

**JOSÉ AFONSO SANTOS SUGAHARA**

*Machine Learning e Data Science na indústria: aplicações e desafios*

Guaratinguetá - SP  
2020

**José Afonso Santos Sugahara**

*Machine Learning e Data Science na indústria: aplicações e desafios*

Trabalho de Graduação apresentado ao Conselho do Curso de Graduação em Engenharia de Produção da Faculdade de Engenharia do *Campus* de Guaratinguetá, Universidade Estadual Paulista, como parte dos requisitos para obtenção do diploma de Graduação em Engenharia de Produção Mecânica.

Orientador (a): Prof. Dr. José Roberto Dale Luche

Guaratinguetá - SP  
2020

S947m Sugahara, José Afonso Santos

Machine Learning e Data Science na indústria: aplicações e desafios /  
José Afonso Santos Sugahara – Guaratinguetá, 2020.

103 f : il.

Bibliografia: f. 97-103

Trabalho de Graduação em Engenharia de Produção Mecânica –  
Universidade Estadual Paulista, Faculdade de Engenharia de  
Guaratinguetá, 2020.

Orientador: Prof. Dr. José Roberto Dale Luche

1. Dados ligados. 2. Armazenamento de dados. 3. Sistemas de  
informação gerencial. I. Título

CDU 681.3.07

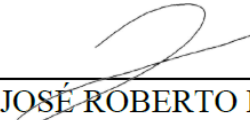
**JOSÉ AFONSO SANTOS SUGAHARA**

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO COMO  
PARTE DO REQUISITO PARA A OBTENÇÃO DO DIPLOMA DE  
“**ENGENHARIA DE PRODUÇÃO**”

APROVADO EM SUA FORMA FINAL PELO CONSELHO DE CURSO DE  
GRADUAÇÃO ENGENHARIA DE PRODUÇÃO

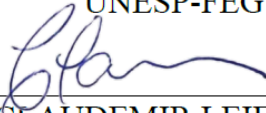
Prof. Dr. Grasielle Augusta Ferreira Nascimento  
Coordenador

**BANCA EXAMINADORA:**

  
\_\_\_\_\_  
Prof. Dr. JOSÉ ROBERTO DALE LUCHE  
Orientador/UNESP-FEG

Afsilva

\_\_\_\_\_  
Prof. Dr. ANEIRSON FRANCISCO DA SILVA  
UNESP-FEG  
UNESP-FEG

  
\_\_\_\_\_  
Prof. Dr. CLAUDEMIR LEIF TRAMARICO  
UNESP-FEG

Dedico este trabalho de modo especial à minha  
família.

## AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, fonte da vida e da graça.

Agradeço pela minha vida, pela minha inteligência, pela minha família e pelos meus amigos.

Ao meu orientador, Prof. Dr. José Roberto Dale Luche, que jamais deixou de me incentivar. Sem a sua orientação, dedicação e auxílio, o estudo, aqui apresentado, seria praticamente impossível.

Aos meus pais, Solange Maria da Conceição dos Santos e Alberto Tadayoshi Sugahara, que, apesar das dificuldades enfrentadas, sempre me incentivaram em relação aos meus estudos.

Às funcionárias da Biblioteca do *Campus* de Guaratinguetá pela dedicação, presteza e principalmente pela vontade de ajudar.

Aos funcionários da Faculdade de Engenharia do *Campus* de Guaratinguetá pela dedicação e alegria nos atendimentos.

“Só sabemos com exatidão quando sabemos pouco; à medida que vamos adquirindo conhecimento, instala-se a dúvida.”

Goethe

## RESUMO

Com o aumento da capacidade computacional, é possível que, agora, em ambientes industriais, sejam usadas novas ferramentas, visando à melhoria de desempenho global das indústrias. Neste projeto, foi realizada uma revisão sistemática de artigos que envolvem as ferramentas de *Machine Learning* e *Data Science* aplicadas na indústria. O artigo estabelece relações entre os termos citados no contexto industrial, apresenta exemplos práticos de uso das ferramentas citadas e potenciais desafios a serem abordados na sua implantação. Este estudo provê uma análise 5W1H (o quê, porque, onde, quando, quem e como) com os artigos mais citados, relacionando os termos *Machine Learning* e *Data Science* com a indústria. O trabalho demonstra que há uma escassez de trabalhos brasileiros que abordam os temas citados no contexto industrial, assim como um baixo número de trabalhos aplicados relacionando os temas.

**PALAVRAS-CHAVE:** *Machine learning*. Aprendizado de máquina. *Big data*. *Data science*. Indústria 4.0.



## **ABSTRACT**

With the improvement of computer processing power, it is now possible the use of new tools with the objective of improving the efficiency of industries. In this project, it has been done a systematic review of articles that uses the tools Machine Learning and Data Science in the industrial sector. This article makes connections between these two terms, presents practical examples and potential challenges when applying them. This study presents a 5W1H (what, why, where, when, who and how) with the most cited articles. These paper shows the lack of studies from Brazil and the low number of applied studies.

**KEYWORDS:** Machine learning. Aprendizaje de máquina. Big data. Data science. Industry 4.0.

## LISTA DE FIGURAS

Figura 1 – Gráfico de patentes de <i>Machine Learning</i> nos EUA.....	14
Figura 2 – Gráfico do Scopus mostrando documentos relacionados à <i>Machine Learning</i> e ... à <i>Data Science</i> .....	14
Figura 3 – Gráfico do Scopus mostrando documentos relacionados à <i>Big Data</i> e à <i>Data Science</i> por país. ....	15
Figura 4 – Gráfico do Scopus mostrando documentos relacionados à <i>Big Data</i> e à <i>Data Science</i> por setor.....	16
Figura 5 – Relação de <i>Machine Learning</i> com inteligência artificial.....	19
Figura 6 – Ilustração das 4 Revoluções Industriais.....	23

## LISTA DE ABREVIATURAS E SIGLAS

ASA	Associação Americana de Estatística
GDPR	<i>General Data Protection Regulation</i>
GTAI	<i>Germany Trade and Investment</i>
IoT	<i>Internet of Things</i>
LGPD	Lei Geral de Proteção de Dados
MIT	<i>Massachusetts Institute of Technology</i>
RFID	<i>Radio-Frequency Identification</i>
WD	<i>Wearable Devices</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	12
1.1	OBJETIVO.....	12
1.2	JUSTIFICATIVA.....	12
1.3	METODOLOGIA DE PESQUISA.....	16
1.4	ESTRUTURA DO TRABALHO.....	17
1.5	DELIMITAÇÃO .....	17
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	18
2.1	<i>DATA SCIENCE</i> .....	18
2.2	<i>MACHINE LEARNING</i> .....	19
<b>2.2.1</b>	<b>Aprendizado supervisionado</b> .....	<b>20</b>
<b>2.2.2</b>	<b>Aprendizado não supervisionado</b> .....	<b>21</b>
<b>2.2.3</b>	<b>Aprendizado por reforço</b> .....	<b>21</b>
2.3	<i>BIG DATA</i> .....	21
2.4	INTELIGÊNCIA ARTIFICIAL .....	22
2.5	INDÚSTRIA 4.0 .....	23
2.6	<i>INTERNET DAS COISAS PARA A INDÚSTRIA</i> .....	23
2.7	SISTEMAS CIBERFÍSICOS.....	24
<b>3</b>	<b>DESENVOLVIMENTO</b> .....	26
3.1	RELAÇÃO DE <i>DATA SCIENCE E MACHINE LEARNING</i> NA INDÚSTRIA.....	26
3.2	USO DE <i>DATA SCIENCE E MACHINE LEARNING</i> NA INDÚSTRIA.....	26
<b>3.2.1</b>	<b>Planejamento logístico</b> .....	<b>26</b>
<b>3.2.2</b>	<b>Manutenção preditiva</b> .....	<b>27</b>
<b>3.2.3</b>	<b>Tecnologias vestíveis</b> .....	<b>28</b>
<b>3.2.4</b>	<b>Veículos não tripulados</b> .....	<b>28</b>
<b>3.2.5</b>	<b>Fábricas inteligentes</b> .....	<b>29</b>
<b>4</b>	<b>DESAFIOS DA IMPLANTAÇÃO DE <i>MACHINE LEARNING E DATA SCIENCE</i> NA INDÚSTRIA</b> .....	<b>30</b>
4.1	PADRONIZAÇÃO .....	30
4.2	SEGURANÇA E PRIVACIDADE DOS DADOS .....	31
4.3	ARMAZENAMENTO E GESTÃO DE GRANDE MASSA DE DADOS.....	32
4.4	MÃO DE OBRA ESPECIALIZADA.....	32
4.5	CONSUMO DE ENERGIA .....	33

5	<b>CONSIDERAÇÕES FINAIS</b> .....	34
	REFERÊNCIAS .....	35
	BIBLIOGRAFIA CONSULTADA .....	38
	APÊNDICE A – QUADRO DE ARTIGOS <i>DATA SCIENCE X MACHINE LEARNING</i> NA INDÚSTRIA .....	39
	APÊNDICE B – QUADRO DE ARTIGOS <i>DATA SCIENCE</i> NA INDÚSTRIA .....	49

# 1 INTRODUÇÃO

A evolução da tecnologia possibilita o acesso a uma elevada quantidade de dados (*Big Data*), é essencial que empresas tirem proveito desses dados para realizar previsões de tendências futuras, como a demanda de um produto. A *Data Science* é uma forma de trabalhar esses dados e obter conhecimentos e *insights* que ajudem na tomada de decisões.

Por causa da grande quantidade de dados produzidos diariamente, o reconhecimento de padrões nesses dados pode ser uma vantagem para grandes empresas. Segundo Waller e Fawcett (2013), a *Big Data* carrega a oportunidade de mudar o modelo de negócios atual e as decisões do dia a dia juntamente com a análise de dados. Para Provost e Fawcett (2013), com uma vasta quantidade de dados disponível, empresas em quase todas as áreas estão focadas em tirar interpretar esses dados para obter vantagens competitivas.

