



Programa de Pós-Graduação em

**Computação Aplicada**

Doutorado Acadêmico

Helder Moreira Arruda

Odisseu: Um Modelo para Serviços Inteligentes na Indústria 4.0  
baseado em Análise de Históricos de Contextos

São Leopoldo, 2024

UNIVERSIDADE DO VALE DO RIO DOS SINOS — UNISINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO APLICADA  
NÍVEL DOUTORADO

HELDER MOREIRA ARRUDA

**ODISSEU: UM MODELO PARA SERVIÇOS INTELIGENTES NA INDÚSTRIA 4.0  
BASEADO EM ANÁLISE DE HISTÓRICOS DE CONTEXTOS**

São Leopoldo  
2024

Helder Moreira Arruda

**ODISSEU: UM MODELO PARA SERVIÇOS INTELIGENTES NA INDÚSTRIA 4.0  
BASEADO EM ANÁLISE DE HISTÓRICOS DE CONTEXTOS**

Tese apresentada como requisito para a  
obtenção do título de Doutor, pelo  
Programa de Pós-Graduação em  
Computação Aplicada da Universidade do  
Vale do Rio dos Sinos - UNISINOS

Orientador:  
Prof. Dr. Jorge Luis Victória Barbosa

Coorientador:  
Prof. Dr. Rafael Kunst

São Leopoldo  
2024

A779o Arruda, Helder Moreira.  
Odisseu : um modelo para serviços inteligentes na indústria 4.0 baseado em análise de históricos de contextos / Helder Moreira Arruda. – 2024.  
118 f. : il. ; 30 cm.

Tese (doutorado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2024.  
“Orientador: Prof. Dr. Jorge Luis Victória Barbosa  
Coorientador; Prof. Dr. Rafael Kunst”

1. Aprendizado de máquina. 2. Ciência de dados. 3. Histórico de contextos. 4. Indústria 4.0. 5. Ontologia. 6. Serviços inteligentes. I. Título.

CDU 004.4

Dados Internacionais de Catalogação na Publicação (CIP)  
(Bibliotecária: Silvana Dornelles Studzinski – CRB 10/2524)

ATA DE BANCA EXAMINADORA DE TESE DE DOUTORADO Nº 02/2024

Aluna: Helder Moreira Arruda

Título da Tese: “ODISSEU: UM MODELO PARA SERVIÇOS INTELIGENTES NA INDÚSTRIA 4.0 BASEADO EM ANÁLISE DE HISTÓRICOS DE CONTEXTOS”

Banca: Prof. Dr. Jorge Luis Victória Barbosa (Orientador) - Unisinos  
Prof. Dr. Rafael Kunst (Coorientador) - Unisinos  
Prof. Dr. Rodrigo Righi (Avaliador) - Unisinos  
Prof. Dr. Fabiano Hessel (Avaliador Externo) - PUCRS  
Prof. Dr. Gustavo Pessin (Avaliador Externo) ITV

Aos quatro dias do mês de abril do ano de 2024, às 14h reuniu-se a Comissão Examinadora de Defesa de Tese composta pelos professores: Prof. Dr. Jorge Luis Victória Barbosa (Orientador) – Unisinos (por webconferência); Prof. Dr. Rafael Kunst (Coorientador) – Unisinos (por webconferência); Prof. Dr. Rodrigo Righi (Avaliador) – Unisinos (por webconferência); Prof. Dr. Fabiano Hessel (Avaliador Externo) - PUCRS (por webconferência) e Prof. Dr. Gustavo Pessin (Avaliador Externo) ITV (por webconferência) para analisar e avaliar a Tese apresentada pelo(a) aluno(a) **Helder Moreira Arruda** (por webconferência).

Considerações da Banca:

Após a apresentação realizada pelo aluno, os professores se reuniram para avaliação do trabalho. Os professores confirmaram os aperfeiçoamentos indicados durante a banca, as quais estão gravados no vídeo e deverão constar na versão final da tese.

Ocorreu alteração do título? ( X ) Não ( ) Sim  
Indicar o novo título:

A Banca Examinadora, em cumprimento ao requisito exigido para a obtenção do Título de Doutor em Computação Aplicada, julga esta tese:

( X ) APROVADA ( ) REPROVADA

Conforme Artigo 75 do Regimento do Programa o texto definitivo, com aprovação do Orientador, deverá ser entregue no prazo máximo de sessenta (60) dias após a defesa. O resultado da banca é de consenso entre os avaliadores. A emissão do Diploma está condicionada a entrega da versão final da Tese.

São Leopoldo, 04 de abril de 2024.



Orientador - Prof. Dr. Jorge Luis Victória  
Barbosa

*À minha esposa.*

## **AGRADECIMENTOS**

Aos meus professores e orientadores Jorge Luis Victória Barbosa e Rafael Kunst, pelo tempo, paciência, confiança, dedicação e oportunidade ofertada.

À HT Micron Semicondutores, pelo suporte e concessão de suas instalações para o trabalho de campo, em especial ao Giovani Cheuiche Pesenti, ao Elvis Fernandes Bugs e aos voluntários que participaram da coleta de dados.

À Universidade do Vale do Rio dos Sinos (Unisinos), pela bolsa de estudos concedida e por todo o suporte ao longo do curso.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela bolsa de Doutorado Acadêmico para Inovação (DAI), concedida através da chamada pública número 23/2018.

Aos colegas do Laboratório de Computação Móvel (Mobilab).

Ao professor Leandro Miletto Tonetto, pelo suporte nas discussões iniciais sobre o conceito de bem-estar subjetivo e suas aplicações.

À minha esposa Manuella, por todo o apoio, compreensão e pelo empréstimo de seus ouvidos ao longo de tantas horas.

Aos meus pais Heliodoro (*in pectore*) e Virginia, por todas as formas de educação.

A Deus, por me iluminar as veredas e sugerir a melhor forma de percorrê-las.

*"Then you better start swimmin'.  
Or you'll sink like a stone.  
For the times they are a-changin'."*  
— ROBERT ALLEN ZIMMERMAN



## RESUMO

A Quarta Revolução Industrial, também chamada de Indústria 4.0, vem alavancando vários campos da computação nos dias de hoje. A Indústria 4.0 compreende tarefas automatizadas na indústria manufatureira, que geram grandes quantidades de dados obtidos por meio de sensores. Esses dados contribuem para a interpretação das operações industriais em prol de tomadas de decisão gerenciais e técnicas. A Ciência de Dados suporta essa interpretação devido a significativos avanços tecnológicos, particularmente métodos de processamento de dados e ferramentas de *software*. Nesse sentido, essa tese apresenta um modelo intitulado Odisseu que tem o foco no suporte ao desenvolvimento de serviços inteligentes voltados para a Indústria 4.0, usando históricos de contextos, que representam dados de uma determinada entidade ao longo de um certo período de tempo. O modelo propõe uma ontologia que atua como elo entre métodos de ciência de dados e serviços inteligentes. Em comparação a outros modelos, o Odisseu busca suprir uma lacuna que envolve o acompanhamento dos dados desde a entrada até o armazenamento em formato de históricos de contextos, além de propor uma ontologia e um modelo para suporte genérico a serviços inteligentes na indústria. Para avaliação do modelo são propostos dois serviços inteligentes, sendo o primeiro voltado para a localização de pessoas em uma indústria e o segundo com o objetivo de estimar o bem-estar subjetivo dos funcionários. Os serviços usaram dados de beacons móveis e fixos, sinais vitais como pulso de volume sanguíneo e atividade eletrodérmica, além de dados originados por questões de autorrelato voltadas para o bem-estar. O serviço de localização alcançou 100% de acurácia tanto com o algoritmo *Random Forest* quanto com o algoritmo *Multilayer Perceptron*. O serviço de bem-estar atingiu o melhor desempenho com o algoritmo *Random Forest*, chegando a 74% de acurácia.

**Palavras-chave:** Indústria 4.0. Históricos de Contextos. Ciência de Dados. Aprendizado de Máquina. Serviços Inteligentes. Ontologia.

## ABSTRACT

The Fourth Industrial Revolution, also called Industry 4.0, has been leveraging many fields of computing today. Industry 4.0 comprises automated tasks in the manufacturing industry that generate large amounts of data obtained through sensors. These data contribute to the interpretation of industrial operations in favor of managerial and technical decision-making. Data Science supports this interpretation due to significant technological advances, particularly data processing methods and software tools. In this sense, this thesis presents a model entitled Odisseu that focuses on supporting the development of intelligent services aimed at Industry 4.0, using context histories, which represent data from a given entity over a certain period of time. The model proposes an ontology that acts as a link between data science methods and smart services. Compared to other models, Odisseu seeks to fill a gap that involves monitoring data from input to storage in context histories format, in addition to proposing an ontology and a model for generic support to intelligent services in the industry. To evaluate the model, two intelligent services are proposed, the first aimed at locating people in an industry and the second aiming to estimate the subjective well-being of employees. The services used data from mobile and fixed beacons, vital signs such as blood volume pulse and electrodermal activity, as well as data from self-report questions focused on well-being. The location service achieved 100% accuracy with both the *Random Forest* and *Multilayer Perceptron* algorithms. The well-being service achieved the best performance with the *Random Forest* algorithm, reaching 74% accuracy.

**Keywords:** Industry 4.0. Context Histories. Data Science. Machine Learning. Intelligent Services. Ontology.

## LISTA DE FIGURAS

Figura 1 – As quatro fases da revolução industrial. . . . .	14
Figura 2 – Foco da pesquisa. . . . .	17
Figura 3 – Exemplo de informação e histórico de contextos. . . . .	21
Figura 4 – As quatro etapas da revisão sistemática. . . . .	28
Figura 5 – Resultados da busca inicial da revisão sistemática. . . . .	31
Figura 6 – Diagrama do banco de dados criado para a revisão sistemática. . .	33
Figura 7 – Artigos por base de dados e os critérios de exclusão da revisão sistemática. . . . .	35
Figura 8 – Métodos de ciência de dados da revisão sistemática agrupados por ano. . . . .	39
Figura 9 – Ferramentas de software da revisão sistemática agrupadas por ano.	41
Figura 10–Quantidade de artigos por ano e banco dados da revisão sistemática.	44
Figura 11–Número de publicações presentes na revisão sistemática por ano. .	45
Figura 12–Tipos de publicação da revisão sistemática por ano. . . . .	45
Figura 13–Os eixos principais da taxonomia da revisão sistemática. . . . .	46
Figura 14–O eixo Métodos da taxonomia da revisão sistemática. . . . .	47
Figura 15–O eixo Aprendizado de Máquina da taxonomia da revisão sistemática. . . . .	48
Figura 16–O eixo Ferramentas da taxonomia da revisão sistemática. . . . .	51
Figura 17–Arquitetura do modelo Odisseu. . . . .	59
Figura 18–Diagrama do banco de dados dos repositórios do modelo Odisseu. .	62
Figura 19–Ontologia do modelo Odisseu. . . . .	64
Figura 20–Planta baixa da sala limpa da HT Micron Semicondutores, onde foram instalados os beacons fixos (retângulos azuis) e definidos os pontos de interesse (círculos verdes). . . . .	75
Figura 21 – Resultados da comparação entre os modelos RF e MLP para o Indoor Location. . . . .	77
Figura 22–Amostra de dados da atividade eletrodérmica coletados durante dez minutos. . . . .	78
Figura 23–Amostra de dados do fotopletismógrafo coletados durante dois minutos. . . . .	79
Figura 24–Matrizes de covariância com as variáveis do índice de bem-estar e satisfação. . . . .	81
Figura 25–Resultados da comparação entre os modelos RF e MLP para o serviço <i>Well-Being</i> . . . . .	82
Figura 26–Tela do protótipo evolutivo Ulisses. . . . .	83
Figura 27–Sentença SQL usada no protótipo. . . . .	84
Figura 28–Sentença SPARQL usada no protótipo. . . . .	84

