

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO

PROPOSTA DE UM MODELO DE BUSINESS INTELLIGENCE PARA O
APOIO À DECISÃO ATRAVÉS DA PERSPECTIVA DA DATA SCIENCE

HEBERT WESLEY PEREIRA ZARONI

ITAJUBÁ

2019

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO

Hebert Wesley Pereira Zaroni

PROPOSTA DE UM MODELO DE BUSINESS INTELLIGENCE PARA O
APOIO À DECISÃO ATRAVÉS DA PERSPECTIVA DA DATA SCIENCE

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para obtenção do título de Mestre em Engenharia de Produção.

Orientador: Prof. Dr. Alexandre Ferreira De Pinho

ITAJUBÁ

2019

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO

Hebert Wesley Pereira Zaroni

Dissertação aprovada por banca examinadora em 08 de Maio de 2019, conferindo ao autor o título de **Mestre em Ciências em Engenharia de Produção.**

Banca Examinadora:

Prof. Dr. Alexandre Ferreira de Pinho (Orientador)

Prof. Dr. Fábio Favaretto (Universidade Federal de Itajubá)

Prof. Dr. André Pimenta Freire (Universidade Federal de Lavras)

ITAJUBÁ

2019

DEDICATÓRIA

À Deus, a Santa Igreja Católica, em
especial aos meus pais Edson e Lídia,
à vó Iolanda, parentes e amigos.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, razão do meu viver, por todas as graças derramadas em minha vida. Para Ele, toda as glórias deste trabalho. Agradeço à Nossa Senhora, Maria minha Mãe do céu, seus anjos e aos Santos, por toda a intercessão, cuidado e bênçãos.

Em especial, agradeço aos meus pais, Edson e Lídia, por toda dedicação, força e empenho que colocaram em mim, mesmo nos momentos mais difíceis. A eles, todo amor do mundo o qual é impossível agradecer, com totalidade, em palavras. Agradeço minha avó Iolanda, por todo carinho e orações direcionadas a mim, além de todos os familiares, pelas preocupações e zelo para comigo. Aproveito para agradecer à minha namorada Diovana, não apenas pelas revisões no trabalho, mas pelo constante apoio e carinho, para o bom êxito dessa jornada.

Agradeço a todos os amigos (em especial aos mais próximos, irmãos que meus pais não deram). Em particular, agradeço a todos que me apoiaram a seguir com a pós-graduação e aos que, de alguma maneira, contribuíram para esse trabalho ser desenvolvido e finalizado (necessito citar alguns nomes: Renata Miranda, Kelly Lobato, Diego Carvalho e Letícia Maciel).

Meu muito obrigado ao meu orientador, Alexandre Pinho, por toda colaboração, por nortear não somente este trabalho, mas também por incentivar-me a melhorar a cada dia, e pela atenção dada durante todos esses anos de pesquisa e desenvolvimento, sem a qual, este trabalho não existiria.

Meus agradecimentos aos amigos e colegas de trabalho da Diretoria de Contabilidade e Finanças da UNIFEI, pela paciência, colaboração e risadas, por fazer do trabalho cotidiano e rotineiro mais leve e menos estressante, além da força dada no Mestrado com a compreensão das ausências e pelos incentivos recorrentes.

Por fim, agradeço a Universidade Federal de Itajubá como um todo, docentes (sobretudo aos que me apoiaram e serviram de exemplo para mim), servidores e colegas discentes que me proporcionaram, proporcionam e proporcionarão momentos verdadeiramente felizes e um ambiente propício a formação humana e intelectual. És para mim, mais que uma instituição.

"O mais profundo dos desejos por conhecimento é o desejo de conhecer o propósito do mundo e de nós mesmos."

(G. K. Chesterton)

RESUMO

A tomada de decisões pelos gestores é um assunto recorrente nas organizações e empresas nos dias atuais, sendo estas privadas ou de gestão pública. Esta última mantém um sistema burocrático, com difícil acesso a informações rápidas e precisas. Com a quantidade massiva de dados disponível dentro das organizações, e também provindos de fontes externas, estas têm buscado novas tecnologias e métodos em meio aos Sistemas de Informação para obtenção de informações com mais qualidade. Os sistemas de *Business Intelligence* (BI) são um dos meios que contribuem para reunião, análise e propagação de dados, resultando em diversos produtos e relatórios administrativos que facilitam a tomada de decisões; e a *Data Science* (DS), ou ciência dos dados, a qual é um campo emergente em meio aos Sistemas de Informação, carrega as características de transformação e análise de dados de forma que ajude a organização, também, no processo decisório. Ambos os conceitos possuem seus métodos, processos, modelos e ciclos de vida para chegarem em um objetivo em comum. Entretanto, na literatura, há uma carência de modelos que agreguem os dois conceitos de forma concomitante, ou aplicando conceitos de um campo no outro. A partir desse ponto de defasagem, o objetivo dessa pesquisa é propor um modelo que aplique os conceitos de BI e DS, colocando os conceitos de cada um em conjunto, conceitualizando-os e identificando os pontos em que são convergentes e divergentes para se ter um modelo eficiente. A metodologia utilizada é a de Modelagem, para desenvolver o modelo proposto, o qual passa por sua Conceitualização, Modelagem, e Solução e Implementação, onde os conceitos e passos dos processos de BI e DS são explorados e colocados conjuntamente, com seus ciclos e fases. Por fim, o modelo desenvolvido foi aplicado em uma ferramenta computacional que possa incorporá-lo, como forma de testá-lo e validá-lo, gerando produtos computacionais para serem utilizados. O resultado foi aplicado na Universidade Federal de Itajubá, mais especificamente no setor de contabilidade e finanças, auxiliando os gestores na tomada de decisões e também por fins de transparência, expondo os relatórios provindos do modelo. Os resultados que surgem do modelo desenvolvido são *dashboards* e produtos de visualização de dados que são disponibilizados no sítio eletrônico da Universidade, ao mesmo tempo que providenciam aos servidores e à gerência uma fonte de informações rápida e eficiente, validando o modelo criado. Pode-se assim concluir que os conceitos são aplicáveis, inclusive na gestão pública, para gerarem modelos auxiliares a tomada de decisão, e que também, podem ser desenvolvidos em outros setores e organizações.

Palavras chave: Sistemas de Informação; *Business Intelligence*; *Data Science*; Modelagem.

ABSTRACT

Management decision-making is a recurring issue in organizations and companies, whether private or publicly managed. The latter type of organization one remains a bureaucratic system with difficult access to quick and accurate information. With the massive amount of data available within organizations, and also from external sources, they have searched new technologies and methods from the Information Systems field to obtain higher quality information. Business Intelligence (BI) systems are one of the means that contribute to gather, analyze and propagate data, resulting in various products and management reports that ease decision making process; and Data Science (DS), which is an emerging field within Information Systems, carries the characteristics of data transformation and analysis in a way that helps the organization in the decision-making process as well. Both concepts have their methods, processes, models, and life cycles that reach a common goal. However, in the literature, there is a lack of models that aggregate the two concepts concomitantly, or apply concepts from one field to another. From this gap point, the objective of this research is to propose a model that applies the concepts of BI and DS, placing the concepts of each one together, conceptualizing them and identifying the points where they are convergent and divergent to have an efficient model. The methodology used is the Modelling one, to develop the proposed model, which encompasses its Conceptualization, Modeling, and Solution and Implementation, where the concepts and steps of the BI and DS processes are explored and placed together, with their own cycles and phases being put together. Finally, the developed model will be applied in a computational software that can incorporate it, as a way to test and validate it, generating computational products to be used. The result will be applied at the Federal University of Itajubá, more specifically in the accounting and finance sector, assisting managers in decision making and also for transparency purposes, exposing the reports coming from the model. The results that emerge from the developed model are dashboards and data visualization products that are available on the University's website. It can be concluded that concepts are applicable, including in public management, to generate decision support models, and that they can be developed in other sectors and organizations.

Key Words: Information Systems; Business Intelligence; Data Science; Modelling.

LISTA DE FIGURAS

Figura 2.1 - Funções de um Sistema de Informação	21
Figura 2.2 - Dimensões de um Sistema de Informação.....	23
Figura 2.3 - Sistema funcional de um SIG	25
Figura 2.4 – Estrutura de um SAD.	26
Figura 2.5 - Infraestrutura de BI	32
Figura 2.6 – Processo de ETL e seu ambiente.	33
Figura 2.7 - Exemplo de modelo de dados multidimensional	36
Figura 2.8 - Metodologias e processos envolvidos em um sistema de BI	40
Figura 2.9 - Popularidade do Data Science no Google Trends.....	41
Figura 2.10 – Os três pilares da Data Science	42
Figura 2.11 – O processo da Data Science.....	46
Figura 2.12 - Processo realizado pelo cientista de dados - exemplo.....	50
Figura 3.1 – Processo de modelagem	52
Figura 4.1 – Ciclo de vida do Business Intelligence.	59
Figura 4.2 – Processo de desenvolvimento da Data Science.....	60
Figura 4.3 – O Modelo Proposto.	62
Figura 4.4 – Primeira fase: Conceitualização detalhada	63
Figura 4.5 – Segunda fase: Modelagem detalhada	66
Figura 4.6 – Terceira fase: Solução pelo modelo e implementação detalhada	69
Figura 5.1 – Exemplo de dados a serem publicados, conforme IN 02 de 06/12/2016	73
Figura 5.2 – Tela retirada do Demonstrativo de Compromissos do SIAFI.....	75
Figura 5.3 – Exemplo de Produto de Visualização de dados, retirado da planilha base para o Modelo de BI através da perspectiva da Data Science	76
Figura 5.4 – Exemplo de Gráfico gerado pela ferramenta Microsoft Excel para ser lançado no dashboard a ser criado, a partir dos dados inseridos.	78
Figura 5.5 – Gráfico gerado: Total pago por lista em R\$	79
Figura 5.6 – Gráfico gerado: Total pago por lista em porcentagem	79
Figura 5.7 – Gráfico gerado: Total pago por fonte de recurso.	80
Figura 5.8 – Gráfico gerado: Total pago em deduções - porcentagem.....	81
Figura 5.9 – Gráfico gerado: Total pago em deduções – valores monetários.....	81
Figura 5.10 – Gráfico gerado: Dias de vencimento.....	82
Figura 5.11 – Gráficos gerados: Notas Vencidas.	82

Figura 5.12 – Planilha que contém os gráficos a serem transformados em um dashboard	83
Figura 5.13 – Planilha que contém os gráficos já transformados em um dashboard.....	84
Figura 5.14 – Publicação no sítio eletrônico da UNIFEI	86

LISTA DE TABELAS

Tabela 2.1 - Comparação entre OLAP e Data Mining	37
--	----

LISTA DE ABREVIATURAS E SIGLAS

BI - *Business Intelligence*

BI&A - *Business Intelligence and Analytics*

CIOs - *Chief Information Officers*

CNPJ – *Cadastro Nacional de Pessoa Jurídica*

CRMs - *Customer Relationship Management Systems*

DCF - *Diretoria de Contabilidade e Finanças*

DM - *Data Marts*

DS - *Data Science*

DW - *Data Warehouses*

EDA – *Exploratory Data Analysis*

ERPs - *Enterprise Resource Planning Systems*

ETL - *Extract, Transform, Load*

IEMI - *Instituto Eletrotécnico e Mecânico de Itajubá*

IN - *Instrução Normativa*

ISSQN – *Imposto Sobre Serviços de Qualquer Natureza*

KMSs - *Knowledge Management systems*

OLAP - *Online Analytical Processing*

SAD - *Sistemas de Apoio à Decisão*

SAE - *Sistemas de Apoio ao Executivo*

SCMs - *Supply Chain Management Systems*

SERPRO - *Serviço Federal de Processamento de Dados*

SI - *Sistemas de Informação*

SIAFI - *Sistema Integrado de Administração Financeira do Governo Federal*

SIG - *Sistemas de Informação Gerenciais*

SIT - *Sistemas de Informação Transacionais*

STN - *Secretaria do Tesouro Nacional*

TI - *Tecnologia da Informação*

UNIFEI - *Universidade Federal de Itajubá*

SUMÁRIO

1. INTRODUÇÃO	14
1.1. Objetivos da dissertação	17
1.2. Estrutura da dissertação	17
2. REFERENCIAL TEÓRICO	19
2.1. Sistemas de Informação	19
2.1.1. Conceitos e definições	20
2.1.2. Sistemas organizacionais	23
2.1.3. SI na gestão pública	28
2.2. Inteligência de Negócios - <i>Business Intelligence</i>	29
2.2.1. Conceitos e definições	30
2.2.2. Conceitos acerca da infraestrutura de BI	31
2.2.3. Ciclo de vida de um sistema de BI	37
2.3. Ciência dos Dados - <i>Data Science</i>	40
2.3.1. Conceitos e definições	41
2.3.2. Cientista de dados	44
2.3.3. Processo de desenvolvimento do <i>Data Science</i>	45
3. METODOLOGIA DE PESQUISA	51
3.1. Características e procedimento da Pesquisa	51
3.2. Objeto de estudo	55
3.2.1. Universidade Federal de Itajubá UNIFEI	55
3.2.2. Diretoria de Contabilidade e Finanças - DCF	56
4. O MODELO PROPOSTO	58
4.1. Conceitualização	62
4.2. Modelagem	65
4.3. Solução pelo modelo e Implementação	68
5. APLICAÇÃO DO MODELO PROPOSTO	72
5.1. Conceitualização	72
5.2. Modelagem	75
5.3. Solução pelo Modelo e Implementação	85
6. CONCLUSÕES	88
6.1. Conclusões e contribuições da pesquisa	88
6.2. Sugestões para trabalhos futuros	90
REFERÊNCIAS BIBLIOGRÁFICAS	91

1. INTRODUÇÃO

Toda organização necessita de um processo de tomada de decisão eficiente. Em sua maioria, os objetivos das organizações são centrados em tomadas de decisão, auxiliando os funcionários e os negócios a atingirem seus propósitos e finalidades (STAIR e REYNOLDS, 2010). A tomada de decisões eficiente é um dos grandes desafios dos gestores e das empresas atualmente (MORABITO NETO e PUREZA, 2012). A partir desses conceitos de tomada de decisão, é possível constatar a importância desse processo decisório dentro das organizações.

As companhias são forçadas nos dias atuais a estabelecerem novas ferramentas de análise de seus dados para melhorarem suas tomadas de decisões, aumentando assim a eficiência e eficácia dentro de seus processos produtivos e de negócios (FLATH e STEIN 2016). As ferramentas devem sempre atender aos tomadores de decisão, bem como possuir o processamento correto de dados para tal atividade.

A massiva disponibilidade de dados em uma empresa tem aumentado o interesse em métodos para se obter informações e conhecimentos pertinentes para a tomada de decisão. Com a disponibilidade de tecnologia, o método de tomada de decisões que era baseado em experiência ou modelos restritos da realidade, hoje pode ser baseado em produtos de dados. Em outras palavras, uma organização pode reunir mais dados que anteriormente e analisá-los, para melhorar cientificamente suas previsões, decisões e por fim, eficácia e produtividade (AMIRIAN, LOGGERENBERG e LANG, 2017). Com a quantidade de dados, há uma necessidade de ferramentas capazes de processá-los de modo eficiente, para transformá-los em informações capazes de agregar valor às organizações.

Um Sistema de Informação (SI) eficaz pode fornecer informações importantes aos seus usuários, quando fornecidas no tempo correto, livres de erros. Informações fornecidas no tempo certo se relacionam à tomada de decisões, para que os gestores as detenham quando são necessárias (LAUDON e LAUDON, 2016). Mesmo em meio à quantidade de dados disponíveis, os SI necessitam ser ágeis e eficientes, para fornecer de modo fácil tais informações sem erros e que agreguem valor aos administradores das empresas.

Segundo Arnott, Lizama e Song (2017), um sistema de *Business Intelligence* (BI) é um leque de processos e *software* que são utilizados para reunir, analisar e propagar dados,

com o objetivo da tomada de decisão. O BI envolve a coleta de informações certas no tempo correto e deixá-las úteis para que os gestores possam utilizar e analisar, trazendo benefícios para a estratégia de negócios e às operações das organizações (STAIR e REYNOLDS, 2010). A partir de um Sistema de Informação voltado ao *Business Intelligence*, a tomada de decisões pode ser facilitada, com os relatórios provindos dos produtos de BI.

O *Business Intelligence* é uma grande facilitadora para as empresas, ajudando-as a se tornarem mais “inteligentes”, e tomarem melhores decisões com a utilização de informações, através do uso em suas rotinas. A cadeia de valor de informações é o processo que se utiliza para se retirar valor dos dados transformados em informação. O BI, portanto, é o centro desse processo (LARSON e CHANG, 2016). Centralizado nesse processo, a transformação de dados em informações pelo SI de BI pode ser de alta relevância à organização.

De acordo com Schutt e O’Neil (2014), no mundo contemporâneo há uma enorme quantidade de dados sendo trafegados simultaneamente – compras, músicas, notícias, tudo sendo compartilhado online. E não é só a quantidade massiva de dados que deixam os dados interessantes aos analistas, mas também a oportunidade de se criar produtos de dados. Não só dados trafegados dentro das organizações as influenciam, mas também os dados externos, os quais crescem muito no mundo presente.

De fato, muitos negócios são invadidos por muitos dados, e muitas organizações estão sempre tentando capitalizar os dados com análises para ganhar vantagens competitivas. A *Data Science* (DS), bem como outras formas de análise de dados, faz parte destas áreas emergentes de competição envolvendo dados, dentre os negócios (HAZEN *et al.*, 2014). Os negócios necessitam a cada dia mais atualizações na área de análise de dados, para posicionar-se de forma competitiva no mercado.

O conceito de *Data Science* é uma extensão das estatísticas capazes de lidar com uma grande quantidade de dados que são produzidos nos dias de hoje, e inclui conceitos de Ciência da Computação, de *Business Intelligence* e capacidade de trabalhar com algoritmos e outras ferramentas computacionais (CIELEN, MEYSMAN e ALI, 2016). Como os negócios necessitam cada vez mais de análise de dados, combinar ferramentas pode ser

uma forma de agregar valor à organização, auxiliando os tomadores de decisão com as informações obtidas pelo processamento de dados.

A *Data Science* é utilizada em quase todos os locais, tanto comercialmente quanto em locais que não visam lucro, para se ganhar conhecimento e levar a seus clientes e usuários uma melhor experiência de seu negócio e organização, incluindo organizações governamentais, as quais estão cada vez mais utilizando a *Data Science* para possuir informações internas ou disponibilizá-las ao público (CIELEN, MEYSMAN e ALI, 2016). Apesar da burocracia existente dentro das organizações governamentais, há grande necessidade de trabalhar com dados dentro delas, incluindo relatórios e ferramentas para auxílio dos servidores públicos e das gerências.

O ponto central dessa pesquisa é a utilização cruzada dos conceitos de *Business Intelligence* e *Data Science*, com a proposta de desenvolvimento de um modelo de BI através da perspectiva da *Data Science*, visando a melhoria na tomada de decisões e assim, agregando valor à organização. A aplicação do modelo é utilizada como meio de validação do modelo proposto, para assim implementar modificações e melhorias.

Na literatura, há uma carência de propostas de modelos que agreguem conceitos de *Business Intelligence* em conjunto com conceitos de *Data Science*, ou então de modelos de BI com pontos de vista da perspectiva da DS. Faltam modelos de BI que se agreguem ao contexto da *Data Science*, este último sendo mais novo e sendo tratado com mais destaque atualmente.

Vale salientar que essa dissertação anseia contribuir cientificamente na área de Sistemas de Informação, desdobrando-se nas áreas específicas de *Business Intelligence* e *Data Science*, trazendo para análise um modelo de apoio à decisão que leve em consideração o BI no contexto do DS, havendo desta maneira, um estudo combinado entre as áreas supracitadas.

A implementação do modelo, em uma ferramenta computacional que incorpore o modelo desenvolvido, permite a validação dos conceitos, além de gerar um produto utilizável e aplicável a diversas organizações, setores e áreas de interesse, proporcionando um modelo reutilizável e aplicável em diversas organizações.

1.1. Objetivos da dissertação

O objetivo principal desta dissertação é propor um modelo de *Business Intelligence* (BI), a partir da perspectiva do *Data Science* (DS), para o auxílio da tomada de decisão.

O modelo proposto combina os conceitos de BI e DS, gerando assim uma contribuição científica, visando novos conhecimentos acerca dos dois principais temas abordados e quando estes são concomitantes, complementares e, também, divergentes.

Sob este objetivo geral, tem-se o seguinte objetivo específico:

Desenvolvimento de uma ferramenta computacional que incorpore o modelo desenvolvido, como forma de testá-lo e validá-lo. Para isso foi realizado o levantamento dos dados e a escolha do *software* que se possa utilizar o modelo; sua aplicação, para o suporte de informações à Diretoria de Contabilidade e Finanças acerca dos pagamentos e da transparência que deve ser apresentada no sítio eletrônico da Universidade Federal de Itajubá. Essa transparência é exigida legalmente, de acordo com a Instrução Normativa (IN) 02 de 06 de dezembro de 2016, pelo Ministério do Planejamento e Gestão – Publicada no Diário Oficial da União em 08/12/2016.

1.2. Estrutura da dissertação

Esta dissertação aborda a proposta de modelo de *Business Intelligence* a partir da perspectiva da *Data Science*, referenciando a teoria de ambos os conceitos e a modelagem e sua metodologia. Sendo assim, para dissertar sobre esse assunto, esta pesquisa de mestrado é composta por 6 (seis) capítulos.

No *Capítulo 1* é realizada uma breve introdução sobre o tema abordado nesta dissertação. Nele encontram-se as justificativas, contextualização e os objetivos do trabalho. Uma breve introdução à metodologia de pesquisa utilizada e a estruturação da pesquisa também fazem parte desse capítulo.

Em seguida, o *Capítulo 2*, traz a base teórica sobre os temas Sistemas de Informação, *Business Intelligence* e *Data Science*. Dentro desses temas, serão tratados os conceitos e definições de Sistemas de Informação e suas vertentes, que são pré-requisitos para o *Business Intelligence*, onde também é explanado seus conceitos, culminando na teoria da *Data Science*, fechando o ciclo da pesquisa.

O **Capítulo 3** apresenta com detalhes a metodologia utilizada - o método modelagem quantitativa em gestão de operações, explicando os passos a serem reproduzidos no capítulo seguinte deste trabalho, em conjunto com os conceitos abordados no capítulo anterior.

No **Capítulo 4**, encontra-se a proposta completa do modelo a ser desenvolvido a partir do método escolhido no capítulo anterior, bem como a aplicação dos conceitos do capítulo 2. Neste capítulo, todos os detalhes do modelo proposto são explanados para depois ser aplicado no capítulo 5 desta dissertação.

No **Capítulo 5** encontra-se a aplicação do método proposto no Capítulo 4, através dos conceitos da revisão da literatura, na Diretoria de Contabilidade e Finanças da Universidade Federal de Itajubá – UNIFEI.

Em seguida, o **Capítulo 6** finaliza este trabalho com sugestões para trabalhos futuros e conclusões. Por fim encontram-se as referências usadas nessa dissertação.

2. REFERENCIAL TEÓRICO

2.1. Sistemas de Informação

Como alguns setores econômicos são muito dependentes do uso correto de dados e informações, esses estão empenhados em aumentar seu conhecimento em como utilizar tecnologias neste sentido. As companhias estão descobrindo que a chave para o sucesso é o desenvolvimento de Sistemas de Informações (SI) propícios para tais setores (SANCHEZ, TERLIZZI e MORAES, 2017). Unindo práticas tradicionais de negócios com SI recentemente desenvolvidos, novos desafios para as organizações são criados, balanceando controle e agilidade (FAYOUMI e LOUCOPOULOS, 2016).

As organizações compreendem que os SI não são somente uma prestação de serviço dentro das companhias, mas contribuem com a performance geral de toda companhia. Os SI estão enraizados no dia-a-dia das organizações e auxiliam na elaboração da visão do negócio, além das estratégias organizacionais (MARNEWICK, 2015).

Para Fayoumi e Loucopoulos (2016), as organizações estão utilizando os Sistemas de Informação para apresentar novos modelos de negócio, mesmo que estes mesmos modelos não tenham sido previamente apresentados, e são um elemento chave para os ambientes de negócio atuais. As organizações, sejam públicas ou privadas, colaboram de diversas maneiras na forma de lidar com sistemas cujos elementos são operacionais e de gestão independente – gerando uma evolução contínua.

Laudon e Laudon (2016), destacam que as companhias Norte-Americanas gastaram um valor de US\$ 814 bilhões em telecomunicações, *software* e *hardware* na área de Sistemas de Informação no ano de 2014. Já Marnewick (2015) destaca que a África do Sul despendeu US\$18,18 bilhões em 2017 e a Holanda € 10,3 bilhões em 2016. Pelo motivo destes custos tão altos existirem, uma organização não pode se permitir gastar dinheiro em projetos que não entreguem resultados.

Nas próximas seções os conceitos acerca dos Sistemas de Informação, bem como suas definições são dissertados com base na literatura. Em seguida, é apresentada uma classificação dos tipos de SI. Por fim, há uma breve conceituação dos SI na gestão pública.

2.1.1. Conceitos e definições

De acordo com Laudon e Laudon (2016), um Sistema de Informação pode ser definido tecnicamente como uma série de componentes que agrupam, transformam, armazenam e distribuem informações para ajudar na tomada de decisões e na gestão de uma organização. Além disso, os sistemas de informação auxiliam a gerência e os trabalhadores da organização a averiguarem problemas, criarem novos produtos e visualizarem assuntos complexos.

Sistemas de Informação são sistemas socioeconômicos que compreendem *hardware*, *software* e o sistema organizacional como um todo (AHLEMANN, 2009). O termo “Sistemas de Informação” foi definido para caracterizar essa ligação entre uma comunicação de tecnologia, *hardware* computacional e *software* desenvolvidos e para lidar com informações referentes a processos da organização. Eles servem para controlar o trabalho de diferentes e diversas funções da companhia, desde o suporte a administração até ferramentas estratégicas. Alguns exemplos de controles efetuados pelos SI são: as ordens de compra, controle de estoque e emissões de contracheque (YEO, 2002).

Um processo de uma organização é o conjunto contínuo de atividades exercidas dentro das empresas, que geram produtos e serviços relacionados a elas, utilizando-se de insumos, recursos e a própria produção em si, envolvendo ou não o cliente de forma direta. As organizações buscam ter seus processos cada vez mais eficientes e eficazes (RAINER e CEGIELSKI, 2016).

O conceito de “informação” é formado quando dados são processados de maneira que se tornem úteis e significativos para os usuários, e também é definido por dados estruturados que possuem um significado contextual, entregando ao usuário o conhecimento necessário para tomada de decisões (YEO, 2002). Para Stair e Reynolds (2010), o termo “informação” também é conceituado como uma coleção de fatos organizados que acrescentam mais valor do que fatos individuais dentro de uma organização, como por exemplo, unindo dados de valores financeiros de venda com os valores de quantidade de vendas.

O termo “dados” é apresentado como um fluxo de fatos brutos que constituem eventos que ocorrem nas organizações antes de serem processados e organizados para que as pessoas possam utilizá-los (LAUDON e LAUDON, 2016). Já o conceito de “sistemas” é formado quando um conjunto de elementos interagem entre si para chegarem a um objetivo em

comum (STAIR e REYNOLDS, 2010). De acordo com Rainer e Cegielski (2016), o termo “conhecimento” é explanado através dos dados e informações que são compiladas e transformadas para que se obtenha a compreensão e *know-how* quando aplicados dentro da organização, de problemas quotidianos.

