

**UNIVERSIDADE DE CAXIAS DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA
CURSO DE MESTRADO**

JEAN MICHEL BAÚ

**IMPLICAÇÕES DO FENÔMENO *BIG DATA* NA TOMADA DE DECISÃO
BASEADA EM DADOS EM UMA COOPERATIVA DE CRÉDITO**

**CAXIAS DO SUL
2019**

JEAN MICHEL BAÚ

**IMPLICAÇÕES DO FENÔMENO *BIG DATA* NA TOMADA DE DECISÃO
BASEADA EM DADOS EM UMA COOPERATIVA DE CRÉDITO**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para obtenção do título de Mestre em Administração.

Orientadora: Prof^a. Dra. Ana Cristina Fachinelli

**CAXIAS DO SUL
2019**

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
Sistema de Bibliotecas UCS - Processamento Técnico

B337i Baú, Jean Michel

Implicações do fenômeno *Big Data* na tomada de decisão baseada em dados em uma cooperativa de crédito / Jean Michel Baú. – 2019.
117 f. : il. ; 30 cm

Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de Pós-Graduação em Administração, 2019.

Orientação: Ana Cristina Fachinelli.

1. Big data. 2. Framework (Arquivo de computador). 3. Processo decisório. I. Fachinelli, Ana Cristina, orient. II. Título.

CDU 2. ed.: 005.311

Catalogação na fonte elaborada pela(o) bibliotecária(o)
Michele Fernanda Silveira da Silveira - CRB 10/2334

JEAN MICHEL BAÚ

**IMPLICAÇÕES DO FENÔMENO *BIG DATA* NA TOMADA DE DECISÃO
BASEADA EM DADOS EM UMA COOPERATIVA DE CRÉDITO**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para obtenção do título de Mestre em Administração.

Aprovado em: 24/05/2019.

Banca Examinadora

Prof^a. Dra Ana Cristina Fachinelli
Universidade de Caxias do Sul

Prof^a. Dra Marta Elisete Ventura da Motta
Universidade de Caxias do Sul

Prof. Dr. Fabiano Larentis
Universidade de Caxias do Sul

Prof. Dra. Priscila Nesello
Universidade Federal de Pelotas

Dedico este trabalho à minha mãe, Maria Angélica, meu exemplo de resiliência e obstinação e à minha companheira de todas as horas, Monique.

AGRADECIMENTOS

Agradeço à empresa pesquisada pela colaboração na execução deste estudo, em especial à Regina, que viabilizou as entrevistas.

Aos professores e à coordenação do PPGA – Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul por todos os momentos de aprendizado.

À minha Orientadora, professora Dr^a Ana Cristina Fachinelli, pelo apoio que sempre visava ampliar os horizontes desta dissertação para muito além da visão utilitarista do conhecimento, por vezes presente no pensamento deste pesquisador.

Aos amigos e colegas de mestrado pelo espírito de companheirismo e colaboração.

À minha mãe, Maria Angélica dos Santos e à minha namorada, Monique Larentis pelo apoio e incentivo incondicionais nesta trajetória.

Enfim, agradeço a todos que, de alguma forma, me apoiaram para realizar este sonho.

“Acharei um caminho. Ou abrirei um para mim ”
Aníbal

RESUMO

Com o crescimento exponencial do volume de dados digitais à disposição das empresas nas últimas décadas, emergem novos desafios à tomada de decisão. Além do volume, também se multiplicam as fontes e formatos dos dados. Este contexto adquire assim contornos do que é chamado de fenômeno *Big Data*. Por sua vez, a tomada de decisão baseada em dados tem potencial para mudar as estratégias empresariais; ao mesmo tempo em que exige maior preparo dos profissionais responsáveis pela tomada de decisão nas organizações. Neste contexto, o objetivo deste trabalho foi analisar as práticas decisórias baseadas em dados associadas à análise de dados caracterizados como *Big Data*. O método adotado foi um estudo de caso único, exploratório e qualitativo, desenvolvido junto a uma instituição cooperativa de crédito. Esta pesquisa revelou que as dimensões do *Big Data* (Volume, Variedade, Velocidade, Veracidade e Valor) são sinergicamente associadas as etapas do processo de extração de insights dos dados para tomada de decisão (Aquisição e Registro; Extração, Limpeza e Comentários; Integração; Agregação e Representação; Modelagem e Análise e Interpretação) no caso estudado. Revelou-se ainda que a tomada de decisão no contexto *data-driven* ocorre considerando elementos subjetivos, como a intuição. Identificou-se ainda a presença de quatro filtros ou etapas presentes no processo de tomada de decisão *data-driven*: alçadas de responsabilidade, fatores situacionais, cultura organizacional e relacionamento político. Dos resultados desenhou-se um *framework* orientativo a prática de tomada de decisão Data-Driven, em organizações que lidam com dados que possuam as dimensões do Big Data.

Palavras-chave: *Big data*. Tomada de decisão baseada em dados. Framework de tomada de decisão. Processo decisório. *Data-Driven Decision Making*.

ABSTRACT

With the exponential growth in the volume of data to be managed and analyzed by companies in the last decades, more complexity is added to the decision making. In addition to volume, they also multiply as data sources and formats. This context acquires the contours of what is called a Big Data phenomenon. In turn, Data-Driven Decision Making has the potential to change as a business strategy; at the same time that it requires greater preparation of the professionals responsible for decision making in organizations. In this context, the objective of this research was to analyze decision-making practices based on data associated with the analysis of data characterized as Big Data. The method adopted was a unique, exploratory and qualitative case study, developed with a cooperative credit institution. This research revealed that the dimensions of the Big Data (Volume, Variety, Velocity, Veracity and Value) are synergistically associated with the steps of the process of extracting insights from the data for decision making (Acquisition and Registration, Extraction, Cleaning and Comments; Aggregation and Representation, Modeling and Analysis and Interpretation) in the case studied. It was also revealed that the decision-making in the data-driven context occurs considering subjective elements, such as intuition. It was also identified the presence of four filters or stages present in the process of data-driven decision making: accountability, situational factors, organizational culture and political relationship. From the results, a Data-Driven Decision-Making framework was designed for organizations that deal with data that has the dimensions of Big Data.

Keywords: Big Data. Data-based decision making. Framework for decision making. Decision-making process. Data-Driven Decision Making.

LISTA DE FIGURAS

Figura 1 – <i>Framework</i> para análise de dados	43
Figura 2 – Fluxo de tomada de decisão	53
Figura 3 – Organograma simplificado das Cooperativas Singulares.....	54
Figura 4 – Fluxo de indicações e entrevistas realizadas	57
Figura 5 – Entrevistados	58
Figura 6 – Categorias a <i>posteriori</i>	61
Figura 7 – Validação de categorias a <i>posteriori</i>	62
Figura 8 – Práticas no preparo do tomador de decisão.....	78
Figura 9 – Relação entre risco e subjetividade.....	82
Figura 10 – <i>Framework</i> proposto para o fluxo de tomada de decisão <i>Data-Driven</i> baseado em <i>Big Data</i>	93

LISTA DE QUADROS

Quadro 1 – Definições de Big Data	24
Quadro 2 – Dimensões do Big Data	28
Quadro 3 – Matriz de decisão	38

SUMÁRIO

1	INTRODUÇÃO	13
1.1	PROBLEMATIZAÇÃO	15
1.2	OBJETIVOS DO TRABALHO	17
1.2.1	Objetivo geral.....	17
1.2.2	Objetivos específicos.....	17
1.3	JUSTIFICATIVA E RELEVÂNCIA.....	17
1.4	DELIMITAÇÃO DO ESTUDO	20
2	REVISÃO DA LITERATURA	22
2.1	<i>BIG DATA</i>	22
2.1.1	Dimensões do <i>Big Data</i>.....	26
2.1.2	Análise em <i>Big Data: background</i>.....	29
2.1.3	Desafios na análise de <i>Big Data</i>.....	31
2.1.3.1	Heterogeneidade	33
2.1.3.2	Acúmulo de ruído.....	33
2.1.3.3	Correlações espúrias.....	34
2.1.3.4	Endogeneidade incidental	34
2.2	<i>DATA-DRIVEN DECISION MAKING</i>	35
2.2.1	Teoria da decisão	36
2.2.2	O paradigma da utilidade esperada	38
2.2.3	Desafios comportamentais à tomada de decisão.....	40
2.3	DADOS NA TOMADA DE DECISÃO.....	41
2.3.1	Processo de extração de informação dos dados para tomada de decisão	43
2.3.2	Desafios à tomada de decisão baseada em dados.....	46
3	PROCEDIMENTOS METODOLÓGICOS.....	50
3.1	CARACTERIZAÇÃO DA UNIDADE DE ANÁLISE.....	51
3.2	INSTRUMENTO E PROTOCOLOS DE PESQUISA.....	54
3.3	PROCEDIMENTOS DE COLETA DE DADOS	58
3.4	PROCESSO DE ANÁLISE	59

4	RESULTADOS: ANÁLISE E DISCUSSÃO	61
4.1	VOLUME, VELOCIDADE, AQUISIÇÃO E REGISTRO.....	64
4.2	VERACIDADE, EXTRAÇÃO E LIMPEZA.....	66
4.3	ALÇADAS.....	69
4.4	VARIEDADE, INTEGRAÇÃO, AGREGAÇÃO E REPRESENTAÇÃO	71
4.5	VALOR, MODELAGEM E ANÁLISE.....	74
4.6	INTERPRETAÇÃO PARA TOMADA DE DECISÃO	76
4.6.1	Preparo dos tomadores de decisão	76
4.6.2	Intuição e subjetividade	79
4.7	FATORES SITUACIONAIS	83
4.8	CULTURA ORGANIZACIONAL.....	86
4.9	RELACIONAMENTO POLÍTICO	88
4.10	<i>FRAMEWORK</i> PROPOSTO.....	91
5	CONSIDERAÇÕES FINAIS	95
5.1	CONTRIBUIÇÕES PARA A TEORIA E PESQUISA.....	95
5.2	CONTRIBUIÇÕES PARA A PRÁTICA ORGANIZACIONAL	99
5.3	LIMITAÇÕES DO ESTUDO E SUGESTÕES PARA PESQUISAS FUTURAS.	100
	REFERÊNCIAS	102
	APÊNDICE A – ROTEIRO APLICADO NAS ENTREVISTAS	114
	ANEXO A – MAPA ESTRATÉGICO DE COOPERATIVA SINGULAR PESQUISADA	117

1 INTRODUÇÃO

Atualmente o fenômeno *Big Data* é onipresente na sociedade conectada. Seja sob a forma de dados estruturados, tais como os encontrados nas bases de dados das empresas tradicionais (em suas plataformas de ERP¹ ou CRM²); ou assumindo a forma de dados não-estruturados, típicos das tecnologias de comunicação e plataformas de interação entre usuários, como texto, imagens e vídeos (LANSLEY; LONGLEY, 2016). Para Mayer-Schönberger e Cukier (2013) o *Big Data* é uma revolução com potencial para mudar a maneira como pensamos sobre muitos assuntos importantes como negócios, saúde, política, educação e inovação nos próximos anos. Para Jin *et al.* (2015, p. 2) o uso do *Big Data* pode contribuir inclusive para o desenvolvimento das nações; “mudando e transformando a maneira como vivemos, trabalhamos e pensamos”.

Para esta pesquisa, adotou-se a definição de *Big Data* a proposta pelo IDC - International Data Corporation (TURNER, V. *et al.*, 2014), que aponta o fenômeno como “uma nova geração de tecnologias projetadas para extrair valor econômico de grandes volumes e ampla variedade de dados, permitindo análises de alta velocidade”. Adicionalmente, o *Big Data* não pode ser entendido apenas como dados em grande volume. Outras propriedades também fazem parte deste conceito, como velocidade, variedade, veracidade e valor (KATAL; WAZID; GOUDAR, 2013; WAMBA *et al.*, 2015); além da disponibilidade dos dados para serem analisados a fim de que se possa tomar uma decisão ou avaliação informada (HUNG, 2016).

O vetor chave para a ascensão do *Big Data* é a digitalização da informação. Em 2000, apenas um quarto da informação armazenada no mundo era digital. Em 2013, esse número subiu para mais de 98% dos cerca de 1200 exabytes (1 exabyte é igual a 1 bilhão de gigabytes) de informações armazenadas em todo o mundo sob todas as formas (MAYER-SCKOENBERGER; CUKIER, 2013). Espera-se que o tráfego global de dados alcance 8,6 zettabytes³ até o final de 2018 (TÖRNGREN, 2018).

De acordo com o Gantz e Reinsel (2012), entre 2005 e 2020, o universo digital deve crescer cerca de 300 vezes; gerando uma média de 5.200 gigabytes de dados

¹ Enterprise Resource Planning – *software* de gestão

² Customer Relationship Management – *software* para gestão de clientes

³ Um *zettabyte* é igual a 1.099.511.627.776 *gigabytes*.

para cada habitante da terra. Tendências apontam que em 2020, apenas pelas antenas de telefonia móvel do planeta transitem cerca de 367 exabytes de dados; o que representa um aumento de quase 5 vezes em relação a 2015 (COMPUTERWORLD, 2016). Em termos financeiros as transações comerciais na internet (*business-to-business* e *business-to-consumer*) atingirão 450 bilhões de dólares por dia em 2020 (TURNER, et al, 2014).

Pela Lei de Moore, há décadas os elementos de computação como armazenamento, memória e processamento tornaram-se mais poderosos e acessíveis; culminando até mesmo em sua virtual gratuidade. Alguns cientistas de dados argumentam que agora pode ser "mais barato manter os dados do que excluir" (ROSE, 2016).

Essa crescente digitalização é impulsionada por múltiplos fatores. Entre eles estão as melhorias significativas na capacidade de armazenamento e capacidade de processamento dos computadores nos últimos anos; o rápido aumento de dispositivos que gera novos fluxos de dados digitais como smartphones, telescópios, monitores de tráfego e câmeras de monitoramento; além do surgimento de canais de comércio eletrônico e plataformas de mídia social (SCHINTLER; FISCHER, 2018). À luz deste cenário, e da influência potencial do *Big Data* ao abranger todas as áreas funcionais de uma organização, pode-se afirmar que seu potencial disruptivo exige o engajamento das empresas em um nível estratégico (MORABITO, 2015).

Nesse contexto, o *Big Data* pode criar valor em termos de transparência, análise experimental, *business intelligence*, análise em tempo real e tomada de decisão baseada em dados. Seus usos podem melhorar a capacidade de predição ou precisão de tomadas de decisão (RADHIKA; KUMARI, 2018). Este trabalho versa justamente sobre este ponto; na relação entre o *Big Data* e a tomada de decisão.

Embora haja uma infinidade de maneiras para que as empresas aproveitem as oportunidades abertas pelo novo cenário tecnológico, um de seus primeiros passos nesse sentido pode ser tornar suas decisões mais orientadas por dados e menos por intuição (WEIGERT, 2017). Este tipo de postura é aferido pela literatura através do conceito *Data-Driven Decision Making*, ou tomada de decisão baseada em dados.

A tomada de decisão baseada em dados pode ser entendida como a utilização de dados para conferir melhor performance e eficiência à gestão de uma organização. Ela diminui a importância de fatores subjetivos, como a intuição, no processo decisório (DAVENPORT, 2010).

Este estudo visou entender o processo decisório e suas etapas, considerando a transformação de informação em decisão, a partir dos conceitos teórico de *data-driven decision making* e do fenômeno *Big Data*. A proposta é contribuir para o entendimento acerca de quais práticas as empresas podem adotar para melhorar sua performance decisória considerando as implicações do contexto *Big Data*. Dados alimentam *insights*, decisões e ações; gerando valor potencial que pode ser criado, capturado e redistribuído por meio da tomada de decisões baseada em dados (WEIGERT, 2017).

1.1 PROBLEMATIZAÇÃO

A necessidade de gerenciamento de grandes volumes de dados, de diversas fontes e em alta velocidade é um desafio presente no atual momento histórico das organizações. Este é o contexto de dados entendido como *Big Data*; uma vez que reúnem características relacionadas a cinco dimensões: velocidade (rapidez na aquisição e processamento de dados), variedade (múltiplas fontes de origem de dados a serem processados), volume (alta quantidade de informação), veracidade (necessidade de aferição dos dados) e valor (extração de insights úteis à organização) Wamba *et al.* (2015).

Essa realidade *Big Data* carrega, por sua natureza, diversos desafios às práticas organizacionais. Os problemas a serem enfrentados envolvem todos os estágios de processamento de dados: aquisição de dados de variadas fontes, integração de dados em formatos únicos que possibilitem a extração de informações, até a correta modelagem para análise e interpretação dos dados coletados (SCHINTLER; FISCHER, 2018).

Se fosse possível considerar apenas os desafios puramente técnicos, relacionados ao tratamento estatístico dos dados, os desafios já seriam suficientemente grandes para serem enfrentados pelas organizações. De acordo com Fan, Han e Liu (2014), entre os problemas a serem superados estão a heterogeneidade, o acúmulo de ruído, a endogeneidade acidental, entre outros. Todos desafios estatísticos que, se não tratados de forma correta podem inutilizar os dados.

Por outro lado, apenas um correto processamento técnico e estatístico dos dados não é capaz de traduzir dados em informação útil à tomada de decisão, ou em qualquer tipo de vantagem competitiva. Faz-se necessário extrair valor econômico

destes grandes volumes advindos de amplas variedades de dados, gerando análises de alta velocidade (TURNER *et al.*, 2014). Nesse ponto, ainda não está totalmente claro como se dá esse processo de extração de valor dos dados, e qual é o valor percebido pelos profissionais responsáveis pela tomada de decisão nas organizações, uma vez que são eles, em última análise, que podem extrair utilidade de fato dos dados e informações. Estudar este contexto é relevante ao entendermos que o *Big Data* proporciona “formas inovadoras de processamento de informações para uma melhor compreensão e tomada de decisão” (FAN; BIFET, 2013, p. 3).

No que tange ao processo de tomada de decisão orientado a dados, ou *Data-Driven Decision Making*, para este trabalho adotou-se o conceito sugerido por Davenport (2010, p. 5), que aponta a decisão baseada em dados como o “uso de dados e análise para entender e gerenciar um negócio de forma mais eficaz”. A literatura aponta que líderes organizacionais veem os dados como um recurso, e a análise como uma capacidade organizacional - ambas ferramentas valiosas que levam ao sucesso competitivo (NNAJI *et al.*, 2018). Por outro lado, Nnaji *et al.* (2018) indicam que enquanto as visões tradicionais no campo da gestão estratégica sugerem que uma estratégia escolhida determina as métricas de valor e a seleção ou aplicabilidade dos dados, várias empresas estão invertendo essa abordagem. Segundo os autores, no contexto contemporâneo, ao invés da estratégia corporativa determinar quais dados devem ser coletados e analisados, são os dados coletados que desempenham influência dramática na estratégia corporativa.

Esta pesquisa visou estudar como (e se) essa lógica inversa é aplicada nos processos de tomada de decisão. Seu foco, portanto, partiu do processo para o fator humano. Assim, este estudo é consonante com o pensamento de Davenport e Patil (2012) através de sua afirmação de que, em essência, executivos que tomam as melhores decisões se concentram nos dados como ponto central de suas estratégias organizacionais, entendendo os fluxos de dados, ao invés de concentrar seu olhar apenas nos estoques de dados.

É da intersecção dos conceitos “*Big Data*” e “*Data-Driven Decision Making*” que emerge a questão central deste estudo: como se dão os processos e práticas adotadas na tomada de decisão baseada em dados, considerando o contexto *Big Data*? Poderia este processo ser mapeado em um *framework* orientativo às organizações que enfrentam os desafios intrínsecos a tomada de decisão e ao *Big Data*? Esta pesquisa objetivou responder a estas questões.

1.2 OBJETIVOS DO TRABALHO

Esta dissertação visa o entendimento das implicações do fenômeno *Big Data* dentro do processo de tomada de decisão.

1.2.1 Objetivo geral

Analisar as práticas decisórias baseadas em dados associadas à análise de dados caracterizados como *Big Data*.

1.2.2 Objetivos específicos

- a) caracterizar o processo de análise de informação de fontes de *Big Data* considerando suas dimensões: volume, velocidade, variedade, veracidade e o valor;
- b) analisar os desafios presentes no processo de tomada de decisão baseado em informação;
- c) analisar se existe e como se dá o papel da intuição no processo decisório;
- d) desenvolver um *framework* para tomada de decisão baseado em análises de informação advindas de fontes *Big Data*.

1.3 JUSTIFICATIVA E RELEVÂNCIA

Através de um exame da discussão sobre *Big Data* na literatura acadêmica sobre gestão, Nnaji *et al.* (2018) argumenta que há espaço para a exploração acerca de como este fenômeno está revolucionando os processos de pensamento de estrategistas e gerentes corporativos. Os autores ainda afirmam que os estudiosos não estão explorando extensamente o *Big Data* como um fenômeno de nível empresarial, com o potencial de mudar a tomada de decisões e a liderança organizacional. Este trabalho se justifica diante desse cenário, pois visa através da análise das práticas compreender como opera o processo de pensamento e tomada de decisão diante de dados, a partir das perspectivas do tomador de decisão.

Além disso, esta pesquisa também se justifica pois há notada necessidade de

mais estudos acerca do uso prático do *Big Data* nas empresas; considerando que apesar de todos os recursos tecnológicos disponíveis, as organizações ainda não contam com práticas claras e consolidadas para lidar com este fenômeno. Estudo global realizado por Baldwin (2015) mostrou que quase 80% das empresas, entre as 226 pesquisadas, não integram totalmente seus dados e 65% consideram fracas suas práticas para o uso de dados. O mesmo estudo relatou que 67% das empresas não tinham “critérios bem definidos para medir o sucesso” dos investimentos em *Big Data*.

O campo de estudo desta pesquisa é um dos maiores sistemas de cooperativas de crédito do Brasil e se caracteriza como caso de interesse científico, uma vez que se utiliza extensamente de dados caracterizados como *Big Data* para tomada de decisão em diversas áreas de operação, em especial na concessão de crédito, que é o cerne do seu negócio e exige tomada de decisão ágil e assertiva. O sistema opera mais de 300 produtos e serviços financeiros, para seus mais de 4 milhões de associados, distribuídos em 22 estados brasileiros e no distrito federal. Estudar a operação de uma empresa com esta escala se traduz em oportunidade de pesquisa com potencial para gerar conhecimento relevante e historicamente contextualizado.

Para levantar subsídios teóricos para esta pesquisa, foram realizadas pesquisas bibliométricas nas bases de dados dos indexadores Scopus, Web of Science (WoS). Essas bases foram escolhidas pois dispõem dos maiores acervos globais de informação científica. A primeira pesquisa buscou publicações que empregassem os termos “*Big Data*” nos campos “título do artigo”, “resumo” e “palavras-chave”, no caso do Scopus e no campo “tópico”, no caso da WoS. Seguindo a mesma metodologia, a segunda pesquisa buscou publicações que empregassem os termos “*Data-Driven Decision Making*”. Na terceira busca utilizaram-se os dois termos de forma simultânea e conjunta, cada um entre aspas. Os dados foram coletados sem filtro de data, de modo a conferir maior amplitude à pesquisa. A coleta dos dados foi realizada no mês de março de 2018, e refeita para fins de atualização e utilizando-se dos mesmos parâmetros em junho de 2019.

Como resultados, o termo “*Big Data*” foi relacionado a 65.263 publicações na base Scopus e 37.087 na WoS, totalizando 102.350 documentos. Já o termo “*Data-Driven Decision Making*” foi relacionado a 623 publicações na base Scopus e 386 na WoS, totalizando 1.009 artigos. Quando pesquisados em conjunto, a busca resultou em 126 publicações na base Scopus e 76 na WoS, totalizando 202 artigos. O número

reduzido de resultados nesta etapa da pesquisa bibliométrica (comparativamente aqueles obtidos através das buscas por cada termo isolado) também é fator que se soma na justificativa a realização deste trabalho, ao percebermos que representa cerca de 0,2% do total de artigos encontrados; denotando espaço para estudos como o aqui realizado.

Destes 202 artigos, resultantes da busca que combinava os termos “*Big Data*” e “*Data-Driven Decision Making*” observou-se o número de publicações nestes temas (somando as publicações nas duas bases de dados pesquisadas) apresenta crescimento, sendo 2 em 2012; 12 em 2013; 13 em 2014; 34 em 2015; 35 em 2016; 44 em 2017; 43 em 2018; e 19 de janeiro a junho de 2019.

Os tipos de publicação que mais se sobressaem neste total são as do tipo *conference paper* (Scopus) com 61 publicações e *article* (WoS) com 37 publicações no período. Observa-se ainda que os periódicos com maior volume de publicações estão são aqueles relacionados as áreas de Ciência da Computação e Engenharia, com 56,5% do total; o que denota a necessidade de maior produção na área de Administração e Negócios, cujas publicações somam 15,4% do total.

Os autores de maior produção detêm apenas 2 artigos publicados diante dos filtros de pesquisa. São eles: Fawcett, T.; Hadjiefthymiades, S.; Kolomvatsos, K.; Provost, F.; Srivastava, D.; Verma, N. e Voids, A. Os periódicos apontados com maior número de publicações foram *Big Data* e *CEUR Workshop Proceedings*, com 5 publicações cada. O periódico *Big Data* é uma publicação americana, com index H 17 pela Scimago, listada como publicação relacionada a Ciência da Computação e Ciências da Decisão. Já o periódico *CEUR Workshop Proceedings* é uma publicação americana, com index H 42 pela Scimago, relacionada a área de Ciência da Computação.

Para esta pesquisa é relevante destacar que a tomada de decisão é um processo produzido a partir de dimensões humanas; e que apenas se utilizar de *Big Data* não necessariamente significa que a decisão será efetivamente caracterizada como *Data-Driven* (GILBOA; ROUZIOU; SIBONY, 2018). Destaca-se ainda que dados e *Big Data* são conceitos diferentes, na medida em que o *Big Data* é composto por dados, mas dados em si não necessariamente possuem as características do *Big Data* (AKERKAR, 2014). Ou seja: contexto *Big Data* exerce implicações específicas ao processo de tomada de decisão? Esta questão também permeia este estudo.

Ganham importância neste contexto organizacional ainda os profissionais que

utilizam dados como fonte de informação para tomada de decisão. Especialmente ao considerarmos que a decisão é uma atividade humana (DAVENPORT *et al.*, 2006). Com as análises de grandes volumes de dados, os analistas de negócios podem, por exemplo, descobrir novos fatos comerciais, antes desconhecidos e inesperados pela empresa (RUSSOM, 2011).

Em face dessa realidade, executivos e líderes empresariais estão adotando a abordagem baseada em dados para conduzir serviços mais direcionados, reduzir riscos e melhorar desempenho organizacional. Eles estão implementando programas de análise de dados para coletar, armazenar, gerenciar e analisar grandes conjuntos de dados; advindos de uma ampla gama de fontes para identificar os principais *insights* de negócios que podem ser explorados para apoiar uma melhor tomada de decisão (FAN; HAN; LIU, 2014).

O crescente poder proporcionado aos diretores de informação, diretores-chefe de tecnologia, diretores de conhecimento e diretores-chefe de dados dentro das organizações terá, sem dúvida, um efeito significativo na estratégia corporativa (MENZ, 2012; PWC, 2015). Porém, apenas a tecnologia de informação e os softwares sozinhos não são suficientes à tomada de decisão. A capacidade humana para utilizar essas infraestruturas de forma a construir análises, interpretando os dados corretamente ainda é fundamental (DAVENPORT *et al.*, 2006).

1.4 DELIMITAÇÃO DO ESTUDO

Os dois temas abordados neste estudo, do ponto de vista teórico são: *Big Data* e tomada de decisão baseada em dados. Considerando a influência dos grandes volumes de dados no processo de tomada de decisão, o presente estudo se propõe a analisar as práticas adotadas pelos tomadores de decisão para examinar, atribuir sentido e decidir com base em dados.

Para este estudo foram considerados o volume, velocidade, variedade, a veracidade e o valor como dimensões características do fenômeno do *Big Data*, conforme Wamba *et al.* (2015). Outros autores sugerem diferentes características ou dimensões, que serão apresentadas ao longo deste trabalho; entretanto estas cinco dimensões são as predominantes na bibliografia utilizada (SCHINTLER; FISCHER, 2018). Esta pesquisa considera o *Big Data* como um fenômeno percebido de contexto de dados.

Em relação a tomada de decisão baseada em dados utilizou-se a definição de Davenport (2010), definindo este conceito como “o uso de dados e análises para entender e gerenciar uma empresa de forma mais eficaz”. Esta definição foi escolhida por ser síntese dentro da literatura pesquisada.

Este estudo possui caráter exploratório e descritivo. Portanto não propõe novas ferramentas ou mecanismos, mas visa descrever um conjunto de práticas adotadas, neste caso, no processo de tomada de decisão orientada a dados.

2 REVISÃO DA LITERATURA

A revisão da literatura desta dissertação está dividida em duas partes. A primeira busca descrever e contextualizar o fenômeno *Big Data*, estudando seus conceitos e dimensões. Nesta seção também serão debatidos os fundamentos e desafios para análises baseadas em dados. Consequente, serão apresentados os conceitos relacionados a tomada decisão baseada em dados. Ainda serão discutidos os desafios que envolvem este tipo de decisão.

2.1 *BIG DATA*

Não há consenso sobre o surgimento do termo “*Big Data*”. O Dicionário Inglês Oxford, ao incluir o termo pela primeira vez em seu índice em 2013, atribuiu sua origem ao sociólogo Charles Tilly, por seu artigo publicado em 1980 “*The old new social history and the new old social history*”, cujo trecho afirma, em tradução livre: “nenhuma das grandes questões realmente cedeu ao desapontamento das pessoas big-data”. Tilly usou o termo em um contexto não direcionado a descrever especificamente a magnitude dos dados, mas como um conjunto de “grandes questões” humanas.

Aproximando-se dos conceitos de informática, uma das primeiras aparições de “*Big Data*” em uma publicação científica ocorreu em 1998, no artigo “*Challenges and trends in processor design*”, na revista americana *Computer*. O artigo relata uma discussão do tipo “mesa redonda” sobre o futuro do design dos processadores de alta velocidade; apresentando “*Big Data*” como a representação de uma grande quantidade de dados.

Já Steve Lohr, jornalista americano no *The Times*, ao realizar extensa pesquisa buscando as origens do termo para seu artigo publicado em 2013, intitulado “*The Origins of ‘Big Data’: An Etymological Detective Story*”, atribui o termo a John Mashey, cientista em chefe de uma empresa de tecnologia chamada *Silicon Graphics* na década de 1990. Mashey empregava o termo “*Big Data*” abrangendo ciência da computação, estatística e econometria em seu trabalho diário e suas palestras; embora não haja documentação científica que suporte de forma definitiva essa afirmação. A mesma conclusão em relação a autoria do termo é encontrada em Diebold (2012) em seu artigo “*A Personal Perspective on the Origin(s) and*

Development of “Big Data”: The Phenomenon, the Term, and the Discipline”.