## LISTA DE TABELAS

Tabela 1 – Trabalhos relacionados e a presença de métodos e ferramentas de ciência de dados em comparação com este trabalho. . . . .	27
Tabela 2 – Questões de pesquisa da revisão sistemática. . . . .	29
Tabela 3 – A string de busca da revisão sistemática. . . . .	30
Tabela 4 – Critérios de exclusão usados na revisão sistemática. . . . .	32
Tabela 5 – Segmentos industriais presentes na revisão sistemática. . . . .	36
Tabela 6 – Métodos de ciência de dados usados nos artigos da revisão sistemática. . . . .	38
Tabela 7 – Ferramentas de software usadas nos artigos da revisão sistemática. . . . .	40
Tabela 8 – Quantidade de dados ao longo do tempo dos artigos da revisão sistemática. . . . .	43
Tabela 9 – Artigos em que os conjuntos de dados estão disponíveis ao público . . . . .	44
Tabela 10–Comparativo dos trabalhos relacionados e modelo proposto. . . . .	58
Tabela 11–Relações da ontologia Odisseia. . . . .	67
Tabela 12–Métricas da ontologia Odisseia. . . . .	68
Tabela 13–Número de leituras do sinal RSSI de cada ponto de interesse. . . . .	71
Tabela 14–Número de leituras do sinal RSSI dos voluntários durante a coleta de dados. . . . .	71
Tabela 15–Arquivos gerados pela Empatica E4. . . . .	72
Tabela 16–Registros coletados por dia com a Empatica E4. . . . .	72
Tabela 17–Registros de sensores de ambiente coletados por dia. . . . .	73
Tabela 18–Escala de Afetos Positivos e Negativos (PANAS). . . . .	74
Tabela 19–Corpus . . . . .	103
Tabela 19–(continuação da tabela Corpus) . . . . .	104
Tabela 19–(continuação da tabela Corpus) . . . . .	105
Tabela 19–(continuação da tabela Corpus) . . . . .	106
Tabela 19–(continuação da tabela Corpus) . . . . .	107
Tabela 20–Métodos . . . . .	108
Tabela 20–(continuação da tabela Métodos) . . . . .	109
Tabela 20–(continuação da tabela Métodos) . . . . .	110
Tabela 20–(continuação da tabela Métodos) . . . . .	111
Tabela 20–(continuação da tabela Métodos) . . . . .	112
Tabela 21–Ferramentas . . . . .	113
Tabela 21–(continuação da tabela Ferramentas) . . . . .	114
Tabela 21–(continuação da tabela Ferramentas) . . . . .	115
Tabela 22–Segmentos Industriais . . . . .	116

## LISTA DE SIGLAS

BVP	<i>Blood Volume Pulse</i>
CSV	<i>Comma Separated Values</i>
EDA	<i>Electrodermal Activity</i>
ERP	<i>Enterprise Resource Planning</i>
GPS	<i>Global Positioning System</i>
GSR	<i>Galvanic Skin Response</i>
ILO	<i>International Labour Organization</i>
IoT	<i>Internet of Things</i>
MCC	<i>Matthew's Correlation Coefficient</i>
ODISSEU	<i>Ontology Driven Industrial Smart Services for Enterprise Use</i>
OWL	<i>Web Ontology Language</i>
RDF	<i>Resource Description Framework</i>
RSSI	<i>Received Signal Strength Indicator</i>
TAM	<i>Technical Architecture Modeling</i>
ULISSES	<i>User Locator for Indoor aSSESs</i>
W3C	<i>World Wide Web Consortium</i>
WBSI	<i>Well-Being and Satisfaction Index</i>

## SUMÁRIO

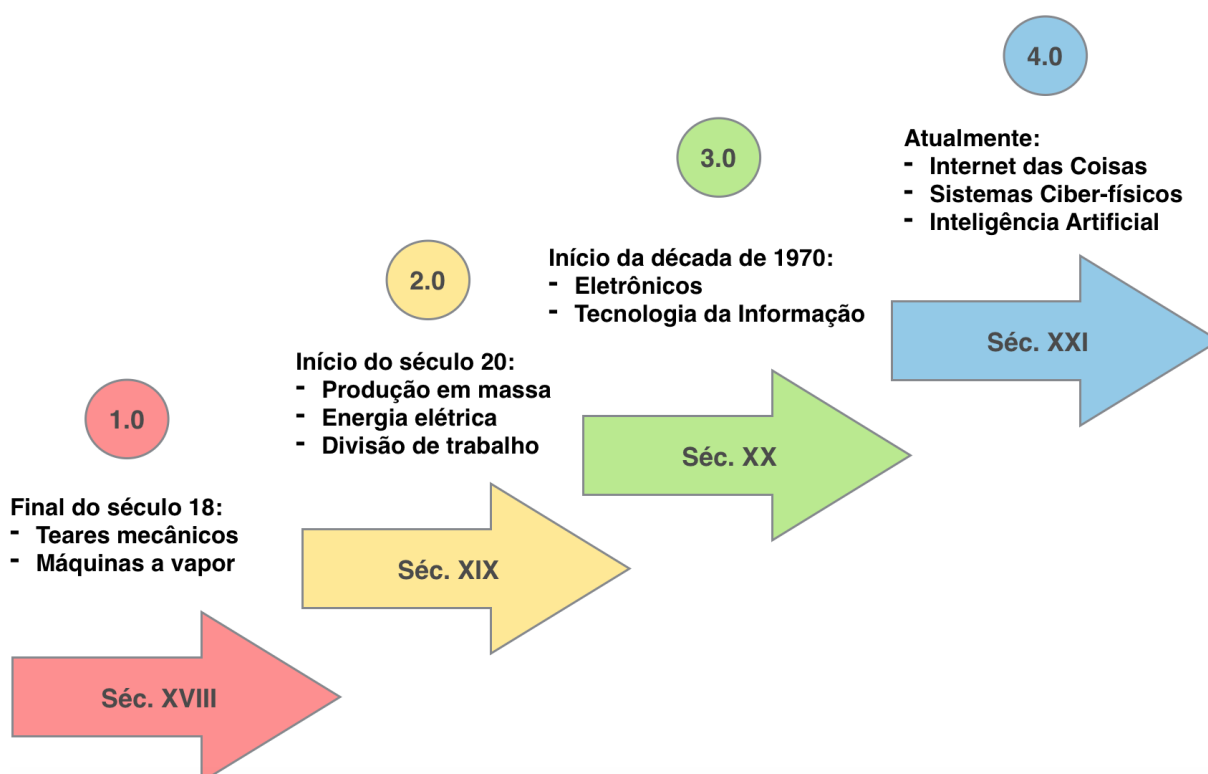
<b>1 INTRODUÇÃO</b>	<b>14</b>
1.1 Definição do Problema e Questão de Pesquisa	15
1.2 Objetivos	16
1.2.1 Objetivo Geral	17
1.2.2 Objetivos Específicos	17
1.3 Método	18
1.4 Organização da Tese	19
<b>2 FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
2.1 Serviços Inteligentes	20
2.2 Contexto e Históricos de Contextos	20
2.3 Ontologias	21
2.4 Beacons	22
2.5 Atividade Eletrodérmica e Pulso de Volume Sanguíneo	23
2.6 Bem-Estar Subjetivo	24
<b>3 TRABALHOS RELACIONADOS</b>	<b>25</b>
3.1 Revisões e Mapeamentos no Tema	25
3.2 Metodologia	28
3.2.1 Plano de Pesquisa	28
3.2.2 Execução	30
3.2.3 Análise	34
3.2.4 Relatórios	34
3.3 Resultados	34
3.3.1 GQ1. Quais segmentos industriais aplicaram técnicas de ciência de dados?	36
3.3.2 GQ2. Quais são os métodos de ciência de dados utilizados nos estudos?	37
3.3.3 GQ3. Quais são as ferramentas de software utilizadas nos estudos?	39
3.3.4 FQ1. Como os estudos empregam séries temporais contextuais?	40
3.3.5 FQ2. Como está a qualidade dos dados utilizados nos estudos ao longo do tempo ?	42
3.3.6 SQ1. Em quais bases de dados os estudos estão publicados	44
3.3.7 SQ2. Qual é o número de publicações por ano?	44
3.4 Taxonomia	46
3.5 Discussão	51
3.6 Conclusão da Revisão Sistemática	54
3.7 Trabalhos Selecionados e Comparativo	56
<b>4 MODELO ODISSEU</b>	<b>59</b>
4.1 Arquitetura	59
4.2 Estruturas de Dados	61
4.3 Ontologia Odisseia	63
<b>5 AVALIAÇÃO E RESULTADOS</b>	<b>69</b>
5.1 Coleta de Dados de Sensores	70
5.1.1 Localização	70
5.1.2 Sinais Vitais	72
5.1.3 Ambiente	73

<b>5.2 Coleta de Dados de Autorrelato</b>	<b>73</b>
<b>5.3 Serviços Inteligentes</b>	<b>74</b>
5.3.1 <i>Indoor Location</i>	75
5.3.2 Well-Being (WBSI)	77
5.3.3 Protótipo Ulisses	82
<b>6 CONCLUSÕES</b>	<b>85</b>
6.1 Contribuições	85
6.2 Limitações	86
6.3 Trabalhos Futuros	86
6.4 Publicações	87
<b>REFERÊNCIAS</b>	<b>88</b>
<b>APÊNDICE A – CORPUS</b>	<b>103</b>
<b>APÊNDICE B – MÉTODOS</b>	<b>108</b>
<b>APÊNDICE C – FERRAMENTAS</b>	<b>113</b>
<b>APÊNDICE D – SEGMENTOS INDUSTRIAIS</b>	<b>116</b>
<b>APÊNDICE E – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)</b>	<b>117</b>
<b>APÊNDICE F – FORMULÁRIO DE APLICAÇÃO DA ESCALA PANAS</b>	<b>118</b>

## 1 INTRODUÇÃO

Uma das formas de melhor entender a civilização atual é por meio da linha do tempo da revolução industrial. A primeira fase desse movimento teve início no final do século XVIII, com base na evolução dos equipamentos mecânicos de fabricação e no surgimento das máquinas a vapor. Então, no início do século 20, a possibilidade de implantar a produção em larga escala baseada na divisão de tarefas deu início à segunda fase da revolução industrial com o advento da eletricidade. Posteriormente, no início dos anos 70, o uso da eletrônica associada à tecnologia da informação possibilitou a automação dos processos de manufatura, estabelecendo a terceira fase desse movimento (KAGERMANN; WAHLSTER; HELBIG, 2013). Atualmente, o mundo vive a chamada nova onda da revolução industrial que começou na Europa e se espalhou pelo resto do mundo. A quarta fase dessa revolução, também chamada de Indústria 4.0, emprega avanços tecnológicos e conceitos como Internet das Coisas (IoT) e Sistemas Ciber-Físicos (CPS) para auxiliar no desenvolvimento de fábricas inteligentes (LIAO et al., 2017).

Figura 1 – As quatro fases da revolução industrial.



Fonte: Elaborado pelo autor.

A Figura 1 mostra as quatro fases da revolução industrial na linha do tempo, começando no final do século 18, com teares mecânicos e máquinas a vapor (1ª fase). Continuando no início do século 20, com o surgimento da produção em massa, ener-



gia elétrica e divisão de trabalho (2ª fase). Seguindo no início da década de 1970, com o advento dos eletrônicos e da tecnologia da informação (3ª fase). Atualmente com o uso da internet das coisas, sistemas ciber-físicos e inteligência artificial (4ª fase).

