

# Big Data Analytics in Supply Chain Management: A Systematic Mapping Study

Thiago Vieira de Souza  
Programa de Pós-Graduação  
em Computação Aplicada  
(PPGCA), UNISINOS  
Avenida Unisinos, 950  
São Leopoldo, RS  
thiago-vs@hotmail.com

Kleinner Farias  
Programa de Pós-Graduação  
em Computação Aplicada  
(PPGCA), UNISINOS  
Avenida Unisinos, 950  
São Leopoldo, RS  
kleinnerfarias@unisinos.br

Vinicius Bischoff  
Programa de Pós-Graduação  
em Computação Aplicada  
(PPGCA), UNISINOS  
Avenida Unisinos, 950  
São Leopoldo, RS  
viniciusbischof@edu.unisinos.br

## ABSTRACT

In recent years, the capacity of big data analytics (BDA) has attracted the significant attention of researchers linked to academia and industry professionals. This capacity is related to the possibility of managing information advanced to reach its supply chain. In other words, information technology uses integrated systems, which facilitates innovation and the diffusion of knowledge throughout this supply chain. However, researchers and professionals still need to explore the capacity potential of BDA, in order to improve supply chain operational decision-making skills. This work classifies the state-of-the-art literature that applied BDA to the supply chain management (SCM). A Systematic Mapping Study was elaborated based on literature guidelines. A total of 50 primary studies were selected through a filtering process from initially 5,437 studies. These primary studies were used to answer the six research questions. The result of the classification showed that 64% of the studies are related to supply-chain management; most of the studies carried out empirical research; and approximately 50% of the primary studies investigated models for optimization process. This research provides to academics and industry professionals the gaps and future challenges related to BDA for SCM.

## Keywords

Big Data Analytics; Supply-Chain; Data Analysis; Information Systems

## 1. INTRODUÇÃO

Cada vez mais as aplicações corporativas são alteradas e evoluem para dar suporte a novos requisitos [24], bem como precisam manipular um grande volume de dados. Neste contexto, o avanço da tecnologia está transformando a maneira de como os dados são gerados e analisados. De acordo com Wamba et al. [30], *Big Data* tem o potencial de revolucionar

as decisões operacionais da cadeia de suprimentos ou *Supply Chain Management (SCM)*. *SCM* refere-se ao gerenciamento do fluxo total de produtos a partir do fornecedor até o cliente final, procurando otimizar sua logística. Uma das formas de auxiliar na otimização da logística seria através do uso de sistemas de informações que gerenciam o fluxo dos dados da cadeia de suprimentos. A consequência é a geração de uma grande quantidade de dados denominado de *Big Data*. A aplicação de *Big Data Analytics (BDA)* nesses dados viabiliza a extração de informações cruciais para a tomada de decisão especificamente para otimizar a cadeia [10]. A aplicação de *BDA* em *SCM* é algo que está em prática nesta última década [17], com uma variada gama de métodos e ferramentas aplicadas nesse contexto. Dessa forma, com o surgimento frequente de desafios em relação a esse tema, a comunidade acadêmica e a indústria despertaram o interesse em resolver potenciais desafios de pesquisa ligadas à esta área [28, 18]. Ou seja, pesquisadores e profissionais necessitam de uma revisão ampla dos estudos primários existentes nesta área, procurando assim identificar evidências sobre a aplicação de *BDA* em conjunto com a cadeia de suprimentos. Através destas evidências será possível promover uma visão geral da área investigada e uma classificação das pesquisas realizadas na área, auxiliando na identificação de possíveis lacunas.

Porém, faltam estudos que forneçam uma visão geral sobre a aplicação de técnicas *BDA* no gerenciamento da cadeia de suprimentos. Hazen et al. [18] revelam a falta de revisões da literatura a partir de uma perspectiva geral de cadeia de suprimentos, e ao mesmo tempo relacionando-se com as técnicas de *BDA*. Além disso, as revisões focaram em analisar aspectos relacionados aos ambientes operacionais, tais como linhas de produção de fábricas e o desenvolvimento de produtos. Por exemplo, Babiceanu Seker [1] realizaram uma revisão da literatura sobre fluxo de material em operações de manufatura, enquanto que Wamba et al. [30] focaram nas investigações de aplicações logísticas. Ainda assim, estes trabalhos não consideraram realizar uma classificação das técnicas de *BDA* relacionadas à cadeia de suprimentos. Outro fator a ser considerado nos trabalhos [1, 30, 18], refere-se a não aplicação de uma metodologia criteriosa e sistemática para auxiliar na seleção destes estudos, impactando na validade do processo de seleção.

Assim, este trabalho procura classificar a literatura e realizar uma análise temática sobre a aplicação de *BDA* no *SCM*, auxiliando o controle operacional logístico, através de

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2020 May 18<sup>th</sup> – 22<sup>th</sup>, 2020, São Bernardo do Campo, São Paulo, Brazil

Copyright SBC 2020.

um mapeamento sistemático. Este método utiliza diretrizes bem definidas na literatura [19, 25], bem como mapeamentos sistemáticos previamente publicados [2, 15, 16]. Deste modo, a busca pelos estudos primários seguiu uma metodologia criteriosa e sistemática em sua condução. Inicialmente 5.437 artigos foram selecionados a partir de oito fontes de pesquisa (bibliotecas digitais), os quais passaram por um criterioso processo de seleção por pares de avaliação. Resultando em uma lista final de 50 artigos relevantes a responder a quatro questões de pesquisa investigadas. Os artigos foram classificados e analisados em relação a área da cadeia de suprimentos, níveis de *BDA*, modelos de *BDA*, técnicas de *BDA* utilizadas, métodos de pesquisa, e locais de publicação.

