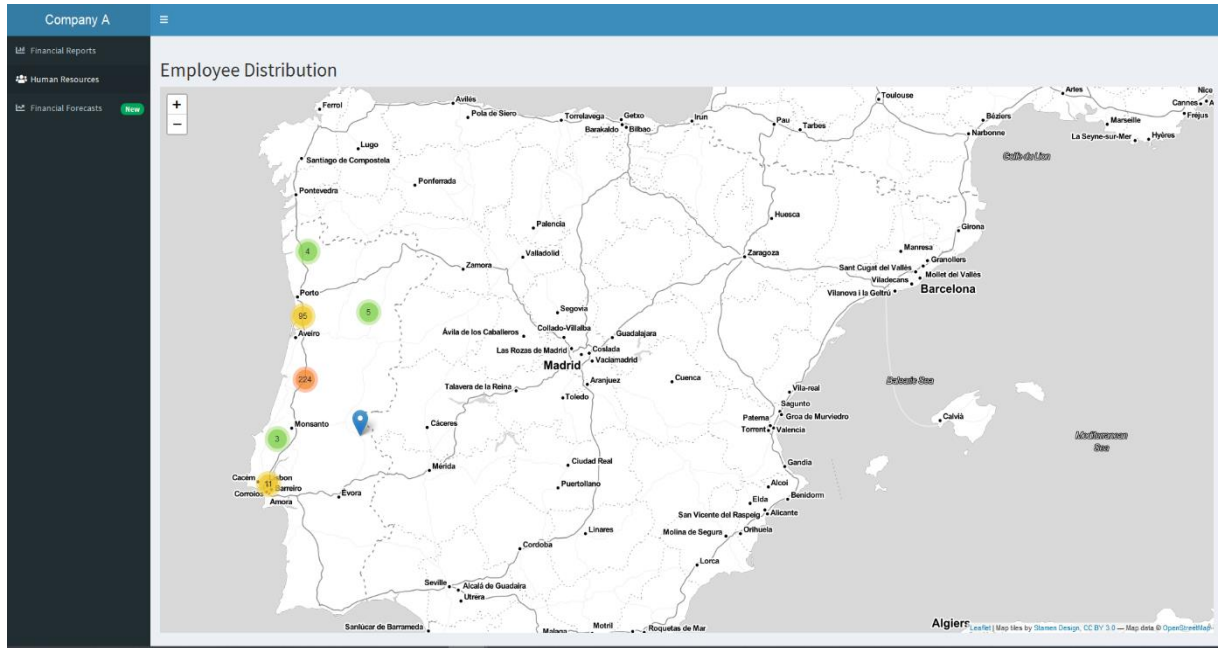


Figura 35 - Distribuição dos colaboradores pelo território nacional

O gráfico da figura 36 permite visualizar a faturação real e prevista por trimestre desde 2010 até 2019, gerada pela RNA. Como se pode verificar, o modelo ajusta-se relativamente bem, prevendo com uma precisão de cerca de 90%. A RNA prevê para o último trimestre de 2019, um volume de faturação na ordem dos dez milhões de euros, revelando-se um ano atípico em comparação com anos transatos. Dado que esta ferramenta de BI funciona em tempo real, a RNA adapta-se às circunstâncias da empresa, pelo que, a informação gerada está em constante atualização.

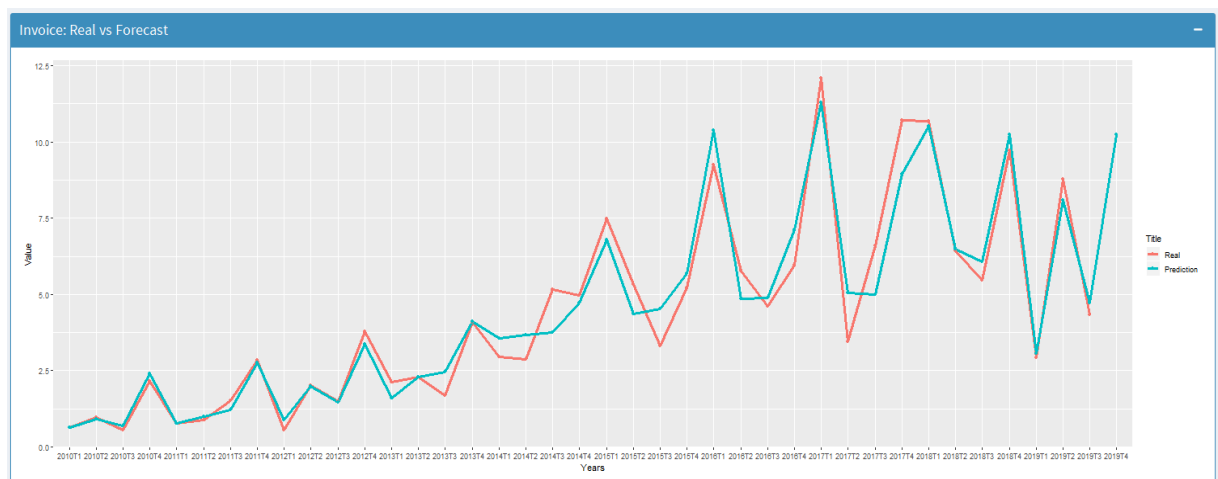


Figura 36 - Faturação da empresa (vermelho) e valor previsto pela rede neuronal (verde).

CONCLUSÃO

Um sistema de BI é um ativo importante nas organizações, para apoiar os gestores e os responsáveis pela gestão de topo a tomar as melhores decisões estratégicas. Os sistemas modernos devem fornecer não apenas as ferramentas para analisar o passado, mas também ajudar o decisor a entender o que provavelmente acontecerá no futuro com base nos dados presentes e passados.