Junto com os avanços citados, a expressão “Ciência de Dados” começou a ser discutida pela comunidade de tecnologia da informação na primeira década do século XXI. Cientistas de dados são pessoas que lidam com quantidades significativas de dados de diferentes fontes para extrair informações relevantes na tomada de decisão (DAVENPORT; PATIL, 2012). Um dos principais objetivos da Ciência de Dados é prever resultados considerando o domínio de conhecimento de interesse (WALLER; FAWCETT, 2013). Um cientista de dados deve ter uma perspectiva de problemas de negócios, além do conhecimento de algoritmos de mineração de dados, métodos computacionais e ferramentas de software para extrair conhecimento e *insights* de grandes conjuntos de dados (PROVOST; FAWCETT, 2013; IGUAL; SEGUÍ, 2017; NASUTION; SITOMPUL; NABABAN, 2020).

Frequentemente, esses conjuntos de dados organizam observações em alta dimensionalidade com vários tipos, formatos e tamanhos de dados. Nesse sentido, uma das formas mais frequentes de lidar com essas informações é no domínio do tempo. As observações amostradas no domínio do tempo constituem uma sequência de informações denominada de série temporal (TORRES et al., 2021). As séries temporais podem receber diversos métodos de processamento para compreender a manutenção das máquinas, o ciclo de vida da produção, os processos industriais e de negócios para gerar resultados valiosos para as empresas. Além disso, as séries temporais permitem agregação, combinação e processamento computacional para criar níveis de informação mais elevados, como dados contextuais (BAVARESCO et al., 2020). O contexto, por sua vez, apresenta uma situação em relação a indivíduos, aplicativos e o ambiente circundante. Os contextos representam o tempo e o estado de algo que pode ser objetos, máquinas, sistemas, pessoas ou grupos (DEY; ABOWD; SALBER, 2001). Históricos de contextos são séries temporais que armazenam informações de contexto (FILIPPETTO; LIMA; BARBOSA, 2021; MARTINI et al., 2021; MACHADO et al., 2021; ARANDA et al., 2021).

## **1.1 Definição do Problema e Questão de Pesquisa**

A necessidade de se estabelecer um modelo padrão voltado para a Indústria 4.0 tem originado iniciativas que buscam integrar dispositivos inteligentes e componentes ubíquos de forma a otimizar a produção e reduzir custos, além de manter uma linha de produção sustentável e eficiente (RAJPUT; SINGH, 2020). Apesar dos esforços para obtenção de um modelo maduro voltado para a Indústria 4.0, ainda há uma forte expectativa das cadeias de valores de diversos segmentos industriais para atingir a

plena capacidade de produção (SANTOS; MARTINHO, 2019).

Não obstante aos estudos de caso envolvendo modelos que usam aprendizado de máquina para realizar manutenção preditiva (CANDANEDO et al., 2018), existe uma lacuna no que tange a modelagens que tratam os dados desde a sua entrada, seja por meio de sensores ou dispositivos móveis, até o estabelecimento de uma estrutura de dados capaz de manter históricos de séries temporais organizadas.

Dessa forma, a presente tese sugere a utilização de dados de históricos de contextos (WIEDEMANN et al., 2020; RENTZ; HECKLER; BARBOSA, 2023) gerados no ambiente industrial para auxiliar o desenvolvimento de serviços inteligentes. Assim, o estudo buscou resposta para a seguinte questão geral de pesquisa:

“Como seria um modelo computacional para gerenciamento de serviços inteligentes em ambientes industriais baseado na análise de históricos de contextos?”

No âmbito dessa questão, surgem questões específicas de pesquisa focadas nas contribuições científicas:

“Como deve ser um modelo computacional para suportar de forma genérica serviços inteligentes baseados em históricos de contextos gerados por diferentes fontes de dados em ambientes industriais?”

“De que forma seria aplicada uma ontologia como base para um modelo computacional baseado em históricos de contextos voltado para a Indústria 4.0?”

“Como seria um serviço inteligente voltado para a localização de funcionários em ambientes industriais *indoor* com base em históricos de contextos?”

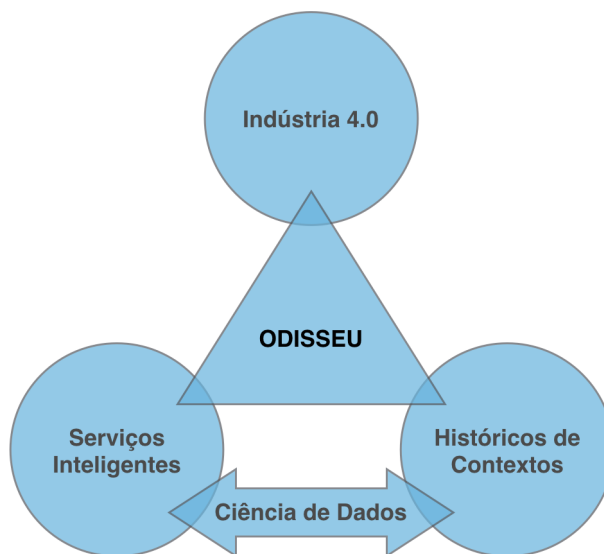
“É possível criar um serviço inteligente baseado em históricos de contextos para acompanhamento do bem-estar, usando um índice que mede o bem-estar subjetivo no ambiente de trabalho por meio de questões de autorrelato e dos sinais vitais dos funcionários de uma indústria?”

A Figura 2 mostra o foco de concentração da pesquisa, que envolve Indústria 4.0, Serviços Inteligentes e Históricos de Contextos. A interseção desses conceitos dá origem ao modelo proposto, chamado de Odisseu, com a Ciência de Dados atuando como elo entre os Serviços Inteligentes e os Históricos de Contextos.

## 1.2 Objetivos

Essa seção apresenta o objetivo geral e os objetivos específicos da tese, tendo o modelo Odisseu como elemento central.

Figura 2 – Foco da pesquisa.



Fonte: Elaborado pelo autor.

### 1.2.1 Objetivo Geral

O objetivo geral desta tese é propor um modelo computacional para a Indústria 4.0 baseado em dados provenientes de históricos de contextos, capaz de suportar serviços inteligentes.

Partindo do tema central, são aprofundados assuntos em áreas específicas como a aplicação de técnicas de ciência de dados em conjunto com históricos de contextos, além da utilização de uma ontologia como base para o modelo.

### 1.2.2 Objetivos Específicos

O trabalho busca atingir os seguintes objetivos específicos:

- a) realizar um estudo sobre o estado da arte dos métodos e ferramentas de ciência de dados utilizados ao longo dos anos na Indústria 4.0 e a sua aplicação em análises de históricos de contextos;
- b) criar um modelo para a Indústria 4.0 capaz de suportar serviços inteligentes baseados em técnicas de aprendizado de máquina, como redes neurais artificiais;
- c) desenvolver uma ontologia capaz de representar as entidades envolvidas no modelo computacional proposto, bem como o conhecimento envolvido nos fluxos de processos até os serviços inteligentes;
- d) implementar um protótipo funcional para avaliar o modelo, onde a partir da análise dos requisitos iniciais seja possível interagir com uma primeira versão do

produto final, agregando novos requisitos ao longo do desenvolvimento;

- e) validar o modelo através da criação de um serviço inteligente focado na localização dos funcionários no interior de uma indústria;
- f) validar o modelo por meio de um serviço inteligente que estima o bem-estar subjetivo, envolvendo tanto os dados de sinais vitais dos funcionários de uma indústria, como questões de autorrelato voltadas para afetos positivos e negativos.

### 1.3 Método

A partir da definição de um modelo e de sua implementação por meio de um protótipo funcional que aplicou os conceitos de históricos de contextos em conjunto com serviços inteligentes, o protótipo foi utilizado para demonstração do funcionamento de dois serviços. O primeiro serviço voltado para a localização de funcionários no interior de uma indústria e o segundo com o objetivo de estimar o nível de bem-estar dos funcionários. Para aplicação do método de pesquisa e avaliação do modelo foram necessários os seguintes passos:

- a) na primeira etapa foram realizados estudos sobre os temas básicos da pesquisa: Indústria 4.0, serviços inteligentes, históricos de contextos, métodos de ciência de dados e medição do bem-estar humano;
- b) a segunda etapa envolveu a pesquisa por trabalhos relacionados a modelos computacionais para a Indústria 4.0;
- c) a terceira etapa envolveu a criação de um modelo que suporte serviços inteligentes baseados em análises de históricos de contextos de dados gerados no ambiente industrial, tendo uma ontologia como base de organização do modelo;
- d) após a definição do modelo, a quarta etapa foi a implementação de um protótipo funcional que suporta os serviços inteligentes;
- e) a partir do protótipo, a quinta etapa ocorreu com a aplicação dos serviços inteligentes de localização e de bem-estar em ambiente de produção da indústria de semicondutores, usando dados de históricos de contextos reais gerados na HT Micron Semicondutores S.A.;
- f) a sexta etapa consistiu na análise dos resultados gerados e validação por parte dos gestores e funcionários envolvidos no processo de coleta de dados e testes.

## 1.4 Organização da Tese

Esta tese encontra-se estruturada em seis capítulos, sendo o primeiro deles a introdução. Os demais são descritos a seguir:

- a) Capítulo 2: Fundamentação Teórica - aborda os conceitos utilizados neste trabalho. No capítulo, são descritos os conceitos de serviços inteligentes, contexto e históricos de contextos, ontologias, beacons, além dos sinais vitais utilizados nos experimentos e a ideia de bem-estar subjetivo;
- b) Capítulo 3: Trabalhos Relacionados - inicialmente é apresentada uma revisão sistemática sobre os métodos e ferramentas de ciência de dados aplicados na Indústria 4.0, além de uma descrição dos trabalhos selecionados e uma comparação dos principais pontos de cada um deles com o modelo proposto;
- c) Capítulo 4: Modelo Odisseu - neste capítulo é descrito o modelo proposto e são definidas sua arquitetura e aspectos de implementação, além de ser apresentada uma ontologia para a Indústria 4.0 e detalhada a possibilidade de sua implementação em diferentes segmentos industriais;
- d) Capítulo 5: Avaliação e Resultados - apresenta uma avaliação do modelo Odisseu e seus resultados em formato de serviços inteligentes voltados para a localização interna e o bem-estar subjetivo dos funcionários de uma indústria de semicondutores;
- e) Capítulo 6: Conclusões - contém as considerações finais da tese, além das contribuições apresentadas.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo aborda os principais conceitos utilizados para a concepção do modelo apresentado nesta tese, intitulado Odisseu. Inicialmente são apresentados os conceitos referentes a serviços inteligentes, em seguida são abordados os conceitos de contexto e históricos de contextos. Também, são explicadas as funcionalidades das ontologias, a definição e propósito de beacons, a funcionalidade da atividade eletro-dérmica e do pulso de volume sanguíneo, e por fim o conceito de bem-estar subjetivo.

### 2.1 Serviços Inteligentes

Por definição, um serviço inteligente é um módulo independente de software capaz de gerar como saída um determinado resultado, baseado em um mecanismo de inteligência do próprio módulo, ou requisitar suporte de um outro serviço inteligente para obter o resultado desejado. O mecanismo de inteligência pode implementar algoritmos de inteligência artificial como redes neurais artificiais, árvores de decisão ou máquinas de vetores de suporte (MARIN et al., 2013).