Os principais resultados apontam que 64% das técnicas de *BDA* são aplicadas na área de *SCM*; 28% das técnicas de *BDA* são preditivas; 32% dos estudos abordam sobre otimização ao aplicar *BDA*, i.e., focam em otimizar processos, custo e tempo; 28% das técnicas de *BDA* são mistas, i.e., as técnicas de *BDA* são compostas por duas ou mais estruturas; 34% dos artigos são estudos de avaliação; 53% dos estudos selecionados foram publicados em conferências.

O artigo está organizado da seguinte forma. A seção 2 apresenta os conceitos básicos em relação a *BDA* e *SCM*. A seção 3 descreve a metodologia deste trabalho. A seção 4 apresenta os resultados obtidos do estudo de mapeamento sistemático. A seção 5 discute alguns desafios e direções futuras em relação aos resultados encontrados. A seção 6 descreve os trabalhos que são relacionados com este mapeamento sistemático. Por fim, a seção 8 apresenta as considerações finais e desafios futuros.

## 2. FUNDAMENTAÇÃO TEÓRICA

Esta seção aborda os conceitos principais do trabalho proposto, incluindo gestão da cadeia de suprimentos, análise de dados, e os conceitos do estudo de mapeamento sistemático.

### 2.1 Gerenciamento da Cadeia de Suprimentos

Conforme exposto por Cooper et al. [7], o termo cadeia de suprimentos é mais comum na literatura que a definição do termo *supply chain management (SCM)*. Normalmente, várias empresas independentes estão envolvidas na fabricação de um produto e para colocá-lo nas mãos do usuário final, formam uma cadeia de suprimentos. Do mesmo modo, Ellram e Coope [9] e Cooper et al. [7] definiram que uma cadeia de suprimentos é o alinhamento das empresas que trazem produtos ou serviços para o mercado. Estas definições podem ser visualizadas na Figura 1.

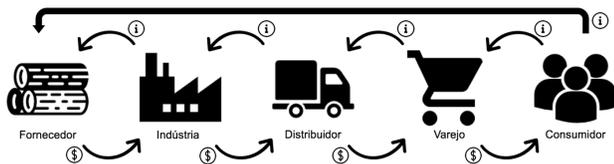


Figura 1: Fluxo da cadeia de suprimentos.

De acordo com Mentzer et al. [23], as definições de *SCM* podem ser classificadas em três categorias: Filosofia de Gerenciamento, Implementação de uma Filosofia de Gerenciamento, e Conjunto de Processos de Gerenciamento. O *SCM* adota uma abordagem de sistemas para visualizar a cadeia

de suprimentos como uma entidade única, e não como um conjunto de partes fragmentadas, cada uma executando sua própria função [5]. Em outras palavras, a gestão da cadeia de suprimentos (ou *SCM*) amplia o conceito de parceria em um esforço incomum para gerenciar o fluxo total de produtos do fornecedor para o cliente final.

Bowersox et al. [3] ressaltaram que as empresas devem estabelecer práticas de gestão que lhes permitam realizar a organização de suas operações resultantes de modo consistente. Os autores também ressaltaram que as empresas devem expandir seu comportamento integrado para incorporar clientes e fornecedores no intuito de obter a máxima eficácia de competitividade. Esta extensão de comportamentos integrados é designada como *SCM* [3]. Neste contexto, o gerenciamento da cadeia se transforma em um conjunto de atividades que precisam ser monitoradas. Além disso, o *SCM* integra um conjunto de atividades realizadas através de um esforço coordenado entre os parceiros desta cadeia de suprimentos para atender dinamicamente as necessidades do consumidor final [21].

### 2.2 Análise de Dados para Extração de Informações

Conforme Wamba et al [30], *Big Data* é definido através de cinco elementos: volume, velocidade, variedade, veracidade e valor. A veracidade e valor representam respectivamente os elementos relacionados a confiabilidade e importância dos dados, os quais são elementos mais relevantes para *Big Data* [1]. Esta relevância é devida a sua capacidade de analisar e extrair dados de alto valor. O conceito de *Big Data Analytics (BDA)* nas cadeias de suprimento se resume na utilização de técnicas de análise para extração de informações relevantes a partir de grandes massas de dados para melhorar a capacidade de auxílio na tomada de decisão [29].

De acordo com Labrinidis e Jagadish [20], para permitir uma tomada de decisão baseada em evidências, as organizações precisam de processos eficientes para transformar grandes volumes de dados em informações diversificadas. O processo geral (Figura 2) para extração de informações de *Big Data* pode ser dividido em cinco etapas: (1) aquisição e gravação; (2) extração, limpeza e anotação; (3) integração, agregação e representação; (4) modelagem e análise; e (5) interpretação. Estas etapas formam dois subgrupos principais, o gerenciamento de dados e a análise [14]. O gerenciamento de dados envolve os processos e tecnologias de suporte para adquirir e armazenar dados, prepará-los e recuperá-los para sua interpretação. O subgrupo de análise refere-se as técnicas utilizadas para investigar e adquirir conhecimento a partir de dados produzidos.

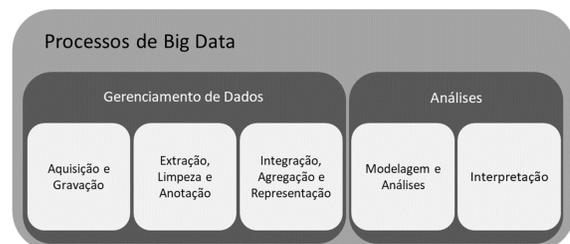


Figura 2: Processos de *Big Data*.

## 2.3 Estudo de Mapeamento Sistemático (SMS)

A metodologia conduzida à aplicação SMS exploram uma área de pesquisa mais ampla e classificam os estudos que são representativos em um assunto específico [19]. O mapeamento sistemático investiga questões genéricas de pesquisa [22, 27, 2], tais como: os tópicos de pesquisa mais abordados; os métodos empíricos frequentemente utilizados; quais técnicas foram amplamente adotadas; quais pesquisadores têm realizado estudos em particular; que tipo de pesquisa foi realizada; em que estudos contextuais foram realizados; e quais os problemas de pesquisa que têm estudos empíricos suficientes para apoiar uma revisão sistemática mais detalhada.