Relativamente a contribuições no decorrer deste projeto, no processo de implementação de uma ferramenta de BI e na avaliação e escolha da plataforma que melhor serviria o presente propósito, foi elaborado um artigo que foi submetido e aceite para a seguinte conferência (ver anexo 1): “*Near Real Time Business Intelligence Framework using R Shiny*” - 19th Portuguese Association for Information Systems Conference (CAPSI'2019), 11 and 12 October 2019, Lisbon, Portugal; encontrando-se também no prelo um capítulo do livro “*Industry 4.0: Principles, Effects and Challenges*”, em coautoria, intitulado “*Business Intelligence Framework with Growth Projection Using Neural Networks*”(Fernandes, Mendes & Almeida, 2020, in press).

A arquitetura e implementação da ferramenta de BI descrita neste projeto, para uma grande empresa que desenvolve produtos de *software* para o setor das telecomunicações, foi proposta para os responsáveis da gestão de topo e para os demais colaboradores que diretamente necessitam de informação em tempo real e que sejam responsáveis por analisar a performance da empresa.

O presente projeto descreve uma solução de BI que importa os dados necessários do ERP-PHC, filtra, trata e processa os mesmos usando o software RStudio e evidencia os resultados num *dashboard* projetado em RShiny. É uma arquitetura simples, com grande potencial e de custos muito reduzidos, sendo de destacar que o RStudio é um software de código aberto. O sistema de análise de informações históricas já está implementado, mostrando em tempo real as receitas e despesas da empresa, bem como a distribuição de recursos humanos. A informação relativa às previsões de faturação para o futuro, usando redes neuronais, está testada e em funcionamento. Como vantagens desta ferramenta, destacamos o baixo custo e de código aberto, sendo possível integrar com outras linguagens de programação que oferecem diferentes pacotes de *software*. A arquitetura proposta pode ser aplicada a qualquer empresa, desde que utilize o mesmo ERP, não sendo necessários intervenções profundas na sua estrutura. Como desvantagens, é importante registar que não há suporte oficial para a linguagem R.

No futuro, espera-se que esta ferramenta de BI seja alargada pelos restantes departamentos da empresa, com maior enfoque nos recursos humanos e logística, com

informações aplicáveis a cada departamento. É expectável que a ferramenta apresente novos separadores com informações relativas aos projetos em vigor na empresa, rácios económicos e financeiros.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abad, F. M. & Chaparro, A. (2016). *Data-mining techniques in detecting factors linked to academic achievement*. <https://www.researchgate.net/publication/309033106> (27 Junho 2019).
- Agrawal, S. & Agrawal, J. (2015). *Neural Network Techniques for Cancer Prediction: A Survey*. 19th International Conference on Knowledge Based and Intelligent Information and Engineering Systems. *Procedia Computer Science* 60. pp. 769 – 774
- Ali, O. T., Nassif, A. B., & Capretz, L. F. (2013). *Business Intelligence Solutions in Healthcare A Case Study: Transforming OLTP system to BI Solution*. <https://www.researchgate.net/publication/259759844> (27 April 2019).
- Al-Maaitah, M. A. (2018). *Impact of Business Intelligence Competencies on the Organizational Capabilities in Jordanian Banks*. *Journal of Computer Science*, Vol. 14, Issue 8, pp. 1144-1154
<https://pdfs.semanticscholar.org/74c3/fddcb070ccc3e2783aef922b551a100a5208.pdf?ga=2.60917292.1869222480.1563882665-184747901.1563091533> (1 Março 2019)
- Barros, R. M. P. C. (2013). *Dashboarding – Projeto e Implementação de Painéis Analíticos*. (Dissertação de Mestrado não editada. Mestrado em Engenharia Informática). Universidade do Minho – Portugal.
- Bisi, B. S., Neto, A. B., & Bonini, C. S. B. (2015). *Redes Neurais Artificiais: Utilização do Algoritmo Retropropagação para Classificação de Grupos em Biosistemas*. XI Fórum Ambiental da Alta Paulista. Vol. 11. n. 2. 15. pp. 345-354.
- Bispo, C. A. F. & Cazarini, E. W. (1998) “A nova geração de sistemas de apoio à decisão”. <https://www.researchgate.net/publication/242619256> (12 fevereiro 2019).
- Blundell, C. Cornebise, J. Kavukcuoglu, K. & Wierstra, D. (2015). *Weight Uncertainty in Neural Networks*. arXiv:1505.05424v2 [stat.ML].
- Carlsson, C. & Walden, P. (2019). *Decision Support Systems - Historical Innovations and Modern Technology Challenges*. THE EWG-DSS 2019 International Conference on Decision Support System Technology. Madeira, Portugal, pp. 229-242.
- Celik, U. & Basarir, Ç., (2017). *The Prediction of Precious Metal Prices via Artificial Neural Network by Using RapidMiner*. *The Journal of Operations Research, Statistics, Econometrics and Management Information*. Volume 5, Issue 1, pp. 46-54.
- Ciaburro, G. & Venkateswaran, B. (2017). *Neural Networks with R*. Published by Packt Publishing Ltd., Livery Place.