Segundo um relatório internacional feito pelo *Massachusetts Institute of Technology* (MIT) em parceria com o *Google Cloud* (2017, *MIT Technology Review Insights*), 60% dos entrevistados já implementou estratégias de *Machine Learning* e quase um terço do público afirmou que está em uma etapa de maturidade com suas iniciativas.

*Machine Learning* vem sendo cada vez mais usado na melhoria da eficiência em várias empresas do mundo. Quais são as tendências mundiais nessa área?

Com esse contexto em mente, esse trabalho irá descrever o cenário internacional para obter uma impressão de como está o estado da arte na pesquisa sobre os temas estudados.

Questão de Pesquisa: Como a *Data Science* e a *Machine Learning* se relacionam com a indústria e quais são suas principais aplicações?

## 1.1 OBJETIVO

O objetivo geral desta pesquisa consiste em identificar tendências de estudos e aplicações de *Machine Learning*, identificar quais áreas possuem essa maior propensão de uso e divulgar qual o estado geral dos temas *Data Science* e *Machine Learning* na indústria. Investiga-se também quais métodos são mais usados e, a partir disso, construir uma tabela com a metodologia 5W1H com os resultados.

## 1.2 JUSTIFICATIVA

Avanços recentes em *Machine Learning* podem ser usados para solucionar diversos problemas, como o serviço ao cliente, a logística, as análises históricas médicas ou até mesmo as escritas de histórias, segundo pesquisas (MCKINSEY, 2016).

*Machine Learning* oferece diversos benefícios para organizações, como por exemplo a melhora da produção e da competição (LI *et al.*, 2015), permitindo, assim, analisar dados gerados por consumidores em vários formatos, como dados em vídeo ou em mídias sociais (MAKLAN *et al.*, 2015).

Um dos principais objetivos da *Data Science*, no contexto da indústria de manufatura, é, efetivamente, encontrar comportamentos anormais no maquinário, nas ferramentas ou nos processos de forma a prever danos e eventos críticos, que podem acarretar eventuais perdas econômicas ou problemas de segurança do trabalho (DIEZ-OLIVAN *et al.*, 2019).

Empresas como *Apple*, *Google*, *Facebook*, *Ebay* e *Amazon*, constantemente, fazem uso de informações digitais, como compras *online*, para melhorar sua tomada de decisão (POUSTTCHI; HUFENBACH, 2014).

*Machine Learning* e aproximações focadas em dados estão se tornando importantes em várias áreas. Suas aplicações bem sucedidas, as quais se destacam os classificadores de *spam* inteligentes, os sistemas de anúncios e a detecção de fraude, são conduzidas por dois fatores: o uso efetivo de modelos estatísticos e os sistemas de aprendizado escaláveis que aprendem o modelo de interesse por grandes bases de dados (CHEN; GUESTRIN, 2016).

Com o intuito de investigar a relevância e a concepção de artigos que envolvem a *Data Science* e a *Machine Learning*, foi gerado um relatório bibliométrico por meio da plataforma *Scopus*.

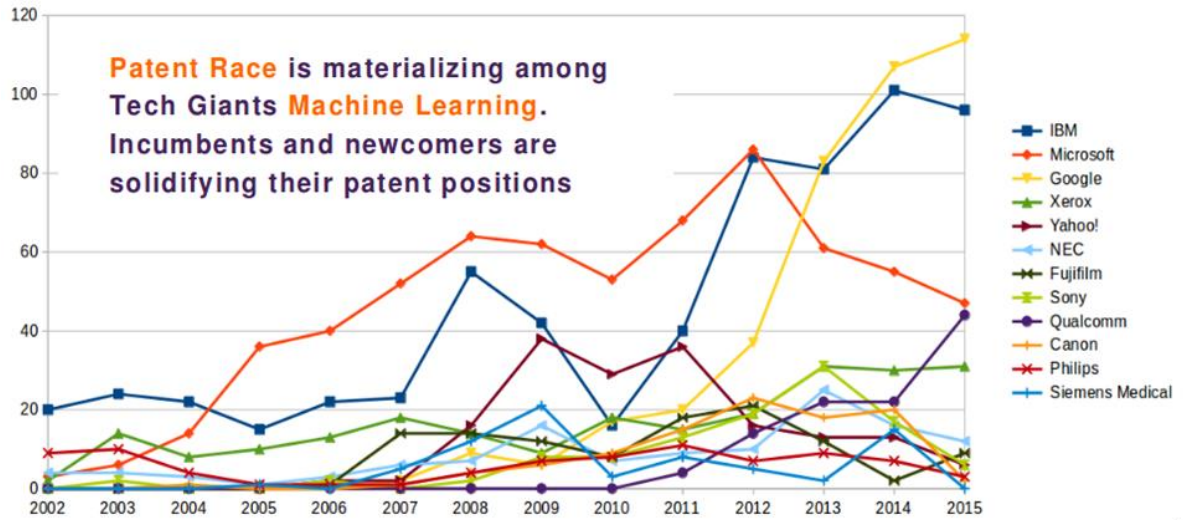
A pesquisa considerou o período de 2013-2019 e a busca foi feita no campo *Topic* (o qual é feito com base no título, no resumo, nas palavras-chave e nos *keyword plus*), e, também, filtrado com o tipo de texto analisado em *Article/Review*. Como palavras-chave para realizar a busca, foi utilizado “*Data Science*”, “*Machine Learning*” e “*Manufacturing*”.

Analisados os apêndices A e B, percebe-se uma tendência crescente tanto na publicação de artigos referentes ao tema *Machine Learning*, *Big Data* e *Manufacturing* quanto no número crescente de patentes na mesma área.

Na Figura 1 é apresentado o número de registro de patentes no tema *Machine Learning* entre várias empresas. Podemos perceber por essa figura o crescimento acelerado nos últimos anos, indicando assim uma alta demanda sobre esse assunto.

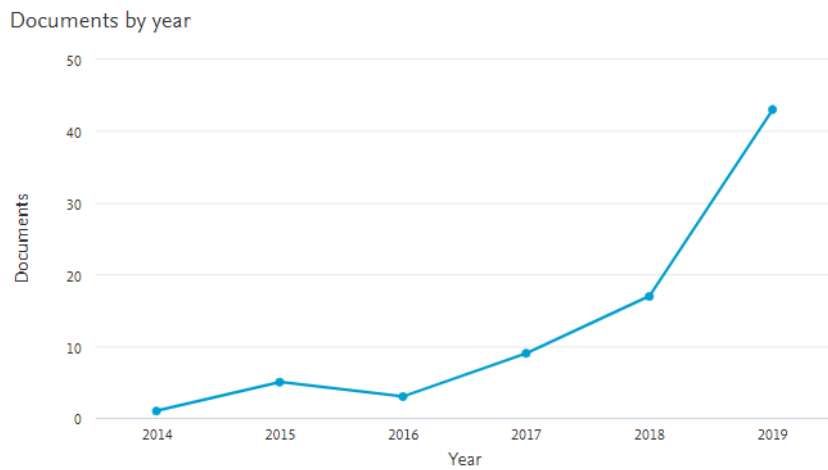
A figura 2 mostra o resultado da busca por artigos publicados no período de 2011-2019, somou-se um total de 78 publicações, pode-se evidenciar que é um tema que ganhou notoriedade em 2019, sendo mais da metade (43) dos artigos publicados somente em 2019.

Figura 1 – Gráfico de patentes de Machine Learning nos EUA.



Fonte: United States Patent and Trademark Office (2020).

Figura 2 – Número de documentos relacionados ao *Machine Learning* e ao *Data Science*.



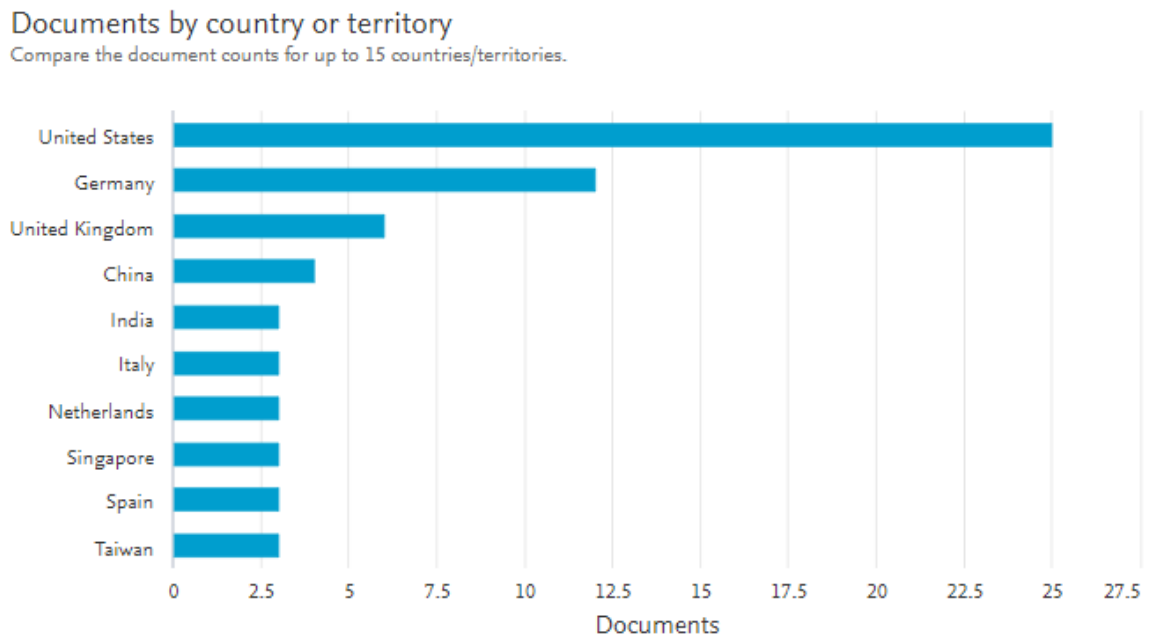
Fonte: Scopus (2020).

