



Análise de Big Data por meio de estatísticas multivariadas na Indústria 4.0: Uma revisão da literatura

Big Data analytics supported by multivariate statistics in Industry 4.0: a literature review

Daniel Ayub¹ 

Wiliam de Assis Silva² 

Marcos Augusto Mendes Marques³ 

Mariana Kleina⁴ 

Fabiano Oscar Drozda⁵ 

Resumo

Com relação ao tema Big Data, observa-se atualmente uma quantidade significativa e um aumento considerável de publicações relacionadas ao assunto, porém ainda os pesquisadores sentem falta de meios para auxiliar na escolha dos referenciais teóricos. Desta forma, o objetivo deste trabalho é demonstrar um processo utilizado para a seleção das publicações relevantes, as quais são o produto de uma revisão sistemática da literatura e que buscam nortear os pesquisadores, agregando conhecimento para conduzir uma pesquisa sobre os métodos analíticos aplicados em Big Data, em ambientes de Manufaturas Inteligentes e apoiadas por abordagens multivariadas. Para buscar esse objetivo foi desenvolvido um roteiro de pesquisa e uma técnica de classificação das publicações mais relevantes. Como principais resultados deste processo, foi possível identificar 14 publicações aderentes, e que permitiram integrar os conceitos sobre Big Data, Indústria 4.0 e abordagens multivariadas, além de demonstrar a análise preditiva de dados como um dos modelos mais utilizados na análise de Big Data.

Palavras-chaves: Big Data; Análise Big Data; Indústria 4.0; Abordagens Multivariadas, Revisão sistemática de Literatura.

Abstract

There is a significant amount of information about Big Data and a considerable increase in publications related to the subject, but researchers still lack the means to help in the choice of theoretical references. In this way, the objective of this work is to demonstrate a process used to select the relevant publications, which are the product of a systematic review of the literature and that seek to guide the researchers, adding knowledge to conduct a research on the analytical methods applied in Big Data, in Intelligent Manufacturing environments that are supported by multivariate approaches. To achieve this objective, a research guide and a classification technique of the most relevant publications were developed. As main results of this process, it was possible to identify 14 adherent publications, which allowed to integrate the concepts on Big Data, Industry 4.0 and multivariate approaches, besides demonstrating the predictive data analysis as one of the most used models in the context of Big Data analysis.

Keywords: Big Data; Big Data Analysis; Industry 4.0; Multivariate Approaches; Literature Review.

Cite as: (APA) Ayub, D., Silva, W. A., Marques, M. A. M., Kleina, M., & Drozda, F. O. (2021). Análise de Big Data por meio de estatísticas multivariadas na Indústria 4.0: Uma revisão da literatura. *Revista Competitividade e Sustentabilidade*, 8(1), 3-16.

¹ Universidade Federal do Paraná - UFPR. Brasil. E-mail: daniel.ayub@ufpr.br

² Universidade Federal do Paraná - UFPR. Brasil. E-mail: wiliamdeassis@gmail.com

³ Universidade Federal do Paraná - UFPR. Brasil. E-mail: marquesammarcos@gmail.com

⁴ Universidade Federal do Paraná - UFPR. Brasil. E-mail: marianakleina11@gmail.com

⁵ Universidade Federal do Paraná - UFPR. Brasil. E-mail: fabiano.drozda@gmail.com

1 INTRODUÇÃO

Com a inclusão das novas tecnologias da informação nos negócios e em seus processos, as organizações estão se deparando com os desafios criados pelo avanço das técnicas de geração, transmissão e armazenamento de dados, as quais estão gerando uma explosão de informações em nível global e nos mais diversos setores da economia (Shao, Shin, Jain, 2015).

Estas novas tecnologias digitais estão mudando a forma de pensar das organizações em relação ao seu desempenho e eficiência de processos, e com o objetivo de integrar estes conceitos recentes, um novo modelo tecnológico de gestão denominado de Manufatura Inteligente está sendo adotado pelas organizações como forma de estratégia competitiva. Este modelo, mais popularmente denominado de Indústria 4.0, é definido de forma genérica como um sistema avançado de inteligência que permite integrações rápidas e dinâmicas dos processos e otimização da tomada de decisão, com o foco na melhoria da eficiência e desempenho organizacional (Shin, Woo, Rachuri, 2014).

Nesta transição, está sendo gerada em alta velocidade uma quantidade variada e extensa de dados, denominada de *Big Data*. Esta rápida expansão de dados poderá gerar desafios relacionados às formas de extração e assimilação de conhecimento, além da instigação da capacidade humana para gerir todos esses dados disponíveis, e assim gerar valor sobre essas informações (Olshannikova, Ometov, Koucheryavy, Olsson, 2015).

Em 2017 eram previstas 1,7 bilhões de máquinas com conexões sem fio (Jung, 2011). Estima-se que o volume de dados gerados até o ano de 2020 em todo o mundo chegará a 44 zettabytes (Idc, 2014).