- Codd, E. F. Codd, S. B. & Salley, C. T. (1989) *Providing OLAP to User-Analysts: An IT Mandate*. <https://www.semanticscholar.org/paper/Providing-olap-to-user-analysts%3A-an-it-mandate-Codd-Codd/0a93e70589fbeb43edf65de61ffbe6cd3696c4a>. (25 Março 2019).
- Costa, M. & Santos, M. Y. (2012). *Sistema de Business Intelligence no suporte à Gestão Estratégica - Caso prático no comércio de equipamentos eletrónicos*. Actas da 12ª Conferência da Associação Portuguesa de Sistemas de Informação.
- Delen, D. & Ram, S. (2018). *Research challenges and opportunities in business analytics*. Journal of Business Analytics Vol. 1, Issue 1, pp 2-12.
- Eswaran, A. (2017). *Competing in 2020: Winners and Losers in the Digital Economy*. A Harvard Business Review Analytic Services Report, pp. 1-16.
- Falez, P., Tirillyb, P., Bilasco, I. M., Devienne, P. & Boulet, P. (2019). *Unsupervised Visual Feature Learning with Spike-timing-dependent Plasticity: How Far are we from Traditional Feature Learning Approaches?*
- Fernandes, B., Mendes, M. & Almeida, J.A. (2019). *Near Real Time Business Intelligence Framework using R Shiny*. 19th Portuguese Association for Information Systems Conference (CAPSI'2019).
- Fernandes, B., Mendes, M. & Almeida, J.A. (2020). *Business Intelligence Framework with Growth Projection Using Neural Networks*. In Ing. Yilmaz Uygun, "Industry 4.0: Principles, Effects and Challenges". New York: Nova Science Publishers, Inc. In press.
- Fortulan, M. & Filho, E. (2005). *Uma Proposta de Aplicação de Business Intelligence no Chão de Fábrica*. Gestão & Produção. v.12. n.1. p.55-66.
- Friedman, T. (2002). *How Secure Is Your Business Intelligence Environment?* <https://www.gartner.com/en/documents/367264> . (2 Fevereiro 2019).
- Gangadharan, G. R. & Sundaravalli N. S., (2004) "*Business intelligence systems: design and implementation strategies*". 26th International Conference on Information Technology Interfaces ITI. June.
- Goodfellow, I., Bengio, Y. & Courville, A. (2016) *Deep Learning*. MIT Press.
- Gray, P. & Watson, H. J. (1996). *The New DSS: Data Warehouses, OLAP, MDD, and KDD*. Americas Conference on Information Systems.
- Gunther, F. & Fritsch, S. (2019). *Neuralnet: Training of Neural Networks*. The R Journal. Vol. 2, Issue 1, pp. 30-38

- Guohua. W. & Francis. T. E. H. (2000). *Data Mining: Concepts, Applications and Techniques*. ASEAN Journal on Science and Technology for Development. Vol. 17, Issue 1, pp. 77-86.
- Hamdi, M., Aloui, C. & Nanda, S. K., (2016) "*Comparing Functional Link Artificial Neural Network And Multilayer Feedforward Neural Network Model To Forecast Crude Oil Prices*". Economics Bulletin, Vol. 36, Issue 4, pp. 2430-2442.
- Han, J., Kamber, M. & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers is an imprint of Elsevier.
- Howson, C., Richardson, J., Sallam, R. & Kronz, A. (2019). Magic Quadrants for Analytics and Business Intelligence Platforms, <https://www.gartner.com/doc/reprints?id=1-68720FP&ct=190213&st=sb> (10 July 2019).
- Huang, H. (2018). *Statistical mechanics of unsupervised feature learning in a restricted Boltzmann machine with binary synapses*. <https://arxiv.org/pdf/1612.01717.pdf>
- Jordan. H. das Neves. J. C. & Rodrigues. J. A. (2008). *O Controlo de Gestão* (8.^a ed.). Lisboa: Áreas.
- Keenan, P. (2019). *DSS and BI: A bibliographic analysis of their links*. THE EWG-DSS 2019 International Conference on Decision Support System Technology. Madeira, Portugal, pp. 53-58.
- Kimball, R. & Ross, M. (2013) *The Data Warehouse Toolkit: the complete guide to dimensional modeling*. 3 ed. John Wiley and Sons. Inc.
- Konrath. A. C., Silva. S. A., Henning. E., Santos. L. M., Mirandas. R. G.. & Samohyl. R. W. (2018). *Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de Estatística*. Abakós Vol. 6, Issue 2, pp. 55-71.
- Lazzarin, L. N. A., Belotti, J. T., Silva, L. V., Marinho, M. H. N., Alves, T. A., Tadano, Y. S. & Siqueira, H. V. (2018). *Redes Neurais Feedforward Aplicadas na Avaliação do Impacto da Poluição Atmosférica e Variáveis Climáticas na Saúde Humana*. 11th National Congress on Experimental Mechanics.
- Linden, I. (2019). *30 years Business Intelligence: from Data Analytics to Big Data*. THE EWG-DSS 2019 International Conference on Decision Support System Technology. Madeira, Portugal, pp 243-253.
- Luckow, A., Matthew, C., Ashcraft, N., Weill, E., Djerekarov, E. & Vorster, B. (2017). *Deep Learning in the Automotive Industry: Applications and Tools*. <https://www.researchgate.net/publication/313458531>. (02 Abril 2019)