Segundo Laudon e Laudon (2016), para que as organizações possam tomar decisões, analisar problemas e controlar suas operações, três atividades em um SI são necessárias para que produzam informações a partir dos dados. O “*input*” (entrada), que colhe os dados brutos de dentro da organização; o “processamento”, que converte esses dados em uma forma compreensível; e, por fim, o “*output*”, que entrega as informações processadas para as pessoas que as utilizarão nas atividades da organização. Os SI também necessitam do “*feedback*”, o qual é retornado a partir do *output*, para que os membros da organização reavaliem os *inputs* de forma correta, em caso de erros. O esquema dessas três funções encontra-se na Figura 2.1.

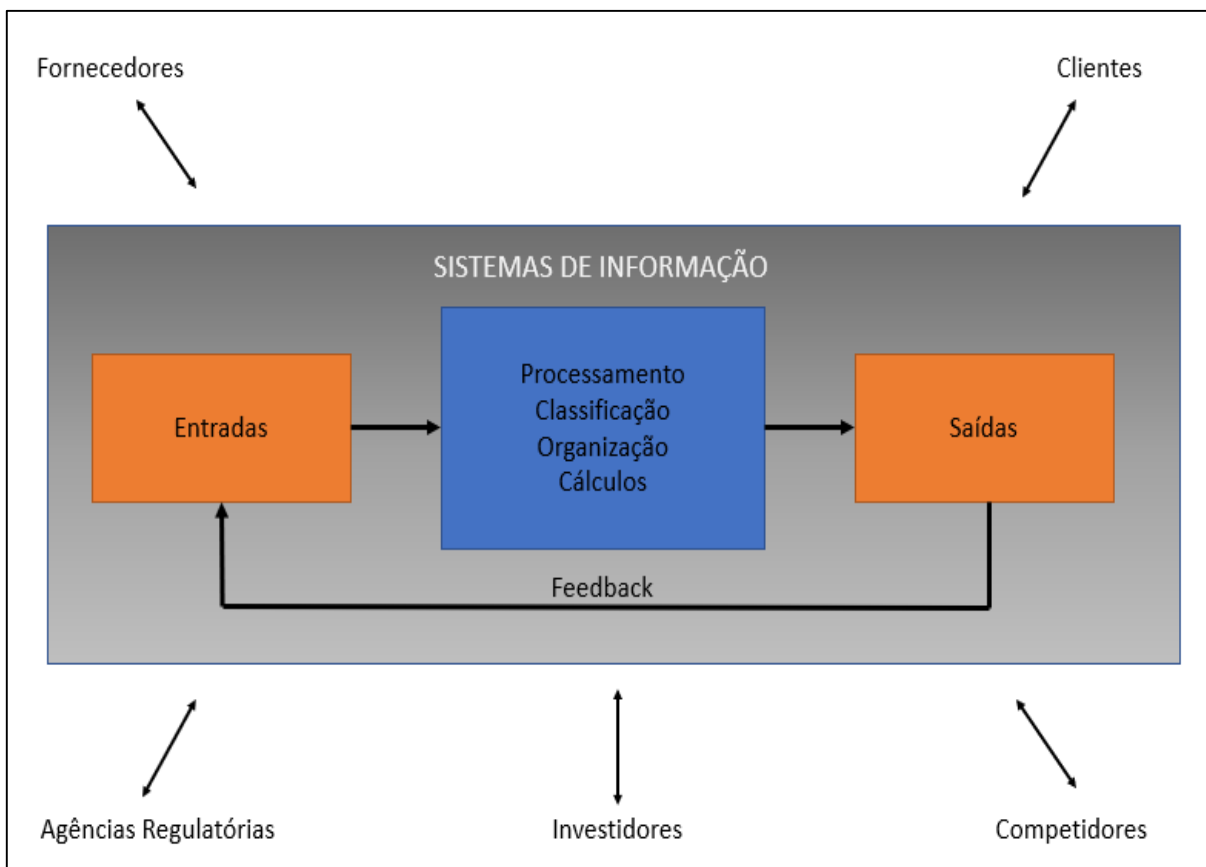


Figura 2.1 - Funções de um Sistema de Informação

Fonte: Adaptado de Laudon e Laudon (2016).

Ainda de acordo Laudon e Laudon (2016), para compreendermos os SI completamente, precisamos entender as dimensões de um SI: a tecnologia, as organizações e a gestão:

- **Organizações:** Os SI são um sistema integral em uma companhia e, dependendo do negócio, este não existiria sem os SI. Os elementos-chaves de uma organização são as pessoas, suas políticas internas, seus processos, estruturas e cultura organizacional. As companhias são subdivididas em níveis diferentes de hierarquia, dependendo de suas responsabilidades e autoridades.
- **Gestão:** Tomar decisões, elaborar planos de ação para solução de problemas e resolver diversas situações são algumas das competências da gerência. Os gestores necessitam de percepção acerca dos desafios das organizações e como responder a eles, além da criação de novos produtos e serviços, “renovando” a empresa de tempos em tempos.
- **Tecnologia da Informação:** uma das ferramentas que os gestores utilizam é a Tecnologia da Informação (TI), desde o *hardware* utilizado até os *software* desenvolvidos. A TI é composta por vários *inputs* e *outputs*, bem como dispositivos de armazenamento de dados, além de dispositivos de telecomunicações e internet e redes que interligam os computadores e aparelhos.

De acordo com Rainer e Cegielski (2016), a TI afeta as organizações de diversas maneiras: reduz o número de gerentes intermediários, deixando-os mais produtivos; transforma o trabalho do gerente, transformando a forma de tomada de decisões por eles; reduz os custos, por vezes diminuindo os postos de trabalho; aumenta a saúde e segurança dos funcionários, melhorando sua qualidade de vida e facilitando a acessibilidade por usuários com deficiência.



Figura 2.2 - Dimensões de um Sistema de Informação

Fonte: Adaptada de Laudon e Laudon (2016).

2.1.2. Sistemas organizacionais

De acordo com Laudon e Laudon (2016), existem três tipos de sistemas, posicionados para cada nível de gerência em uma organização, os SIT – Sistemas de Informação Transacionais, os Sistemas que dão suporte ao *Business Intelligence* (BI), e os Sistemas que interligam a organização.

Os SIT – Sistemas de Informação de Processamentos Transacionais são os SI responsáveis por gerir atividades e transações cotidianas em uma companhia, como vendas, receitas, depósitos em dinheiro e fluxo de materiais em um estoque, por exemplo (LAUDON e LAUDON, 2016). As “transações” são trocas relacionadas aos negócios das organizações, e são as primeiras aplicações computacionais a serem desenvolvidas pelas empresas (STAIR e REYNOLDS, 2010).

A principal função dos SIT é responder questionamentos rotineiros com informações, em sua maioria, fáceis de compreender. Em um nível operacional, os SIT são essenciais, pois

os objetivos e tarefas são bem estruturados e pré-definidos, ajudando o gestor de um nível inferior de uma organização (LAUDON e LAUDON, 2016).

Os SIT coletam geralmente os dados em tempo real, assim que os dados são gerados, fornecendo *inputs* para os bancos de dados das empresas. Eles são muito importantes para as empresas, pois fornecem auxílio às suas principais operações (RAINER E CEGIELSKI, 2016).

Já os Sistemas que suportam o *Business Intelligence* (BI), segundo Laudon e Laudon (2016), são divididos em três, os SIG – Sistemas de Informação Gerenciais, os SAD – Sistemas de Apoio à Decisão e os SAE, Sistemas de Apoio ao Executivo. O BI, que é abordado no próximo tópico deste capítulo, é um termo contemporâneo para ferramentas de dados e *software* que analisam, organizam e providenciam acesso aos dados, ajudando aos gestores na tomada de decisão:

- SIG – Sistemas de Informação Gerenciais: os SIG auxiliam os gestores de nível médio na tomada de decisão, com o fornecimento de relatórios da performance atual da organização, bem como é utilizado para monitorar o negócio e até mesmo fazer previsões sobre a performance futura da empresa (LAUDON e LAUDON, 2016).

O SIG é uma reunião de pessoas, procedimentos, *software*, bases de dados e dispositivos que fornecem informações rotineiras aos tomadores de decisão, e em vários casos são utilizados para relatórios periódicos em bases diárias, semanais, mensais e anuais. Os SIG se proliferaram pois possuem um grande valor para os gestores das empresas (STAIR e REYNOLDS, 2010).

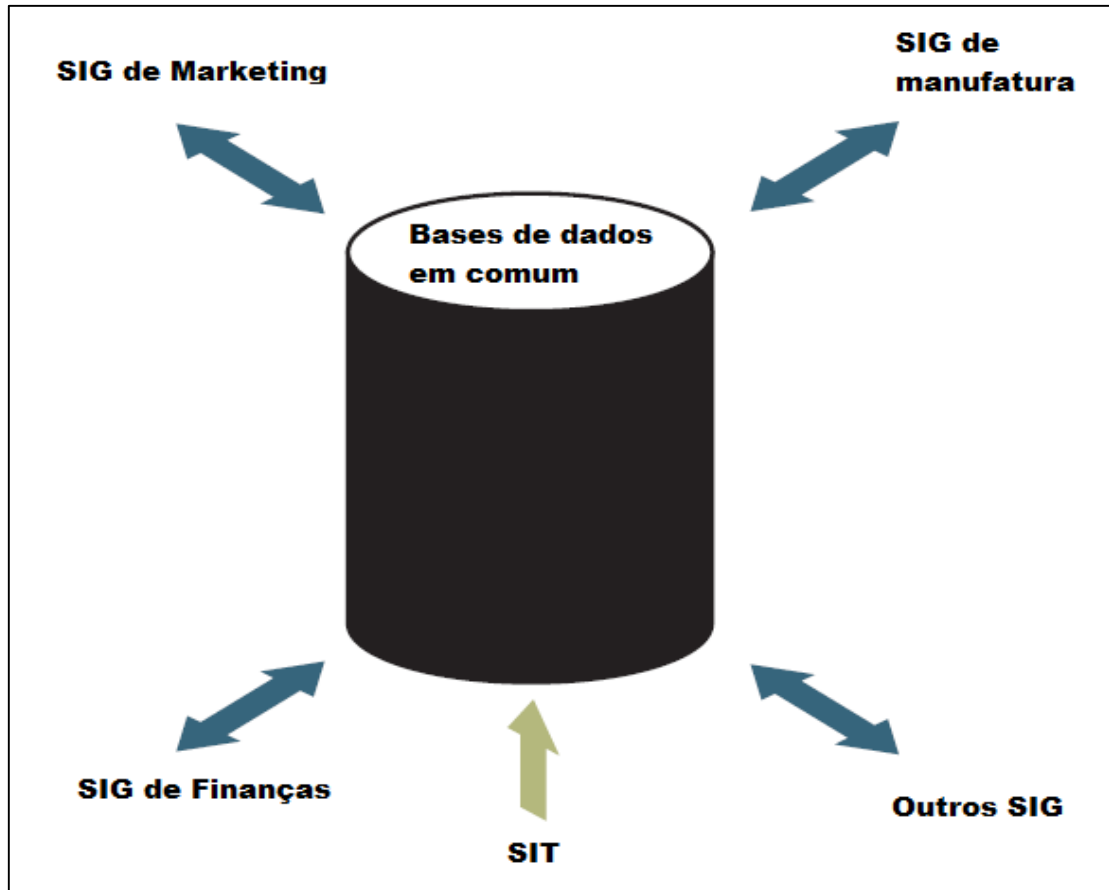


Figura 2.3 - *Sistema funcional de um SIG*

Fonte: Stair e Reynolds (2010).

- SAD - Sistemas de Apoio à Decisão: Os SAD possuem seu foco em problemas que são únicos e que possuem uma mudança rápida, e por esse motivo, a solução não pode ser pré-definida antes. Os SAD, além de utilizarem dados provindos dos SIG e dos SIT, também utilizam dados externos às organizações e são utilizados por gestores que necessitam de análises sofisticadas e modelos para análise de dados (LAUDON e LAUDON, 2016).

Além de auxiliar a tomada de decisão em todos os aspectos, um SAD pode ajudar aos usuários indo além de um SIG tradicional, com uma assistência imediata na solução de problemas, que por vezes são complexos e a informação chave para solução, de difícil obtenção, oferecendo alternativas ótimas para tomada de decisão. Gestores exercem um papel importante na elaboração e implementação de um SAD (STAIR e REYNOLDS, 2010).

Para Arnott, Lizama e Song (2017), um dos conceitos principais no desenvolvimento de um SAD é que os tomadores de decisão, que são os usuários primordiais de um SAD, podem livremente escolher se utilizam ou não o sistema. Ou seja, criar o interesse e comprometimento dos usuários que demandam um projeto em SAD é crucial para a utilização desse sistema. Um sistema SAD pessoal, que seria um sistema moderadamente simples, desenvolvido para um gerente utilizar individualmente, ou para um grupo pequeno de gestores, é a forma original de um SAD. A estrutura de um SAD pode ser vista na Figura 2.4.

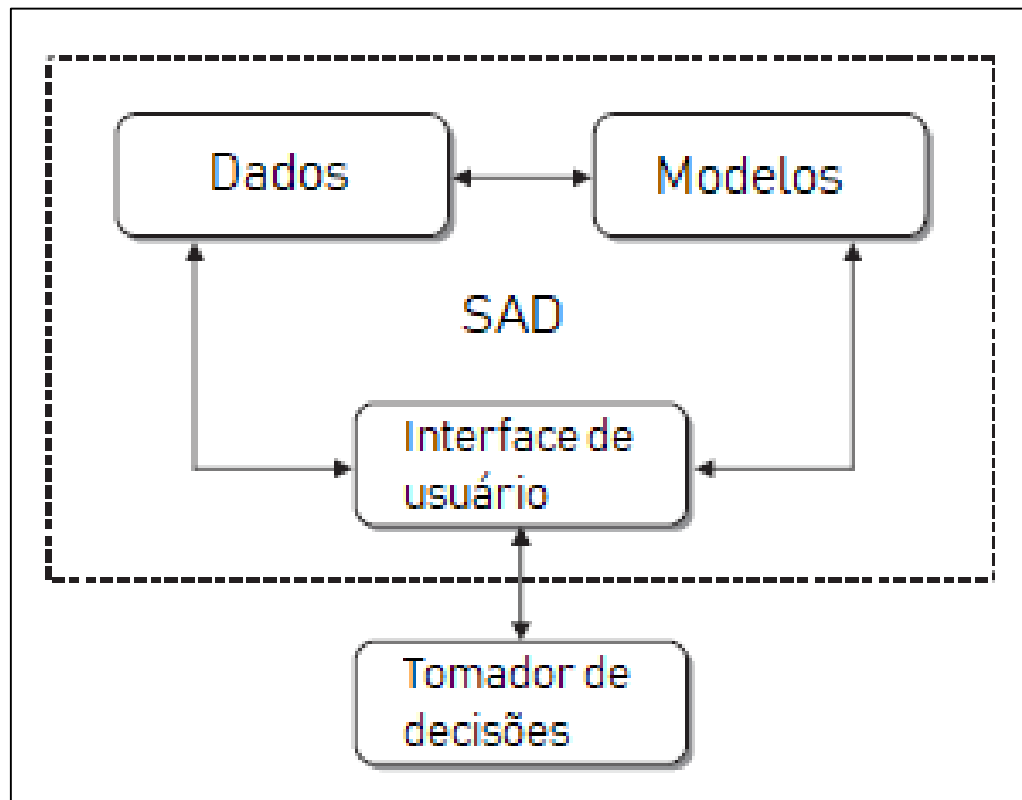


Figura 2.4 – Estrutura de um SAD.

Fonte: Vercellis (2009).

Segundo Stair e Reynolds (2010), um SAD pode ter uma coleção de modelos usados no auxílio a decisão, uma reunião de fatos e informações para auxílio a decisão e sistemas e procedimentos, que interagem com o gestor, para auxílio a decisão, além de interação com outros SAD.

- SAE - Sistemas de Apoio ao Executivo: ajudam a alta gestão na tomada de decisões. Esse tipo de SI ajuda na tomada de decisões quando não há procedimento específico que ajude na solução de problemas. Os SAE apresentam gráficos e dados de muitas

maneiras e fontes, com uma interface que facilitam ao gestor de alto nível na sua utilização. Por vezes os gestores utilizam SAD com *dashboards* digitais, uma tela com vários gráficos e quadros com indicadores para ajudarem na tomada de decisão, utilizando-se de dados provenientes de SIG e SAD, além de outras fontes externas (LAUDON e LAUDON, 2016).

Ainda conforme Laudon e Laudon (2016), os sistemas que interligam a organização, são compostos dos chamados ERPs (*Enterprise Resource Planning Systems*), SCMs (*Supply Chain Management Systems*), CRMs (*Customer Relationship Management Systems*) e KMSs (*Knowledge Management systems*):

- ERPs (*Enterprise Resource Planning Systems*): os ERPs são uma série de programas integrados que controlam operações vitais ao negócio, para todas as plantas das companhias. Um sistema ERP podem substituir uma gama de programas, unificando-os em apenas um, fazendo todo o sistema de mais fácil utilização e mais eficaz. Apesar de um ERP variar de empresa para empresa, a maioria deles contém *software* que auxiliam na produção e em finanças (STAIR e REYNOLDS, 2010).
- SCMs (*Supply Chain Management Systems*): os SCMs ajudam os gestores em sua relação com os fornecedores de sua companhia. Esses sistemas auxiliam os fornecedores, distribuidores e empresas de logística a obter informações compartilhadas acerca das ordens de produção, da produção em si e dos serviços prestados em conjunto, além de ajudar no controle de estoque e na entrega dos produtos (LAUDON e LAUDON, 2016).
- CRMs (*Customer Relationship Management Systems*): Os programas de CRM auxiliam as empresas no controle e gestão de relacionamento com os clientes, incluindo setores de *marketing* e publicidade, vendas e serviço ao consumidor. Os CRMs ajudam a empresa na coleta de dados dos clientes, além de conseguir informações acerca de novos produtos e suas vendas, bem como atrair *feedback* dos consumidores de uma forma mais ágil, através de e-mails, *surveys* e hábitos de Internet dos consumidores (STAIR e REYNOLDS, 2010).
- KMSs (*Knowledge Management systems*): Já os KMSs são utilizados por algumas empresas para criar, produzir e entregar serviços e produtos através da gestão do

conhecimento exercido nela. Os KMSs auxiliam as empresas melhorando os processos de coleta de dados acerca de conhecimento e *expertise* (LAUDON e LAUDON, 2016).

2.1.3. SI na gestão pública

Os Sistemas de Informação são importantes para aumentar a eficiência e eficácia no setor público. Todavia, os projetos grandes em SI ainda são um desafio grande na administração pública. Nos anos recentes, os governos dos países estão utilizando do poder da TI e dos SI para que, inclusive os cidadãos, possam participar em um processo mais democrático (KLIER, KLIER e MUSCHTER, 2017).

De acordo com Pang (2014), as organizações públicas são grandes consumidores de TI. O governo dos EUA usa a TI para vários propósitos, desde a manutenção as infraestruturas operacionais, até a interação com os cidadãos. Os pesquisadores na área de SI estão sempre identificando novas práticas organizacionais a partir dos investimentos em TI que os governos praticam, agregando valor a partir deles.

Klier, Klier e Muschter (2017) afirmam que a partir dos orçamentos governamentais nos EUA, os CIOs (*Chief Information Officers* – alta gerência de tecnologia da informação) estão cada vez mais investindo em TI, em especial nos setores de educação. Na Alemanha, as instituições investiram € 20 Bilhões em 2013 na área de TI governamental, e o crescimento desse investimento na Europa tende a aumentar mais, com 141 projetos propostos pela Comissão Europeia, entre 2007 e 2013.

Para Pang (2014) a burocracia existe de diversas formas na administração de TI nas organizações públicas, e há poucas pesquisas acerca desse tema, além de afirmar que muitos gestores de TI se apegam a uma infraestrutura de TI obsoleta e inflexível. Essa infraestrutura possui um grande custo de manutenção, levando meses ou anos para que uma decisão acerca de investimentos e desenvolvimentos de TI seja tomada, pois envolve a hierarquia e processos complexos.

É importante destacar que os órgãos públicos não são uniformes, variando em sua estrutura e níveis administrativos. Em um nível geral, existem muitas características rudimentares. Além disso, diferem das organizações privadas em três características: sua liderança é uma liderança que foi eleita, possuindo uma cadeia de comando, com uma série de regras que

variam de setor para setor. A segunda característica é como lidam com o público, pois são monopolizadas em certas áreas. Por fim, diferem em relação a como enxergam o público, pois não são clientes, são cidadãos, que possuem direitos constitucionais e leis a seguir (LINDGREN e JANSSON, 2013).

Apesar da importância da utilização da TI e de SI nos setores públicos e seu contínuo crescimento, a gestão de projetos nessas áreas é desafiadora. A responsabilidade da gestão de TI nos órgãos públicos é geralmente muito dividida, especialmente em um nível federal, com muitos projetos, portanto, necessitando de financiamentos por *stakeholders* (KLIER, KLIER e MUSCHTER, 2017).

2.2. Inteligência de Negócios - *Business Intelligence*

De acordo com Arnott, Lizama e Song (2017), o termo Business Intelligence (BI) – Inteligência de Negócios, é comumente utilizado para designar Sistemas de Apoio à Decisão (SAD) de larga escala dentro das organizações. Entretanto, os sistemas de BI não são apenas sistemas SAD focados em dados, eles são um *mix* destes dados e sua análise. Muitas *surveys* realizadas por analistas de empresas frequentemente resultam que o desenvolvimento de BI é uma das prioridades dos CIOs e continuam com essa prioridade até os dias atuais.

Investimentos em BI foram os maiores investimentos de TI nas organizações em 2015 e tem sido o maior desde 2009. Portanto, o investimento em BI é um dos mais importantes dentro de uma organização, e continuará sendo por algum tempo (ARNOTT, LIZAMA e SONG, 2017).

Chen, Chiang e Storey (2012) destacam que o termo “Inteligência” (*Intelligence*) é utilizado por pesquisadores desde a década de 1950, para denotar inteligência artificial. Já o termo “Inteligência de Negócios” - *Business Intelligence*, se tornou popular nas organizações e comunidades de TI apenas na década de 1990. Após os anos 2000, o termo “*Business Analytics*” foi empregado para representar o componente de Análise do BI – utilizando-se do termo “*Business Intelligence and Analytics*” (BI&A), para unificar os termos. Nesta dissertação é considerado apenas o termo “*Business Intelligence*”, definindo o termo análogo “BI&A”.

Para Arnott, Lizama e Song (2017), em termos de níveis de gestão que utilizam sistemas de BI, os níveis superiores os utilizam como um fornecedor de informações estratégicas e táticas para

tomada de decisões. Já para os níveis inferiores, os sistemas de BI ajudam nas tarefas do dia a dia, além de ajudarem a gestão intermediária a cumprirem objetivos estratégicos, com o resultado de 22,5% de pesquisas em BI, que são realizadas nas organizações, são focadas em tarefas de tomada de decisões estratégicas. Portanto, muitas companhias utilizam dos sistemas de BI principalmente para tomada de decisões estruturadas, com dados internos das organizações.

Os conceitos sobre *Business Intelligence* são abordados nas próximas seções, assim como as definições de BI, conceitos sobre sua infraestrutura e seu ciclo de desenvolvimento.

2.2.1. Conceitos e definições

De acordo com Larson e Chang (2016), o termo “BI” é definido na literatura de diversas formas, e descrevem o BI como a habilidade de prover uma vantagem de informação para a organização, para que o negócio realize suas tarefas cotidianas de maneira mais eficiente. Segundo Laudon e Laudon (2016), BI é um termo utilizado que descreve a infraestrutura para armazenamento, integração, geração de relatórios e análise de dados dentro de um ambiente organizacional. Já para Vercellis (2009), BI pode ser definido como um leque de modelos matemáticos e metodologias de análise que avaliam a disponibilidade de dados necessários para gerar informações e conhecimento úteis para tomada de decisão.

Para Stair e Reynolds (2010), o conceito de BI envolve a coleta de informações corretas em uma maneira temporal e em uma forma utilizável para análise, possibilitando um efeito favorável na estratégia de um negócio e suas operações. O BI transforma dados em informações úteis para serem distribuídas pela companhia, oferecendo auxílio para soluções de problemas e melhorando os processos.

O *Business Intelligence* tem por objetivo auxiliar os usuários a tomarem decisões oferecendo métodos, ferramentas para um acesso rápido para que possam gerenciar as informações. A evolução das necessidades dentro dos negócios e a disponibilidade de muitas ferramentas de BI, motivaram as empresas a medirem suas performances minuciosamente, alinhando o BI aos seus objetivos a longo prazo (BRICHNI *et al.*, 2017).

Os sistemas de BI desempenham um papel importante para permitir o conhecimento da performance das atividades de uma organização através da análise e coleta de uma série de dados. Existe uma grande diversidade de ênfases que podem ser feitas (NEWMAN *et al.* 2016).

De acordo com Vercellis (2009), a vantagem de um tomador de decisões, que pode contar com um sistema de BI auxiliando suas atividades, é a qualidade de um processo de tomada de decisão mais avançado, o qual é beneficiado de modelos matemáticos e algoritmos que ajudam a análise de uma quantidade grande de possibilidades de ação, com conclusões mais certeiras para que escolha a decisão mais eficaz.

Segundo Brichini *et al.* (2017), é inquestionável que, se um sistema de BI existir para auxiliar a tomada de decisões em uma companhia, os desenvolvedores do sistema devem possuir um conhecimento das tarefas que envolvam decisões na organização, além de trabalhar em conjunto com a gerência (tomadores de decisão) para melhorarem a efetividade do processo de tomada de decisão.

Para Larson e Chang (2016), um sistema de BI não é um sistema comum de Tecnologia da Informação, apesar de possuírem características similares em termos de estrutura. A implementação de um sistema de BI é uma tarefa difícil, a qual necessita de *hardware*, *software* e demais recursos. A empresa pode complementar o sistema incluindo um *Data Warehouse* (que será abordado adiante neste trabalho) integrado às estruturas de dados, dependendo do volume de dados.

Para Chen, Chiang e Storey (2012), o BI é altamente aplicável, apesar de sua dificuldade, e pode alavancar oportunidades através da análise de dados, necessária a diversas áreas que vão desde *e-commerce* e inteligência de *marketing*, até na política, passando por ciência e tecnologia, áreas de saúde e bem-estar além de segurança pública.

2.2.2. Conceitos acerca da infraestrutura de BI

A infraestrutura de BI é a base que agrega valor aos dados da organização. BI é menos sobre o processo e suas ferramentas, e mais sobre a utilidade da informação. Uma metodologia de BI de sucesso necessita ter seu foco maior na cadeia de valor da informação e menos no desenvolvimento de Tecnologia de Informação. Pesquisadores tem demonstrado que pesquisas em *software* para desenvolvimento de BI não são tão bem sucedidas, e sim a utilização da informação que agregam valor as organizações (LARSON e CHANG, 2016).