O analista Doug Laney, da empresa de pesquisa Gartner, introduz em 2001 o conceito de “três V’s do *Big Data*”, em seu artigo “*3D data management: Controlling data volume, variety and velocity*”. Os “V’s” representavam três características: volumes extremamente grandes de dados, alta velocidade de dados e variedade extrema de dados. Volume, Velocidade e Velocidade foram designados comumente a partir daí como os “três V’s do *Big Data*” (HURWITZ *et al.*, 2013).

Em 2008 a revista Nature foi publicada em edição especial intitulada “*Big Data: science in the petabyte era*”; apresentando o pensamento de pesquisadores de diversas áreas do conhecimento afim de discutir as implicações do *Big Data* para a ciência. A revista apresenta o *Big Data* como uma “inundação de dados sendo gerada neste momento”. O debate central girava em torno dos diferentes tipos de informação disponíveis na internet, em um amplo universo que compreende desde informações científicas de fontes renomadas até dados individuais e pessoais, disponibilizados por usuários e corporações.

Em 2010 foi a vez da revista The Economist publicar “*Data, data everywhere*”, onde *Big Data* é apresentado como um fenômeno relacionado a grandes quantidades de dados e informação em diferentes níveis (acadêmico, social, sobreposto ao ecossistema de negócios) geradas por diversas fontes digitais como sensores, computadores ou celulares; e que podem gerar comunicação entre máquinas.

Segundo o IDC - International Data Corporation (TURNER, V. *et al.*, 2014) as tecnologias de *Big Data* descrevem uma nova geração de tecnologias e arquiteturas, projetadas para extrair valor econômico de grandes volumes e ampla variedade de dados, permitindo análises de alta velocidade. O ponto central desta definição versa sobre as dimensões do que pode ser considerado “*Big Data*”, como a quantidade e a velocidade da coleta de dados, além da necessidade de capacidade de processamento para sua análise. Em consonância, Manyika *et al.* (2011), afirmam que *Big Data* é o pacote de dados de dimensões que extrapolam as possibilidades de captura, armazenamento, e análise de ferramentas típicas.

“*Big Data: The next frontier for innovation, competition, and productivity*” foi o nome dado ao relatório apresentado pelo McKinsey Global Institute em junho de 2012, traduzindo “*Big Data*” como “conjuntos de dados cujo tamanho está além da capacidade de ferramentas de software de banco de dados para capturar, armazenar, gerenciar e analisar”. A McKinsey estimou que 7 exabytes de dados haviam sido

globalmente armazenados em 2010. O relatório estendia o debate ao afirmar que mais de 50% dos dados gerados não eram diretamente resultado de operações digitais conduzidas por humanos, mas resultado de transações entre máquinas.

Acompanhando a proeminência do termo, *Big Data* foi o centro das discussões no Fórum Econômico Mundial em 2012. O relatório "*Big Data, Big Impact*" apresentado pela Vita Wave Consulting, em conjunto com o governo federal americano, anunciava um investimento de mais de 200 milhões de dólares em pesquisa sobre o assunto (WEF, 2012).

Boyd e Crawford (2012) propõem uma abordagem mais holística na definição de *Big Data*, que implica em: tecnologia (armazenamento, poder de computação), análise (identificação de padrões para reivindicações econômicas, sociais, técnicas e legais) e mitologia (por exemplo, a crença generalizada de que "*Big Data*" oferece um nível mais alto de geração de informações valiosas).

Loshin (2013) discute o conceito focando o ambiente das organizações empresariais. Segundo o autor, *Big Data* é centralmente a aplicação de técnicas com boa razão custo/benefício voltadas para a resolução de problemas que ultrapassam a capacidade da organização de fazê-lo com seus atuais ativos tecnológicos. A evolução das proposições acerca das dimensões do *Big Data* é descrita no quadro abaixo, adaptado de Wamba *et al.* (2015).

Quadro 1 – Definições de Big Data

(continua)

Autores e data	Definição de <i>Big Data</i>
(JACOBS, 2009)	Dados muito grandes para serem colocados em um banco de dados relacional e analisados com a ajuda de estatística requerendo software maciço funcionando em dezenas, centenas ou mesmo milhares de servidores.
(MANYIKA <i>et al.</i> , 2011)	Conjuntos de dados com um tamanho que está além da capacidade de ferramentas de software de banco de dados típicas para capturar, armazenar, gerenciar e analisar.
(ROUSE, 2011)	Descrição da quantidade volumosa de dados não estruturados e semi-estruturados, cuja análise em um banco de dados relacional traria altos custos em tempo e dinheiro.
(IBM, 2012)	Dados capturados por sensores, postagens em sites de redes sociais, fotos digitais e vídeos, registros de transações de compra e sinais de GPS do telefone celular, etc.
(JOHNSON, 2012)	Conjuntos de dados extremamente grandes, relacionados ao comportamento do consumidor, postagens em redes sociais, geotagging, outputs de sensores.

(conclusão)

Autores e data	Definição de <i>Big Data</i>
(DAVENPORT; PATIL, 2012)	Dados de tudo, incluindo dados de fluxo de clique da Web para genômica e dados de pesquisas biológicas e médicas.
(FISHER <i>et al.</i> , 2012)	Dados que não podem ser tratados e processados de forma direta.
(HAVENS <i>et al.</i> , 2012)	Dados que você não pode carregar na memória seu computador pessoal
(BOYD; CRAWFORD, 2012)	Fenômeno cultural, tecnológico e acadêmico que se baseia na interação de: (1) Tecnologia: maximizando o poder computacional e a precisão algorítmica para reunir, analisar, vincular e comparar grandes conjuntos de dados. (2) Análise: desenhando grandes conjuntos de dados para identificar padrões econômicos, sociais, técnicas e legais. (3) Mitologia: a crença generalizada de que grandes conjuntos de dados oferecem uma forma mais elevada de inteligência e conhecimento que pode gerar insights anteriormente impossíveis, com a aura de verdade, objetividade e precisão.
(TURNER, V. <i>et al.</i> , 2014)	<i>Big Data</i> possui três características principais: os dados em si, a análise dos dados e a apresentação dos resultados da análise. Depois, existem os produtos e serviços que podem ser envolvidos em torno de um ou todos esses elementos.
(MARZ; WARREN, 2015)	Descobrimos que o processamento e o gerenciamento de grandes volumes de dados formam uma nova disciplina de ponta chamada <i>Big Data</i> .
(GHOSH, 2016)	A caracterização atualmente aceita de <i>Big Data</i> é dada em termos de cinco 'V's: volume; velocidade; variedade; veracidade e valor.
(AL-SAI; ABUALIGAH, 2017)	<i>Big Data</i> pode ser definido como “um termo que descreve grandes volumes de dados de alta velocidade, complexos e variáveis que requerem técnicas e tecnologias avançadas para permitir a processo de captura, armazenamento, distribuição, gerenciamento e análise da informação”.
(HAND, 2018)	O termo não tem uma definição aceita universalmente, mas podemos definir <i>Big Data</i> como o resultado da coleta automática de dados [...] e a capacidade de armazená-los.
(JIA, 2019)	<i>Big Data</i> refere-se à soma de todos os tipos de dados de informação. Refere-se aos dados que excedem a capacidade de processamento dos bancos de dados tradicionais.

Fonte: Adaptado pelo autor a partir de Wamba *et al.* (2015).

Através do quadro acima percebe-se que a definição de *Big Data* evolui ao longo dos anos. O que antes era visto como simples característica de dados, passa a ser compreendido como fenômeno cultural e tecnológico.

2.1.1 Dimensões do *Big Data*

Laney (2001) apresenta os 3 V's do *Big Data* (Volume, Variedade e Velocidade) introduzindo a ideia de que o fenômeno deveria ser estudado e entendido como uma soma de partes. Tendo na infraestrutura de TI para *e-commerces* a inspiração para suas projeções, a consultoria foi precursora na discussão sobre as dimensões do *Big Data*. Assim eram propostos os 3 "V's": (i) volume: indicando o tamanho crescente dos dados, na ordem dos terabytes; (ii) velocidade: relacionado à taxa crescente em que a informação é produzida; e (iii) variedade: referente aos diversos tipos e formatos que a informação digital pode tomar, como fotos, vídeos e texto.

Avançando sobre estes conceitos, pode-se definir "Volume" como a grande quantidade de dados que consomem capacidade de armazenamento; "Velocidade" como a frequência de geração, transferência ou entrega de dados e, "Variedade" como o fato de que os dados assumem diferentes formatos, incluindo dados estruturados e não-estruturados (KWON; SIM, 2013; McAFEE *et al.*, 2012).

Outros conceitos foram adicionados posteriormente, ancorados na percepção de que apenas 3 dimensões seriam insuficientes para explicar e categorizar o *Big Data*. "Valor" é sugerido como um quarto "V", relacionado a necessidade de extrair benefícios econômicos de grandes volumes de dados (TURNER, V. *et al.*, 2014; ORACLE, 2012; FORRESTER, 2012). Essa dimensão denota a capacidade de gerar benefícios e valor através de insights obtidos por análises (por exemplo: milhões de dólares economizados pelos fabricantes de motores de aeronaves usando análises para prever eventos de motores que levam a disputas dispendiosas nas companhias aéreas) (CHEN; ZHANG, 2014).

Uma quinta dimensão ainda seria sugerida pela IBM (2012), chamada de "Veracidade". Esta dimensão guarda relação com o nível de confiança das variadas fontes de dados, uma vez que a internet se configura como um campo quase ilimitado para a proliferação de dados e informação sem qualquer tipo prévio de censura, análise ou filtro.

A proliferação de "V's" para explicar as dimensões do *Big Data* é também razão de escárnio por parte da imprensa especializada. Um dos exemplos disso é o artigo cômico "*The 42 V's of Big Data*" (SHAFER, 2017).

Khan, Uddin e Gupta (2014) destacam o potencial de *Big Data* nas áreas de

pesquisa e negócios, considerando seu potencial na solução de problemas do mundo real. Os autores destacam as 7 dimensões do termo, a seguir apresentadas:

- a) volume (o tamanho dos dados que são manipulados);
- b) velocidade (a rapidez com que os dados chegam às organizações);
- c) variedade (diversidade em relação à apresentação dos dados);
- d) veracidade (o grau de fidelidade dos dados);
- e) validade (a precisão que os dados se encontram para sua utilização);
- f) volatilidade (quanto tempo é necessário o armazenamento dos dados); e
- g) valor (o objetivo final, ou seja, aquilo que se deseja extrair dos dados).

Firican (2017) propõe sete dimensões para o fenômeno *Big Data*, apresentando algumas variações acerca das propostas em estudos anteriores:

- a) variabilidade: significa o número de inconsistências nos dados devido a multiplicidade de variáveis de dados e dimensões resultantes de vários tipos e origens de dados diferentes;
- b) veracidade: se refere à proveniência ou confiabilidade dos dados no contexto de uma análise significativa;
- c) validade: refere-se à precisão e exatidão dos dados para o uso pretendido;
- d) vulnerabilidade: refere-se às preocupações com a segurança e integridade dos dados;
- e) volatilidade: refere-se às regras para circulação e disponibilidade de dados;
- f) visualização: significa o formato e o layout dos dados resultados da análise;
- g) valor: refere-se aos significados contextuais trazidos pela análise dos dados.

São 5 as dimensões destacadas na literatura pesquisada. A evolução das proposições acerca destas dimensões é apontada no quadro abaixo, adaptado de Wamba *et al.* (2015):

Quadro 2 – Dimensões do Big Data

Atributo	Natureza	Aplicações Práticas
Volume	Grande volume de dados consumindo enorme armazenamento, ou consistindo de grande número de registros (RUSSOM, 2011).	<ul style="list-style-type: none"> • A Tesco gera mais de 1,5 bilhões de novos dados por mês (MANYIKA <i>et al.</i>, 2011). • O <i>data warehouse</i> do Wal-Mart inclui cerca de 2,5 petabytes de informação (MANYIKA <i>et al.</i>, 2011).
Variedade	Dados gerados a partir de maior variedade de fontes e formatos, contendo dados multidimensionais (RUSSOM, 2011).	<ul style="list-style-type: none"> • A Procter & Gamble criou um grupo composto por mais de 100 analistas para analisar as inter-relações entre as funções da organização (DAVENPORT <i>et al.</i>, 2006). • A Tata Motors analisa 4 milhões de mensagens de texto todos os meses, abrangendo desde reclamações de produtos até lembretes sobre serviços, (AGARWAL; WEILL, 2012).
Velocidade	Frequência de geração e entrega de dados (RUSSOM, 2011).	<ul style="list-style-type: none"> • Os revendedores da Amazon podem rastrear dados individuais do cliente, incluindo dados de fluxo de cliques da Web, podendo realizar análises comportamentais. (MANYIKA <i>et al.</i>, 2011).
Veracidade	A necessidade de alguns dados requererem análise para ganhar confiabilidade (BEULKE, 2011).	<ul style="list-style-type: none"> • eBay Inc. enfrentou um enorme problema de replicação de dados, gerando de 20 a 50 versões dos mesmos dados. (DAVENPORT; PATIL, 2012).
Valor	A medida de ganho econômico gerado pelos dados, através da extração da informação (WAMBA <i>et al.</i> , 2015)	<ul style="list-style-type: none"> • Premier Healthcare Alliance usou o compartilhamento aprimorado e análise de dados para melhorar o tratamento dos pacientes, reduzindo os gastos em US\$ 2,85 bilhões (IBM, 2012). • Match.com reportou aumento de mais de 50% na receita, com mais de 1,8 milhão de assinantes pagos em sua plataforma, a maioria dos quais impulsionados através de análise de dados (KIRON; SHOCKLEY, 2011).

Fonte: Adaptado pelo autor a partir de Wamba *et al.* (2015).

Nota-se, no quadro acima, que as dimensões do fenômeno *Big Data* têm tradução prática e empírica. O Gartner (2012) resume estas cinco dimensões na sua definição de *Big Data* como "alto volume, velocidade e variedade recursos de informação demanda econômica, formas inovadoras de processamento de informações para uma melhor compreensão e tomada de decisão" (FAN; BIFET, 2013).

Segundo Akerkar (2014), um fato interessante sobre *Big Data* é que nem tudo que é considerado grande volume de dados é, na verdade, *Big Data*. Há uma necessidade de explorar e aprofundar aspectos científicos como análise, processamento e armazenamento de grandes volumes de dados. Em consonância, definiu-se "*Big Data*" para este estudo como uma abordagem holística para gerenciar, processar e analisar 5 V's (ou seja, volume, variedade, velocidade, veracidade e valor), a fim de criar informações acionáveis para a entrega de valor sustentado, medição de desempenho e estabelecendo vantagens competitivas.

2.1.2 Análise em *Big Data*: background

O uso efetivo do *Big Data* encontra diversos desafios ao considerar-se a análise de dados e informação. O peso e a gravidade destes desafios variam dependendo do tipo de dados em análise, e do próprio tipo de análise realizada, e sobre cada matriz de dados (UN GLOBAL PULSE, 2012).

A Organização das Nações Unidas, em seu artigo "*Big Data for Development: Challenges & Opportunities*" (UN GLOBAL PULSE, 2012) apresenta os desafios em análises de *Big Data* em 3 esferas: 1) Observar a imagem completa, ou seja, resumir os dados; 2) Compreender e interpretar os dados através de inferências; e 3) Definir e detectar anomalias.

O Instituto de Estatística na União Europeia (Eurostat) aponta que "é preciso reconhecer que o *Big Data* é um fenômeno que está impactando em muitas áreas políticas. Portanto, é essencial desenvolver uma estratégia oficial de estatística em *Big Data*, em nível internacional" (DGINS, 2013).

A expressão *Big Data* significa "grandes dados" em tradução livre. E existem duas fontes comuns de dados agrupados sob o conceito de *Big Data*. A primeira diz respeito aos dados internos da organização (dados estruturados, não-estruturados ou semiestruturados). A segunda é relacionada a dados externos à organização, tais

como informações disponíveis em sites de mídia social, ou ainda literatura distribuída gratuitamente pelos concorrentes (SATHI, 2012).

Estima-se que apenas 5% de todos os dados digitais são "estruturados". Dados estruturados são aqueles disponíveis em um mesmo formato, dentro de um mesmo bando de dados. Os dados internos de uma empresa, que constam sem seu sistema eletrônico de gestão, por exemplo, são considerados dados estruturados uma vez que podem ser agrupados e analisados através de uma mesma tecnologia (KAPLAN; HAENLEIN, 2010).

Os outros 95% resultantes são compostos de dados semiestruturados e não-estruturados. Dados semiestruturados não se enquadram em uma estrutura única de banco de dados. No entanto, eles contêm *tags* (etiquetas) que permitem o agrupamento de elementos semânticos de modo a possibilitar a construção de hierarquias dentro dos dados (MINELI; CHAMBERS; DHIRAJ, 2013).

As mídias sociais estão entre os geradores mais prolíficos de grandes volumes de dados não-estruturados e semiestruturados. Isso ocorre pois permitem que bilhões de pessoas em todo o mundo interajam diariamente, publicando e compartilhando informações em diferentes formatos de dados. Define-se "mídia social" como um grupo de aplicações baseadas na web que permitem a criação e introdução de conteúdo (KAPLAN; HAENLEIN, 2010).

Este conteúdo pode ser definido como um conjunto de informações publicadas em um website, implicando uma certa quantidade de esforço criativo (OCDE, 2015). Grande parte de seus dados são considerados não-estruturados pois são compostos de diferentes formatos (como vídeos, fotos, áudios e textos), o que dificulta sua categorização e análise (KAPLAN; HAENLEIN, 2010). Para retirar efetivo conhecimento de grandes volumes de dados de mídias sociais, são requeridas metodologias de diferentes disciplinas, tais como: mineração de dados, aprendizado de máquinas (*machine learning*), análise de redes sociais (grafos), sociologia, estatística e matemática (KAPLAN; HAENLEIN, 2010).

Zhao (2013) ainda complementa que as técnicas de análise relacionadas ao *Big Data* são diversas, e não exclusivas desta área. Desenvolvidas ao longo da história, elas são relacionadas a diversas áreas científicas, como matemática, estatística, economia, ciências sociais e computação. Embora tenham avançado ao longo do tempo, eles são instrumentos analíticos que estão embutidos em soluções de *Big Data*; como mineração de dados, testes A/B, e análise de redes.

A triangulação dos dados disponível em tempo real com informações previamente armazenadas e analisadas permite a geração de insights não possíveis pelo uso de outras técnicas (NUNAN; DOMENICO, 2013); fato que tem potencial para inserir novas formas de entender a análise científica. Em 2008, o editor chefe da revista Wired, Chris Anderson, afirmou quão ineficiente é usar o método científico, se comparado ao *Big Data*. Em seu artigo, "*The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*", defendeu que, com a possibilidade de analisar uma enorme quantidade de dados, o método científico estaria desatualizado. Anderson afirmou ainda que o modelo baseado em observação, formulação de hipóteses, condução de experimentos com coleta de dados gerando uma posterior interpretação de dados, será completamente substituído por modelos estatísticos e análises de correlações, sem qualquer teoria.

O autor completa seu raciocínio ao afirmar que todos os modelos antigos ou teóricos são inválidos, uma vez que usando um volume maior de informações a etapa de modelagem poderia ser ignorada. Desse modo, métodos estatísticos poderiam ser usados para encontrar padrões sem contar com hipóteses prévias; valorizando a correlação acima da causalidade (ANDERSON, 2008). Metodologicamente, a análise de *Big Data* pode envolver vários estágios distintos: da "aquisição e gravação de dados" sobre "extração de informações" e "integração de dados", para "modelagem e análise de dados" e "interpretação"; cada uma das quais introduz desafios que precisam ser resolvidos (SCHINTLER; FISCHER, 2018).

De acordo com Mayer-Schönberger e Cukier (2013), a inteligência corporativa e a indústria de softwares analíticos estão mudando sua visão acerca da análise de dados: da concentração de todos os dados em uma mesma estrutura para a dispersão; uma vez que os dados estão dispersos em diversos locais, e sob diferentes formas. Os autores avançam afirmando que em troca de convívio com a confusão da proliferação de dados, é possível alcançar serviços extremamente valiosos, impossíveis pelos métodos de análise tradicional.

2.1.3 Desafios na análise de *Big Data*

De acordo com Bickel (2008), os objetivos principais das análises de grandes volumes de dados é o desenvolvimento de métodos eficazes para previsão de fenômenos futuros, proporcionando uma visão sobre a relações entre suas

características para fins científicos. Adicionalmente, levando em consideração os grandes tamanhos das amostras de dados, pode-se incluir ainda outros objetivos para suas análises, como entender a heterogeneidade e a comunalidade entre diferentes subpopulações (FAN; HAN; LIU, 2014). Neste sentido, as análises em *Big Data* são capazes de: (i) explorar as estruturas ocultas de cada subpopulação dos dados, o que tradicionalmente não é viável, podendo até serem confundidas com "outliers" quando a amostra é pequena; e (ii) extrair importantes características comuns em muitas subpopulações, mesmo quando há grandes variações individuais (FAN; HAN; LIU, 2014).

Estes objetivos carregam consigo diferentes desafios para as análises em *Big Data*, uma vez que este fenômeno é caracterizado pelo grande volume e variedade de dados a serem analisados. Considera-se ainda que um processo de análise que busque a verdade nos dados pressupõe precisão estatística. Fan, Han e Liu (2014) apresentam os principais desafios enfrentados pelos analistas em *Big Data*: (i) o alto volume de dados heterogêneos propicia o acúmulo de ruído, as correlações espúrias e a homogeneidade incidental; (ii) o alto volume combinado com a grande proporção da amostra aumenta o custo computacional, além de ser suscetível a instabilidades; (iii) as amostras massivas em *Big Data* são tipicamente um conjunto de múltiplas fontes de dados, em diferentes pontos de tempo, usando diferentes tecnologias.

Este último desafio exige ainda o enfrentamento de questões relacionadas a heterogeneidade dos dados, variações experimentais e preconceitos estatísticos; demandando procedimentos de análise cada vez mais adaptativos e robustos. O autor ainda afirma que para lidarmos com os desafios inerentes às análises em *Big Data*, precisamos de novos tipos de pensamento estatístico, aliados a métodos computacionais modernos; uma vez que o imenso volume de dados a se analisar pode não ser suportado por métodos estatísticos e computacionais tradicionais.

Embora não seja o foco deste trabalho, é válido destacar que o processamento de *Big Data* exige ferramentas de computação poderosas. A mais proeminente delas é a tecnologia de código aberto *MapReduce*, que processa grandes volumes de dados através de uma série de compactações. Outras tecnologias semelhantes também ganham destaque na literatura, como *Hadoop*, *Haloop*, *SailFish* (RADHIKA; KUMARI, 2018).

Fang, Han e Liu (2014) aprofundam sua explanação sobre os principais desafios que devem ser superados em análises de *Big Data*. Os autores apontam a

heterogeneidade, o acúmulo de ruído, as correlações espúrias, e a endogeneidade incidental como alguns dos principais obstáculos no processo de análise de dados.

2.1.3.1 Heterogeneidade

Fan, Han e Liu (2014) explicam que dados em *Big Data* são tipicamente compostos por meio da aglutinação de muitas fontes de dados, correspondentes a diferentes subpopulações. Cada subpopulação pode exibir alguns recursos exclusivos não compartilhados por outros. Em configurações clássicas onde o tamanho da amostra é pequeno ou moderado, pontos de dados de pequenas subpopulações são geralmente categorizados como “*outliers*”. No entanto, na era do *Big Data* o grande tamanho da amostra nos permite compreender melhor a heterogeneidade, tornando possível a exploração mais profunda da associação entre certas covariáveis (genes, por exemplo) e resultados raros (como doenças raras), permitindo ao analista compreender por qual razão determinadas ações (como tratamentos baseados em quimioterapia) beneficiam uma subpopulação enquanto tem potencial para prejudicar outra subpopulação. Assim, o que poderia ser considerado simplesmente como “*outlier*” em uma análise de amostras moderadas, pode ser facilmente incorporado ao corpo de dados em *Big Data*.

Em suma, um dos grandes desafios e vantagens trazidos pelo *Big Data* é entender a heterogeneidade dos dados e de suas subpopulações, o que é inviável para amostras pequenas ou moderadas (FAN; HAN; LIU, 2014). Desse modo, o analista em *Big Data* pode revelar semelhanças fracas em toda a população, graças à grande amostra tamanhos.

A heterogeneidade do *Big Data* também apresenta desafios significativos para inferência estatística. Para baixos volumes de dados, técnicas estatísticas padronizadas podem ser aplicadas, enquanto que dentro da realidade do *Big Data* é necessária a regulação cuidadosa dos procedimentos para evitar o *overfitting* ou acúmulo de ruído (STÄDLER; BÜHLMANN; VAN DE GEER, 2010).

2.1.3.2 Acúmulo de ruído

Análises em *Big Data* são caracterizadas por cálculos simultâneos, ou testes

de muitos parâmetros. Erros de estimativas podem se acumular quando uma regra de decisão ou previsão depende de um grande número desses parâmetros (FAN; HAN; LIU, 2014). Este efeito de acumulação de ruído é mais provável a medida em que a quantidade de dados analisados aumenta, de modo que o ruído pode até mesmo substituir os verdadeiros sinais.

Desse modo a discriminação do que é ruído e do que é sinal é essencial para a extração de conhecimento dos dados em análise. A seleção de variáveis desempenha um papel fundamental na superação da acumulação de ruído e na classificação de previsão de regressão. Em estatística, este tipo de problema geralmente é tratado pela suposição de dispersão (BÜHLMANN; VAN DE GEER, 2011).

2.1.3.3 Correlações espúrias

Análises de grandes volumes de dados também são suscetíveis a correlações espúrias, uma vez que muitas variáveis aleatórias não-correlacionadas podem apresentar alta correlação em amostras em altas dimensões (FAN; HAN; LIU, 2014). As correlações espúrias podem causar falsas descobertas científicas através de inferências estatísticas erradas.

Para altos volumes de dados, mesmo para um modelo estatístico simples a seleção de variáveis é um desafio devido a possibilidade da presença de algum grau de correlação espúria. Estudos mostraram que, quando a dimensionalidade é alta, as variáveis importantes podem ser altamente correlacionadas, gerando várias variáveis espúrias que são cientificamente não-relacionadas (FAN; LV, 2008).

Em estatística, uma das formas para minimizar o acúmulo de ruído é, através de uma suposição de dispersão, realizar uma correta seleção de variáveis, de modo a melhorar o desempenho preditivo e a interpretabilidade dos dados. A realização de métodos de validação cruzada também pode atenuar o problema (FAN; GUO; HAO, 2012).

2.1.3.4 Endogeneidade incidental

Endogeneidade incidental é outra questão que se apresenta na análise de grandes volumes de dados. O termo “endogeneidade” pode ser entendido, em uma

condição estatística de regressão, como o fato de que alguns dados preditores se correlacionam com o ruído residual (ENGLE; RICHARD, 1983).

A suposição de que o ruído residual não é correlacionado com os dados preditores é crucial para a validade da maioria dos procedimentos estatísticos existentes (FAN; HAN; LIU, 2014). Quando isso acontece, invalida completamente todos os métodos estatísticos aplicados na análise de dados. A endogeneidade cria, portanto, vieses estatísticos causando inconsistência na seleção dos modelos e dados (FAN; LIAO, 2012).

Diferentemente da correlação espúria, a endogeneidade incidental refere-se à existência genuína de correlações entre as variáveis, devido à alta dimensionalidade de dados. Uma analogia em relação à correlação espúria poderia ser descrita como se encontrássemos duas pessoas bastante parecidas fisicamente, mas sem relação genética entre si; enquanto que a endogeneidade incidental seria semelhante a esbarrar em um conhecido em uma grande cidade. Geralmente a endogeneidade ocorre como resultado da seleção de vieses, erros de medição e variáveis omitidas (FAN; HAN; LIU, 2014).

O autor ainda explica que este tipo de problema ocorre com frequência na análise de *Big Data*, principalmente devido a dois motivos: (i) os cientistas são capazes e tendem a coletar tantos dados quanto possível, o que aumenta a possibilidade de que alguns deles estejam correlacionados com o ruído residual, incidentalmente; e (ii) dados em *Big Data* geralmente são advindos de várias fontes, em diferentes formatos, o que aumenta a possibilidade de seleção de erros de polarização e medição, que também causam potencial endogeneidade incidental. O problema de lidar com este tipo de problema ainda não está bem entendido em estatística para altos volumes (FAN; LIAO, 2012).

A ação humana sobre a análise de *Big Data* também é um dos desafios, uma vez que os problemas estatísticos podem ser resolvidos mediante equações e testagens; enquanto que apenas dados geralmente são insuficientes para a tomada de decisão (AGRAWAL *et al.*, 2012). Este tema será discutido na seção 2.3 deste estudo.

2.2 DATA-DRIVEN DECISION MAKING

A inevitabilidade da mudança é um aspecto fundamental do ambiente

empresarial moderno. Agora, mais do que nunca, as mudanças parecem mais significativas e não menos imprevisíveis (ALEXANDER; KUMAR; WALKER, 2018).

Há abundantes evidências de que empresas de todas as indústrias alcançam melhoras significativas em seu desempenho através do uso de dados ou métodos analíticos (GLASS; CALLAHAN, 2014; SIMON, 2013; WALKER, 2015). Estudos sugerem que a tomada de decisão orientada por dados aumenta o desempenho (LAVALLE *et al.*, 2011), a produção e a produtividade (BRYNJOLFSSON; HITT; KIM, 2011) em empresas com fins lucrativos, além da eficácia das decisões gerenciais (LEROUX; WRIGHT, 2010). Pesquisa da Accenture (2014) com mais de 1.000 empresas que completaram pelo menos uma implementação de soluções de *Big Data* mostrou que 92% estavam satisfeitas com os resultados.

O uso de dados na tomada de decisão leva as empresas a uma nova realidade, tornando possíveis práticas que até mesmo os funcionários mais experientes não poderiam adotar sozinhos. Para exemplo, a empresa americana Amazon deu a tarefa de fixação de preços de seus produtos a um algoritmo; de forma a praticamente excluir a presença humana nesta função. Essa prática resultou em menor cedência de descontos, menor acúmulo de estoque e maior rapidez na introdução de novos produtos (CHAN; PEYNE, 2017).

Davenport (2010) define a tomada de decisão baseada em dados (*Data-Driven Decision*, ou DDD) como o "uso de dados e análises para entender e gerenciar uma empresa de forma mais eficaz". Este conceito inclui a disponibilidade e o uso de dados para apoiar decisões e para a criação de novos produtos (BRYNJOLFSSON; HITT; KIM, 2011). Definição semelhante é proposta por Ikemoto e Marsh (2007), na qual a tomada de decisão baseada em dados é o processo pelo qual os administradores coletam e analisam dados para orientar uma série de decisões.