- Luhn. H. P. (1958) "A *business intelligence system*." IBM Journal of Research and Development. vol. 2. nº 4. pp. 314-319.
- Mussa. M. S., Souza. S. C., Freire. E. F. S., Cordeiro. R. G. & Hora. H. R. M. (2018). *Business Intelligence in Education: An application of Pentaho Software*. Revista Produção e Desenvolvimento Vol. 4, Issue 3, pp. 29-41.
- Negash. S., (2004) "*Business Intelligence*" in Communications of the Association for Information Systems. vol. 13. pp. 177-195.
- Newton. P. N. & Reader. S. S. (2013). Data Mining in Decision Support System. <https://pdfs.semanticscholar.org/45ea/05c2075d8ca2ccc0ed0a46b4416a2125c22e.pdf> (8 May 2019).
- Nwufo. O. C., Okwu. M. O., Nwaiwu. C. F., Igbokwe. J. O., Nwafor. O. M. I. & Anyanwu. E. E. (2017). The Application of Artificial Neural Network in Prediction of the Performance of Spark Ignition Engine Running on Ethanol-Petrol Blends. International Journal of Engineering and Technologies (12). pp 15-31.
- Olszak. C. & E. Ziemia. (2007) "*Approach to Building and Implementing Business Intelligence Systems*." Interdisciplinary Journal of Information. Knowledge. and Management vol. 2. pp. 135-148
- Power, D. J. (2007). *Uma Breve História dos Sistemas de Apoio à Decisão*. <http://dssresources.com/history/dsshhistory.html> (24 Abril 2019).
- Quintela, H. (2005). *Sistemas de conhecimento baseados em data mining : aplicação à análise da estabilidade de estruturas metálicas*. Dissertação de Mestrado em Sistemas de Informação (não editada), Universidade do Minho.
- Rao, B. J., Babu, M. S. P. & Sastry, S. H. (2019). *A New Methodology to Perform Big Data Analytics on Business Warehouse Data*. International Journal of Innovative Research in Science, Engineering and Technology. Vol. 8, Issue 1, pp. 432-443.
- Rasmussen. N. H., Goldy. P. S. & Solli. P. O. (2002). Financial Business Intelligence. John Wiley & Sons, Inc.
- Sang. G. M., Xu. L. & Vrieze. P. (2016). Implementing a Business Intelligence System for small and medium-sized enterprises. <https://www.researchgate.net/publication/313037031> (1 May 2019)
- Santos. M. Y. & Ramos, I. (2006). *Business Intelligence: Tecnologias da Informação na Gestão de Conhecimento*. FCA – Editora de Informática. (LIVRO)
- Scaff. V. L. R. & Almeida. D. (2005). "*Sistemas de Informação como ferramenta de apoio à decisão na logística*". XII SIMPEP – Bauru. SP. Brasil. 7 a 9 de Novembro.

- Sezões. C., Oliveira. J. & Baptista. M. (2006). *Business Intelligence*. Porto: Príncipe
- Shim, J. P., Warkentin, M., Courtney, J. F., Daniel, J. P., Sharda, R. & Carlsson, C. (2002). *Past present and future of decision support technology*. Published by Elsevier Science B.V.
- Turban. E., Sharda. R. & Delen. D. (2010). *Decision Support and Business Intelligence Systems*. Prentice Hall.
- Vieira, A. S. & Silva, P. C. (2018) *Neural Networks For Financial Domain: A Bibliographic Review*. 15th International Conference on Information Systems & Technology Management - CONTECSI – 2018 pp. 4132-4146.
- Zeng. L., Xu. L. Shi. Z. Wang. M. & Wu. W. (2006) "*Techniques. process. and enterprise solutions of business intelligence.*" Conference on Systems. Man. and Cybernetics October 8-11. Taipei. Taiwan. Vol. 6. pp. 4722.

ANEXOS

**ANEXO 1 – ARTIGO “NEAR REAL TIME BUSINESS INTELLIGENCE
FRAMEWORK USING R SHINY”**

Near Real Time Business Intelligence Framework using R Shiny

Bruno Fernandes, Polytechnic Institute of Coimbra – ISEC, Portugal, a21270987@alunos.isec.pt
Mateus Mendes, Polytechnic Institute of Coimbra – ESTGOH and Institute of Systems and Robotics
of the University of Coimbra, Portugal, mmendes@isr.uc.pt
Jorge Alexandre Almeida, Polytechnic Institute of Coimbra – ISEC, Portugal, jalmeida@isec.pt

Abstract

Modern industry deals with large amounts of data, which is often difficult for humans to process and use for decision making. Industry 4.0 proposes the automation of different procedures in enterprises, aiming to reduce human errors, operation time and costs. That includes analysis of different operation parameters in near real time, in order to facilitate management to make the right decisions at the right time. That requires the use of tools that are simple and fast to use and provide the necessary information. The present paper describes an architecture of a Business Intelligence system proposed for a Telecommunications software company. The system draws information from a proprietary ERP and is all developed using free open source software. The architecture proposed uses the power of R for statistical computing, data mining and artificial intelligence. Financial information is shown in a dashboard in near real time.

Keywords: Business Intelligence, Financial Information Dashboard, Data Mining, Neural Networks

1. INTRODUCTION

Nowadays, companies are faced with enormous amounts of information, produced internally and externally at a fast pace. That information is derived from their internal processes, such as production, accounting, logistics, commercial, and financial area, as well as their external relationship with customers, suppliers and banking entities. Storing and analyzing data derived from those relationships, companies know that they have information vital to their survival in the market, that can bring them competitive advantages. Most of that information is supported by enterprise management systems, called ERP (Enterprise Resource Planning). The ERP's main task is to consolidate company information into a single database, combining the functions of production control, projects, inventory, purchases, human resources, billing and treasury, among others, and thus eliminating the difficulty of dealing with data stored in various systems.

The amount of information stored over time can be very large, even for small enterprises. Many companies still process the data using spreadsheets, but that quickly becomes impractical for big companies or old companies which want to process data stored over several years. Therefore, the results obtained are only part of what could be discovered using more powerful tools. To overcome the difficulty and extract a good amount of knowledge about past events, there are more powerful technologies and tools that form the basis of Business Intelligence (BI) systems. Those include tools of Data Warehouse (DW), On-Line Analytical Processing (OLAP), and Data Mining (DM) (Shim et al., 2002).