A infraestrutura de BI conta com Bases de Dados, com *Data Warehouses* (DW), *Data Marts* (DM), *Hadoop*, plataformas de análise, ferramentas de análise como *Online Analytical Processing* (OLAP), estatísticas e modelos como *Data Mining* (LAUDON e LAUDON, 2016).

De acordo com Vercellis (2009), a infraestrutura de BI é composta por três itens primordiais. Essa estrutura é vista na Figura 2.5:

1. As fontes de dados – onde é necessária coleta e integração de dados. As fontes podem ser sistemas operacionais, que podem incluir documentos, e-mails e dados de fontes externas às organizações. O maior desafio dessa etapa é unificar diferentes fontes de dados.
2. Os *Data Warehouses* e *Data Marts* – aonde armazenam os dados após as ferramentas de ETL (*Extract, Transform, Load* – extração, transformação e carregamento dos dados).
3. As metodologias e ferramentas de BI, nas quais os modelos matemáticos e ferramentas de análise são utilizadas para desenvolver metodologias para as tomadas de decisão.

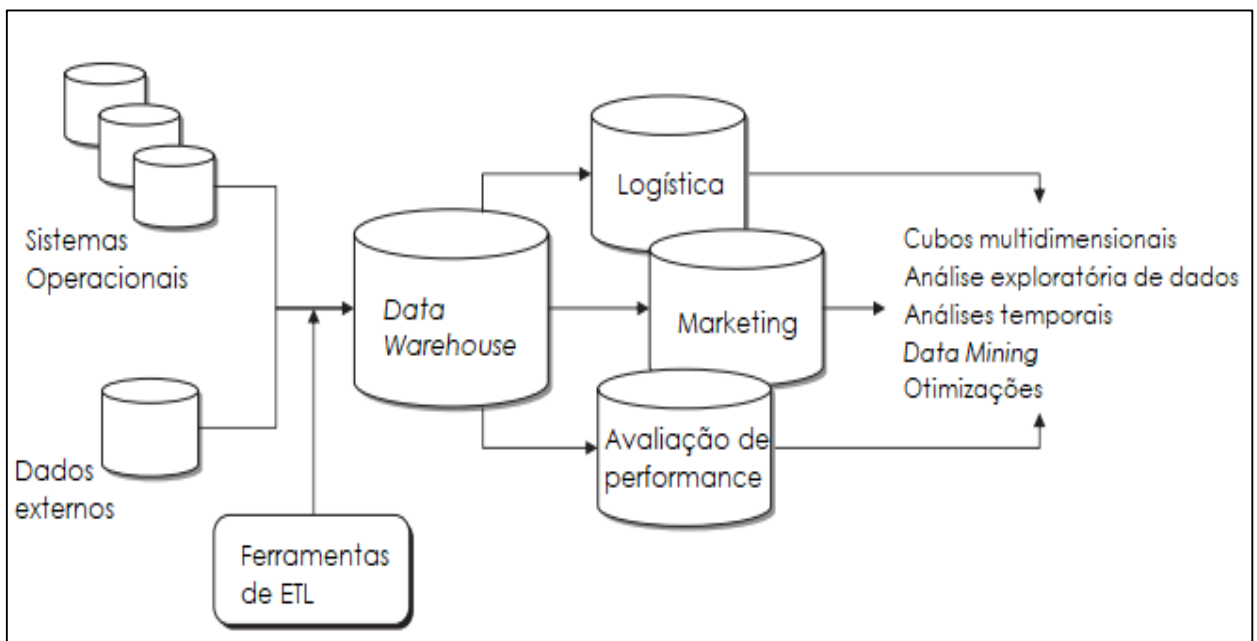


Figura 2.5 - Infraestrutura de BI

Fonte: Vercellis (2009).

As ferramentas de BI auxiliam as decisões tomadas pela inteligência das organizações, analisando os dados da companhia. *Data Warehouses* são utilizados para armazenar uma grande quantidade de dados, enquanto ferramentas de OLAP são aplicadas nos DW para responder perguntas analíticas. O processo de ETL é a base do processo de um DW, com um dispêndio de tempo de cerca de 80% em um projeto de DW (NATH *et al.* 2017). De acordo com

Gallinucci, Golfarelli e Rizzi (2015), o objetivo das ferramentas de BI é alcançado quando a informação estocada no DW podem ser acessadas pelas técnicas de OLAP.

O processo de ETL - *Extract, Transform, Load* – extração, transformação e carregamento dos dados, se refere às ferramentas (*software*) que realizam essas ações no *Data Warehouse*. A “Extração” se dá na coleta de dados internos e externos. A “Transformação” se dá na limpeza e transformação dos dados para melhorar a qualidade deles pela correção de imprecisões, inconsistências e dados faltantes/duplicados. Já o “Load”, se dá no carregamento dos dados transformados no *Data Warehouse* para disponibilizá-los aos analistas e aos SAD (VERCELLIS, 2009).

Já de acordo com Bala, Boussaidb e Alimazighic (2017), o processo de ETL é comumente complexo, pois envolve fontes de dados de qualidade fraca e de grande heterogeneidade. A extração se dá no acesso de várias fontes de dados para escolher aquilo que será analisado. Para ser padronizado, esses dados passam por ações de transformação como, por exemplo, filtros, conversões, limpezas, uniões, divisões, agregações de dados – chamadas “funcionalidades” do ETL; para assim serem carregados nos *Data Warehouses*. O processo e ambiente de ETL, pode ser visto na Figura 2.6.

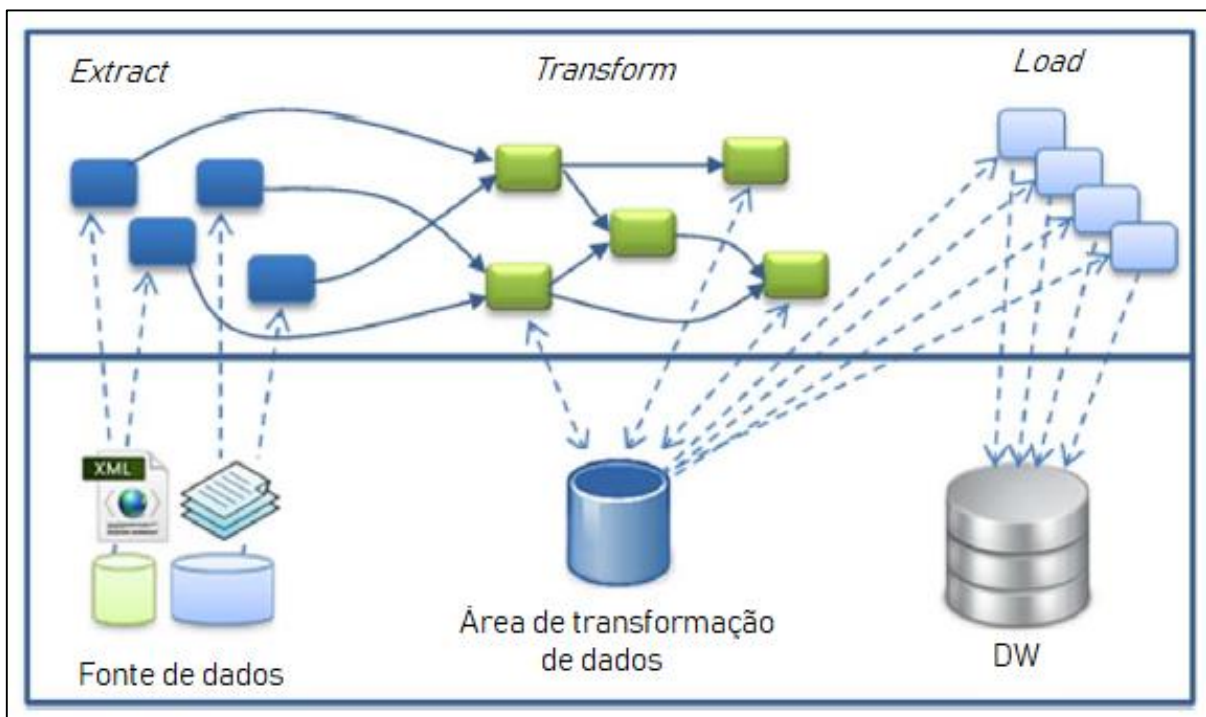


Figura 2.6 – Processo de ETL e seu ambiente.

Fonte: Bala, Boussaidb e Alimazighic (2017).

Theodorou *et al.* (2017), destacam que o processo de ETL é encarregado de enviar frequentemente os dados disponíveis de uma fonte de dados para o *Data Warehouse*. Além disso, as ferramentas de ETL disponíveis para comercialização não fornecem um auxílio automatizado para garantir uma qualidade nos dados, portanto, a maioria do trabalho é manual. O modelamento de uma ferramenta ETL não é uma área estudada com profundidade, seja em meios acadêmicos ou na indústria, logo, não há um modelo padrão de processo de ETL.

De acordo com Laudon e Laudon (2016), um *Data Warehouse* é uma base de dados que armazena dados históricos e atuais que possuem um potencial para os tomadores de decisão de uma companhia se interessarem. Para Stair e Reynolds (2010), o *Data Warehouse* é uma base de dados que possui informações e dados de diversos setores e fontes de dentro de uma organização, contendo dados de todos os processos da empresa, bem como dados de produtos e clientes.

Os dados disponíveis em um sistema de BI de uma organização, em sua maioria, provêm de um *Data Warehouse* (ARNOTT, LIZAMA e SONG, 2017). Um *Data Warehouse* é uma peça central para a arquitetura de um sistema de BI, o qual viabiliza um conjunto de dados integrados para análise (LARSON e CHANG, 2016). Os *Data Warehouses* podem possuir milhões e até centenas de milhões de arquivos de dados. Como esses dados são coletados de diversas fontes e sistemas de produção, uma base de dados histórica acaba sendo construída, para que os analistas possam utilizar delas (STAIR e REYNOLDS 2010).

Pode-se identificar três categorias principais de dados que são alimentados a um DW: dados internos, externos e pessoais. Os dados internos são provenientes de bases de dados de SITs, e demais sistemas operacionais, e são coletados a partir de operações transacionais que acontecem nas operações da organização (um ERP por exemplo, que interliga a organização). Os externos vêm de diversas fontes externas da empresa e os pessoais provém de dados dos próprios tomadores de decisão que, muitas vezes, estão em seus computadores pessoais, em planilhas e demais bases de dados (VERCELLIS, 2009).

Um *Data Mart* (DM), é uma subdivisão de um DW, o qual resume ou foca uma parte dos dados da organização, separando a base de dados para uma porção em particular de usuários (LAUDON E LAUDON, 2016). Um *Data Mart* é mais útil para um grupo de pessoas que objetiva dados mais detalhados. Para o desenvolvimento de um DM, não são necessários

hardware muito avançados, poupando muitos gastos para a organização (STAIR e REYNOLDS, 2010).

Para Laudon e Laudon (2016), após a captura dos dados e sua organização utilizando-se de tecnologias de BI, os dados estão disponíveis para análise (nos DWs), em *software* que geram relatórios, análise de dados multidimensionais e *Data Mining*.

De acordo com Stair e Reynolds (2010), os chamados *Online Analytical Processing* (OLAP) são sistemas que são utilizados para fornecer informações com eficiência, a partir do armazenamento nos *Data Warehouse*. Grandes empresas que desenvolvem esses sistemas são exemplificadas: Microsoft, Cognos, SAP, Business Objects, MicroStrategy, Applix, Infor, and Oracle.

OLAP auxilia a análise de dados multidimensionais, para que seus usuários sejam capazes de operar os mesmos dados de formas distintas, usando múltiplas dimensões. Cada perspectiva da informação (como por exemplo: produto, preço, custo, região) demonstra uma dimensão diferente. Um modelo multidimensional pode ser criado para representar cada dimensão, como um cubo, como representado na Figura 2.7 (LAUDON e LAUDON, 2016).

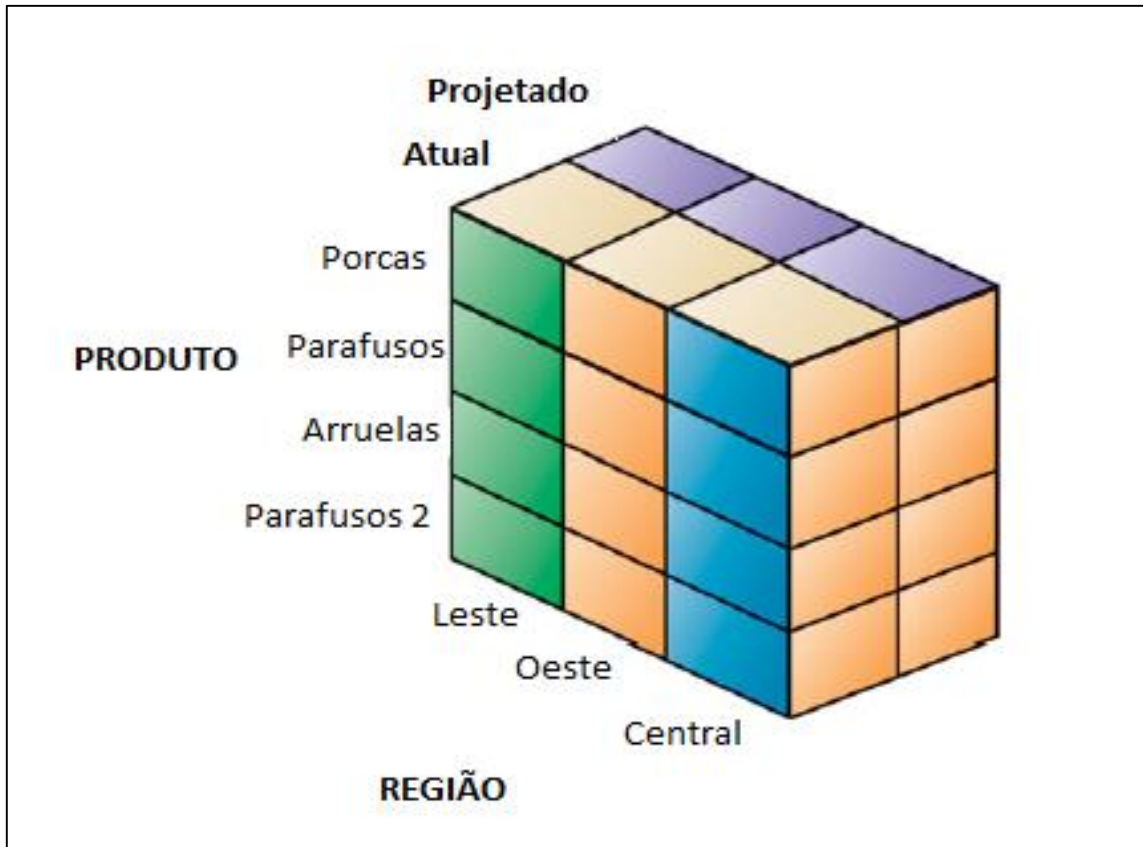


Figura 2.7 - Exemplo de modelo de dados multidimensional

Fonte: Laudon e Laudon (2016).

As técnicas de OLAP são geralmente utilizadas em cubos multidimensionais que armazenam dados estruturados da organização. Os cubos são utilizados para computar dados multidimensionais em documentos usando medidas como palavras-chave e contagem de documentos (GALLINUCCI, GOLFARELLI e RIZZI, 2015). Os cubos são dados extraídos para serem transformados, sendo candidatos perfeitos para integrar um algoritmo de *Data Mining*. (MANSMANN *et al.*, 2014).

O *Data Mining*, é uma ferramenta que fornece conhecimento dos dados de uma organização que não podem ser obtidos de uma simples ferramenta de OLAP. Os tipos de informação obtidas pelo *Data Mining* são: associações, provisões, sequências e classificações. Esses sistemas fazem análises de alto nível em padrões e tendências nos dados, mas também podem ir mais profundamente para fornecer mais detalhes (LAUDON e LAUDON, 2016).

O *Data Mining* é uma ferramenta que envolve a descoberta automática de padrões e relacionamentos de dados dentro de um DW. A análise preditiva é um tipo de *Data Mining* que combina dados históricos com previsões sobre eventos e condições futuras, como por exemplo,

a quantidade de vendas prováveis no futuro (STAIR e REYNOLDS, 2010). A diferença entre OLAP e *Data Mining* pode ser vista na Tabela 2.1.

Tabela 2.1 - *Comparação entre OLAP e Data Mining*

Característica	OLAP	Data Mining
Proposta	Auxilia análise de dados e tomada de decisão	Auxilia análise de dados e tomada de decisão
Tipo de análise suportada	De cima para baixo, análise de dados direcionada à consulta	De baixo para cima, análise de dados direcionada à descoberta
Habilidades do usuário	Deve possuir conhecimento da organização e do contexto dos dados	Deve confiar nas ferramentas de <i>data-mining</i> para validar hipóteses

Fonte: Adaptado de Stair e Reynolds (2010).

A tecnologia de *Data Warehouse*, em conjunto com a OLAP e *Data Mining*, é conhecida não somente como pela sua universalidade e alta performance no processamento de dados, mas também pela sua rigidez e algumas limitações quando lida com dados complexos e semiestruturados. Uma solução para tais limitações é estender as capacidades das tecnologias dominantes existentes, como por exemplo o uso de bases de dados multidimensionais e relacionais (MANSMANN *et al.*, 2014).

2.2.3. Ciclo de vida de um sistema de BI

Para Larson e Chang (2016), o ciclo de vida do BI começa com algumas perguntas: “Quais questões da organização precisam ser respondidas? Quais as fontes de dados serão utilizadas? Quais dados serão usados? ”. Estes questionamentos são necessários para começar a examinar o processo que será feito para criar um sistema de BI, pois este inclui uma série de componentes como Sistemas de Informação, ETL, bases de dados e ferramentas computacionais, além da

infraestrutura de BI, que é o meio o qual os dados da organização serão trabalhados e os indivíduos interagirão com as ferramentas desenvolvidas.

Ainda para Larson e Chang (2016), os ciclos são fases que compreendem a progressão de um conceito até seu fim, e para os sistemas de BI, não é diferente. Os passos são: A Descoberta (*Discovery*) – focada nos pré-requisitos do sistema, o *Design* (ou Projeto) – focado em definir o trabalho a ser realizado com os dados, o Desenvolvimento (*Development*) – focado na criação de bases de dados e códigos, a Implantação (*Deployment*) e a agregação de valor:

1. **Descoberta** – Nesta fase do ciclo, a alta gerência determina os requisitos de informação e questões relacionadas a organização, fornecendo *insights* sobre as fontes de dados a serem coletados, além dos fatos necessários para o desenvolvimento do sistema. A entrega de um sistema funcional de BI tende a ser um processo onde o cliente final possui uma expectativa desde a etapa de Descoberta, pois possui muitos pré-requisitos para executar com eficácia. Transformar dados em informação não é um processo simples e os pré-requisitos não são facilmente definidos, mesmo quando os usuários possuem experiência (LARSON e CHANG, 2016). Para Vercellis (2009), a primeira fase é chamada de “Análise”, onde há exatamente a Descoberta dos objetivos e necessidades da organização, para desenvolver um sistema de BI.
2. **Projeto** – Nesta segunda fase, o maior foco é a modelagem, começando com a infraestrutura do sistema de BI, que conta com pessoal técnico, processos, dados e componentes do projeto. Esta fase também centra em definir os objetivos e estratégias da organização, para alcançar as informações que ela precisa. A arquitetura do sistema de BI inclui como os dados serão coletados, os DWs e DMs que serão implantados, redes e *software* que auxiliarão o sistema de BI. Nesta fase, os modelos criados podem incluir modelos conceituais, lógicos e físicos de dados como, por exemplo, modelos para o processo de ETL (LARSON e CHANG, 2016). A segunda fase, portanto, envolve reconhecer a estrutura que a organização possui, e projetar o sistema. (VERCELLIS, 2009).
3. **Desenvolvimento** – O desenvolvimento do sistema de BI inclui um grande leque de ações e a principal delas é a produção de um sistema que, quando aplicado à organização, possa providenciar informação a partir dos dados, para que os usuários possam analisar. As demais atividades podem incluir ETL e configurar camadas em uma ferramenta de BI, por exemplo (LARSON e CHANG, 2016). Com os detalhes planejados a serem realizados, os modelos matemáticos, e até mesmo um protótipo de

baixo custo, podem ser criados nessa fase, a partir dos dados armazenados nos DW e DM (VERCELLIS, 2009).

4. **Implantação** – Esta fase é uma fase formal e controlada, dado a complexidade que um sistema de BI pode atingir, por fatores como mudanças constantes nas organizações e por depender de muitos fatores independentes. Algumas ações desta fase abrangem a integração de novas funcionalidades e capacidade na produção, por exemplo – focando na estabilidade da produção do sistema de BI (LARSON e CHANG, 2016). Nesta fase, o sistema que foi desenvolvido já pode ser lançado para testes e uso (VERCELLIS, 2009).
5. **Agregação de valor** – a entrega de valor à organização, inclui a estabilização, manutenção, mudanças de gerência e o *feedback* dos usuários do sistema de BI. Sistemas de BI de sucesso, em suma, possuem uma longa vida, e por essa razão, necessitam de mudanças – que impactam as fontes de dados, processos, *upgrades* de *software* além de mudanças na estratégia organizacional. O *feedback* dos usuários é de grande importância, pois são utilizados para entender como a informação está sendo usada, validando o valor que está sendo agregado pela utilização do sistema de BI (LARSON e CHANG, 2016).

Para Chen, Chiang e Storey (2012), em adição às técnicas de aplicação do sistema de BI, as empresas necessitam de conhecimento acerca da organização, além de uma comunicação eficaz entre os usuários, para que a implementação do projeto de BI seja um sucesso. Os departamentos que utilizarão o sistema terão tanto oportunidades quanto desafios em sua aplicação, e também no ensino de usuários para sua utilização.

Na Figura 2.8, pode-se ver as metodologias e processos que podem ser envolvidos em um sistema de *Business Intelligence*, envolvendo pessoas, tecnologia e recursos das organizações:

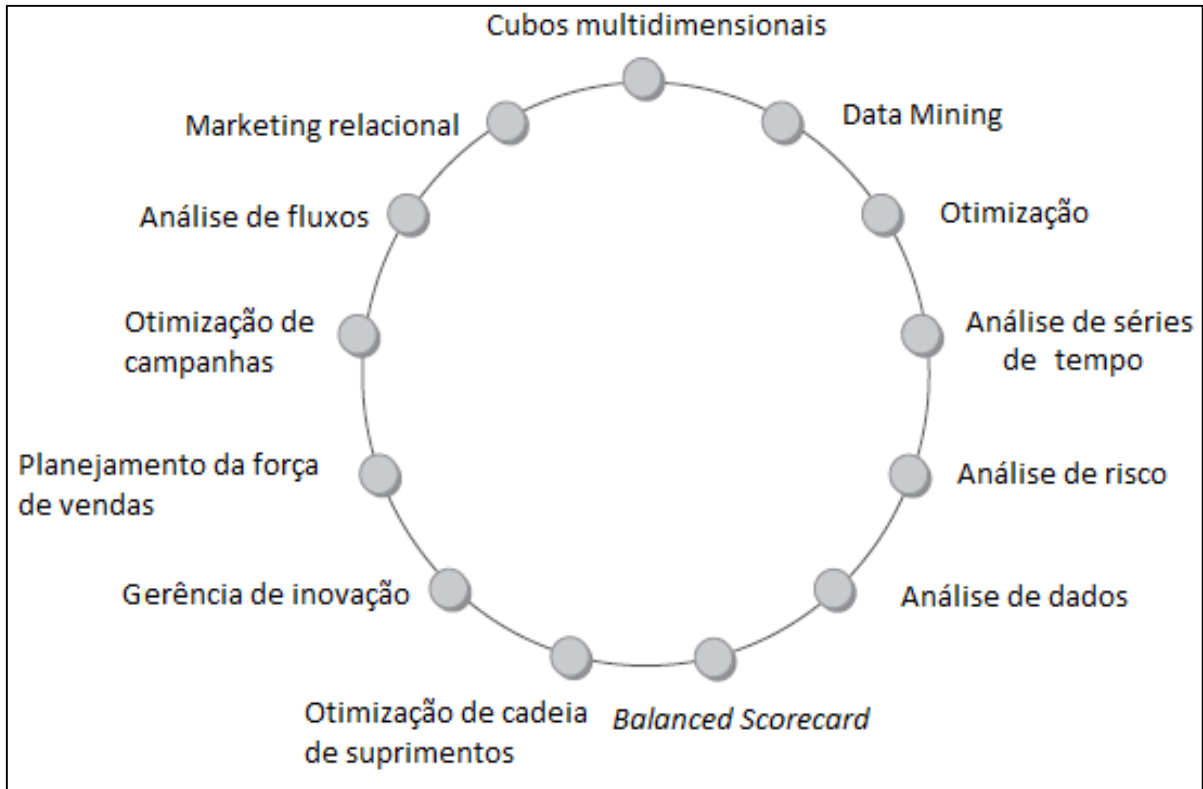


Figura 2.8 - Metodologias e processos envolvidos em um sistema de BI

Fonte: Vercellis (2009).

2.3. Ciência dos Dados - *Data Science*

Para Flath e Stein (2016), com a ascensão da onipresença de dados, o desejo de se ter conhecimento sobre os negócios e o valor dos dados está crescendo. Portanto, a ideia de “análise de dados” descreve a *Data Science* (Ciência dos Dados) dentro de um contexto de uma organização, e essa ideia tem crescido rapidamente dentro delas nos últimos anos.

A ascensão do conceito de *Data Science* é um acontecimento recente e pragmático. Muitos aspectos de diversas organizações são agora potencialmente abertos à coleta de dados. Esta disponibilidade de dados tem levado a um crescente interesse em métodos de extrair informações e conhecimento dos dados para auxílio na tomada de decisão (AMIRIAN, LOGGERENBERG e LANG, 2017).

Chen, Chiang e Storey (2012) destacam que o BI está gerando oportunidades não apenas nas pesquisas, mas para educação. Em 2012, a Universidade de Columbia e a Cidade de Nova Iorque investiram US\$ 80 milhões em um novo centro de *Data Science*, gerando centenas de

vagas de trabalho em 100 *startups* pelos próximos 10 anos. BI é o *Data Science* nos negócios e organizações.

Nos dias de hoje, vive-se em um mundo onde há diversas funcionalidades que são focados em dados. A internet utilizada é formada por serviços de dados e por bases de dados. O papel de um cientista dos dados nunca foi mais importante, gerando oportunidades de criações de produtos de dados, que possuem seu valor dos próprios dados e gerando mais dados como resultado, aumentando ainda mais seu valor (COSTA e SANTOS, 2017).

O *Business Intelligence* pode integrar com a *Data Science* nos negócios para criar impactos positivos na análise de uma grande quantidade de dados, criando *outputs* na visualização e na análise interativa, e a alta gerência pode entender esses *outputs* sem um conhecimento técnico (NEWMAN *et al.*, 2016).

De acordo com Costa e Santos (2017), a *Data Science*, nos últimos anos, tem atraído uma grande atenção, como visto na Figura 2.9, surgindo como um casamento de Estatísticas e de Ciência da Computação. Porém, se assume que as habilidades de um cientista de dados vão além dessas duas áreas.