A tomada de decisão baseada em dados pode envolver análises abrangentes e multidimensionais de uma miríade de informações, sem quaisquer suposições ou restrições. Assim, as decisões obtidas têm maior potencial de imparcialidade e objetividade. Portanto, a tomada de decisão baseada em dados tem vantagens óbvias sobre a tomada de decisão baseada em modelos intuitivos (LONG, 2018).

2.2.1 Teoria da decisão

A teoria da decisão oferece uma abordagem formal para a tomada de decisão,

sendo frequentemente revisitada em estudos que analisam os processos decisórios no ambiente corporativo; sendo considerada um ramo da ciência comportamental (GILBOA; ROUZIOU; SIBONY, 2018). Os autores ainda explicam que os fundamentos para Teoria da Decisão nascem em meados do século XVII, na afirmação de Pascal, ao considerar a necessidade da escolha racional como fundamento para a entrada na vida eterna. Esta linha de pensamento introduziu ideias como a matriz de decisão, estratégia dominante, probabilidade subjetiva e maximização esperada da utilidade para lidar com a incerteza (HACKING, 1975; BERNOULLI, 1738; BAYES, 1763). Deste ponto de partida surgiram as abordagens de matematização da economia, a invenção da teoria dos jogos e o desenvolvimento de ferramentas estatísticas de análise, buscando entender matematicamente a tomada de decisão humana (GILBOA; ROUZIOU; SIBONY, 2018).

Um dos trabalhos seminais da teoria da decisão é de Simon (1947), "*Administrative Behaviour: A study of decision-making processes in administrative organization*". Seu estudo trouxe uma abordagem alternativa à gestão científica de Taylor (1911), ao demonstrar que os gestores não possuem um controle perfeito sobre as organizações, uma vez que não detém conhecimento perfeito; introduzindo, portanto, uma maior ênfase a dimensão humana nas tomadas de decisão.

Desse modo, a maioria das situações de decisão envolve algum nível de incerteza ou risco. Por isso faz-se necessário o entendimento de que existem dois tipos de elementos envolvidos no processo de tomada de decisão: aqueles que estão sob o controle do decisor, e os que não estão; configurando-se como variáveis aleatórias. Estes elementos podem ser fontes externas de incerteza (como as forças da natureza ou a ação concorrencial) ou ainda as próprias escolhas de outros tomadores de decisão (GILBOA; ROUZIOU; SIBONY, 2018).

Nesse sentido o tomador de decisão se depara com os "estados do mundo"; possíveis cenários que podem se desdobrar e exigir outras decisões. Os estados do mundo podem ser mais ou menos prováveis, onde cada possibilidade pode ser medida por probabilidades. No entanto, o decisor não tem qualquer tipo de controle sobre este tipo de variável, de modo que não pode escolher de fato entre eles. E se pudesse, sua decisão incorreria em erro, uma vez que estaria apenas aplicando suas preferências diante de cenários inteiramente aleatórios; isso porque apesar de "previstos" pelo tomador de decisão, eles podem simplesmente acontecer ou não (GILBOA; ROUZIOU; SIBONY, 2018). Os estados de mundo representam, portanto,

a própria incerteza, uma vez que as combinações das ações que podem ser tomadas sob estes cenários constituirão um determinado resultado único, representando os benefícios líquidos de escolher uma única ação dentro de um conjunto de ações alternativas (POLASKY *et al.* 2011).

Gilboa, Rouziou e Sibony (2018) sugerem três tipos distintos de decisões. No primeiro tipo, o problema é bem especificado e permite uma solução algorítmica e lógica. Nesse caso a intuição é irrelevante. Desse modo o tomador de decisões só precisa possuir os dados necessários acerca do problema a ser resolvido.

O segundo tipo de decisão é aquele em que não há um único “melhor” resultado desejável. Aqui a intuição do decisor desempenha um papel proeminente. Nesse caso, enquanto métodos teóricos tem a função de suportar e dar sentido desta intuição; formulando metas e restrições ao tomador de decisão, conferindo racionalidade as suas linhas de pensamento. Há também situações nas quais um tomador de decisão se depara com problemas mal estruturados, de difícil articulação lógica. Nestes casos, a teoria da decisão auxilia o tomador fornecendo modelos de decisão que podem ser utilizados, como a matriz de decisão para construção de cenários futuros.

2.2.2 O paradigma da utilidade esperada

Gilboa, Rouziou e Sibony (2018) denominam como “atos” as escolhas possíveis em um processo de decisão. Os autores desenham uma matriz de decisão, onde as linhas demonstram os atos possíveis, as colunas listam os estados de mundo possíveis, e seu conteúdo são os resultados esperados. Para cada ato (linha) e cada estado (coluna) existe um resultado único. Uma matriz de decisão típica pode ser ilustrada conforme o Quadro 3, abaixo.

Quadro 3 – Matriz de decisão

	Cenário 1	Cenário 2
Ação 1	Resultado esperado	Resultado esperado
Ação 2	Resultado esperado	Resultado esperado

Fonte: Adaptado pelo autor a partir de Gilboa, Rouziou e Sibony (2018).

Diante da matriz, um tomador de decisão deve especificar os resultados desejados por uma função de utilidade (aferindo algum valor sobre cada resultado

esperado, não necessariamente monetário), e seu julgamento (sobre os cenários), aferindo então percentuais para cada resultado provável. A partir daí ele pode estabelecer uma série de critérios de decisão com base nesses conceitos, sendo que o critério mais popular é o *paradigma de utilidade esperada* (GILBOA; ROUZIOU; SIBONY, 2018).

Este paradigma sugere que cada escolha (ato) seja avaliada de acordo com um único número, que é a expectativa matemática de utilidade (ao tomador de decisão) que aquele resultado pode gerar, considerando os cenários prováveis. Logo, o paradigma sugere que a escolha com maior utilidade esperada seja a escolhida; comparativamente com as demais opções (GILBOA, 2010). Para os autores, esta abordagem metodológica mostra-se aplicável de forma especial às decisões baseadas em dados, uma vez que cada vez mais algoritmos são desenvolvidos para encontrar estatísticas e técnicas rápidas e úteis para análises de dados; de modo a ampliar a dimensão racional dos processos decisórios

Por outro lado, Gilboa, Rouziou e Sibony (2018) afirmam que o Paradigma da Utilidade Esperada não é utilizada de forma extensiva nas tomadas de decisão organizacionais. Os autores explicam que as causas para isso são os desafios práticos cotidianos que impedem os tomadores de decisão de adotar o *paradigma de utilidade esperada*. Entre eles:

- a) frequentemente, as probabilidades não podem ser atribuídas aos cenários (estados de mundo), uma vez que eles estão sujeitos a eventos únicos. Nessas situações, o paradigma clássico sugere o uso da subjetividade, mas não oferece as ferramentas para aferir estas probabilidades de forma subjetiva. Em situações que envolvam mais de um tomador de decisão, concordar com as probabilidades subjetivas parecer exercício em futilidade ou uma fonte de divergências sem fim;
- b) os tomadores de decisão podem não conseguir estabelecer uma função de utilidade para quantificar seus julgamentos sobre os resultados desejados. Em alguns casos, os teóricos da decisão podem oferecer questionários úteis que podem "calibrar" a função de utilidade. Por exemplo, se os resultados esperados forem apenas monetários pode-se fazer perguntas simples sobre a decisão sob risco a partir de cada decisão. No entanto, existem problemas de decisão nos quais os resultados devem ser julgados de acordo com uma multiplicidade de critérios, e estes nem

sempre se prestam a fórmulas matemáticas simples;

- c) às vezes, a própria estrutura da matriz de decisão é difícil de imaginar. Em muitos casos o conjunto de cenários e de escolhas é múltiplo e complexo, e o tomador de decisão pode simplesmente perder muito tempo tentando descrever uma lista exaustiva de cenários. Em alguns casos, como no nível estratégico das organizações, o conjunto de escolhas possíveis é teoricamente infinito.

Tendo em vista estes desafios, nota-se que características intrínsecas do processo de tomada de decisão exigem a compreensão situacional de seus efeitos por parte dos responsáveis por ela. Somam-se a isso os desafios comportamentais.

2.2.3 Desafios comportamentais à tomada de decisão

Gilboa, Rouziou e Sibony (2018) ainda listam fatores comportamentais como desafios para o uso da teoria da decisão, apoiando-se em evidências da psicologia comportamental. Segundo os autores, os seres humanos consideram impossível definir a utilidade de vários resultados de uma maneira que esteja em conformidade com os princípios axiomáticos da tomada de decisão racional. Mesmo quando os resultados esperados que podem ser totalmente quantificados em termos monetários, somos inconsistentes na atribuição de valores a resultados que envolvem perdas percebidas. Um problema ainda mais fundamental é que nossas preferências não são fixas: dependendo de como o problema é apresentado, nossas preferências podem mudar, e com isso também muda a forma de classificação das escolhas e os resultados esperados de diferentes formas.

O medo da análise posterior da decisão (especialmente no contexto organizacional) é outro fator que afeta a tomada de decisão podendo reduzir a tolerância ao risco, uma vez que crédito e culpa serão atribuídos após o fato. Outro fator comportamental diz respeito ao fato de que não temos esperança de quantificar as probabilidades. Acreditamos que conjunções plausíveis de eventos sejam mais prováveis do que suas causas, o que viola um princípio básico de probabilidades. Tendemos a negligenciar taxas de base, resultando em estimativas amplamente distorcidas quando da tentativa de inferências bayesianas. Além disso, superestimamos consistentemente nossas chances de sucesso, e até mesmo a probabilidade de eventos de sorte fora do nosso controle, mesmo que o tomador de

decisão esteja bem-intencionado. E, finalmente, ainda segundo Gilboa, Rouziou e Sibony (2018), não somos muito eficazes em definir os atos que devemos considerar. Inércia - e pura falta de imaginação - pode nos levar a uma excessivamente satisfação com o *status quo*, de modo que deixemos de gerar escolhas alternativas por completo. Não tendemos a gerar opções suficientes, nos satisfazemos prematuramente; assim nos tornamos propensos a desconsiderar dados negativos sobre as opções que estamos considerando.

Segundo Gilboa, Rouziou e Sibony (2018), a crença na intuição também é pode ser entendida como um desafio à tomada de decisão. Para os autores há um subgênero abundante e bem-sucedido de literatura de gestão popular exaltando as virtudes da intuição como um guia para tomada de decisão gerencial; encontrado em estudos como os de Gladwell (2007) e Klein (2007). Porém os autores defendem que praticamente todos os executivos provavelmente concordariam, em princípio, com os axiomas da racionalidade e com a tomada de decisão baseada em lógica e na utilidade esperada. Sendo igualmente improvável que um tomador de decisão não queira adotar uma opção de escolha que resulte potencialmente em maiores retornos utilitários em todas as situações.

Contribuindo com essa discussão, Weigert (2017) afirma que decisões tácitas que envolvem criatividade, ou um grande número de fatores de entrada exigiria um enorme poder computacional para seu processamento. Para o autor é neste ponto que os humanos venceriam os computadores, utilizando seu instinto, que "armazena" milhões de anos de situações comparáveis. Porém este tipo de decisão é inerentemente difíceis de padronizar e reproduzir.

2.3 DADOS NA TOMADA DE DECISÃO

Tradicionalmente as organizações conduzem a maior parte suas decisões diárias usando regras simples, baseadas na experiência ou prática (CALLINGHAM, 2004). Porém, o uso de dados está mudando completamente o processo de tomada de decisão nas empresas, configurando-se em um ativo para a melhora no desempenho organizacional (CHAN; PEYNE, 2017). Em estudo realizado em 2015, 63% das empresas relataram usar dados caracterizados como *Big Data*, enquanto em 2012 eram apenas 5% (BEAN, 2016). A mesma pesquisa revelou que, em 2015, 53% empresas afirmaram haverem contratado um diretor de dados, contra apenas 12% na

mesma situação em 2012.

Estudos sugerem que existe uma conexão entre a decisão baseada em dados e o desempenho das organizações (BRYNJOLFSSON; HITT; KIM, 2011). Com o rápido desenvolvimento das tecnologias de computação, os recursos de coleta e processamento de dados ganharam avanços significativos em campos como análise de *Big Data*, *business intelligence* e *data mining*. Uma análise de dados abrangente e multidimensional pode melhorar a eficiência da tomada de decisões; logo, o paradigma de tomada de decisão orientado à dados torna-se predominante (XU *et al.*, 2014). Este paradigma refere-se à prática de basear decisões na análise de dados, e não apenas na intuição (PROVOST; FAWCETT, 2013). No entanto, a utilização incorreta dos dados pode levar a decisões equivocadas, ocasionando perdas financeiras para as empresas. A tomada de decisão baseada em dados ajuda as empresas a evitar desperdícios de tempo e dinheiro (CHAN; PEYNE, 2017).

O uso de grandes volumes de dados nas empresas permite ainda uma maior transparência em seus processos decisórios. Com a alta competitividade, há crescente desconfiança dos mercados em decisões baseadas puramente na intuição; logo, os dados tornam a tomada de decisão aparentemente mais neutra (CAPGEMINI, 2011).

A possibilidade de análise em tempo real é outra característica dos dados orientando o processo decisório. Segundo Chan e Payne (2017) no passado os dados revelavam fatos apenas através de algum tipo de relatório ou feedback. No entanto, as tecnologias de dados evoluíram, permitem as organizações seu uso para obter dados em tempo real, e até mesmo para o futuro através de extrapolações.

Usar dados para tomar decisões de forma eficaz requer pelo menos dois elementos-chave: cultura e processos (MAXWELL; ROTZ; GARCIA, 2016). A cultura da organização deve apoiar a decisão baseada em dados, de forma que seu pessoal valorize e abrace o uso de dados para desenvolver e implementar estratégias (PRESKILL; BOYLE, 2008). A cultura influencia as percepções e a inclinação da equipe em relação ao uso de dados, conduzindo o processo de tomada de decisão (HUFFMAN *et al.*, 2006). Um processo coordenado de atividades sequenciais de decisão baseada em dados também é requerido, uma vez que a organização coleta, analisa e usa dados. A qualidade de cada estágio de uso dos dados determina a qualidade das decisões tomadas. Se os dados não forem coletados corretamente, as organizações não poderão analisar as informações para tirar conclusões; e se não

forem analisados de forma consistente e correta, a equipe poderá obter conclusões incorretas (MAXWELL; ROTZ; GARCIA, 2016).

Portanto, extrair valor dos dados não é uma tarefa simples, e requer muito mais do que o investimento costumeiro das empresas em tecnologia da informação (WEIGERT, 2017).

2.3.1 Processo de extração de informação dos dados para tomada de decisão

Para extrair informações dos dados de modo suportar a tomada de decisão, é necessário um processo com pelo menos 5 etapas: coleta, armazenamento, processamento, gerenciamento e análise final dos recursos de dados (KUDYBA, 2014). Já para Jeffery (2010), o sucesso da tomada de decisões baseada em dados depende da qualidade da organização dos dados, sua análise, bem como do processo de desenhar informações significativas a p dos dados.

Labrinidis e Jagadish (2012) propõem um *framework* – também apresentado por Agrawal *et al.* (2012) – para a decisão baseada em grandes volumes de dados, dividindo o processo de extração de insights dos dados em cinco estágios, subprocessos de duas etapas: gerenciamento de dados e análise. O gerenciamento de dados envolve processos e tecnologias de suporte para aquisição, armazenamento, preparação e recuperação de dados. Enquanto que a análise se refere as técnicas usadas para analisar e extrair inteligência dos dados (GANDOMI; HAIDER, 2015).

Figura 1 – *Framework* para análise de dados



Fonte: Adaptado de Gandomi e Haider (2015).

A extração de insights de grandes volumes de dados se estrutura nos

processos de gerenciamento e de análise de dados. A primeira etapa trata do processo de coleta e armazenamento de dados de diferentes fontes (SEBEI; TAIEB; AOUICHA, 2018). O primeiro ponto a ser observado por analistas e tomadores de decisão é a qualidade da fonte de origem dos dados. Para Agrawal *et al.* (2012), apenas com a proveniência adequada, será possível definir todos os processamentos subsequentes que dependem dessa etapa. Desse modo podem ser gerados com precisão os metadados (dados sobre dados) que suportem os processos posteriores. Algumas fontes de dados podem produzir quantidades surpreendentes de dados brutos, e muitos desses dados são inúteis e devem ser filtrados. O desafio é definir esses filtros de tal forma que eles não descartem informações úteis (LABRINIDIS; JAGADISH, 2012). Isso porque dependendo da variedade e da granularidade (ou seja, o nível de detalhe) dos dados coletados pode ser difícil prever quais insights podem ser obtidos (GÜNTHER *et al.*, 2017). Considerando que uma alta quantidade de dados é de fácil obtenção pelas empresas, é necessária a obtenção de dados corretos, uma vez que a qualidade não depende apenas dos dados, mas também do próprio processo de coleta dos dados (CHAN; PEYNE, 2017). Se os dados coletados não estiverem corretos, não poderão apoiar o processo de tomada de decisão (JANSSEN; VAN DER VOORT; WAHYUDI, 2017). A utilização de grandes volumes de dados contribuiu para a criação de uma atitude empresarial que coleta dados sem um propósito pré-definido, promovendo uma abordagem indutiva (GÜNTHER *et al.*, 2017). Ou seja, por vezes os dados podem ser coletados livremente *a priori*, enquanto que a busca por significados é executada *a posteriori* (BHOLAT, 2015).

A segunda etapa visa a preparação e limpeza dos dados para a etapa de processamento (SEBEI; TAIEB; AOUICHA, 2018). Frequentemente, as informações coletadas não estarão em um formato pronto para análise, uma vez que podem advir também de diversas fontes diferentes. Por esse motivo faz-se necessário um processo de extração de informações fontes disponíveis, convertendo-as para um formato adequado para análise (AGRAWAL *et al.*, 2012). Já a limpeza de dados pressupõe restrições de processamento de dados, de forma a permitir a extração de dados válidos ou modelos de erro bem compreendidos (AGRAWAL *et al.*, 2012).

A terceira etapa é de integração, agregação e representação dos dados. Aqui os dados são convertidos em formatos adequados para a análise (SEBEI; TAIEB; AOUICHA, 2018). Para uma análise eficaz em larga escala, os processos precisam acontecer de maneira completamente automatizada; pois, dada a heterogeneidade do

fluxo de dados, não é suficiente apenas registrá-lo e lançá-lo em um repositório (AGRAWAL *et al.*, 2012). As diferenças na estrutura de dados e na semântica devem ser expressas em formulários que sejam compreensíveis por computador e, em seguida, "roboticamente" resolvíveis, sugerem Agrawal *et al.* (2012). Os autores ainda explicam que mesmo para análises mais simples, que dependem de apenas um conjunto de dados, o design de banco de dados adequado deve ser pensado, uma vez que há muitas maneiras alternativas de armazenar informações. Como tal, os dados coletados para uma finalidade também podem ser usados para muitas outras finalidades, porque esses dados podem ser combinados e analisados de novas maneiras (AALTONEN; TEMPINI, 2014).

No processo de análise dos dados inicialmente ocorre a modelagem. Nesta etapa são realizadas consultas aos dados minerados, com o objetivo analisar os dados (SEBEI; TAIEB; AOUICHA, 2018). Aqui o grande volume de dados mostra suas vantagens frente a pequenas amostras, uma vez que as estatísticas gerais obtidas a partir de padrões de dados e análises de correlação geralmente superam as flutuações individuais, frequentemente revelando conhecimentos e padrões ocultos (AGRAWAL *et al.*, 2012). É neste ponto que a organização desenvolve e aplica seus próprios modelos para análise de dados, de forma a explorar a redundância de informações como forma de compensar os dados ausentes, cruzar casos conflitantes, validar relações confiáveis entre dados, analisar clusters de informação e descobrir inter-relações de dados e modelos ocultos; destacam Agrawal *et al.* (2012). Os autores também afirmam que as análises realizadas a partir de dados integrados, limpos, confiáveis e de acesso eficiente utilizando interfaces de consulta adequadas, podem ajudar a melhorar a qualidade e confiabilidade dos dados, através da compreensão correta de sua semântica. Diversas técnicas e ferramentas podem ser aplicadas na análise de dados, como análise de texto, análises de vídeo, de sentimentos, de conteúdo, análises em tempo real e análises de grafos (GANDOMI; HAIDER, 2015). Estas técnicas podem ser usadas tanto para análise retroativa, isto é, para entender acontecimentos passados quando para análises preditivas, que buscam projetar situações futuras através de técnicas de extrapolação. Agrawal *et al.* (2012) ainda lembram que a visualização das análises de dados deve estar contida em uma interface que facilite o entendimento dos dados. Para os autores os analistas de hoje precisam embalar e apresentar resultados em visualizações poderosas que auxiliem a interpretação e deem suporte à colaboração do usuário. Neste sentido, os

tomadores de decisão precisam ser capazes de não apenas ver os resultados, mas também entender por que estão vendo esses dados.

A etapa final, da interpretação dos dados, consiste em entender a análise realizada, considerando a complexidade dos dados, de modo a tomar as decisões corretas sobre um problema particular enfrentado (SEBEI; TAIEB; AOUICHA, 2018). Essa etapa é necessária pois possui a capacidade de analisar grandes volumes de dados é de valor limitado se os usuários não puderem entender a análise (AGRAWAL *et al.*, 2012). Em última análise, o decisor precisa interpretar esses resultados das análises de dados. Geralmente esse trabalho envolve examinar todas as suposições encontradas, refazendo a análise, considerando que mesmo os melhores processos de tratamento de dados são passíveis de erro (AGRAWAL *et al.*, 2012). Por essas razões Agrawal *et al.* (2012) defendem que nenhum usuário responsável concederá autoridade suprema a um sistema de computador. Os autores sugerem que usuário tentará entender e verificar os resultados produzidos pelo computador, realizando suas próprias análises e tomando as decisões.

Agrawal *et al.* (2012) comentam que apenas os resultados de uma análise de dados raramente são suficientes para a tomada de decisão. Em vez disso, é preciso considerar informações suplementares que expliquem como cada resultado foi alcançado, além de informações diversas que possam ter impacto na decisão.

Algoritmos não podem substituir a perícia tradicional, a sensibilidade e a criatividade advinda da experiência pessoal; mas devem complementar estas características sendo subservientes a elas (ANDERSON, 2009). Os desafios cognitivos para análise de grandes volumes de dados exigem alguma margem de manobra para a intuição, e um padrão de julgamento que somente pode ser adquirido a partir do conhecimento humano e do senso comum (ROSE, 2017).

2.3.2 Desafios à tomada de decisão baseada em dados

Alguns desafios permeiam o processo de tomada de decisão baseada em dados, em suas dimensões técnicas e humanas. Em muitas organizações as habilidades analíticas estão concentradas em poucos colaboradores, o que pode limitar a assertividade das decisões (CAPELLÁ; HORNE; SHAH, 2012). Apenas uma equipe multifuncional pode garantir a integridade dos dados, a calibração do modelo de decisão e a correta interpretação dos dados coletados (SPILLECKE; UMBLIJS,

2013).

As razões para incluir a contribuição humana no processo de tomada de decisão baseada em dados são variadas. O acréscimo de conhecimento tácito às decisões pode ampliar o valor dos resultados obtidos através dos dados, ou ainda agindo mitigar riscos potenciais advindos de dados mal classificados ou coletados (WEIGERT, 2017).

As empresas cada vez mais procuram equipes flexíveis e multifuncionais, a fim de para expandir as habilidades de análise para todos os empregados, sem restringi-las a um pequeno grupo (CHAN; PEYNE, 2017). Um dos fatores de sucesso em projetos que utilizem grandes volumes de dados é justamente a combinação de diferentes habilidades em equipes multidisciplinares (GAO; KORONIOS; SELLE, 2015).

No momento, um dos profissionais de maior responsabilidade nesse processo são os cientistas de dados. A Harvard Business Review (2012) dedicou um artigo a esta profissão: "*Data Scientist: The Sexiest Job of the 21st Century*", descrevendo o cientista de dados como o único que pode "descobrir o tesouro na bagunça", o que ajuda os decisores "conversar continuamente com os dados". Isso destaca o fato de que as equipes presentes nas organizações ainda não estão preparadas para lidar com grandes volumes de dados, o que dificulta o uso de *Big Data* nos processos de decisão. Especialistas na análise de *Big Data* para auxiliar a tomada de decisão ainda são raros, e ativamente procurados pelas empresas (CHAN; PEYNE, 2017).

Para Günther *et al.* (2017), embora seja esperado que os tomadores de decisão sejam suficientemente capazes de executar suas próprias análises, eles geralmente ainda requerem a ajuda de analistas para entender e significar dados (TAMM; SEDDON; SHANKS, 2013; NAMVAR; CYBULSKI, 2014). Na era do *Big Data*, "os analistas precisam envolver-se verdadeiramente nos negócios" (OLBRICH, 2014); fato que implica a necessidade de colocar os analistas de dados próximos às decisões corporativas (CHATFIELD; REDDICK; AL-ZUBAIDI, 2015).

Para Alavi e Leidner (2001) existe uma clara ligação entre dados e conhecimento. Dados são apenas unidades informacionais brutas que apenas podem fornecer informações utilitárias quando devidamente processados, o que exige conhecimento por parte dos profissionais que lidam com estes dados. A aplicação ativa de seu conhecimento atua como uma fonte de vantagem competitiva para as empresas (ALAVI; LEIDNER, 2001). Tomadores de decisão se deparam com

informações advindas de fontes cada vez mais variadas e complexas; porém não há um crescimento proporcional em habilidades cognitivas humanas ou em orçamentos de pesquisa que sejam capazes de auxiliá-los a extrair a máxima eficiência de cada informação (LUTZ; BOUCHER, 2017).

Nesse sentido, a própria natureza das tecnologias se configura em um desafio para a tomada de decisão baseada em dados. Isso porque parte importante do trabalho do decisor é classificar suas alternativas, de forma a reduzir as incertezas. O desafio desse contexto são as lacunas de dados e informações imprecisas, naturais no uso de tecnologias emergentes (PROMENTILLA *et al.*, 2018)

A necessidade de síntese é outro desafio. Isso se deve ao fato de que tomar decisões baseadas em múltiplas fontes de dados é uma tarefa complexa. Os processos de redução de dados, classificação, limpeza, e apresentação por vezes não são acompanhados (tampouco bem compreendidos) pelos tomadores de decisão o (LUTZ; BOUCHER, 2017). Lim (2016) lembra que grandes volumes de dados, como no caso do *Big Data*, requerem metodologias de análise que considerem justamente grandes volumes de informação, variadas fontes e até uma certa bagunça. Porém, para o autor, isso compensa a precisão exigida de amostras menores de informação em modelos tradicionais, com os quais alguns profissionais podem estar acostumados.

Nesse sentido, Bell (2014) alerta que muitos tomadores de decisão não possuem motivos claros e estratégicos para a implementação da tomada de decisão baseada análises de grandes volumes de dados. É o caso de executivos que coletam dados apenas porque “todos estão fazendo isso”, ou ainda por temor de perder eventuais vantagens competitivas, através do raciocínio “se outras empresas coletarem dados e a minha não? ”.

Face a esta realidade, uma grande parcela dos dados coletados pelas empresas sequer é analisada, pois soma-se a esse cenário a escassez de profissionais decisores com habilidades para processar dados em massa e transformá-los em informação significativa; bem como as dificuldades de julgamento inerentes aos processos analíticos (BELL, 2014). Outro motivo para desperdício de dados é apontado por Günther *et al.* (2017), ao relatar que muitos analistas gastam mais da metade de suas horas de trabalho “limpando dados”, ou seja, preparando estes dados para inseri-los nas plataformas de software das suas organizações. Soma-se a esse contexto a própria burocracia das organizações, uma vez que neste

tipo de sistema todos os setores desejam controlar as funções de análise, de modo a requerer com isso crescimento orçamentário, de modo que a estrutura ideal de tomada de decisão pode ser prejudicada (BELL, 2014). Ainda existe o desafio cultural e as preocupações com a privacidade (MCAFEE *et al.*, 2012).

Chan e Payne (2017) chamam a atenção para o fato de que mesmo em empresas que usam dados de forma avançada, a dimensão humana da intuição ainda desempenha um papel decisivo no processo decisório. Eles citam o caso do Google, uma das empresas de maior uso extensivo de *Big Data* no mundo. Mesmo ali, a ideia de desenvolvimento de seu automóvel autônomo veio de um gerente que havia perdido um de seus amigos em um acidente de trânsito. Ele tinha a intuição de que aquele poderia ser um produto novo, e isso antes da emergência do fenômeno *Big Data*.

Para os autores, integrar dados e intuição talvez seja um dos maiores desafios das organizações em seus processos decisórios; tendo em vista que o ser humano tende a colocar maior crença em suas intuições, que são o resultado de seus sentimentos e experiências. Os dados devem ser ferramentas para orientar a criatividade e a intuição, reduzindo custos e melhorando a produtividade ao fornecer faixas de reflexão. A criatividade e a intuição lidam com o aspecto humano que o *Big Data* não consegue ainda entender, como desejo e a emoção (CHAN; PEYNE, 2017).

3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa realizou um estudo de caso único, de natureza exploratória para identificar e documentar o conjunto de práticas recorrentes realizadas para análise de dados caracterizados como *Big Data*. A proposta é consonante com Gil (2011), que apresenta o estudo de caso como a análise exaustiva de um ou poucos objetos, resultando em formas ampliadas e profundas de conhecimento; e com Stake (1995) que define “caso” como uma unidade específica, um sistema delimitado cujas partes são integradas. Posicionamento presente também em Vergara (2014), ao afirmar que um estudo de caso é o circunscrito a uma ou poucas unidades; como uma única pessoa, uma única comunidade ou mesmo um país.

Pesquisas conduzidas por estudos de casos são aplicáveis em uma ampla variedade de situações, tendo como objetivo principal a contribuição com o conhecimento acerca dos fenômenos estudados (YIN, 2015). Nesse sentido, este tipo de metodologia de pesquisa estuda a vida real, propiciando análises causais de fenômenos em contextos nos quais outros tipos de levantamentos de dados ou realizações e experimentos não seriam possíveis (GIL, 2011).

Nas palavras de Clark (2006):

Os estudos de caso são a base para os resultados de pesquisa em locais específicos e eras específicas. [...] os estudos de caso têm seu conceito no mundo real ao invés de em imagens soltas de um mundo ideal. Para gerar e confirmar conceitos eles servem como experimentos de vida real (CLARK, 2006, p. 41).

Este método de pesquisa possibilita a análise da complexidade e especificidade de um fenômeno ativo. Um caso deve ser estudado quando ele próprio é de interesse excepcional, de modo que os detalhes envolvidos na interação do caso com os seus próprios contextos são de precioso interesse científico (STAKE, 1995).