1.1. Data Warehouse

In the last few years, databases have been improved for operational and analytical data handling, with users tending to the operational level, even though they also need analytical tools. That is different from operational databases. DWs are designed to store gigabytes and even terabytes of information and can sometimes be divided into smaller logical drives called Data Marts (DM). According to Kimball & Ross (2013), DM are DW but contain data from a particular area or department of the organization. Data warehouses are considered as key tools of the BI infrastructure. To be effectively implemented, they need to be embedded in a complete BI architecture. Data warehouses have a solution for data access, reconciliation problems, and quality. To achieve this goal and power a data warehouse, significant work must be done by a process called ETL (Extract-Transform-Load) commonly called (Linden, 2019).

1.2. OLAP

According to Zeng, Xu, Shi, Wang & Wu (2006), OLAP refers to the techniques of execution and complex analysis of information, stored in a DW or not. Those techniques transform raw data into information, which is more easily perceived, analyzing the various dimensions and combinations, aiming to recognize trends in the data. Costa & Santos (2012) affirm that OLAP systems are used to analyze DW. They allow the creation of cubes analyzing the information generated from several angles. After the creation of cubes, those cubes represent integrated data with some measure of interest, so they can have several dimensions to analyze and represent a particular business.

1.3. Data Mining

Data mining is the action of extracting information from a database, using sophisticated analysis techniques such as those coming from Artificial Intelligence, Statistics and Mathematics, in order to produce valuable knowledge used in the strategic management of an organization, according to Bispo & Cazarini (1998) and Turban et al. (2010). Applying the most appropriate Data Mining techniques to achieve the goals defined is considered an information modelling phase. The main techniques used are decision trees, artificial neural networks and rules of induction. The parameters of each model are adjusted in order to optimize the results (Bispo & Cazarini, 1998).

Neuronal networks are modified regression models that make approximations, according to Abad & Chaparro (2016). They are networks trained using inductive reasoning, with a set of input and output data. They consist of nodes and connections, where each connection has a weight which is adjusted during the training phase. They have the advantage of being able to abstract an interpretation of complex data in order to represent patterns or trends. The rules of induction are an exploratory method, with the purpose of finding models or categories of data that are grouped in a natural way.

1.4. Structure of the paper

Section 2 describes the related work. Section 3 describes BI model proposed. Section 4 shows results produced by the system. Sections 5 and 6 draw some conclusions and perspectives of future work.

2. RELATED WORK

In a given period of time, a company generates a large amount of data related to its production, billing, stocks, employees, materials, products, etc. The raw data, will have little utility to withdraw any conclusion or extract any knowledge.

According to Mussa, Souza, Freire, Cordeiro and Hora (2018), a BI tool has as main objective to provide information in the right way and in the right time so that the company can apply the best decisions in a faster and more reliable way. It is a set of tools and applications that allow decision makers to organize, analyze, distribute and thus act on information relevant to the company's business. BI systems at a strategic level enable to accurately define objectives and monitor the achievement of such objectives, allowing to make comparative reports on an organization's profitability or the effectiveness of the distribution channels by simulating or predicting future results. At a tactical level they offer a basis for optimizing marketing, sales and financial decisions in order to help achieve strategic objectives in a more effective way. At the operational level, BI systems are used to perform *ad-hoc* analysis and answer questions related to the current operations of the organization's activity, typically focusing on suppliers and customers (Olszak et al. 2007).

Nowadays, BI systems are being regarded as a powerful set of tools which allow to create value for a company. In that context, the present paper aims to briefly summarize some architectures proposed by different authors for different activities, with diverse goals and using various tools. According to Zeng et al. (2006), more and more business sectors are deploying advanced BI solutions to increase their competitiveness, due to their importance for the efficient functioning of the organizations in managing their resources and developing the businesses. Selecting and implementing a BI tool can be a time-consuming and difficult task. There are now producers that offer a wide range of solutions, from low-tech products that offer just simple reporting, to sophisticated platforms. Some platforms are open source, others are proprietary software.

According to Howson, Richardson, Sallam and Kronz (2019), modern (BI) platforms are characterized by easy-to-use tools for flowing from a complete analytical workflow, from data preparation to visual appearance. They stand out from traditional BI platforms, not requiring significant involvement of IT teams to predefine data storage models.

A comparison of the classification of BI platforms is presented in Figure 1. The figure plots the platform's ability to execute as a function of the completeness of vision. Platforms like Tableau, Microsoft and other emerging vendors offer a restricted set of features, but are widely used for a variety of BI functions and custom reports. In general, they are easy to use and deploy.

The market leader, Microsoft, offers data preparation, visual-based data discovery, interactive dashboards and enhanced analysis through a single product, Power BI. Power BI Desktop can be used as a standalone, free self-service analysis tool. Power BI can be used by people with minimal skills in data science, among many others. Important script items include support for existing Reporting Services, a common data model, and open data preparation with data streams.

BI models are present in all sectors of activity. This article intends not only to deal with models of BI implementations in software companies, but also to identify some BI implementations in the most diverse sectors.

A BI system is never implemented by a single technology, product, or vendor. There are several models and implementations of BI systems. Costa and Santos (2012) propose a model that allows an organization to obtain information that supports decision making at all levels of planning, whether strategic, tactical or operational. By leveraging the entire information history of its database, the BI solution allows the organization to access information on purchases and sales of products,

accounting and financial information, as well as current accounts of suppliers. Using data from the ERP, it stores the necessary information in a local database management system. The author-defined architecture consists of a first server with its ERP database, a second server with the Data Warehouse, the ETL process and the OLAP and reporting servers. Report analysis is supported by a web environment and a spreadsheet software. Because of the complexity of the database, the authors chose to create views of the tables considered essential so as not to overload the BI system. The Data Warehouse environment, besides allowing the multidimensional modelling of the data, may allow the introduction of Data Mining techniques. The authors implemented three interface scenarios, i) using a front end reporting solution; ii) using a spreadsheet software; and iii) using the Business Intelligence tool Pentaho.