O estudo de mapeamento sistemático tende a ter um amplo alcance em termos de escopo [25]. Mesmo que uma ampla gama de artigos relacionados a uma área de tópicos seja considerada em um mapeamento sistemático, geralmente há um foco maior na síntese dos dados de classificação em relação a esses estudos. Os pesquisadores geralmente elaboram essas análises em relação a classificação realizada para promover uma visão quantitativa do estado da arte, e assim, encontrar novas lacunas de pesquisa, e identificar os pontos da área de pesquisa que estão mais evoluídos [31].

## 3. METODOLOGIA

Esta seção tem como objetivo apresentar a metodologia de estudo de mapeamento sistemático. A seção 3.1 apresenta a definição dos objetivos de pesquisa, e das questões de pesquisa que direcionam essa classificação. A seção 3.2 apresenta as estratégias de busca. A seção 3.3 define os critérios de inclusão e exclusão. A seção 3.4 descreve os passos do processo de seleção dos artigos. Por fim, a seção 3.5 apresenta a execução do processo de seleção.

### 3.1 Objetivos e questões de pesquisa

Os dois principais objetivos deste trabalho são: (1) fornecer uma classificação da literatura contemporânea sobre a utilização de *BDA* nos processos da cadeia de suprimentos; e (2) identificar os processos da cadeia de suprimentos onde *BDA* tem sido aplicado com maior frequência. Para explorar várias facetas deste objetivo, quatro questões de pesquisa (QP) foram elaboradas.

As questões de pesquisa nos estudos de mapeamento devem ser genéricas, para que se descubra os futuros direcionamentos destas e os tópicos investigados na literatura [25]. A Tabela 1 resume as questões de pesquisa investigadas neste trabalho.

Tabela 1: Questões de Pesquisa SMS

Questões de Pesquisa	Motivação	Variável
QP1: Em quais áreas do gerenciamento da cadeia de suprimentos é aplicado <i>Big Data Analytics</i> ?	Descobrir em quais áreas do gerenciamento da cadeia de suprimentos as técnicas de <i>Big Data Analytics</i> são aplicadas.	Áreas da cadeia de suprimentos
QP2: Em que nível de análise <i>Big Data Analytics</i> é utilizado nessas áreas do gerenciamento da cadeia de suprimentos?	Listar os principais níveis de <i>Big Data Analytics</i> utilizados para o gerenciamento da cadeia de suprimentos.	Níveis de <i>Big Data Analytics</i>
QP3: Que tipos de modelos de <i>Big Data Analytics</i> são usados no gerenciamento da cadeia de suprimentos?	Classificar os modelos de <i>Big Data Analytics</i> utilizados para a cadeia de suprimentos.	Modelos de <i>Big Data</i>
QP4: Quais técnicas são utilizadas para desenvolver os modelos de <i>Big Data Analytics</i> ?	Listar as principais técnicas utilizadas nos modelos de <i>Big Data Analytics</i> .	Técnicas de <i>Big Data</i>

## 3.2 Estratégia de Busca

Esta seção descreve os critérios definidos para realizar o método de busca pelos estudos primários. O método utilizado é procedente de uma combinação de busca automática em bases de dados digitais, e busca manual. Primeiramente realizou-se uma busca manual por palavras-chave para construção do termo de busca. Este conjunto de palavras procede de grupos de pesquisa e autores que tem atuado na área em questão. A formação do termo de busca considerou publicações relevantes da área (de acordo com o Qualis/CAPEs, A1), bem como contou com uma revisão de pesquisadores e profissionais da área investigada sobre *BDA* e *SCM*, com os seus respectivos sinônimos. Observa-se que o termo de busca deve ser adaptado para funcionar adequadamente em cada base de dados investigada [31]. Assim, identificou-se um conjunto de palavras-chave relacionadas a literatura, para a formulação do termo de busca ilustrado nesta pesquisa como referência inicial.

((“*Supply Chain*” OR *Manufacturing* OR “*Order Picking*” OR *Logistics*) AND (“*Big Data Analytics*” OR “*Predictive Analytics*”))

O termo de busca foi aplicado nas seguintes bases de dados digitais: (1) *ACM Digital Library*, (2) *Google Scholar*, (3) *IEEE Explore*, (4) *Inspec*, (5) *Science Direct*, (6) *Scopus*, (7) *Springer Link*, e (8) *Wiley Online Library*. Procurou-se obter uma cobertura abrangente dos trabalhos publicados no maior número de bases digitais. Entretanto, o comportamento dos mecanismos de busca é alterado por atualizações [2]. Assim, ressalta-se que a maior parte das bases mencionadas não são projetadas para auxiliar eficientemente a realização *SMS*.

### 3.3 Critérios de inclusão e exclusão

Esta seção tem como objetivo estabelecer critérios de exclusão e inclusão usados para filtrar os estudos recuperados das bases de dados digitais selecionadas.

A lista a seguir especifica os Critérios de Exclusão (CE) definidos:

- **CE 1:** publicações que não mencionem as palavras-chave da pesquisa no título, resumo ou palavras-chave do artigo;
- **CE 2:** artigos que relatam apenas um resumo, teleconferências ou patentes;
- **CE 3:** estudos corresponde às palavras-chave definidas no termo de busca, mas o contexto é diferente dos objetivos da pesquisa,
- **CE 4:** não estavam dentro do contexto do domínio de *ABD* e *SCM*;
- **CE 5:** estudos duplicados, será considerada a versão atual; e
- **CE 6:** não atenda à motivação das questões de pesquisa descritas na seção 3.1.