O estudo de caso, com metodologia qualitativa se justifica para esta pesquisa, pois tem a capacidade de cortar transversalmente disciplinas, fornecendo ampliada visão sobre o objeto de estudo (DENZIN; LINCOLN, 2008); fato potencialmente aplicável no instrumento de pesquisa. Stake (2011) salienta que a abordagem qualitativa é resultado de raciocínio prioritariamente baseado na percepção e na profundidade da compreensão humana. Adicionalmente, o estudo qualitativo é experiencial, interpretativo e situacional, não buscando a generalização, mas

provendo conhecimento através de experiências reais (CROWE *et al.*, 2011). Pensamento semelhante advém de Järvensivu e Törnroos (2010), na afirmação de que estudos de caso não devem ser medidos por sua contribuição generalista, mas sim considerando suas contribuições contemporâneas à ciência. Os autores ainda defendem que o objetivo verdadeiro dos estudos de caso não é revelar verdades universais, mas prover conhecimento historicamente contextualizado. Este entendimento está presente na pesquisa proposta, uma vez que busca analisar processos através de experiências contemporâneas reais e historicamente contextualizadas.

Para Yin (2015) o estudo de caso único é adequado sob três justificativas: quando o caso é crítico, único ou revelador. A pesquisa aqui desenvolvida caracteriza-se como caso único; uma vez que para o autor o caso único reúne características instigantes e extremamente raras, presentes apenas naquela situação ou objeto de estudo, o que justifica sua análise em profundidade. A unidade de análise pesquisada reúne características instigantes à pesquisa de tomada de decisão baseada em dados, pela própria complexidade e volume de dados inerentes a operações financeiras. Neste sentido, o que confere características únicas ao caso estudado é a natureza de sua organização em sistema cooperativo; o que revelou idiosincrasias específicas em seu processo de tomada de decisão.

Stake (1995) caracterizou três tipos principais de estudo de caso: intrínseco, instrumental e coletivo. A pesquisa realizada configurou-se como estudo de caso instrumental, uma vez que, esse tipo de estudo utiliza-se de um caso particular para apreciar de maneira mais ampla um fenômeno. Este estudo utilizou-se desse método para entender melhor o fenômeno *Big Data* aplicado a tomada de decisão baseada em dados.

Yin (2015) enumera seis fontes de evidências para realização de estudos de casos: documentos, registros em arquivos, entrevistas, observações diretas, observação participante e artefatos físicos. Para o caso da pesquisa aqui descrita, utilizou-se a entrevista semiestruturada como forma primária de coleta de dados, a observação direta do pesquisador *in loco*, e análise documental.

3.1 CARACTERIZAÇÃO DA UNIDADE DE ANÁLISE

O estudo de caso teve como unidade de análise um sistema de cooperativas

de crédito. Mais precisamente duas Cooperativas Singulares e um Centro Administrativo integrantes deste sistema. O sistema opera mais de 300 produtos e serviços financeiros, para seus mais de 4 milhões de associados, distribuídos em 22 estados brasileiros e no distrito federal. A empresa conta com mais de 1.600 agências, contidas em mais de 100 Cooperativas Singulares; empregando mais de 24 mil colaboradores, segundo dados de relatórios internos. Sua organização em um sistema que promove a descentralização da tomada de decisão nos permite analisar os processos de tomada de decisão em nível regional e em alto nível hierárquico; situação que favoreceu os esforços de pesquisa.

Em volume de ativos, o sistema administra mais de 96 bilhões de Reais. Neste momento, a instituição opera em forte crescimento. Segundo relatórios financeiros, seu resultado líquido total em 2018 foi mais de 15% maior do que ano anterior, ultrapassando os 2 bilhões de Reais em lucro. O sistema administra mais de 50 bilhões de Reais em crédito concedido.

Por se tratar de uma cooperativa, seus valores e propósito ocupam lugar de destaque em sua cultura. Segundo o site da empresa:

Entre os diversos tipos de associação cooperativa estão as cooperativas de crédito. Criadas para oferecer soluções financeiras de acordo com as necessidades dos associados. Elas são um importante instrumento de incentivo para o desenvolvimento econômico e social. Isto porque utilizam seus ativos para financiar os próprios associados, mantendo os recursos nas comunidades onde eles foram gerados (SITE DA EMPRESA, 2018).

Em 2018 a empresa recebeu diversos prêmios. Entre eles o “Melhores Empresas para Você Trabalhar” promovido pela revista Você S/A; “Melhores & Maiores” da revista Exame; “Valor 1000” promovido pelo jornal Valor Econômico; entre outros.

O sistema é composto basicamente por 5 entes: Confederação e Banco Cooperativo (além de outras empresas de serviços financeiros); Cooperativas Centrais (responsáveis pelas estratégias em macrorregiões do país), Cooperativas Singulares (de atuação regional nos estados da federação), Agências (para atendimento direto ao público associado) e Centros Administrativos (responsáveis por atividades de *staff* e apoio administrativo).

Um esquema visual do sistema, que permite a identificação do fluxo de tomada de decisão geral, tema deste trabalho, pode ser encontrado na imagem

abaixo, elaborada pelo pesquisador. Este esquema foi validado com um dos entrevistados.

Figura 2 – Fluxo de tomada de decisão



As Cooperativas Singulares podem ser entendidas como o coração de todo o sistema. Elas contam com regramentos impostos pelas Cooperativas Centrais e pela Confederação (além de obedecer a marcos regulatórios inerentes a atividade financeira); porém gozam de considerável autonomia. Elas podem estabelecer as próprias políticas de concessão de crédito; definir quantas e onde serão instaladas suas agências; além de criar e gerenciar dados e informações próprias que podem ou não ser compartilhadas com o sistema. Em termos de estrutura organizacional, contam com Conselho Administrativo, Presidência, e com um corpo de diretores e gestores operacionais, responsáveis pela gestão conjunta e individual de suas agências.

As decisões estratégicas são definidas em assembleias, com a participação e voto dos seus associados. Essas decisões depois serão traduzidas para o seu corpo operacional. Um organograma simplificado foi desenvolvido para facilitar a compreensão acerca da estrutura de gestão das Cooperativas Singulares.

Figura 3 – Organograma simplificado das Cooperativas Singulares



Fonte: Site da empresa (2019).

Duas Cooperativas Singulares e um Centro Administrativo foram utilizados como recorte de pesquisa pelos seguintes motivos: surgiram como resultado da técnica de pesquisa utilizada (*snowball*); permitiram uma análise potencial mais aprofundada e específica, pois pode-se realizar a confrontação de entrevistas de respondentes de diferentes posições na organização; facilidade de acesso ao pesquisador; e, pelo alto trânsito e uso de dados ser uma característica inata nestas unidades de negócios.

3.2 INSTRUMENTO E PROTOCOLOS DE PESQUISA

Para Yin (2015), a utilização de um protocolo é essencial para garantir a confiabilidade da pesquisa, orientando o pesquisador na coleta de dados. Martins (2008) define o protocolo de pesquisa como um conjunto de códigos, menções e procedimentos suficientes para se replicar o estudo, ou aplicá-lo em caso semelhante ao estudo original, oferecendo condições para testar a confiabilidade do estudo.

Para este estudo utilizou-se a entrevista semiestruturada como instrumento

primário de pesquisa (MERTENS, 1998). Entrevistas qualitativas semiestruturadas são guiadas por um roteiro de questões de ordenação flexível, permitindo a inserção de novos questionamentos durante a aplicação das entrevistas (FUJISAWA, 2000). Alencar (2001) explica que a entrevista semiestruturada é sustentada por um questionário composto de perguntas abertas, permitindo que o respondente forneça argumentos que propiciem o desdobramento de questões, de modo a ampliar a compreensão do fenômeno estudado. Desse modo, o roteiro de entrevista orienta a coleta de dados, porém ainda permite que o pesquisador considere novas perspectivas, capturando informações emergentes (YIN, 2015). A opção pela técnica de entrevista semiestruturada ainda proporciona ao entrevistador melhor entendimento da perspectiva dos entrevistados do em que entrevistas livres, ou seja, sem estrutura, onde os participantes da pesquisa falam livremente, e por vezes “resultam num acúmulo de informações difíceis de analisar que, muitas vezes, não oferecem visão clara da perspectiva do entrevistado” (ROESCH, 1999).

Dentro dessa perspectiva, a criação dos roteiros para esta pesquisa contemplou a construção de categorias *a priori*, permitindo ainda a inclusão de categorizações *a posteriori*. Para Moraes (1999), a categorização é um ato de agrupar dados considerando semelhanças ou analogia. O autor explica que categorias definidas *a priori* já devem atender aos critérios de classificação de antemão, isto é, antes de proceder à classificação propriamente dita do conteúdo.

Para esta pesquisa foram definidas categorias de análise *a priori* de acordo com a literatura: *Big Data* e *Data-Driven Decision Making*. Sobre *Big Data*, as categorias foram definidas por suas dimensões teóricas definidas como os 5 V's do *Big Data* (RUSSOM, 2011; BEULKE, 2011; WAMBA *et al.*, 2015): Volume, Velocidade, Veracidade, Variedade e Valor. Sobre o processo decisório baseado em dados, as categorias *a priori* foram definidas a partir do *framework* para análise de dados de Gandomi e Haider (2015): Aquisição e Registro, Extração e Limpeza, Agregação e Representação, Modelagem e Análise, e Interpretação. A partir destas categorias desenvolveu-se um roteiro com 12 questões. As questões visaram o cruzamento entre as dimensões do fenômeno *Big Data* com os desafios à tomada de decisão baseada em dados sugerida pela literatura.

Este roteiro foi revisado e validado por dois professores doutores. A versão final do roteiro, aplicada nas entrevistas pode ser encontrada no Apêndice A.

Segundo Carvalho e Souza (2013) as categorias adquirem o papel de

recipientes e os seus elementos são considerados representações de uma realidade existente *a priori*. Os autores completam seu raciocínio explicando que a categorização deve ser apreendida como uma relação cujos elementos são acomodados em seus devidos espaços em função de suas características formais. Em seguida, Carvalho e Souza (2013) esclarecem que o estabelecimento de sentido às informações não se dá *a priori*, e sim na interação; daí a necessidade de categorizações *a posteriori*.

Na pesquisa aqui descrita foram criadas categorias *a posteriori*, que emergiram da percepção do autor durante as entrevistas e no processo de análise. Assim, como as categorias advém dos dados e são sustentadas por eles, são consideradas fundamentadas (CARVALHO; SOUZA, 2013). As entrevistas foram realizadas seguindo os protocolos e instrumentos de pesquisa explicitados na sessão 3.3, 3.4 e no apêndice deste documento.

O procedimento para seleção dos entrevistados se deu através da técnica denominada “bola de neve”, ou *snowball*. Nesta técnica de amostragem não é possível determinar a probabilidade de seleção de cada participante na pesquisa, mas torna-se útil para estudar determinados grupos difíceis de serem acessados (HANDCOCK, GILE; 2011). Ela foi escolhida para este estudo exatamente por este fato como característica da amostra: profissionais de alto nível hierárquico, implicando em dificuldade de acesso. Foram entrevistados profissionais detentores de cargos de alto nível hierárquico, com prerrogativas de tomada decisão: em nível de presidência, diretoria e gerência.

De acordo com Vinuto (2016), há momentos em que a amostragem em bola de neve pode ser a melhor – e, em muitos casos, a única - forma disponível para se estudar determinado grupo. A grande vantagem dessa forma de amostragem é o fato dos entrevistados serem recrutados a partir da relação pessoal das pessoas dispostas a indicar contatos, o que pode emprestar confiabilidade ao entrevistador.

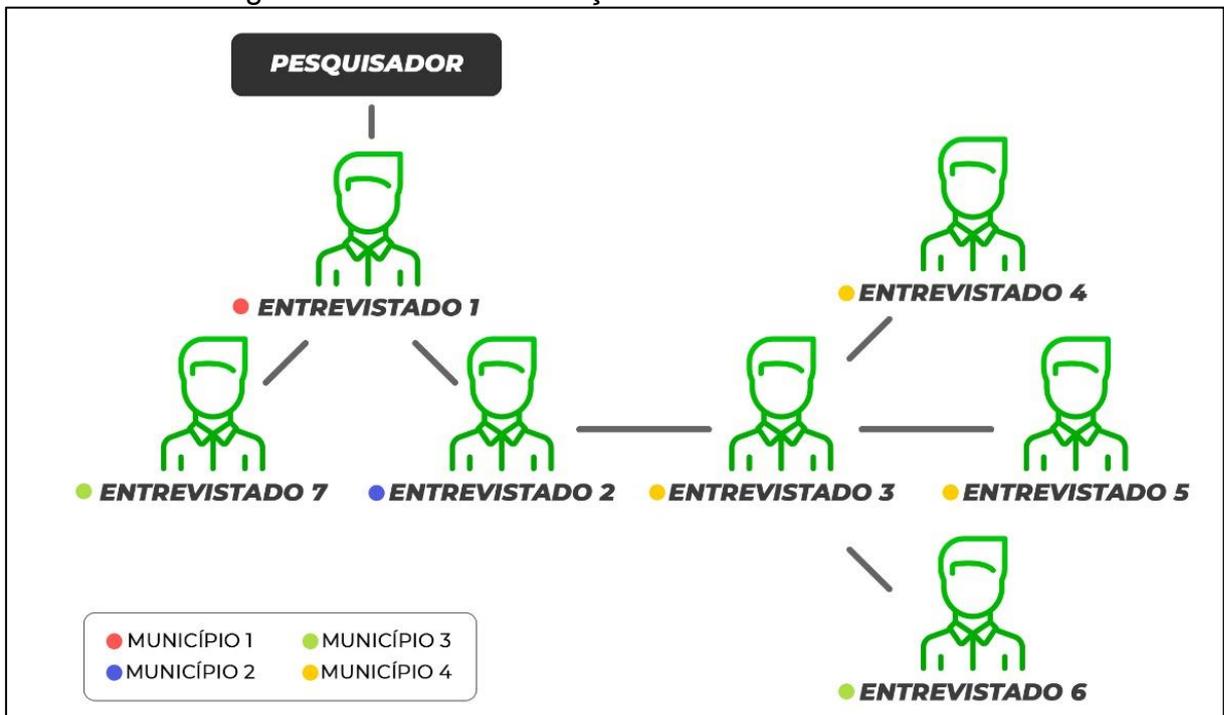
Como na pesquisa por bola de neve o tamanho da amostra é definido a posteriori, a saturação se dá nesse contexto. Para Holmberg e Thorogood (2004) a saturação teórica se dá quando a coleta deixa de revelar dados inéditos. Entendimento consonante com o de Mason (2010), ao afirmar que o corpo de uma amostra de pesquisa precisa ser grande o suficiente para capturar uma vasta gama de experiências, mas não tão grande a ponto de ser repetitivo.

Não existem instrumentos matemáticos ou lógicos de delimitação prévia do

ponto de saturação e, por consequência, do número de observações requerido (MORSE; HOLMBERG, 2008). Por outro lado, Creswell (2014) indica que a saturação teórica começa a ocorrer em pesquisas a partir do sexto participante; fato também descrito por Morse (1995).

No estudo realizado, o ponto percebido de saturação foi alcançado na sétima entrevista, conforme procedimentos sugeridos por Silva e Fossá (2015), descritos na seção 3.4 deste documento. A saturação foi alcançada confrontando-se os dados advindos das entrevistas através de sua transcrição.

Figura 4 – Fluxo de indicações e entrevistas realizadas



Fonte: Elaborado pelo autor (2019).

Foram entrevistados 7 profissionais de duas Cooperativas Singulares, uma agência e um Centro Administrativo. A entrevistas se realizaram em 4 municípios diferentes, no Rio Grande do Sul.

Figura 5 – Entrevistados



Fonte: Elaborado pelo autor (2019).

O tempo de duração de cada entrevista variou de 52 minutos a 1 hora e 16 minutos. O áudio das entrevistas foi gravado e transcrito posteriormente.

Para triangulação dos dados, somaram-se as entrevistas a observação direta do pesquisador, aferida nas visitas aos espaços da empresa para a realização das entrevistas, nos 5 prédios visitados nos 4 municípios onde esta pesquisa foi realizada. Na observação foi possível identificar nuances dos processos e dinâmicas de trabalho da empresa pesquisada, aqui descritas no Capítulo.

Ainda como esforço de triangulação foi realizada a análise documental de documentos como relatórios e mapas estratégicos da empresa, além de notícias e informações divulgadas na imprensa. Estes documentos foram selecionados por sua relação com o tema pesquisado, além de serem aqueles disponibilizados pelos entrevistados ou pela empresa de modo público, sem confidencialidade.

O estudo de caso realizado apresenta objetivos exploratórios e descritivos, justificados uma vez que a pesquisa proposta almeja analisar e descrever um conjunto de práticas. Para Dencker e Viá (2002) as pesquisas exploratórias objetivam formular e esclarecer questões, expandindo o conhecimento do pesquisador sobre o fenômeno ou ambiente investigado.

3.3 PROCEDIMENTOS DE COLETA DE DADOS

Segundo Mayring (2002), todos os procedimentos de uma pesquisa devem ser explicitados e documentados, além de seguir regras e critérios fundamentados de modo a garantir validade e confiabilidade aos resultados. O presente documento visa atender as proposições do autor.

O primeiro procedimento realizado foi o contato, via visita presencial, com um dos pesquisados para dar início as indicações dentro da metodologia bola de neve. A medida que as indicações ocorriam, os indicados entravam em contato com o pesquisador, via contato telefônico ou por e-mail, colocando-se à disposição para este estudo.

Assim as entrevistas eram agendadas diretamente com os entrevistados. Cada entrevista foi individual, em sessão única, com áudio gravado, seguindo a afirmação de Creswell (2014) que considera que a forma primária de coleta de dados é em geral a entrevista e, na pesquisa qualitativa, é indicado o uso de gravador em sua realização. As entrevistas foram realizadas nos meses de janeiro e fevereiro de 2019.

Os entrevistados foram informados sobre os objetivos da pesquisa e suas condições de sigilo. Neste momento foram dirimidas eventuais dúvidas acerca da entrevista que seria aplicada. Ao longo e ao final das entrevistas foram solicitados documentos citados na entrevista, que pudessem ser cedidos para as análises deste estudo. A observação direta se deu nos períodos em que o pesquisador acessava as empresas, antes durante e depois das entrevistas. Em alguns casos os entrevistados conduziram o pesquisador em uma visita guiada a departamentos da empresa.

3.4 PROCESSO DE ANÁLISE

A análise de conteúdo foi realizada nesta pesquisa por meio da codificação e categorização, conforme propõe Flick (2009). Gibbs (2009) define que a codificação é um procedimento de indexação das passagens de texto ou de outros dados relacionados a uma ideia teórica ou a um aspecto descritivo. O autor ainda difere a codificação de categorização, ao afirmar que a primeira possui um caráter mais descritivo enquanto a segunda está relacionada a um nível mais analítico e teórico, exercendo influência na organização de uma hierarquia das codificações. Para o autor, a teoria fundamentada é uma abordagem para codificação, cujo foco central está em gerar de forma indutiva ideias teóricas novas ou hipóteses a partir dos dados, em vez de testar teorias especificadas de antemão.

Bardin (2011) define a análise de conteúdo como um conjunto de técnicas de análise realizado através de procedimentos sistemáticos de descrição do conteúdo de mensagens que permitam a inferência de conhecimentos relativos às condições de

produção e recepção destas mensagens. Câmara (2013) explica que neste tipo de análise, o pesquisador busca compreender as características e estruturas que estão por trás dos fragmentos de mensagens estudados. O autor ainda afirma que o esforço do pesquisador é duplo: entender o sentido das mensagens (como se fosse o receptor normal) e, principalmente, desviar o olhar, buscando outra significação, passível de se enxergar por meio ou ao lado da primeira. Quanto a interpretação, a análise de conteúdo transitam entre dois polos: o rigor da objetividade e a fecundidade da subjetividade. É uma técnica refinada, que exige do pesquisador, disciplina, dedicação, paciência e tempo. Faz-se necessário também, certo grau de intuição, imaginação e criatividade, sobretudo na definição das categorias de análise (SILVA; FOSSÁ, 2015).

O primeiro procedimento realizado para análise dos dados coletados foi a transcrição dos áudios. A transcrição foi realizada de forma manual. Uma vez transcritas, as entrevistas geraram arquivos que foram submetidos, junto com outros documentos coletados, para análise de conteúdo no software NVIVO12®. O uso de software para análises qualitativas é prática sugerida por Gil (2010). Através do NVIVO12® geraram-se figuras descritas e analisadas na seção a seguir.

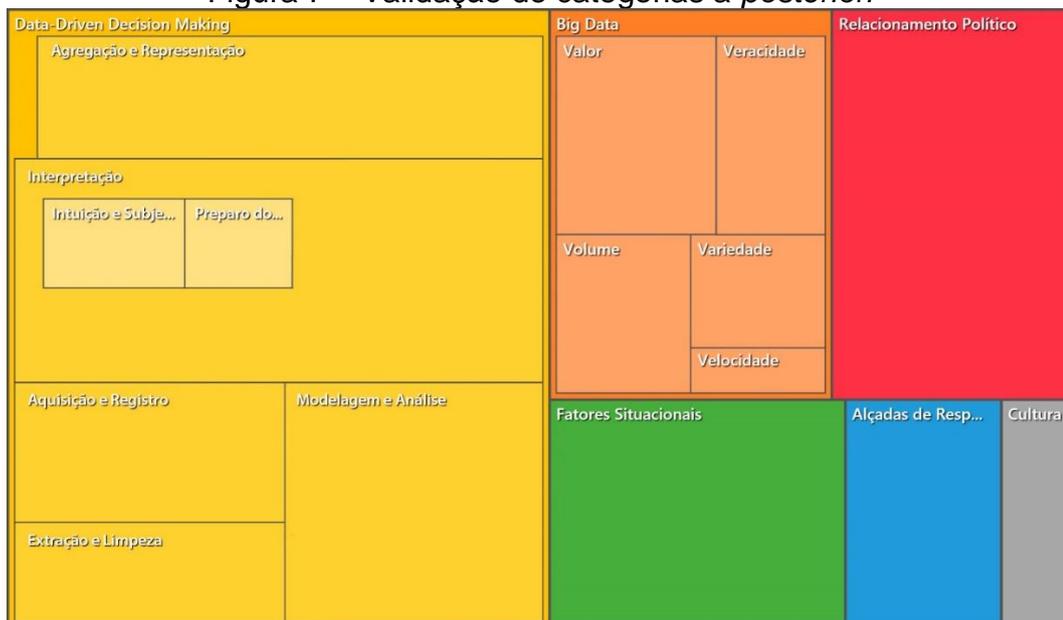
Desse modo, utilizaram-se os protocolos sugeridos por Silva e Fossá (2015) para a realização da triangulação dos dados e identificação da saturação. O primeiro passo realizado foi a leitura geral do material coletado (entrevistas transcritas, documentos e anotações). Após, realizou-se a codificação das categorias *a priori* citadas na seção 3.2 desta pesquisa, utilizando o quadro referencial teórico, com o auxílio do software NVIVO12®. Realizou-se então uma análise de frequência de palavras, com a exportação de um gráfico de nuvem de palavras a ser utilizado nas análises. Consequente, foram realizadas as codificações das quais emergiram as categorias *a posteriori*. Após foi realizada a exportação de documentos com as classificações identificadas pelo software. Concluídos estes procedimentos, seguiu-se a análise explicitada na seção a seguir.

Observa-se que a nuvem de palavras dá indícios da formação de categorias a *posteriori*. As palavras “cooperativa”, “sistema” e “conselho”, por exemplo, têm forte relação com características únicas do negócio, não encontradas em outras instituições presentes no mercado financeiro; componentes, portanto de sua cultura organizacional. As palavras “quando” e “mercado” indicam a presença de fatores situacionais relacionados aos dados e a tomada de decisão, tema desta pesquisa e das entrevistas. “Gerente”, “agência” e “limite” são palavras que denotam a formação da estrutura organizacional, fortemente baseada em alçadas de responsabilidade.

Não menos relevantes são as palavras “comunidade”, “pessoas”, “relacionamento” e “política”; indicando a presença de algum tipo de relação política presente no processo de tomada de decisão. Destacam-se ainda “intuição” e “subjetividade”, uma vez que fazem um contraponto ao conceito de *Data-Driven*; além de “equipe” e “autonomia”, indicando a importância da figura do tomador de decisão.

A partir da observação do conteúdo das entrevistas, e considerando os dados apontados pela nuvem de palavras, emergiram as categorias de análise a *posteriori*: Relacionamento Político, Fatores Situacionais, Alçadas de Responsabilidade e Cultura Organizacional; além das subcategorias Intuição e Subjetividade e Preparo do Tomador de Decisão. A relevância dessas categorias foi identificada no *software* NVIVO12®, através de seu gráfico de hierarquia baseado na frequência total de palavras presentes nas transcrições das entrevistas.

Figura 7 – Validação de categorias a *posteriori*



Fonte: Elemento gráfico gerado pelo *software* NVIVO12®.

Observando a figura acima, verifica-se a relevância das categorias e subcategorias emergidas *a posteriori*, quando comparada sua representatividade com as categoriais *a priori*; como *Big Data*, por exemplo. A presença dessas novas categorias, evidenciadas nas falas dos entrevistados, será descrita e analisada nas subseções a seguir.

No processo de análise, identificou-se que as categorias concernentes ao tema *Big Data* possuíam sinergia com as relacionadas ao tema *data-driven decision making*. Essa sinergia adveio da percepção acerca do teor das entrevistas: verificou-se nas falas dos respondentes a convergência dos dois temas centrais deste estudo, mesmo quando as perguntas consideravam apenas um deles.

Desse modo, a análise descrita nas subseções a seguir foi assim organizada: Volume, Velocidade, Aquisição e Registro (combinando as categorias *a priori* acerca do *Big Data* com as categorias *a priori* relativas ao *data-driven decision making*); Veracidade, Extração e Limpeza (combinando as categorias *a priori*); Alçadas (identificada *a posteriori*); Variedade, Integração, Agregação e Representação (combinando as categorias *a priori*); Valor, Modelagem e Análise (combinando as categorias *a priori*); Interpretação para tomada de decisão (combinando as categorias *a priori*); Fatores Situacionais (identificada *a posteriori*); Cultura Organizacional (identificada *a posteriori*); Relacionamento Político (identificada *a posteriori*); e *Framework* proposto.

Esta convergência relaciona-se ainda a um dos objetivos específicos dessa pesquisa, ao permitir a caracterização do processo de análise de informação de fontes de *Big Data* considerando suas dimensões. A forma como a associação entre os conceitos foi identificada, revela-se como uma descoberta deste estudo; uma vez que emergiu *a posteriori*.

Uma das causas possíveis para essa sinergia entre os conceitos ocorrer na fala dos entrevistados pode ser relativa à natureza dos processos de tomada de decisão no nível hierárquico pesquisado. Segundo a fala dos respondentes, as dimensões do fenômeno *Big Data* apenas geram resultados práticos quando aplicadas aos processos de extração de insights dos dados para tomada de decisão. Nesse contexto, um grande volume de dados não tem relevância se não puder ser tecnicamente absorvido e processado pela organização. No mesmo sentido, dados de fontes variadas se tornam um problema sem um processo de aglutinação útil ao

tomador de decisão. Pode-se observar ainda que o inverso também é verdadeiro na visão dos entrevistados: sem um grande volume de dados as análises ficam prejudicadas pela falta de triangulação; ou seja, de dados que permitam a confrontação e checagem de informações. Ao mesmo tempo que, para que o tomador de decisão possa se aproveitar de uma aglutinação de fontes de dados, essas fontes precisam ser processadas através da estrutura de tecnologia da empresa. Essa lógica se alinha a afirmação teórica de que apenas dados geralmente são insuficientes para a tomada de decisão (AGRAWAL *et al.*, 2012).

Com base nesse raciocínio as categorias e subcategorias foram aproximadas por suas similaridades e reorganizadas no texto a seguir, resultando em uma análise unificada dos temas *Big Data* e *Data-Driven Decision Making*.

4.1 VOLUME, VELOCIDADE, AQUISIÇÃO E REGISTRO

O uso extensivo de dados administrado pela organização foi citado por todos os respondentes. Isso se dá pela própria natureza da atividade da organização:

[...] uma infinidade de dados, que vai desde renda, faturamento, idade, sexo, quanto ganha, quanto gasta, os volumes de produtos e serviços, os volumes de outras soluções, cartão, seguro, previdência. Enfim, uma base de dados significativa, só nossa Cooperativa (Singular) (ENTREVISTADO 1).

Cabe ainda ressaltar que a aquisição de um alto volume de dados é um processo que se dá através tempo. E no caso de empresas de serviços financeiros, isso pode se traduzir em subsídios para decisões mais assertivas:

No nosso BI, com a base de dados dos últimos 5 anos que a gente tem, eu pedi para ele verificar para mim quais são os associados que, nesses últimos 5 anos mais deram retorno para a Cooperativa (Singular) [...]. Porque a gente fica míope quando olha um ano só, porque o associado, ele tem uma vida aqui dentro. Porque se olha o associado em um ano só: "Ele é excelente! ". Mas o melhor pode ser o associado do lado, porque pode ser que faça 10 anos que ele é associado aqui e ele já deu muito retorno para a cooperativa (ENTREVISTADO 3).

Por outro lado, adquirir e registrar um alto volume de dados não necessariamente significa que eles serão utilizados plenamente. Embora faça sentido no que tange a filtragem: quanto maior o volume, maiores são as possibilidades de filtragem e combinações para análise posterior. E o desperdício no consumo de

dados, antecipado por Bell (2014) parece ser uma realidade relevante na empresa pesquisada:

O que a gente entrega para as Cooperativas (Singulares), eles consomem 50%. [...] até fechou o ano (2018) com 51%, e agora está entre 47%, 48%. [...] eu tenho um esforço gigante de processamento estatístico, [...] a gente investiu bastante em time para trazer inteligência para dentro da máquina comercial, que muitas vezes isso não é muito visto porque fica atrás do balcão (ENTREVISTADO 7).

Mas as afirmações dos entrevistados permitem a conclusão de que a organização objetiva a aquisição constante de ainda mais dados. Isso porque mais dados representam maior possibilidade de triangulação na busca por veracidade. A cooperativa inclusive vem contratado empresas para trazer ainda mais dados adicionais sobre mercado, o que é chamado de “enriquecimento da base de dados”, além de investir em novas tecnologias (BAGUETE, 2018).

Outro ponto que a gente está trabalhando ainda é enriquecer a informação que eu tenho de cada pessoa com dados externos. Tipo dados de redes sociais, dados de relacionamento com outras instituições do tipo: “Tu tem conta de telefone aonde?”. Eu posso saber seu telefone dentro do cadastro, mas será que é esse mesmo? Ou eu posso falar com a Vivo e adquirir a base dos meus CPFs de onde as pessoas têm operação, (ou) onde é que recebem a conta, para mim ter um endereço mais qualificado. Qual o volume de pagamento da conta dele? “Ah, eu tenho aqui que a renda dele é 2 mil Reais, mas ele gasta 700 Reais de telefone”. Não está compatível (ENTREVISTADO 7).