Outro dado relevante é o interesse de países onde são publicados trabalhos referentes aos termos “*Data Science*” e “*Machine Learning*”, de acordo com a bibliometria na base de dados da SCOPUS, que é um portal onde é disponibilizado diversos artigos do mundo todo. A pesquisa realizada com o termo “*Machine Learning*” e “*Data Science*”, limitando-se aos



últimos 9 anos, agrupou, em ordem decrescente, o número de documentos publicados por país. Como mostrado na Figura 3, nas primeiras posições, encontram-se os Estados Unidos (com cerca de 25 publicações), seguido pela Alemanha (com cerca de 12 publicações) e pelo Reino Unido (com pouco mais de 6 publicações). No período analisado, o Brasil possuía uma única publicação.

Figura 3 – Número de documentos relacionados à *Big Data* e à *Data Science* por país.

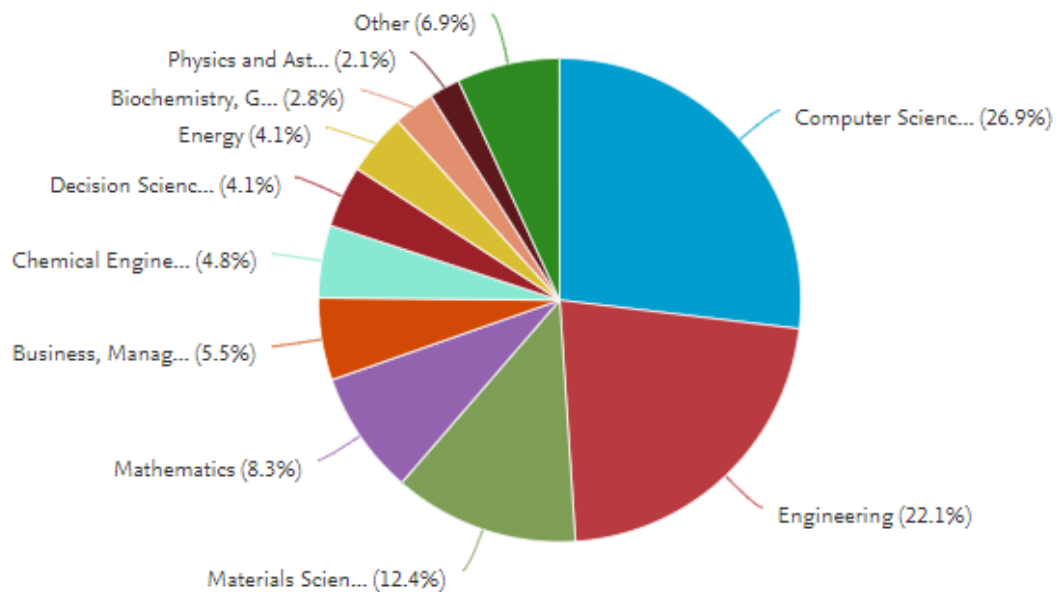


Fonte: Scopus (2020).

A *Data Science* e a *Machine Learning* são tecnologias que podem ser empregadas em um escopo de áreas bem amplo. A figura 4 mostra em quais áreas a *Internet Industrial* está sendo aplicada até o ano de 2019. Seguindo da seguinte forma:

- Ciência da Computação (26,9%)
- Engenharia (22,1%)
- Ciência dos Materiais (12,4%)
- Matemática (8,3%)
- Negócios (5,5%)
- Engenharia Química (4,8%)
- Ciência da Decisão (4,1%)
- Energia (4,1%)
- Bioquímica (2,9%)
- Física e Astronomia (2,1%)
- Outros (6,9%)

Figura 4 – Gráfico do Scopus mostrando documentos relacionados à *Big Data* e à *Data Science* por setor.



Fonte: Scopus (2020).

### 1.3 MÉTODO DE PESQUISA

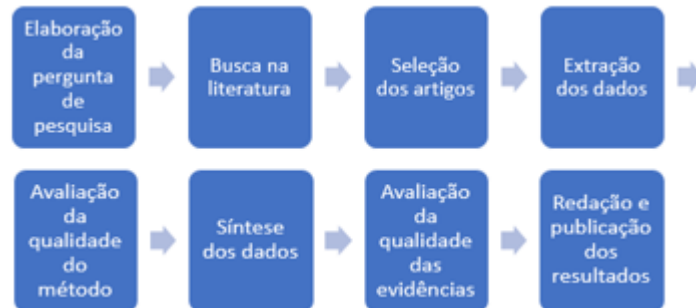
Este trabalho possui caráter exploratório, buscando na literatura as palavras-chave *Machine Learning* e *Data Science*, em busca de tendências e *insights* que sirvam de sugestões para trabalhos futuros na área. Trata-se de uma revisão sistemática.

Uma revisão sistemática, assim como outros tipos de estudo de revisão, é uma forma de pesquisa que utiliza como fonte de dados a literatura sobre determinado tema. Esse tipo de investigação disponibiliza um resumo das evidências relacionadas a uma estratégia de intervenção específica, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e síntese da informação selecionada (SAMPAIO; MANCINI, 2007).

Os métodos para a elaboração de revisões sistemáticas preveem: (1) elaboração da pergunta de pesquisa; (2) busca na literatura; (3) seleção dos artigos; (4) extração dos dados; (5) avaliação da qualidade metodológica; (6) síntese dos dados (metanálise); (7) avaliação da qualidade das evidências e (8) redação e publicação dos resultados (GALVÃO; PEREIRA, 2014).

A Figura 5 mostra o fluxograma da elaboração desse trabalho.

Figura 5 – Fluxograma do método utilizado para elaboração desse trabalho.



Fonte: Scopus (2014).

Pode-se, também, afirmar que esta monografia tem caráter explicativo e qualitativo, pois busca analisar e interpretar fatos para, assim, identificar sugestões de pesquisa e trabalhos futuros por métodos qualitativos.

#### 1.4 ESTRUTURA DO TRABALHO

No Capítulo 2, apresenta-se o embasamento teórico para os conceitos utilizados neste trabalho, que indicará algumas definições, ferramentas e técnicas fundamentais da *Machine Learning* e da *Data Science*.

No Capítulo 3, apresenta-se o estudo dos artigos selecionados para a pesquisa. Foram escolhidos os 10 artigos mais citados da base Scopus. Os termos de busca dos artigos foi *Machine Learning*, *Data Science* e *Industrial*. Foram somente considerados artigos aqueles publicados a partir do ano de 2013. A análise dos artigos apresenta-se por uma adaptação do método 5W1H (o que foi estudado o porquê do estudo [resultado], o seu ano, em que área ele é focado e como ele foi feito). O número de citações também será relevante nessa análise.

No Capítulo 4, então, são apresentadas as considerações finais do trabalho, contendo as análises finais, conclusões e sugestões de trabalhos futuros.

#### 1.5 DELIMITAÇÃO

O estudo em questão irá tratar das palavras-chave *Machine Learning*, *Data Science* e *Manufacturing*. Esta monografia se limitará a uma base científica, a Base Scopus.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 DATA SCIENCE

O termo *Data Science* tem origem em 1974 com a publicação do livro *Concise Survey of Computer Methods* (1974) e vem evoluindo até assumir a definição atual. Dhar (2013) ressalta que a *Data Science* difere da estatística clássica pelo fato de os dados usados serem heterogêneos e não estruturados, podendo ter origem de vários locais diferentes e estarem em formas variadas, como em textos, em imagens e em vídeos.

Cao (2016) faz um resumo da história da *Data Science*:

Jeff Wu, professor no Instituto de Tecnologia da Geórgia, em 1997, propõe a mudança do termo “estatística” para *Data Science*. O objetivo dessa mudança consistia em alterar o foco da estatística clássica de questões, tais como: coleta de dados, modelagem, análises, entendimento/resolução de problemas e tomada de decisão, para questões mais complexas, tais como: dados complexos, abordagem empírico-dialética e representação e exploração do conhecimento obtido dos resultados.

William S. Cleveland sugeriu, em 2001, alterar o campo de estatística para *Data Science* e juntar as áreas mais técnicas do campo da estatística com a computação, de modo a favorecer a parceria de estatísticos e cientistas da computação.

Leo Breiman sugeriu, em 2001, que era necessário se distanciar de uma dependência exclusiva de modelos estatísticos clássicos, alegando ser necessário adotar ferramentas mais diversificadas, como por exemplo a modelagem de algoritmos, sendo esta a ferramenta que consegue trabalhar com dados que não possuem uma relação aparente (aprendizado não supervisionado, aprendizado de reforço).

Em 2015, uma declaração sobre o papel da estatística na *Data Science* foi feita por líderes da Associação Americana de Estatística (ASA), dizendo que a Estatística e o *Machine Learning* têm papel fundamental na *Data Science*.

Observa-se, portanto, que a *Data Science* difere da estatística clássica por trabalhar dados menos estruturados, fazendo uso de bases de dados maiores e uso de ferramentas, como algoritmos iterativos e programação.

Temos algumas definições, pois, para *Data Science*:

- É a ciência de estudo de dados;

- É uma área multidisciplinar, que faz uso de computação e estatística para a criação de modelos que visam buscar conhecimento de dados tanto estruturados quanto não estruturados, visando ao apoio à tomada de decisão;

- É a integração de diversas áreas, como: Estatística, Informática, Computação, Comunicação, Sociologia e Gestão de dados (DHAR, 2013).

## 2.2 MACHINE LEARNING

Uma das primeiras definições de *Machine Learning* foi feita por Arthur Samuel, ex-engenheiro do *Massachusetts Institute of Technology*. Para ele, *Machine Learning* é

um campo de estudo que dá aos computadores a habilidade de aprender sem terem sido programados para tal”. Em seu artigo “*Some Studies in Machine Learning Using the Game of Checkers*” ele foi um dos primeiros a demonstrar o conceito de inteligência artificial (SAMUEL, 1959, p. 211).