Por outro lado, os Critérios de Inclusão (CI) foram usados para selecionar os trabalhos inerentes a esta pesquisa.

- **CI 1:** estudos relacionados ao termo de busca descritos na seção 3.2 e na seção 3.1;

- **CI 2:** trabalhos escritos em inglês;
- **CI 3:** artigos disponíveis em bibliotecas digitais eletrônicas;
- **CI 4:** estudos publicados entre janeiro/2010 e dezembro/2017.

### 3.4 Procedimentos de Seleção

Esta seção apresenta o processo de execução dos filtros dos estudos selecionados, descrevendo os passos para alcançar os estudos primários. Aplicaram-se oito etapas para filtrar os estudos relevantes depois de realizar a pesquisa inicial. A busca e filtragem foram executadas entre janeiro e março de 2017. A Figura 3 ilustra as etapas do processo de seleção dos estudos relevantes, bem como da execução dos filtros, que são descritos da seguinte forma:

- **Etapa 1: busca inicial.** Coletar os resultados iniciais das pesquisas logo após a utilização do termo de busca nas bases de dados digitais selecionadas.
- **Etapa 2: filtro inicial.** Remover discrepâncias obtidas nos resultados da pesquisa. Para isso, os critérios de exclusão CE 1 e CE 2 foram aplicados.
- **Etapa 3: filtro por título e resumo.** Filtrar estudos aplicando o critério de exclusão CE3. Consiste na remoção dos estudos que apresentam-se fora do contexto do objetivo de pesquisa.
- **Etapa 4: filtro por texto.** Filtrar estudos considerando o critério de exclusão CE4. Remove os estudos que não fazem parte do domínio especificado.
- **Etapa 5: combinação.** Agrupa-se todos os estudos selecionados em um único local. Remove-se os estudos selecionados dos diretórios (bases digitais).
- **Etapa 6: remoção de artigos duplicados.** Consiste na aplicação do filtro CE5, permanecendo apenas a versão mais atual dos artigos.
- **Etapa 7: filtro baseado no texto completo.** Os estudos são filtrados aplicando o CE6. Realiza-se uma leitura completa dos textos selecionados procurando analisar se estes respondem as questões de pesquisas relacionadas neste estudo.
- **Etapa 8: trabalhos representativos.** A lista final dos estudos primários é definida após revisar todas as etapas anteriores e os critérios de inclusão e exclusão.

### 3.5 Filtragem

Esta seção apresenta os resultados de cada uma das oito etapas da execução do processo de filtragem. A Figura 3, ilustra o resultado obtido em cada etapa do processo de seleção.

A busca inicial dos artigos resultou em 5.437 artigos, em seguida, aplicou-se o filtro inicial, que resultou 2,55% (139 de 5.437) dos artigos foram removidos após aplicação dos critérios de exclusão CE 1 e CE 2. Na etapa 3, os artigos foram filtrados pelo título e resumo conseqüentemente 91,77% (4.862 de 5.298) dos artigos foram removidos aplicando o critério CE 3. Em seguida 63,3% (276 de 436) dos artigos foram descartados após uma revisão superficial do texto aplicando o critério CE 4.

Os estudos restantes (160 artigos) foram combinados em um único diretório, e esta combinação gerou uma amostra de 134 artigos, ou seja, 16,25% dos artigos foram filtrados nesta etapa. A próxima etapa consistiu na remoção de artigos duplicados, e após aplicação do critério de exclusão CE 5, 28,35% (38 de 134) artigos foram removidos. Em seguida foi aplicado o CE 6, filtragem por texto completo, o qual descartou 26,04% (25 de 96) artigos.

Ao examinar os 71 trabalhos restantes, observou-se algumas situações em que os assuntos foram abordados de forma similar, ou seja, trabalhos produzidos com bases nos artigos anteriores, dessa forma 29,57% (21 de 50) dos artigos foram descartados. Por fim 50 artigos foram selecionados como os mais representativos para este estudo. A relação de estudos faz parte do material de estudos do grupo de pesquisa, e está disponível no Google Docs<sup>1</sup> os quais são atribuídos pelos mesmos identificadores deste estudo, i.e., entre A1 até A50.

## 4. RESULTADOS

Esta seção apresenta os resultados em relação as questões de pesquisa definidas na metodologia desse mapeamento sistemático. Ao total, quatro questões de pesquisa são investigadas neste mapeamento sistemático. Assim, aborda-se cada questão de pesquisa conforme sua descrição na seção 3.1.

### 4.1 QP1: Áreas da cadeia de suprimentos

Esta questão investiga em quais áreas da cadeia de suprimentos *Big Data Analytics* foi aplicada com maior frequência. Os principais resultados são demonstrados na Tabela 2.

Cerca de 78% (39/50) dos artigos selecionados estão concentrados nas áreas de gerenciamento da cadeia e de demanda. Estes resultados indicam as áreas onde a gestão e as tomadas de decisão estão constantemente sendo exercidas. Processos de *Big Data Analytics* são aplicados para trazer visibilidade e informação, com o intuito de auxiliar as tomadas de decisão destas áreas. Além disso, o entendimento das necessidades de mercado tem sido um diferencial de muito valor. Dominar esta variável implica em ter um controle eficaz da produção, evitando grandes produções sem demanda, o que reduz grandes números de estoques e gastos com matérias primas desnecessárias. O contrário também se aplica nesses casos, isto é, pouca produção com muita demanda de mercado. Em seguida, 22% (11/50) dos artigos avaliados se concentram nas áreas de manufatura, transporte/logística e armazenamento/depósito. Estas áreas estão ligadas diretamente a produção, e não a gestão. São áreas que focam na otimização, seja por redução de tempo, custo, ou matéria prima.