Em relação a velocidade de aquisição de dados existem desafios inerentes ao setor, que limitam a entrega de dados pelas fontes. Um dos exemplos são informações, normas e regulamentos advindos de órgãos governamentais. Uma das fontes citadas é o BACEN, que disponibiliza baterias de dados técnicos a cada 3 meses, o que por si só já gera alguma defasagem se considerar a necessidade de tomada de decisão em tempo real; mesmo que essa condição seja igual para todas as empresas deste mercado. Esses *gaps* temporais tendem a prejudicar a extração de *insights* através da triangulação em tempo real (NUNAN; DOMENICO, 2013);

A maioria das falas dos entrevistados relacionadas a velocidade, tanto na aquisição e tratamento de dados quanto relacionadas aos processos da organização, denotam um sentimento negativo; em consonância com Chan e Payne (2017). Isso é: a velocidade é um desafio, uma vez que os dados não são entregues na velocidade desejada.

[...] (nossos maiores desafios são) velocidade de processamento e tomada de decisão. Nós temos muita decisão para tomar, nós temos muita coisa para processar (ENTREVISTADO 3).

[...] (os processos) ocorrem em uma velocidade menor (do que se gostaria), muito ainda por questões políticas e outra por questões culturais (ENTREVISTADO 5).

A necessidade percebida por maior velocidade de processamento e por processos mais ágeis é também relacionada com as expectativas dos associados e do mercado. E as expectativas crescentes por maior instantaneidade de processos mostra-se intrinsecamente ligada ao *mainstream*⁴ atual do ambiente de negócios, onde há um contexto social no qual a velocidade ganha relevância comercial:

[...] porque para todas aquelas operações que a análise poderia ser de forma pré-aprovada, qualquer tempo de análise é muito demorado. [...] a gente não vai conseguir chegar em um tempo preciso entre o cara que espera e o cara que entrega, sempre vai ter um descasamento de expectativas. O cara que espera quer que seja muito rápido e o cara que entrega precisa de um prazo para fazer bem feito. [...] o mundo ideal é que quando eu chame um Uber no aplicativo, ele chegue em um segundo. Apertar e ele chegar: essa é a minha expectativa. Mas para o Uber administrar tudo isso, o mundo ideal para ele seria que eu tolerasse 15 minutos numa boa (ENTREVISTADO 6).

Entende-se, portanto que a aquisição de dados caminha na direção de um volume cada vez maior; enquanto que a velocidade se move na direção da instantaneidade. Sendo a velocidade o desafio mais premente, uma vez que a própria complexidade das organizações e suas operações atuam como mecanismos limitantes de agilidade.

4.2 VERACIDADE, EXTRAÇÃO E LIMPEZA

Apesar do grande volume, pode-se aferir nas entrevistas que existem lacunas nos dados. Elas são relacionadas tanto com o processo de aquisição dos dados quanto com seu tratamento, no sentido de lhes conferir veracidade. Nesse contexto, os entrevistados apontaram a importância de estruturas e departamentos locais para o processamento de dados. Uma vez que essas estruturas estejam mais próximas dos tomadores de decisão, acredita-se que a tarefa de suplantando essas lacunas com

⁴ Aquilo que se considera habitual ou normal porque é feito ou aceito pela maioria. Disponível em: <https://www.dicio.com.br/>. Acesso em: 07 abr. 2019.

cruzamentos de informações é facilitada.

Existe um conceito na área de negócios, a sigla é CLV (Customer Lifetime Value) o Sistema (Cooperativo) não tem essa informação. No nosso BI5, com a base de dados dos últimos 5 anos que a gente tem, eu pedi para ele verificar para mim quais são os associados que nesses últimos 5 anos mais deram retorno para a Cooperativa (Singular) (ENTREVISTADO 3).

A inserção manual de dados em softwares internos é um desafio percebido para a veracidade, uma vez que é passível de uma miríade de erros; e, como alerta Promentilla *et al.* (2018), informações imprecisas. Uma das características do processo manual é a impossibilidade de padronização total para inserção de informações; o que gera alguma tolerância a erros. Os ruídos são atenuados em fontes de dados externas, como no caso de empresas terceiras; uma vez que se espera que os dados já tenham passado por um processo de limpeza na fonte.

Na verdade, hoje todas as informações que a gente recebe tem erros: erros de cadastro, erros de informação. Porque tem dados que a gente pega direto com o associado, tem dados que a gente pega com o Serasa [...] (ENTREVISTADO 4).

Então tudo aquilo que tem inserção manual, eu diria que são umas das principais fontes de ruído, de problemas de dados (ENTREVISTADO 6).

O (dado) “endereço” ainda tem muitos erros, porque pode ou não ter comprovante, é dispensável, [...] (ENTREVISTADO 1).

No caso do setor financeiro, foco deste estudo, há ainda uma particularidade: a alta informalidade dos clientes relacionada a declaração de suas fontes de receita. Ela se apresenta nos casos em que o associado precisa comprovar sua renda para efeitos de tomada de crédito, por exemplo. Nesse sentido, a informalidade é responsável por situações de difícil comprovação documental, o que prejudica a veracidade dos dados coletados para tomada de decisão de concessão de crédito.

Para diminuir a incidência erros nas informações utilizadas na organização, os responsáveis em nível operacional buscam informações e documentos adicionais, para executar uma triangulação; objetivando a maior veracidade. Essa triangulação é tratada como “aderência”. E, segundo os entrevistados, essa não é prática comum nas empresas do setor financeiro; que muitas vezes optam por não realizar esse procedimento:

⁵ *Business Intelligence*: departamento de tratamento de dados.

Por exemplo, no Itaú eu nunca pude usar a informalidade de alguém, e isso a gente sabe que existe. Então como é que você vai provar se uma pessoa diz: “Ah, eu ganho por fora mais mil Reais”? Isso a gente chama de aderência. [...] com extrato de conta, mostrando que entram esses mil Reais, ou pelo patrimônio da pessoa. Principalmente nas empresas que tem um ganho enorme: “Do meu faturamento, 30% é informal”. Aí você vai perguntar onde está esse dinheiro e percebe que a pessoa tem bastante patrimônio. Então são aderências que a gente faz, são conclusões que a gente chega (ENTREVISTADO 2).

Também existe a preocupação com a manipulação equivocada dos dados, que pode reduzir sua veracidade.

Porque eu recebo alguns questionamentos aqui dos números que a gente gera. Que se chama veracidade do dado; que é quando alguém trabalhou e massageou a informação, torturou aquela informação para mostrar um dado. [...] uma premissa de não ferir o indicador corporativo às vezes não é bem respeitada (ENTREVISTADO 7).

Nesse contexto, um procedimento adotado pela organização para conferir maior veracidade aos dados é o de eliminação do Gauss. Ou seja: realizam-se processos para limpeza das informações, eliminando dados que estejam nas extremidades da análise, mantendo-a concentrada nos dados e valor médio. Dessa forma busca-se aumentar a veracidade através do exame dos dados centrais, eliminando dados discrepantes que, por sua natureza, podem induzir a análise do todo ao erro:

[...] em uma atualização cadastral, nem todos eles estão totalmente atualizados enquanto a renda, então a gente busca se centrar no intervalo mais de confiança, que a gente acredita dar limitadores; para que a gente possa não ficar refém somente aquele dado [...] quando a gente olha essa mostra, a gente busca desqualificar dados que estão nas extremidades, tanto acima, como abaixo, e trabalhar com uma média mais real interna, e fazer outros contrapontos de desclassificação de dados, outros e critérios que também consiga eliminar o cara [...] (ENTREVISTADO 1).

Outra prática adotada para conferir maior veracidade aos dados vem da padronização de processos internos. O departamento responsável no Centro Administrativo pelo apoio de dados direto as agências é o único que possui ISO9001 na organização. Essa certificação de qualidade em processos confere diversas instruções procedimentais e padronizações, que mitigam boa parte dos erros potenciais no tratamento de dados.

Pesquisas qualitativas também são utilizadas como forma de triangulação posterior dos dados. As fontes de dados qualitativos exercem a função tanto de trazer novos insights ao processo de tomada de decisão, quanto de validação de dados internos:

Hoje de manhã mesmo eu recebi um e-mail de um casal, no caso um associado e a esposa, dizendo assim “Olha gente eu estou encantado e admirado em como foi processo do meu crédito imobiliário. Então eu desejo super sucesso para vocês e o meu agradecimento de coração, porque foi uma experiência muito legal.” Então esse termômetro que a gente acaba levando de volta para a área de negócios: “Olha isso está funcionando” “Isso não está funcionando, “Esse processo aqui está muito burocrático” (ENTREVISTADO 6).

[...] e a gente foi buscar dados qualitativos, de porquê que a gente abre tantas contas. A gente achava que era fruto das folhas de pagamento que a gente estava buscando, que era o esforço de buscar. A gente queria ver se era isso ou se tinha mais a ver com a proposta de valor; enfim saber o que era. E quando nós fomos ouvir o associado, a gente fez uma entrevista com os associados, qualitativas, com perguntas programadas [...] (ENTREVISTADO 1).

Nota-se a importância dos dados qualitativos como uma relevante fonte de insights estratégicos e operacionais. O uso desta fonte de dados também se relaciona com a cultura organizacional; tópico analisado em seção posterior.

Em relação aos processos relacionados com a veracidade dos dados, é possível notar a relevância dos processos de triangulação e eliminação de dados extremos como iniciativas fundamentais de extração e limpeza. Apenas a partir de aferida a sua veracidade, os dados podem avançar para processos de análise.

Infere-se que tanto os processos de Aquisição e Registro, quanto de Extração Limpeza e comentários fazem parte do macroprocesso de gerenciamento de dados, conforme expresso por Gandomi e Haider (2015). Na empresa pesquisada, porém, diferentes dados são analisados (e diferentes análises são feitas) dependendo da posição hierárquica do tomador de decisão. Nesse sentido a organização é estrutura em divisões hierárquicas chamadas de alçadas de responsabilidade.

4.3 ALÇADAS

A organização por alçadas é umas das principais características da instituição pesquisada em seu processo de tomada de decisão. As alçadas estabelecem as funções, responsabilidades e limites de atuação de cada pessoa dentro da empresa.

Elas atuam também como um marco divisor entre o nível estratégico e o operacional.

Existe uma frase que representa bem o papel do Conselho (Administrativo): o Conselho é aquele órgão que chega dentro da Cooperativa, e ele coloca o nariz, mas não coloca a mão. A partir do momento que o Conselho colocar a mão, ele está tirando a autoridade e a responsabilidade da diretoria (ENTREVISTADO 5).

Isso pode ser identificado também na observação *in loco* nos prédios onde as entrevistas foram realizadas. Em uma das Cooperativas Singulares, por exemplo, a sala da presidência fica em um andar distinto daquele reservado às salas dos diretores e gerentes executivos. Essa distância física pode ter algum impacto simbólico, e prático, no respeito às alçadas.

O mesmo foi observado no Centro Administrativo. Mas aqui, além de diferentes andares para diferentes departamentos, o acesso é controlado por cartões eletrônicos de identificação. O que significa que funcionários têm o acesso liberado apenas ao seu setor de trabalho.

As alçadas são um instrumento que viabiliza a presença da intuição e subjetividade no processo de tomada de decisão, sem prejudicar a saúde da organização. Assim são estabelecidos limites seguros para a atuação de cada profissional, protegendo o negócio como um todo.

Chama a atenção o processo de tomada de decisão aplicado à concessão de crédito. As Cooperativas Singulares conferem, de acordo com a alçada, níveis de liberdade operacional para seus gerentes na concessão de crédito. Isso significa que mesmo com a ausência de dados internos (ou de dados adicionais para conferir maior aderência) os gerentes têm a autorização para a concessão de valores de crédito total dentro da sua alçada, além do limite pré-aprovado (estabelecido pelos algoritmos dos softwares operacionais). Ou seja: os gerentes têm garantido o direito de não seguir os dados, dentro dos limites de sua alçada de responsabilidade.

Peraí, o papel está dizendo que não dá para dar (o crédito), mas é o “Jean”. Sabe o “Jean”? É aquele cara ali da esquina; que todo mundo vai todo dia ali no restaurante dele. Todo mundo sabe quanto ele é sério. Ele já se atrapalhou? Já se atrapalhou, mas ele é um cara do bem. E o nosso papel como cooperativa é desenvolver as comunidades, muito disso por conta de fornecer crédito. Então a gente consegue fazer a avaliação subjetiva de um crédito, porque a gente conhece o associado. Então aí vem um pouco de combinar as duas coisas em uma tomada de decisão: dados são fundamentais e no nosso negócio cada vez mais complexos, eles são cada vez mais importantes, a prova disso é que nós temos uma área de BI, porque

a gente não se achava mais com tanta informação para tomar uma decisão; mas o subjetivo na cooperativa ainda faz muito sentido (ENTREVISTADO 3).

Existe um espaço para a subjetividade sim, espaço para *feeling*, para cheiro. A gente diz que crédito tem que cheirar. Não vai só emprestar para uma empresa. Tem que ir lá cheirar a empresa. Não dá para olhar só papel, tem que ir lá sentir o negócio, feeling. Por isso o nosso valor é relacionamento, porque feeling vem de relacionamento, quanto mais proximidade eu tenho com alguém mais eu consigo sentir as coisas (ENTREVISTADO 3).

Os entrevistados percebem que abordagem se configura como um diferencial competitivo da organização frente aos concorrentes do mercado financeiro. E como empresa estudada opera em sistema regionalizado, os limites de alçada na concessão de crédito variam, seguindo a realidade de mercado percebida por cada Cooperativa Singular. Nas duas Cooperativas Singulares analisadas, o limite de alçada de um gerente de agência para a concessão de crédito variou em 100%.

O Itaú hoje, por exemplo, eu sei porque a gente conversa com muitos ex-colegas e hoje estão trabalhando conosco. Se o computador, a máquina gerou o limite lá, se não deu, não adianta nem perder tempo (ENTREVISTADO 4).

Depreende-se que a organização por alçadas de responsabilidade também determina a quais tipos de dados um tomador de decisão terá acesso. Nesse sentido as alçadas agem como um filtro a tomada de decisão. Uma vez determinada a alçada, o tomador de decisão terá a sua disposição a variedade de dados disponíveis exclusivamente para seu cargo na instituição; já como uma primeira etapa do macroprocesso de análise dos dados. Esse fato se configura como relevante procedimento no processo de tomada de decisão baseada em dados, pois orienta e distribui os fluxos de dados dentro da organização.

4.4 VARIEDADE, INTEGRAÇÃO, AGREGAÇÃO E REPRESENTAÇÃO

O alto volume de dados na organização pesquisada é proveniente de diferentes fontes. Cada Cooperativa Singular lida com dados nacionais do Sistema Cooperativo, dados de órgãos reguladores, e dados internos; advindos de suas agências e associados. Embora o Centro Administrativo seja responsável, *a priori*, pela concessão de dados as Cooperativas Singulares, algumas delas já contam com departamentos próprios de *BI*, para tratamento de dados desde sua aquisição. De acordo com Agrawal *et al.* (2012) a seleção das fontes de dados é vital ao processo

de análise de dados, e neste sentido não foram citados problemas pelos entrevistados.

As duas Cooperativas Singulares pesquisadas contam com algum tipo de estrutura ou departamento interno de tratamento de dados desde sua aquisição. Isso parece favorecer análises locais:

Então a gente consegue aqui, pela nossa área de BI, identificar melhor quais produtos são mais aderentes para cada segmento (de mercado). A gente consegue afinar mais ainda os dados aqui, coisa que o Centro Administrativo, pelo volume de dados gigantescos, eles não conseguem chegar a um nível menor que a gente chega hoje aqui. A gente chama de “minerar”. Então a gente minera mais os dados aqui, e a gente consegue ser mais pontual (ENTREVISTADO 3).

No que tange a variedade, o principal procedimento citado pelos entrevistados é o de integração: concatenar grandes volumes de dados em repositórios únicos, para representação em painéis simplificados e úteis a tomada de decisão. Esse desafio é consonante na literatura com Agrawal *et al.* (2012). Para isso tanto as Cooperativas Singulares quanto o Centro Administrativo vêm mobilizando esforços e criando estruturas e departamentos.

[...] (os dados são reunidos) em um grande arquivo de dados, e a gente alimenta, não sei se é PHP que eles estão fazendo, ou uma outra linguagem que a gente faz. E o que a gente busca fazer: a gente estabelece os critérios, bota os critérios, na regra do sistema, e o sistema pega aquele banco de dados gerando os resultados. [...] são dados que são integrados no sistema (software), que podem vir do Sistema (Cooperativo) por exemplo, que eles nos subsidiam, e dados externos que daí a gente pega: Serasa, SPC, essas fontes, SCR que é Banco Central. Então essas aqui são três fontes, geralmente que a gente usa externas, confiáveis, para fazer a capacidade de pagamento, por exemplo. Então esses três dados são externos e os outros dados eles são todos internos, sejam internos próprios locais aqui da Cooperativa (Singular), ou internos próprios do sistema (Cooperativo) (ENTREVISTADO 1).

Há uma particularidade na organização pesquisada: a idade e a complexidade da organização, aliada ao alto volume de operações. Isso significa que existe um histórico de uso de diferentes softwares, que ao longo do tempo se tornaram grandes demais para uma fácil transposição de dados para tecnologias mais modernas. Além disso, a ampla gama de produtos e serviços oferecidos muitas vezes é operada por softwares distintos e fornecidos por terceiros.

Esse fator é um desafio na integração de dados de variadas fontes, que vem sendo enfrentado pela organização com a criação de departamentos e contratação de

profissionais específicos para esta tarefa.

O que eu tenho problema, talvez, é a falta de informação no local esperado. [...] uma vez eu queria ver se tu tinha seguro. Você tinha que entrar no sistema do seguro. Se você queria ver se tinha consórcio tinha que entrar em 6 coisas para ver. Então hoje eu tenho em uma tela só, que a gente chama de Visão do Associado. Isso tudo é uma ferramenta caseira da nossa Cooperativa (Singular), ela contratou um rapaz que faz essa integração[...] No Itaú, nós tínhamos isso há 10 anos atrás, mas nós estamos falando de tecnologias muito diferentes, de investimento diferentes (ENTREVISTADO 2).

Inclusive diferentes fontes é um problema [...] porque tem muitas fontes que não temos ainda dentro do Sistema. Está em evolução, isso melhorou muito desde que eu estou aqui, mas nós não temos ainda uma única base de dados. Então nós temos sistema de cartão, sistema de consórcio, sistema de conta corrente, sistema de crédito, sistema de investimentos, são bases distintas. Então quando a gente precisa buscar uma informação, hoje o único local que está conseguindo ser um repositório de dados de associados, não estou falando de dados da estratégia, a gente tem um CRM agora, que funciona muito bem (ENTREVISTADO 3).

Os desafios na integração dos dados também têm impacto direto na velocidade da tomada de decisão, retirando a agilidade de alguns processos operacionais. Esse é um dos principais problemas resultantes das dificuldades de integração de dados, segundo os entrevistados:

Hoje a gente é muito consumido pelo dia a dia, porque a gente faz muitas coisas. E muitas coisas são várias operações em vários sistemas diferentes [...] e faz com que a gente dedique menos tempo que eu gostaria para atender as pessoas (ENTREVISTADO 6).

Essa relação entre integração de dados e velocidade na tomada de decisão também tem impacto estratégico quando considera-se o ambiente de mercado atual. Empresas jovens do setor financeiro, com menor carteira de produtos e serviços, são formadas sobre plataformas tecnológicas modernas. E isso pode se traduzir em vantagem competitiva para nichos específicos:

Porque não é tudo (integrado) em um sistema (software) só? Na minha opinião: [...] eu diria que os anos de história eles fazem isso com a gente. Quando você começa uma fintech⁶ nova, você monta tudo do zero, uma tecnologia super nova. [...] quando você vê ali Nubank, essas coisas mais modernas. Fintechs. Mas esses caras operam cartão de crédito, uma conta corrente às vezes, e algum crédito pessoal, então coisas muito simples. Quando você entra em uma instituição financeira mais complexa, ou mais completa digamos assim, você tem ainda muita necessidade, porque não vamos esquecer que estamos no Brasil, então tem toda uma questão de

⁶ Empresas de tecnologia financeira.

regramento e isso tem um peso (ENTREVISTADO 6).

A organização está avançando na integração de dados através da instalação de departamentos de tratamento de dados e análises de *BI*; e da gradual substituição de softwares defasados por gerações mais novas. Na observação em agência, o pesquisador teve acesso as telas utilizadas no sistema que demonstram essa realidade. Embora alguns *softwares* observados usem tecnologia mais recente, uma parte das operações aparentemente é feita em sistemas que utilizam linguagens antigas, como a MS-DOS⁷.

Considerando os desafios concernentes a variedade de dados, observa-se que sua aglutinação em painéis úteis a tomada de decisão é fator determinante a tomada de decisão em um contexto *Data-Driven*. Uma vez que os dados coletados para uma finalidade também podem ser usados para muitas outras finalidades (AALTONEN; TEMPINI, 2014), pode-se inferir que esta etapa seja diretamente responsável pela qualidade da decisão futura; uma vez que estas estruturas permitem a extração de valor dos dados.

4.5 VALOR, MODELAGEM E ANÁLISE

Na organização pesquisada, o processo de conferir valor aos dados, e modelá-los para análise é feito através de regras estabelecidas de acordo com a necessidade dos tomadores de decisão. Sua modelagem obedece a objetivos estratégicos e operacionais, considerando políticas internas e marcos regulatórios legais. A partir daí são criados indicadores para análise e posterior suporte a tomada de decisão:

São duas fontes: sistemas e regulamentários. Então existem os indicadores que são obrigatórios por parte de resoluções do Banco Central, a própria 4557 que é Gestão Integrada de Riscos, apresenta indicadores obrigatórios e dentro desses indicadores existem indicadores mínimos e máximos que o Banco Central impõe para as instituições financeiras, inclusive as Cooperativas. [...] O Sistema (Cooperativo) pegou esses indicadores do Banco Central que são obrigatórios, [...] com outros indicadores de provisão ou de atrasos do pagamento e outros de liquidez, e a gente tem esses 11 indicadores, que todos os meses são analisados. O Sistema (Cooperativo) apertou um pouquinho a torneira dos limites mínimos e máximos [...] aí a gente já é cobrado para fazer alguma correção antes (de alguma inconformidade), isso é monitorado mensalmente. O que o conselho faz é

⁷ Sistema operacional da empresa Microsoft, descontinuado em 2001.

olhar como estão esses indicadores, (e entender) porque eles estão verdes (sistema interno de classificação por cores), ou porque algum deles foi para o amarelo, hoje nós não temos nenhum no vermelho (ENTREVISTADO 5).

Porque a pesquisa de satisfação no ano (anterior) foi 72 (índice em uma escala de 0 a 100); e no ano seguinte foi 88, isso é dado. E as cooperativas que queriam ir embora, não queriam mais administrar o serviço aqui (no Centro Administrativo), (e depois) eu passei a administrar cooperativas querendo entrar (ENTREVISTADO 6).

O procedimento para criação de modelos de análise comporta a tentativa e erro. Isso porque os modelos são criados por profissionais de TI^8 , que muitas vezes não sabem exatamente qual o objetivo prático da organização com o tratamento daqueles dados. Quando um modelo eventualmente não atende as necessidades dos tomadores de decisão, ele passa por processos de revisão:

[...] é muito interessante em tecnologia montando relatórios; que o pessoal da T.I acham muito importante e não que o negócio precisa. E aqui eu tenho as três coisas acontecendo em conjunto, ou seja, eu gero a inteligência, distribuo para a máquina comercial e mensuro o resultado. O que não for tão bom eu calibro de novo e coloco na mesa, e o que for legal eu vou aprimorando os modelos [...] (ENTREVISTADO 7).

[...] a gente foi aprendendo onde a gente acertou e errou também. A gente vem fazendo o pré-aprovado desde 2009, mais ou menos. A gente começou era bem pequenininho e foi evoluindo nossa regra. E aí onde tinha uma perda daquela amostra, nós dávamos uma mergulhada, uma aprofundada: Por que será que perdemos? Porque será que esse cara foi para um pré-aprovado e depois ele não conseguiu solver operação? A gente identificava um critério ou outro, então assim, temos que reponderar (ENTREVISTADO 1).

Cabe ainda pontuar que o valor dos dados também tem relação com o tempo, especialmente em termos qualitativos. Isso significa que quanto mais dados históricos são reunidos, maior é o valor percebido e sua utilidade para a tomada de decisão e processos operacionais é aumentada:

Está tudo ali registrado. Se amanhã vier outra pessoa atender, já está tudo registrado [...] então o gerente tem que estar preparado, deixar limites pré-aprovados (de crédito) para o associado, porque se ele precisar ele tem na mão. O associado quer fazer uma viagem? Tem que saber que ele precisa de um cartão de crédito internacional. O que acontece é que quando o associado vê que você pergunta tudo isso e você vai dizendo para ele o que tu tem e o que não tem, consequentemente quando surge uma oportunidade, ou até uma necessidade, ele vem automaticamente aqui (ENTREVISTADO 2).

⁸ Tecnologia da Informação.

No processo de tomada de decisão, os dados também são utilizados em momentos posteriores a tomada de decisão em si. Seu valor consiste tanto em usos para avaliação da decisão, quanto para proteção do tomador perante outras esferas internas. Em nível operacional, estes dados históricos também podem não integrar os softwares oficiais, mas fazer parte de uma estrutura individual e informal de análise:

Na verdade, ele fica escaneado na pastinha do cliente, é interno, da agência, para observar e para a nível de auditoria dizer “De onde você tirou isso? ”. [...] Quando uma operação dá um atraso, que não paga, eu costumo pegar a operação e ver onde é que eu errei. Eu não decido sozinha, mas hoje quem tem alçada na agência sou eu, então junto com a analista de crédito que tem na agência, que é a gerente administrativa, a gente toma uma decisão (ENTREVISTADA 2).

A necessidade de uso de dados para tornar a decisão tomada aparentemente mais neutra é citada na literatura por Capgemini (2011). Fato relevante é que na fala dos respondentes isso se dá não apenas antes, mas após a tomada da decisão.

Uma vez que os dados passam pelos processos de modelagem e análise, é a vez do tomador de decisão exercer a sua interpretação sobre eles. Sua interpretação é base para a tomada de decisão. Como processo mental individual, ela define a tomada de decisão. Discutiremos o processo de interpretação de dados na seção a seguir.

4.6 INTERPRETAÇÃO PARA TOMADA DE DECISÃO

A interpretação dos dados, por sua vez, é diretamente dependente do preparo do tomador de decisão. Nesse contexto consideram-se ainda fatores que interferem nessa interpretação, como a intuição do tomador de decisão aliada a questões de ordem subjetiva, conforme citado por Gilboa, Rouziou e Sibony (2018).

4.6.1 Preparo dos tomadores de decisão

O corpo de tomadores de decisão precisa estar preparado para sua tarefa, considerando a realidade de mercado, o contexto *Data-Driven* e a autonomia esperada pelo responsável de cada função dentro da organização. A percepção do pesquisador diante dos entrevistados é que a organização conta com equipes preparadas para isso.

Das entrevistas emergiu a identificação de práticas adotadas pela empresa para a preparação dos seus tomadores de decisão. A primeira delas reside na contratação e na retenção de pessoas. Todos os entrevistados possuíam sólida experiência no setor financeiro, atuando inclusive em outras instituições. A contratação de profissionais com ampla experiência setorial mostrou-se prática recorrente na organização pesquisada.

Pelo currículo dos entrevistados também foi possível observar que eles possuem laços de proximidade (amigos, familiares, conhecimento regional) com os municípios onde atuam; traduzidos aqui como conhecimento de campo. Esta prática concorda com Rose (2017), ao entender que a o conhecimento humano adquirido ao longo do tempo é importante para o processo decisório. Nesse sentido, os tomadores de decisão são capazes de captar e entender as nuances culturais e regionais de onde estão inseridos; fato subjetivo relacionado ao seu processo individual de tomada de decisão.

...nós fazemos negócios que entendemos com pessoas que conhecemos [...] (ENTREVISTADO 3).

Eu tenho 22 anos de Cooperativa. Eu comecei como estagiário, caixa, depois de caixa eu fui atendente, assistente de negócios, depois eu virei gerente de negócios PF e PJ, depois eu fui gerente de agência [...] (ENTREVISTADO 4).

Por exemplo, eu tenho 33 anos de atividade financeira [...] (ENTREVISTADO 3).

[...] eu tenho 15 anos de HSBC [...] (ENTREVISTADO 6).

Mas a experiência precisa ser complementada com outras iniciativas; do contrário a equipe de gestores poderia ficar envelhecida, fato que poderia prejudicar o processo de tomada de decisão, considerando o dinamismo das mudanças de mercado e tecnológicas. A segunda forma de preparação da equipe de tomadores de decisão se dá através de treinamentos formais e viagens em missões de estudos.

Então o conselho fez uma viagem de estudos para a Alemanha, passou uma semana lá na Alemanha conhecendo a forma de relacionamento do cooperativismo alemão nas comunidades [...] (ENTREVISTADO 5).

[...] nós fechamos uma parceria com uma escola [...] formato de um MBA, só que não tem titulação de MBA, mas é no formato de MBA (ENTREVISTADO 3).

O desenho dos processos aliado as ferramentas para a tomada de decisão

são outros fatores que integram a formação destes profissionais. Essas ferramentas contribuem no desenvolvimento diário das capacidades cognitivas dos profissionais.

[...] nesse mundo que a gente vive hoje da informação fluida é relativamente fácil conseguir as informações, o que precisa é treinar o tomador de decisão a buscar as informações e cruzar elas. Essa habilidade de cruzar as informações; porque algumas a gente consegue cruzar por sistemas, o BI está aí para fazer isso (ENTREVISTADO 3).

[...] porque a gente construiu um modelo de análises, que daí ele é mais quantitativo, mais baseado em dados que a gente consegue, considerando tudo que a gente tem de informações, dos balanços, das demonstrações. [...] e de novo ele não precisa ser um expert em todos os setores. [...] mas ele tem as análises setoriais ali, e inclusive os resumos das análises setoriais. Então quando ele tiver analisando isso tudo ele consegue acessar essas coisas todas que permitem que ele consiga se orientar mais para tomar essa decisão (ENTREVISTADO 6).

Percebe-se, portanto, um conjunto de práticas relativas ao preparo do tomador de decisão. Algumas delas são adotadas diretamente pela empresa, como a criação de processos e ferramentas de suporte, aliados a treinamentos do seu corpo de funcionários, o que inclui em alguns casos missões internacionais de estudo; enquanto outras práticas são requeridas por ela, como o conhecimento de campo e a experiência no setor. Uma vez analisado o conteúdo das entrevistas, foi possível identificar as principais práticas adotadas para o preparo dos tomadores de decisão na instituição estudada, na forma de quatro dimensões.