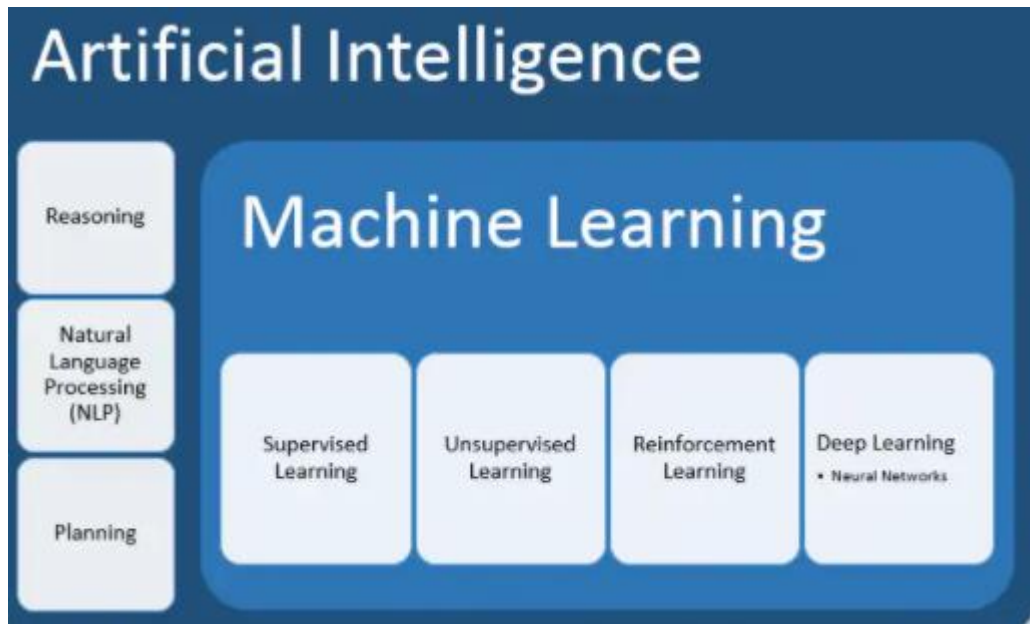
Para Irizarry (2020), *Machine Learning* é um termo guarda-chuva, que é usado para descrever um processo complexo que envolve não somente um cientista de dados que possui todo o conhecimento, mas um time de cientistas de dados com várias especializações distintas.

Já para Amershi *et al.* (2019), *Machine Learning* se refere aos métodos estatísticos de modelagem, que, recentemente, causaram euforia no mercado de *software* e de serviços digitais. A *Microsoft*, por exemplo, fez uso do *Machine Learning* para a criação do seu sistema de pesquisa, o *Bing*, e para a sua assistente virtual, a Cortana.

Podemos, nessa perspectiva, perceber que não há um conceito único de *Machine Learning*, e que definições e conceitos novos na área vêm surgindo a cada ano. Isso se deve, possivelmente, ao fato de o *Machine Learning* ser uma área que está sendo cada vez mais explorada, devido ao progresso de sistemas de computação e de armazenamento de grande quantidade de dados.

A Figura 6 mostra a relação de *Machine Learning* e Inteligência artificial, que embora conceitos parecidos, eles possuem algumas diferenciações. Podemos afirmar que *Machine Learning* sempre é Inteligência Artificial, porém a recíproca não é verdadeira, existem conceitos como o Processamento de Linguagem Natural (NLP) que se enquadra como inteligência artificial, porém não se enquadra no conceito de *Machine Learning*.

Figura 6 – Relação de *Machine Learning* com inteligência artificial.



Fonte: IBM (2020).

Com relação ao seu uso na indústria de manufatura, Wuest *et al.* (2016) comenta que esse setor enfrenta comportamentos cada vez mais caóticos, complexos e dinâmicos, e, para manter a demanda de produtos de alta qualidade de forma eficiente, é necessário o uso do *Machine Learning* para abordar tais desafios.

### 2.2.1 Aprendizado supervisionado

A maioria dos sistemas de *Machine Learning* utiliza o aprendizado supervisionado. Ele consiste em definir variáveis de entrada ( $x$ ) e variáveis de saída ( $y$ ) e utilizar um algoritmo para aprender a mapear funções baseadas nessas variáveis, de forma que  $y = f(x)$ .

O objetivo é estimar a função de tal forma que ela consiga prever resultados da variável de saída.

Chama-se aprendizado supervisionado pela forma como o algoritmo aprende baseado em um “professor”, no caso a base de dados utilizada. As variáveis de saída e de entrada já estão correlacionadas, o algoritmo realiza previsões iterativas e, então, é corrigido pela base usada para treinamento. As iterações cessam quando o algoritmo obtém o desempenho desejado.

Algoritmos supervisionados podem ser agrupados em dois grupos:

- Classificação: onde as variáveis utilizadas são discretas, como por exemplo cores (azul, preto, vermelho), sexo (masculino ou feminino) ou estado de saúde (saudável ou não saudável) e

- Regressão: onde as variáveis são contínuas, como por exemplo altura, peso ou custo.

### 2.2.2 Aprendizado não supervisionado

Aprendizado não supervisionado é quando há somente variáveis de entrada, não há variáveis de saída para o aprendizado. O objetivo desse tipo de algoritmo é modelar uma estrutura ou distribuição de dados para, assim, obter algum padrão.

Chama-se aprendizado não supervisionado, pois, diferente do modelo supervisionado, não há associação prévia de variáveis, então fica a cargo do algoritmo agrupar os dados de uma forma lógica.

Algoritmos não supervisionados podem ser agrupados em dois grupos:

- Clusterização: um problema onde os dados são agrupados em grupos. Por exemplo, agrupar consumidores baseado no seu histórico de compras e
- Associação: onde se procura alguma regra lógica que descreva parte dos seus dados. Por exemplo, pessoas que compram X tendem também a comprar Y.

### 2.2.3 Aprendizado por reforço

O aprendizado por reforço se baseia nas perspectivas psicológicas e neurocientíficas do comportamento animal. Assim como humanos e outros animais, o aprendizado por reforço procura imitar o aprendizado obtido com as métricas de reforço e de punição existentes nos cérebros de humanos e de outros animais (MNIH *et al.*, 2015).

No aprendizado por reforço, é definido um problema e, por tentativa e erro, o algoritmo tenta solucioná-lo. Dependendo de como é feita essa solução, o algoritmo muda sua função de punição e recompensa, convergindo, então, para a resposta ótima do modelo.

Um exemplo análogo seria o adestramento de animais, incentivando comportamentos desejados e punindo comportamentos indesejados, de forma a melhor treinar o animal.

## 2.3 *BIG DATA*

Segundo o relatório da *International Data Corporation*: a quantidade de dados que serão criados nos próximos 3 anos será superior à quantidade criada nos últimos 30 anos.

É possível perceber o crescimento da quantidade de dados existentes. Com isso, percebe-se também o potencial dentro da análise dessa vasta quantidade de dados, para isso, é necessária a integração de diversas tecnologias, um exemplo delas é a *Big Data*.

Não há um consenso sobre o termo *Big Data*, mas podemos resumir seu significado em alguns pontos:

- Dados que são tão abundantes e complexos que modelos tradicionais de análise não são capazes de usar todos eles;
- Trabalho com dados muitas vezes não estruturados e massivos e
- Segundo a organização *Apache Hadoop*, *Big Data* é conjunto de dados que não podem ser capturados, geridos e testados por computadores normais eficientemente.

Dentro dessas definições, existe a característica em comum de que *Big Data* seria o trabalho com dados que são tão complexos a ponto de modelos tradicionais, sejam estatísticos ou computacionais, não serem capazes de analisar de forma satisfatória.

Grover e Kar (artigo 3) definem 5 atributos à *Big Data*, conhecidos como os 5 Vs da *Big Data*:

- Volume: se relaciona com a grande quantidade de dados gerados;
- Velocidade: se relaciona à rapidez em que os dados são gerados;
- Variedade: diz respeito ao ser comum o uso de dados estruturados e não estruturados;
- Valor: se refere à utilidade que pode ser obtida desses dados e
- Veracidade: em relação aos dados serem consistentes e representarem de forma confiável a realidade.

## 2.4 INTELIGÊNCIA ARTIFICIAL

Em 1950, o matemático britânico Alan Turing, considerado como o pai da inteligência artificial, afirmou que um dia haveria máquinas que poderiam duplicar a inteligência humana em cada aspecto existente.

Segundo Kalogirou (2003), a inteligência artificial é, de forma geral, uma forma de máquinas se adaptarem à solução de problemas do mesmo modo que a inteligência humana. O termo inteligência artificial também pode ser usado para sistemas que operam tarefas mais complexas.

Já para Barr e Feigenbaum (2014), inteligência artificial é uma parte das ciências da computação que se dedica a projetar sistemas inteligentes de computação, por exemplo,



características que estão associadas à inteligência humana, como análise de comportamentos, linguagem, aprendizado e solução de problemas.