A partir dos resultados, é evidente que *Big Data Analytics* está sendo aplicado para áreas de gestão e controle. Possuir informações consistentes nessas áreas indica a busca pelo suporte na tomada de decisão para obter vantagem competitiva. Além disso, essas áreas seriam beneficiadas ao serem combinadas com aplicações de *Big Data Analytics*, pois vários processos poderiam ser otimizados, agregando valor ao produto.

### 4.2 QP2: Níveis de Big Data Analytics

Esta questão de pesquisa investiga quais níveis os conceitos de *Big Data Analytics* estão sendo aplicados com maior

<sup>1</sup><http://tiny.cc/l8iejz>

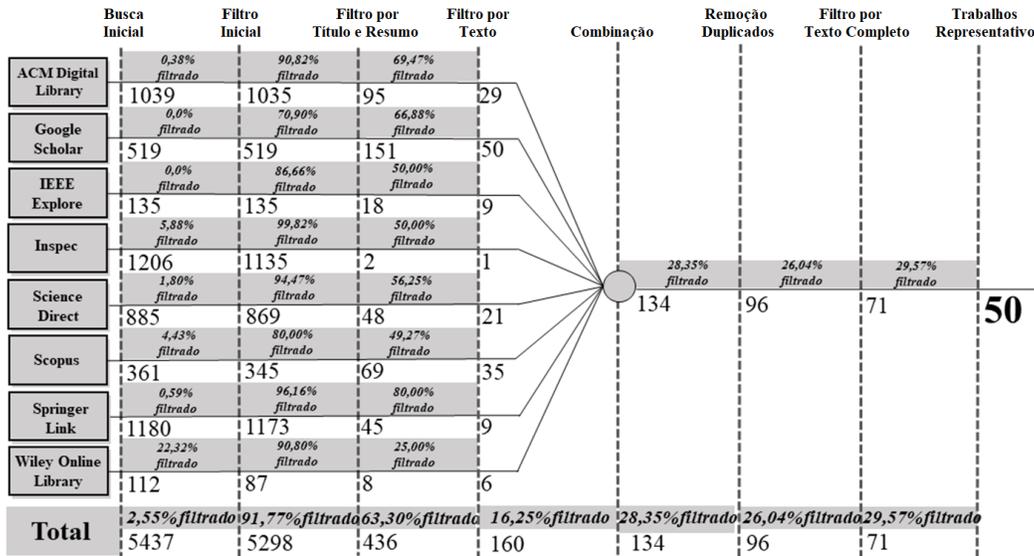


Figura 3: Processos de seleção dos estudos primários.

Tabela 2: Áreas da cadeia de suprimentos vinculadas aos estudos selecionados

Áreas da Cadeia de Suprimentos	Número de Estudos	Percentual	Lista de Artigos
Gerenciamento da Cadeia	32	64%	[A1], [A2], [A3], [A4], [A5], [A6], [A7], [A8], [A9], [A10], [A11], [A12], [A13], [A14], [A15], [A16], [A17], [A18], [A19], [A20], [A21], [A22], [A23], [A24], [A25], [A26], [A27], [A28], [A29], [A30], [A31], [A32], [A33], [A34], [A35], [A36], [A37], [A38], [A39], [A40], [A41], [A42], [A43], [A44], [A45], [A46], [A47], [A48], [A49]
Gerenciamento de Demanda	7	14%	[A5], [A6], [A10], [A11], [A21], [A46], [A50]
Manufatura	5	10%	[A2], [A3], [A19], [A23], [A29]
Transporte / Logística	3	6%	[A9], [A22], [A34]
Armazenamento / Depósito	3	6%	[A8], [A12], [A42]
<b>Total</b>	<b>50</b>		

freqüência. Os principais resultados são apresentados demonstrados na Tabela 3.

A maioria dos artigos, isto é, 66% (37/50) aplicam os conceitos *Big Data Analytics* nos níveis de análise preditiva e descritiva. A análise preditiva busca encontrar informações que indicam movimentos futuros baseados nos dados que estão sendo informados pelas mais diversas plataformas. Enquanto que na análise descritiva, tenta-se encontrar os relacionamentos e/ou associações entre os dados históricos para prever movimentos futuros.

Além disso, os níveis de *BDA* misto e prescritivo representam 34% (17/50) dos artigos avaliados. Tratam-se de níveis genéricos. O misto aplica mais de uma técnica de *BDA* para obtenção de resultados, de modo análogo como é feito com a integração de ferramentas para medir a qualidade de software [26]. A técnica prescritiva busca a implementação de ferramentas e mecanismos utilizados para análise e apresentação de informações obtidas pelas técnicas de *BDA*.

Os resultados apontam para a concentração de aplicação de *Big Data Analytics* nos níveis de análises de preditiva e descritiva. Analisar nesses níveis é ideal para situações onde a quantidade de informações aumenta exponencialmente e que são geradas por mais de um meio. Essas análises possibilitam

Tabela 3: Níveis de *Big Data Analytics* relacionados aos estudos selecionados.

Níveis de Big Data Analytics	Números de Estudos	Percentual	Lista de Artigos
Preditiva	18	36%	[A1], [A3], [A4], [A5], [A8], [A12], [A14], [A19], [A22], [A33], [A36], [A37], [A39], [A40], [A42], [A44], [A46], [A50]
Descritiva	15	30%	[A6], [A7], [A9], [A10], [A11], [A17], [A18], [A23], [A24], [A26], [A27], [A32], [A38], [A48], [A49]
Mista	15	30%	[A2], [A13], [A15], [A16], [A20], [A22], [A25], [A29], [A30], [A34], [A35], [A41], [A43], [A45], [A47]
Prescritiva	2	4%	[A28], [A31]
<b>Total</b>	<b>50</b>		

que processos possam ser criados para analisar, interpretar, e prever situações nesse grande volume de informações.

### 4.3 QP3: Modelos de *Big Data Analytics*

Esta questão de pesquisa realiza um levantamento sobre os modelos de *BDA* que estão sendo aplicados com maior freqüência. Tabela 4 apresenta os principais resultados.