Figura 8 – Práticas no preparo do tomador de decisão



Fonte: Elaborado pelo autor (2019).

As práticas apontadas na figura acima foram percebidas na trajetória profissional descrita por cada um dos entrevistados, ao contextualizar sua história de vida e presença na organização. Estas práticas podem se traduzir em dimensões de preparo presentes nos profissionais de alto nível hierárquico.

Nesse sentido, todos os profissionais entrevistados possuíam alta experiência no setor. Infere-se, portanto, que sem esse preparo um profissional não ascende na organização. O conhecimento de campo tem direta relação com a experiência setorial, embora possa não ser relacionado a ela, uma vez que pode advir de outras experiências do profissional, ao entendermos “campo” como espaço geográfico e antropológico. Os treinamentos e missões são necessários uma vez que os mercados estão em constante mudança; fato frisado pelos entrevistados. Os processos e ferramentas formam uma dimensão de suporte, traduzindo o preparo do tomador de decisão em ações práticas.

As falas dos entrevistados indicam que o preparo dos tomadores de decisão é de especial relevância. Isso porque, em última instância, a tomada de decisão não é realizada por uma máquina ou software. Ela é realizada por um ser humano. Que carrega consigo suas experiências, traduzidas em um conjunto de subjetividades.

4.6.2 Intuição e subjetividade

Quanto mais experiente o tomador de decisão, mais dados subjetivos ele precisa enfrentar cognitivamente na hora de executar suas tarefas. Isso porque as probabilidades com frequência extrapolam aquelas sugeridas pelos dados; considerando que os cenários projetados pós-decisão são, em si, subjetivos segundo Gilboa, Rouziou e Sibony (2018). Os autores esclarecem que o ser humano enfrenta problemas de avaliação para quantificar as probabilidades, ao acreditar que determinadas conjunções de eventos são mais prováveis do que outras; fato nem sempre suportado pelos dados, afetando a sua decisão. Essa subjetividade se traduz em um verdadeiro desafio, reconhecido para os entrevistados, em acordo com a literatura:

Essa bagagem decisória, ela fica registrada na memória do gestor, infelizmente. E daqui a 2 ou 3 anos, quando tiver uma decisão similar ele vai resgatar isso na memória e vai pensar: “3 anos atrás a gente tinha que fazer algo parecido, a decisão foi essa e deu tudo errado. O meu voto é contra”, independente de qualquer dado. Por isso que a decisão por intuição é

perigosa. [...] (ENTREVISTADO 5).

[...] nós temos muito tempo de experiência, de acompanhamento, de dados, informações, a gente conhece a região. [...] Então quando a gente vai tomar uma decisão, a gente também tem feeling, talvez a gente não tenha a informação aqui do BI, mas a gente tem uma noção aqui de mercado, de praça, de potencial. E a gente vê que isso faz sentido: “se a gente for por esse caminho tenho quase certeza que é o caminho certo a ser percorrido”. Então tem essa particularidade também (ENTREVISTADO 4).

[...] mas eu não posso tapar os ouvidos para minha intuição. Faça; ou não faça que vai dar errado. Eu acabo usando ela. Nas vezes que eu não usei eu sabia, eu sentia que ia dar algum problema. E deu problema, porque que eu não ouvi (ENTREVISTADO 3).

Analise-se aqui o processo de concessão de crédito, um dos pilares do negócio da instituição pesquisada. Uma vez que identificou-se que as alçadas de responsabilidade são um mecanismo que estabelece os limites da subjetividade na concessão ao crédito, é válida a pergunta: ela exerce um papel relevante nesse processo? Ou seja: quanto do crédito disponível pela empresa é concedido tendo por base a subjetividade, e não estritamente dados objetivos?

Uma das formas que pode ser adotada para possível aferição dessa questão é considerar o crédito concedido acima do limite pré-aprovado. Uma vez que o pré-aprovado considera estritamente indicadores construídos a partir de dados já tratados e analisados por algoritmos; o crédito que ultrapassa esse limite está sujeito a análise subjetiva do responsável operacional que o concede. De acordo com os entrevistados, a média das Cooperativas Singulares do Sistema é de cerca de 70% de crédito concedido dentro dos limites pré-aprovados.

Então 70% de todo crédito hoje que é concedido pela Cooperativa, ele é automático ou ele já tem a gestão de limites, que é um limite já pré-analisado [...] E o restante é via gestão de limites, onde o gerente encaminha o crédito para análise de um comitê, que daí é os 30% que não estão aprovados (ENTREVISTADO 4).

[...] a gente deve estar na casa dos 70% do que a gente faz é pré-aprovado aproximadamente. O mercado já roda com números maiores, tem instituições financeiras na casa de 80%, 85% já pré-aprovado (ENTREVISTADO 6).

Então podemos dizer que hoje, 30% da nossa carteira é subjetividade, é confiar no associado, é acreditar no negócio dele, é estar próximo, ter essa condição de validar o que ele não tem confirmado, declarado (ENTREVISTADO 4).

O que eu acho? Eu acho mais justo do jeito que a gente faz, tem mais riscos? No modelo que a gente trabalha não, não aumenta o risco e nem diminui, se você fizer direitinho como tem que ser, dando aderência, você conhece o associado, você tem o histórico (ENTREVISTADO 2).

Poder-se-ia considerar, portanto, que de acordo com as afirmações dos entrevistados, cerca de 30% de todo o crédito concedido tolera alguma subjetividade. Fato que, se traduzido em cifras pode ser considerado de grande relevância. E essa é a percepção dos entrevistados. Porém, em uma análise mais profunda esse raciocínio prova-se apenas parcialmente verdadeiro. Isso porque a instituição conta com uma série de procedimentos e normas para a concessão de crédito, de modo a ampliar a concessão apenas para clientes de menor risco. E esses procedimentos atuam como instrumentos de contenção da subjetividade na tomada de decisão.

Para essa conclusão analisou-se o exemplo de uma das Cooperativas Singulares pesquisadas. Nesta Cooperativa, de acordo com os entrevistados, cerca de 94% do crédito concedido é destinado a associados que possuem uma classificação de risco baixa.

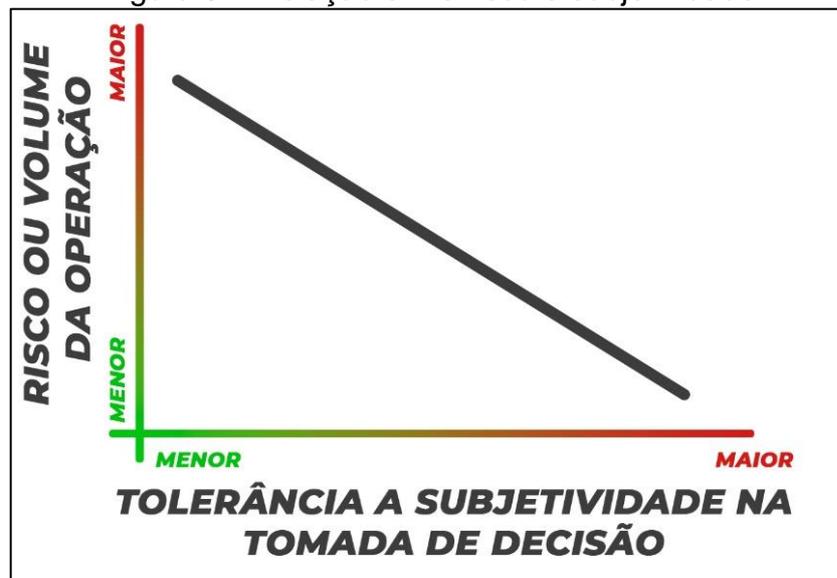
Hoje se a gente olhar a concessão de crédito da Cooperativa (Singular), ela está em 94% dos melhores scores que é o baixíssimo, baixo e médio. Mas mesmo assim nós temos ainda 6% que são para os scores não tão bons, que é o alto e o altíssimo (ENTREVISTADO 4).

Observa-se, portanto, que na organização pesquisada existe sim uma abertura para a subjetividade no processo de tomada de decisão de concessão de crédito. Porém essa subjetividade é contida por processos e indicadores gerenciais. Segundo a fala citada acima, a subjetividade está presente de fato em apenas 6% dos negócios gerados, uma vez que os outros 94% levam em consideração indicadores de classificação de risco baseado em dados.

Quanto menor o valor, no subjetivo melhor. Quanto maior o valor o subjetivo cai. Vai oferecer uma operação de crédito de um milhão de Reais para uma empresa? Ele tem que ter balanço. O subjetivo tem pouquíssimo peso, e quando maior, quanto mais o valor vai subindo, chega um ponto que o subjetivo chega a ser nulo. Não se olha mais para o subjetivo, porque é muito dinheiro em risco e também nós temos muitas informações. [...] quanto menor o valor que a gente está emprestando, mais o subjetivo tem peso (ENTREVISTADO 3).

Neste sentido, identifica-se uma razão entre a presença de subjetividade na tomada de decisão e o risco. Para ilustrar, apresenta-se a Figura 9.

Figura 9 – Relação entre risco e subjetividade



Fonte: Elaborado pelo autor (2019).

Deduz-se, nesse contexto, que embora exista um limite de alçada que tolera a subjetividade, ela tenha resultados práticos menores do que o imaginado pelos entrevistados. Essa conclusão é relevante pois a liberdade de alçada para decisões subjetivas foi citada por todos os entrevistados deste estudo; contextualizada como fator de vantagem competitiva. Porém a realidade mostrou-se sutilmente diferente do cenário identificado pelos entrevistados.

Isto é: na prática, a subjetividade é suplantada pela triangulação de dados que compõe a chamada “aderência”. Em apenas 6% dos casos a intuição e a subjetividade são consideradas na concessão de crédito, apesar do pensamento dos entrevistados seja de que ela possuía importância muito maior. Por outro lado, a percepção do pesquisador é de que mesmo o percentual sendo aparentemente pequeno, para o setor financeiro talvez esse número represente uma parcela significativa de negócios, se comparada com outras instituições desse mercado.

Mesmo assim é interessante notar que a sensibilidade criativa do tomador de decisão (ANDERSON, 2009) está presente na fala dos entrevistados. Nesse sentido a organização por alçadas ocupa papel fundamental.

A baixa tolerância a subjetividade, e conseqüente necessidade por dados para triangulação pode ser explicada também pela condição de conferir segurança ao tomador de decisão. O medo da análise posterior da decisão (especialmente no contexto organizacional) é um fator que afeta a capacidade de estabelecer uma função de utilidade esperada (GILBOA; ROUZIOU; SIBONY, 2018) podendo reduzir

a tolerância ao risco, uma vez que crédito e culpa serão atribuídos após o fato.

Autonomia é bom, mas a gente tem medo de usar, né? (ENTREVISTADO 2).

Só que autonomia vem com um sobrenome que se chama responsabilidade. Então ele (o gerente, concessor de crédito) tem que saber o que está fazendo. Então “Tu conhece o “Jean”?” Tu é o gerente. Você vai lá e pode dar (o crédito), porque tu conhece, porque a gente respeita essa avaliação subjetiva. Mas se der problema tu vai ter que correr atrás e resolver isso aí, porque o relatório dizia que não era para dar (ENTREVISTADO 3).

Observa-se, de acordo com o ponto de vista dos entrevistados, que são poucos os casos em que as decisões desconsideram os dados na concessão de crédito. Porém, na organização pesquisada existem fatores que podem fazer com que o tomador de decisão opte por um veredito diferente daquele sugerido pelos dados: os fatores situacionais, a cultura organizacional e os relacionamentos políticos.

4.7 FATORES SITUACIONAIS

No caso da organização pesquisada, sua operação em sistema gera inconformidades nos dados devido as diferenças regionais econômicas em um país das dimensões geográficas do Brasil. Essas diferenças exigem correções de interpretação.

Em função desse fator, os dados são interpretados e reinterpretados considerando o contexto no qual a tomada de decisão deverá ser realizada. E isso tem impactos estratégicos e operacionais:

O Sistema (Cooperativo) faz uma estratégia sistêmica [...] ele criou um consenso: “Eu aceito esse risco (na concessão de crédito) olhando todos os perfis de renda, propriedades e tudo. [...] (então o Sistema define como limites de crédito) Noventa mil Reais para um agricultor aqui, e noventa mil Reais para um agricultor lá do Mato Grosso que tem um milhão de hectares. Então, os noventa para nós é assertivo, para o Mato Grosso não é. [...] Nós temos crédito, por exemplo, pré-aprovado que vem do Sistema (Cooperativo) [...] eles têm uma amostra mais conservadora, eles eliminam mais situações, porque eles pegam amostras nível Brasil, e faz o seu credscore⁹ [...] nós fizemos um próprio, aqui dentro da Cooperativa (Singular), [...] é um plus, esse aqui (do Sistema), por exemplo, vai até 90 mil Reais. Nós, vamos até 200 mil Reais no nosso aqui. É mais agressivo no nosso [...] (ENTREVISTADO 1).

O que não está totalmente errado [...] cada Cooperativa (Singular) tem o seu momento de vida, tem as suas prioridades, as suas dificuldades e as suas

⁹ Índice interno utilizado para classificação de risco na concessão de crédito.

estratégias específicas. Então é difícil para o Sistema criar um dashboard que atenda todas as Cooperativas, não tem como. Cada cooperativa acaba criando os seus dashboards (ENTREVISTADO 3).

O contexto situacional, aplicado a decisão *Data-Driven* acontece em ordem inversa a proposta por Davenport (2010); uma vez que são as estratégias corporativas que estabelecem quais dados devem ser coletados e analisados, e não emergem deles. Fórmulas de análise são criadas para os dados, e não a partir deles:

Eles foram para o software (critérios de análise desenvolvidos a partir de aprendizado interno). [...] dentro dessa caminhada, também toda essa maturação, para também evoluir, a gente criou um conceito próprio de credscore aqui também, que é diferente do sistêmico (ENTREVISTADO 1).

Uma explicação possível para este procedimento está relacionada com o valor no tempo dos dados. A maioria dos dados disponíveis para as empresas fala mais sobre o passado do que sobre o futuro. Porém as decisões são tomadas para o futuro. Como implicação deste fato, ganha relevância as diretrizes culturais, de planejamento estratégico da empresa além dos aspectos situacionais, como fatores que influenciam as projeções de cenários futuros.

Em janeiro do ano passado a nossa carteira de (crédito em) energia solar era de 3 milhões (de Reais). A gente encerrou dezembro com 30 milhões. Algum orçamento de receita diria que a carteira cresceria 10 vezes? E se a gente fosse olhar a média histórica a gente iria de 3 milhões para 3,3 ou 3,4 milhões. Só que é um mercado novo. Como é que tu vais tomar de base a história, se a história não diz nada de energia fotovoltaica? (ENTREVISTADO 5).

[...] não existe nenhum relatório que vai me dizer isso, porque é futuro. Todas essas questões que forem mais complexas, eu não posso decidir estritamente pelos relatórios, eu preciso dessa entrevista, cada caso é um caso (ENTREVISTADO 2).

A imprevisibilidade inerente a tomada de decisão é enfrentada com a criação de cenários pela organização pesquisada. Esta é uma ferramenta utilizada para a interpretação dos dados e consequente tomada de decisão, e está em acordo com a literatura no contexto da teoria da decisão (GILBOA; ROUZIOU; SIBONY, 2018).

Então foi muito estratégico esse movimento, a gente enxergou as oportunidades e foi para o lado certo. Enxergando, claro por feeling, por conhecer o mercado, por conhecer a região, por conhecer a tendência de mercado. Escutando nosso pessoal do Centro Administrativo que tem uma área específica de análise econômica, dos cenários econômicos, dizendo quais mercados iam continuar aquecidos, que iriam ter oportunidades, e a

gente foi para isso (ENTREVISTADO 4).

Se nós flexibilizássemos o capital social de quem já tem 20 anos de associação a Cooperativa? ou 15, ou 10, quais seriam as faixas? Quanto comprometeria dos nossos demonstrativos dos nossos limites operacionais? [...] Aí a gente solicita simulações para a área de BI da cooperativa (ENTREVISTADO 5).

Além dos resultados esperados na formulação de cenários (GILBOA; ROUZIOU; SIBONY, 2018), as entrevistas mostraram a importância da consideração de fatores situacionais. O momento de vida da empresa, nos seus aspectos mais amplos, estabelece as contingências técnicas e estratégicas da tomada de decisão.

A própria realidade de mercado também se configura como uma contingência a tomada de decisão, especialmente em nível estratégico. Os movimentos da concorrência por vezes são dados relevantes a serem considerados, ganhando mais valor (ou tanto quanto) do que situações internas.

E quando eu assumi, foi bem no momento que a gente tava com a casa começando a ficar organizada. E o que eu fiz, o meu trabalho principal, foi visitar as políticas, rever as estratégias comerciais, [...] de poder conceder crédito com mais velocidade sim, mas com mais qualidade. Conceder para quem era bom, para quem tivesse menos riscos e talvez liberar de novo uma política mais subjetiva, onde desse para dar para aquele cara que era bom, que tinha capacidade, mas não tinha comprovado. Então isso tudo aconteceu. Em 18 meses, a cooperativa cresceu 450 milhões (de Reais) na carteira, porque a gente tinha um espaço
[...] Primeiro pela capacidade de alavancagem da Cooperativa (Singular), a gente tem que saber se a gente tem funding; que é ter recursos. Para ver se a gente tem condições de alocar na carteira de crédito, em um prazo de 10 anos
Outra questão: os prazos. Eles eram condizentes com a tendência de mercado? Com a realidade de mercado? A gente viu que algumas instituições financeiras de ponta, Itaú, Bradesco, outras instituições públicas também, estavam trabalhando com prazos maiores. Então porque a gente não pode (ampliar os prazos) se a gente tem funding? (ENTREVISTADO 4).

Em uma organização concebida como sistema, os fatores situacionais podem ser ainda mais relevantes. Tanto no sentido de fornecer subsídios a tomada de decisão local, através de observação de seus pares; quanto pela consideração de maturidade organizacional de cada unidade.

E a gente viu que algumas Cooperativas do Sistema, principalmente do Paraná, estavam atuando com 100 meses, então a gente teve que olhar para isso (ENTREVISTADO 4).

Inferese, portanto, que os fatores situacionais, expressos ou não pelos

dados, são filtros a tomada de decisão. Isso porque, no ponto de vista dos entrevistados, as decisões devem considerar uma série de situações internas (como capacidade de alavancagem financeira) e externos (como perspectivas de mercado). Seu efeito é de filtro, pois uma contingência situacional pode levar uma decisão ao caminho inverso daquele apontado pela análise dos dados históricos coletados pela organização. Ou seja, uma decisão apenas ocorre no sentido dos dados se concordar com os fatores situacionais percebidos pelo corpo de tomadores de decisão. Ainda em relação ao momento de tomada de decisão, outros fatores possuem o mesmo efeito: a cultura organizacional e o relacionamento político.

4.8 CULTURA ORGANIZACIONAL

Na organização pesquisada, a cultura organizacional pode ser entendida como um filtro necessário, e por vezes determinante, para a tomada de decisão, fato suportado na literatura por Huffman *et al.* (2006). Apesar de a organização atuar em um setor com alta regulação como o financeiro, de onde se poderia imaginar que os dados tenham papel determinante na tomada de decisão, a cultura desempenha uma função importante.

Em alguns casos, a cultura se sobrepõe aos dados de tal maneira que eles são desconsiderados. Ou seja: tomam-se decisões à revelia do que os dados indicam tecnicamente como melhor caminho.

Tudo que a diretoria executiva faz, ela olha para o mapa (estratégico) e vê se conversa ou não conversa com o mapa. Tem 100% de convergência? Então dá para fazer, porque tem link com o pensamento estratégico do conselho. Vamos investir em cervejarias? Não? Mas no mapa manda pensar em turismo; turismo como proposta de valor, como diferencial para o público da nossa região. Então conversa com o mapa. Então nós vamos fazer (ENTREVISTADO 5).

Conversa com o mapa? Não? Mesmo que dê dinheiro, a gente não faz. Dói. Mas é assim (ENTREVISTADO 2).

E dentro da Cooperativa (Singular) nós temos três elementos para tomada de decisão que são base para nós, que são elementos que depois de analisar os dados a gente olha se tem coerência com isso. Um é o nosso propósito enquanto cooperativa; a nossa proposta de valor, e nossos valores. [...] Pode acontecer dos dados me dizerem uma coisa, que está ferindo uma das propostas de valor. Mas aí eu não abro mão da proposta de valor, dos valores propostos e não faço a decisão (ENTREVISTADO 1).

Nota-se que a cultura organizacional pode ser entendida como um filtro a

tomada de decisão. Aqui entendemos cultura como a expressão do conjunto de valores e práticas exercidas pela organização, identificadas na fala dos entrevistados. Nesse contexto, a cultura carrega em si alguma carga de subjetividade, uma vez que, em última análise, uma mesma situação pode ser avaliada em diferentes prismas; dependendo da pessoa e do contexto no qual está o tomador de decisão.

Nas Cooperativas Singulares pesquisadas, o instrumento que traduz a cultura é o mapa estratégico. Este documento é construído em assembleias de associados e gestores, reunindo informações de direcionamento estratégico como princípios de atuação, bandeiras sociais e proposta de valor. Um destes mapas consta no Anexo A.

[...] a cooperativa construiu o planejamento estratégico visão 2018-2020 olhando o mercado. [...] trabalhando muito forte a proposta de valor, o que nós queremos que os nossos associados enxerguem de diferente e agregador para eles, na parceria que eles têm conosco. E através disso a gente direcionou a nossa energia, o nosso foco para esses mercados, para esses segmentos (ENTREVISTADO 4).

Um dos pontos centrais da cultura da empresa pesquisada se baseia na ideia de “conhecer o associado”. Embora este conceito carregue uma parcela de subjetividade, ele se relaciona com a coleta e análise de dados; operando especialmente através de visitas diretas e entrevistas conduzidas junto aos associados.

A gente foi conhecer o sistema alemão de cooperativa, e o sistema alemão ele se promete assim: “Nós fizemos negócios que entendemos, com pessoas que conhecemos”. Parece uma frase bonita, simples, mas a gente já fez isso agora [...] Porque o que acontece, essa proximidade, essa visita. Aí o (dado) que vem pronto sistematicamente não consegue ter, é frio. [...] mas tem situações, que são coisas setoriais e regionais. Por exemplo, você não encontra em praticamente nenhum mercado, nenhum analista financeiro, que fale sobre o segmento da uva e do vinho. Economistas eles falam da soja, do aço, das commodities. [...] e a nossa cadeia é 80% uva e vinho aqui (ENTREVISTADO 1).

A cultura da organização se apresenta também como um desafio a implantação do *Data-Driven*. E a forma encontrada pela empresa para gradualmente romper os dogmas internos é a apresentação de dados e seus benefícios, comparativamente as limitações humanas de retenção e cruzamento de informação.

Eu tive algumas agendas que eu precisava quebrar, resistência forte de

gerentes muito antigos. Aquele cara que era o big boss de uma agência, que tinha a carteira mais premium da agência. Todos os associados como maior rentabilidade da agência que ele visitava os caras para conversar. E eu fiz simulações de atendimento [...] me botando no lugar do cliente, porque eu sabia quem era, eu estudei a pessoa antes e comecei a fazer perguntas para o gerente que teoricamente sabia tudo sobre aquela pessoa. [...] comecei mostrar para uma plateia que o cara, mesmo sendo o cara top da agência não conhecia. E que ele tinha uma carteira no passado de 50 pessoas e agora ele tinha 300 pessoas. Então mesmo o cara mais premium, ele achava que conhecia. E quando eu comecei a mostrar coisas que nem essa a gente começou a quebrar a resistência de adoção de uma ação de CRM para apoiar a máquina comercial. [...] somaria isso a resistência a mudança. O ser humano naturalmente é resistente à mudança. E segundo, um pouco de amarra ao modelo anterior que nós tínhamos [...] quem vem de mercado já vem sedento por esse tipo de coisa. Como o as Cooperativas formaram outro cenário, as pessoas estão ainda muito amarradas nesses modelos mais antigos do que em um modelo que acredita muito nos modelos dos números que estão vendo para a tomada de decisão (ENTREVISTADO 7).

A tomada de decisão *Data-Driven*, portanto, parece fazer parte do cotidiano da organização em diferentes níveis, quando olha-se para o contexto organizacional e hierárquico. Por outro lado, os arranjos da cultura organizacional podem admitir a tomada de decisão baseada não em dados, mas em fatores culturais internos. O mesmo ocorre no âmbito das relações externas, quando considera-se o relacionamento político.

4.9 RELACIONAMENTO POLÍTICO

Uma característica relevante da organização pesquisada é sua composição em Cooperativas Singulares. Essa forma descentralizada de atuação possui impactos diretos na sua atuação junto as comunidades onde o Sistema está inserido. E talvez um dos principais impactos seja o foco acurado no relacionamento com os seus associados.

Isso ocorre desde a contratação de profissionais locais, quanto na sua participação em entidades e órgãos participativos locais. Essa participação estende o relacionamento político das Cooperativas com as lideranças e comunidade local. E esse relacionamento é reconhecido pelos entrevistados como forte diferencial de mercado, induzindo uma vantagem competitiva.

A gente está presente nas entidades, nos conselhos. No Observatório Social, na CICS. Com parcerias (ENTREVISTADO 2).

Qual é a nossa vantagem? Nós conhecemos as pessoas que vivem aqui, nós falamos o mesmo sotaques deles. Como diz um executivo de uma

cooperativa da Alemanha que visitei: “Os meus filhos vão nas mesmas escolas que os filhos dos meus associados” [...] a gente conhece as dores das comunidades, dos associados. E isso faz com que a gente tenha estratégias muito pontuais que atendem exatamente aquilo que eles estão precisando (ENTREVISTADO 3).

A organização já reserva inclusive fundos para esse relacionamento político com as comunidades. Além de possuir fundos formais, com recursos destinados a entidades locais que inscrevem seus projetos para competir por verbas; os gerentes locais possuem verbas livres para uso a seu critério. Em 2018, dados do relatório anual de uma das Cooperativas Singulares pesquisadas aponta mais de meio milhão de Reais distribuídos a mais de 100 projetos sociais nas cidades cobertas por ela.

Eu vou te dar um exemplo de decisão simples: hoje um gerente nosso tem total autonomia para tomar a decisão, por exemplo, se ele quer aplicar 5 mil (Reais) em uma festa da comunidade. Ele pode tomar essa decisão. [...] Porque a decisão de investir 5 mil Reais em uma festa de qualquer legume, não é uma decisão de dados, é uma decisão subjetiva, uma decisão política. Vale a pena aquela comunidade? É uma comunidade parceira de um modo geral? Como a nossa marca vai ser vista em estar presente ou não estar presente? (ENTREVISTADO 3).

É interessante notar o vulto financeiro que pode alcançar uma decisão tomada tendo por base razões políticas subjetivas, em detrimento da racionalidade dos dados. Grandes investimentos, até mesmo como a instalação de agências inteiras passam por um filtro político de tomada de decisão. E essa realidade é aferível: em um dos municípios de atuação de uma das Cooperativas Singulares pesquisadas, com menos de cem mil habitantes, operam três de suas agências. No mesmo município todas as outras instituições financeiras presentes possuem apenas uma agência física cada.

Existe espaço sim para isso, nas cooperativas existem mais fortes até que nos bancos: existem as decisões que precisam ser puramente técnicas, mas sempre há um tempero que é a decisão política. Tem política em uma decisão? Sempre tem, inclusive em uma cooperativa. Inclusive essa decisão política, ela tem um peso maior, por ser cooperativa, por ter relação com várias entidades, por ter uma preocupação real com o desenvolvimento das comunidades que a gente está presente. Então a gente algumas vezes releva algum dado que aparecia. Vou dar um exemplo bem típico: tem localidades que a gente tem agência que não faz sentido ter, o resultado é pequeno. Se é um banco, é só tu olhar que os bancos realmente não têm. Os bancos não têm agências em muitos municípios que tu conhece, porque é uma decisão puramente técnica. Nós não podemos tomar uma decisão técnica. Nós temos que tomar uma decisão política, que daí vai ter que considerar que a Cooperativa tem um papel social, é a única instituição financeira presente. Se não tiver a Cooperativa onde é que os aposentados vão receber as aposentadorias deles? Onde é que o pequeno comércio vai depositar o dinheiro dele? [...] Então a cooperativa tem uma função social, que é o campo

político da decisão que a gente leva em conta muitas vezes. Nisso uma cooperativa é muito mais forte que um banco (ENTREVISTADO 1).

“Poxa, onde é que eu vou abrir a minha próxima agência?” Porque eu preciso de uma ação de marketing que financeiramente eu não sustento, mas o efeito da imagem é extremamente importante. Ou política na região de uma cooperativa. Se eu não botar a minha bandeira naquela cidade um outro concorrente vai vim colocar, e eu preciso politicamente me estabelecer naquele lugar, mesmo sabendo que ela é financeiramente viável com um prazo mais extenso. O número mostra que não deveria [...]. Hoje eu diria que, na minha visão, é 90/10: os dados têm peso de 90% nesse processo de tomada de decisão, em 10% eles acabam escorregando por motivos de outra situação. Não vejo que é muito diferente em outras empresas, não imagino que seja 100% dados; acho que isso é meio que, na minha visão, uma utopia. E eu estou bem feliz com o 90/10 (ENTREVISTADO 5).

A estrutura da organização pesquisada ainda prevê a participação do associado através de assembleias e conselhos que tem poder de decisão. Dali emergem decisões que vão desde a aprovação para a construção de uma nova agência quanto aumentos de salários dos presidentes das Cooperativas Singulares. O mesmo ocorre em âmbito interno: decisões estratégicas das mais variadas são tomadas em comitê.

[...] quando a gente decide uma estratégia ou modelo de atuação, a gente sempre senta os três juntos (diretores) com os assessores. Eu tenho mais dois assessores [...] então são 5 pessoas que se reúnem para definir uma estratégia e decidir de forma colegiada, ver qual é o melhor caminho a se tomar (ENTREVISTADO 4).

A criação da agência agro é um exemplo. Eu tive a ideia, baseada em uma realidade que eu estava vendo. E porque ia ter vantagens para a minha agência também, embora eu perderia alguns clientes. Aí eu conversei com o nosso gerente de desenvolvimento. E ele me disse: “Isso conversa com o mapa estratégico, vamos fazer”. E em alguns meses a agência estava de pé (ENTREVISTADA 2).

É importante considerar ainda que decisões tomadas em assembleias, conselhos ou comitês exigem o convencimento de seus integrantes, uma vez que é o voto coletivo que decide. E isso exige percepção e condução política por parte das lideranças das Cooperativas, tanto na tentativa de que a sua decisão seja a escolhida pelos conselhos; quanto para o convencimento dos níveis operacionais para a execução do que foi decidido.