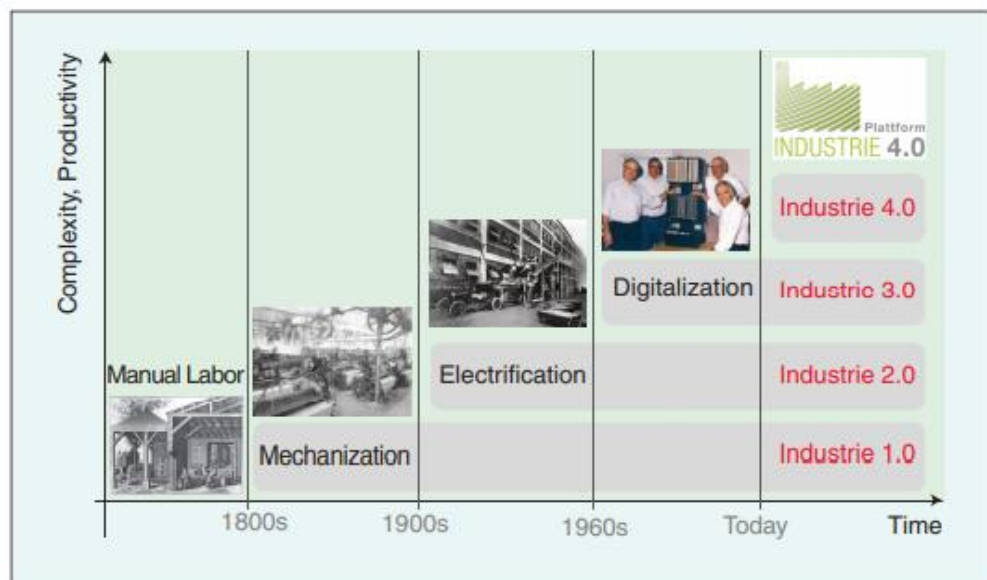
## 2.5 INDÚSTRIA 4.0

O termo indústria 4.0 foi cunhado por três engenheiros chamados Henning Kagermann (físico e fundador da empresa SAP), Wolfgang Wahlster (professor de inteligência artificial) e Wolf-Dieter Lukas (físico e oficial sênior do Ministério da Educação e Pesquisa da Alemanha).

Indústria 4.0 é um termo usado para se referir a produtos, processos e tecnologias que fazem parte do fluxo de valor de uma empresa. Ela se refere às fábricas inteligentes, onde sistemas ciberfísicos monitoram os processos dentro do fluxo e se comunicam entre si e com os funcionários em tempo real, portanto, Indústria 4.0 significa mudar o processo de produção até um ponto em que tudo está sistemicamente interligado e controlado (CIFFOLILLI; MUSCIO, 2018).

A Figura 6 apresenta um pequeno resumo de no que constitui cada revolução industrial anterior a indústria 4.0. Sendo a primeira revolução constituída da mecanização do trabalho, a segunda do uso de máquinas elétricas, da terceira do uso de máquinas digitais até finalmente chegarmos à quarta revolução industrial.

Figura 5 – Ilustração das 4 Revoluções Industriais.



Fonte: Drath e Horch (2020).

## 2.6 INTERNET DAS COISAS PARA A INDÚSTRIA

O objetivo da Indústria 4.0 é a integração de todos os sistemas de uma indústria, com o objetivo de automatizar, melhorar a confiança e o controle do sistema como um todo, que implica no desenvolvimento da *Internet das Coisas (Internet of Things – IoT)* para a indústria. Mais especificamente, a Indústria 4.0 é voltada para permitir a conexão de tudo, em qualquer lugar e a qualquer hora, no contexto da manufatura, de forma a melhorar a produtividade, a eficiência e a segurança (HANSONG *et al.*, 2018). Há algum problema nessas referências!

*Internet das Coisas* para a indústria visa a conectar e a sincronizar um grande número de máquinas por meio de ferramentas de *software*. Essas máquinas podem ser de robôs industriais até de pequenos sensores nas esteiras de produção.

Para Lin *et al.* (2017), a *Internet das Coisas* pode ser separada em três camadas:

- Camada de percepção, também chamada de camada dos sensores, interage com dispositivos físicos por meio de dispositivos inteligentes (RFID, sensores ou atuadores). Seu principal objetivo é fazer a conexão dos instrumentos à rede e medir, coletar e processar as informações obtidas;

- Camada de rede pode ser considerada a camada do meio na arquitetura da *Internet das Coisas*. Ela é usada para receber as informações recebidas da camada dos sensores e as transmitir para a rede central. Ela pode ser considerada a camada mais importante, pois integra diversas tecnologias (Centrais de rede, *Wi-Fi*, *Bluetooth*, *Gateways*);

- Camada de aplicação, que recebe a informação da camada de rede e usa essa informação para gerar serviços ou operações. Por exemplo, guardar os dados recebidos das camadas anteriores ou analisar esses dados para possíveis previsões.

Alguma das ferramentas que são usadas na *Internet das Coisas* para a indústria é o uso de tecnologia de *Radio-Frequency Identification (RFID)*, o uso de códigos de barra, *Qr Code* e de sistemas inteligentes.

## 2.7 SISTEMAS CIBERFÍSICOS

Para Lin *et al.* (2017), Sistemas Ciberfísicos são sistemas que podem, de maneira eficiente, integrar componentes físicos e digitais ao integrar tecnologias de computação e comunicação, visando a mudar a interação entre pessoas e os mundos físicos e digitais. Os Sistemas Ciberfísicos têm ênfase na interação entre os componentes digitais e físicos, e almeja tornar o controle de componentes físicos seguro, eficiente e eficaz por meio de componentes digitais.

Sistemas ciberfísicos são definidos como tecnologias disruptivas que visam gerir sistemas interconectados entre suas formas físicas e suas capacidades computacionais.

Para Lee *et al.* (2015), os sistemas Ciberfísicos consistem basicamente em dois sistemas:

- Uma conexão que permite a obtenção de dados em tempo real do mundo físico e o recebimento de informações vindas do meio digital, e
- Gestão de dados inteligentes, análise e capacidade computacional para construir o ciberespaço (LEE; BAGHERI; KAO, 2014).

As seguintes áreas foram identificadas como promissoras para o uso de Sistemas Ciberfísicos pela organização americana (OKOLIE, 2018):

- Agricultura;
- Controle de construções;
- Defesa;
- Energia;
- Saúde;
- Manufatura e Indústria;
- Sociedade e
- Transporte.

### 3 DESENVOLVIMENTO

Neste capítulo, será tratada a relação dos termos *Big Data* e *Machine Learning* com a indústria de manufatura, de acordo com os artigos pesquisados, possíveis aplicações, e, em seguida, desafios na aplicação de tais tecnologias na indústria.

#### 3.1 RELAÇÃO DE *DATA SCIENCE* E *MACHINE LEARNING* NA INDÚSTRIA

Com o crescimento da quantidade de dados disponíveis e o aumento do poder de processamento dos computadores, surge oportunidade e necessidade de estudar esses dados, aqui definidos como *Big Data*, com os métodos e ferramentas disponíveis da *Data Science*. Uma das ferramentas que vem ganhado bastante notoriedade é a *Machine Learning*, que consiste em métodos estatísticos capazes de organizar e obter informações de grandes volumes de dados.

Na indústria de manufatura, os termos se aplicam de forma a integrar melhor todo o sistema de produção, bem como possibilitar acesso aos dados de produção em tempo real, possíveis falhas nos sistemas, previsão do melhor momento para manutenções preditivas, controle logístico etc.

Essa integração de sistemas faz parte dos pilares da Indústria 4.0, que consiste em fazer uso de novas tecnologias, como por exemplo sensores e sistemas mais integrados de *Enterprise Resource Planning* (ERP), para uma maior eficiência das operações realizadas nas fábricas.

#### 3.2 USO DE *DATA SCIENCE* E *MACHINE LEARNING* NA INDÚSTRIA

Há diversos usos práticos de *Data Science* e de *Machine Learning* na indústria, aqui, iremos citar alguns deles.

##### 3.2.1 Planejamento logístico

Varejo, finanças, ciências, indústria e segurança são algumas áreas que dependem de previsões providas por técnicas, como *Machine Learning* (MACKENZIE, 2016).

Para Li *et al.* (2017), sistemas logísticos inteligentes podem proporcionar diminuição de custos logísticos, melhora na eficiência da gestão dos armazéns e, por fim, redução do custo logístico de ponta a ponta. Setores logísticos terão uma maior eficiência e nível de automação

por meio de tecnologias como a *Internet das Coisas* para a indústria, *Radio Frequency Identification* (RFID), robótica etc. Logística auto-organizável é um exemplo de logística inteligente que provê respostas tempestivas às situações inesperadas no processo de manufatura.

É essencial para um bom controle de custos e planejamento da produção prever as demandas de, por exemplo, matéria prima, uso do maquinário, operários e manutenção. Com a *Machine Learning*, é possível criar modelos que fazem previsões da demanda futura dos consumidores e, baseado nisso, adaptar a produção de forma a produzir somente o necessário. Como, normalmente, há acesso aos dados de consumo de produtos, esse problema pode ser abordado por algoritmos de regressão, como regressão logística.

Com o uso de ferramentas da indústria 4.0, é também possível integrar melhor os sistemas da produção, a ponto de, caso haja gargalos no meio produtivo, o sistema como um todo se adapte para que não haja interrupção na produção.

### 3.2.2 Manutenção preditiva

A manutenção preditiva é uma área essencial que deve ser sempre bem endereçada em indústrias. A estratégia padrão para o planejamento de manutenções preditivas é a utilização de indicadores simples, como horas de operação, número de produtos processados ou tempo desde a última manutenção (KHAN; TUROWSKI, 2016).

Pode-se fazer uso de sensores para monitorar em quais condições, por exemplo de temperatura ou vibração, as máquinas performam mais eficientemente, e quais condições indicam potencial falha. Para isso, é necessário tanto a obtenção de dados de desempenho do maquinário, como formas de modelar tais comportamentos, visando à obtenção de padrões.

Havendo uma grande massa de dados disponíveis para análise, é recomendado que se use modelos supervisionados de *Machine Learning*, onde tais modelos procuram relacionar variáveis X com o estado ou desempenho do maquinário Y (*output*). Quando se trata de *outputs* contínuos, é requerido modelos de regressão, como por exemplo regressão logística e regressão polinomial. Quando se trata de *outputs* categóricos, o problema deve ser tratado com modelos de classificação, como por exemplo árvore de decisões e *support vector machine*.

Tradicionalmente, modelos de classificação são mais apropriados para a determinação de equipamentos com grande ou baixo potencial de falha, porém, esses modelos não se adaptam bem quando é necessário definir se deve ou não ser realizada a manutenção, já modelos contínuos conseguem simular melhor o nível de saúde do maquinário, sendo assim possível

uma melhor classificação de quando seria a melhor oportunidade para a sua manutenção (SUSTO *et al.*, 2019).