Observando tal tendência, a análise de *Big Data*, com o intuito de melhoria do desempenho das organizações está sendo cada vez mais frequente, pois está proporcionando novas visões e cenários a respeito dos negócios, levando a um ambiente mais competitivo, por meio da redução de custos, redução de ineficiências e consequente aumento de lucratividade (Libes, Shin, Woo, 2015).

A análise de *Big Data* tende a ser um fator decisivo para o sucesso das organizações, ainda mais para as fábricas inteligentes (Shao, Shin, Jain, 2015). Neste cenário, o número de hipóteses a serem consideradas aumenta exponencialmente, sendo desta forma relevante um modelo para a análise de dados que permita a extração de informações da forma mais eficiente e prática (Kudo, Akitomi, Moriwaki, 2015). Neste contexto, as abordagens estatísticas multivariadas podem viabilizar diferentes categorias de análises para este novo ambiente tecnológico das manufaturas inteligentes.

Sabendo que existem milhares de publicações científicas sobre o tema e que estas estão disponíveis em bases acadêmicas e na internet, este trabalho propõe apresentar uma revisão sistemática e conhecer o que a literatura científica apresenta sobre o processo de análise de *Big Data*, integrados aos conceitos da Indústria 4.0 e aplicados com o apoio de abordagens estatísticas multivariadas. O objetivo é pesquisar em diversas bases materiais teóricos sobre o tema, buscando fontes confiáveis e apresentando assim um processo estruturado e capaz de criar para os pesquisadores um conhecimento introdutório sobre o assunto.

2 MATERIAIS E MÉTODOS

Para a revisão sistemática da literatura foi desenvolvido um roteiro, onde foram utilizadas sequências lógicas de pesquisa, desdobramento das informações e segregação de

publicações voltadas ao tema de pesquisa. As etapas realizadas estão representadas no fluxo da Figura 1.

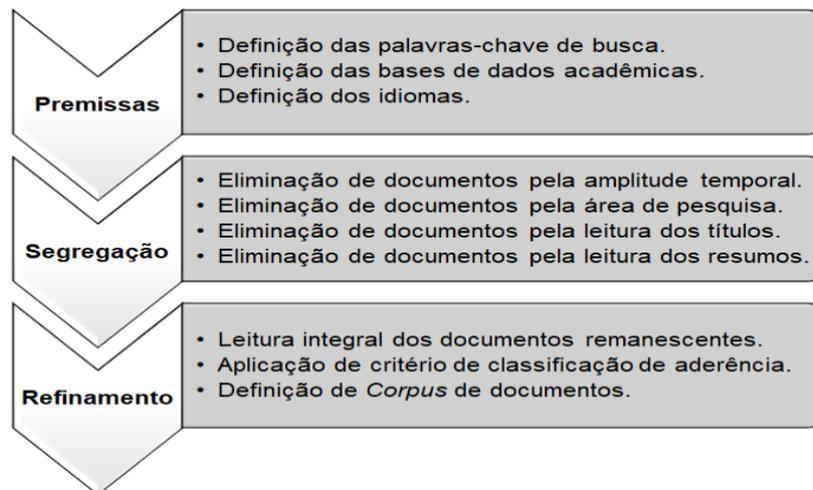


Figura 1. Roteiro utilizado para a revisão da literatura.

Fonte: Os Autores, 2019.

Seguindo as premissas de pesquisa, primeiramente definiram-se as palavras-chave mais associadas ao conteúdo da pesquisa. Para tal, foram combinados diferentes termos atrelados ao tema, porém correlacionados à pesquisa e que agregam valor. Ao todo foram elaboradas seis *keywords* para estender a abrangência de busca, conforme Tabela 1, sendo tais termos transformados em parâmetros de pesquisa por meio do operador lógico booleano “AND”.

Tabela 1: Parâmetros de Busca

Termos de Busca	Palavras-Chave
<i>Multivariate Analysis Applied in Big Data</i>	“Multivariate Analysis” AND “Big Data”
<i>Multivariate Analysis Applied in Industrial Internet</i>	“Multivariate Analysis” AND “Industrial Internet”
<i>Multivariate Analysis Applied in Industry 4.0</i>	“Multivariate Analysis” AND “Industry 4.0”
<i>Analytical Methods Applied in Big Data</i>	“Analytical Methods” AND “Big Data”
<i>Big Data Applied in Industry 4.0</i>	“Big Data” AND “Industry 4.0”
<i>Big Data Applied in Smart Manufacturing</i>	“Big Data” AND “Smart Manufacturing”

Fonte: Os Autores, 2019.

Na sequência, estabeleceram-se as bases de dados *Scopus*, *Science Direct*, *Web of Science* e Portal CAPES para a realização da busca. Tais bases possuem grande referência nos meios acadêmicos do Brasil.

Como idioma de pesquisa, optou-se pelo inglês por ser a língua mais utilizada em publicações, e também pelo português, por ser a língua nativa do local de aplicação.

Na segunda etapa da pesquisa foram definidos os critérios de segregação das publicações resultantes:

- Amplitude temporal: Publicações efetivadas nos últimos cinco anos (2013 a 2017).