Uma coisa é quando tu vai lá e diz: “Eu decidi fazer assim, assim e assim”. Eu não tenho consenso, eu não tenho participação, aí as pessoas começam a dizer: “Ah, é fácil. Então o cara lá sentado na salinha dele decidiu e agora manda a gente fazer. Quero ver fazer isso, não vai ser bem assim”. Aí a gente não tem o engajamento, a participação, o próprio convencimento da equipe.

E quando a gente monta um projeto, monta uma ideia, uma estratégia de forma construtiva, com todo mundo participando, dando ideias, quando a gente vai lançar ela: “Pessoal para nós chegar aqui, nessa estratégia, nessa decisão, a gente consultou o cara da revenda, esse, esse e esse, todo mundo ajudou na construção da estratégia. Então esse é o melhor caminho, vocês concordam? ”. Então a gente percebe que a forma para se construir, se planejar, demora mais tempo, mas quando a gente vai para a implementação, quando a gente vai para executar essa estratégia; o resultado, o engajamento, o comprometimento, ele é muito maior (ENTREVISTADO 4).

Então qual é o grande desafio das lideranças que estão a frente? É política, é eu saber que o “Jean” é um cara importante para mim dentro do conselho. Preciso do apoio dele em um determinado assunto e ele é contra esse assunto. E se eu trazer esse assunto para ele, ele vai ser contra. Mas o “Jean”, ele é um cara inteligente. Então eu preciso saber qual indicador que vai contra-argumentar a intuição dele; e em que momento eu vou apresentar esse indicador para ele. Muito antes da reunião. Como é que eu vou preparar o “Jean” para quando ele receber a pauta, ele cair a ficha que eu estava correto, é bem político (ENTREVISTADO 5).

Infere-se, portanto, que o relacionamento político atua como um relevante filtro a tomada de decisão, no sentido de que pode conduzi-la para um caminho diverso daquele proposto pelos dados. Em outras palavras, o relacionamento político pode invalidar completamente as análises de dados presentes em todo o processo decisório. Esse filtro pode ter mais ou menos importância, dependendo das dimensões de risco atinentes a decisão a ser tomada.

4.10 *FRAMEWORK* PROPOSTO

Esta seção apresenta na Figura 10, o esquema conceitual construído a partir da relação entre os referenciais teóricos e os resultados deste estudo. Os pressupostos teóricos relacionados a este estudo são os conceitos de *Big Data*, tendo como base suas dimensões Volume, Velocidade, Variedade, Veracidade e Valor, conforme identificadas por Wamba *et al.* (2015); e *Data-Driven Decision Making*, analisado sob o prisma do processo de extração de insights dos dados, conforme expresso por Gandomi e Haider (2015).

O *framework* concatena as descobertas e percepções que emergiram nas análises. Como uma primeira etapa no macroprocesso de decisão, aqui denominada “Gerenciamento de Dados”, seguindo o modelo de Gandomi e Haider (2015), estão as duas etapas iniciais do processo de tomada de decisão Data-Driven: a Aquisição e o Registro dos dados, relacionado diretamente com seu volume e com a velocidade de processamento; além de Extração, Limpeza e Comentários dos dados. Essa última

etapa se relaciona diretamente com o objetivo de conferir veracidade aos dados.

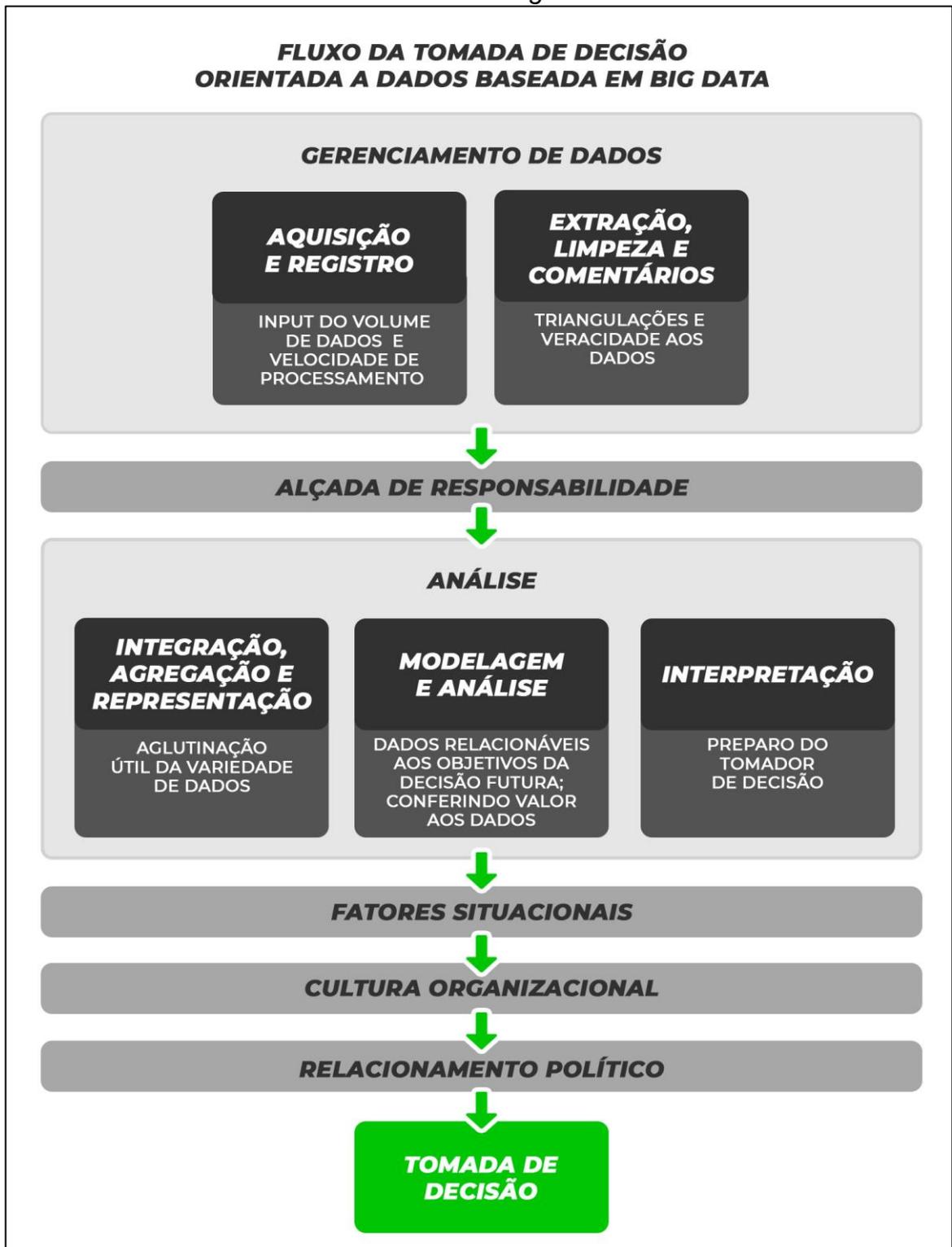
Uma vez que os dados foram preparados, são então submetidos a um filtro de alçadas de responsabilidade. Isso é, cada tomador de decisão terá acesso aos dados concernentes apenas a sua alçada de decisão.

Já dentro de sua alçada, o tomador de decisão terá acesso aos dados aglutinados em painéis úteis ao seu trabalho. Essa aglutinação de dados lhe permite realizar a sua interpretação, e tem direta relação com a variedade de tipos e fontes de informações disponíveis. Diferente do modelo desenhado por Gandomi e Haider (2015), aqui propõe-se que a etapa de Integração, Agregação e Representação seja a primeira no macroprocesso de análise dos dados. Isso porque na pesquisa identificou-se que a aglutinação dos dados varia de acordo com a alçada do tomador de decisão. Ou seja: é alçada que determina quais dados serão analisados, e quais as suas combinações relevantes.

Também se faz necessário criar modelos de análise dos dados, estabelecendo relações com os objetivos da decisão futura. Aqui residem processos para conferir diferentes camadas de valor aos diferentes tipos de dados. A interpretação, por sua vez, é a dimensão humana da análise. E essa dimensão é correlacionada diretamente com o preparo do tomador de decisão.

Porém, a decisão só pode ser tomada após ser avaliada através de três últimos filtros. O primeiro deles tem a ver com fatores situacionais da organização. Aqui são consideradas, por exemplo, as contingências da organização naquele momento, bem como análises de mercado. O segundo filtro é o da cultura organizacional. Uma decisão precisa contar com algum nível de concordância com os valores e diretrizes culturais da empresa. E, por fim, há o filtro das relações políticas. Nessa etapa são consideradas as implicações e efeitos políticos da tomada de decisão; sejam eles relativos à marca, a imagem percebida da organização ou de seus membros pelas comunidades nas quais a empresa está inserida, entre outras questões deste contexto.

Figura 10 – *Framework* proposto para o fluxo de tomada de decisão *Data-Driven* baseado em *Big Data*



Fonte: Elaborado pelo autor (2019).

O *framework* aqui proposto revela que há um processo estruturado na tomada de decisão baseada em dados no contexto *Big Data*. As etapas são dispostas dentro de uma lógica de ordenação prática que se relaciona com a literatura, combinando o

modelo de Gandomi e Haider (2015) com as dimensões do fenômeno *Big Data* presente em Wamba *et al.* (2015). Nota-se então que de fato existe um fluxo que torna o *Big Data* intrinsecamente ligado à tomada de decisão.

A particularidade dos achados da pesquisa está no fato de que os resultados revelaram a presença de filtros, que ocupam função condicionante no processo decisório, como as Alçadas de Responsabilidade, os Fatores Situacionais, a Cultura Organizacional e o Relacionamento Político. Ou seja: uma decisão só é tomada na medida em que supera os filtros. Por outro lado, para que uma determinada decisão não seja tomada (ou seja tomada à revelia daquela indicada pelos dados) basta que um dos filtros seja intransponível. A posição da alçada de responsabilidade no *framework* foi determinada de acordo com a análise dos processos da empresa pesquisa relatados nas entrevistas. Da mesma forma foi determinada posição dos demais filtros.

Nesse sentido, após as análises e interpretação dos dados, o tomador de decisão considera a viabilidade geral de sua decisão analisando os fatores situacionais. Uma vez identificada a viabilidade, a decisão deve possuir algum grau de consonância com elementos culturais da organização, como seus valores e diretrizes de planejamento. Por fim, uma análise do impacto nos relacionamentos políticos da organização também é levada em consideração.

Na organização estudada, de acordo com o ponto de vista dos entrevistados, se os dados apontam para um lado e a cultura organizacional aponta para outra, decisões passadas mostram que a cultura organizacional terá maior peso e poderá conformar a decisão a ela. O mesmo vale para o relacionamento político. Se uma decisão acertada do ponto de vista financeiro e gerencial afetar de modo relevante os relacionamentos políticos da organização, ela não será levada adiante. Obviamente a tomada de decisão se dá em um contexto que se movimenta, porém, os filtros existem de fato.

Nesse sentido, os filtros contingenciam a tomada de decisão. Apresentam-se na forma de condições internas da organização, como sua estrutura de alçadas e sua cultura; mas também como fatores relacionados ao ambiente externo a ela, como situações de mercado e relações políticas institucionais.

5 CONSIDERAÇÕES FINAIS

Esta seção apresenta as considerações finais da pesquisa, tendo em consideração os objetivos específicos e, por conseguinte ao objetivo geral deste estudo, que buscou analisar as práticas decisórias baseadas em dados no contexto do *Big Data*. No decorrer deste estudo foram identificadas as relações existentes entre os conceitos de *Big Data* e *Data-Driven Decision Making*.

Apresentam-se a seguir as contribuições deste estudo para a teoria e pesquisa e para a prática organizacional, e ao final são discutidas as limitações da pesquisa e sugestões para estudos futuros.

5.1 CONTRIBUIÇÕES PARA A TEORIA E PESQUISA

Este estudo pode contribuir com a teoria através de suas descobertas, extraídas da pesquisa de campo. A primeira descoberta relevante é sinergia entre as subcategorias estudadas dos dois conceitos principais, o que possibilitou inclusive uma análise unificada.

Esta pesquisa demonstrou que as dimensões do Big Data (*Volume, Variedade, Velocidade, Veracidade e Valor*) identificadas por Wamba *et al.* (2015), são sinergicamente associadas as etapas do processo de extração de insights dos dados para tomada de decisão; conforme expresso por Gandomi e Haider (2015) nas etapas de Aquisição e Registro; Extração, Limpeza e Comentários; Integração; Agregação e Representação; Modelagem e Análise e Interpretação. E isso se dá em um contexto *Data-Driven*, em acordo com Davenport (2010).

A convergência entre os conceitos foi identificada a partir das falas dos entrevistados. Além de responder a um dos objetivos específicos desta pesquisa, uma vez que é resultante da caracterização do processo de análise de informação de fontes de *Big Data* considerando suas dimensões; este fato também se revela importante no contexto de tomada de decisão. Isso porque através de das análises fica claro que as dimensões do *Big Data* não têm um valor em si para os tomadores de decisão de alto nível hierárquico. Elas precisam ser tratadas pela organização de modo a fornecer insumos para o processo de tomada de decisão. É com base nisso que se estabelece uma relação sinérgica entre os conceitos e teorias estudadas.

Essa relação é expressa ainda quando percebe-se que, para os tomadores

de decisão, as dimensões do *Big Data* são entendidas como desafios que exigem respostas imediatas sob a forma de processos, estruturas e departamentos. Ao mesmo tempo, são também seu objeto de desejo; uma vez que observa-se, em suas falas, que esperam que a organização adquira cada vez mais dados, em um volume crescente e constante. Também buscam ampliar as fontes de dados, de modo a trazer maior variedade de informações para fins de triangulação. Essa triangulação se traduz como importante procedimento no sentido de conferir maior veracidade aos dados; fator relevante para a organização. A necessidade de velocidade crescente, tanto na aquisição e tratamento de dados quanto em todo o processo decisório também é expressa pelos entrevistados; em especial quando citam as expectativas do mercado. Em face dessa realidade os dados possuem grande valor para a organização, e o valor dos dados é extraído após o tratamento correto de todas as suas dimensões.

Outra característica desta relação é expressa no fato de que as dimensões do *Big Data* convergem, cada uma, com diferentes etapas do processo de extração de *insights* dos dados para a tomada de decisão. Poder-se-ia, por exemplo, supor que todas as dimensões do *Big Data* estivessem presentes em cada uma das etapas do processo decisório. Ou ainda, que alguma fosse mais ou menos relevante do que outras. Porém essa relação se deu de forma ordenada, conforme apresentado no *framework* proposto (respondendo também a outro objetivo específico desta dissertação): Volume e Velocidade, Veracidade, Variedade e Valor. A presença de uma ordenação sequencial é consonante com a literatura (MAXWELL, ROTZ, GARCIA, 2016). Por outro lado, é importante destacar que essa estrutura específica surgiu considerando estritamente o contexto pesquisado, não permitindo sua generalização.

Embora algumas dimensões fossem mais citadas nas entrevistas do que outras, bem como algumas etapas do processo decisórios mereceram maior relevância por parte dos entrevistados; sua importância nas falas deveu-se mais a posição hierárquica do respondente e sua proximidade com os processos de tratamento de dados do que com a relevância dos conceitos em si. Adicionalmente, nota-se que a interdependência das etapas de extração de *insights* dos dados com as dimensões do *Big Data* ocorre de forma integral em toda a extensão do processo de tomada de decisão.

Outro objetivo específico deste estudo foi analisar os desafios presentes no processo de tomada de decisão baseado em informação. O primeiro ponto relevante

a ser citado é que a lógica proposta por Nnaji *et al.* (2018), de que dados coletados determinam as estratégias corporativas (importante elemento na interseção do *Big Data* com a decisão *data-driven*) não foi confirmada, pelo menos não em sua integralidade. Isso pode ser explicado em parte porque na organização pesquisada, em muitos casos, ainda são as estratégias que determinam quais dados serão coletados, como serão analisados e qual valor será atribuído a eles, e não o oposto. Nesse sentido, infere-se que a organização pesquisada ainda não se aproveita do potencial total do fenômeno *Big Data* para a criação de suas estratégias e operações; uma vez que diferentes cenários poderiam ser extraídos de combinações criativas de dados. Essas combinações estatísticas poderiam, por exemplo, extrair importantes características comuns em muitas subpopulações de dados, mesmo quando há grandes variações individuais (FAN *et al.*, 2014).

Vieses estatísticos, endogeneidade e inconsistências na seleção dos dados (FAN, LIAO, 2012) não se apresentaram como grandes preocupações dos entrevistados; embora a presença de ruído e lacunas informacionais seja uma realidade reconhecida por eles. Isso pode ser reflexo de sua posição na organização, uma vez que não participam diretamente dos processos de limpeza de dados, por exemplo. Os entrevistados demonstraram elevada crença na qualidade de seus processos de tratamento de dados, relacionada a veracidade das informações utilizadas por eles. Destaca-se ainda que, ao contrário do que é apontado por Lutz e Boucher (2016), os tomadores de decisão na empresa pesquisada demonstraram compreensão, pelo menos superficial, acerca dos processos de classificação e limpeza dos dados.

Já o uso de extrapolações de dados para simulação de cenários futuros é corrente na organização; prática apontada na literatura por Chan e Payne (2017). Por outro lado, os tomadores de decisão reconheceram, em muitas de suas falas, que há uma imprevisibilidade inerente ao futuro. Diante desse entendimento, a realização de análises preditivas ainda está, na organização pesquisada, mais relacionada com dados de ordem qualitativa do que quantitativa.

Entendendo o processo de tomada de decisão como um todo, Spillecke e Umblijs (2013) apontam a necessidade de uma equipe multifuncional para garantir a calibração do modelo de decisão e a correta interpretação dos dados; fato que foi parcialmente identificado na empresa pesquisada. Embora existam assembleias compostas por pessoas com diferentes *backgrounds* técnicos e culturais, elas atuam

apenas sobre algumas decisões estratégicas. No caso estudado, os tomadores de decisão entrevistados (envolvidos com as operações diárias da organização) possuem um forte conjunto de similaridades relativas a sua carreira profissional. Se por um lado isso remete a empresa a decisões teoricamente mais técnicas, tomadas por um corpo de especialistas, também pode limitar a assertividade das decisões; fato apontado por Shah, Horne e Capellá (2012).

Nas análises percebeu-se que as decisões na organização pesquisada tendem a ser dirigidas aos dados, mas não completamente dirigidas por eles. Essa conclusão tem direta relação com outro objetivo específico desta pesquisa, onde buscou-se entender o papel das características humanas, como a intuição no processo decisório.

Demonstrou-se que a instituição pesquisada oferece ambiente para que a tomada de decisão ocorra considerando, em algum momento, a intuição e a subjetividade; mesmo em níveis operacionais. Por outro lado, a empresa instituiu processos e diretrizes, como as alçadas de responsabilidade, no sentido de controlar seu impacto nas decisões. Nesse ponto este estudo contribui com a teoria ao identificar a presença de filtros aplicados no processo de tomada de decisão, mesmo no contexto *Data-Driven*. Os filtros são uma forma possível para traduzir de modo prático o entendimento de Chan e Payne (2017), ao explicarem que mesmo em empresas que usam dados de forma avançada, a dimensão humana da intuição ainda desempenha um papel decisivo no processo decisório.

Esta pesquisa demonstrou que a organização pesquisada estabelece quatro filtros a tomada de decisão: sua estruturação por alçadas, fatores situacionais, sua cultura organizacional e os relacionamentos políticos. Em perspectiva, isso significa que uma decisão só será tomada (seguindo os dados) desde que ocorra na alçada correta, seja viável considerando as contingências, esteja de acordo com os preceitos culturais da organização e não fira eventuais interesses políticos. Obviamente uma decisão pode ser tomada desconsiderando algum destes critérios, porém eles são considerados.

A organização por alçadas destacou-se por traduzir em mecanismo prático a proposição de Davenport (2010) ao considerar que o contexto *data-driven* reduz a presença de fatores subjetivos como a intuição nas decisões. Os fatores situacionais por sua vez são responsáveis por estabelecer as contingências operacionais. A cultura organizacional é um traço relevante na instituição pesquisada, e sua

importância no contexto *data-driven* é destacada na literatura (MAXWELL, ROTZ, GARCIA, 2016). Já a presença do relacionamento político como componente do processo de tomada de decisão foi uma descoberta inesperada desta pesquisa, não encontrada na literatura pesquisada.

E este estudo apontou, através do ponto de vista dos entrevistados, que a tomada de decisão *data-driven* é uma realidade na organização pesquisada; porém este processo é abalado pelo relacionamento político. Ou seja: tomam-se decisões contrárias ao diagnóstico puro dos dados, por razões completamente subjetivas; que tem mais a ver com marca e relacionamento institucional do que com performance financeira e operacional. Cabe notar que esta pesquisa não foi conduzida em uma empresa privada comum, mas em uma que se organiza como cooperativa. E a composição de uma cooperativa, por si só já denota algum nível de política na organização; uma vez que é este tipo de empresa é composta por diversos mecanismos de participação de seus associados, como assembleias e conselhos.

5.2 CONTRIBUIÇÕES PARA A PRÁTICA ORGANIZACIONAL

O estudo desenvolvido aqui contribui com a prática organizacional através do *framework* proposto. Este esquema tem potencial de cunho orientativo a prática de tomada de decisão *Data-Driven*, em organizações que lidam com dados que possuam as dimensões do *Big Data*.

O *framework* apresenta indícios dos procedimentos a serem adotados em cada fase do processo. Relativamente a Aquisição e Registro dos dados, é importante destacar a necessidade de uma infraestrutura de tecnologia adequada para o processamento do volume e variedade dos dados.

Em relação a Extração, Limpeza e Comentários, é primordial a escolha correta das fontes de dados, uma vez que elas serão a matéria-prima para a tomada de decisão posterior. Ressalta-se ainda a necessidade de dados de triangulação, como forma de mitigar potenciais erros e lacunas.

A definição de alçadas de responsabilidade é indispensável ao processo decisório, como forma de assegurar o alinhamento dos dados com os diferentes níveis estratégicos e operacionais. Na visão do pesquisador, este ponto é nevrálgico para o processo de decisão; na medida em que a divisão por alçadas orienta toda a organização, e permite a análise compartimentalizada de grandes problemas, o que

pode facilitar inclusive a implantação de medidas operacionais.

Segue-se a etapa de Integração, Agregação e Representação; tendo como ponto central a aglutinação útil da variedade de dados. Esta pesquisa demonstrou que este ponto é encarado como um desafio relevante, e vital para a qualidade da decisão. Assim recomenda-se especial atenção a este ponto, além do direcionamento constante de esforços no sentido de manter as tecnologias utilizadas pela empresa sempre atualizadas. Isso é importante uma vez que o desafio da aglutinação de dados é potencializado pelo tempo: quanto mais dados, mais difícil a transposição de tecnologias.

A fase de Modelagem e Análise opera a partir dos indicadores e fórmulas elaborados pela empresa para aferir seu desempenho e orientar suas estratégias. Ao estabelece-los, é importante considerar o histórico da organização e as tendências e planos que apontam para o futuro. Esse ponto é diretamente relacionado com a etapa de Interpretação, uma vez que tem como crucial dimensão o preparo do tomador de decisão.

Aqui identifica-se outra contribuição potencial deste estudo, ao perceber um conjunto de práticas relacionadas ao processo de preparo dos tomadores de decisão. As dimensões relacionadas podem orientar processos seletivos e de construção cognitiva dos profissionais com responsabilidade decisória.

A consideração dos filtros “Fatores Situacionais”, “Cultura Organizacional” e “Relacionamento Político” também contribuem a prática organizacional, no sentido de trazer luz para sua importância. Além disso, as empresas podem criar procedimentos para normatizar estes filtros, uma vez que sua existência é um fato, provado por sua presença mesmo em instituições financeiras.

5.3 LIMITAÇÕES DO ESTUDO E SUGESTÕES PARA PESQUISAS FUTURAS

Entre as limitações deste está sua realização em caso único, resultando em uma visão restrita do tema. Estudos de múltiplos casos podem apresentar dados diferentes ou adicionais aos aqui apresentados. A falta de maior transversalidade no tempo, permitindo maior observação direta é outra limitação incorrida nesta pesquisa. Uma pesquisa conduzida com técnicas que permitam uma análise ao longo de maiores períodos de tempo poderia apresentar novas descobertas sobre o tema pesquisado.

Como sugestões para estudos futuros aponta-se ainda a validação do *framework* aqui proposto em diferentes organizações. A instituição financeira possui um conjunto de idiosincrasias que certamente não estão presentes em outros tipos de organização. Um estudo de casos múltiplos, por exemplo, poderia revelar se o processo identificado em uma organização financeira pode ser extrapolado para outros tipos de empresas; identificando ainda se o contexto *Big Data* muda dependendo do contexto da organização.

Também sugere-se estudos mais aprofundados sobre a dimensão política da tomada de decisão em empresas orientadas a dados. Especialmente ao entender-se que os relacionamentos políticos possuem desdobramentos externos e internos na organização. Uma pesquisa específica sobre este tema poderia trazer maior luz a este filtro a tomada de decisão no contexto *Big Data*.

REFERÊNCIAS

- AALTONEN, A.; TEMPINI, N. Everything counts in large amounts: A critical realist case study on data-based production. **Inform. Technol.**, v. 29, n. 1, p. 97–110, 2014.
- ACCENTURE ANALYTICS. **Accenture Big Success with Big Data Survey**. 2014. Disponível em: https://www.accenture.com/t00010101T000000Z__w__/_mx-es/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Dualpub_7/Accenture-Big-Data-POV.ashxla=es-LA. Acesso em: 01 abr. 2018.
- AGARWAL, R.; WEILL, P. The benefits of combining data with empathy. **MIT Sloan Management Review**, v. 54, n. 1, p. 35, 2012.
- AGRAWAL, D. *et al.* Challenges and opportunities with big data. **A community white paper developed by leading researchers across the United States**, v. 5, n. 1, p. 34-43, 2012.
- AKERKAR, R. **Big data computing**. Boca Raton: Taylor & Francis Group, LLC, 2014.
- AL-SAI, Z. A.; ABUALIGAH, L. M. Big data and E-government: A review, **8th Int. Conf. Inf. Technol.**, p. 580–587, 2017.
- ALAVI, M.; LEIDNER, D. E. Knowledge management and knowledge management systems: Conceptual foundations and research issues. **MIS quarterly**, v. 25, n. 1, p. 107-136, 2001.
- ALENCAR, E. **Introdução à metodologia de pesquisa**. Lavras: UFLA, 2001.
- ALEXANDER, A.; KUMAR, M.; WALKER, H. A decision theory perspective on complexity in performance measurement and management. **International Journal of Operations & Production Management**, v. 38, n. 11, p. 2214-2244, 2018.
- ANDERSEN, T. B. E-Government as an anti-corruption strategy. **Information Economics and Policy**, v. 21, n. 3, p. 201-210, 2009.
- ANDERSON, C. The end of theory: The data deluge makes the scientific method obsolete. **Wired magazine**, v. 16, n. 7, p. 16-07, 2008.
- BAGUETE. **Sicredi lança plataforma Woop**. 2018. Disponível em: <https://www.baguete.com.br/noticias/22/01/2018/sicredi-lanca-plataforma-woop>. Acesso em: 01 abr. 2018.
- BALDWIN, H. **When Big Data projects go wrong**. Forbes. Disponível em: <https://www.forbes.com/sites/howardbaldwin/2015/01/22/when-big-data-projects-go-wrong/#617d6e806427>. Acesso em: 01 abr. 2018.
- BARDIN, L. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BAYES, T. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. **Philosophical transactions of the Royal Society of London**, v.1, n. 53, p. 370-418, 1763.

BEAN, R. **Just using big data isn't enough anymore**. Harvard Business Review. 2016. Disponível em: <https://hbr.org/2016/02/just-using-big-data-isnt-enough-anymore>. Acesso em: 30 mar. 2018.

BELL, P. C. "Big Data" and the Analytics Course'. *In: OR/MS Today*, p. 28–31. MD, USA: INFORMS, 2014.

BERNOULLI, D. Exposition of a New Theory on the Measurement of Risk. **Econometrica**, v. 22, n. 1954, p. 23-36, 1738.

BEULKE, D. **Big Data Impacts Data Management: The 5 Vs of Big Data**. 2011. Disponível em: <http://davebeulke.com/big-data-impacts-data-management-the-five-vs-of-big-data/>. Acesso em: 21 abr. 2018.

BEYER, M. A.; LANEY, D. **The importance of 'big data': a definition**. Stamford, CT: Gartner, 2012.

BHOLAT, D. Big data and central banks. **Big Data & Society**, v. 2, n. 1, p. 1-6, 2015.

BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. **Information, communication & society**, v. 15, n. 5, p. 662-679, 2012.

BRYNJOLFSSON, E.; HITT, L. M.; KIM, H. H. Strength in numbers: How does data-driven decisionmaking affect firm performance?. **Available at SSRN 1819486**, v. 1, n. 1, p. 1-33, 2011.

BÜHLMANN, P.; VAN DE GEER, S. **Statistics for high-dimensional data: methods, theory and applications**. New York: Springer Science & Business Media, 2011.

CALLINGHAM, M. **Market intelligence: how and why organizations use market research**. London: Kogan Page Publishers, 2004.

CALLINGHAM, R. Issues for the assessment and measurement of statistical understanding in a technology-rich environment. *In: Data and context in teaching statistics*. Proceedings of Eighth International Conference on Teaching Statistics. Voorburg, Netherlands: International Statistical Institute. 2010.

CALLINGHAM, R. Issues for the assessment and measurement of statistical understanding in a technology-rich environment. *In: Data and context in teaching statistics. Proceedings of Eighth International Conference on Teaching Statistics*. Voorburg, Netherlands: International Statistical Institute. 2010.

CÂMARA, R. H. Análise de conteúdo: da teoria à prática em pesquisas sociais aplicadas às organizações. Gerais: **Revista Interinstitucional de Psicologia**, v. 6, n. 2, p. 179-191, 2013.

CAPELLÁ, J.; HORNE, A.; SHAH, S. **Good Data Won't Guarantee Good Decisions**'. Harvard Business Review, 2012. Disponível em: <http://www.boojazz.com/wp-content/uploads/Good-Data-isnt-Good-Decisions.pdf>. Acesso em: 21 abr. 2018.

CAPGEMINI. **World Payments**. REPORT, 2011. Disponível em: https://www.capgemini.com/br-pt/wp-content/uploads/sites/8/2017/07/World_Payments_Report_2011.pdf. Acesso em: 21 mar. 2018.