Tabela 4: Modelos de *Big Data Analytics* relacionados aos estudos selecionados

Modelos de Big Data Analytics	Números de Estudos	Percentual	Lista de Artigos
Otimização	16	32%	[A2], [A3], [A6], [A9], [A10], [A16], [A17], [A19], [A20], [A22], [A26], [A32], [A33], [A34], [A37], [A49]
Previsão	8	16%	[A4], [A5], [A12], [A21], [A27], [A36], [A46], [A50]
Classificação	8	16%	[A23], [A25], [A31], [A35], [A41], [A43], [A45], [A48]
Simulação	8	16%	[A1], [A8], [A13], [A14], [A15], [A39], [A40], [A44]
Visibilidade	4	8%	[A7], [A11], [A18], [A29]
Misto	3	6%	[A28], [A30], [A42]
Outros	3	6%	[A24], [A38], [A47]
<b>Total</b>	<b>50</b>		

Os modelos de otimização e previsão representam 48% (24/50) dos artigos classificados. Os modelos de otimização procuram otimizar uma etapa da cadeia através de informações, tais como, redução de tempo, custo, ou processos.

O modelo de previsão procura fornecer previsões para um melhor posicionamento, e geralmente está relacionado a gestão e tomada de decisão, por exemplo, na decisão de um melhor ajuste da cadeia de suprimentos ao analisar e prever o comportamento de clientes [26]. Modelos de classificação e simulação representam 32% (16/50) dos artigos avaliados. O modelo de classificação elenca os principais processos da cadeia para priorizar os processos mais relevantes. Os processos de simulação concentram esforços em prever e/ou evitar possíveis problemas geralmente em processos de manufatura. Os modelos de visibilidade, misto e outros representam apenas 20% (10/50) dos artigos avaliados. Tratam-se de modelos que apresentam técnicas relacionadas a gestão da informação.

Analisando os modelos de *BDA* mais utilizados nos artigos, possível perceber uma concentração dos modelos de otimização e previsão. A gestão da informação no auxílio na tomada de decisão utilizando *Big Data Analytics* tem sido a maior preocupação dos pesquisadores.

#### 4.4 QP4: Técnicas de Big Data Analytics

Esta questão de pesquisa apresenta o levantamento das técnicas de *Big Data Analytics* que estão sendo aplicadas com maior frequência. Os principais resultados estão resumidos na Tabela 5.

**Tabela 5: Técnicas de *Big Data Analytics* relacionadas aos estudos selecionados**

Técnicas de <i>Big Data Analytics</i>	Números de Estudos	Percentual	Lista de Artigos
Misto	14	28%	[A1], [A2], [A7], [A19], [A22], [A24], [A25], [A28], [A30], [A39], [A44], [A47], [A49], [A50]
Visualização	11	22%	[A12], [A15], [A18], [A20], [A23], [A29], [A31], [A32], [A35], [A38], [A40]
Abordagem Heurística	9	18%	[A4], [A16], [A17], [A21], [A27], [A41], [A43], [A45], [A48]
Algoritmo Genérico	8	16%	[A3], [A9], [A10], [A11], [A13], [A26], [A34], [A42]
Regra de mineração	3	6%	[A33], [A36], [A37]
Árvore de decisão	3	6%	[A6], [A8], [A14]
Análise de Sentimentos	2	4%	[A5], [A46]
Total	50		

As técnicas de *BDA* Misto, Visualização e Abordagem Heurística representam 68% (34/50) dos artigos selecionados. Estas técnicas tem uma alta representatividade, pois a técnica mista utiliza entre duas ou mais técnicas para compor uma estrutura de *BDA*. A técnica de visualização geralmente é utilizada para complementar técnicas de mineração de dados. Este tipo de técnica encontra-se na maioria dos artigos que possuem classificação mista. A abordagem heurística é amplamente utilizada para otimização.

As demais técnicas representam 32% (16/50) dos artigos. Apesar de serem técnicas relevantes, estão ligadas a processos específicos. Geralmente são utilizadas em conjunto com outras técnicas de *Big Data Analytics*. Isso é relacionado a uma tendência na utilização de técnicas genéricas, pois podem ser aplicadas em todas as etapas da cadeia. Desse modo, combinando técnicas focadas em fornecer informações com técnicas de otimização de processos. Enquanto isso, as técnicas específicas vêm sendo pouco utilizadas devida à sua baixa capacidade de integração com outros tipos de técnicas.

## 5. DISCUSSÃO E DESAFIOS

Esta seção apresenta algumas discussões e desafios que foram derivados dos estudos selecionados.

**Contraste entre pesquisa acadêmica e prática industrial.** Os resultados da Tabela 3 mostram que os resultados se concentraram em investigar recursos de *BDA* voltados para gestão de cadeia de suprimentos, ou seja, métodos que auxiliem na tomada de decisão dos mais diversos processos. Por outro lado, profissionais da área de sistemas de informação alegam que possuem uma percepção de que utilização de *Big Data Analytics* impacta na melhoria de gestão das áreas de estoque e ruptura [6]. Isto é uma evidência de que parte das pesquisas selecionadas focaram em uma área que tem menos relevância na prática. Portanto, seria interessante de que os próximos estudos preencham essa lacuna de pesquisa: aplicar, executar e desenvolver técnicas utilizando *Big Data Analytics* na gestão das áreas de estoque e ruptura.