- Área de pesquisa: Engenharia, matemática / estatística, ciências da computação e ciências da decisão, as quais são as mais correlacionadas ao tema e ao foco do trabalho.
- Leitura dos títulos e resumos: Leitura de todos os títulos e resumos das publicações resultantes da pesquisa, visando à eliminação de trabalhos repetidos e não correlacionados ao tema de pesquisa.

Com a fase final de refinamento, procurou-se por meio da leitura integral das publicações remanescentes e da classificação de aderência destas publicações, a criação de um *Corpus* de documentos que demonstrará as publicações mais aptas há se tornarem a base de pesquisa mais correlacionada ao tema do trabalho.

A classificação de aderência será verificada por meio de um *ranking* de classificação das publicações. Essa pontuação classificatória será originada pela somatória da multiplicação de dois fatores, um em relação à importância dos temas para a presente pesquisa, conforme Tabela 2, e o outro pela abrangência dos assuntos dentro dos documentos, conforme Tabela 3.

Tabela 2: Pesos em relação aos temas de pesquisa

Tema atrelado à pesquisa	Pesos
Métodos analíticos de <i>Big Data</i>	10
Conceitos de <i>Big Data</i>	10
Abordagem estatística ou multivariada	10
Modelos de análise ou mineração de dados	7
Indústria 4.0 ou Smart Manufacturing	5

Fonte: Os Autores, 2019.

Tabela 3: Pontuações em relação à abrangência dos temas nos artigos

Abrangência do tema dentro do texto	Pontuação
Elemento de pesquisa detalhado no texto	5
Elemento de pesquisa explanado no texto	3
Elemento de pesquisa apenas citado no texto	1
Elemento de pesquisa não citado no texto	0

Fonte: Os Autores, 2019.

3 RESULTADOS E DISCUSSÃO

Após a realização da busca nas bases acadêmicas, utilizando-se as configurações padrões de pesquisas oferecidas por elas, foi possível eliminar algumas palavras-chave, pois as mesmas não possuíam grande influência nos resultados da pesquisa. Os termos *Multivariate Analysis Applied in Industrial Internet* e *Multivariate Analysis Applied in Industry 4.0* não apresentaram resultados significativos nas bases de dados selecionadas, conforme demonstrado na Tabela 4, sendo assim, estas palavras-chaves não foram consideradas no total das 2.666 publicações.

Tabela 4: Resultado da pesquisa nas bases acadêmicas

Palavra-Chave	Scopus	Science Direct	Web Science of	Portal Capes	Total
<i>Multivariate Analysis AND Big Data</i>	47	225	25	434	731
<i>Multivariate Analysis AND Industrial Internet</i>	(0)	(2)	(0)	(6)	(8)
<i>Multivariate Analysis AND Industry 4.0</i>	(0)	(0)	(0)	(3)	(3)
<i>Analytical Methods AND Big Data</i>	74	342	32	565	1.013
<i>Big Data AND Industry 4.0</i>	149	(0)	83	314	546
<i>Big Data AND Smart Manufacturing</i>	64	125	33	154	376
Total	334	692	173	1.467	2.666

Fonte: Os Autores, 2019.

Mediante tal análise também pôde ser observado outros parâmetros de classificação para o tema, como a quantidade de publicações por país de origem, o resultado das publicações nos últimos cinco anos, por área de pesquisa e por principais palavras-chave utilizadas.

Em relação aos países que mais produziram materiais científicos no período destacam-se os Estados Unidos e a China, com cerca de 39% das publicações, conforme resultado na Figura 2. Dentre as quatro palavras-chave utilizadas, os Estados Unidos se destacam nos campos da Análise Multivariada e Métodos de Análise em *Big Data*. Já a China possui grande evidência nas publicações referentes à Indústria 4.0. O *ranking* foi elaborado contemplando os países que contribuíram com até 80% das publicações pesquisadas. O Brasil aparece na 25ª colocação deste ranking de publicações, com 5 documentos e impactando em 0,9% do total, demonstrando assim que não possui grande notoriedade em pesquisas sobre o tema.

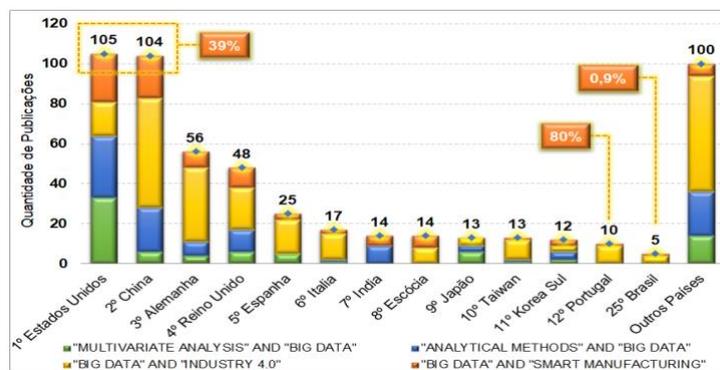


Figura 2. Quantidade de publicações por país de origem
Fonte: Os Autores, 2019.