CARVALHO, M. L. G.; SOUZA, M. **Categorização/Classificação**. Cadernos Cespuc, nº 23. Universidade Estadual de Montes Claros. 2013

CHAN, A.; PEYNE, B. **Data-Driven decision making in Marketing**: A theoretical approach. Halmstad University, School of Business, Engineering and Science. 2017

CHATFIELD, A.; REDDICK, C.; AL-ZUBAIDI, W. Capability challenges in transforming government through open and big data: Tales of two cities. *In: Proceedings of the Thirty-Sixth International Conference on Information Systems*, 2015.

CLARK, B. Em busca da universidade empreendedora. *In: AUDY, J. L. N.; MOROSINI, M. C. (Org.). Inovação e empreendedorismo na Universidade*. Porto Alegre: Ed. PUCRS., 2006.

COMPUTERWORLD. Cisco projeta expansão intensa dos dados móveis (afetando custos de TI). 2016. Disponível em: <http://computerworld.com.br/cisco-projeta-expansao-intensa-dos-dados-moveis-afetando-custos-de-ti>. Acesso em: 01 abr. 2018.

CRESWELL, J. W. **Investigação Qualitativa e Projeto de Pesquisa**: Escolhendo entre Cinco Abordagens. Porto Alegre: Penso Editora, 3rd edição, 2014.

CROWE, S. *et al.* The case study approach. **BMC Medical Research Methodology**, v. 11, n. 1, p. 19, 2011.

CUKIER, K.; SCHÖNBERGER-MAYER, V. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

CUKIER, K.; SCHÖNBERGER-MAYER, V. **Big data**: como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana. Rio de Janeiro: Elsevier, 2013.

DAVENPORT, T. H. Are you ready to reengineer your decision making. **MIT Sloan Management Review**, v. 51, n. 2, p. 2-6, 2010.

DAVENPORT, T. H. *et al.* Competing on analytics. **Harvard business review**, v. 84, n. 1, p. 98, 2006.

DAVENPORT, T. H.; PATIL, D. J. Data scientist. **Harvard business review**, v. 90, n. 5, p. 70-76, 2012.

DENCKER, A. F. M.; VIÁ, S. C. **Pesquisa empírica em Ciências Humanas**. São Paulo: Futura, 2002.

DENZIN, N. K.; LINCOLN, Y. S. **Collecting and interpreting qualitative materials**. Califórnia: Sage, 2008.

DGINS. **Directors General of the National Statistical Institutes**, Scheveningen–The Hague from 25 to 27 September 2013, Netherlands

ENGLE, R. F.; HENDRY, D. F.; RICHARD, J. F. Exogeneity”: *Econometrica*. **Journal of the Econometric Society**, v. 51, n. 2, p. 277-304, 1983.

FAN, J.; GUO, S.; HAO, N. Variance estimation using refitted cross-validation in ultrahigh dimensional regression. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 74, n. 1, p. 37-65, 2012.

FAN, J.; HAN, F.; LIU, H. Challenges of big data analysis. **National science review**, v. 1, n. 2, p. 293-314, 2014.

FAN, J.; LIAO, Y. **Endogeneity in ultrahigh dimension**. Available at SSRN 2045864, 2012. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2045864. Acesso em: 01 abr. 2018.

FAN, J.; LV, J. Sure independence screening for ultrahigh dimensional feature space. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 70, n. 5, p. 849-911, 2008.

FAN, W.; BIFET, A. Mining big data: current status, and forecast to the future. **ACM SIGKDD Explorations Newsletter**, v. 14, n. 2, p. 1-5, 2013.

FIRICAN, G. **The 10 vs of big data**. Upside where Data means business. 2017. Disponível em: <https://goo.gl/XuJ9M8>. Acesso em: 22 set. 2018.

FISHER, D. *et al.* Interactions with big data analytics. **interactions**, v. 19, n. 3, p. 50-59, 2012.

FLICK, U. **Introdução à pesquisa qualitativa**. 3. ed. Porto Alegre: Artmed, 2009.

FREDERICK, W. T. **The principles of scientific management**. New York: Harper & Brothers, 1911.

FREDERICK, W. T. **The principles of scientific management**. New York: Harper & Brothers, 1911.

FUJISAWA, D. S. **Utilização de jogos e brincadeiras como recurso no atendimento fisioterapêutico de criança**: implicações na formação do fisioterapeuta. 2000. Dissertação (Mestrado em Educação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2000.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International journal of information management**, v. 35, n. 2, p. 137-144, 2015.

GANTZ, J.; REINSEL, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. **IDC iView: IDC Analyze the future**, v. 2007, n. 2012, p. 1-16, 2012.

GAO, J.; KORONIOS, A.; SELLE, S. Towards a process view on critical success factors in big data analytics projects. IN: **Proceedings of the Twenty-First Americas Conference on Information Systems**, Puerto Rico, 2015.

GIBBS, G. R. **Análise de dados qualitativos**. Porto Alegre: Bookman, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5.ed. São Paulo: Atlas, 2010.

GIL, A. C. **Método e técnicas de pesquisa social**. 6 ed. São Paulo: Atlas, 2011.

GILBOA, I. **Rational choice**. London: MIT press, 2010.

GILBOA, I.; ROUZIOU, M.; SIBONY, O. Decision theory made relevant: Between the software and the shrink. **Research in Economics**, v. 72, n. 2, p. 240-250, 2018.

GLASS, R.; CALLAHAN, S. **The Big Data-driven business**: How to use big data to win customers, beat competitors, and boost profits. New Jersey: John Wiley & Sons, 2014.

GOGIA, S. *et al.* The big deal about big data for customer engagement. **Forrester Research**, v. 1, n.1, p. 1-17, 2012.

GHOSH, J. (2016). Big Data Analytics: A Field of Opportunities for Information Systems and Technology Researchers. **Journal of Global Information Technology Management**, 19(4), 217–222. doi:10.1080/1097198x.2016.1249667

GÜNTHER, W. A. *et al.* Debating big data: A literature review on realizing value from big data. **The Journal of Strategic Information Systems**, v. 26, n. 3, p. 191-209, 2017.

HACKING, I. **The Emergence of Probability**. US: Cambridge University Press. 1975.

HAND, D. J. Statistical challenges of administrative and transaction data. **Journal of the Royal Statistical Society: Series A (Statistics in Society)**, 181(3), 555–605. doi:10.1111/rssa.12315, 2018.

HANDCOCK, M. S.; GILE, K. J. On the Concept of Snowball Sampling. **Sociological Methodology**, v. 41, n. 1, p. 367-371, 2011.

HANDCOCK, M. S.; GILE, K. J. On the Concept of Snowball Sampling. **Sociological Methodology**, v. 41, n. 1, p. 367-371, 2011.

HARYADI, A. F. *et al.* Antecedents of big data quality: An empirical examination in financial service organizations. In: **2016 IEEE International Conference on Big Data (Big Data)**. IEEE, 2016. p. 116-121.

HAVENS, T. C. *et al.* Fuzzy c-means algorithms for very large data. **IEEE Transactions on**

HOLMBERG, C. J. G; THOROGOOD N. **Qualitative Methods for Health Research**. London: Sage, 2004.

HUFFMAN, D. *et al.* Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. **New Directions for Evaluation**, v. 2006, n. 109, p. 73-85, 2006.

HUNG, P. C. K. **Big data applications and use cases**. Switzerland: Springer, 2016.

HURWITZ, J. *et al.* **Big Data For Dummies, a Wiley Brand**. New Jersey: John Wiley & Sons, inc., 2013.

IBM. **What is big data**: Bring big data to the enterprise, 2012. Disponível em: <http://www01.ibm.com/software/data/bigdata/>. Acesso em: 01 abr. 2018.

IKEMOTO, G. S.; MARSH, J. A. Cutting through the "Data-Driven" Mantra: Different Conceptions of Data-Driven Decision Making. Reprints. In: MOSS, P. A. (Ed.), **Evidence and decision making: 106th Yearbook of the National Society for the Study of Education**. Malden, MA: Blackwell, 2007.

JACOBS, A. The pathologies of big data, **Communication of the ACM**, v. 52, n. 8, p. 36-44, 2009.

JANSSEN, M.; VAN DER VOORT, H.; WAHYUDI, A. Factors influencing big data decision-making quality. **Journal of Business Research**, v. 70, n. 1, p. 338-345, 2017.

JANSSON, I.; ISAKSSON, A. **Big Data in Official Statistics Production**. Paper for discussion, Advisory Scientific Board, SCB, 2013.

JÄRVENSIVU, T.; TÖRNROOS, J-Å. Case study research with moderate constructionism: Conceptualization and practical illustration. **Industrial Marketing Management**, v. 39, n. 1, p. 100–108, 2010.

JEFFERY, M. **Data-driven marketing**: the 15 metrics everyone in marketing should know. New Jersey: John Wiley & Sons, 2010.

JIA, D. Research on the Integration of Marketing Management and Big Data Technology. **Cyber Security Intelligence and Analytics**, 633–639. doi:10.1007/978-3-030-15235-2_88, 2019.

JIN, X., *et al.* Significance and challenges of Big Data research. **Big Data Research**, v. 2, n. 2, p. 59-64, 2015.

JOHNSON, J. E. Big Data + Big Analytics = Big Opportunity. **Financial Executive**, v. 28, n. 6, p. 50-53, 2012

KAPLAN, A. M.; HAENLEIN, M. Users of the world, unite! The challenges and opportunities of Social Media. **Business horizons**, v. 53, n. 1, p. 59-68, 2010.

KATAL, A; WAZID, M; GOUDAR, R. H. Big data: issues, challenges, tools and good practices. *In: 2013 Sixth international conference on contemporary computing (IC3)*. IEEE, 2013.

KHAN, M. A. U. D.; UDDIN, M. F.; GUPTA, N. Seven V's of Big Data understanding Big Data to extract value. **American Society for engineering education (ASEE Zone 1)**, 2014 zone, v. 1, p. 1-5, 2014.

KIRON, D.; SHOCKLEY, R. Creating business value analytics. **MIT Sloan Management Review**, v. 53, n. 1, p. 57-63, 2011.

KLEIN, G. A. **The power of intuition**: How to use your gut feelings to make better decisions at work. New York: Crown Business, 2004.

KUDYBA, S. **Big data, mining, and analytics**: components of strategic decision making. New York: CRC Press, 2014.

KWON, O.; SIM, J. M. Effects of data set features on the performances of classification algorithms. **Expert Systems with Applications**, v. 40, n. 5, p. 1847-1857, 2013.

LABRINIDIS, A.; JAGADISH, H. V. Challenges and opportunities with big data. **Proceedings of the VLDB Endowment**, v. 5, n. 12, p. 2032-2033, 2012.

LANEY, D. 3D data management: Controlling data volume, velocity and variety. **META group research note**, v. 6, n. 70, p. 1, 2001.

LANSLEY, G.; LONGLEY, P. Deriving age and gender from forenames for consumer analytics. **Journal of Retailing and Consumer Services**, v. 30, n. 1, p. 271-278, 2016.

LAVALLE, S. *et al.* Big data, analytics and the path from insights to value. **MIT sloan management review**, v. 52, n. 2, p. 21-31, 2011.

LEROUX, K.; WRIGHT, N. S. Does performance measurement improve strategic decision making? Findings from a national survey of nonprofit social service agencies. **Nonprofit and Voluntary Sector Quarterly**, v. 39, n. 4, p. 571-587, 2010.

LIM, K. Big data and strategic intelligence. **Intelligence and National Security**, v. 31, n. 4, p. 619-635, 2016.

LIM, M. K.; BAHR, W.; LEUNG, S. C. H. RFID in the warehouse: A literature analysis (1995–2010) of its applications, benefits, challenges and future trends. **International Journal of Production Economics**, v. 145, n. 1, p. 409-430, 2013.

LONG, Q. Data-driven decision making for supply chain networks with agent-based computational experiment. **Knowledge-Based Systems**, v. 141, n. 1, p. 55-66, 2018.

LOSHIN, D. **Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph**. USA: Elsevier, 2013.

LOSHIN, D. **Big data analytics: from strategic planning to enterprise integration with tools, techniques, NoSQL, and graph**. USA: Elsevier, 2013.

LUTZ, M.; BOUCHER, X. Data-driven decision-making for IT capacity: beyond statistical analyses. **Journal of Decision Systems**, v. 26, n. 1, p. 1-24, 2017.

MANYIKA, J. *et al.* **Big data: The next frontier for innovation, competition, and productivity**. USA: McKinsey Global Institute, 2011.

MARTINS, G. A. **Estudo de Caso: Uma estratégia de pesquisa**. 2. ed. São Paulo: Atlas, 2008.

MARZ, N., WARREN, J. **Big Data: Principles and Best Practices of Scalable Real-time Data Systems**. **Manning Publications Co.**, Shelter Island, 2015.

MASON, M. Sample size and saturation in PhD studies using qualitative interviews. *In: Forum qualitative Sozialforschung/Forum: qualitative social research*. 2010.

MAXWELL, N. L.; ROTZ, D.; GARCIA, C. Data and decision making: Same organization, different perceptions; different organizations, different perceptions. **American Journal of Evaluation**, v. 37, n. 4, p. 463-485, 2016.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data: A revolution that will transform how we live, work, and think**. New York: Houghton Mifflin Harcourt, 2013.

MAYRING, P. **Introdução à pesquisa social qualitativa**. 5 ed. Weinheim: Beltz, 2002.

McAFEE, A. *et al.* Big data: the management revolution. **Harvard business review**, v. 90, n. 10, p. 60-68, 2012.

MENZ, M. Functional top management team members: A review, synthesis, and research agenda. **Journal of Management**, v. 38, n.1, p. 45-80, 2012.

MERTENS, D. **Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches**. Thousand Oaks, California: Sage Publications, 1998.

MINELI, M.; CHAMBERS, M.; DHIRAJ A. **Big data, big analytics: emerging business intelligence and analytic trends for today's businesses**. New Jersey: John Wiley & Sons, Inc., 2013.

MORABITO, V. Big data and analytics. **Strategic and organisational impacts**, Switzerland: Springer, 2015.

MORAES, R. Análise de conteúdo. **Revista Educação. Porto Alegre**, v. 22, n. 37, p. 7-32, 1999.

MORSE, J. M. Styles of collaboration in qualitative inquiry. **Qualitative Health Research**, v. 18, n. 1, p. 3-4, 2008.

MORSE, J. M.; FIELD, P. A. **Qualitative Research Methods for Health Professionals** (2nd edition). London: Sage, 1995.

NAMVAR, M.; CYBULSKI, J. BI-based organizations: a sensemaking perspective. *In: Proceedings of the Thirty-Fifth International Conference on Information Systems*, Auckland, New Zealand, 2014.

NNAJI, C. *et al.* Developing a Decision-Making Framework to Select Safety Technologies for Highway Construction. **Journal of Construction Engineering and Management**, v. 144, n. 4, p. 0401-8016, 2018.

NUNAN, D.; DI DOMENICO, M. L. Market research and the ethics of big data. **International Journal of Market Research**, v. 55, n. 4, p. 505-520, 2013.

NUNAN, D.; DI DOMENICO, M. L. Market research and the ethics of big data. **International Journal of Market Research**, v. 55, n. 4, p. 505-520, 2013.

OCDE/Estudos. **Estudos Econômicos do Brasil**. Paris, 2015.

OLBRICH, S. Madness of the crowd—how big data creates emotional markets and what can be done to control behavioural risk. *In: Proceedings of the Twenty-Second European Conference on Information Systems*, Tel Aviv, Israel, 2014.

ORACLE. **Big data for the enterprise**. Oracle White Paper, 1–14, 2012. Disponível em: <http://www.oracle.com/us/products/database/big-data-forenterprise-519135.pdf>. Acesso em: 01 set. 2018.

POLASKY, S. *et al.* Decision-making under great uncertainty: environmental management in an era of global change. **Trends in ecology & evolution**, v. 26, n. 8, p. 398-404, 2011.

PRESKILL, H.; BOYLE, S. A multidisciplinary model of evaluation capacity building. **American journal of evaluation**, v. 29, n. 4, p. 443-459, 2008.

PROMENTILLA, M. A. B. *et al.* A stochastic fuzzy multi-criteria decision-making model for optimal selection of clean technologies. **Journal of cleaner production**, v. 183, n. 1, p. 1289-1299, 2018.

PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big data**, v. 1, n. 1, p. 51-59, 2013.

PWC. **Great expectations: The evolution of the chief data officer**. 2015. Disponível em: <https://www.pwc.com/us/en/financial-services/publications/viewpoints/assets/pwc-chief-data-officer-cdo.pdf>. Acesso em: 02 set. 2018.

RADHIKA, D.; KUMARI, D. A. Adding Big Value to Big Businesses: A Present State of the Art of Big Data, Frameworks and Algorithms. *In: ICT Based Innovations*. Springer, Singapore, 2018.

ROESCH, S. M. A. **Projetos de estágio e de pesquisa em administração: guias para estágios, trabalhos de conclusão, dissertações e estudos de casos**. São Paulo: Atlas, 1999.

ROSE, D. **Data Science: Create Teams That Ask the Right Questions and Deliver Real Value**. New York: Apress, 2016.

ROSE, K. Data on demand: A model to support the routine use of quantitative data for decision-making in Access Services. **Journal of Access Services**, v. 14, n. 4, p. 171-187, 2017.

ROUSE, M. **Big data**, 2011. Disponível em: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>. Acesso em: 11 mar. 2018.

RUSSOM, P. **The three vs of Big Data Analytics**, TDWI, 2011. Disponível em: <http://tdwi.org/blogs/philip-russom/2011/06/three-vs-of-big-data-analytics-3-data-velocity.aspx>. Disponível em: 03 ago. 2018.

SATHI, A. **Big data analytics: disruptive technologies for changing the game**. USA: IBM Corporation, 2012.

SCHINTLER, L. A.; FISCHER, M. M. Big Data and Regional Science: Opportunities, Challenges, and Directions for Future Research. *In: Working Papers in Regional Science, 2018/02*. WU Vienna University of Economics and Business, Vienna, 2018.

SEBEI, H.; TAIEB, M. A. H.; AOUICHA, M. B. Review of social media analytics process and big data pipeline. **Social Network Analysis and Mining**, v. 8, n. 1, p. 30, 2018.

SHAFER, T. **The 42 V's of big data and data science**, 2017. Disponível em: <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>. Acesso em: 02 set. 2018.

SITE DA EMPRESA. **Não autorizado pela empresa**. Acesso em: 21 fev. 2019.

SILVA, A. H.; FOSSÁ, M. I. T. Análise de conteúdo: Exemplo de Aplicação da técnica para análise de dados qualitativos. **Qualit@s Revista Eletrônica**, v. 17, n. 1, p. 1-14, 2015.

SIMON, H. A. **Administrative behavior: A Study of Decision-making Processes in Administrative Organization**, 1st ed. New York: The Macmillan Company, 1947.

SIMON, I. **Too big to ignore: the business case for big data**. New Jersey: John Wiley & Sons, 2013.

SPIILLECKE, D.; UMBLIJS, A. **Making data-Driven marketing decisions**, 2013. Disponível em:

https://www.mckinsey.com/~/media/McKinsey/Industries/Retail/Our%20Insights/Making%20data%20driven%20marketing%20decisions/Making_data_driven_marketing_decisions.ashx. Acesso em: 20 out. 2018.

STÄDLER, N.; BÜHLMANN, P.; VAN DE GEER, S. ℓ_1 -penalization for mixture regression models. **Test**, v. 19, n. 2, p. 209-256, 2010.

STAKE, R. E. **Pesquisa qualitativa: estudando como as coisas funcionam**. Porto Alegre: Penso, 2011.

STAKE, R. E. **The art of case study research**. Londres: Sage Publications, 1995.

TAMM, T.; SEDDON, P.; SHANKS, G. Pathways to value from business analytics. *In: Proceedings of the Thirty-Fourth International Conference on Information Systems*, Milan, Italy, 2013.

TÖRNGREN, O. **Mergers in big data-Driven markets: is the dimension of privacy and protection of personal data something to consider in the merger review?**. 2018, 73 f. Dissertação (Mestrado em Direito) - Stockholm University. 2018. Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-153509>. Acesso em: 03 fev. 2019.

TURNER, V. *et al.* The digital universe of opportunities: Rich data and the increasing value of the internet of things. **IDC Analyze the Future**, v. 16, 2014.

UN GLOBAL PULSE. **Big Data for Development: Challenges & Opportunities**, 2012. Disponível em: <http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf>. Acesso em: 01 out. 2018.

VERGARA, S. C. **Projetos e relatórios de pesquisas em administração**. 15. ed. São Paulo: Atlas, 2014.

VINUTO, J. A amostragem em bola de neve na pesquisa qualitativa: um debate em aberto. **Temáticas**, v. 22, n. 44, p. 203-220, 2016.

WALKER, R. **From big data to big profits: Success with data and analytics**. New York: Oxford University Press, 2015.

WAMBA, S. F. *et al.* How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. **International Journal of Production Economics**, v. 165, n. 1, p. 234-246, 2015.

WAMBA, S. F. *et al.* How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. **International Journal of Production Economics**, v. 165, n. 1, p. 234-246, 2015.

WEF - World Economic Forum. **Big Data, Big Impact: New Possibilities for International Development**, 2012. Disponível em: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf. Acesso em: 03 mar. 2018.

WEIGERT, T. **Data-Driven Decision Making: An Adoption Framework**. 2017. 43 f. Tese (Doutorado em Administração) - Massachusetts Institute of Technology - School of Management. 2017.

XU, Z. B. *et al.* The frontiers in big-data-driven management and decision making. **Manage. World China**, v. 11, n.1, p. 158-163, 2014.

YIN, R. K. **Estudo de Caso: Planejamento e Métodos**. 5 ed. Porto Alegre: Bookman, 2015.

ZHAO, D. Frontiers of big data business analytics: patterns and cases in online marketing. *In*: LIEBOWITZ, J. **Big data and business analytics**. Boca Raton: Taylor & Francis Group, LLC, 2013.

APÊNDICE A – ROTEIRO APLICADO NAS ENTREVISTAS

Instrumento de Pesquisa

1) Fenômeno Big Data e suas dimensões (6 questões)

As perguntas foram orientadas por duas matrizes teóricas: 5 V's do Big Data (RUSSOM, 2011; BEULKE, 2011 E WAMBA *et al.*, 2015) com os desafios inerentes ao fenômeno, propostos por Fan, Han e Liu (2014).

Os objetivos dessa seção são: validar que a empresa pesquisada utiliza-se de dados caracterizados como Big Data em seu cotidiano, e identificar como cada lidam com os desafios concernentes a cada dimensão do fenômeno.

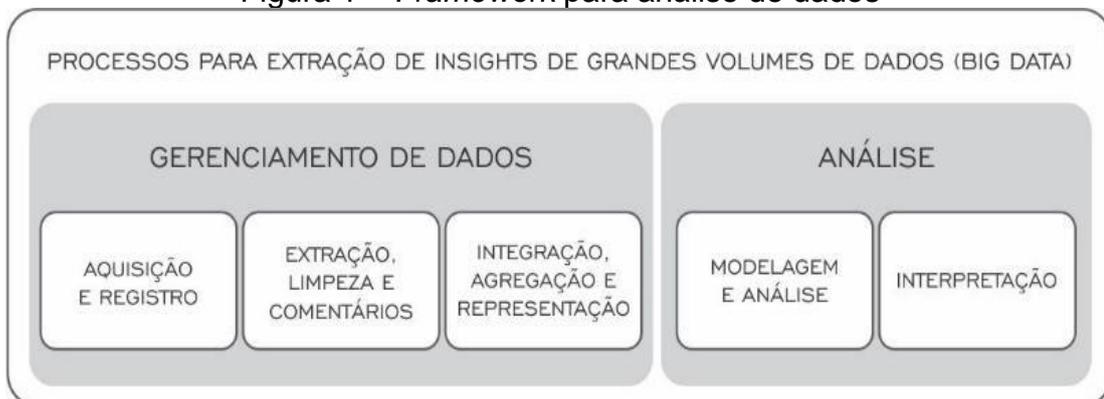
QUESTÕES	RELAÇÃO TEÓRICA
Sua empresa lida com grandes volumes de dados, de diferentes fontes, captados em alta velocidade para as diferentes tomadas de decisão?	<i>“Portanto, definimos “Big Data” para este estudo como uma abordagem holística para gerenciar, processar e analisar os 5 V’s (ou seja, volume, variedade, velocidade, veracidade e valor), a fim de criar informações acionáveis para a entrega de valor sustentado, medição de desempenho, estabelecendo vantagens competitivas (GARTNER, 2012; WAMBA et al., 2015).”</i>
Como sua empresa lida com o alto volume de dados? Quais procedimentos são adotados para evitar ruídos (dados falsos, vieses, lacunas de dados)?	<i>“Grande volume de dados consumindo enorme armazenamento, ou consistindo de grande número de registros (Russom, 2011)” e “Geralmente a endogeneidade ocorre como resultado da seleção de vieses, erros de medição e variáveis omitidas (FAN; HAN; LIU, 2014).”</i>
Como sua empresa lida com a variedade de dados, vindos de diferentes fontes? Quais procedimentos adotados para unificar ou aglutinar os dados em um formato interpretável? Como os dados são limpos?	<i>“Dados gerados a partir de maior variedade de fontes e formatos, contendo dados multidimensionais (RUSSOM, 2011). “ e “Fan, Han e Liu (2014) explica que dados em Big Data são tipicamente compostos por meio da aglutinação de muitas fontes de dados, correspondentes a diferentes subpopulações.”</i>
Como sua empresa lida com a velocidade dos dados? Já houve algum caso de erro de interpretação em função da escolha do período de tempo selecionado para análise?	<i>“Frequência de geração e entrega de dados (RUSSOM, 2011)” e “Análises em Big Data são caracterizadas por cálculos simultâneos, ou testes de muitos parâmetros. Erros de estimativas podem se acumular quando uma regra de decisão ou previsão depende de um grande número desses parâmetros (FAN; HAN; LIU, 2014).”</i>

<p>Existem procedimentos adotados para aferir a veracidade aos dados em análise? Como ele são colocados em prática? Quais dificuldades aparecem?</p>	<p><i>“A necessidade de alguns dados requererem análise para ganhar confiabilidade (BEULKE, 2011)” e “Um processo de análise que busque a verdade nos dados pressupõe precisão estatística. Fan, Han e Liu (2014)”</i></p>
<p>Como sua empresa confere valor aos dados? Existe algum procedimento para extrair insights dos dados? Como isso se dá?</p>	<p><i>“A medida de ganho econômico gerados pelos dados, através da extração da informação (WAMBA et al., 2015)” e “Eles estão implementando programas de análise de dados para coletar, armazenar, gerenciar e analisar grandes conjuntos de dados de uma gama de fontes para identificar os principais insights de negócios que podem ser explorados para apoiar uma melhor tomada de decisão (FAN; HAN; LIU, 2014).”</i></p>

2) Tomada de decisão baseada em dados: processo, insights e desafios (6 questões)

As perguntas foram orientadas por duas matrizes teóricas: o Framework para análise de dados de Gandomi e Haider (2015), e os desafios à tomada de decisão baseada em dados propostos por Chan e Peye (2017)

Figura 1 – *Framework* para análise de dados



Fonte: Adaptado de Gandomi, Haider (2015)

Os objetivos dessa seção são: identificar qual o processo utilizado para análise de dados, e como os desafios na tomada de decisão baseada em dados são superados, e insights são extraídos nesse processo – além de identificar potenciais novos desafios não registrados pela literatura pesquisada.

QUESTÕES	RELAÇÃO TEÓRICA
<p>Como é o processo de tomada de decisão no seu setor? Você diria que a decisão tomada geralmente é baseada em dados e fatos apurados?</p> <p>Quais os principais desafios na tomada de decisão? Sua equipe está preparada para lidar com o alto volume de dados? Como isso tem evoluído?</p> <p>Caso negativo, por que não há esse preparo? Caso positivo, como é esse preparo?</p>	<p><i>“Porém, o uso de dados está mudando completamente o processo de tomada de decisão nas empresas, configurando-se em um ativo para a melhora no desempenho organizacional (CHAN, PEYNE, 2017).” e “A tomada de decisão baseada em dados ajuda as empresas a evitar desperdícios de tempo e dinheiro (CHAN, PEYNE, 2017).” e “Isso destaca o fato de que as equipes presentes nas organizações ainda não estão preparadas para lidar com grandes volumes de dados, o que dificulta o uso de Big Data nos processos de decisão. Especialistas na análise de Big Data para auxiliar a tomada de decisão ainda são raros, e ativamente procurados pelas empresas (CHAN, PEYNE, 2017)”</i></p>
<p>AQUISIÇÃO: Como é feita a mineração de dados na empresa? Como as fontes de dados são escolhidas?</p>	<p><i>“Considerando que uma alta quantidade de dados é de fácil obtenção pelas empresas, é necessária a obtenção de dados corretos, uma vez que a qualidade não depende apenas dos dados, mas também do próprio processo de coleta dos dados (CHAN, PEYNE, 2017).”</i></p>
<p>EXTRAÇÃO: Uma vez que os dados são coletados, eles são divididos para análise? Como se decide o que analisar e o que não analisar?</p>	<p><i>“Nesse sentido Chan e Payne (2017) afirmam que o principal desafio de usar os dados não está na coleta, mas na escolha dos dados que devem ser utilizados e como visualizá-los.”</i></p>
<p>INTEGRAÇÃO: Como é feita a análise de dados? É realizada por uma equipe? Quais as características dessa equipe – tamanho, perfil funcional?</p>	<p><i>“As empresas cada vez mais procuram equipes flexíveis e multifuncionais, a fim de para expandir as habilidades de análise para todos os empregados, sem restringi-las a um pequeno grupo (CHAN, PEYNE, 2017).” E “Especialistas na análise de Big Data para auxiliar a tomada de decisão ainda são raros, e ativamente procurados pelas empresas (CHAN, PEYNE, 2017).”</i></p>
<p>ANÁLISE: Como os cenários futuros são formulados para tomada de decisão? Como é a relação entre projetar o futuro através de dados passados?</p>	<p><i>“Segundo Chan e Payne (2017) no passado os dados revelavam fatos apenas através de algum tipo de relatório ou feedback. No entanto, as tecnologias de dados evoluíram, permitem as organizações seu uso para obter dados em tempo real, e até mesmo para o futuro através de extrapolações.”</i></p>
<p>INTERPRETAÇÃO: Qual é o papel da experiência e da intuição na tomada de decisão frente aos dados? Como os insights são gerados?</p>	<p><i>“Chan e Payne (2017) chamam a atenção para o fato de que mesmo em empresas que usam dados de forma avançada, a dimensão humana da intuição ainda desempenha um papel decisivo no processo decisório.” e “A criatividade e a intuição lidam com o aspecto humano que o Big Data não consegue ainda entender, como desejo e a emoção (CHAN, PEYNE, 2017).”</i></p>

ANEXO A – MAPA ESTRATÉGICO DE COOPERATIVA SINGULAR PESQUISADA