**Arquitetura para *Big Data Analytics* em tempo real.** Grande parte dos trabalhos utilizaram técnicas preditivas e de aprendizagem de máquina, onde as características mais fortes são a utilização de dados históricos para extração e predições de informações relevantes [8]. No entanto, há uma deficiência em técnicas para utilização de *Big Data Analytics* em tempo real. É importante que as informações sejam instantaneamente coletadas e logo analisadas, pois isso antecipa o auxílio da tomada de decisão. Esta é uma forte tendência pois as fontes de dados que podem auxiliar no gerenciamento da cadeia de suprimentos tendem a aumentar, como, por exemplo, nos conceitos de cidades inteligentes onde toda a cidade é conectada por sensores e tais informações são disponibilizadas em tempo real [18]. Por exemplo, os sensores instalados em um semáforo podem ser utilizados para antecipar atualizar o caminho de entrega de mercadorias em um cliente. Essas informações, entre muitas outras tornam cada vez mais os sistemas de informação que fazem parte da cadeia de suprimentos ficarem mais completos e dinâmicos. Porém, por outro lado, com o aumento exponencial de informações que é obtido em um tempo cada vez menor, é necessário que técnicas de análise sejam desenvolvidas para acompanhar essa demanda. Neste contexto, Cadaviz et al. [4] reforçam a importância de se ter arquiteturas robustas para dar suporte a sistemas de informação que lidam com grande volume de dados e em tempo real.

## 6. TRABALHOS RELACIONADOS

Esta seção apresenta os trabalhos relacionados a esta pesquisa.

Wamba et al. [30] conduziram uma revisão sistemática sobre o impacto de Big Data na economia. Ao total, foram analisados 68 artigos entre 2009 a 2012. Wamba et al. [30] os autores focaram no setor da indústria onde o *Big Data* é aplicado, os métodos de pesquisa, local de publicação, e a distribuição cronológica dos estudos. Ao contrário, este estudo adotou a metodologia de mapeamento sistemático, e analisou 50 artigos publicados até 2017. Além disso, trata-se de uma mapeamento focado na aplicação de *Big Data Analytics* no gerenciamento de cadeias de fornecedores.

Babiceanu e Seker [1] realizaram uma revisão da literatura sobre *Big Data Analytics* em sistemas de operação de manufatura. Foram analisados trabalhos publicados até 2016. Contudo, este trabalho aborda o tema de *Big Data Analytics* aplicado a sistemas de manufatura. Ao contrário, este trabalho procurou classificar a utilização de *Big Data Analytics*

por sistemas de informação aplicadas em todas as cadeias de fornecedores.

Hazen et al. [18] apresentaram uma lista de desafios futuros em relação a aplicação de *Big Data Analytics* na sustentabilidade para cadeias de fornecedores. Contudo essas cadeias de fornecedores são específicas as teorias de capital social, institucionais, dependências de recursos, e custo de transação econômica. Essa agenda de pesquisa foi baseada em trabalhos publicados até 2016.

Em geral, este trabalho fornece uma classificação e uma visão panorâmica sobre a aplicação de *Big Data Analytics* em cadeia de suprimento. Além disso, essa classificação foi realizada a partir de um mapeamento sistemático fornecendo uma perspectiva quantitativa, e discussões sobre o tema.

## 7. AMEAÇAS À VALIDADE

Esta seção discute as estratégias utilizadas para mitigar algumas ameaças à validade. Aplicou-se a metodologia do estudo de mapeamento sistemático para a execução desta pesquisa [2, 19, 31]. Estudos anteriores destacam a importância de identificar as ameaças à validade de estudos empíricos [11, 12, 13].

Este método fornece protocolos para extrair dados, a fim de garantir resultados detalhados do uso da metodologia. Assim, as palavras-chave, o termo de busca, e as questões de pesquisa foram definidas. No entanto, alguns fatores podem ameaçar a validade do estudo: (1) a dificuldade de identificar uma amostra representativa da literatura atual sobre as questões de pesquisa exploradas, bem como de recuperar uma quantidade gerenciável de estudos para análise; (2) a dificuldade de filtrar os trabalhos potencialmente relevantes de forma manual, mitigando a subjetividade dos autores; (3) a realização da extração de dados dos artigos artigos heterogêneos, em termos de método empírico utilizado, propósito e estrutura de análise dos dados; (4) a inclusão de teses e dissertações publicadas on-line que não são revisadas por pares; e (5) a dificuldade de extração e síntese de dados de artigos heterogêneos, sem o suporte de qualquer taxonomia ou estudo de revisão da literatura anterior.

## 8. CONCLUSÃO

Este trabalho apresentou uma classificação da literatura sobre o uso de *Big Data Analytics* em gerenciamento de cadeia de suprimentos, bem como identificou desafios e tendências na área. Quatro questões de pesquisa foram exploradas através de um rigoroso processo de identificação, seleção, extração de dados e síntese de dados. Levando em consideração o método de filtragem elaborado para o estudo de mapeamento, ao final do processo foram extraídos 50 artigos para análise (de uma amostra inicial de 5.437), que foram classificados de acordo com as questões de pesquisa elaboradas. A partir destas questões de pesquisa foi possível identificar os direcionamentos dos estudos selecionados. Conforme os objetivos principais deste estudo, identificou-se que mais de 70% das áreas onde conceitos de *Big Data Analytics* estão sendo aplicados em gestão e a análise de dados. Especificamente, a grande maioria voltada para gestão de processos da cadeia produtiva, e a uma minoria dos trabalhos focaram em gestão de demanda.

As principais contribuições deste trabalho foram: (1) revisão e classificação da literatura atual sobre aplicação de *Big Data Analytics* no gerenciamento de cadeia de suprimentos;

(2) identificação das áreas da cadeia de suprimentos mais impactadas por *Big Data Analytics*, tanto no âmbito acadêmico como no âmbito profissional; (3) apresentação das principais técnicas de *Big Data Analytics* utilizadas. Entende-se que estas contribuições são importantes a comunidade de sistemas de informação, pois apontam trabalhos relacionados a esta área, e além disso traz um entendimento sobre como que as técnicas de *Big Data Analytics* foram adotadas no gerenciamento de cadeia de suprimentos.

## 9. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, e Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), código 313285/2018-7.