Conforme demonstrado na Figura 3, em relação às publicações realizadas de 2013 a 2017, percebe-se uma tendência no crescimento de materiais científicos sobre o tema. Considerando o período entre 2015 a 2017, é possível observar um crescimento de mais de 116% na quantidade de publicações.

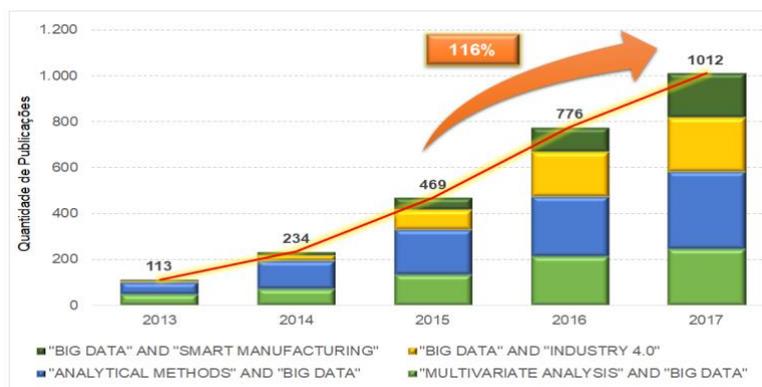


Figura 3. Quantidade de publicações por ano e por palavra-chave
Fonte: Os Autores, 2019.

Quanto às áreas de pesquisa, expõe-se uma tendência nas áreas da Ciência da Computação e Engenharia, tendo também impacto nas áreas de Matemática / Estatística e Ciências da Decisão. Esta observação confirmou o foco das áreas de pesquisa, conforme observado na Figura 4.

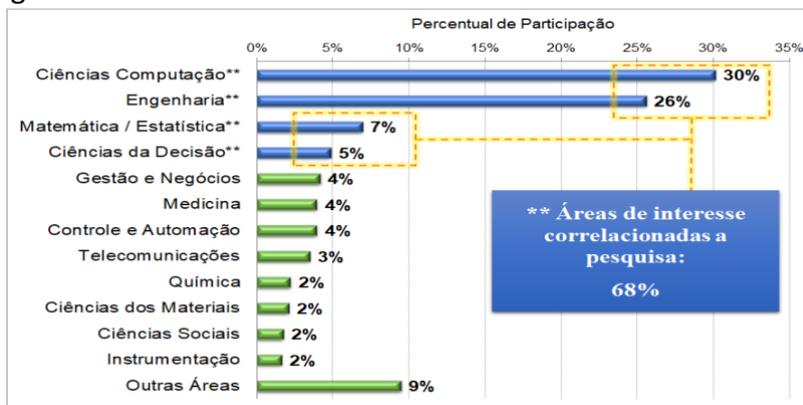


Figura 4. Publicações por área de pesquisa
Fonte: Os Autores, 2019.

Conforme Figura 5, durante a pesquisa, também se observou as principais palavras-chave utilizadas nas publicações. Dentre elas, o termo *Big Data* representou um total de 23% em relação às demais *keywords* utilizadas.

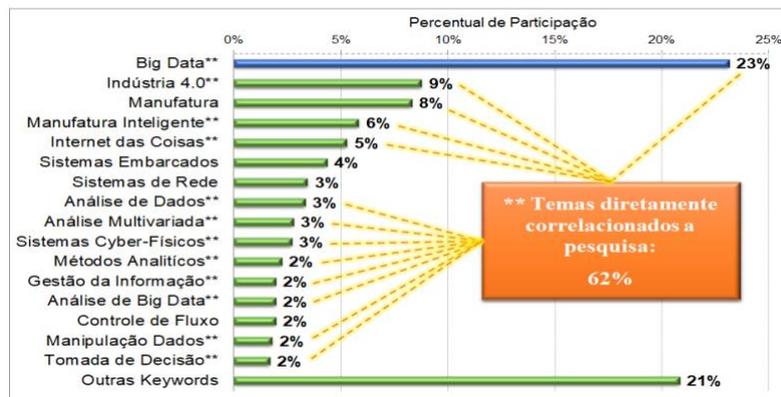


Figura 5. Publicações por palavras-chave
Fonte: Os Autores, 2019.

Como resultado da busca identificou-se um total de 2.666 documentos, os quais foram reduzidos para 2.604 devido a primeira faixa de corte, referente aos materiais publicados entre os anos de 2013 a 2017. Na sequência foram acrescentados mais níveis de segregação.

O próximo corte refere-se às áreas de interesse demonstradas na Figura 4 (Ciências da Computação, Engenharia, Matemática / Estatística e Ciências da Decisão), o que gerou uma redução de 36% das publicações originais, tendo um resultado de 1.705 publicações resultantes, conforme demonstrado na Figura 6.