Uma das empresas que faz uso desse tipo de manutenção preditiva é a *ABB Robotics*, ela faz uso de sensores interconectados para monitorar a necessidade de manutenção de seus maquinários e sinalizar uma manutenção antes mesmo que haja uma falha.

### 3.2.3 Tecnologias vestíveis

Tecnologias vestíveis são objetos como cartões, óculos ou monitores que são portáteis e estão integrados com todo um sistema de produção. Esses objetos fazem uso de *Machine Learning* e procuram facilitar as operações de uma fábrica, fornecendo informação de uma maneira simples. Um exemplo seria o uso de realidade aumentada; ela se define como a integração de informações virtuais com informações físicas através de visores ou telas portáteis. Os RFIDs (Radio Frequency Identification), junto com sensores portáteis, podem também ser utilizados em ferramentas ou em lotes de peças para facilitar a sua localização e a sua checagem.

A empresa Airbus faz uso de óculos inteligentes para reduzir erros e aumentar a segurança no chão de fábrica. Eles auxiliam no entendimento de desenhos industriais complexos e na conversão de medidas imperiais para medidas métricas, ajudando, assim, no posicionamento dos equipamentos nos aviões.

A empresa Caterpillar faz uso de realidade aumentada para dar aos operários de maquinário pesado, como retroescavadeiras, uma visão geral de todo o equipamento, dos níveis de combustível, até a possível necessidade de troca de filtros de ar. Com esse mesmo equipamento, é possível mandar instruções simples de como trocar os filtros, por exemplo.

### 3.2.4 Veículos não tripulados

Veículos não tripulados fazem uso de comunicação entre eles mesmos e o *layout* da fábrica, de forma a poder transitar nela. Isso pode ser utilizado para o transporte de produtos em armazéns, por exemplo, aumentando, esse modo, a quantidade de material transportado e a diminuição de mão de obra humana.

Segundo a revista Business Insider, um exemplo de uso dessa tecnologia é a Amazon, que usa diversos robôs Kiva interconectados por *Wi-Fi* nos seus armazéns para transporte de produtos dentro de seus espaços. A Amazon também está testando o uso de *Drones* para a entrega de mercadorias de seus clientes diretamente em suas casas.

### 3.2.5 Fábricas inteligentes

Para Khan e Turowski (2016), o termo fábricas inteligentes é frequentemente usado para se referir aos sistemas de produção da próxima geração. Avanços na tecnologia da informação possibilitaram a evolução de tais sistemas. Para se manter relevantes no mercado, é de interesse das empresas fazerem uso de tais avanços tecnológicos, almejando resolver desafios e atender seus consumidores de formas que não eram possíveis anteriormente.

O objetivo de fábricas inteligentes é prover produtos e serviços sustentáveis para consumidores, de acordo com seu comportamento, pedido e *feedback*. Com base em cadeias e fornecedores inteligentes, as fábricas inteligentes podem organizar o seu processo industrial, com a ajuda das tecnologias de *Big Data* e computação em nuvem (LI *et al.* 2017).

## 4 DESAFIOS DA IMPLANTAÇÃO DE MACHINE LEARNING E DATA SCIENCE NA INDÚSTRIA

Embora, como foi citado no capítulo anterior, haja diversas utilidades no uso de *Machine Learning* e de *Data Science* na indústria, para o melhor aproveitamento de seus potenciais, é necessário abordar alguns desafios com relação à sua implantação.

Esse capítulo abordará os principais desafios citados nos artigos estudados, assim como possíveis abordagens para tais desafios.

Segundo Lu Yang (2017), para garantir que os processos dentro da indústria, no contexto de *Machine Learning* e *Data Science*, sejam usados de maneira eficiente, são necessários oito princípios: acessibilidade, *multilingualism*, segurança, privacidade, subsídios, uso de padrões abertos ao público, *softwares “open source”* e decisões multilaterais.

Acessibilidade diz respeito ao oferecimento de oportunidades iguais de acesso aos participantes do processo, sem nenhum tipo de discriminação. *Multilingualism* tem relação ao uso de vários tipos de linguagem de programação para o transporte eficiente de informação e conhecimento dentro do sistema produtivo. Já Segurança se refere às políticas de uso, que devem ser bem definidas para manter o processo e as informações em segurança e confiáveis. Por seu lado, Privacidade se refere ao uso dos dados com o devido consentimento de seus participantes. Enquanto Subsídios remete à necessidade de investimentos para guiar essa transformação. Por sua vez, o uso de padrões, assim como o uso de *softwares “open source”*, busca a melhor interação entre os processos usados. Decisões multilaterais, por fim, permitem a interoperabilidade das ferramentas citadas, por meio do alinhamento de requisitos entre os participantes do processo.

### 4.1 PADRONIZAÇÃO

A indústria 4.0 tem o potencial de se tornar a linguagem global da produção (*Germany Trade and Investment – GTAI 2014*). Cada processo usado na indústria 4.0 integra outras tecnologias para abordar problemas na indústria. Assim, a devida criação de padrões é extremamente importante (KUSIAK, 2017). Um esforço global é necessário para padronizar devidamente os processos da indústria 4.0 de forma a seguir os objetivos estratégicos da indústria 4.0. Por exemplo, é necessário um conjunto de normas técnicas para que a rede de diferentes fábricas e empresas possam estar interligadas (LI; LIAN, 2018).



Para Khan e Turowski (2016), os processos envolvidos na manufatura de produtos podem envolver uma ou mais fábricas, a integração e a padronização desses processos acabam sendo um desafio por usar diversas tecnologias, interfaces, padrões, métodos e características únicas de cada meio.

Li *et al.* (2017) cita que na indústria 4.0, diversos componentes e sistemas estão integrados pela mesma rede, isso resulta em controles mais inteligentes e uma produção mais colaborativa. Porém isso acarreta também no desafio de integrar sistemas diferentes com criticidades diferentes. Para assegurar confiança e disponibilidade na produção industrial, essas funções com diversos níveis de criticidade devem ser isoladas de forma que as de menor criticidade não interfiram com as de maior.

A padronização de processos é necessária para o desenvolvimento e a popularização da *Internet* das Coisas. Ela diminui as barreiras de entrada tanto para os provedores desses serviços quanto para os seus clientes. Ela também permite a interoperabilidade de aplicações e sistemas, tornando, assim, a qualidade dos serviços e dos produtos prestados é maior. Um processo minucioso de padronização e de boa coordenação é essencial para que sistemas de países diferentes possam trocar informações entre si (XU *et al.*, 2014).

## 4.2 SEGURANÇA E PRIVACIDADE DOS DADOS

A *Internet* das Coisas para a indústria acaba sendo vulnerável a ataques por alguns motivos. Primeiro, a maior parte dos seus componentes tem pouco ou nenhum monitoramento, isso acaba os tornando vulneráveis a ataques físicos às ferramentas. Segundo, a maior parte dos componentes usados são caracterizados por baixo uso energético e computacional, por isso, não é possível implementar protocolos mais complexos de segurança (ATZORI; IERA; MORABITO, 2010).

Li *et al.* (2017) afirma que a garantia da segurança dos sistemas é essencial para sistemas industriais. Por causa dos vários sensores e máquinas conectados que formam o sistema de produção, torna-se complexo assegurar a autenticidade e a confidencialidade dos dados coletados no sistema. A característica autogerenciável desses sistemas abre margem para potenciais ataques que visam controlar o sistema, e, além desse caso, é um desafio manter a privacidade e a segurança em sistemas de armazenamento de dados em nuvem.

Para Boyes *et al.* (2018), segurança é imprescindível na prevenção e na minimização de potenciais danos aos funcionários, ao maquinário e ao meio ambiente. A criação de sistemas de

segurança acaba sendo um grande desafio para a indústria 4.0 e, por outro lado, essencial para o uso consciente e responsável dos seus instrumentos.

Podemos citar exemplos de legislações, que já estão em vigor no mundo, procurando cumprir essa demanda por segurança e privacidade, como a Regulação Geral de Proteção de Dados (*General Data Protection Regulation - GDPR*) de origem europeia, e a Lei Geral de Proteção de Dados (LGPD) de origem brasileira<sup>1</sup>. Essas leis visam proteger o uso dos dados dos consumidores e dos cidadãos, de uma forma que fique claro o consentimento do uso desses dados por parte dos seus donos, e regulamentar parâmetros para proteção desses dados,

#### 4.3 ARMAZENAMENTO E GESTÃO DE GRANDE MASSA DE DADOS

No mundo atual movido a dados, há diversas formas de eles serem gerados. No contexto industrial, dados são gerados e coletados por meio de vários sensores. Citando apenas alguns como dados de processo, dados de produto e dados de qualidade, vemos que há o risco de um não armazenamento apropriado deles.

Khan e Turowski (2016) propõem que, para trabalhar essa massa de dados, são necessários novos métodos para armazenamento, processamento e gestão. Novos algoritmos, modelos, produtos e métodos de visualização são necessários para o uso e benefício dos dados. Por exemplo, há o problema comum as quais várias soluções intermediárias são usadas para a gestão de dados de uma companhia, partindo de armazenamento e troca de dados em impressões, *e-mails*, tabelas de *Excel* até o uso de bases sem parametrização única entre departamentos.

Um exemplo que Khan e Turowski (2016) cita é o armazenamento de bases redundantes em vários locais de uma mesma companhia, com formatação diferente e somente leves diferenças. Tal comportamento aumenta a redundância dos dados, inconsistências e interpretações para as mesmas informações.

#### 4.4 MÃO DE OBRA ESPECIALIZADA

É comum, no cenário atual, empresas enfrentarem problemas na obtenção de mão de obra especializada. Um dos fatores dessa lacuna é o envelhecimento da população. Pessoas que trabalharam anos na produção estão se aposentando e levando consigo conhecimento e experiência obtidos nas suas carreiras, outro problema é a tentativa de manter os funcionários

---

<sup>1</sup> Lei nº 13.709, de 14 de agosto de 2018 vigente desde 18/09/2020.

que já estão na empresa. E, por fim, como boa parte da força de trabalho é composta de pessoas que estão menos abertas a mudanças, a introdução de novas tecnologias acaba sendo um desafio (KHAN; TUROWSKI, 2016).