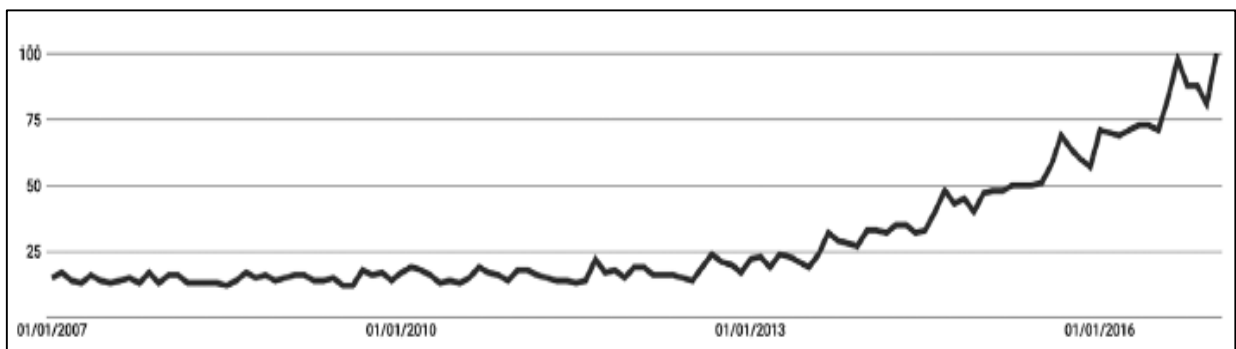


Figura 2.9 - Popularidade do Data Science no Google Trends

Fonte: Costa e Santos (2017).

Os conceitos sobre *Data Science* são abordados nas próximas seções, como as definições e seu processo de desenvolvimento.

2.3.1. Conceitos e definições

Data Science é um campo interdisciplinar o que tem foco em extrair conhecimento de dados de diversas maneiras, sendo a nova geração de conceitos como *Data Mining* e análise de dados. A partir da *Data Science*, a análise de dados é auxiliada por estatísticas e

algoritmos de aprendizado de máquina (*Machine Learning Algorithms*), para produzir produtos de dados ou modelos que funcionam como análises descritivas, preditivas e prescritivas (LARSON e CHANG, 2016).

Data Science é um conceito multidisciplinar, o qual permite a diversos especialistas de múltiplas áreas a trabalharem e estudarem em conjunto. A principal razão do *Data Science* ser tão atraente às organizações é sua associação ao processo, análise, interpretação dos dados (NEWMAN *et al.*, 2016). A área em evolução da *Data Science* combina campos como matemática, estatística, ciência da computação, ciência do comportamento e análise preditiva (HAZEN *et al.*, 2014).

A Figura 2.10, mostra os três pilares da *Data Science*: dados, tecnologia e pessoas. “Dados” são todos os dados utilizados, estruturados ou não, “Tecnologia” são as utilizadas para processar os dados e as “Pessoas” podem incluir os cientistas de computação, estatísticos, cientistas de dados e analistas de negócios envolvidos com o *Data Science* (SONG e ZHU, 2016).

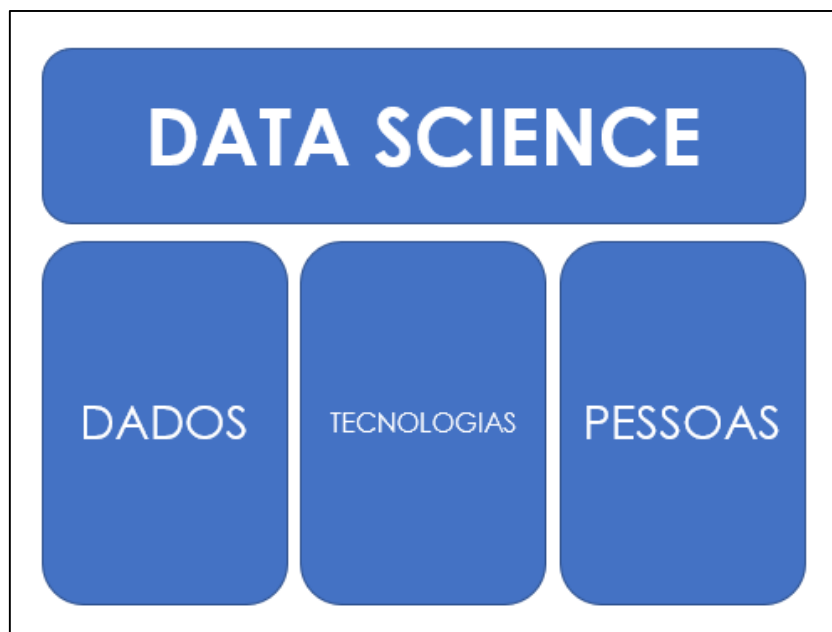


Figura 2.10 – Os três pilares da *Data Science*

Fonte: Song e Zhu (2016).

Para Schutt e O’Neil (2014), *Data Science* é uma série de ações de melhoria utilizadas nas companhias para lidar com uma ampla gama de problemas que podem ser solucionados com dados, e por isso podendo merecer o nome de ciência, apesar de muitas vezes ser

tratado com ânímos exagerados, o que deve ser evitado. Os resultados de um produto de *Data Science* podem ser:

- Análise exploratória de dados;
- Visualizações (relatórios);
- *Dashboards* e métricas;
- Resultados de conhecimento da organização;
- Tomada de decisão baseada nos dados;
- Engenharia de dados e *Big Data*;
- Pesquisas e investigações;
- Otimizações diversas;
- Encontro de correlações com dados;
- Dentre muitos outros produtos.

Segundo Provost e Fawcett (2014), a *Data Science* abrange uma série de princípios e técnicas que auxiliam a análise automatizada de dados, com o objetivo principal de melhoria na tomada de decisão. A tomada de decisão baseada nos dados se refere a prática de basear as decisões na análise dos dados, e não apenas por intuição.

Para Amirian, Loggerenberg e Lang (2017), *Data Science* é a ciência que usa de métodos computacionais para identificar e encontrar padrões nas séries de dados. Seu objetivo principal se dá no conhecimento adquirido que afeta as decisões, fazendo estas últimas mais consistentes e eficientes, ajudando os tomadores de decisão.

Ainda para Amirian, Loggerenberg e Lang (2017), os dados são necessariamente uma medida de informação histórica e, por definição, a *Data Science* analisa dados históricos. Porém, os dados utilizados na *Data Science* podem ser coletados há alguns anos ou há alguns milissegundos, com um processo contínuo. Portanto, seu processo pode ser em tempo real ou próximo do tempo real.

De acordo com Newman *et al.* (2016), a *Data Science* nos negócios inclui os conceitos dos 5Vs da *Big Data*: Volume, Velocidade, Variedade, Veracidade e Valor. Volume, se refere à quantidade e volume propriamente dito de dados que são processados. Velocidade significa a rapidez com que os dados são gerados para análise nos negócios, e variedade pois existem inúmeros tipos de dados que os negócios trabalham. Veracidade parte do pressuposto que a análise correta desses dados resulte em dados que reflitam a realidade,

aumentando o desenvolvimento do negócio, e por fim, Valor é a agregação de valor que a utilização de um modelo de *Data Science* traz as organizações.

O papel da *Data Science* se tornou muito importante para as organizações. Primeiramente, a *Data Science* proporcionou as companhias a adquirir e analisar dados de suas operações, além de sua performance e estratégia. Em segundo lugar, permitiu a elas a melhoria de suas operações e serviços, baseado nos *outputs* das análises. Em terceiro lugar, o negócio pode ampliar a qualidade de suas previsões, para os tomadores de decisão planejarem suas estratégias (NEWMAN *et al.*, 2016).

2.3.2. Cientista de dados

O termo “Cientista de Dados”, aquele que lida com a *Data Science*, foi estabelecido para definir as pessoas que trabalhavam com aplicações de dados, e que tinham um impacto nas organizações, descobrindo e entendendo melhor questões de negócios a partir dos dados, explorando-os de uma maneira científica (COSTA e SANTOS, 2017).

Conforme Schutt e O’Neil (2014), o Cientista de Dados é um cientista que pode provir de diversos campos, de ciências sociais até mesmo biólogos, que trabalham com uma grande quantidade de dados e necessitam trabalhá-los computacionalmente, com seus respectivos problemas, estruturas, tamanhos e complexidades, solucionando problemas do mundo real. Um Cientista de Dados deve estar alinhado à estratégia da organização em que está inserido, desde a engenharia e infraestrutura de dados da empresa até preocupações acerca de tomada de decisões dos gestores.

Segundo Costa e Santos (2017), para melhor distinguir os termos de “Cientista de Dados” e “Analista de Dados”, é considerado que o papel do Cientista de Dados seja uma evolução do papel de Analista de Dados, pois possui habilidades em negócios e comunicações, para superar desafios dentro das organizações, agregando valor a elas. Além disso, o Cientista de Dados combina conhecimentos de estatísticos, Cientistas da Computação, porém se encaixa melhor sob do leque dos Sistemas de Informação.

De acordo com Song e Zhu (2016), os principais papéis do Cientista de Dados são: Extrair conhecimento dos dados para resolver problemas dentro das organizações; realizar as perguntas corretas e necessárias para alinhar seus resultados com os objetivos do negócio em questão; identificar os dados corretos a se usar ou reutilizar; selecionar as melhores

tecnologias e ferramentas; analisar, avaliar e visualizar os dados; e, por fim, ajudar a tomada de decisão relacionada aos dados, entre outras funções menores.

2.3.3. Processo de desenvolvimento do *Data Science*

De acordo com Larson e Chang (2016), a análise rápida de dados, bem como o *Data Science* possuem sua origem devido ao *Big Data*, e é considerável afirmar que suas formas já sejam existentes há algum tempo – Análise Visual é tida como um sinônimo da análise rápida de dados, assim como *Data Mining* seria um sinônimo para *Data Science*, porém, estas são novas versões para aquelas, que são métodos conhecidos.

Enquanto a *Big Data* foi cada vez mais sendo utilizada em análises, ferramentas de visualização puseram-se em prática na mineração e exploração de dados brutos, auxiliando a Análise Exploratória de Dados no processo de *Data Science*. Portanto, o escopo da *Data Science* é adquirir dados rapidamente para serem analisados, com a dependência do problema a ser avaliado (LARSON e CHANG, 2016).

Seguindo uma abordagem estruturada de *Data Science*, como um processo, a chance de sucesso em um projeto é maximizada, e o custo minimizado. Também se faz possível projetar em conjunto com uma equipe, onde cada membro foca no que cada um pode agregar de melhor (CIELEN, MEYSMAN e ALI, 2016).

Conforme Schutt e O’Neil (2014), o processo de desenvolvimento do método *Data Science* se dá em sete etapas (que podem ser condensadas em seis):

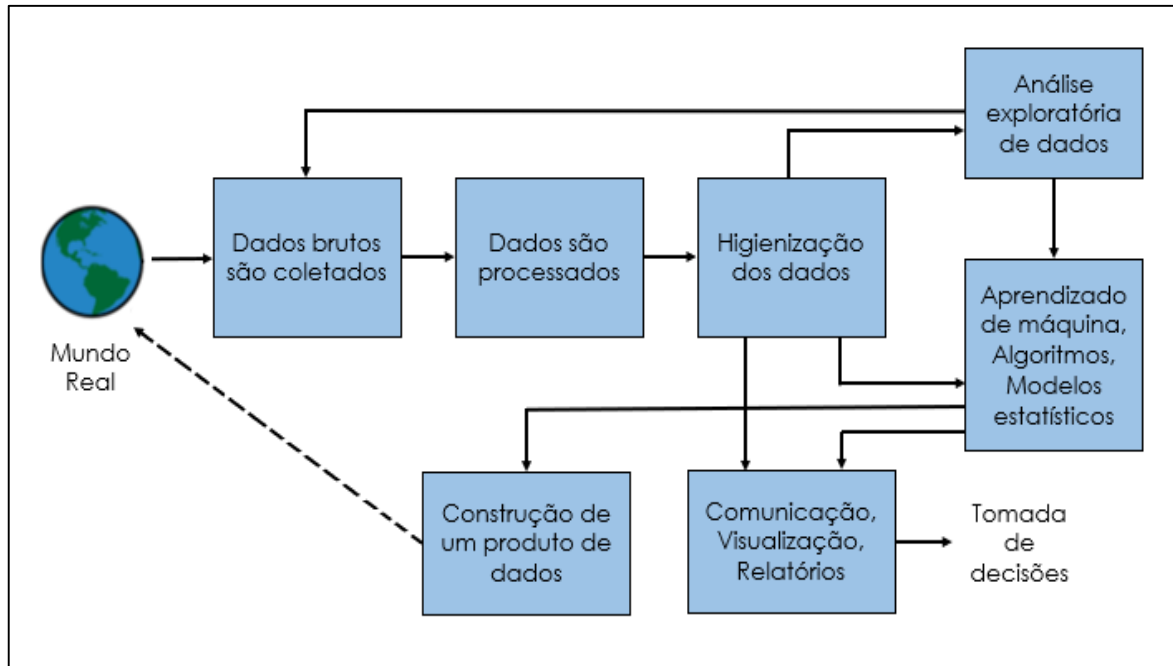


Figura 2.11 – O processo da Data Science

Fonte: Schutt e O’Neil (2014).

Tem-se o mundo real em primeiro lugar, onde as pessoas realizam inúmeras atividades. Em seguida, tem-se os dados brutos, os quais devem ser processados primeiramente, para que seja aplicada uma higienização para que enfim, possam ser analisados. Após os dados serem higienizados, é necessário o passo de Análise Exploratória de Dados (EDA – *Exploratory Data Analysis*), onde pode-se concluir que os dados podem ou não estar de fato higienizados, com valores duplicados, dados perdidos ou *outliers* discrepantes – incorrendo no retorno da etapa de coleta bruta e dados, para se verificar a consistência dos dados (SCHUTT e O’NEIL, 2014).

Para Larson e Chang (2016), ferramentas de visualização de dados são usadas para dar suporte na coleta de dados brutos para a Análise Exploratória de Dados no processo de *Data Science*, resultando em novas funcionalidades como gráficos complexos e quadros que conectam diferentes tipos de dados provindos de fontes heterogêneas.

Segundo Schutt e O’Neil (2014), tem-se em seguida o *design* do modelo: é utilizado algum algoritmo, dependendo do tipo do problema que é buscada a solução. Dessa maneira se pode interpretar, visualizar, comunicar e gerar relatórios para a gerencia de uma organização, para se tomar decisões. Alternativamente, é possível criar um “produto de dados” (como por exemplo

um algoritmo de ranqueamento de busca), retornando aos usuários do mundo real. Os passos são detalhados a seguir:

1. **Coleta de dados brutos** - O primeiro ato deve ser o acesso aos dados que tenham uma certa qualidade e relevância, os quais estão disponíveis na organização analisada. Muitas empresas possuem dados já higienizados, e armazenados em *Datas Warehouses* ou bancos de dados, porém, por vezes o acesso a esses dados pode ser um desafio para o cientista de dados, pois muitas companhias entendem o valor e a sensibilidade dos dados, possuindo, portanto, políticas de acesso a eles (CIELEN, MEYSMAN e ALI, 2016).

Conforme Larson e Chang (2016), a coleta de dados brutos, sem o conhecimento de suas estruturas ou significados, é possível devido às novas tecnologias emergentes. Depósitos de dados criados recentemente como, por exemplo, os “lagos de dados” (*data lakes*), estão emergindo, disponibilizando um armazenamento e um poder de processamento grande, ajudando a análise de uma grande quantidade de dados não estruturados.

Para Amirian, Loggerenberg e Lang (2017), a descoberta de informação nas bases de dados inclui atividades multidisciplinares, para descobrir, validar, deixar útil e entender o conhecimento que será proveniente dos projetos.

2. **Processamento e higienização de dados** - O segundo passo é a higienização/processamento de dados, que são um subgrupo do processo de *Data Science* que foca em remover os erros e vícios para que os dados se tornem uma representação eficiente do processo de origem. Há dois tipos de erros existentes: o primeiro consiste em um erro de interpretação, no qual o dado é suposto; já o segundo erro é um erro de inconsistência entre as fontes de dados ou aos dados padronizados da companhia/organização (CIELEN, MEYSMAN e ALI, 2016).

Portanto, a higienização é feita com junções, arredondamentos, descartes e transformações. Para isso são utilizadas algumas ferramentas e *software* como o R, Python ou SQL (SCHUTT e O’NEIL, 2014). De acordo com Amirian, Loggerenberg e Lang (2017), outras ferramentas também podem ser utilizadas, como por exemplo Microsoft Excel, além de linguagens como Java, VBA, MATLAB, C#, Perl, e Scala, bem como múltiplos gerenciadores de base de dados.

Muitas das análises necessitam de dados elencados em um formato bidimensional, o qual é formado por linhas e colunas, e cada linha representa uma unidade de análise – diferenciando de observações e medidas. Os dados são coletados de fontes distintas, e que na fase da preparação e higienização, são transformados em unidades de análise (AMIRIAN, LOGGERENBERG e LANG, 2017)

3. **Análise exploratória de dados/visualização** - No processo de *Data Science*, a análise rápida e a visualização são finalizadas como parte da fase de Análise Exploratória de Dados (EDA – *Exploratory Data Analysis*), onde a análise descritiva é utilizada como identificação de parâmetros e para elucidar a relação entre variáveis, para se utilizar no modelo analítico. Se o resultado dessa análise exploratória resultar em um produto de BI, como um *dashboard*, no final do processo este é validado (LARSON e CHANG, 2016).

De acordo com Schutt e O’Neil (2014), a EDA é uma parte crucial para o processo de *Data Science*. Os principais artifícios utilizados pela EDA são os gráficos e resumos estatísticos. Em termos genéricos, a EDA é um método sistemático no qual há transformação de variáveis, procura de relação entre variáveis, utilização de gráficos e geração de dados estatísticos, pelo menos alguns básicos, como média, mínima, máxima, além da identificação de *outliers*.

As técnicas de visualização utilizadas nessa etapa do processo vão de gráficos simples como histogramas ou gráficos de linha, até diagramas mais complexos. Por vezes é mais vantajoso que se crie gráficos compostos de vários gráficos simples para que compreenda os dados de uma maneira mais clara (CIELEN, MEYSMAN E ALI, 2016). Mais do que um leque de ferramentas, a EDA é uma *mindset*, na qual se quer entender os dados, suas formas e tentar conectar o entendimento ao processo que gerou os dados (SCHUTT E O’NEIL, 2014).

Para Newman *et al.* (2016), as análises podem ajudar os negócios a serem mais competitivos, devido ao grande número de *outputs* que geram, como análises visuais, gráficos, relatórios e demais produtos. As tecnologias que a envolvem incluem o *Business Intelligence* e visualizações.

4. **Criação do modelo** - Para Cielen, Meysman e Ali (2016), a partir do momento em que se possui o conjunto de dados higienizados e com um bom entendimento acerca deles,

o próximo passo a se dar é a modelagem do modelo de *Data Science*, com o objetivo de se fazer melhores previsões, classificações de objetos de pesquisa e ganhar conhecimento do sistema como um todo, com o auxílio de ferramentas estatísticas e de *Data Mining*.

Segundo Schutt e O'Neil (2014), o modelo de *Data Science* é produzido em duas fases: a modelagem para análise do *Data Science* e a modelagem de dados para descrevê-los em análise rápida, sendo esta última uma análise descritiva, preditiva e prescritiva, utilizando-se de algoritmos automatizados, como por exemplo, regressões e classificações.

O modelo pode ser baseado em algoritmos como *Naïve Bayes*, ou outro qualquer, dependendo do tipo do problema a ser analisado e solucionado, o qual pode ser um problema de previsão, classificação ou apenas um problema básico de descrição (SCHUTT e O'NEIL, 2014).

5. ***Interpretação dos dados/implementação dos resultados*** - Após a etapa de modelagem, se pode visualizar, interpretar, comunicar e gerar relatórios para a gerência de uma organização, para a tomada de decisões e também, em meio acadêmico, para a publicação de artigos (SCHUTT e O'NEIL, 2014).

Para Larson e Chang (2016), após a validação dos dados, os sistemas, modelos de análise, *dashboards* e sistemas originados, bem como qualquer outra ferramenta de visualização tem pouco valor a não ser que sejam de fato utilizadas. Esses produtos finais devem ser acrescentados no ambiente de produção, para prover novas funcionalidades, como em uma implementação de *Business Intelligence*.

Segundo Cielen, Meysman e Ali (2016), os valores produzidos no modelo poderão ter a necessidade de serem repetidos, por esse motivo devem ser automatizados, não implicando que devam ser refeitos a todo momento. Por vezes, basta a implementação de planilhas de Excel ou aplicativos que possam atualizar os relatórios automaticamente.

6. ***Criação de um produto de dados*** - Conforme citado por Schutt e O'Neil (2014), alternativamente à geração de relatórios para visualização e comunicação dos resultados, é possível criar um “produto de dados”, retornando os dados aos usuários

do mundo real, como por exemplo um algoritmo de ranqueamento de busca, ou um classificador de *spam*, para ser comercializado, onde os usuários podem interagir com o produto, diferentemente de ferramentas estatísticas.

O modelo apresentado necessita da intervenção do ser humano (o cientista de dados), seja para a coleta de dados, seja para as decisões que serão tomadas com suas saídas, além de ser envolvido com um nível alto de todo o processo (SCHUTT e O'NEIL, 2014).

De acordo com Schutt e O'Neil (2014), o processo o qual o cientista de dados está inserido e os passos que ele deve seguir é visto como no exemplo da Figura 2.12:

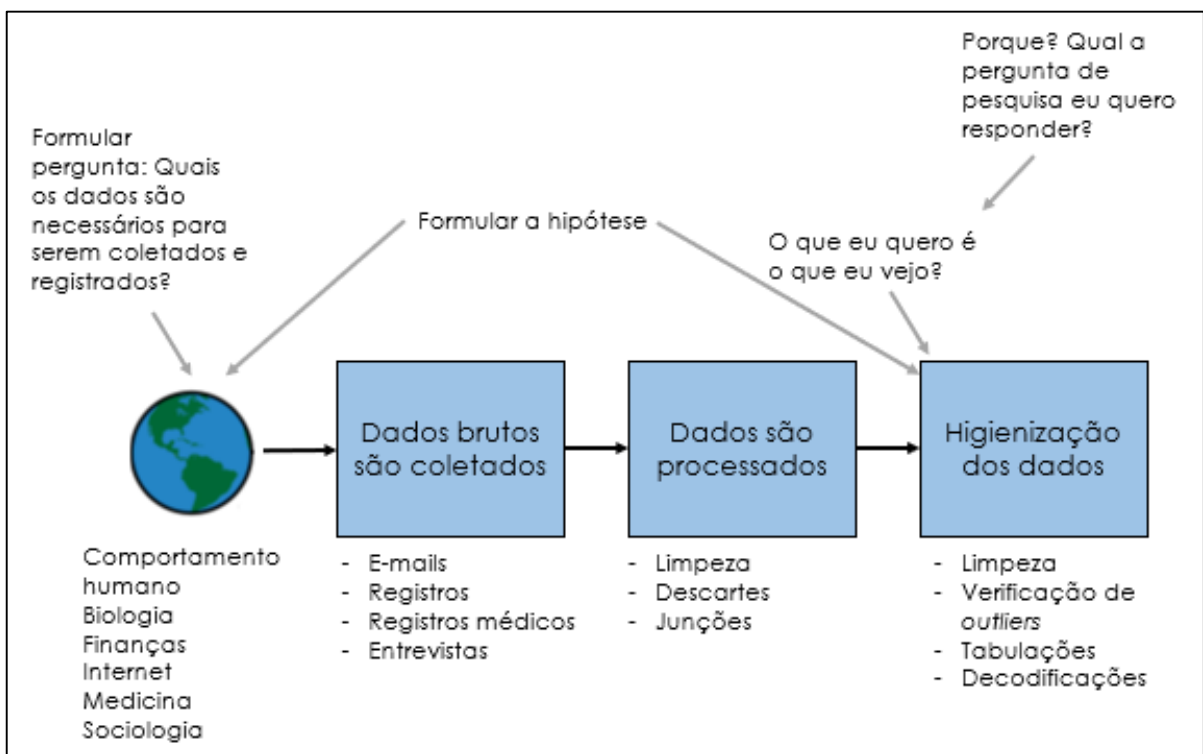


Figura 2.12 - Processo realizado pelo cientista de dados - exemplo

Fonte: Adaptado de Schutt e O'Neil (2014).

De acordo com Song e Zhu (2016), as organizações podem adquirir computadores, servidores e ferramentas precisas para processar os dados, porém as habilidades do ser humano não podem ser substituídas, sendo uma peça-chave para os desafios da era do *Big Data*.

3. METODOLOGIA DE PESQUISA

O objetivo primordial de uma pesquisa científica é encontrar respostas para problemas de um sistema, através da utilização de métodos científicos. A pesquisa científica necessita ser concisa, original e coerente. Sendo assim, a definição desse tipo de pesquisa é o desenvolvimento de novos conhecimentos, a partir de exigências científicas existentes (GIL, 1999).

Neste capítulo são abordados, portanto, os pontos relativos à metodologia de pesquisa empregada, detalhando o objeto de estudo, tipo de abordagem e método quantitativo utilizado, para que haja o desenvolvimento de novos conhecimentos, a partir do método empregado.

3.1. Características e procedimento da Pesquisa

Com fim de classificar a pesquisa realizada nesse trabalho, foram usados os conceitos apresentados por Turrioni e Mello (2012), conforme sua natureza, objetivo e abordagem. Assim sendo, essa pesquisa foi classificada da seguinte forma:

- Natureza Aplicada, uma vez que se espera que o modelo proposto possa ser aplicado e implementado, como forma de validação e também para utilização do mesmo, na organização escolhida como objeto de estudo, visando a solução de problemas imediatamente após o desenvolvimento do modelo.
- Objetivo Normativo, pois foi desenvolvida uma proposta de modelo de *Business Intelligence* no contexto da *Data Science*, sistematizando e preenchendo desta maneira, a lacuna existente na literatura referente à carência de modelos com essa perspectiva, utilizando os conceitos concomitantemente, trazendo também um benefício na estratégia de relativas a um problema específico do objeto de estudo. No caso da pesquisa normativa, se procura a otimização da solução do problema estudado (BERTRAND e FRANSOO, 2002).
- Abordagem Quantitativa, pois o modelo proposto será um modelo quantitativo, de acordo com o “modelo de pesquisa de modelagem quantitativa em gestão de operações” adaptado de Mitroff *et al.* (1974) e Arenales *et al.* (2007) e aplicando os

conceitos de BI e DS, quantificando, transformando e traduzindo as informações para classificá-las e assim poder analisá-las.

Em relação ao procedimento de pesquisa, foi utilizado o conceito de modelagem adaptado de Arenales *et al.* (2007) e Mitroff *et al.* (1974), que consiste em uma metodologia de pesquisa de modelagem quantitativa em gestão de operações, e seu processo é descrito conforme Figura 3.1.