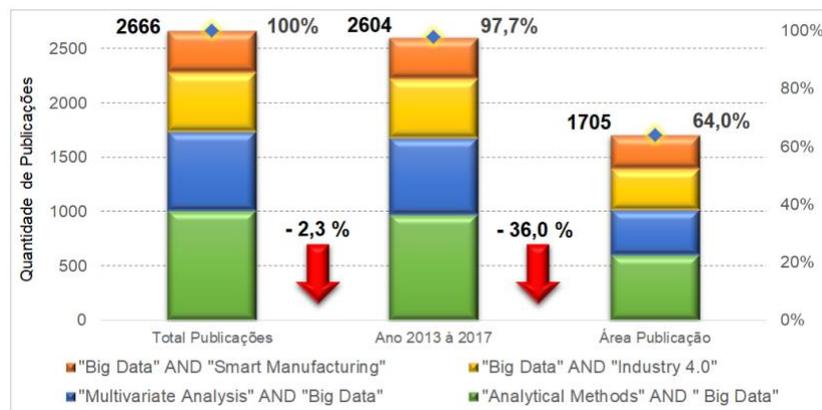


Figura 6. Estratificação por ano e área de publicação da pesquisa
Fonte: Os Autores, 2019.

Na etapa seguinte, selecionaram-se os documentos que realmente se encaixavam ao tema de pesquisa, por meio da busca de palavras-chave em tabelas dinâmicas geradas com base nos títulos e resumos de todas as publicações selecionadas até a presente fase. Como resultado desta leitura, foi elaborado um *Corpus* preliminar com 60 publicações aderentes ao tema e que podem contribuir de maneira significativa para a pesquisa, representando uma redução de 97,7% em referência a quantidade original de documentos.

Conforme apresentado na seção anterior, as 60 publicações ganharam uma classificação para os documentos mais relevantes ao tema, sendo o primeiro com 160 pontos e o último com 25 pontos. Para determinar uma linha de corte foi utilizado o critério estatístico da média e do desvio padrão em relação às pontuações registradas, conforme Tabela 5.

Tabela 5: Análise estatística do ranking de aderência

Pontuações do ranking de aderência		
Média dos pontos	Desvio padrão	Linha de corte
59,8	44,1	$59,8 + 44,1 = 103,9$

Fonte: Os Autores, 2019.

Para o ranking de aderência foi definida a linha de corte de 103,9 pontos, referente à soma da média dos pontos mais um desvio padrão. Após a aplicação deste corte restaram 12 artigos mais impactantes.

Analisando as demais publicações, observou-se que existiam algumas com menores pontuações de aderência, mas com potencial para participação desta seleção devido ao número expressivo de citações.

Desta forma foi criada mais uma linha de corte, porém esta em relação às citações. Foi utilizado o número de citações oriundas do *Google Acadêmico*. Utilizando a mesma metodologia estatística anterior, conforme Tabela 6, foi possível definir o critério de seleção, que neste caso definiu-se como os documentos que possuíam mais do que 201 citações.

Tabela 6: análise estatística das citações

Citações <i>Google Acadêmico</i>		
Média dos pontos	Desvio padrão	Linha de corte
48,5	152,5	$48,5 + 152,5 = 201$

Fonte: Os Autores, 2019.

Nesta triagem foram selecionados mais dois documentos, finalizando em 14 a quantidade de artigos com grande impacto para a pesquisa, conforme demonstrado na Figura 7.

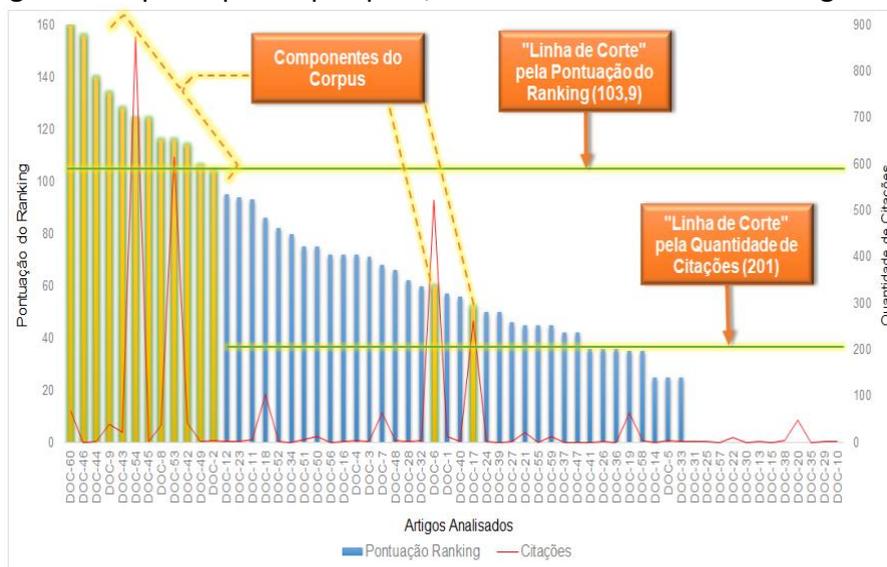


Figura 7. Linhas de corte para ranking e citações

Fonte: Os Autores, 2019.

Como resultado final, a Figura 8 demonstra a redução das publicações até a criação do Corpus final de documentos aderentes ao tema de pesquisa.