#### 4.5 CONSUMO DE ENERGIA

Xu *et al.* (2014) cita que, conforme as técnicas de *Machine Learning* e *Data Science* na indústria avançam, o consumo de energia aumenta, devido aos bilhões de sensores que se comunicam por redes sem fio. Assim, tornando uso da energia um limitador para o uso abrangente dessas tecnologias. A economia de energia deve se tornar uma prioridade no planejamento de novos sensores. Há uma necessidade de desenvolver técnicas eficientes que visem reduzir esse consumo.

Há diversos desafios energéticos a serem abordados por empresas que desejam aplicar as técnicas de *Machine Learning* e *Data Science* na indústria. De um lado, o mundo enfrenta desafios energéticos graças ao crescimento populacional, ao aumento da população urbana, e à baixa eficiência energética. Porém, do outro lado, o aumento da consciência ecológica, e do preço da energia e as mudanças no comportamento do consumidor são forças motoras para a maior exigência de produtos que consumam o mínimo de energia possível (LI *et al.* 2017).

## 5 CONSIDERAÇÕES FINAIS

Esta revisão sistemática identificou qual o papel dos temas *Data Science* e *Machine Learning* são compreendidos e abordados no setor industrial. Dentre essas compreensões, foram encontrados alguns termos que estão diretamente relacionados, como *Big Data*, Inteligência Artificial, Indústria 4.0, *Internet* das Coisas e Sistemas Ciberfísicos.

Esta pesquisa também expôs quais áreas do conhecimento estão relacionadas, os seus usos práticos e potenciais desafios a serem encontrados ao abordar o tema no contexto industrial. É interessante ressaltar que alguns desses termos estudados possuem várias definições, mostrando, talvez, que uma melhor fundamentação científica dos termos possa padronizá-los de forma melhor.

Evidencia-se o vasto potencial de utilização das ferramentas de *Data Science* e *Machine Learning* na indústria, visando à melhoria de processos, seja de forma a gastar menos recursos ou de forma a deixar os processos mais dinâmicos.

A maioria dos artigos analisados são revisões bibliográficas, apontando uma área potencial para trabalhos futuros, focada em trabalhos aplicados. Também se observou uma escassez de trabalhos feitos em universidades brasileiras que abordam os temas *Machine Learning* e *Data Science*. Isso também demonstra outra área potencial para a criação de trabalhos futuros. Tal revisão acaba por ser relevante por tratar de um tema ainda pouco abordado em estudos brasileiros.

Destaca-se que esta revisão apresenta limitações, pois a seleção de artigos compromete somente a base de dados Scopus, e o trabalho aborda poucos trabalhos aplicados.

Esta revisão sistemática produziu tabelas que visam auxiliar trabalhos futuros em possíveis abordagens de temas relevantes nas áreas de estudo abordadas.

## REFERÊNCIAS

- AMERSHI, S. *et al.* Software engineering for machine learning: a case study. *In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING*, 41., 2019, Montreal. **Anais [...]** Montreal: ICSE, 2019. p. 291-300.
- ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: a survey. **Computer Networks**, v.54, p.2787-2805, 2010.
- BARR, A.; FEIGENBAUM, E. A. **The handbook of artificial intelligence**. Oxford: Butterworth-Heinemann, 2014. v.1.
- BOYES, H. *et al.* The industrial internet of things (IIoT): an analysis framework. **Computers in Industry**, v.101, p.1-12, 2018.
- CAO, L. Data Science: a comprehensive overview. **ACM Computing Surveys**, v.1, n.1, p.1-42, jan. 2016.
- CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. *In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 22., 2016, San Francisco. **Anais [...]** San Francisco: ACM SIGKDD, 2016. p. 785-794.
- CIFFOLILLI, A.; MUSCIO, A. Industry 4.0: national and regional comparative advantages in key enabling technologies. **European Planning Studies**, v.26, n.12, p.2323-2343, 2018.
- DHAR, V. Data science and prediction. **Communications of the ACM**, v.56, n.12, p. 64-73, 2013.
- DIEZ-OLIVAN, A. *et al.* Data fusion and machine learning for industrial prognosis: trends and perspectives towards Industry 4.0. **Information Fusion**, v.50, p.92-111, oct.2019.
- DRATH, R.; HORCH, A. Industrie 4.0: hit or hype? **Industrial Electronics Magazine**, p.56-58, jun. 2014.
- GALVÃO, T. F.; PEREIRA, M. G. Revisões sistemáticas da literatura: passos para sua elaboração. **Epidemiologia e Serviços de Saúde**, Brasília, v.23, n.1, p. 183-184, jan./mar. 2014.
- GERMANY TRADE AND INVESTMENT. **Industrie 4.0: smart manufacturing for the future**. Berlin: GTAI, 2014.
- GROVER, P.; KAR, A. K. Big data analytics: a review on theoretical contributions and tools used in literature. **Global Journal of Flexible Systems Management**, v.18, n.3, p.203-229, sep. 2017.
- HANSONG, X. *et al.* A survey on industrial internet of things: a cyber-physical systems perspective. **Access**, v.6, p.78238-78259, dec. 2018.
- IRIZARRY, R. A. The role of academia in data Science education. **Harvard Data Science Review**, v.21, p.1-8, 2020.

KALOGIROU, S. A. Artificial intelligence for the modeling and control of combustion processes: a review. **Progress in Energy and Combustion Science**, v.29, p.515-566, 2003.

KHAN, A.; TUROWSKI, K. A perspective on industry 4.0: from challenges to opportunities in production systems. *In: INTERNATIONAL CONFERENCE ON INTERNET OF THINGS AND BIG DATA*, 2016, Rome. **Anais [...]** Rome: IoTBD, 2016. p.441-448.

KIM, E. **Business insider**. Disponível em: <https://www.businessinsider.com/kiva-robots-save-money-for-amazon-2016-6>. Acesso em: 16 dez. 2020.

KUSIAK, A. Smart manufacturing must embrace big data. **Nature**, v.5, n.544, p. 23-25, apr. 2017.

LEE, J.; BAGHERI, B.; KAO, H. A cyber-physical systems architecture for industry 4.0: based manufacturing systems. **Manufacturing Letters**, v.3, p.18-23, 2015.

LEE, J.; BAGHERI, B.; KAO, H. Recent advances and trends of cyber-physical systems and big data analytics in industrial informatics. *In: INTERNATIONAL CONFERENCE ON INDUSTRIAL INFORMATICS*, 12., 2014, United States. **Proceedings [...]** United States: INDIN, 2014. p.1-6.

LI, D. X.; LIAN, D. Big data for cyber physical systems in industry 4.0: a survey. **Journal Enterprise Information Systems**, v.13, n.2, p.148-169, 2018.

LI, J. *et al.* Industrial internet: a survey on the enabling technologies, applications and challenges. **IEEE Communications Surveys & Tutorials**, v.19, n.3, p.1504-1526, 2017.

LIN, J. *et al.* A survey on internet of things: architecture, enabling technologies, security and privacy, and applications. **IEEE Internet of Things Journal**, v.4, n.5, p.1125-1142, oct. 2017.

LU, Y. Industry 4.0: a survey on technologies, applications and open research issues. **Journal of Industrial Information Integration**, v.6, p.1-10, 2017.

MACKENZIE, A. The production of prediction: what does machine learning want? **European Journal of Cultural Studies**, v.18, n.4-5, p. 429-445, 2015.

MAKLAN, S.; PEPPARD, J.; KLAUS, P. Show me the Money: improving our understanding of how organizations generate return from technology-led marketing change. **European Journal of Marketing**, v.49, n.3-4, p.561-595, 2015.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. Aprendizado de máquina (ML): o novo campo de testes para a vantagem competitiva. **MIT Technology Review**, v. 1, n. 10, p.1-10, 2017.

MCKINSEY GLOBAL INSTITUTE. **The age of analytics: competing in a data-driven world**. [S.l.]: Mckinsey & Company, 2016.

MNIH, V. *et al.* Human-level control through deep reinforcement learning. **Nature**, v.5, n.518, p.529-533, feb. 2015.

POUSTTCHI, K.; HUFENBACH, Y. Engineering the value network of the customer interface and marketing in the data-rich retail environment. **International Journal of Electronic Commerce**, v.18, n.4, p.17-41, 2014.

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, feb. 2013.

ROLLINS, J. **Foundational methodology for data science**. Disponível em: <https://www.ibm.com/br-pt/analytics/machine-learning>. Acesso em: 30 out. 2020.

SAMPAIO, R. F.; MANCINI, M. C. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia**, São Carlos, v.11, n.1, p.83-89, jan./fev. 2007.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal**, p.211-229, jul. 1959.

SCOPUS. **Relação de estudos sobre Machine Learning**. Disponível em: [https://www.scopus.com/search/form.uri?display=basic&zone=h eader&origin=](https://www.scopus.com/search/form.uri?display=basic&zone=h+eader&origin=) Acesso em: 27 nov. 2020.

SUSTO, G. A. *et al.* Machine learning for predictive maintenance: a multiple classifier approach. **IEEE Transactions on Industrial Informatics**, v. 11, n. 3, p.1-9, 2019.

UNITED STATES PATENT AND TRADEMARK OFFICE. **Inventing AI**: tracing the diffusion of artificial intelligence with U.S. patents. United States: USPTO, 2020.

WALLER, M. A.; FAWCETT, S. E. Data science, predictive analytics and big data: a revolution that will transform supply chain design and management. **Journal of Business Logistics**, v.34, n.2, p. 77-84, 2013.

WUEST, T. *et al.* Machine learning in manufacturing: advantages, challenges, and applications. **Production & Manufacturing Research**, v.4, n.1, p.23-45, 2016.