Figure 1 - Magic Quadrant for Analytics and Business Intelligence Platforms according to Howson et. al (2019)

The authors, Christian, Akujobi, Saha and Korzaan (2018) report the implementation of a business intelligence and analytics system in a software production company for the gaming industry using Tableau. The first step was to clean and migrate the data from the different databases, using R programming for this integration. The software was used in this case to aid in the analysis of statistics and repeater data of complex algorithm patterns. The second step consisted in designing the data model and defining its visualization using Tableau.

Another model of BI implementation is proposed by Ali, Nassif and Capretz (2013), more specifically in the area of healthcare. With a significant increase in infection rates, entailing very high costs to the organization, the intention of Ali *et al.* was to define and recommend an appropriate data and reporting management structure to support the decision-making strategy. The authors argue that implementing a robust framework to obtain, manage and report data and information is a key success factor in those projects, with the aim of reducing the number of infections and the associated costs.

The BI implementation project uses the OLTP (On-Line Transaction Processing) tools responsible for handling routine data that is generated daily through the company's information systems. It also uses analytics tools for decision support, allowing to solve problems in a simple and economical way through a straightforward and integrated view of the data. The authors defined the following development steps: 1) Design the database; 2) Define the ETL process; 3) With the desired data, the structure of the OLAP cubes was drawn, creating and displaying the data source, associated dimensions, the multidimensional cube and the KPIs (Key Performance Indicators) to display monthly metrics; 4) Define and construct information reports; and 5) Add Data Mining components. With healthcare organizations experiencing difficulties in extracting valuable information from their data, the tool allows them to increase strategic thinking through modifications to the classic information tools for BI solutions.

Artificial neural networks have already been used data mining processes and BI solutions, and numerous examples can be found in the literature. Nwuf *et al.* (2017) defined a neural network model to analyze the performance of a combustion engine. The model was designed to predict fuel consumption, maximum pressure achieved and engine efficiency by using the engine speed register, fuel ratio and exhaust gas temperature as input. They found that the neural network model provided accurate information for the analysis of complex equipment and was considered useful for predicting the performance of combustion engines. Rodrigues *et al.* (2019) use artificial neural networks to model the condition of diesel engines' oil. Based on twenty one variables that represent the oil condition, the authors conclude that the ANN can learn and make decisions similar to human experts and, therefore, the process can be automated.

3. DESCRIPTION OF THE BI MODEL PROPOSED

The present paper describes a BI solution implemented in a Portuguese company that operates in the field of software development for the mobile telecommunications and mobile internet sectors. The company has verified over the years a remarkable growth in revenue, number of employees and facilities. It is classified as a big company, according to the classes defined in the Portuguese Decree Law no. 158/2009 of July 13, which approves the Portuguese accounting standardization system. It has continually firmed new partnerships and penetrated new markets, namely with telecommunications companies with a strong presence in the national and European markets. The focus of the company will be to maintain this trend of evolution, conquering emerging markets in the telecommunications sector. Due to its continually growing size, and the projected growth, a support system at the decision-making level, through the implementation of a BI system, is appropriate and necessary. In order for managers to simplify the processes when making strategic management, the company aims to suppress the spreadsheets used to analyse cost and performance indicators, with sometimes extensive information. To address these constraints, the company needs a system that gives historical and current data information available in the ERP database, and which also makes projections for the future and thus assists the management in decision making.

The development of the BI solution proposed started with meetings with the management of the company, as well as the technical department, in order to understand the requirements of the BI model aimed by the management and the technical infrastructure already in place.

The company currently uses an ERP system as an integrated management system, including all accounting, financial and human resources information. The management showed interest in a BI

solution that would show, with a time gap as small as possible, the evolution and trends of revenue and expenses, per project, company department and in total.

The goal is therefore to extract the necessary billing information, market positioning, expenses with employees and total expenses, from the ERP, to use it for the BI system for further analysis and calculation of the economic and financial indicators. The architecture of the solution includes open source software that allows to extract and transform the information, being prepared to receive artificial intelligence algorithms, namely neural networks, for making the projections of future trends. Sang, Xu and Vrieze (2016), and the Free Software Foundation¹, define Open Source Software (OSS) as the software that can be freely used, modified and distributed by anyone. Therefore, OSS can always be improved, so it is the preferred choice for the present system.

The presentation of the information allows the visualization of the data in real time. It is generated through the same software where the user interface is created as dashboards, containing tables, graphs and iterative tools of multidimensional analysis that allow to evaluate the current state of the business. The BI tool was expected to provide treatment and analysis of the information considered relevant for those responsible for top management, in order to help them make a more efficient and effective decision making.