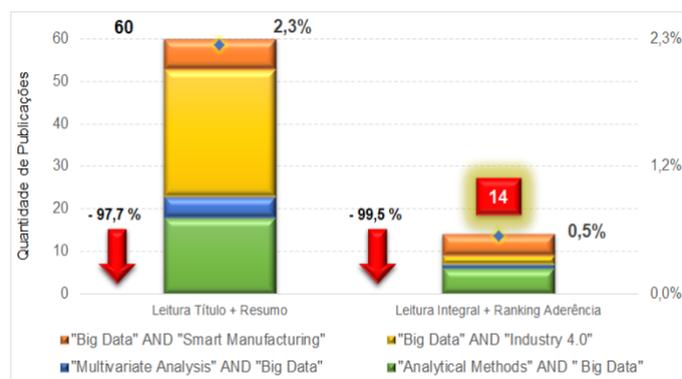


Figura 8. Estratificações por leitura e *ranking* de aderência
Fonte: Os Autores, 2019.

3.1. Publicações mais impactantes atreladas aos temas de pesquisa

Na Tabela 7 são demonstrados por ordem do ranking de aderência os 14 artigos componentes do Corpus de publicações significativas para a pesquisa acadêmica.

Tabela 7: Corpus de publicações aderentes à pesquisa

Artigo	Título	Ano
60	<i>Predictive modeling with big data: Is bigger really better?</i>	2013
46	<i>Statistical process monitoring in the era of smart manufacturing.</i>	2017
44	<i>Performance assessment and uncertainty quantification of predictive models for smart manufacturing systems</i>	2015
09	<i>Critical analysis of Big Data challenges and analytical methods</i>	2016
43	<i>Data analytics using simulation for smart manufacturing</i>	2015
54	<i>Data-intensive applications, challenges, techniques and technologies: A survey on Big Data</i>	2014
45	<i>Considerations and recommendations for data availability for data analytics for manufacturing</i>	2015
08	<i>Visualizing Big Data with augmented and virtual reality: challenges and research agenda</i>	2015
53	<i>The rise of "big data" on cloud computing: Review and open research issues</i>	2015
42	<i>Predictive analytics model for power consumption in manufacturing</i>	2014
49	<i>Big-data-driven Anomaly Detection in Industry (4.0): an approach and a case study</i>	2016
02	<i>An Artificial Intelligence Computer System for Analysis of Social-Infrastructure Data</i>	2015
06	<i>Beyond the hype: Big data concepts, methods, and analytics</i>	2014
17	<i>Service innovation and smart analytics for Industry 4.0 and big data environment</i>	2014

Fonte: Os Autores, 2019.

A Tabela 8 demonstra os artigos presentes no *Corpus* de publicações aderentes divididos em grupos, os quais estão correlacionados aos temas de pesquisa abordados, sendo o Grupo I com foco em *Big Data*, o Grupo II em estatística multivariada e o Grupo III em Indústria 4.0.

Tabela 8: divisão dos grupos de pesquisa

Grupos	Temas abordados	Artigos
I	Análises, desafios e métodos analíticos praticados em <i>Big Data</i>	09 / 54 / 45 / 08 / 06 / 17
	Modelagem preditiva aplicada em <i>Big Data</i>	60 / 44 / 42 / 06
II	Aplicações de modelos estatísticos aplicados em <i>Big Data</i>	46 / 44 / 06
	Aplicações de abordagens multivariadas praticadas em <i>Big Data</i>	60 / 46 / 54 / 49 / 02
III	Análise de <i>Big Data</i> integrados as tecnologias aplicadas na Indústria 4.0	46 / 44 / 43 / 54 / 45 / 08 / 53 / 42 / 49 / 02 / 17

Fonte: Os Autores, 2019.

A Figura 9 representa a relação destes grupos em referência aos artigos em que neles estão inseridos.

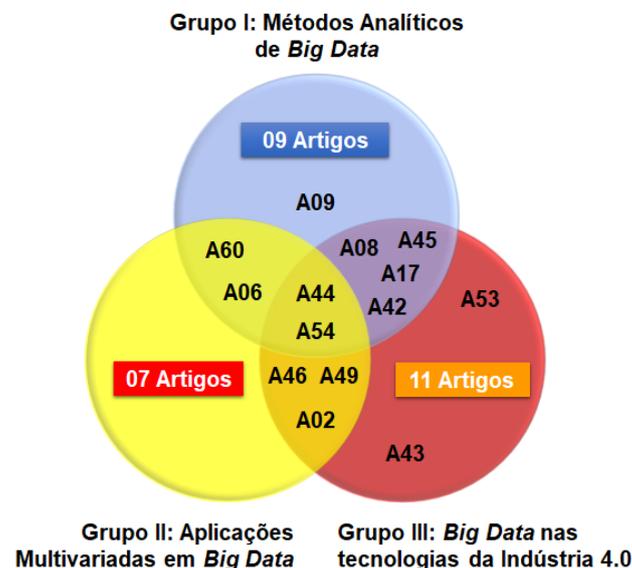


Figura 9. Diagrama de relação entre os artigos e os grupos de pesquisa

Fonte: Os Autores, 2019.