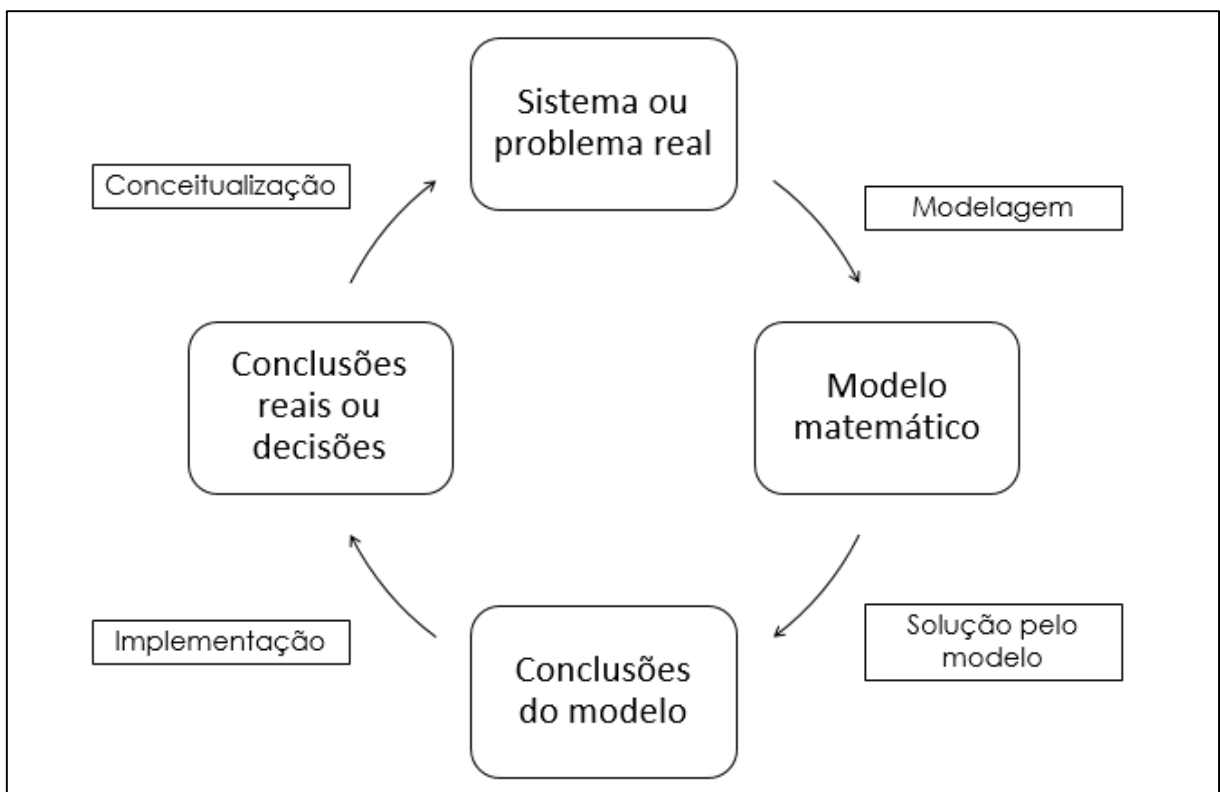


Figura 3.1 – Processo de modelagem

Fonte: Adaptado de Arenales *et al.* (2007) e Mitroff *et al.* (1974).

Neste modelo, a abordagem operacional de pesquisa possui quatro etapas: “Conceitualização”, “Modelagem”, “Solução pelo modelo” e “Implementação”. Na primeira etapa, a de “Conceitualização”, o pesquisador realiza decisões acerca das variáveis que devem ser acrescentadas no modelo e cria o modelo conceitual do problema sob estudo. No segundo passo, de “Modelagem”, o pesquisador de fato cria o modelo quantitativo, determinando as relações causais entre as variáveis. Na etapa de “Solução pelo modelo”, a matemática desempenha um papel determinante na maioria das vezes. Por fim, na etapa de “Implementação”, os resultados do modelo são realizados e pode-se começar um novo ciclo (BERTRAND e FRANSOO, 2002).

Um ciclo de pesquisa pode ainda começar e terminar em qualquer uma das etapas do ciclo, dando ao pesquisador o conhecimento de partes específicas do processo de solução de problemas que ele deseja e, conseqüentemente, nos objetivos da pesquisa que ele procura (BERTRAND e FRANSOO, 2002).

Utilizar-se de modelos ajuda o pesquisador a melhorar seu entendimento para a identificação de tais problemas e desenvolver estratégias afim de solucioná-los, e melhorar a tomada de decisão (MORABITO NETO e PUREZA, 2012). Os modelos ajudam a alta gerência na tomada de decisões em diferentes aspectos das organizações (FAYOUMI e LOUCOPOULOS, 2016).

O modelo proposto neste trabalho foi criado seguindo o método adaptado a partir do conceito de modelagem definido por Arenales *et al.* em 2007 e Mirtroff *et al.* em 1974, passando por cada uma de suas etapas, que são detalhadas a seguir. No Capítulo 4, foi detalhado cada passo da metodologia concomitantemente com os conceitos definidos no Capítulo 2. Dessa forma, o Modelo Proposto foi criado, gerando o produto trabalhado no Capítulo 5 deste trabalho.

1. **Conceitualização:** O modelo baseado pela pesquisa empírica não pode ser realizado sem uma ênfase sistemática para identificação do processo real da operação a ser estudada – o modelo conceitual do sistema. A Conceitualização define as variáveis mais relevantes do sistema a ser estudado bem como a natureza de seus relacionamentos e medidas, e o pesquisador define o alcance do problema e do modelo (BERTRAND e FRANSOO, 2002).

O modelo conceitual a ser trabalhado nessa dissertação teve o modelo desenvolvido no *software* Microsoft Excel, após a coleta, limpeza e análise de dados, de acordo com a perspectiva da *Data Science*, seguindo seus passos de desenvolvimento, que se faz como fase de Conceitualização e preparação para o modelo, além do problema a ser analisado, envolvendo os passos de descoberta e projeto do ciclo de vida do BI.

Nesse momento, os dados necessários foram catalogados, verificados em questão a sua necessidade, coletados e armazenados. A partir dessas ações, verifica-se a identificação do processo real, definindo as variáveis mais importantes.

2. **Modelagem:** A fase de Modelagem e da Solução pelo Modelo requerem habilidades de análise, para que o pesquisador deve selecionar quais variáveis devem entrar no modelo. O pesquisador constrói nesta fase, de fato, o modelo quantitativo (MORABITO NETO e PUREZA, 2012).

A modelagem e sua solução foram desenvolvidas no *software* Microsoft Excel, o qual serviu para trabalhar os dados e se retirar as informações necessárias na aplicação, na fase de implementação – utilizando-se dos conceitos de BI e de *Data Science* para criação do modelo a ser utilizado.

O Modelo Proposto, nessa fase, passou pela higienização dos dados, além da organização dos mesmos. Nesse momento, se fez necessária a análise dos dados, resultando em um produto de visualização de dados, permitindo já nessa fase, a ter-se um produto utilizável pelos usuários do objeto de estudo.

3. **Solução pelo Modelo:** Nessa etapa, a habilidade de retirar conclusões e decisões a partir do modelo elaborado é necessária, já encaminhando para a implementação – o modelo é resolvido utilizando o método matemático (MORABITO NETO e PUREZA, 2012).

O Modelo Proposto desenvolvido em Microsoft Excel foi analisado para perceber se de fato está de acordo com o exigido e se está mostrando as informações necessárias ao objeto estudado. Com essa interpretação do resultado alcançado, já se pode preparar para a fase seguinte, de Implementação, onde foi produzido um produto final, com *dashboard* e informações relevantes, definidos na primeira fase de Conceitualização.

O Modelo Proposto, portanto, foi “solucionado”, a partir das fases anteriores, em conjunto com o *software* escolhido, o tratamento e operacionalização dos dados convertidos em informações relevantes à organização estudada.

4. **Implementação:** Para Bertrand e Fransoo (2002), adaptado de Mitroff *et al.* (1974), um modelo representa sempre um conceito da realidade que não inclui todos os elementos dessa realidade. A implementação é a habilidade de ter um impacto social com o modelo sendo utilizado; é a habilidade, de fato, de ter os resultados sendo aplicados. Geralmente são os gestores quem podem julgar se os resultados são de

fato eficientes ou não. Necessita de tempo e é a fase mais difícil, pois precisa de colaboração entre diversas atividades e pessoas, além disso, se houver erros nas fases anteriores, acabam refletindo nesta fase. Para Amirian, Loggerenberg e Lang (2017), a chave para ter-se um modelo eficiente é a habilidade do modelo ter um impacto positivo na tomada de decisões pelos gestores da organização que teve o problema modelado e desenvolvido.

Nesta fase, o modelo desenvolvido nas fases anteriores foi aplicado ao objeto estudado, tendo sua publicação feita com transparência no sítio eletrônico da Universidade.

Nos itens seguintes deste Capítulo são apresentadas as atividades referentes à Figura 3.1, trazendo “O Modelo Proposto” em conjunto com os conceitos de BI e DS, definidos no Capítulo 2. Adaptado dos moldes do conceito de modelagem definido por Arenales *et al.* (2007) e Mitroff *et al.* (1974), os quatro passos são utilizados, e no Capítulo 5, há a aplicação do modelo proposto, como forma de validação do mesmo, o qual será aplicado no objeto de estudo.

3.2. Objeto de estudo

3.2.1. Universidade Federal de Itajubá UNIFEI

Segundo o sítio eletrônico oficial da Universidade Federal de Itajubá – UNIFEI, a instituição foi fundada em 23 de novembro de 1913, com o nome de Instituto Eletrotécnico e Mecânico de Itajubá – IEMI. O então IEMI, foi fundado pelo advogado Theodomiro Carneiro Santiago, sendo a décima Escola com ensino superior em Engenharia do Brasil.

Passando por diversos nomes e modificações durante os anos, a efetivação do projeto de se transformar a escola em universidade veio em 24 de abril de 2002, através da sanção da lei número 10.435, pelo presidente da República, Fernando Henrique Cardoso, além de sua expansão, com a inauguração do *campus* fora de sede, na cidade de Itabira-MG.

Ainda de acordo com o site oficial da UNIFEI, a instituição possui 656 servidores, entre servidores técnico-administrativos e docentes; conta com 34 cursos de graduação, 21 cursos de pós-graduação e cerca de 5000 alunos matriculados no ano de 2016.

3.2.2. Diretoria de Contabilidade e Finanças - DCF

A Diretoria de Contabilidade e Finanças – DCF, de acordo com seu regimento interno disponibilizado no site oficial da Universidade Federal de Itajubá – UNIFEI, é um departamento vinculado à Pró-Reitoria de Administração, e tem por meta a coleta de dados financeiros, sua organização e gestão, bem como a execução e controle de atividades referentes à contabilidade e execução orçamentária e financeira da UNIFEI – em consonância com a legislação vigente no Brasil.

É de competência da DCF prestar informações sobre toda a execução orçamentária, financeira e contábil da Universidade. De acordo com a Instrução Normativa 2 do Ministério do Planejamento, Desenvolvimento e Gestão, de 06 de dezembro de 2016:

“[...] Com o fim de salvaguardar a transparência administrativa, nos termos da Lei nº 12.527, de 18 de novembro de 2011, o órgão ou entidade deverá disponibilizar, mensalmente, na seção específica de acesso à informação de seu sítio na Internet, a ordem cronológica de seus pagamentos, bem como as justificativas que fundamentam a eventual quebra da ordem. [...]”

Portanto, todas as informações provenientes dos relatórios devem ser publicadas no sítio eletrônico da Universidade. O sistema que provém os dados referentes à pesquisa, é o sistema do Governo Federal – Secretaria do Tesouro Nacional (STN), chamado SIAFI - Sistema Integrado de Administração Financeira do Governo Federal. Este sistema integrado, de acordo com o site oficial do Tesouro Nacional, é o mecanismo mais relevante empregado para registro, acompanhamento e controle da execução orçamentária, financeira e patrimonial do Governo Federal.

Ainda de acordo com seu sítio eletrônico oficial, a STN desenvolveu o SIAFI em conjunto com o SERPRO (Serviço Federal de Processamento de Dados - empresa pública de desenvolvimento de Tecnologias de Informação), implementando-o em 1987, solucionando inúmeros problemas que até então existiam, com a gestão de informações e dados que competiam ao financeiro e orçamentário do País. Seus objetivos circulam desde a provisão de mecanismos de controle diário das áreas citadas, até permitir o acesso a informações por todos os níveis da Administração Pública Federal e auditorias internas e externas aos órgãos.

Apesar da existência do software governamental Tesouro Gerencial, este é de difícil acesso, necessitando um treinamento ou cursos específicos, gerando gastos para a

Universidade. Utilizando-se de um *software* de fácil acesso e *user-friendly* (amigável ao usuário), facilita-se o acesso às informações internas ao departamento, evitando gastos.

O *software* Microsoft Excel foi a ferramenta escolhida para inserção dos dados de entrada, processamento e saída de dados. De acordo com o site oficial da Microsoft, empresa detentora do *software*, ele é considerado um *software* de planilhas eletrônicas, o qual é utilizado para criação, cálculos e análise de dados tabulados em planilhas, além de trabalhar com a linguagem VBA e a ferramenta de OLAP, criando gráficos e relatórios. A ferramenta é detalhada na criação do modelo, que foi produzido, aplicado e desenvolvido, de forma a utilizar-se da ferramenta e seus componentes gráficos e interfaces.

A DCF é subdividida em 3 setores: Contabilidade, Orçamento e Finanças, e a aplicação dos métodos foi, em específico, na divisão Financeira da referida diretoria, onde existe a necessidade de obtenção, transformação e disseminação das informações, de maneira clara e eficaz. Todos os dados obtidos no SIAFI têm como ponto de partida as ações realizadas dentro deste departamento, sejam pagamentos de fornecedores até impostos quitados ao governo ou outro ente da federação.

4. O MODELO PROPOSTO

Nesta seção desta dissertação, é demonstrada a aplicação do método, a partir dos conceitos formulados na seção de fundamentação teórica – colocando-os em prática e em conjunto. No capítulo seguinte, é abordada a aplicação do modelo criado neste capítulo, utilizando-se dos passos definidos para a conclusão do trabalho.

A presente dissertação segue uma abordagem de pesquisa quantitativa, criando um modelo computacional de *Business Intelligence* na perspectiva da *Data Science*, o qual é utilizado para coleta e análise de informações, com o objetivo de representar de maneira adequada as principais realidades de pagamento de fornecedores da Universidade Federal de Itajubá, e assim retirar relatórios, *dashboards* e demais produtos de *BI/Data Science*, a partir do modelo criado, e por fim, publicando no sítio eletrônico da referida Universidade para fins de transparência e cumprimento das leis.

O modelo proposto é abordado nesta seção, bem como a forma como foi criado e sua teoria. Ele aborda, em suma, os conceitos de “Ciclo de Vida do *Business Intelligence*” e o “Processo de desenvolvimento da *Data Science*”, sendo agregados concomitantemente, através da Metodologia definida no capítulo 3 dessa dissertação, com base na modelagem definida por Arenales *et al.* (2007), através dos quatro passos definidos e adaptados também de Mitroff *et al.* (1974). A partir do modelo desenvolvido, sua aplicação é dada no Capítulo 5 deste trabalho, no objeto de estudo, como forma de validação e confirmação do mesmo.

A princípio, os cinco passos do Ciclo de Vida do *Business Intelligence*, exposto na Seção 2.2.3 do Capítulo 2 são apresentados, para que o modelo comece a ser desenvolvido. O ciclo é composto de 5 fases, como supracitado e mostrado na Figura 4.1.



Figura 4.1 – Ciclo de vida do Business Intelligence.

Fonte: Adaptado de Larson e Chang (2016).

Para Larson e Chang (2016), a Descoberta (*Discovery*) servirá como o início do processo, onde a alta gerência fornecerá os primeiros *insights* e os dados serão coletados. Na fase de *Design* (ou Projeto) a definição do trabalho a ser realizado com os dados acontece, com a estrutura de BI, *softwares* e pessoal. Na fase de Desenvolvimento (*Development*), a criação de bases de dados e códigos, produzindo um sistema que produzirá valor agregado à organização, a partir do processo de ETL aplicado. Por fim, tem-se a Implantação (*Deployment*) e a agregação de valor, onde o ciclo termina com o sistema criado a partir do modelo, visando a melhoria contínua.

Para Larson e Chang (2016), as organizações estão expandindo suas competências no campo de BI, para incluir cientistas de dados (*Data Scientists*). O ciclo de vida do *Business Intelligence* é centrado na utilidade da informação e no desenvolvimento do *software* que gerará o BI. O ciclo de vida do *Data Science*, em contraste, foca na velocidade da tecnologia

e na aquisição dos dados, envolvendo desenvolvimento de modelos analíticos que são criados, validados e alterados até que os objetivos das empresas sejam atingidos.

Para Amirian, Loggerenberg e Lang (2017), o BI acaba sobrepondo a *Data Science*. Ambos utilizam estatísticas e ferramentas computacionais, porém o foco do BI é explicar o que houve ou está ocorrendo na empresa. A *Data Science* também utiliza de dados históricos, mas em contraste com o BI, ela tem seu foco em achar padrões em termos de modelo, para descrever alguma variável esperada, baseada nos *inputs*.

Na sequência, tem-se os passos para o “Processo de desenvolvimento da *Data Science*”, exposto no ponto 2.3.3 do Capítulo 2. O ciclo é composto de 6 fases, como mostrado na Figura 4.2.

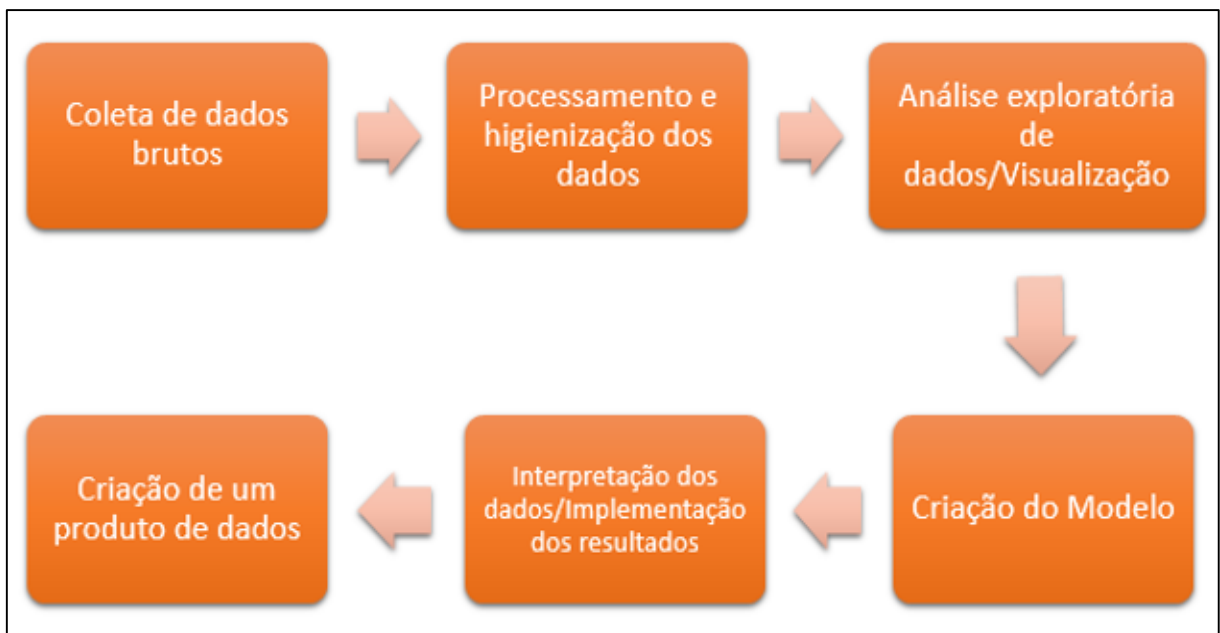


Figura 4.2 – Processo de desenvolvimento da *Data Science*.

Fonte: Elaborado pelo autor.

De acordo com Cielen, Meysman e Ali (2016), para o processo de desenvolvimento da *Data Science*, o acesso aos dados é o primeiro passo, o que consiste na Coleta de Dados Brutos, que serão selecionados conforme alinhado com a organização, para depois serem processados e higienizados. Nesse momento, são removidos erros dos dados coletados, já preparando para a próxima fase, inseridos na ferramenta que gerará um produto de *Data Science*, conforme Amirian, Loggerenberg e Lang (2017).

No próximo passo, de acordo com Larson e Chang (2016), há a Análise Exploratória de Dados, já envolvendo BI e visualizações, passando para a criação do modelo, onde Schutt e O’Neil (2014) trazem o modelo de dados para uma análise descritiva, para assim progredir para os próximos passos, interpretando e implementando o modelo final.

De acordo com Newman *et al.* (2016), a habilidade de uma empresa ao administrar bem o *Data Science* e dominá-lo nos negócios tem se tornado de grande importância para uma empresa que adota o *Business Intelligence* como ferramenta de abordagem de análise.

A *Data Science* e a análise rápida de dados são retiradas dos desenvolvimentos tradicionais em *Business Intelligence*, porém não são as únicas tendências a emergirem do BI. A *Data Science* é considerada como uma inclusão embaixo de um “guarda-chuva” dos procedimentos que englobam o BI (LARSON e CHANG, 2016).

Para a junção dos conceitos, portanto, o modelo desenvolvido passa pelas fases explanadas na seção da Metodologia: Conceitualização, Modelagem, Solução pelo Modelo e, por fim, a Implementação, adaptados de Arenales *et al.* (2007). Porém todos com a perspectiva do desenvolvimento do *Data Science* gerando um produto de *Business Intelligence*. As fases de Solução pelo Modelo e Implementação serão abordadas conjuntamente, por se tratar da finalização do Modelo, trazendo conceitos complementares.

Na Figura 4.3 a seguir, pode-se ver a interação entre todos esses conceitos, bem como os passos que serão seguidos na execução da metodologia escolhida. Em azul, tem-se o “Ciclo de vida do BI”, em laranja, o “Processo de desenvolvimento da *Data Science*”, e externamente, encontra-se por fim as 4 fases da Metodologia empregada, explanada no Capítulo 3 deste trabalho, resultando no Modelo Proposto, que permite uma compreensão melhor de todo o trabalho a ser realizado no último capítulo, bem como facilita e esclarece a interação entre os conceitos definidos.

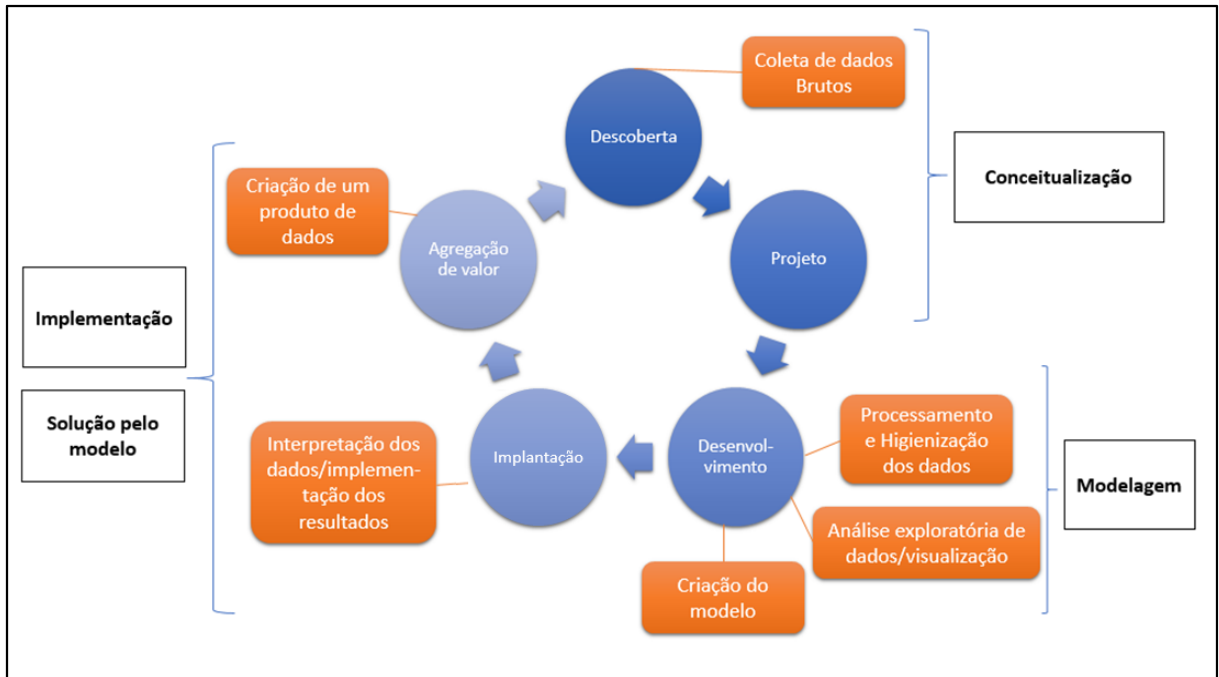


Figura 4.3 – O Modelo Proposto.

Fonte: Elaborado pelo autor.

O resultado e objetivo a ser alcançado pelo Modelo Proposto consiste em um modelo conceitual validado e efetivado na Diretoria de Contabilidade e Finanças da Universidade Federal de Itajubá, bem como publicado no sítio eletrônico da Universidade – nos moldes da Instrução Normativa 2 do Ministério do Planejamento, Desenvolvimento e Gestão, de 06 de dezembro de 2016. Este modelo para validação é apresentado no Capítulo 5 deste trabalho, a partir dos passos definidos no capítulo presente.

4.1. Conceitualização

Conforme conceituado no capítulo referente à Metodologia, a partir da Figura 4.2 pode-se retirar com detalhes cada passo, dentro da fase de Conceitualização: **Descoberta e Projeto** (na perspectiva do processo de desenvolvimento do *Data Science: Coleta de dados Brutos*). A figura supracitada demonstra esses passos com detalhes, de como será feita essa primeira fase do desenvolvimento do modelo, contendo todos os passos desta primeira fase. Para todas as demais fases, há fluxos como este, como uma continuação, em um progresso contínuo de elaboração do modelo.

O fluxo na Figura 4.4 engloba toda a fase de Conceitualização, do início ao fim, que é explanado neste capítulo.