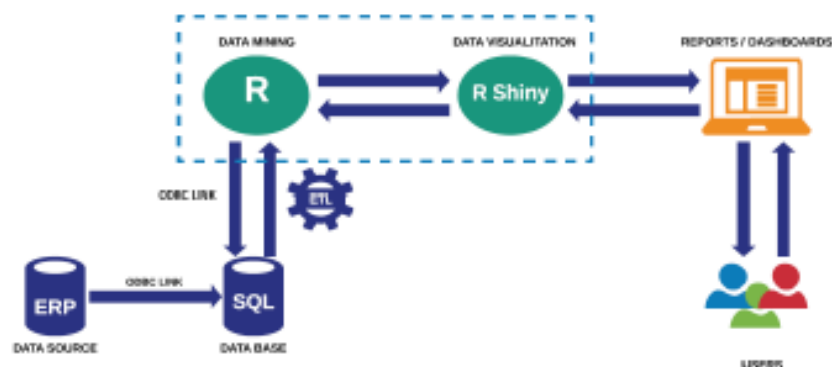


Figure 2 - Architecture of the Business Intelligence system developed

After the initial phase in defining the company's requirements, the decision was to make a copy of the relevant information from the ERP database to a separate database, where data could be processed using Rstudio. RStudio software is an open source tool, which meets all the requirements intended for the system. R is a programming language and computational environment, specialized in data manipulation, statistical analysis and graphical visualization. It is available for different operating systems. For user interface, RStudio software was chosen. It has a package name RShiny, which facilitates the creation of interactive web applications that can be used to display data in an interactive way with innovative visualization features. Figure 2 shows the architecture of the system, highlighting the components and technologies used for its construction.

The above mentioned architecture comprises a server that stores information copied from the ERP. Copies of the data from the ERP to the server are done automatically every five minutes, through an ODBC link. This method ensures that the ERP information is available in near real-time for the BI system, while the BI data and the ERP data are separated for integrity reasons. The BI system has read-only access to the ERP database, while the ERP has no access to the BI database.

¹ <http://www.fsf.org> (last visited 2019-05-09).

RStudio has then access to the BI database through another ODBC connection, which grants read and write access, as shown in Figure 2.

After the dataset is extracted from the database, it is processed in R and sent to R Shiny. R Shiny prepares the processed data for visualization in the dashboard. The dashboard offers different visualization options to the user, so communications between R, R Shiny and the user are bi-directional.

Figure 3 shows the data mining process in more detail. Selected data are extracted from the database and fed to an R module responsible for applying the necessary statistical and artificial intelligence algorithms. This module is where the financial parameters are calculated, based on the information retrieved from the database, and the charts are produced. The module is also prepared to include a feed forward neural network, namely supervised predictive models, available in R. The neural network planned is three-layered model. The number of neurons will be determined experimentally in future work. The network will be trained and used to project the future growth of the company's billing. In the process of handling historical billing data, users select the year and billing grouped by quarter within each year. This way the network is expected to provide a good result for the forecast of billing for the next quarters. The output of this module is then forwarded to the next module, which prepares the output and user interface. The user interface was designed in R Shiny, as stated previously.

The user interface was designed in the form of a dashboard, accessed through a web browser. A dashboard is a display area that can contain various elements such as tables, graphs, indicators or maps. It is a visual representation that makes it possible to display the performance of the organization, according to the latest available data. Barros (2013) claims the dashboards assume a fundamental importance in the supervision of the organization, supporting the decision making process. They should be organized graphically, and should have graphical views appropriate to the information that is intended to display. In the construction of web applications and according to Konrath et al. (2018), Shiny presents two scripts that are written to a directory of the R environment. The first script, ui.R, refers to the User Interface (UI). The UI has the function of controlling the layout and the appearance of the dashboard. The second script is the server and is named server.R.

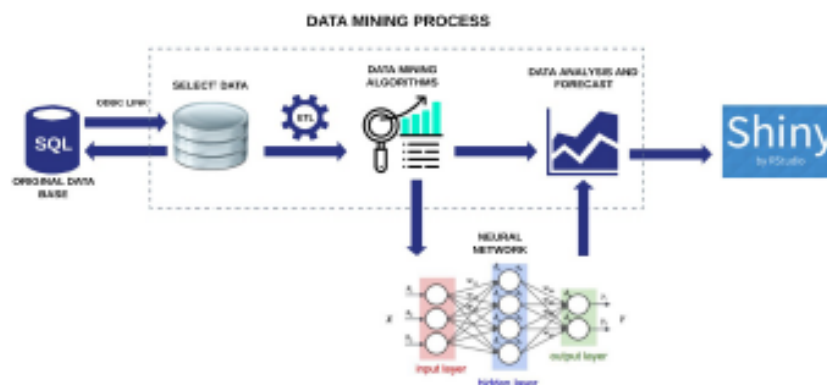


Figure 3 - Details of the data mining process

4. OUTPUTS OF THE BI SYSTEM

Figure 4 shows a screenshot of an example of a dashboard produced by the BI system proposed. One of the objectives of the prototype is to provide a BI tool that makes it possible to analyze data in an easy and intuitive way. As shown in the figure, in the first tab there are charts that display the evolution of billing per year, by market, personnel expenses and revenues over the years. There is a second tab that shows employee distribution, facilitating the analysis of geo-referenced data. The third tab will show forecasts, produced using AI algorithms. The AI algorithms are presently being implemented. They will allow to make projections for the future.



Figure 4 - Example of the dashboard produced by the BI system, showing financial reports

The first tab, called “Financial Reports”, shows on top the situation of the company at a glance, regarding income and expenses. The green boxes show the invoicing and revenue for the current year, compared to the all time highest invoicing and revenue. The yellow boxes show the costs of employees and the total costs for the current year, compared to the previous year. The second row shows the evolution of invoicing and revenue, and their distribution per market.

The second tab shows a map of the distribution of the company’s human resources, as depicted in Figure 5. This visual information allows, if the company decides to change offices, define the new location according to the largest number of employees in a given area, reducing travel expenses. Currently the largest workforce is in Coimbra, with the second largest in Lisbon and a small representation in the north of Portugal.