Os serviços inteligentes vêm transformando a infraestrutura de cidades ao longo do mundo, de modo a facilitar o dia a dia de seus habitantes, criando as chamadas cidades inteligentes. Um exemplo é o sistema de transportes urbano que se beneficia da interligação entre dispositivos capazes de fornecer informações sobre o tráfego e os sinais de trânsito, além de ter a capacidade de analisar informações de câmeras de segurança instaladas nas avenidas. Com base nessas informações, um serviço inteligente pode informar a um controlador de tráfego que um congestionamento de veículos que esteja ocorrendo na avenida A pode ser atenuado aumentando o tempo de parada do sinal de trânsito da avenida B (DOBRE; XHAFA, 2014).

A união de dispositivos móveis com serviços inteligentes amplia a possibilidade de coletar e processar dados em diversos locais ao mesmo tempo. Dispositivos móveis com sensores embarcados e conectados a internet (IoT) geram grandes quantidades de dados que precisam ser armazenados de forma padronizada e organizada. Os dispositivos capazes armazenar dados relativos a localização e tempo, além de identificar de forma única um determinado provedor de informações, são referenciados como sensíveis ao contexto (GIL et al., 2016).

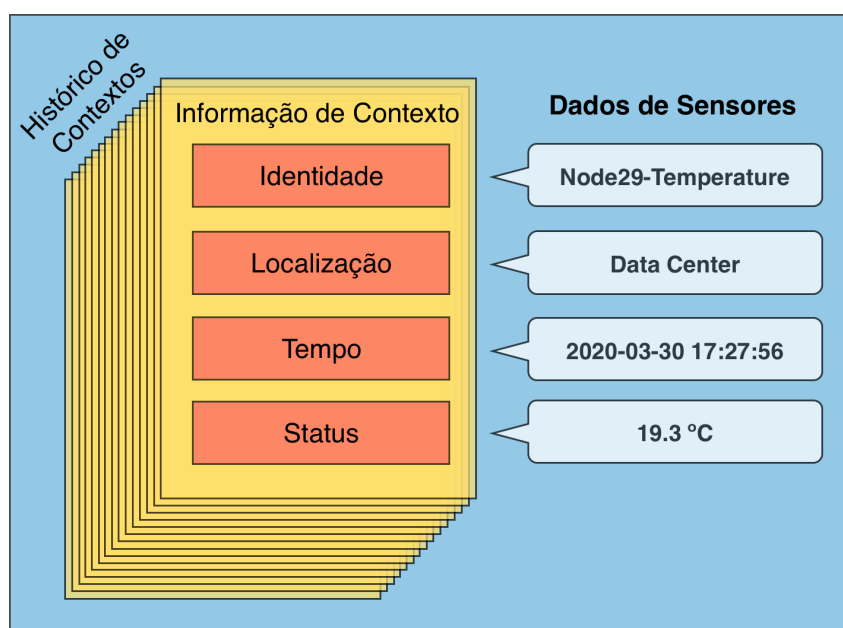
### 2.2 Contexto e Históricos de Contextos

A ideia de contexto, de acordo com Dey, Abowd e Salber (2001), compreende a identificação única de uma informação de contexto, juntamente com o local de ocorrência, o momento no tempo em que ocorreu, além de um *status* ou valor dessa in-

formação. A Figura 3 mostra um exemplo de informação de contexto destacada em um retângulo amarelo (Informação de Contexto), composto por retângulos vermelhos menores com a identidade da informação (Identidade), o local (Localização), o tempo (Tempo) e o valor da informação (Status). No exemplo, foram utilizados dados gerados por sensores (Dados de Sensores) para coletar o valor da temperatura do ar, que marcava 19.3 °C, no dia 30/03/2020 às 17:27:56, no *Data Center* de uma indústria, sendo que o sensor de temperatura que coletou a informação pertence ao Nó 29. Com isso, mesmo que haja um outro sensor de temperatura no *Data Center*, a informação coletada pelo Nó 29 naquele momento do tempo é única.

Ainda na Figura 3, atrás do retângulo amarelo que representa a informação de contexto do exemplo supracitado, existe uma série de retângulos amarelos representando outras informações de contexto coletadas ao longo do tempo (Histórico de Contextos). Séries temporais contendo informações organizadas de contexto são chamadas de históricos de contextos (ARANDA et al., 2021).

Figura 3 – Exemplo de informação e histórico de contextos.



Fonte: Elaborado pelo autor.

## 2.3 Ontologias

O uso de ontologias para compartilhar o entendimento sobre domínios é uma prática comum na gestão do conhecimento. Entidades, instâncias, relações, funções e axiomas compõem um domínio. O compartilhamento de conhecimento, as inferências lógicas e a reutilização de conhecimento são algumas razões para desenvolver modelos de contexto baseados em ontologia. A linguagem OWL (*Web Ontology Lan-*

*guage*) é uma opção para desenhar e gerenciar ontologias amplamente utilizada neste contexto (WANG et al., 2004).

O W3C (*World Wide Web Consortium*) propôs a linguagem OWL para estender as limitações do framework RDF (*Resource Description Framework*), também usado para representar ontologias. A linguagem OWL possui três sub-linguagens distintas: OWL Full, OWL DL e OWL Lite. Todas as sub-linguagens usam a mesma sintaxe XML do framework RDF (ANTONIOU; HARMELEN, 2004).

Giustozzi, Saunier e Zanni-Merk (2018) trabalharam em uma proposta baseada em ontologia para modelagem de contexto na Indústria 4.0, composta por ontologias centrais e ontologias de domínio. As ontologias centrais são a ontologia de sensor (rede de sensores semânticos), a ontologia de localização (lugar físico espacial) e a ontologia de tempo (hora atual e carimbos de data / hora). Ontologias de domínio são ontologia de recursos (entidades humanas e objetos físicos), ontologia de situação (estado abstrato) e ontologia de processo (tarefas executadas por um recurso).

## 2.4 Beacons

Beacons são dispositivos que emitem ondas de rádio para verificar a proximidade de seu portador em relação a um determinado ponto de referência. Podem ser usados para localizar pessoas em ambientes *indoor*, onde muitas vezes a capacidade de uso do GPS (Sistema de Posicionamento Global) é limitada (NEWMAN, 2014).

Para a utilização de Beacons como ferramenta de localização de pessoas em ambientes *indoor*, é necessário que o local esteja equipado com um número adequado de emissores de rádio capazes de indicar a posição de um determinado indivíduo com precisão. A indicação da posição de cada indivíduo é obtida por meio de um conjunto de valores de RSSI (*Received Signal Strength Indicator*), que é o indicador de potência do sinal recebido de cada um dos transmissores. O RSSI é um indicador numérico medido em dBm (decibel miliwatt) e varia de acordo com a distância da posição de recepção até o local de emissão do sinal de rádio. Uma das técnicas de disposição dos dispositivos emissores é ter um equipamento posicionado no centro do ambiente e os outros nas extremidades, preferencialmente usando o teto para fixação com o objetivo de atenuar interferências no sinal de rádio ocasionadas por obstáculos presentes no ambiente como móveis e maquinário industrial. Com os valores de RSSI gerados pelos dispositivos de rádio é possível identificar a posição de uma pessoa que esteja usando um dispositivo de rádio portátil. Para isso é usada uma técnica chamada de trilateração que usa a distância euclidiana entre os pontos fixos como forma de obter a localização de um ponto móvel (HUH; SEO, 2017).

Como alternativa ao método de trilateração e com o objetivo de localizar pessoas de forma mais precisa em um ambiente *indoor*, é possível utilizar redes neurais artifi-



ciais capazes de aprender com base em dados. As técnicas de aprendizado profundo (*Deep Learning*) são baseadas na capacidade que os neurônios do cérebro humano tem de aprender, ajustando a força de suas conexões. Para que isso ocorra em redes neurais artificiais, são utilizadas múltiplas camadas de conexões entre neurônios artificiais que têm o objetivo de identificar a melhor força de conexão entre as camadas, também chamada de peso (BENGIO; LECUN; HINTON, 2021).

A ideia proposta no serviço de localização interna desta tese é baseada na técnica implementada por Ferreira et al. (2017), onde foi utilizada a estrutura de um prédio comercial com três roteadores localizados em quatro salas da edificação com o propósito de verificar a possibilidade do uso de Redes Neurais Convolucionais (CNN) como instrumento de localização *indoor*. A entrada de dados inserida na rede neural é um conjunto de séries temporais com a intensidade do sinal de cada um dos roteadores recebido pelo alvo a ser localizado. O resultado apresentado é a sala em que o alvo se encontra.

Usando a mesma edificação do trabalho supracitado, o experimento conduzido por Carvalho et al. (2019) ampliou o número de roteadores para cinco e a quantidade de salas para sete. Foram utilizadas Redes Neurais Recorrentes (RNN) no processo de localização, incluindo as variações *Gated Recurrent Unit* (GRU) e *Long Short-Term Memory* (LSTM).

## 2.5 Atividade Eletrodérmica e Pulso de Volume Sanguíneo

A Resposta Galvânica da Pele (GSR - *Galvanic Skin Response*), como também é conhecida a atividade eletrodérmica (EDA), mede continuamente a atividade elétrica da pele sem interferência do controle humano consciente. Dessa forma, é possível detectar alterações nas emoções que podem levar ao estresse, provocadas por reações do sensor ao suor (DZEDZICKIS; KAKLAUSKAS; BUCINSKAS, 2020).

Dispositivos vestíveis como a pulseira Empatica são capazes de medir a atividade eletrodérmica de seres humanos, que é controlada pelo sistema nervoso e obtida por meio de sensores de suor (MAJUMDER; MONDAL; DEEN, 2017). A Empatica possui também outros sensores capazes de gerar dados para estimar o estresse por meio da pulsação, temperatura da pele e dados de movimentação obtidos através de acelerômetros (SCHMIDT et al., 2018).

Além da atividade eletrodérmica, o pulso de volume sanguíneo (BVP) também pode ser utilizado para detecção de alterações emocionais como o estresse, raiva e diversão. Tanto EDA como BVP podem ser coletados usando abordagens não intrusivas. O BVP analisa os vasos sanguíneos próximos a região do sensor, medindo a variação do fluxo de sangue por meio da emissão de luz refletida na pele, de modo que cada pico máximo corresponde a um batimento cardíaco. A EDA indica variações

elétricas na superfície da pele, ocasionadas por transpiração, causada por glândulas sudoríparas (ALVES et al., 2013).

## **2.6 Bem-Estar Subjetivo**

O bem-estar subjetivo é o estudo de fatores positivos e negativos que influenciam o pensamento das pessoas sobre a vida, incluindo reflexões acerca da satisfação de viver e respostas emocionais como emoções positivas e agradáveis, em contraste com emoções negativas e desagradáveis. Afetos positivos e negativos são relacionados inversamente de forma moderada, porém não necessariamente exatamente opostos. Devido a natureza subjetiva, escalas de autorrelato constituem uma das principais formas de medir o bem-estar subjetivo, levando em consideração experiências pessoais. Apesar de diferentes escalas de autorrelato apresentarem convergência, é importante usar métodos diversificados para predição dos valores, como relatórios de informantes e medidas biológicas (DIENER; OISHI; TAY, 2018)

O estabelecimento de limites saudáveis no trabalho e na vida pessoal é competência fundamental e está relacionado ao bem-estar. Questões de saúde mental, como ansiedade e depressão, apresentam correlação frequente com a falta de equilíbrio entre a vida pessoal e profissional. Ademais, problemas de saúde mental que já existiam foram exacerbados pela pandemia de COVID-19, o que aumentou a atenção da sociedade nesse sentido (HALL, 2023).

O bem-estar dos trabalhadores está associado a indicadores do desempenho da organização, como produtividade e rotatividade voluntária, e pode ser considerado um indicador de ética organizacional. Estudos psicológicos conceituais e empíricos indicam que a satisfação no trabalho e o engajamento estão associados ao bem-estar do trabalhador. O humor e as emoções estão ligados ao bem-estar subjetivo e podem se manifestar de diferentes formas, como atividade cerebral, comportamento e respostas fisiológicas (WIJNGAARDS et al., 2022).