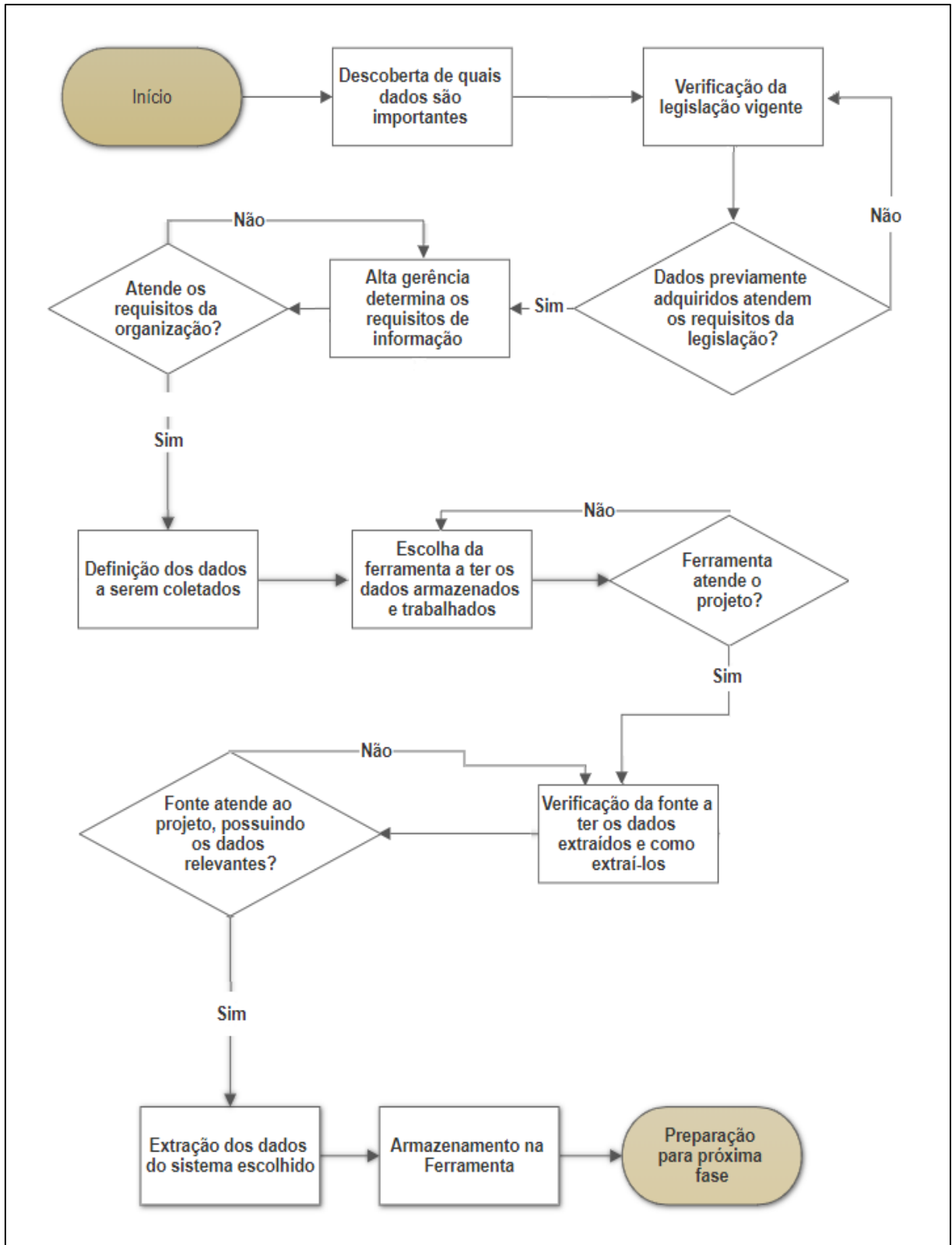


Figura 4.4 – Primeira fase: Conceitualização detalhada

Fonte: Elaborado pelo autor.

O primeiro passo da fase de Conceitualização se dá na descoberta de quais dados são de fato importantes, se servirão para extrair as informações necessárias do modelo a ser elaborado, através da futura transformação dos dados. Essa descoberta se dá no foco do departamento da empresa a ser implementado o modelo, com as necessidades de completar o objetivo final que conforme Larson e Chang (2016), consiste na fase de Descoberta, onde a alta gerência dá os *insights* sobre fontes de dados para que haja futura agregação de valor através dos dados que serão coletados.

Após o primeiro passo, há a necessidade da verificação da legislação vigente, pois o objeto de estudo é uma organização governamental. Se os dados atenderem os requisitos da legislação, pode-se seguir para o próximo passo, onde a alta gerência decide, dentre esses dados, as informações que serão utilizadas pelo departamento. Caso sejam úteis, eles são definidos para serem coletados nos próximos passos.

A alta gerência deve definir os dados que serão relevantes e que possuirão uso contínuo dentro do departamento, para que o modelo tenha uma utilidade sempre constante e crescente pelas pessoas envolvidas, além de no final do processo, agregar valor à organização. Conforme Pang (2014) destaca, existe uma burocracia quando se trata de administração de TI nas organizações públicas, e afirma que muitos gestores se apegam a uma infraestrutura de obsoleta e inflexível, o que pode gerar certa demora nos *insights* da alta gerência.

A partir dessas decisões que são a base do modelo, a definição de qual ferramenta será usada é o próximo passo. A ferramenta deverá atender ao projeto, criando uma interface amigável ao usuário e que possua ferramentas de OLAP para uma futura ETL, para assim haver o desenvolvimento de um produto de *Business Intelligence*, e onde poderão ser aplicados os passos de desenvolvimento de *Data Science*.

Para Nath (2017), as ferramentas de OLAP irão responder as perguntas analíticas que virão a seguir, e para Bala, Boussaidb e Alimazighic (2017), o processo de ETL contribuirá para higienização e transformação dos dados para assim atingir o objetivo proposto. Caso a ferramenta ofereça tais recursos, e que possa ser facilmente operada pelas pessoas envolvidas na organização, pode-se passar para o próximo passo.

Após o passo de definição da ferramenta a ser utilizada, a definição de quais sistemas governamentais serão retirados os dados é a próxima etapa. A facilidade de obtenção dos

dados, bem como a qualidade dos mesmos, é levada em consideração na escolha. Dentre diversos sistemas, a opção é a de menos fontes possíveis para se obter os resultados esperados, facilitando a escolha dos dados, pela quantidade existente. Caso o sistema analisado não seja compatível, outro deverá ser analisado até que supra a necessidade de extração de dados de forma otimizada.

Após essa definição, há a extração dos dados, de fato, e movimentação dos mesmos para a planilha, onde haverá o armazenamento, preparando assim para a próxima fase da metodologia, que se dá na Modelagem de fato, onde ocorrerá o trabalho com os dados, através de sua transformação, higienização e processamento, para assim começar o desenvolvimento de uma análise exploratória de dados e futuramente se tornar o modelo de BI esperado, concluindo o objetivo.

4.2. Modelagem

A partir da Figura 4.5 pode-se detalhar a segunda fase de criação do modelo, que é a Modelagem, que engloba o passo “*Desenvolvimento*” do ciclo de vida do *Business Intelligence* e, concomitantemente na perspectiva da *Data Science*, os seguintes passos: *Processamento e Higienização dos dados, Análise Exploratória dos Dados/Visualização* e a *Criação do modelo*.

Para Cielen, Meysman e Ali (2016), este é o momento onde se remove os erros e vícios dos dados coletados na fase anterior. Para isso, se fará higienizações, junções, descartes e transformações dos dados. De acordo com Amirian, Loggerenberg e Lang (2017), nessa fase que os dados são transformados em unidades de análise, para poderem começar a se transformar em informações.

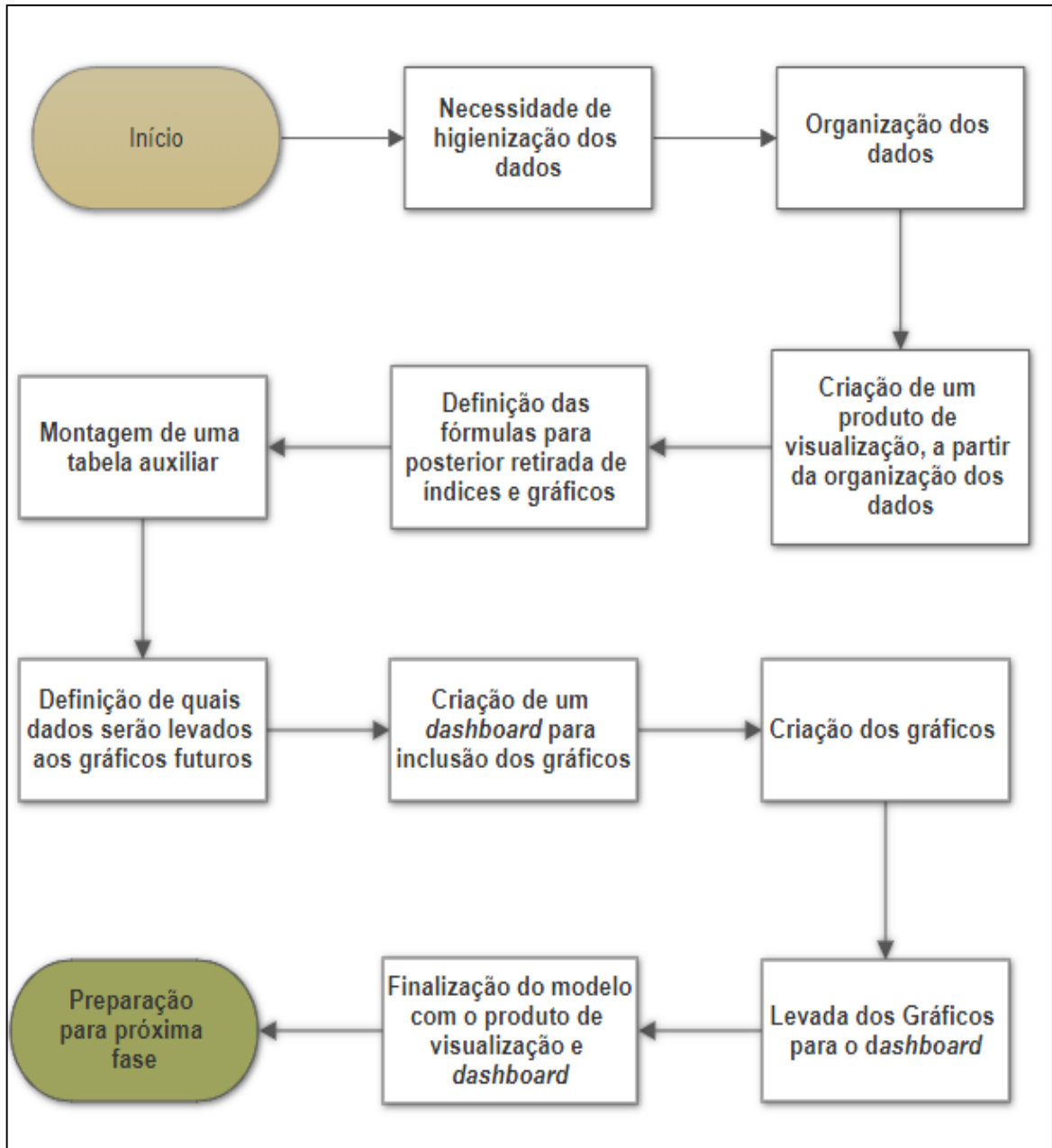


Figura 4.5 – Segunda fase: Modelagem detalhada

Fonte: Elaborado pelo autor.

Para Morabito Neto e Pureza (2007), utilizar modelos auxilia o pesquisador a melhorar sua percepção acerca de problemas, estes definidos na Conceitualização. Dessa maneira, a fase de Modelagem é a fase mais importante que será feita neste trabalho, pois se não for realizada conforme a fase de Conceitualização, e não for efetivada de forma correta, os dados podem se perder, e os produtos finais de BI podem não agregar valor à organização e usuários que o utilizarão.

Após definidos todos os passos da fase anterior, a Conceitualização, com a extração dos dados e aplicação no *software* adequado, a segunda fase inicia-se fazendo a Transformação dos dados, visto no processo de ETL. A higienização é realizada, retirando todos os dados que não sejam necessários, aplicáveis ou úteis para o modelo a ser desenvolvido.

Para Amirian, Loggerenberg e Lang (2017), como os dados são obtidos de fontes diferentes, é no passo da preparação e higienização, que são transformados em unidades de análise. Essa higienização é feita com base nos objetivos a serem atingidos com o modelo, seguindo o que foi definido na fase anterior, a partir das bases de dados que serão adequadas ao modelo.

Em seguida a higienização, a organização dos dados é necessária, pois alguns deles podem ficar deslocados ou confusos, dificultando a compreensão. Essa organização, em conjunto com a higienização, é feita com junções, arredondamentos, e transformações (SCHUTT e O'NEIL, 2014). Como a Visualização dos Dados é um subproduto de BI a ser obtido, a organização é mais necessária, com a formatação correta dos dados, conforme a proposta definida na fase de Conceitualização.

Após a criação de um produto de Visualização de Dados, há a necessidade da definição de fórmulas para a criação de um futuro *dashboard*. De acordo com Schutt e O'Neil (2014), os *dashboards* e métricas são um produto de BI e servirão como Análise Exploratória de Dados. Com as fórmulas definidas, será possível inserir gráficos no *dashboard*, além de se retirar alguns índices importantes para o setor a ser analisado. As fórmulas são componentes essenciais ao *software* escolhido para desenvolvimento do modelo, trazendo novas informações com os dados extraídos.

Posteriormente a essa etapa, a criação de tabelas adicionais para ajudarem na extração de informações no modelo pode ser necessária. Além dessas criações de tabelas auxiliares, a definição de quais informações serão desenhadas graficamente deve ser feita nessa etapa, com as informações definidas na primeira fase de Conceitualização.

Neste momento, é necessária a seleção de quais dados serão graficamente expostos no futuro *dashboard*. Para a definição destes dados, a definição na primeira fase de Conceitualização é consultada – com as informações e demandas provindos dos gestores. Alguns dados não necessitam ser graficamente expostos (onde poderão ser consultados com a ferramenta de visualização de dados), ou seja, estes não farão parte do *dashboard*.

No próximo passo, a criação do *dashboard* se inicia. Dentro dessa etapa, os gráficos são desenvolvidos, de acordo com a demanda do passo anterior. A criação de gráficos na ferramenta escolhida pode variar, de acordo com tipo, necessidade, demanda e tipo da informação. Todos são unidos em uma única interface, para poderem ser trabalhados e exporem as informações que serão utilizadas pela organização.

Nesse momento, são criados botões e comandos auxiliares, dependendo da ferramenta escolhida, para que se facilite a navegação no *dashboard*, e assim, facilite o acesso as informações que os usuários necessitarão. Para Larson e Chang (2016), a identificação de parâmetros nessa etapa serve para relacionar as variáveis a se utilizar no modelo analítico, finalizando o *dashboard* como um produto de BI, resultado de uma Análise Exploratória de Dados bem realizada, para validação futura.

Nesse passo se dá o *Load* do processo de ETL dos dados. Conforme Vercellis (2009), o *Load*, se aplica quando os dados carregados são disponibilizados aos analistas e aos SAD. O carregamento dos dados para disponibilização aos usuários se dá na criação do *dashboard* e seus gráficos dinâmicos, com acessibilidade fácil e rápida das informações disponíveis. O objetivo é atingido quando a gerência ou outro usuário, consegue acessar de modo fácil as informações úteis e precisas provindas do produto de dados.

Após o carregamento dos dados no *dashboard*, haverá última fase do modelo a ser desenvolvido, a Solução e Implementação do modelo criado. Destarte, haverá a agregação de valor ao projeto, a partir da execução do modelo desenvolvido e sua revisão constante, trazendo à organização, o valor agregado que se espera ao fim do projeto.

4.3. Solução pelo modelo e Implementação

Nos últimos passos do ciclo de BI, com a aplicação de *Data Science*, percebe-se a implementação do modelo desenvolvido, com a interpretação dos dados e efetivação dos resultados, passando para a fase de agregação de valor, criando por fim, um produto de dados eficiente e eficaz à organização, em seus diversos níveis, da gerência que está designada a aplicação até as pessoas de outros departamentos que terão acesso às informações, se necessário. Nesta última parte, são necessárias revisões, para implementação de melhorias, para que o ciclo se feche.

A Figura 4.6 informa com detalhes essa fase do processo, chamada “Solução pelo Modelo e Implementação”, seguindo a metodologia de Modelagem aplicada, englobando os seguintes passos do Ciclo de Vida do BI: “*Implantação*” e “*Agregação de valor*”, além dos passos “*Interpretação dos dados/implementação dos resultados*” e “*Criação de um produto de dados*” do processo de desenvolvimento da *Data Science*.

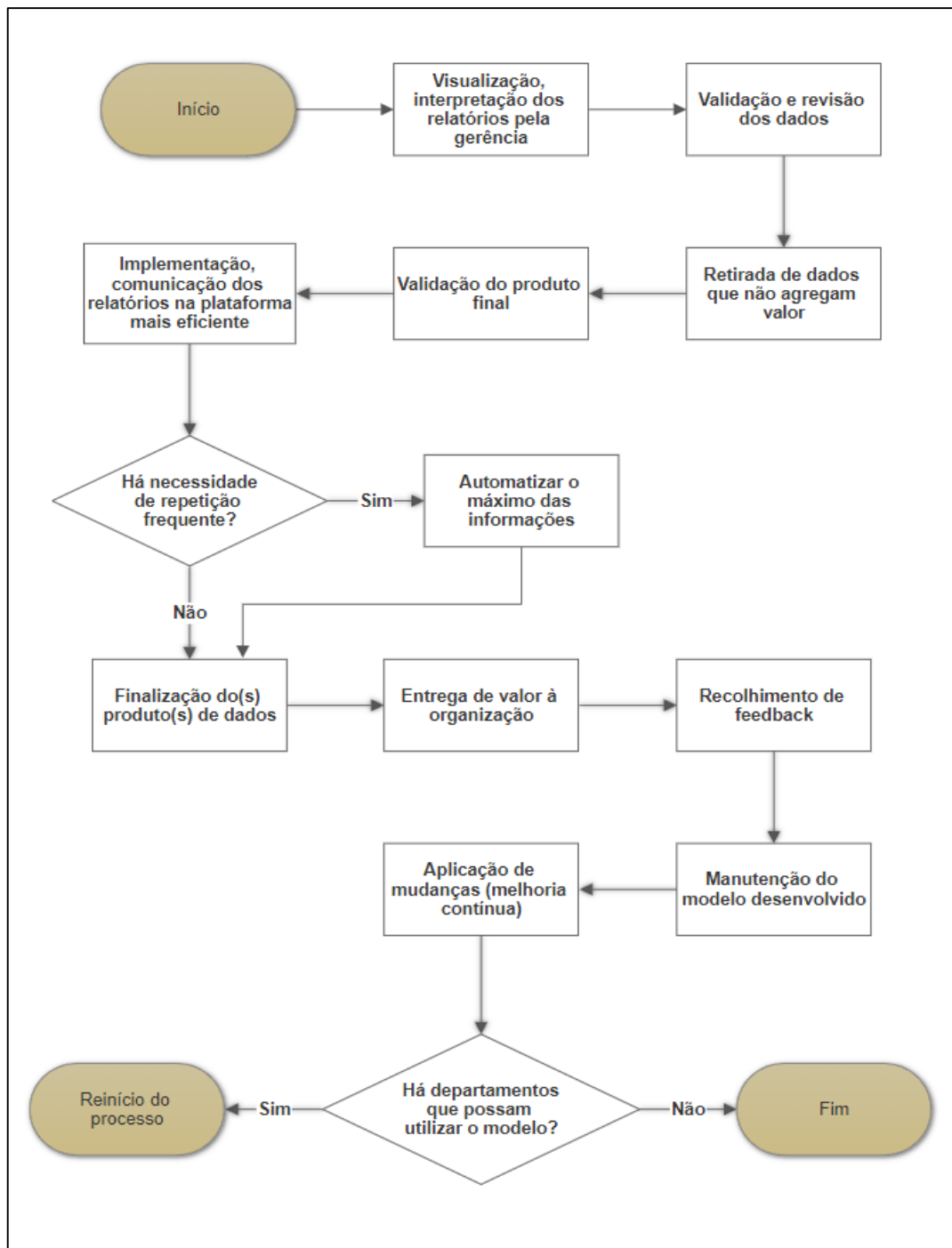


Figura 4.6 – Terceira fase: Solução pelo modelo e implementação detalhada

Fonte: Elaborado pelo autor.

Precedendo a implementação de fato, a gerência fará sua análise pessoal dos relatórios gerados a partir da fase anterior, a Modelagem. Provindos dos *dashboards* e demais produtos resultantes do modelo, os relatórios indicam aos gerentes as informações que eles necessitarão e dessa maneira, conseguem discernir se o proposto na primeira fase de Conceitualização foi realizado de maneira correta.

A partir dessa visualização e interpretação, a gerência consegue validar os dados, revisando-os e retirando as informações que não agregam valor aos relatórios do departamento, nem às pessoas que utilizaram do resultado final do modelo. Desse modo, a gerência consegue terminar a validação para assim encaminhar para o passo de implementação do modelo criado/produto de BI. O produto desenvolvido, para Larson e Chang (2016), precisa ter estabilidade na sua produção, com capacidade produtiva alta.

Enfim, se chega no ponto de implementação do modelo criado, e comunicando os relatórios provenientes do modelo nas plataformas que forem mais eficientes à organização (e-mail, rede intranet, *website*), caso haja necessidade. Se houver necessidade de repetição com frequência, da distribuição de relatórios ou de visualizações derivadas do produto de BI, as informações e dados necessitam ser automatizados ao máximo, para que não haja excesso de modificações manuais no modelo, e assim poupar tempo e ser o mais eficiente e eficaz possível, de acordo com Cielen, Meysman e Ali (2016). Caso não haja necessidade da repetição com frequência, já se pode finalizar os produtos de dados, agregando assim o valor esperado à organização, finalizando concretamente a fase de implementação.

Após os passos de implementação, existe a imprescindibilidade de um recolhimento de *feedback* de todos os usuários envolvidos, conforme Larson e Chang (2016). Essa etapa é de suma importância, pois o modelo deve ter uma manutenção contínua, com constantes aplicações de mudanças, melhorando o sistema e agregando mais valor à organização, através das opiniões dos próprios usuários, sejam eles da gerência ou apenas os usuários finais, caso existam, que apenas verificarão os relatórios afim de transparência e conhecimento. Para Chen, Chiang e Storey (2012), além de toda a aplicação do sistema de BI, as empresas precisam de uma comunicação eficaz entre os usuários, para que a implementação seja um sucesso.

Schutt e O'Neil (2014) afirmam que o modelo necessita da operação do ser humano desde a coleta de dados até a finalização. Com manutenção constante em conjunto com melhorias

no modelo e na saída de informações do mesmo, pode-se avaliar a aplicação do modelo em outros departamentos, para agregar ainda mais valor dentro da organização.

Caso o resultado seja positivo e conveniente, pode-se aplicar, começando todo o processo do início, desde a Conceitualização até a Implementação para atender a demanda de informações de forma correta, além de verificar se outro departamento pode aplicar o modelo desenvolvido, inclusive verificando sua aplicabilidade.

5. APLICAÇÃO DO MODELO PROPOSTO

A aplicação do modelo criado no Capítulo 4 se dá no presente capítulo, já utilizando de todos os conceitos daquele, para desenvolvimento do modelo proposto. A intenção da aplicação do modelo se dá na validação dele, finalizando o ciclo proposto e verificando a eficiência e eficácia do mesmo. A partir dos conceitos trazidos nos Capítulos anteriores, podemos aplicar, portanto, o Modelo Proposto, gerando produtos finais utilizáveis pela organização estudada, incluindo usuários e alta gerência.

5.1. Conceitualização

No primeiro capítulo desta dissertação foram apresentados os objetivos a serem atingidos, de forma sucinta. Como objetivo específico, se tem a aplicação do modelo que foi desenvolvido na DCF da UNIFEI, definindo o sistema a ser analisado e desenvolvido. Os problemas iniciais a serem solucionados foram: quais as informações que necessitam ser publicadas no sítio eletrônico da UNIFEI, assim como a obtenção dos dados, provenientes dos sistemas de informação do governo federal brasileiro, para dessa maneira, ter o modelo desenvolvido para se obter as informações e serem publicadas. Além disso, as informações poderão ser utilizadas pela gestão para futuras visualizações e tomadas de decisão.

Portanto, identificar quais os dados a serem descobertos foi o primeiro passo. Como o modelo a ser desenvolvido se remete aos pagamentos que a UNIFEI realiza, para melhor disposição de informações incluindo transparência, os dados são também referentes aos pagamentos. No que se remete a estes, em primeiro lugar, como citado no capítulo 3, o Sistema de Informação a ser utilizado para futura coleta dos dados brutos é o SIAFI - Sistema Integrado de Administração Financeira do Governo Federal.

Neste Sistema de Informação, são realizados todos os pagamentos de fornecedores da UNIFEI, unificando as informações. Os pagamentos são feitos de acordo com os repasses financeiros, provindos da Subsecretaria de Planejamento e Orçamento do Ministério da Educação. Esses repasses são sazonais, e não possuem previsibilidade de envio, nem mesmo a quantidade de dinheiro a ser enviada. Todos os pagamentos devem respeitar uma ordem cronológica de pagamentos, definida pela Instrução Normativa 02 de 06 de dezembro de 2016, assim como a publicação.

Os dados brutos que se possui, a partir do SIAFI, são:

- Valores;
- Datas de pagamento e ateste;
- CNPJs;
- Se são pagamentos líquidos aos fornecedores ou deduções.

Logo, os dados básicos que devem ser obtidos, para o modelo são definidos pela IN 02 de 06 de dezembro de 2016, já adiando o passo de verificação da legislação vigente. O exemplo a seguir foi retirado do sítio eletrônico governamental oficial, onde se encontra a própria IN e outras informações, como detalhado na Figura 5.1.

CATEGORIA: I - FORNECIMENTO DE BENS							
MÊS	Nº	CNPJ	Razão social	Data de exigibilidade ¹	Data de pagamento	Justificativa ²	Pagamento efetuado?
MARÇO	01	000.000.000/0000-00	XXXX	08/02/2017	01/03/2017		SIM
	02	000.000.000/0000-00	XXXX	12/02/2017	03/03/2017		SIM
	03	000.000.000/0000-00	XXXX	15/02/2017		Irregularidade xxxxxxx	NÃO
	04	000.000.000/0000-00	XXXX	20/02/2017	05/03/2017		NÃO
Abril	05
	06

Figura 5.1 – Exemplo de dados a serem publicados, conforme IN 02 de 06/12/2016

Fonte: Sítio eletrônico Compras Governamentais – último acesso: 30/04/2018.

Assim sendo, os principais dados que são necessários são:

- Mês de pagamento do documento fiscal (em sua maioria, Notas Fiscais).
- CNPJ – (Cadastro Nacional de Pessoa Jurídica) e Razão Social de cada fornecedor.

- Valores dos pagamentos e impostos.
- Data de exigibilidade – que se refere à data de ateste pelo responsável pela Nota Fiscal.
- Data do Pagamento efetivado.
- Justificativas de alterações na ordem cronológica.
- Divisão por categorias de pagamentos.

A partir desses dados, pode-se retirar diversas informações do modelo de BI, para futura tomada de decisão, não apenas a título de publicação. Com os dados pré-definidos pela própria IN, pode-se partir do princípio de coleta de dados a partir do Sistema de Informação SIAFI, após a aprovação da coordenação/diretoria. Como há necessidade de transparência e de um sistema informações/BI, o aval foi concedido.

Após a definição dos dados, a ferramenta escolhida foi o *software* Microsoft Excel, que atenderá ao projeto, fornecendo ferramentas de ETL e as de OLAP, para desenvolvimento do produto de *Business Intelligence*, possuindo todos os pré-requisitos e todos os recursos para aplicação do modelo.

Os dados serão retirados do sistema SIAFI, e transportados para uma planilha do *software* Microsoft Excel para que sirva como uma pequena base de dados para a seguinte fase de Modelagem. O SIAFI possui um comando chamado “Demonstrativo de Compromissos”, que traz os dados referentes aos pagamentos em um *range* de datas específico, porém os pagamentos necessitam ser publicados na ordem que foram pagos, de acordo com a IN supracitada, e não possui uma forma automatizada para obtenção de tais informações, de modo simplificado e fácil.

Para tal, a problemática é a organização dos dados, a qual não somente é necessária para a publicação, mas também para utilização dentro da DCF em algum futuro relatório ou tomada de decisão pela alta diretoria, o que será abordado na próxima fase, de Modelagem. A tela do sistema é mostrada na Figura 5.2.