## 10. REFERÊNCIAS

- [1] R. F. Babiceanu and R. Seker. Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. *Computers in Industry*, 81:128–137, 2016.
- [2] V. Bischoff, K. Farias, L. J. Gonçalves, and J. L. V. Barbosa. Integration of feature models: A systematic mapping study. *Information and Software Technology*, 105:209–225, 2019.
- [3] D. J. Bowersox, D. J. Closs, and O. K. Helferich. *Logistical management*, volume 6. McGraw-Hill New York, NY, 1996.
- [4] M. K. Cadaviz, K. Farias, L. J. Gonçalves, and V. Bischoff. Doric: An architecture for data-intensive real-time applications. In *Proceedings of the XIV Brazilian Symposium on Information Systems*, 2018.
- [5] C. R. Carter and L. M. Ellram. Reverse logistics: a review of the literature and framework for future investigation. *Journal of business logistics*, 19(1):85, 1998.
- [6] D. Q. Chen, D. S. Preston, and M. Swink. How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, 32(4):4–39, 2015.
- [7] M. C. Cooper, D. M. Lambert, and J. D. Pagh. Supply chain management: more than a new name for logistics. *The international journal of logistics management*, 8(1):1–14, 1997.
- [8] E. de Paiva and K. Revoredo. Big data e transparência: Utilizando funções de mapreduce para incrementar a transparência dos gastos públicos. In *XII Simpósio Brasileiro de Sistemas de Informação*, pages 025–032. SBC, 2016.
- [9] L. M. Ellram and M. C. Cooper. Supply chain management, partnership, and the shipper-third party relationship. *The International Journal of Logistics Management*, 1(2):1–10, 1990.
- [10] J. Fan, F. Han, and H. Liu. Challenges of big data analysis. *National science review*, 1(2):293–314, 2014.
- [11] K. Farias, A. Garcia, and C. Lucena. Effects of stability on model composition effort: an exploratory study. *Software & Systems Modeling*, 13(4):1473–1494, 2014.
- [12] K. Farias, A. Garcia, J. Whittle, C. v. F. G. Chavez, and C. Lucena. Evaluating the effort of composing

- design models: a controlled experiment. *Software & Systems Modeling*, 14(4):1349–1365, 2015.
- [13] K. Farias, A. Garcia, J. Whittle, and C. Lucena. Analyzing the effort of composing design models of large-scale software in industrial case studies. In *International Conference on Model Driven Engineering Languages and Systems*, pages 639–655, 2013.
- [14] A. Gandomi and M. Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137–144, 2015.
- [15] L. Gonçalves, K. Farias, B. da Silva, and J. Fessler. Measuring the cognitive load of software developers: a systematic mapping study. In *IEEE/ACM 27th International Conference on Program Comprehension (ICPC)*, pages 42–52, 2019.
- [16] L. J. Gonçalves, K. Farias, T. C. D. Oliveira, and M. Scholl. Comparison of software design models: An extended systematic mapping study. *ACM Computing Surveys (CSUR)*, 52(3):1–41, 2019.
- [17] A. Gunasekaran, T. Papadopoulos, R. Dubey, S. F. Wamba, S. J. Childe, B. Hazen, and S. Akter. Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research*, 70:308–317, 2017.
- [18] B. T. Hazen, J. B. Skipper, J. D. Ezell, and C. A. Boone. Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. *Computers & Industrial Engineering*, 101:592–598, 2016.
- [19] B. A. Kitchenham, D. Budgen, and O. Pearl Brereton. Using mapping studies as the basis for further research - a participant-observer case study. *Information Software Technology*, 53(6):638–651, June 2011.
- [20] A. Labrinidis and H. V. Jagadish. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12):2032–2033, 2012.
- [21] X. Li, K.-S. Chang, J. J. Dooley, A. A. Deshpande, T. A. Greene, and D. J. Hakeman. Supply chain visibility for real-time tracking of goods, Nov. 14 2006. US Patent 7,136,832.
- [22] V. Maretta, P. A. Júnior, and H. Costa. Uma revisão sistemática da literatura sobre comunicação no contexto da gerência de projetos de sistemas de informação. In *XII Simpósio Brasileiro de Sistemas de Informação*, pages 084–091. SBC, 2016.
- [23] J. T. Mentzer, W. DeWitt, J. S. Keebler, S. Min, N. W. Nix, C. D. Smith, and Z. G. Zacharia. Defining supply chain management. *Journal of Business Logistics*, 22(2):1–25, 2001.
- [24] A. Oliveira, V. Bischoff, L. Goncales, K. Farias, and M. Segalotto. Brcode: An interpretive model-driven engineering approach for enterprise applications. *Computers in Industry*, 96:86–97, 2018.
- [25] K. Petersen, S. Vakkalanka, and L. Kuzniarz. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64:1–18, 2015.
- [26] D. Raitz, J. L. de Aguiar, and K. A. L. Godarth. Comércio eletrônico: Comportamento do consumidor na internet. *iSys-Revista Brasileira de Sistemas de Informação*, 10(2):153–170, 2017.
- [27] P. Santos, K. Delgado, M. Lauretto, and M. Ribeiro. Revisão sistemática da literatura sobre ranking de relacionamentos na web semântica. In *XIII Simpósio Brasileiro de Sistemas de Informação*, pages 32–39. SBC, 2017.
- [28] K. H. Tan, Y. Zhan, G. Ji, F. Ye, and C. Chang. Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. *International Journal of Production Economics*, 165:223–233, 2015.
- [29] C.-Y. Tsai and S.-H. Huang. A data mining approach to optimise shelf space allocation in consideration of customer purchase and moving behaviours. *International Journal of Production Research*, 53(3):850–866, 2015.
- [30] S. F. Wamba, S. Akter, A. Edwards, G. Chopin, and D. Gnanzou. How big data can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165:234–246, 2015.
- [31] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.