### 3 TRABALHOS RELACIONADOS

Para seleção dos trabalhos relacionados foi realizada uma revisão sistemática da literatura sobre as pesquisas que envolvem Indústria 4.0, históricos e séries temporais de contextos, além de métodos de ciência de dados que possam ser implementados em serviços inteligentes. Revisão sistemática é uma forma estruturada de buscar no estado da arte tendências sobre o tema pesquisado. A revisão sistemática foi publicada como artigo no periódico *Sensors* (ARRUDA et al., 2023a).

A literatura apresenta revisões sistemáticas englobando o escopo semelhante ao deste estudo. A manufatura tem recebido pesquisas para lidar com problemas de tomada de decisão usando técnicas analíticas, mineração de dados e aprendizado de máquina (WOLF et al., 2019). Além disso, uma revisão das ferramentas e aplicativos de big data para manufatura apresentou componentes essenciais para criar soluções completas (CUI; KARA; CHAN, 2020). Ademais, estudos de caso aplicados a uma empresa química apresentaram como resultado uma revisão de mineração de dados e categorias analíticas como preditiva, inquisitiva, descritiva e prescritiva, focada em processos de manufatura (BELHADI et al., 2019). No entanto, essas revisões não recuperam e analisam métodos de ciência de dados e ferramentas de software focadas em aplicações industriais gerais. Este capítulo propõe uma revisão sistemática da literatura sobre métodos e ferramentas de ciência de dados empregados em distintos segmentos da indústria. Além disso, analisa o uso de diferentes níveis de séries temporais e qualidade de dados em aplicações de ciência de dados. Nesse sentido, fornece respostas a três questões gerais, duas específicas e duas estatísticas para sintetizar a literatura por meio de uma taxonomia, favorecendo a representação dos resultados.

O capítulo encontra-se organizado da seguinte forma. A Seção 3.1 mostra revisões e mapeamentos no tema e como o estudo se diferencia. A Seção 3.2 explica a metodologia empregada na revisão sistemática. A Seção 3.3 apresenta os resultados e as descobertas com base nas questões de pesquisa, destacando segmentos industriais, métodos de ciência de dados e ferramentas de software. A Seção 3.4 descreve a taxonomia proposta para representar as descobertas cobertas pela literatura, e a Seção 3.5 discute as descobertas. A Seção 3.6 aborda as limitações, trabalhos futuros e conclusões da revisão sistemática. Finalmente, a Seção 3.7 apresenta os trabalhos selecionados e mostra um comparativo entre eles e com o modelo proposto.

#### 3.1 Revisões e Mapeamentos no Tema

Esta seção analisa outras pesquisas e revisões em comparação com a revisão sistemática. Nos últimos anos, alguns autores têm revisado a literatura, visando ex-

plorar as melhores técnicas utilizadas pelas fábricas inteligentes que correspondem ao campo da ciência de dados. Isso porque a Indústria 4.0 permite o emprego de múltiplos tipos de tecnologias em diferentes segmentos de produção.

Mazzei e Ramjattan (2022) usaram técnicas de processamento de linguagem natural para revisar métodos de aprendizado de máquina usados em casos da Indústria 4.0. Os autores estabeleceram questões sobre os principais problemas da Indústria 4.0, quais métodos de aprendizado de máquina foram utilizados nessas situações e como as áreas focaram na literatura acadêmica e nos *white papers*. A revisão sistemática focou em duas bases de dados utilizando a técnica de modelagem de tópicos BERTopic. Os problemas mais recorrentes diziam respeito à segurança, produção inteligente, conectividade IoT, otimização de serviços, automação robótica e otimização logística. As redes neurais convolucionais foram o método de aprendizado de máquina mais frequente.

Wolf et al. (2019) estudaram a falta de ferramentas de gestão orientadas para problemas de tomada de decisão no domínio da manufatura. O trabalho apresenta um mapeamento sistemático que identificou sete áreas de aplicação para análise de dados e contou com técnicas analíticas avançadas associadas a cada área. O mapeamento originou uma nova ferramenta para facilitar a tomada de decisões que identificou projetos analíticos promissores. Além disso, a ferramenta de gestão utilizou técnicas de mineração de dados e algoritmos de aprendizado de máquina.

Cui, Kara e Chan (2020) publicaram uma revisão sistemática da literatura com o objetivo de classificar ferramentas de big data com semelhanças e identificar as diferenças entre elas. O trabalho levou em consideração dados industriais, tecnologias de big data e aplicações de dados em manufaturas. A estrutura conceitual da revisão sistemática da literatura apresentou três perspectivas: fonte de dados, ecossistema de big data e consumidor de dados. Tipos de dados, dispositivos de origem, dinâmica de dados, formatos de dados e sistemas compuseram a perspectiva da fonte de dados. A perspectiva do ecossistema de big data apresentou aspectos de dados como armazenamento, gerenciamento de recursos, visualização, análise, banco de dados, *data warehouse*, pesquisa, consulta, processamento, ingestão, fluxo de dados, fluxo de trabalho e gerenciamento. Previsão, otimização, monitoramento, *design*, suporte à decisão, análise de dados, agendamento, gerenciamento de dados, simulação e controle de qualidade faziam parte dos componentes da perspectiva do consumidor de dados. Quatro questões de pesquisa apresentaram os motivadores e requisitos para aplicações de big data, os componentes essenciais do ecossistema de big data, as capacidades dos ecossistemas de big data e as direções futuras das aplicações de big data. Concluindo, os autores encontraram seis impulsionadores principais e nove componentes essenciais do ecossistema de big data. O estudo não encontrou na literatura nenhuma solução de big data pronta para uso empresarial.

Belhadi et al. (2019) revisaram sistematicamente a literatura sobre análise de big data em processos de fabricação, além de vários estudos de caso aplicados a uma empresa química líder de mercado. Os três casos fizeram parte de um projeto de transformação digital, sendo o primeiro caso uma implementação de *big data analytics* numa fábrica de fertilizantes, o segundo numa empresa de ácido fosfórico e o terceiro, uma unidade de produção inteligente e autocontrolada. O artigo classificou os trabalhos selecionados de acordo com as categorias de mineração e análise de dados: preditiva, inquisitiva, descritiva e prescritiva. Além disso, as técnicas implementadas categorizaram os artigos em *offline* e *online* em tempo real. Além disso, o trabalho estabeleceu as seguintes tendências de pesquisa: abordagens de mineração de dados em tempo real, arquitetura facilitadora de análise de big data, inteligência integrada de dados humanos e análise prescritiva. Cada tendência de pesquisa apontou para as questões de pesquisa relativas à gestão de desempenho, controle de produção e manutenção em processos de manufatura. Os autores perceberam que o surgimento de tecnologias avançadas, principalmente sensores, gerou dados com ampla variabilidade, grande variedade, alta velocidade, intensa volatilidade, alto volume, veracidade não verificada e baixo valor. Além disso, o estudo propôs uma estrutura de análise de big data no processo de fabricação, que apresentou os desafios do processo, as faculdades e as capacidades da análise de big data. Porém, nenhum dos trabalhos relacionados recuperou e analisou, ao mesmo tempo, métodos de ciência de dados e ferramentas de software focadas em aplicações industriais (Tabela 1).

Tabela 1 – Trabalhos relacionados e a presença de métodos e ferramentas de ciência de dados em comparação com este trabalho.

Artigo	Métodos	Ferramentas
Mazzei e Ramjattan (2022)	Sim	Não
Wolf et al. (2019)	Não	Sim
Cui, Kara e Chan (2020)	Não	Sim
Belhadi et al. (2019)	Sim	Não
Este trabalho	Sim	Sim

Fonte: Arruda et al. (2023a).

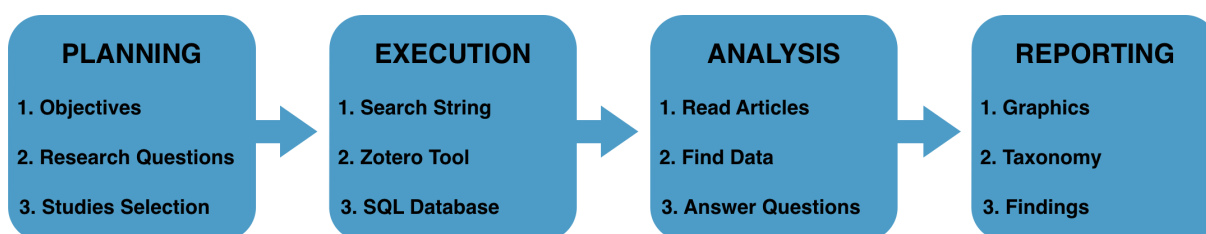
Portanto, este capítulo identifica e organiza segmentos industriais, métodos de ciência de dados e ferramentas de software empregadas em ambientes industriais para produzir uma taxonomia. Por sua vez, a taxonomia sintetiza a literatura favorecendo a representação dos achados. Para isso, o capítulo apresenta uma revisão sistemática da literatura convergindo para três temas principais: Indústria 4.0, ciência de dados e séries temporais. Esses temas são a base para a criação de questões gerais, focadas e estatísticas que moldam a investigação deste trabalho. Nesse sentido, o trabalho

investiga abordagens específicas derivadas desses temas, particularmente o uso do contexto e a qualidade dos dados empregados nos estudos. Esses aspectos proporcionam uma abordagem diferencial em relação às revisões citadas.

## 3.2 Metodologia

Esta seção apresenta os métodos de pesquisa empregados neste trabalho. A estrutura segue a metodologia proposta por Petersen, Vakkalanka e Kuzniarz (2015). A figura 4 resume as etapas organizadas em quatro com três subetapas cada. Primeiramente, as etapas abrangem o planejamento da pesquisa, seguido pela execução da revisão sistemática, análise dos dados e relato dos resultados.

Figura 4 – As quatro etapas da revisão sistemática.



Fonte: Arruda et al. (2023a).

### 3.2.1 Plano de Pesquisa

O planejamento da pesquisa estabelece os objetivos, define as questões de pesquisa e planeja a seleção dos estudos. As subseções a seguir explicam cada etapa detalhadamente.

#### 3.2.1.1 Objetivos da Revisão Sistemática

O aspecto central deste capítulo é uma revisão sistemática do estado da arte em métodos e ferramentas de ciência de dados empregados na Indústria 4.0. O objetivo foi encontrar estudos que envolvessem Indústria 4.0, ciência de dados e séries temporais, para produzir *insights* úteis para o campo industrial. Após a coleta dos artigos, os objetivos focaram na classificação de cada estudo de acordo com os segmentos industriais, métodos de ciência de dados e ferramentas de software. Posteriormente, este trabalho sintetizou os resultados com gráficos, tabelas e uma taxonomia dos resultados para facilitar a análise dos dados.

### 3.2.1.2 Questões de Pesquisa

As questões de pesquisa focaram nos três temas principais da revisão: “Indústria 4.0”, “Ciência de Dados” e “Séries Temporais”. As sete questões de pesquisa tiveram a seguinte divisão: três questões gerais (GQ), duas questões focadas (FQ) e duas questões estatísticas (SQ), conforme apresentado na Tabela 2.

Tabela 2 – Questões de pesquisa da revisão sistemática.

Ref.	Questões de Pesquisa
GQ1	Quais segmentos industriais aplicaram técnicas de ciência de dados?
GQ2	Quais são os métodos de ciência de dados utilizados nos estudos?
GQ3	Quais são as ferramentas de software utilizadas nos estudos?
FQ1	Como os estudos empregam séries temporais contextuais?
FQ2	Como está a qualidade dos dados utilizados nos estudos ao longo do tempo ?
SQ1	Em quais bases de dados os estudos estão publicados?
SQ2	Qual é o número de publicações por ano?