XU, L. D.; HE, W.; LI, S. Internet of things in industries: a survey. **IEEE Transactions on Industrial Informatics**, v.10, n.4, p.2233-2243, nov. 2014.

**BIBLIOGRAFIA CONSULTADA**

NAUR, P. **Concise survey of computer methods**. New York: Petrocelli Books, 1974.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 3, n. 3, p. 210-229, 1988.



**APÊNDICE A – Quadro de artigos *Data Science* x *Machine Learning* na indústria**

(continua)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally?	Venkat Venkatasubramanian	Revisar o progresso feito até então na área de <i>Machine Learning</i> na Química, destacando esforços passados para identificar lições aprendidas. Dessas lições, identificar oportunidades presentes e futuras na engenharia química. Rever marcos históricos dessa indústria.	2019	Indústria Química	Revisão Bibliográfica	Um resumo histórico da situação da <i>Machine Learning</i> na Indústria Química, com sugestão de trabalhos futuros e potenciais desafios na sua implantação,	Trabalhos nas áreas de criação de materiais, processos e diagnóstico de falhas.	78

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0	Alberto Diez-Olivana, Javier Del Ser, Diego Galar, Basilio Sierra	Revisar o progresso feito na área de fusão de dados e <i>Machine Learning</i> na indústria, com ênfase na identificação de tendências de pesquisa, oportunidades e desafios.	2018	Química, Ferrovias, Manufatura, Energia Eólica, Agricultura, Engenharia Civil, Marítima, Aeroespacial, Energia, Petróleo e Gás	Revisão Bibliográfica	Um quadro resumo com diversas áreas, algoritmos e aplicações de <i>Machine Learning</i> no contexto industrial.	Uso da tabela fornecida para explorar possíveis oportunidades de uso da <i>Machine Learning</i> .	77

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature	Purva Grover, Arpan Kumar Kar	Analisar artigos sobre <i>Big Data</i> publicados em periódicos de Gestão de Negócios, subsequentemente, a categorização dos estudos foi feita para o entendimento dos casos.	2017	Saúde, Gestão de empresas, Finanças, Industrial, Administração Pública, <i>Consumer Discretionary</i>	Revisão Bibliográfica	Foi encontrado que o principal uso foi em análise de mídias sociais, mineração de texto, marketing e <i>Supply Chain</i> .	Uso de sistemas de <i>Machine Learning</i> em áreas ainda embrionárias como Saúde e Finanças.	54

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Extracting Knowledge from Data through Catalysis Informatics	Andrew J. Medford, M. Ross Kunz, Sarah M. Ewing, Tammie Borders, and Rebecca Fushimi.	Revisar o passado e o futuro da área de Química Computacional usando modelos de <i>Machine Learning</i> .	2018	Indústria Química	Modelagem	Criação de uma <i>Chemical Master Equation</i> e Modelos <i>Microkinetic</i> como representação dos catalisadores, que ajudam no entendimento e na criação de novos materiais catalisadores,	Integração entre modelos computacionais e dados experimentais, Avanços em <i>Microkinetic Modeling</i> , e Desenvolvimento de <i>softwares "open source"</i> voltados para a área de química computacional.	50

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Software Engineering for Machine Learning: A Case Study	Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., Zimmermann, T.	O estudo de como é feita a construção de <i>softwares</i> com foco no consumidor na <i>Microsoft</i> .	2019	Engenharia de <i>Software</i>	Estudo de Caso	Processos como descobrir e gerir aplicações de <i>Machine Learning</i> são mais complexos que em outras áreas da Engenharia de <i>Software</i> . Adaptação e reuso de modelos de <i>Machine Learning</i> requerem muitas vezes habilidades que não estão presentes nos times da <i>Microsoft</i> . Componentes de Inteligência Artificial são mais complexos na sua utilização que componentes de <i>software</i> tradicionais.	Expansão do modelo de estudo aplicado para outras empresas relevantes na área, como <i>Google</i> e <i>Facebook</i> .	34

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Data-Driven Evolutionary Optimization: An Overview and Case Studies localization linkages in high-contrast composite materials	Yaochu Jin , Fellow, IEEE, Handing Wang , Member, IEEE, Tinkle Chugh , Dan Guo , Kaisa Miettinen	Modelagem de cinco estudos de casos com computação evolucionária.	2019	Metalurgia, Saúde, Aeronáutica e Agricultura	Modelagem	Uma lista de algoritmos e sugestões de seus usos em casos práticos para problemas onde funções analíticas objetivas não estão disponíveis.	Uso dos algoritmos criados no artigo em outras indústrias.	32

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Machine learning approaches for elastic localization linkages in high-contrast composite materials	Ruoqian Liu, Yuksel C. Yabansu, Ankit Agrawal, Surya R. Kalidindi, Alok N. Choudhary	Explorar experimentos e estratégias de mineração de dados para estabelecer modelos estatísticos, para a captura de relações de <i>elastic localization</i> em compostos de alto contraste.	2015		Modelagem	Foi demonstrado que um modelo feito com <i>features</i> de descrição simples junto com <i>features</i> construídas é capaz de melhorar o desempenho. Para casos em que foi usado um número reduzido de descritores por meio do ranqueamento deles, o desempenho ficou ainda melhor.		31

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Data-Driven Materials Science: Status, Challenges, and Perspectives	Lauri Himanen, Amber Geurts, Adam Stuart Foster, and Patrick Rinke	Analisar o histórico da ciência dos materiais guiado por dados, desde os primórdios da <i>open science</i> até a grande expansão de infraestrutura de dados de materiais.	2019	Ciência dos Materiais	Revisão Bibliográfica	Um resumo do estado atual da ciência dos materiais guiado por dados, assim como seus desafios.	Usando as conclusões do artigo, criar uma maior integração entre o setor industrial e o setor acadêmico para a criação de trabalhos mais robustos.	30



**Quadro de artigos *Data Science x Machine Learning* na indústria**

(continuação)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
SolarData: An R package for easy access of publicly available solar datasets Discovery in Chemical Engineering	Dazhi Yang	Facilitação do acesso à base com dados solares	2018	Energia	Modelagem	Pacote em linguagem R, SolarData, que facilita o acesso à base de dados solares.	Uso da biblioteca criada em trabalhos relacionados à energia solar.	26

**Quadro de artigos *Data Science x Machine Learning* na indústria**

(conclusão)

Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
Towards a data science toolbox for industrial analytics applications	Christoph M. Flath, Nikolai Stein Sean T. Osis, Reed Ferber	Criar ferramentas que conciliem a pesquisa em <i>Machine Learning</i> e seu uso prático em indústrias.	2017	Manufatura	Modelagem, Estudo de caso	Um guia de 5 passos para uma boa implantação de modelos de <i>Machine Learning</i> na indústria de manufatura.	Uso da ferramenta criada em outras indústrias, como a química ou saúde.	21

Fonte: Produção do Próprio Autor (2020)

**APÊNDICE B – Quadro de artigos *Data Science* na indústria**

(continua)

Número	Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
1	Smart Manufacturing	Andrew Kusiak	Discorrer sobre a origem, o estado presente e o futuro da manufatura inteligente.	2018	Meio Acadêmico	Revisão bibliográfica	Uma divisão em seis pilares dos conceitos importantes para a manufatura inteligente: Processos e Tecnologia de Manufatura, materiais, dados, engenharia preditiva, sustentabilidade e compartilhamento de recursos e redes.	Uso dos pilares abordados no trabalho em trabalhos aplicados.	321

**Quadro de artigos *Data Science* na indústria**

(continuação)

Número	Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
2	Big data for cyber physical systems in industry 4.0: a survey	Li Da Xu & Lian Duan	Trazer mais atenção aos temas <i>Big Data</i> e Sistemas Ciberfísicos, sugestões de trabalhos futuros	2018	Meio Acadêmico	Survey	Um agrupamento de artigos em duas categorias: robustes e inteligência, assim como os objetivos da <i>Big Data</i> e suas aplicações nas categorias citadas	Uso dos conceitos abordados em companhias e organizações distintas de setores diferentes	128

**Quadro de artigos Data Science na indústria**

(continuação)

Número	Nome	Autoria	O que	Quando	Onde	Como	Por que (Resultados)	Sugestões	Número de Citações
3	<a href="#">Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement</a> SupplyChain	Matthew A. Waller & Stanley E. Fawcett	Desenvolvimento de um modelo que explica a função das análises preditivas na teoria de desenvolvimento de processos.	2013	Meio Acadêmico	Revisão bibliográfica	Uma matriz dois por dois que busca enquadrar os métodos utilizados para predição de forma a torná-los mais simples de serem estudados.	Revisão bibliográfica de estudos usando a matriz criada no artigo.	99

**Quadro de artigos *Data Science* na indústria**

(continuação)

Número	Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
4	Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0	Alberto Diez-Olivana, Javier Del Ser, Diego Galar & Basilio Sierra	Prover uma pesquisa sobre o uso de <i>Machine Learning</i> para prognósticos industriais, com ênfase na identificação de tendências, nichos e oportunidades na área.	2019	Setor Industrial	Pesquisa	Uma discussão de três categorias que compõem a integração entre algoritmos de <i>Machine Learning</i> e a indústria 4.0, sendo eles os prognosticos descritivos, preditivos e prescritivos.	Aplicação das categorias criadas no artigo em trabalhos aplicados.	85

**Quadro de artigos *Data Science* na indústria**

(conclusão)

Número	Nome	Autoria	O quê	Quando	Onde	Como	Porque (Resultados)	Sugestões	Número de Citações
5	Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters	Ankit Agrawal, Parijat D Deshpande, Ahmet Cecen, Gautham P Basavarsu, Alok N Choudhary & Surya R Kalidindi	Descrever as ferramentas disponíveis para prever o efeito da fadiga no aço.	2014	Setor Industrial	Modelagem	Sugestões de modelos de regressão capazes de prever falhas por fadiga em aços.	Criação de modelos para outros tipos de materiais.	84

Fonte: Produção do Próprio Autor (2020)