Figure 5 - Example of the dashboard produced by the BI system, showing employee distribution

5. MAIN CONTRIBUTION AND DISCUSSION

The BI tool described has been proposed to financial and administration users for validation. In the future, it is expected that it will be used by the department of human resources and operations with information applicable to each department. As advantages of this tool, one is the use of open source software, being possible to integrate other programming languages, which offer different software packages. The architecture proposed can be applied to any company, as long as it uses the same ERP or the necessary changes are implemented. The neural models can be trained with existing data, as long as the format is preserved. As disadvantages, it should be noted there is no official support for the R language and R Shiny.

6. CONCLUSION AND FUTURE WORK

A Business Intelligence system is an important asset in organizations, to help managers make the best strategical decisions. Modern systems should provide not only the tools to analyze the past, but also help management understand what will probably happen in the future based on the present and past data.

The present paper describes a BI solution, proposed for a big telecommunications software company. The solution proposed imports the relevant data from the ERP, processes the data using R Studio and shows the results in a dashboard designed in RShiny. It is a simple, powerful and inexpensive architecture. The system for analyzing past information is already implemented, showing the company income and expenses at a glance, as well as the distribution of human resources. The module for making projections for the future, using neural networks, is currently under development.

Author contributions: Bruno Fernandes developed the software and outlined the first draft of the paper. Jorge Almeida and Mateus Mendes proposed the architecture of the system, supervised the project and contributed equally to the paper.

REFERENCES

- Abad, F. M., & Chaparro, A. (2016). Data-mining techniques in detecting factors linked to academic achievement, <https://www.researchgate.net/publication/309033106> (21 April 2019).
- Ali, O. T., Nassif, A. B., & Capretz, L. F. (2013). Business Intelligence Solutions in Healthcare A Case Study: Transforming OLTP system to BI Solution, <https://www.researchgate.net/publication/259759844> (27 April 2019).
- Barros, R. M. P. C. (2013). Dashboarding – Projeto e Implementação de Painéis Analíticos. (Dissertação de Mestrado não editada, Mestrado em Engenharia Informática). Universidade do Minho – Portugal.
- Bispo, C. A. F., & Cazarini, E. W. (1998). A nova geração de sistemas de apoio à decisão, <https://www.researchgate.net/publication/242619256> (30 April 2019).
- Costa, M., & Santos, M. Y. (2012). Sistema de Business Intelligence no suporte à Gestão Estratégica - Caso prático no comércio de equipamentos eletrónicos. Actas da 12ª Conferência da Associação Portuguesa de Sistemas de Informação.
- Christian, N., Akujobi, I., Saha, M., Korzaan, M. (2018). ACS: Bringing Business Intelligence and Analytics to a Massive Multiplayer Online Gaming Company. *Information Systems Education Journal (ISEDJ)*, 16(2), pp. 11-17.
- Decreto Lei n.º 158/2009 de 13 de julho. Diário da República n.º 133/2009, Série I de 2009-07-1
- Howson, C., Richardson, J., Sallam, R., Kronz, A. (2019). Magic Quadrants for Analytics and Business Intelligence Platforms, <https://www.gartner.com/doc/reprints?id=1-68720FP&ct=190213&st=sb> (10 July 2019).
- Kimbal, R., & Ross, M. (2013). The Data Warehouse Toolkit: the complete guide to dimensional modeling. (3 ed.), John Wiley and Sons, Inc.
- Konrath, A. C., Silva, S. A., Henning, E., Santos, L. M., Mirandas, R. G., & Samohyl, R. W. (2018). Desenvolvimento de Aplicativos Web Com R e Shiny: inovações no ensino de Estatística. *Abakós* 6 (2) pp. 55-71.
- Linden, I. (2019) 30 years Business Intelligence : from Data Analytics to Big Data. 5th *International Conference on Decision Support System Technology - ICDSST 2019*, University of Madeira, Portugal, pp. 243-251.
- Mussa, M. S., Souza, S. C., Freire, E. F. S., Cordeiro, R. G., Hora, H. R. M., (2018). Business Intelligence in Education: An application of Pentaho Software. *Revista Produção e Desenvolvimento* 4, (3), pp. 29-41.
- Nwufo, O. C., Okwu, M. O., Nwaiwu, C. F., Igbokwe, J. O., Nwafor, O. M. I., Anyanwu, E. E. (2017). The Application of Artificial Neural Network in Prediction of the Performance of Spark Ignition Engine Running on Ethanol-Petrol Blends. *International Journal of Engineering and Technologies* (12), pp 15-31.
- Olszak, C. M., & Ziemba, E. (2007). Approach to Building and Implementing Business Intelligence Systems, <https://www.researchgate.net/publication/255612097> (21 April 2019).
- Rodrigues, J., & Costa, I., & Farinha, J. T., & Mendes, M., & Margalho, L. (2019). Modelling Diesel Engine Oil Condition Using Artificial Neural Networks, eMaintenance 2019, Stockholm, Sweden, May 2019.
- Sang, G. M., Xu, L., & Vrieze, P. (2016). Implementing a Business Intelligence System for small and medium-sized enterprises, <https://www.researchgate.net/publication/313037031> (1 May 2019).

- Shim, J.P., Warkentin, M., Courtney, J.F., Power, D.J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems* (931), pp. 1-16.
- Turban, E., Sharda, R., & Delen, D. (2010). *Decision Support and Business Intelligence Systems* (9 ed.), Prentice Hall.
- Zeng, L., Xu, L., Shi, Z., Wang, M., & Wu, W. (2006). Techniques, process, and enterprise solutions of business intelligence. *IEEE International Conference on Systems, Man, and Cybernetics* (6), pp. 4722-4726.