Demonstrativo de Compromissos - DEMCOMP

Filtros

* Natureza: Pagamento | * Tipo de Pesquisa: UG

Informações Gerais

Órgão: [] | UG Pagadora: 153030 []

* Status: Realizável | Credor: []

Número do Documento Hábil: Ano: [] | Tipo: [] | Número: [] | UG Emitente: []

Período de Pagamento: De: 24/08/2018 | Até: 24/08/2018

Período de Vencimento: De: [] | Até: []

Tipo do Compromisso: [] | Sistema de Origem: []

Documento de Realização: []

Dados de Itens Orçamentários e Financeiros Dados de Pagamento

Plano Interno: [] | Natureza da Despesa: [] | PTRES: [] | UG Responsável: [] | Unidade Orçamentária: []

Fonte de Recurso: [] | Categoria de Gasto: [] | Vinculação: []

Agrupador e Ordenador

* Agrupar por: Data Pagamento/Recebimento | Ordenar por: []

Pesquisar | Limpar

Figura 5.2 – Tela retirada do Demonstrativo de Compromissos do SIAFI

Fonte: Sítio eletrônico SIAFI – último acesso: 24/08/2018.

O SIAFI possui alguns filtros que facilitam a obtenção de dados, como por exemplo de qual fonte de recursos cada pagamento foi efetuado, eliminando previamente alguns pagamentos que não necessitam de informações relativas a eles. Apesar desses filtros serem efetivos, os dados necessitam ser extraídos manualmente para a planilha, para assim começar a modelagem, com a higienização e os passos seguintes.

5.2. Modelagem

Com os dados extraídos do SIAFI, há a necessidade de higienização dos dados. Essa é a primeira etapa da fase de Modelagem. Os dados são retirados manualmente do sistema governamental, e não há necessidade de utilização de alguns, como por exemplo se os dados que são relacionados ao que foi pago ainda estão realizáveis ou não. Além dessa exclusão de dados, os dados importantes que são extraídos, quando transportados para a planilha, acabam ficando transpostos de maneira não organizada, levando assim à necessidade da organização dos dados, seguindo as etapas de modelagem.

Para organizar os dados, deve-se movê-los de acordo com o exemplo citado na primeira fase de Conceitualização, para que fique o mais fiel possível ao exemplo dado pelo próprio governo federal, respeitando também as necessidades da UNIFEI e assim extraindo futuramente um Produto de Visualização de Dados. Os dados adicionais que foram inseridos são para ajudar nos futuros questionamentos que possam existir por parte tanto de pessoas da própria UNIFEI, quanto por pessoas externas que verão o produto de dados, incluindo os valores que foram pagos.

Com esse produto de dados, se cumpre a premissa da IN 02 de 06 de dezembro de 2016, onde há a necessidade de transparência desses dados (com a futura publicação em sítio eletrônico da organização). Com o *dashboard* e produto de BI futuro, a contribuição se dá na própria organização, além de ajudar na interpretação dos dados que estão no Produto de Visualização, e assim obter ajuda na tomada de decisões estratégicas pela organização.

O exemplo de Produto de Visualização que foi retirado pela organização dos dados é dado na Figura 5.3.

Mês	Nº	DATA ATESTE	DATA PGTO	CNPJ	DOC. HÁBIL	RAZÃO SOCIAL	FONTE	VALOR	LISTAS
Julho	1	11/07/2018	24/07/2018	EXAAH3000	153030/2018RP000108	Monitoria	8100000000	R\$ 3.840,00	1-Auxílios
Julho	2	06/06/2018	03/07/2018	00047002648888	153030/2018RP000052	Bolsa	8100000000	R\$ 1.402,10	1-Bolsas
Julho	3	06/06/2018	03/07/2018	00013191587659	153030/2018RP000053	Bolsa	8100000000	R\$ 400,00	1-Bolsas
Julho	4	06/06/2018	03/07/2018	00000000000191	153030/2018RP000054	Bolsa	8100000000	R\$ 1.600,00	1-Bolsas
Julho	5	06/06/2018	03/07/2018	00000000000191	153030/2018RP000055	Bolsa	8100000000	R\$ 4.000,00	1-Bolsas
Julho	44	15/06/2018	03/07/2018	00010139887601	153030/2018FL000041	Bolsa	8100000000	R\$ 800,00	1-Monitoria
Julho	45	20/06/2018	03/07/2018	00000000000191	153030/2018FL000042	Bolsas	0100000000	R\$ 800,00	1-Monitoria
Julho	46	25/06/2018	03/07/2018	00000000000191	153030/2018FL000044	Bolsa	8100000000	R\$ 2.420,00	1-Monitoria
Julho	47	26/06/2018	03/07/2018	00000000000191	153030/2018FL000046	Bolsa	8100000000	R\$ 880,00	1-Monitoria
Julho	48	26/06/2018	03/07/2018	00000000000191	153030/2018FL000047	Bolsa	8100000000	R\$ 440,00	1-Monitoria
Julho	49	26/06/2018	03/07/2018	00000000000191	153030/2018FL000048	Bolsa	8100000000	R\$ 1.760,00	1-Monitoria
Julho	50	26/06/2018	03/07/2018	00000000000191	153030/2018FL000049	Bolsa	8100000000	R\$ 1.320,00	1-Monitoria
Julho	51	26/06/2018	03/07/2018	00000000000191	153030/2018FL000050	Bolsa	8100000000	R\$ 2.200,00	1-Monitoria
Julho	52	26/06/2018	03/07/2018	00000000000191	153030/2018FL000051	Bolsa	8100000000	R\$ 2.420,00	1-Monitoria
Julho	53	29/06/2018	09/07/2018	00000000000191	153030/2018FL000052	Bolsa	8100000000	R\$ 14.080,00	1-Monitoria
Julho	54	06/06/2018	04/07/2018	00003971083625	153030/2018RB000024	Reembolso	8100000000	R\$ 97,00	1-Reembolsos
Julho	55	08/06/2018	04/07/2018	00048232653604	153030/2018RB000022	Reembolso	8100000000	R\$ 2.412,30	1-Reembolsos
Julho	56	28/06/2018	04/07/2018	00009118366678	153030/2018RB000030	Reembolso	8100000000	R\$ 6.638,31	1-Reembolsos
Julho	57	29/06/2018	04/07/2018	00000533778859	153030/2018RB000031	Reembolso	8100000000	R\$ 2.902,04	1-Reembolsos
Julho	58	24/06/2018	26/07/2018	00009753707665	153030/2018RB000039	Reembolso	8100000000	R\$ 160,00	1-Reembolsos
Julho	1	10/07/2018	04/07/2018	02169700000130	153030/2018NP000655	MARCELO REGINALDO SILVA & CIA LTDA - ME	8100000000	R\$ 3,00	2-Tributos
Julho	2	10/07/2018	04/07/2018	28039445000102	153030/2018NP000662	ELETRO CHAVES IDEAL DE ITAUBA LTDA - ME	8100000000	R\$ 3,24	2-Tributos
Julho	3	10/07/2018	04/07/2018	03281000000103	153030/2018NP000663	TRANSAREIA BOA VISTA LTDA	0112000000	R\$ 13,44	2-Tributos
Julho	4	10/07/2018	04/07/2018	12266220000140	153030/2018NP000689	EMAV COPIADORA LTDA	0112000000	R\$ 102,36	2-Tributos
Julho	5	10/07/2018	04/07/2018	05285121000140	153030/2018NP000693	IVISOM PRODUÇÕES LTDA - ME	8100000000	R\$ 114,95	2-Tributos
Julho	6	10/07/2018	05/07/2018	01507200000106	153030/2018NP000696	BOA VISTA AUTO POSTO LTDA	0112000000	R\$ 0,34	2-Tributos
Julho	25	10/08/2018	10/07/2018	17862639000189	153030/2018NP000757	S-PAIVA ASSESSORIA E CONSULTORIA EM SAUDE OCUPA	0112000000	R\$ 1.192,46	2-Tributos
Julho	1	18/05/2018	04/07/2018	05823840000178	153030/2018NP000645	ELEVADORES ORION LTDA	8100000000	R\$ 564,63	3-Pequenos Credores
Julho	2	25/05/2018	04/07/2018	08804362000147	153030/2018NP000640	FACHINELLI COMUNICACAO LTDA - ME	8100000000	R\$ 4.769,22	3-Pequenos Credores
Julho	3	29/05/2018	04/07/2018	00836942000538	153030/2018NP000671	VIA MONDO AUTOMOVEIS E PECAS LTDA	0112000000	R\$ 160,79	3-Pequenos Credores
Julho	4	04/06/2018	04/07/2018	18928698000175	153030/2018NP000646	SAFE CERT EIRELI - ME	8100000000	R\$ 1.485,00	3-Pequenos Credores
Julho	5	04/06/2018	04/07/2018	83431593000178	153030/2018NP000648	Associação Brasileira de Engenharia e Ciências Mecânicas	8100000000	R\$ 902,50	3-Pequenos Credores
Julho	6	04/06/2018	04/07/2018	02169700000130	153030/2018NP000653	MARCELO REGINALDO SILVA & CIA LTDA - ME	8100000000	R\$ 2.150,26	3-Pequenos Credores
Julho	7	04/06/2018	04/07/2018	02169700000130	153030/2018NP000655	MARCELO REGINALDO SILVA & CIA LTDA - ME	8100000000	R\$ 136,42	3-Pequenos Credores
Julho	8	05/06/2018	04/07/2018	01088055000168	153030/2018NP000652	SOBRAL-CHAVES E CARIMBOS LTDA - ME	8100000000	R\$ 521,55	3-Pequenos Credores
Julho	9	05/06/2018	04/07/2018	24366038000140	153030/2018NP000675	ARYEL AUGUSTO DE MARCHI	0112000000	R\$ 7.592,00	3-Pequenos Credores
Julho	10	06/06/2018	04/07/2018	53585212000144	153030/2018NP000660	ASSOCIACAO NACIONAL DE PESQUISA E DESENVOLVIM	8100000000	R\$ 3.344,00	3-Pequenos Credores
Julho	11	07/06/2018	04/07/2018	28039445000102	153030/2018NP000662	ELETRO CHAVES IDEAL DE ITAUBA LTDA - ME	8100000000	R\$ 158,76	3-Pequenos Credores
Julho	12	07/06/2018	04/07/2018	03281000000103	153030/2018NP000663	TRANSAREIA BOA VISTA LTDA	0112000000	R\$ 466,56	3-Pequenos Credores
Julho	13	07/06/2018	04/07/2018	21268634000108	153030/2018NP000669	MULTIPLIER ELEUSA DONIZETI DA SILVA	0112000000	R\$ 2.395,94	3-Pequenos Credores
Julho	14	07/06/2018	04/07/2018	14183614000160	153030/2018NP000672	SAFE - SUPPLY & IT SUPRIMENTOS LTDA	8100000000	R\$ 231,90	3-Pequenos Credores
Julho	15	07/06/2018	04/07/2018	12579490000292	153030/2018NP000752	SAINT EMILION AUTOMOVEIS PEÇAS E SERVIÇOS LTDA - VIA MONDO	0112000000	R\$ 706,12	3-Pequenos Credores

Figura 5.3 – Exemplo de Produto de Visualização de dados, retirado da planilha base para o Modelo de BI através da perspectiva da Data Science

Fonte: Elaborado pelo autor.

Após a Criação desse Produto de Visualização, a definição de fórmulas para posterior retirada de dados para a criação de gráficos é necessária. As fórmulas dentro da ferramenta Excel são as fórmulas “=PROCV” e a fórmula “=SE” que auxiliam a tabela formada trazendo informações úteis à classificação das listas referentes às categorias de pagamento, bem como o transporte de informações de tabelas auxiliares. No caso do modelo desenvolvido, a partir do CNPJ da empresa, a Razão Social que está em uma planilha auxiliar, é trazida para a tabela principal, além de uma tabela auxiliar para informar o mês que foram pagas as Notas Fiscais, utilizando a fórmula “=PROCV” em conjunto com a fórmula “=MÊS”.

Nesse momento, se escolhe quais os dados seguirão para os gráficos. A partir da solicitação definida na Conceitualização, a alta gerência, define quais informações são úteis e interessantes para a organização para ter graficamente reproduzidas. No caso da UNIFEI, os gráficos sugeridos são: o Total pago por lista classificatória, o Total por fonte de recurso provindo do Governo Federal e Deduções diversas, sabendo assim o quanto foi retido de impostos ou de demais deduções dos pagamentos realizados, para que a gerência possa trabalhar com essas informações.

Os gráficos são gerados dentro da ferramenta “gráficos dinâmicos” do Microsoft Excel, após se abrir uma “Tabela Dinâmica” a partir da planilha principal onde os dados foram inseridos. Essa ferramenta de OLAP permite o trabalho com os dados, para que se possa obter as informações desejadas. Neste caso, as tabelas dinâmicas realizadas após a inserção do comando no Excel, gerarão planilhas adicionais.

Dentro dessas planilhas, se geram os gráficos dinâmicos, para assim serem transportados no passo seguinte, para o *dashboard*, e assim trabalhados de forma que fiquem amigáveis ao usuário. Na Figura 5.4, tem-se os exemplos de gráficos criados.

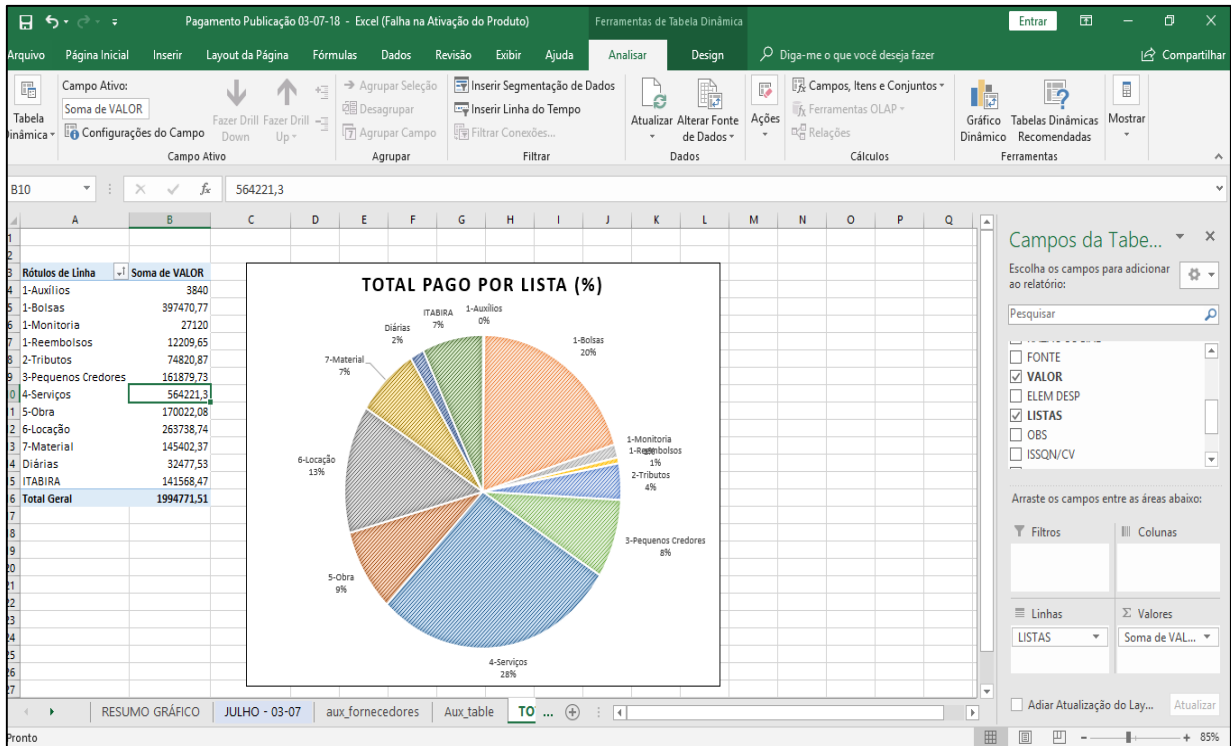


Figura 5.4 – Exemplo de Gráfico gerado pela ferramenta Microsoft Excel para ser lançado no dashboard a ser criado, a partir dos dados inseridos.

Fonte: Elaborado pelo autor.

Após os gráficos criados, uma planilha exclusiva para o *dashboard* é criada, e todos os gráficos são transportados para essa planilha. Os gráficos desenvolvidos para este trabalho são os seguintes:

- Total pago por Lista (R\$): Este gráfico permite ver o dispêndio de dinheiro feito dentro do período, para cada tipo de Lista de ordem cronológica exigida pela IN 02 de dezembro de 2016. Consegue-se perceber qual a fonte de recursos, em comparação com outra.

Estas informações são úteis em nível de transparência administrativa, bem como para ter a percepção de quanto de recurso orçamentário foi utilizado em recursos financeiros, podendo traçar estratégias para diminuição de gastos, bem como a realocação para atender as necessidades da Universidade de maneira mais eficiente, como visto na Figura 5.5.

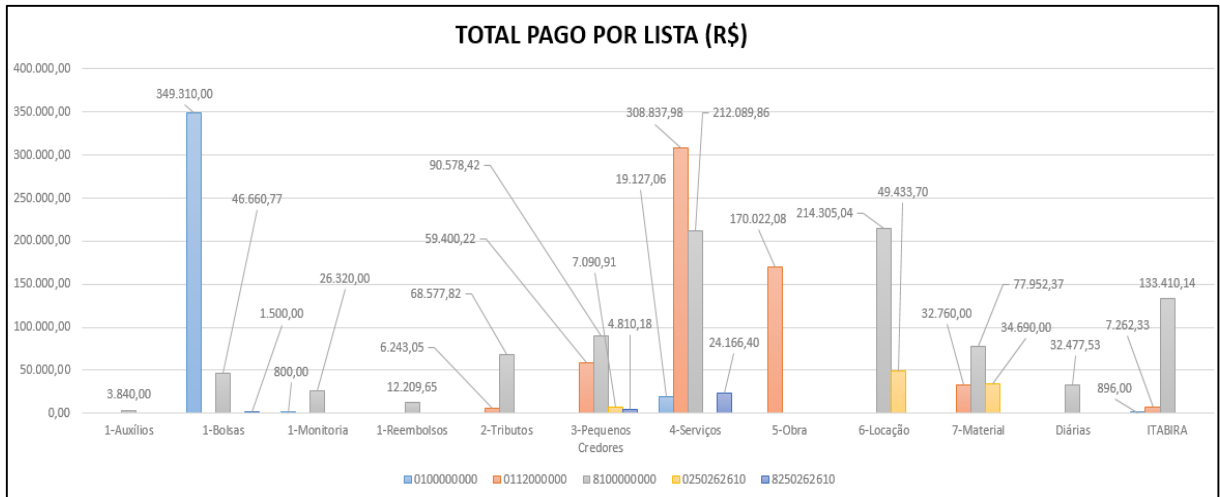


Figura 5.5 – Gráfico gerado: Total pago por lista em R\$

Fonte: Elaborado pelo autor.

- Total pago por lista (%): Este gráfico permite ver a porcentagem dispendida por lista, no total das fontes. Com o mesmo princípio da representação gráfica anterior, permite ter a percepção de quanto em porcentagem foi dispendida para cada lista, e assim verificar possibilidades de melhoria e eficiência no gasto do dinheiro público.

Pode-se ter uma percepção maior que o gráfico anterior no que se compete aos gastos dos 2 *campi* da UNIFEI, para que providências possam ser tomadas em relação a ordem de pagamento e, conforme supracitado, ter a eficiência no dispêndio de dinheiro como visto na Figura 5.6.

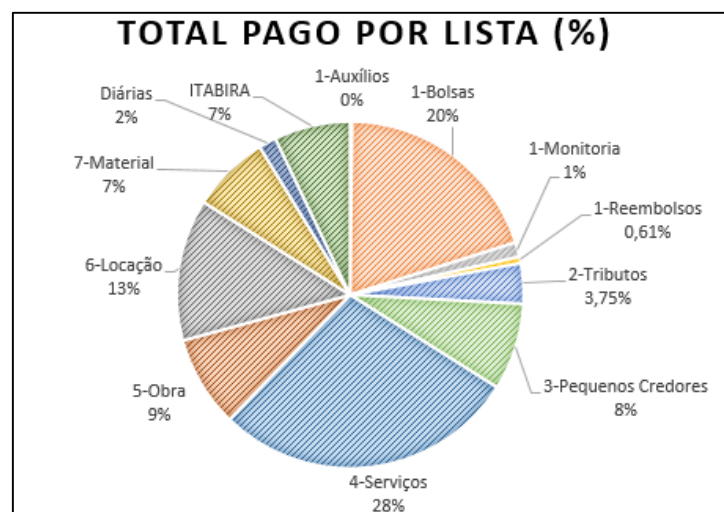


Figura 5.6 – Gráfico gerado: Total pago por lista em porcentagem

Fonte: Elaborado pelo autor.

- Total por fonte de recurso: Esta representação gráfica permite ver a porcentagem de recurso gasto por cada fonte de recurso provindo do Governo Federal. Aqui se percebe como o orçamento foi preparado, para que gastos sejam melhores alocados, como visto na Figura 5.7.

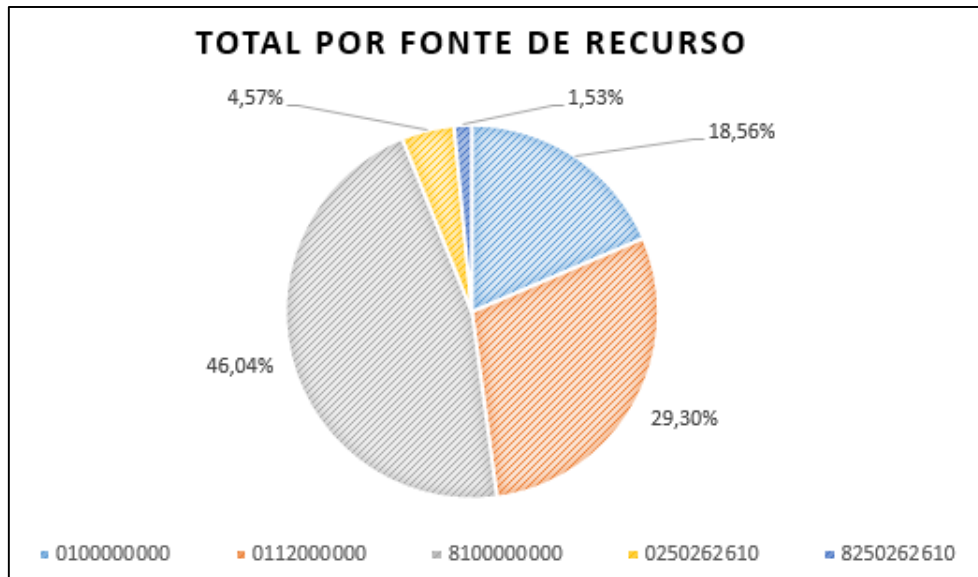


Figura 5.7 – Gráfico gerado: Total pago por fonte de recurso.

Fonte: Elaborado pelo autor.

- Total de deduções (gastas com o recurso enviado) – 2 gráficos: Pode-se perceber, especialmente, quanto de dinheiro gasto com pagamento de impostos municipais (ISSQN – Imposto Sobre Serviços de Qualquer Natureza) e até mesmo federais, além de depósitos em Contas de Garantia firmadas em contrato, para evitar algum transtorno com as empresas fornecedoras.

Aqui se tem a percepção da quantidade de dinheiro e a porcentagem dispendida com impostos, para que se compare com os valores líquidos pagos pelas empresas. A vantagem é ter o valor dispendido com impostos municipais, onde pode-se perceber a quantidade de serviço prestado no município em favor da Universidade Federal de Itajubá, como visto nas Figuras 5.8 e 5.9.

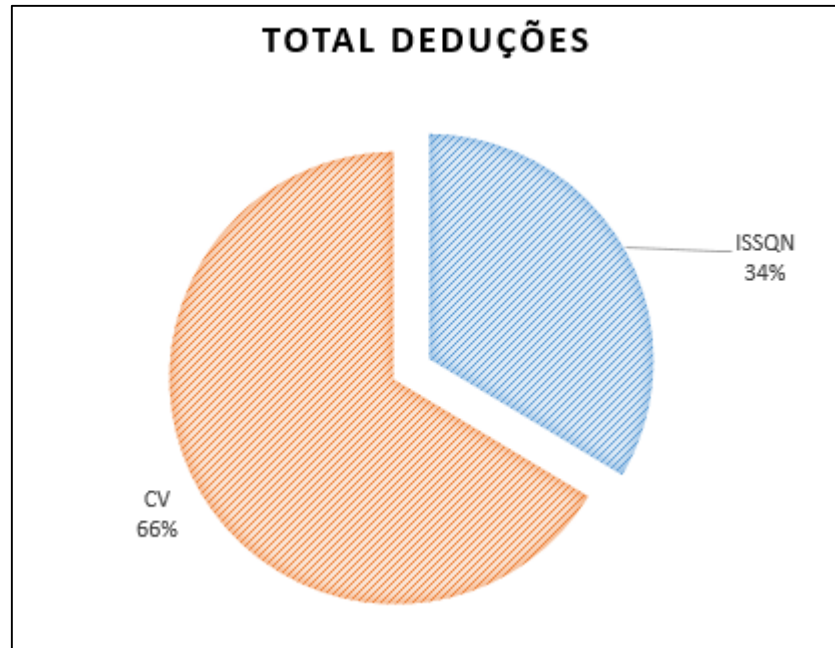


Figura 5.8 – Gráfico gerado: Total pago em deduções - porcentagem.

Fonte: Elaborado pelo autor.

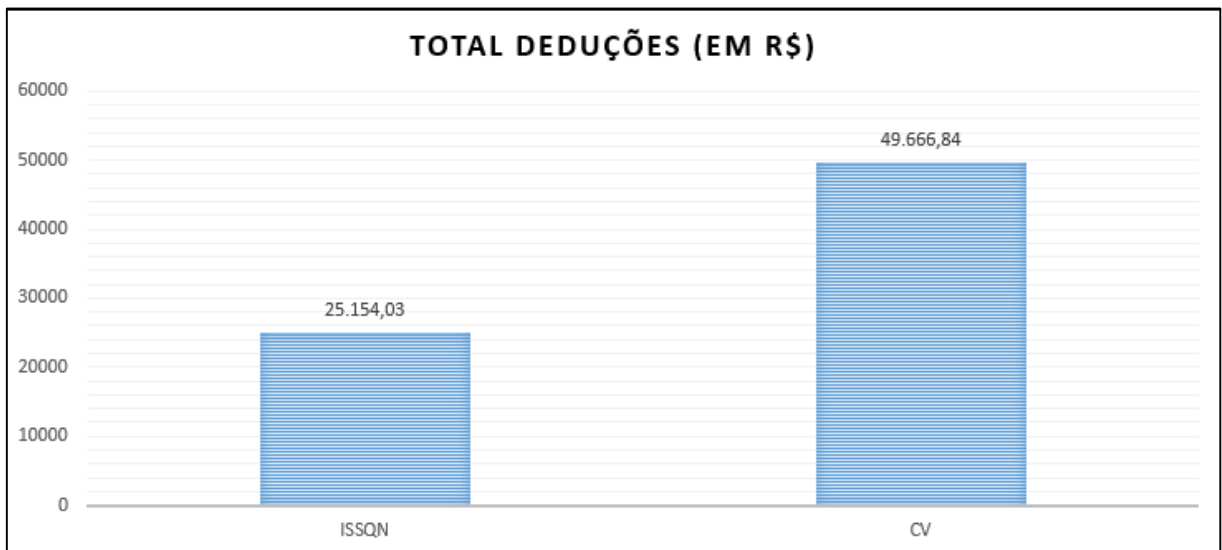


Figura 5.9 – Gráfico gerado: Total pago em deduções – valores monetários.