Fonte: Arruda et al. (2023a).

A motivação para buscar segmentos industriais envolvidos com ciência de dados foi descobrir onde grandes quantidades de dados precisavam ser analisadas e mostrar novas oportunidades de trabalho (GQ1), os tipos de métodos utilizados para esse fim (GQ2) e quais eram as ferramentas empregadas na indústria (GQ3). Além disso, compreender como os dados são utilizados ao longo do tempo é fundamental para escolher a melhor técnica a utilizar em situações específicas (FQ1). Também, a qualidade dos conjuntos de dados disponíveis é importante para analisar o desempenho de um algoritmo em relação às lacunas e ao equilíbrio dos dados (FQ2). Por fim, as fontes (SQ1) e o número de publicações ao longo do tempo (SQ2) auxiliam no processo de pesquisa.

### 3.2.1.3 Seleção de Estudos

O processo de seleção dos estudos envolveu cinco bases de dados relevantes na área de pesquisa: ACM, IEEE, Scopus, Springer e Wiley. Um estudo sobre as questões de pesquisa ajudou a definir a string de busca. Além disso, a utilização de sinônimos e palavras relacionadas permitiu que a busca obtivesse resultados mais abrangentes. A tabela 3 mostra a organização da string de busca considerando três temas.

Tabela 3 – A string de busca da revisão sistemática.

<b>Tema</b>	<b>Termos de Busca</b>
Industry 4.0	( “industry 4.0” OR “industrie 4.0” OR “cyber physical systems” ) AND
Data Science	( “data science” OR “machine learning” OR “big data” OR “data analytics” OR “data mining” ) AND
Time Series	( “time series” OR “context histories” OR “contexts histories” OR “context history” OR “trails” )

Fonte: Arruda et al. (2023a).

O refinamento da busca ocorreu por meio de seis critérios de exclusão (EC). Primeiro, o processo de filtragem desconsiderou os artigos não escritos em inglês (EC1) e não encontrados em periódicos, conferências ou *workshops* (EC2). Em seguida, a análise dos títulos (EC3) e resumos (EC4) considerou apenas os trabalhos que estavam de acordo com as questões de pesquisa. Depois, a filtragem excluiu os artigos duplicados (EC5). Por fim, o último critério de filtragem (EC6) foi a abordagem de três passagens. Esta abordagem utiliza a análise do título, resumo, introdução, título das seções e subseções, conteúdo matemático e conclusões na primeira passagem. A segunda passagem é a observação das imagens, diagramas e ilustrações. Por fim, a terceira passagem pesquisa todo o texto (KESHAV, 2007).

### 3.2.2 Execução

Após a fase de planejamento, a execução das etapas planejadas ocorreu de acordo com a inserção da string de busca nas bases de dados selecionadas. Além disso, a utilização da ferramenta Zotero e de um banco de dados SQL permitiu uma melhor organização dos resultados.

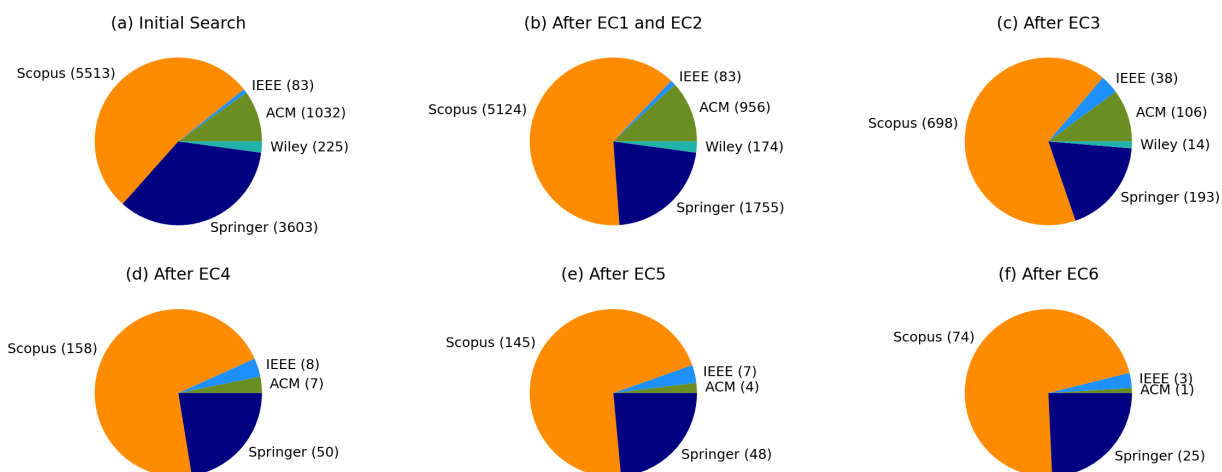
#### 3.2.2.1 Aplicação da String de Buscas

A busca inicial nas bases de dados ocorreu sem filtros, utilizando a string de busca proposta e organizando os dados coletados em coleções nomeadas de acordo com cada base de dados. O processo de filtragem ocorreu todo no arquivo “zotero.sqlite”, que é o banco de dados SQL gerado pelo Zotero. As bases de dados de busca escolhidas foram ACM, IEEE, Scopus, Springer e Wiley. A Figura 5 mostra o nome das bases de dados e o número de artigos recuperados na busca inicial após a aplicação



de cada critério de exclusão.

Figura 5 – Resultados da busca inicial da revisão sistemática.



Fonte: Arruda et al. (2023a).

### 3.2.2.2 Ferramenta Zotero

A utilização de uma única ferramenta de gerenciamento visa agilizar o processo de coleta, facilitando a busca e classificação dos artigos. É preferível uma ferramenta com acesso aberto ao seu banco de dados. No início deste estudo, foram realizados testes com o Mendeley (<https://www.mendeley.com>; acessado em 22 de maio de 2023) e Zotero (<https://www.zotero.org>; acessado em 22 de maio de 2023) ferramentas de gerenciamento de referências. O Zotero foi escolhido devido à necessidade de acessar o banco de dados SQL sem restrições, por se tratar de um banco de dados de acesso aberto. Zotero é uma ferramenta gerenciadora de referências que oferece uma forma prática de coleta de artigos. A ferramenta organiza os resultados da pesquisa graças à possibilidade de utilizar um conector de navegador que agiliza o processo, ao permitir a coleta de metadados de um conjunto de artigos ao invés de um por um. Além disso, a utilização do *plugin* de navegador ZotFile (<http://zotfile.com>; acessado em 22 de maio de 2023) na análise individual dos artigos selecionados facilitou a extração das sentenças destacadas (VANHECKE, 2008).

A tabela 4 apresenta os critérios de exclusão utilizados no processo de filtragem com a ferramenta Zotero. Na tela principal do Zotero, o campo denominado “Extra” permite ao usuário inserir informações adicionais sobre os trabalhos. A anexação do símbolo de barra vertical (“|”) ao final do campo “Extra” criou um novo campo a ser usado por consultas SQL chamado “*Status*”. Este novo campo utilizado ao longo do processo de filtragem atribuiu um “*Status*” diferente a cada artigo após a aplicação de cada critério de exclusão. Antes da aplicação dos critérios de exclusão, todos os

artigos tinham o “Status” definido como vazio (“ ”). A utilização de sentenças SQL na base de dados Zotero proporcionou uma forma prática de aplicar os dois primeiros critérios de exclusão ao mesmo tempo, filtrando artigos não escritos em inglês (EC1) e não encontrados em periódicos, conferências ou workshops (EC2). Os artigos que atenderam a esses critérios de exclusão tiveram seu “Status” definido como “ec”, o que significa excluídos por EC1 ou EC2. Os demais artigos com *status* vazio foram filtrados pelo terceiro critério de exclusão, a análise do título (EC3). Os trabalhos descartados tiveram seu *status* alterado para “ec3”, e os aceitos para a próxima etapa ganharam o *status* “ec3\_next”. O processo de filtragem continuou com os artigos com *status* “ec3\_next”, que tiveram seus resumos analisados no quarto critério de exclusão (EC4), e aceitos para a próxima fase (“ec4\_next”) ou rejeitado (“ec4”). O próximo filtro eliminou trabalhos duplicados, representando o quinto critério de exclusão (EC5), definindo o *status* como “ec5” ou mantendo o artigo na fase seguinte, definindo o *status* como “ec5\_next”. O último critério de exclusão (EC6) aplicou a abordagem de três passagens e alterou o *status* dos artigos descartados para “ec6” e dos artigos aceitos para “final”.

Tabela 4 – Critérios de exclusão usados na revisão sistemática.

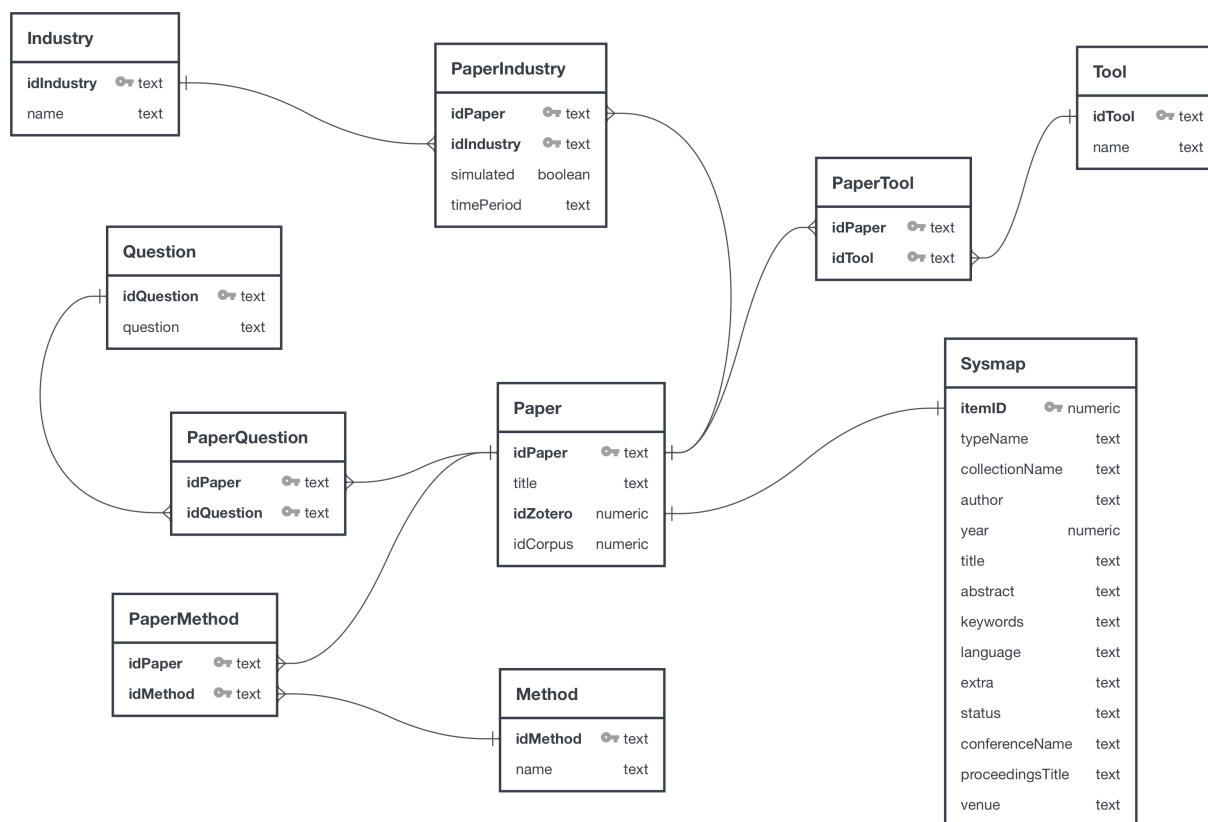
Abrev.	Critério de Exclusão	Status	Excluído	Próximo Critério
EC1	Não escrito em inglês	“ ”	“ec”	-
EC2	Não encontrado em jornais, conferências ou workshops	“ ”	“ec”	-
EC3	Análise de título	“ ”	“ec3”	“ec3_next”
EC4	Análise do resumo	“ec3_next”	“ec4”	“ec4_next”
EC5	Artigos duplicados	“ec4_next”	“ec5”	“ec5_next”
EC6	Abordagem de três passagens	“ec5_next”	“ec6”	“final”

Fonte: Arruda et al. (2023a).