3.2. Síntese dos artigos componentes do Grupo I:

Critical analysis of Big Data challenges and analytical methods (Artigo 09): Contribuiu com uma análise crítica dos principais desafios e métodos analíticos aplicados em *Big Data*, como os modelos de análise preditiva (Sivarajah, Kamal, Irani, Weerakkody, 2017).

3.3. Síntese dos artigos componentes do Grupo II:

Não foram observados artigos que possuíam como tema principal apenas aplicações de abordagens estatísticas multivariadas em *Big Data*.

3.4. Síntese dos artigos componentes do Grupo III:

Data analytics using simulation for smart manufacturing (artigo 43): Análise de *Big*

Data utilizando simulações em fabricações inteligentes, por meio de coletas de dados originárias do chão de fábrica, os quais auxiliam na análise dos dados e no processo de tomada de decisão nos processos de fabricação e controle de produção (Shao, Shin, Jain, 2015).

The rise of "big data" on cloud computing: Review and open research issues (artigo 53): Demonstra a integração da computação em nuvem com *Big Data*, focando nos benefícios da análise de grandes dados no ambiente da nuvem, em relação ao armazenamento e processamento dos dados, além dos benefícios de disponibilidade, integridade, transformação, qualidade, privacidade, segurança e governança dos dados (Hashem, 2015).

3.5. Síntese dos artigos componentes dos Grupos I e II:

Considerations and recommendations for data availability for data analytics for manufacturing (artigo 45): Trata sobre a disponibilidade de dados, desde a qualidade, confiabilidade, eficiência e formatos em um processo de produção inteligente. Descreve as principais dificuldades que existem para o acesso e utilização de grande quantidade de dados (Libes, Shin, Woo, 2015).

Visualizing Big Data with augmented and virtual reality: challenges and research agenda (artigo 08): Apresenta diversas técnicas e ferramentas para visualização de *Big Data*, apresentando os principais métodos analíticos aplicados em sua análise, objetivando a comparação com a capacidade humana de interação e visualização (Olshannikova, Ometov, Koucheryavy, Olsson, 2015).

Predictive analytics model for power consumption in manufacturing (artigo 42): Modelo de análise preditiva de dados em um sistema inteligente de fabricação. Demonstra algoritmos e sistemas de redes neurais para aplicação e solução de análises de *Big Data* (Shin, Woo, Rachuri, 2014).

Service innovation and smart analytics for Industry 4.0 and big data environment (artigo 17): Explana as inovações aplicadas na Indústria 4.0, como a importância da análise de *Big Data* neste contexto, o qual requer a utilização de técnicas avançadas, de forma que os dados possam ser sistematicamente transformados em informações para a tomada de decisão (Lee, Kao, Yang, 2014).

3.6. Síntese dos artigos componentes dos Grupos I e III:

Predictive modeling with big data: Is bigger really better? (artigo 60): Aborda sobre a modelagem preditiva de *Big Data*, destacando como uma das mais reconhecidas técnicas de compreensão e uso dos dados. Demonstra que quanto maior a massa de dados, melhor se torna a análise, resultando em um melhor aproveitamento das informações disponíveis (Junqué de Fortuny, Martens, Provost, 2013).

Beyond the hype: Big data concepts, methods, and analytics (artigo 06): Apresenta amplas definições e conteúdo teórico sobre *Big Data*, destacando a análise de dados não estruturados, além de abordar sobre o método preditivo de análise e as abordagens estatísticas para inferir os dados (Gandomi & Haider, 2014).

3.7. Síntese dos artigos componentes dos Grupos II e III:

Statistical process monitoring in the era of smart manufacturing (artigo 46): Apresenta o monitoramento estatístico de processos multivariado como ferramenta para a análise de

Big Data nas manufaturas inteligentes, evidenciando nesta técnica como uma das fundamentais para o avanço da fabricação inteligente (He & Wang, 2018).

Big-data-driven Anomaly Detection in Industry (4.0): an approach and a case study (artigo 49): Apresenta uma abordagem para a gestão da qualidade com detecção em tempo real, e aplicação de análise de dados em uma manufatura inteligente. Destaca que as análises não podem ser desenvolvidas com estatística convencional, mas com a aplicação de abordagens multivariadas (Stojanovic, 2016).

An Artificial Intelligence Computer System for Analysis of Social-Infrastructure Data (artigo 02): Sistema de inteligência artificial para análise de *Big Data*, automatizando a análise dos dados. Demonstra aplicações multivariadas como o modelo de regressão para realizar previsões (Kudo, Akitomi, Moriwaki, 2015).

3.8. Síntese dos artigos componentes dos Grupos I, II e III:

Performance assessment and uncertainty quantification of predictive models for smart manufacturing systems (artigo 44): Quantificação de modelos preditivos para sistemas de produção inteligentes. Utiliza a teoria de aprendizagem estatística para a avaliação de desempenho e quantificação destes modelos quando aplicados em *Big Data* (Oneto, 2015).