Fonte: Elaborado pelo autor.

- Dias de vencimento: São dois gráficos que trazem os dias em que as Notas Fiscais ficaram atrasadas, de acordo com o tempo de ateste até o seu pagamento. O primeiro traz quantas Notas Fiscais demoraram para pagar em relação aos dias. O segundo, traz a quantidade de Notas Fiscais vencidas, com seu pagamento realizado após 30 dias.

Essas informações são de extrema importância, visto que a quantidade de dias de atraso pode trazer transtornos a administração pública, incluindo de ordem judicial, além de problemas licitatórios futuros com as empresas, visto nas Figura 5.10 e 5.11.



Figura 5.10 – Gráfico gerado: Dias de vencimento.

Fonte: Elaborado pelo autor.



Figura 5.11 – Gráficos gerados: Notas Vencidas.

Fonte: Elaborado pelo autor.

Pode-se ver o esboço dos gráficos, já transportados para a planilha que serão acrescentados botões e comandos para que se transforme em um *dashboard* completo e detalhado, na Figura 5.12.

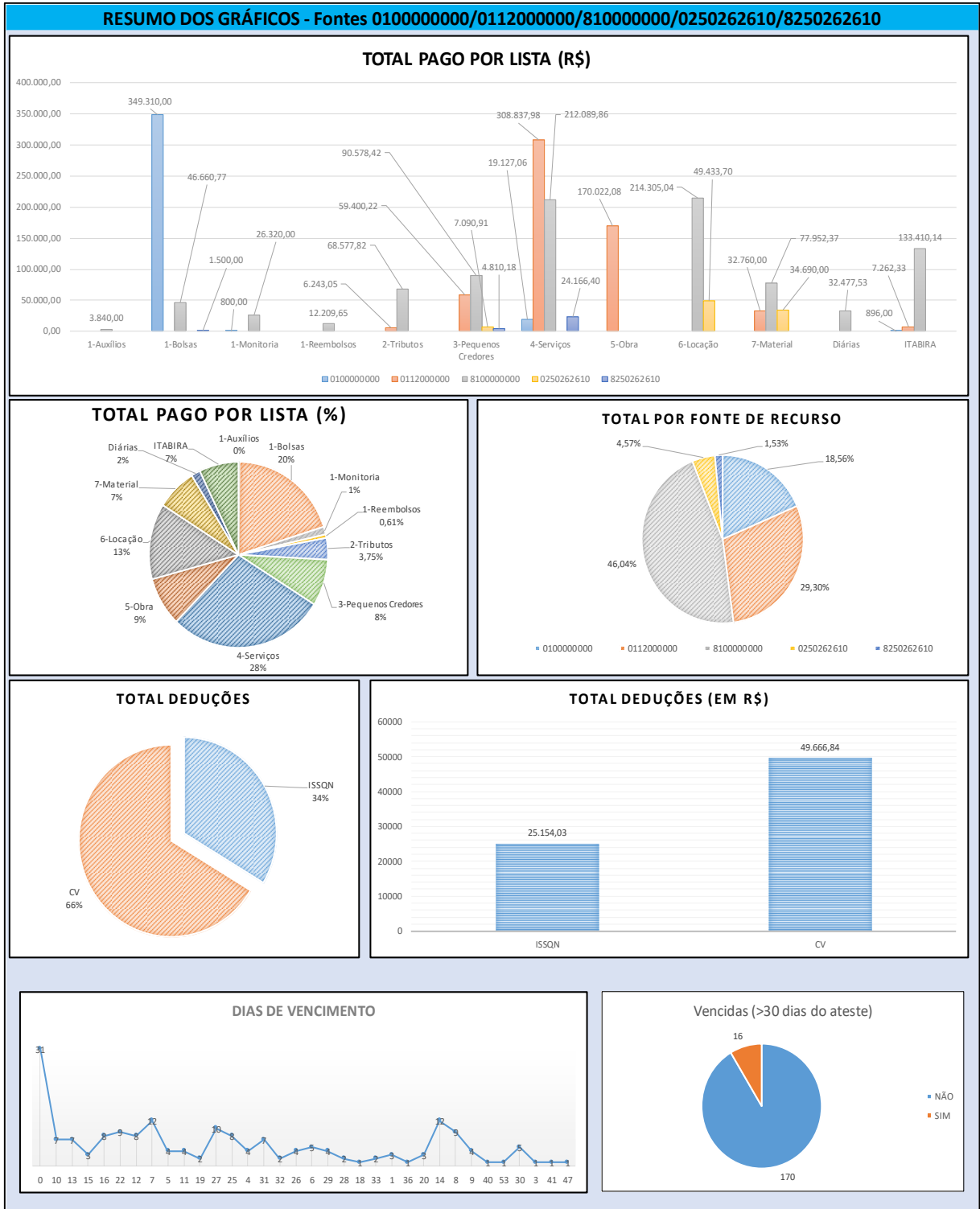


Figura 5.12 – Planilha que contém os gráficos a serem transformados em um dashboard

Fonte: Elaborado pelo autor.

A partir dos gráficos elaborados, os botões e comandos são acrescentados para que se gere um *dashboard* completo. Com os botões mostrados na Figura 5.13, os dados podem ser manipulados dentro da própria planilha do Microsoft Excel, para que as informações

possam ser filtradas, de acordo com a vontade do gestor ou usuário que utilizar a ferramenta. Os botões e comandos são inseridos a partir do comando “Inserir Segmentação de dados”, que os cria, e assim se selecionados, filtram as informações a partir dos dados que o usuário desejar:

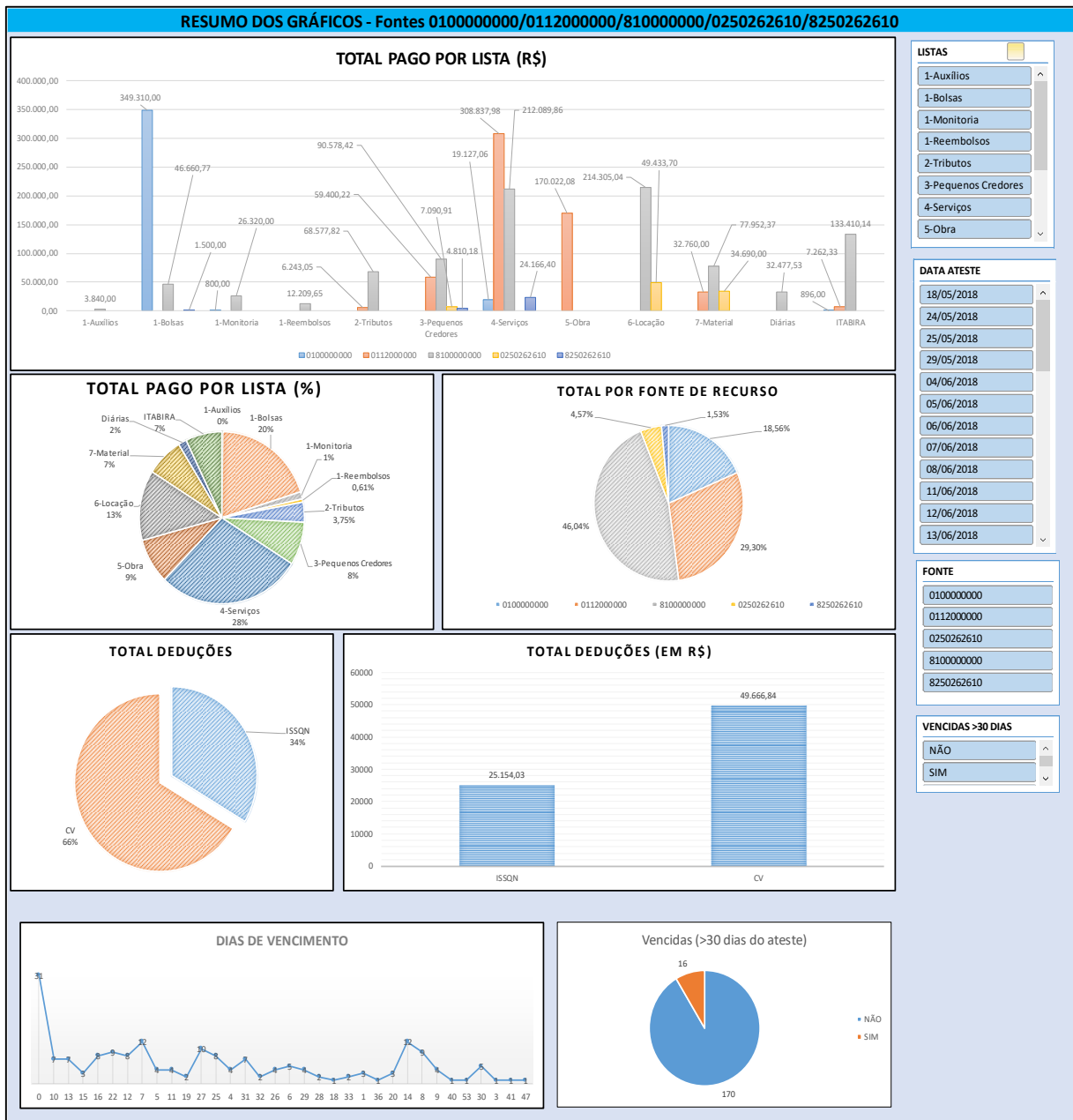


Figura 5.13 – Planilha que contém os gráficos já transformados em um dashboard

Fonte: Elaborado pelo autor.

A partir da planilha completamente pronta, a fase de Modelagem chega ao fim, e assim se inicia a próxima fase de Solução pelo Modelo e Implementação.

5.3. Solução pelo Modelo e Implementação

Na última fase, se dá a Solução pelo modelo e sua Implementação. Nesta fase, se dão todas correções necessárias, bem como a validação pela diretoria da DCF, para que haja uma melhoria constante e inclusive, caso haja necessidade e um estudo acerca dessa necessidade, aplicação em outros setores da organização.

Após a finalização do modelo, o primeiro passo é enviar a planilha completa com o Modelo pronto para a gerência. A diretoria da DCF analisa a eficiência do modelo, se está de acordo com os conceitos definidos na fase de Conceitualização, bem como se os relatórios e produtos de BI resultantes são de fato úteis para o setor. Houve assim uma interpretação por parte da gerência, para que os dados estejam de acordo com o esperado – possuindo desta maneira, um produto finalizado e validado para que seja implementado e visualizado por servidores do setor, gerentes da organização e disponibilizado para todas as pessoas, com propósito de transparência.

Com o produto finalmente validado pela diretoria da DCF, necessita-se implementar o produto de BI, para acesso a todos os usuários supracitados. Como haverá uma frequência periódica mensal, o desenvolvimento da planilha analisada deverá ser cada vez mais automatizada possível, para assim não haver a frequente demora em sua elaboração e, além disso, para não haver atraso em sua disponibilização. Essa automatização se dá na não repetição de comandos todas as vezes que a planilha for gerada e, também, na não extração de dados que não estejam conformes o passo de Conceitualização, para que a análise da alta gerência não seja necessária com tanta frequência.

A plataforma para disponibilização para os funcionários usuários do setor, é o servidor interno da UNIFEI, onde se acrescenta as planilhas modeladas mensalmente. Cada funcionário pode acessar as planilhas mensais de seu posto de trabalho, em seu computador. A plataforma escolhida para disponibilização ao público é o sítio eletrônico da UNIFEI, especificamente na página que se refere à Diretoria de Contabilidade e Finanças. De acordo com a IN 2 de 06 de dezembro de 2016, as informações devem ser disponibilizadas no sítio eletrônico da organização a qual ocorre o dispêndio de dinheiro. Na Figura 5.14 segue o demonstrativo da página onde se encontram os produtos do modelo, bem como algumas informações adicionais:



Figura 5.14 – Publicação no sítio eletrônico da UNIFEI

Fonte: Elaborado pelo autor.

Após as publicações e disponibilizações serem realizadas, o *feedback* dos usuários pode ser recolhido, para assim aplicar alguma melhoria necessária. Esta melhoria deve ser sempre contínua, aperfeiçoando o modelo aplicado, para que sempre atenda, de uma maneira cada vez mais eficiente, as necessidades do setor analisado, bem como a legislação vigente.

O *feedback* pode ser recolhido através dos usuários pertencentes à gerência, à comunidade que visualiza o sítio eletrônico, bem como de demais funcionários da organização, visando o melhor fluxo de informações e uma melhor tomada de decisões por parte das gerências da UNIFEI.

A partir do modelo pronto, pode-se analisar a necessidade de outros departamentos para que se utilize o modelo criado, para agregar valor a eles, adaptando-o conforme as necessidades citadas. Neste caso, o processo deverá ser reiniciado, para que os passos sejam novamente desenvolvidos, desde a Conceitualização, até a Solução pelo Modelo e Implementação.

Caso haja esse interesse, novos dados serão extraídos, de acordo com o departamento analisado em questão, para que o modelo de BI desenvolvido sob à perspectiva da *Data Science* seja um modelo de agregação de valor de igual ou maior eficácia que no setor da Diretoria de Contabilidade e Finanças.

6. CONCLUSÕES

6.1. Conclusões e contribuições da pesquisa

O presente trabalho de mestrado tratou de conceitos que se referem aos Sistemas de Informação (SI), mais precisamente os sistemas de *Business Intelligence* (BI), ou Inteligência de Negócios, aprofundando os conhecimentos em SI especificamente nesta área. Para explorar esses conceitos, a pesquisa se deu explorando os conceitos de BI através da perspectiva da *Data Science*, ou Ciência dos Dados, cujo enfoque é utilizar-se de diversos conceitos científicos para trabalhar com os dados e assim obter melhoria na tomada de decisões por parte dos gestores de uma organização, além da obtenção de melhoria no fluxo de informações internas e externas.

Na área de Sistemas de Informação, a tomada de decisões através de um sistema preciso é uma grande dificuldade encontrada pelos gestores, devido a imensa quantidade de dados e informações existentes atualmente, oriundas de dentro da empresa e, também, providas de meios externos às organizações.

O avanço da tecnologia e internet providenciam um cenário cada vez mais complexo, com um emaranhado de dados que necessitam de um trabalho de transformação em informações eficientes e eficazes igualmente complexo. Apesar dessa grande dificuldade, sistemas de BI estão emergindo cada vez mais, e nesse trabalho, há o auxílio da *Data Science* para se aplicar o modelo e, também, para alcançar o objetivo específico de coleta de dados e escolha de ferramentas a se utilizar.

Para explorar esses tópicos, foi escolhida uma Instituição Federal de Ensino Superior (IFES), de gestão pública governamental, para aplicar um modelo de *Business Intelligence* através da perspectiva da *Data Science*, utilizando dos conceitos metodológicos de Modelagem, abrangendo seus passos e fases, baseados no modelo adaptado de Arenales *et al.* (2007).

De acordo com Pang (2014), há uma burocracia acerca da administração de um SI nas organizações públicas, pois os gestores são apegados à infraestrutura obsoleta de TI. Desta maneira, toda Conceitualização para futura aplicação, de fato, possui uma dificuldade, demandando tempo para envolver a hierarquia e alguns processos mais complexos. Mesmo

sendo desafiadora, a utilização de SI nos setores públicos continua crescendo, conforme Klier, Klier e Muschter (2017) proporcionando uma oportunidade para aplicação de modelos como o desenvolvido neste trabalho.

Para efetuar a modelagem do Sistema de Informação de *Business Intelligence*, conforme os componentes de *Data Science* característicos, os passos dos conceitos metodológicos de Modelagem foram aplicados concomitantemente com os conceitos de elaboração e ciclo de vida de um produto de BI e, de maneira análoga, os passos e fases de desenvolvimento do *Data Science* dentro de uma organização, reunindo conceitos de Larson e Chang (2016), Amirian, Loggerenberg e Lang (2017), Schutt e O'Neil (2014) e Cielen, Meysman e Ali (2016).

Todos esses conceitos, juntos, formam o ciclo desenvolvido neste presente trabalho, e assim sua aplicação. Os pontos acerca do BI e da DS, quando reunidos, são deveras novos, e suas aplicações em conjunto são escassas, agregando assim, valor à presente pesquisa. Essa pesquisa traz contribuições relevantes para a área de BI, DS e inclusive SI, porque os principais elementos do sistema de informação utilizado no modelo computacional em estudo foram capturados, modelados e aplicados para verificar sua usabilidade.

Com todos esses conceitos definidos, a aplicação foi realizada na Instituição Federal de Ensino Superior (IFES) UNIFEI – Universidade Federal de Itajubá, como forma de validação do modelo desenvolvido. Para efetuar a Solução pelo Modelo e Implementação, os passos iniciais de Conceitualização e Modelagem foram seguidos, desenvolvendo dessa maneira, um produto de BI de Visualização de Dados bem como um *dashboard* interativo, onde gestores e demais usuários possam percorrer extraíndo informações úteis a tomada de decisão, bem como a transparência administrativa. O intuito de ajudar na tomada de decisão foi extensivamente buscado com o modelo, bem como a intensão da utilização dos relatórios de Visualização de Dados para a transparência administrativa, onde se encontra em uso no sítio eletrônico da universidade, para acesso de qualquer pessoa.

O modelo desenvolvido, após transcorrer os passos da Metodologia para sua criação, foi aplicado de forma que os objetivos dos gestores e usuários fossem cumprido, para que a aplicabilidade seja útil e replicável todas as vezes que for necessário ao setor, a instituição e aos usuários externos. Por fim, esta dissertação também traz contribuições para a área de

tomada de decisão, auxiliando gestores nesta área e propondo aplicações em demais setores e organizações.

6.2. Sugestões para trabalhos futuros

Para trabalhos futuros, o modelo desenvolvido traz uma grande aplicabilidade para dessa maneira ser útil a outros setores da organização ao qual foi inserido – a nível micro. A validação do modelo é necessária, como parte do método escolhido, visando o constante uso do produto final desenvolvido, inclusive em outras áreas e com diversos procedimentos e processos. A aplicação do modelo desenvolvido, traz a possibilidade de se aplicar em outros setores, e também a possibilidade de melhoria no futuro, para que os conceitos possam ser aplicados de forma diferente, agregando valor a tipos diferentes de organização.

Logicamente, o processo deverá ser totalmente refeito, desde a Conceitualização até as fases finais, para que seja adaptado para o novo setor em questão, incluindo os objetivos a serem atingidos. Já a nível macro, o modelo desenvolvido pode ser aplicado em diversas outras organizações, como empresas privadas e demais organizações onde haja tomadas de decisão.

Outra sugestão para trabalhos futuros é aprofundar ainda mais os conceitos de *Data Science* concomitantemente com os conceitos de *Business Intelligence*. Alguns pontos de exploração seria previsões, uso maior de ferramentas estatísticas e de ciência de computação, bem como pesquisas acerca de correlação de dados e desenvolvimento de algoritmos para melhoria da tomada de decisão por futuros e diferentes produtos desenvolvidos.

Para futuros trabalhos, os pesquisadores podem explorar as áreas supracitadas e, também, áreas de dados, aplicando modelos mais complexos incluindo desenvolvimento de algoritmos e outras ciências que abrangem dados, como por exemplo *Big Data* e *Data Mining*, citadas neste presente trabalho. Além desta área, campos de gestão que envolvam dados, desde a própria tomada de decisão em si, até a gestão do conhecimento, podem ser auxiliados com ferramentas de BI e DS, projetadas de um modelo de ambas áreas concomitantes.

REFERÊNCIAS BIBLIOGRÁFICAS

- AHLEMANN, F. Towards A Conceptual Reference Model For Project Management Information Systems. **International Journal of Project Management**. N.27, 19–30, 2009.
- AMIRIAN P., LOGGERENBERG F., LANG T. **Big Data in Healthcare – Extracting Knowledge from Point-of-Care Machines**. UK, University of Oxford: Springer, 2017. 108p.
- ARENALES, M.; ARMENTANO, V.; MORABITO, R.; YANASSE H. **Pesquisa Operacional**. Rio de Janeiro: Elsevier, 2007.
- ARNOTT, D.; LIZAMA, F.; SONG, Y. Patterns of Business Intelligence Systems Use in Organizations. **Decision Support Systems**. N.97, 58-68, 2017.
- BALA, M.; BOUSSAIDB, O.; ALIMAZIGHIC, Z. A Fine-Grained Distribution Approach for ETL Processes in Big Data Environments. **Data & Knowledge Engineering**. N.11, p. 114-136, 2017.
- BERTRAND, J. W.; FRANSOO, J. C. Modelling And Simulation: Operations Management Research Methodologies Using Quantitative Modeling. **International Journal of Operations & Production Management**. V.22, n.2. 2002, p. 241-264.
- BRICHNI, M.; DUPUY-CHESSA, S.; GZARA, L., MANDRAN, N.; JEANNET, C. BI4BI: A Continuous Evaluation System For Business Intelligence Systems. **Expert Systems With Applications**. N. 76, p. 97–112, 2017.
- CHEN, H.; CHIANG, R.H.; STOREY, V.C.; Business Intelligence and Analytics: From Big Data to Big Impact. **MIS Quartely**. N.36, p.1165–1188, 2012.
- CIELEN D., MEYSMAN A. D. B., ALI M. **Introducing Data Science**. NY-United States of America, 2016. 300p.
- COSTA, C.; SANTOS, M. Y. The Data Scientist Profile and its Representativeness in the European E-Competence Framework and the Skills Framework for the Information Age. **International Journal of Information Management**. N. 37, p. 726–734, 2017.
- FAYOUMI, A.; LOUCOPOULOS, P. Conceptual Modeling For The Design Of Intelligent And Emergent Information Systems. **Expert Systems With Applications**. N.59, 174–194, 2016.
- FLATH, C. M.; STEIN, N. Towards a Data Science Toolbox for Industrial Analytics Applications. **Computers in Industry**. N. 94, p.16–25, 2017.
- GALLINUCCI, E.; GOLFARELLI, M.; RIZZI, S. Advanced Topic Modeling For Social Business Intelligence. **Information Systems**. N.53, p. 87-106, 2015.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 1999.
- HAZEN, B. T., BOONE, C. A., EZELL, J. D., JONES-FARMER L.A. Data Quality For Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction To The Problem And Suggestions for Research And Applications. **Int. J.Production Economics**. N. 154, p. 72–80, 2014.

KLIER, J.; KLIER, M.; MUSCHTER S. How to Manage IS Requirements in Complex Public Sector Structures: Toward an Action Design Research Approach. **Requirements Eng.** N. 22, p. 419–432, 2017.

LARSON D., CHANG V. A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science. **International Journal of Information Management.** USA, n. 36, 700-710, 2016.

LAUDON, J.; LAUDON, K. **Management Information System – Managing the Digital Firm.** 14th edition – Global Edition. Pearson. United States of America. 2016. 675p.

LINDGREN I.; JANSSON G. Electronic Services in The Public Sector: A Conceptual Framework. **Government Information Quarterly.** N. 30, p. 163–172, 2013.

MANSMANN, S.; REHMAN, N.; WEILER, A.; SCHOLL, M. Discovering OLAP Dimensions In Semi-Structured Data. **Information Systems.** N.44, p. 120-133, 2014.

MARNEWICK, C. Benefits Of Information System Projects: The Tale of Two Countries. **International Journal of Project Management.** N.34, 748-760, 2015.

MITROFF I., BETZ F., PONDY L. R., SAGASTI F. On Managing Science in the Systems Age: Two Schemas for the Study of Science as a Whole Systems Phenomenon. **Interfaces.** V.4, 1974. P.46-58.

MORABITO NETO, R.; PUREZA, V. Abordagens Quantitativa e Qualitativa. In: MIGUEL, P. A. C. (Org). **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações.** 2. ed. Rio de Janeiro: Elsevier, 2012. Cap.8, p.169-198.

NATH, R. P. D.; HOSE, K.; PEDERSEN, T. B.; ROMERO, O. SETL: A Programmable Semantic Extract-Transform-Load Framework for Semantic Data Warehouses. **Information Systems.** N.68, p. 17-43, 2017.

NEWMAN, R.; CHANG, V.; WALTERS, R. J.; WILLS, G.B. Model and Experimental Development for Business Data Science. **International Journal of Information Management.** N. 36, p. 607–617, 2016.

PANG, M. IT Governance and Business Value in the Public Sector Organizations — The Role of Elected Representatives in IT Governance and its Impact on IT Value in U.S. State Governments. **Decision Support Systems.** N. 59, p. 274-285, 2014.

PROVOST, F. FAWCETT, T. **Data Science for Business.** O’Reilly Media, Inc., United States of America: 2013, p. 409.

RAINER, R. K.; CEGIELSKI, C. G. **Introdução a Sistemas de Informação.** Elsevier/Campus editora. São Paulo, Brasil. 2016. 463p.

SANCHEZ, O.P; TERLIZZI, M.A.; MORAES, H. R.O.C. Cost And Time Project Management Success Factors For Information Systems Development Projects. **International Journal of Project Management.** N.35, 1608–1626, Setembro 2017.

SCHUTT, R.; O’NEIL, C. **Doing Data Science.** O’Reilly Media, Inc., United States of America: 2014, 405p.

Site oficial da Imprensa Nacional – Disponível em: <http://www.impresanacional.gov.br/web/guest>. Último Acesso em 30/04/2018.

Site oficial Compras Governamentais - publicação da Instrução Normativa 2 de 06 de dezembro de 2016 – Disponível em: <https://www.comprasgovernamentais.gov.br/index.php/legislacao/instrucoes-normativas/291-instrucao-normativa-n-2-de-6-de-dezembro-de-2016>. Último acesso em 23/04/2018.

Site oficial da Universidade Federal de Itajubá – UNIFEI. Disponível em: <https://unifei.edu.br>. Último acesso em 23/04/2018.

Site oficial do Tesouro Nacional - Disponível em: <http://www.tesouro.fazenda.gov.br>. Último acesso em 23/04/2018.

SONG I., ZHU Y. Big data and Data Science: What Should We Teach? **Expert Systems**. Wiley Publishing Ltd. Drexel University, Philadelphia. N.4 Vol. 33, 2016.

STAIR, R.M.; REYNOLDS, G.W. **Principles Of Information Systems**. Cengage Learning. Boston, United States of America. 2010. 705p.

THEODOROU, V.; JOVANOVIC, P.; ABELLÓ, A.; NAKUÇI, E. Data Generator for Evaluating ETL Process Quality. **Information Systems**. N.63, p. 80-100, 2017.

TURRIONI, J. B.; MELLO, C. H. P. **Metodologia de pesquisa em engenharia de produção**. Itajubá: Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá, 2012.

VERCELLIS, C. **Business Intelligence. Data mining e Otimização para Tomada de Decisão**. 1st edition. John Wiley & Sons. Milano, Italy. 2009. P. 420.

YEO, K.T. Critical Failure Factors In Information System Projects. **International Journal of Project Management**. N.20, 241-246, 2002.