### 3.2.2.3 Banco de Dados SQL

O banco de dados SQL permitiu uma organização dos dados extraídos durante o processo. Além disso, o modelo relacional possibilitou organizar os dados coletados ao longo do desenvolvimento da revisão sistemática e facilitou a geração de gráficos e a extração de informações. Nove tabelas e uma visualização do banco de dados da ferramenta Zotero compuseram o modelo. A Figura 6 representa o modelo relacional, desenvolvido com a ferramenta QuickDBD (<https://app.quickdatabasediagrams.com>; acessado em 22 de maio de 2023).

Figura 6 – Diagrama do banco de dados criado para a revisão sistemática.



Fonte: Arruda et al. (2023a).

A tabela “Paper” possui quatro atributos, um identificador único do artigo (campo “idPaper”), um campo para armazenar o título do trabalho (“title”), um código identificador do trabalho na ferramenta Zotero (“idZotero”), e um campo com a ordem do artigo no corpus (“idCorpus”). Esta tabela tem um relacionamento um-para-um com a visão “Sysmap”, que representa os dados mais relevantes utilizados da base de dados do Zotero.

O campo “itemID”, da visão “Sysmap”, é o identificador único do artigo utilizado pelo Zotero e está relacionado ao campo “idZotero”, da tabela “Paper”. O campo “typeName” representa o tipo de publicação (seção de livro, artigo de periódico, artigo de conferência, manuscrito, livro ou relatório). Este trabalho considerou apenas artigos de periódicos, trabalhos de conferências e workshops, que são uma variante de conferências. O campo “collectionName” é o nome da coleção escolhida para organizar os documentos. Este trabalho utilizou os nomes das bases de busca e um identificador representando a rodada de busca. O campo “author” é o nome do primeiro autor. O campo “year” é o ano de publicação, “title” é o título do artigo e “abstract”, o resumo do artigo. O campo “palavras-chave” organizou as palavras-chave do trabalho separadas por vírgula. O campo “language” é a linguagem de escrita do artigo. O campo “extra” foi utilizado para definir um status para cada artigo usando uma barra vertical seguida

de um código. Outro atributo chamado “status” mostra o código de status. Artigos de uma conferência ou workshop usaram os campos “conferenceName” e “proceedings-Title” para armazenar o nome da conferência ou workshop e o título dos anais. Por fim, o campo “local” indica se o artigo é de periódico, conferência ou workshop.

As tabelas principais “Industry”, “Question”, “Tool” e “Methods” estão relacionadas à tabela “Paper” em um relacionamento disjuncto muitos-para-muitos e em relacionamento um-para-muitos com as tabelas auxiliares. A tabela “Industry” contém o cadastro dos segmentos industriais utilizados na revisão. “Question” armazena as questões de pesquisa do artigo. A tabela “Tool” contém as ferramentas de software utilizadas nos artigos selecionados. A tabela “Method” tem os métodos de ciência de dados implementados pelos trabalhos. As tabelas auxiliares “PaperIndustry”, “PaperQuestions”, “PaperTool” e “PaperMethod” possuem as chaves primárias das tabelas principais. A tabela auxiliar “PaperIndustry” possui dois campos extras. Um deles foi responsável por indicar quando determinado segmento industrial atuou em um ambiente simulado (campo “simulated”) e o outro por armazenar o período de tempo dos dados utilizados no trabalho (campo “timePeriod”).

### 3.2.3 Análise

Os trabalhos selecionados foram investigados em busca de dados que respondessem às questões de pesquisa e classificassem cada trabalho em um segmento específico da indústria. Além disso, a investigação permitiu identificar os métodos de ciência de dados e ferramentas de software aplicadas nos estudos. Embora alguns artigos tenham mencionado o segmento industrial, seus dados, na verdade, resultaram de um ambiente de simulação. Além disso, o tempo de duração dos dados utilizados nos estudos, quando disponíveis, apareceu em horas, dias, meses ou anos.

### 3.2.4 Relatórios

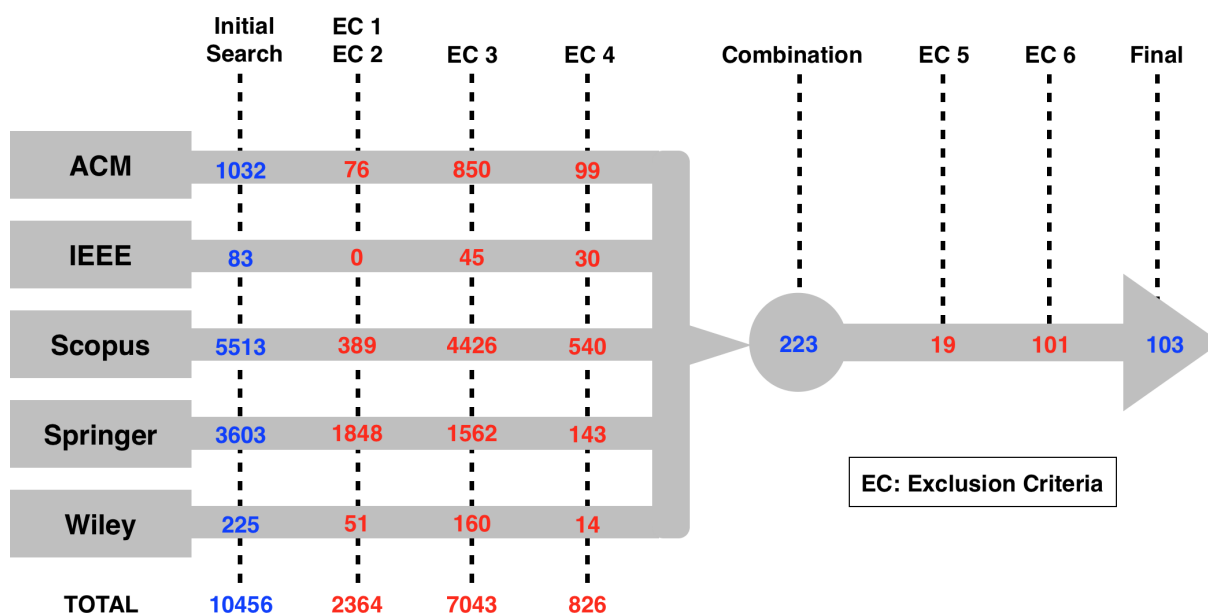
Os relatórios forneceram resultados de diferentes maneiras. A criação de gráficos favoreceu o processo de análise fornecendo informações em figuras com dados agrupados e organizados. Além disso, a criação de uma taxonomia sintetizou uma visão geral dos resultados. Ademais, as questões de pesquisa tiveram as respostas discutidas o que produziu destaques da pesquisa.

## 3.3 Resultados

Esta seção apresenta os resultados da revisão sistemática da literatura. A Figura 7 mostra cada etapa do processo com a quantidade de artigos de cada base de dados

utilizada ao longo do processo. Além disso, a figura mostra a quantidade de artigos descartados pelos critérios de exclusão.

Figura 7 – Artigos por base de dados e os critérios de exclusão da revisão sistemática.



Fonte: Arruda et al. (2023a).

Inicialmente, a busca retornou 10.456 artigos das cinco bases de dados. Com o objetivo de encontrar os primeiros anos que correspondiam à string, a busca não utilizou nenhum filtro além das palavras-chave presentes na string de busca, o que significa que não houve corte por anos. Em seguida, os dois critérios de exclusão iniciais (EC1 e EC2) eliminaram os artigos não escritos em inglês e os que não foram encontrados em periódicos, conferências ou workshops (22,61%). O terceiro critério de exclusão (EC3) retirou os artigos que não passaram na análise do título (67,36%). O quarto critério de exclusão (EC4) excluiu artigos de acordo com a análise do resumo (7,90%). A combinação dos demais artigos resultou em 223 trabalhos, representando 2,14% da busca inicial. O quinto critério de exclusão (EC5) removeu 19 estudos duplicados. Por fim, o sexto critério de exclusão (EC6) excluiu 101 artigos que utilizaram a abordagem de três passagens, restando 103 trabalhos no *corpus*, o que correspondeu a 0,99% da busca inicial. A Tabela 19, do Apêndice A, apresenta os artigos selecionados e os códigos de identificação do *corpus*.

A etapa seguinte consistiu na análise minuciosa do *corpus* visando responder cada questão de pesquisa, apresentando os resultados com gráficos e tabelas. O restante desta seção apresenta as questões de pesquisa e respectivas respostas.

### 3.3.1 GQ1. Quais segmentos industriais aplicaram técnicas de ciência de dados?

Visando padronizar os segmentos industriais presentes no *corpus*, esses resultados consideraram a classificação proposta pela Organização Internacional do Trabalho (<https://www.ilo.org>; acessado em 22 de maio de 2023), uma agência das Nações Unidas. Essa classificação apresenta 22 segmentos industriais, dos quais 15 estão no *corpus*. A tabela 5 apresenta os segmentos industriais e o código de identificação do *corpus* de cada artigo, além de um segmento extra para artigos com segmentos enquadrados no segmento de uso geral.

Tabela 5 – Segmentos industriais presentes na revisão sistemática.

Segmento Industrial	ID no Corpus
Agricultura, plantações, outros setores rurais	35, 59
Produção de metais básicos	13, 14, 19, 25, 50, 66, 103
Indústrias químicas	89
Construção	86
Comida, bebida, tabaco	22, 46, 88
Silvicultura, madeira, celulose e papel	53
Serviços de saúde	9, 18, 49, 58
Engenharia mecânica e elétrica	4, 11, 12, 16, 20, 26, 45, 52, 56, 61, 62, 72, 76, 78, 79, 80, 82, 87, 90, 102
Mídia, cultura, gráfica	100
Mineração (carvão, outras minerações)	23, 54, 70, 91
Produção de petróleo e gás; Refinaria de óleo	3, 15, 27, 74, 75, 81
Serviços postais e de telecomunicações	63, 85
Têxteis, vestuário, couro, calçado	48, 51
Fabricação de equipamentos de transporte	5, 6, 10, 31, 34, 40, 41, 44, 55, 57, 60, 68, 84, 92, 98, 101
Utilidades (água, gás, eletricidade)	8, 28, 33, 39, 41, 47, 71, 93, 96
Propósito geral / outros	1, 2, 7, 17, 21, 24, 29, 30, 32, 36, 37, 38, 42, 43, 64, 65, 67, 69, 73, 77, 83, 94, 95, 97, 99

Fonte: Arruda et al. (2023a).

O segmento industrial *Propósito geral / outros* representou o maior número de artigos com 24,04% referentes ao total do *corpus*. Depois, *Engenharia mecânica e elétrica* foi o segundo segmento industrial com 19,23%, seguida pela *Fabricação de equipamentos de transporte* com 15,38%. Os demais segmentos representaram menos de 10% do total cada. Luo, Liu e Qi (2019) usaram dois segmentos industriais: *Fabricação de equipamentos de transporte* e *Utilidades (água, gás e eletricidade)*. Esse artigo foi contabilizado duas vezes para fins de análise percentual.