Data-intensive applications, challenges, techniques and technologies: A survey on Big Data (artigo 54): Um estudo sobre as técnicas, tecnologias, desafios e aplicações empregados em *Big Data*. Demonstra várias técnicas aplicadas, além de citar a abordagem multivariada como técnica estatística empregada ao tema (Chen & Zhang, 2014).

4 CONCLUSÃO

Neste artigo pôde ser observado um significativo crescimento de publicações a respeito dos temas pesquisados sobre *Big Data*, análise multivariada e Indústria 4.0. Desta forma a técnica de revisão sistemática da literatura trouxe inúmeros benefícios para a pesquisa, pois devido ao grande número de informações e acesso as mesmas, a técnica possibilitou a seleção das publicações mais relevantes e impactantes.

Por meio da revisão de literatura foi possível identificar publicações que possuem aderência com a pesquisa, além de possibilitar a seleção criteriosa de um material bibliográfico, utilizando um raciocínio lógico para a sistematização da pesquisa.

Foi apresentada uma técnica de auxílio de forma estruturada para a seleção das publicações mais relevantes, tornando-a mais ágil, confiável e eficaz, desta forma justificando cientificamente a base teórica da pesquisa. Além disso, sintetizou-se o resultado baseado em evidências, e seus passos metódicos para serem possíveis de reprodução.

A aplicação da revisão sistemática na pesquisa possibilitou contribuir não apenas para os pesquisadores e comunidade científica e prática, as quais estudam o tema de análise de *Big Data*, mas também como um método para destacar as publicações reconhecidas e relevantes ao tema, com base em critérios julgados como relevantes pelos pesquisadores.

O processo estruturado realizado permitiu a identificação de um *Corpus* de 14 artigos relevantes e que de acordo com os pesquisadores estão alinhados ao tema, além de possuírem caráter e destaque científico.

Inserido nestes artigos, também foram identificados dois trabalhos que se sobressaíram em relação ao número de citações na classificação de relevância. É importante ressaltar que publicações com baixo número de citações não podem ser descartadas apenas

por esse fato, uma vez que podem existir artigos muito recentes e que ainda não tiveram a oportunidade de serem referenciados.

As abordagens multivariadas não estão explicitamente creditadas nos trabalhos, porém aparecem e são citadas como as ferramentas estatísticas mais apropriadas para a criação de algoritmos e desenvolvimento das análises de *Big Data*.

Pôde ser observado que o método de análise preditiva de dados é um dos mais utilizados no âmbito dos métodos analíticos aplicados em *Big Data*, com foco na melhoria da eficiência e da competitividade organizacional, sendo esse um possível modelo de aplicação em estudos futuros e mais aprofundados.

Por fim, de todos os artigos analisados, ainda são poucos os que focam em análise de *Big Data* com o detalhamento das técnicas multivariadas em suas aplicações.

A partir do exposto, como sugestão para trabalhos futuros indica-se avaliar os principais métodos para análise *Big Data* com estatística multivariada em contextos reais de produção para avaliar possíveis ganhos de produtividade e melhoria de processos industriais.

REFERÊNCIAS

- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information systems*, 47, 98-115.
- He, Q. P., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, 67, 35-43.
- Idc Analyze the Future (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. EMC Digital Universe with Research & Analysis.
- Jung, H. (2011). Cisco visual networking index: global mobile data traffic forecast update 2010–2015. Technical Report, Cisco Systems Inc. 2011. Available online: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html> (accessed on October 2019).
- Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: is bigger really better?. *Big Data*, 1(4), 215-226.
- Kudo, F., Akitomi, T., & Moriwaki, N. (2015, July). An artificial intelligence computer system for analysis of social-infrastructure data. In 2015 IEEE 17th Conference on Business Informatics (Vol. 1, pp. 85-89). IEEE.
- Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for industry 4.0 and big data environment. *Procedia Cirp*, 16, 3-8.
- Libes, D., Shin, S., & Woo, J. (2015, October). Considerations and recommendations for data availability for data analytics for manufacturing. In 2015 IEEE International Conference on Big Data (Big Data) (pp. 68-75). IEEE.

- Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing Big Data with augmented and virtual reality: challenges and research agenda. *Journal of Big Data*, 2(1), 22.
- Oneto, L., Orlandi, I., & Anguita, D. (2015, October). Performance assessment and uncertainty quantification of predictive models for smart manufacturing systems. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 1436-1445). IEEE.
- Shao, G., Shin, S. J., & Jain, S. (2014, December). Data analytics using simulation for smart manufacturing. In *Proceedings of the Winter Simulation Conference 2014* (pp. 2192-2203). IEEE.
- Shin, S. J., Woo, J., & Rachuri, S. (2014). Predictive analytics model for power consumption in manufacturing. *Procedia Cirp*, 15, 153-158.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- Stojanovic, L., Dinic, M., Stojanovic, N., & Stojadinovic, A. (2016, December). Big-data-driven anomaly detection in industry (4.0): An approach and a case study. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 1647-1652). IEEE.