*Utilidades* representou 8,65% do *corpus*. *Produção de metais básicos* aproximou-se de 6,73% do *corpus*. *Petróleo e gás* representou 5,77% do *corpus*. *Serviços de saúde* e *mineração* abrangeram 3,85% cada. *Alimentação* representou 2,88% do *corpus*. *Agricultura*, *serviços postais e de telecomunicações* e *Têxteis* abrangeram 1,92% do *corpus* cada. *Indústrias químicas*, *Construção*, *Silvicultura* e *Mídia* aproximaram-se

de 0,96% do *corpus* cada.

### 3.3.2 GQ2. Quais são os métodos de ciência de dados utilizados nos estudos?

Um aspecto primordial do uso bem-sucedido da ciência de dados é a escolha de métodos adequados. A Tabela 6 mostra as abreviaturas dos métodos de ciência de dados usados em cada artigo, ordenados pelo código de identificação do *corpus*, e a Tabela 20 do Apêndice B contém os nomes dos métodos. *Long short-term memory* (LSTM) foi o método de ciência de dados mais utilizado, aparecendo em 22 artigos, seguido por *Support Vector Machine* (SVM), com 19 aparições, e *Random Forest* (RF), que apareceu 14 vezes. *Rede Neural Convolucional* (CNN) apareceu 11 vezes. *Rede Neural Recorrente* (RNN) apareceu nove vezes. *Perceptron multi-camadas* (MLP) e *Análise de componentes principais* (PCA) apareceram oito vezes cada. *Rede Neural* (NN) apareceu sete vezes. *Média móvel integrada autoregressiva* (ARIMA) e *Regressão Logística* (LR) apareceram seis vezes cada. *Autoencoder* (AE), *Deep Neural Network* (DNN), *Local Outlier Factor* (LOF) e *Synthetic Minority Over-sampling Technique* (SMOTE) apareceram cinco vezes cada. *Rede neural convolucional - memória longa e de curto prazo* (CNN-LSTM), *Clustering espacial de aplicações com ruído baseado em densidade* (DBSCAN), *Gated Recurrent Unit* (GRU), *K-means* (KM), *K-vizinho mais próximo* (KNN), *SVM de uma classe* (OCSVM), *Support Vector Regression* (SVR) e *XGBoost* (XGB) apareceu quatro vezes cada. *AdaBoost* (AB), *Memória bidirecional de longo e curto prazo* (BLSTM), *Rede Neural de retropropagação* (BPNN), *Árvore de decisão* (DT), *Decisão de reforço de gradiente Tree* (GBDT), *modelos de mistura gaussiana* (GMM), *modelos ocultos de Markov* (HMM), *modelo de regressão linear* (LRM) e *Isolation Forest* (iForest) apareceram três vezes cada. *Aglo-merativo Hierárquico Clustering* (AHC), *Memória longa de curto prazo baseada em atenção* (ALSTM), *Rede Neural Artificial* (ANN), *Unidade recorrente controlada bidi-recional* (BGRU), *Bayesian Ridge/Regularização* (BR), *Árvore de Classificação e Regressão* (CART), *Rede Neural Convolucional de Detecção e Classificação de Falhas* (FDC-CNN), *Gradient Boosting Machine* (GBM), *Algoritmo/Análise de Clustering Hierárquico* (HCA), *Análise Discriminante Linear* (LDA), *Perfil de Matriz* (MP), *Ontologia* (Ontologia), *Auto- Organizing Maps* (SOM), *Short-term Fourier Transform* (STFT), *Visual Analytics* (VA) e *Wide-first kernel e Deep Convolutional Neural Network* (WDCNN) apareceram duas vezes cada. Os demais métodos de ciência de dados apareceram apenas uma vez no *corpus*.

Além disso, para melhor acompanhar a evolução ao longo da linha do tempo, a Figura 8 mostra quantas vezes um método de ciência de dados apareceu ao longo dos anos de publicação. *Long Short-Term Memory* (LSTM) foi o método que mais apareceu no *corpus*, com 22 ocorrências. *Support Vector Machine* (SVM) teve 19

Tabela 6 – Métodos de ciência de dados usados nos artigos da revisão sistemática.

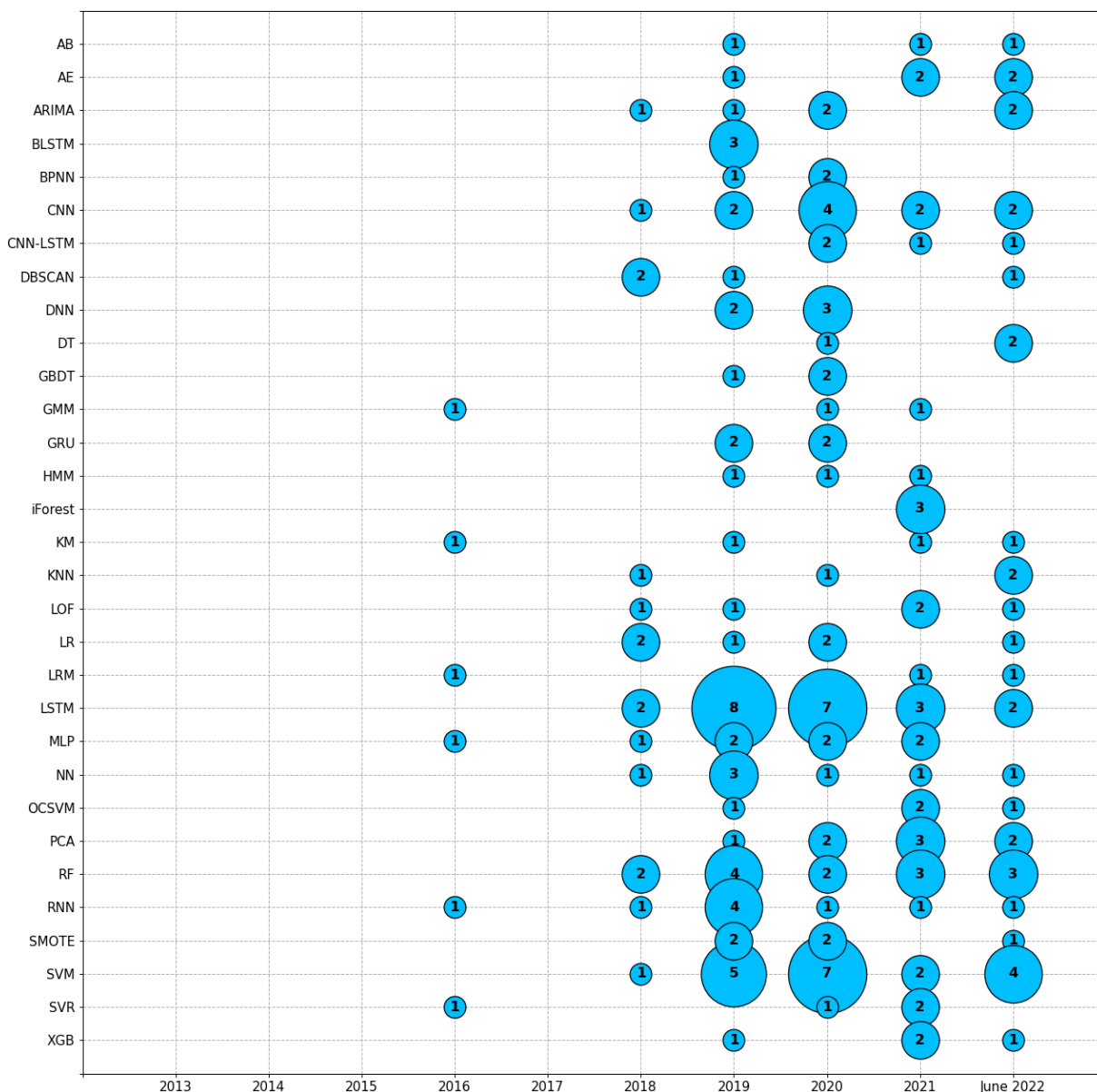
ID	Método(s)	ID	Método(s)
1	CNN, GAF	52	BDA, CNN, DT, GFK, JDA, KNN, LDA, SVM, TCA
2	DWT, LRM, NN, STFT	53	MLP
3	RMS	54	ARIMA, DES
4	AFDC-CNN, CDSAE-AD, CSAE-AD, EncDec-AD, FDC-CNN, LSTM-AD	55	LRM, MLP
5	ANFIS, MLP, NHPP, RBF, SVR, Weibull	56	RNN
6	AE, LOF, RF, iForest	57	CNN
7	DPCA, GBDT	58	ANN, SVM
9	RF	59	EEMD-DL-LSTM
10	ARIMA	61	LSTM, OntoLSTM
11	BINN, I-Forest, OCSVM, PCA	62	MCOD, MP
12	BPNN, DBN, DNN, KNNC, SVM, WPD	63	LOF
13	LSTMAE	64	1D-CNN-LSTM, AAE, DADA, DANN, DCTLN, NLT, STFT, WDCNN
14	2-DConvLSTMAE, ARIMA, CNN-LSTM, DeepLSTM, PersistenceModel, RSNet	65	CST, GA, KM
15	ARIMA, CNN, LSTM, ResNet	66	RNN, SOM
16	LSTM, RNN, SBA	67	CNN, CNN-LSTM, LSTM
17	GBM, RF, SVM, TCN	68	CRISP-DM, DT, KNN, LRM, Prophet, RF, SVM
18	BR	69	LR, LSTM, RF, SVM, TNN
19	GHMM, HMM, MCU	70	CART, GST, LDA, SDM, SVM
20	AE, VGG	71	CNN-LSTM, LSTM
21	AGRU, ALSTM, FFT-MLP, FFT-SVM, GRU, LSTM, RNN-WDCNN, SRDCNN, WDCNN	72	GBT, LR, RF, SVC
22	AOD	73	AE, CMD, CNN, CNN-MMD, KNN, MDDAN, MDIAN, MLCAE, MLCAE-KNN, SVM
23	AE, BGRU, BLSTM, BRNN, GRU, LSTM, RNN	74	MLP, SMOTE, SVM
24	AML, FFNN, RF, XGB	75	AB, CART, GBDT, LGBM, NN, RF, XGB
25	HMM, LSTM, MDP	76	AHC
26	AE, LOF, TSO, iForest	77	LSTM, MORL
27	VQS	78	AHC, SOM, Ward
28	VA	79	BGRU, BLSTM, CNN, GRU, LSTM, RNN
29	AnoGAN, FFT, LMS, LSTM, OCSVM, PCA, Tikhonov, UKF, t-SNE	80	1NN-DTW, FDC-CNN, MC-DCNN, MTS-CNN, SAX-VSM, SF
30	PCA, SSA-BLSTM	81	CDT, DBSCAN, GEC, KNN, NN
31	AE-GRU, DNN, GRU, LSTM, MLP, RNN	82	AEWGAN, LR, RF, SMOTE, SVM, WGAN
32	CNN, PCA, SVM	83	HCA, KM
33	CNN, LSTM	84	DBSCAN, LR, MLP, NB, RF
35	Methontology	85	WSM
36	BGM, GMM, HDBSCAN, MP, PCA	86	ANOVA, SVM, VR
37	CNN, OCSVM, RNN, iForest	87	CNN-LSTM
38	LSTM	88	LSTM
39	GMM, KM, SPIRIT, SVR	89	AB, GBM, MLP, PCA, RF, SVR, XGB
40	GMM, LSTM	90	DF, LR, NN, SVM
41	BNN, GLM, NN, SGB, SVM	91	HCA
42	ARMA, BPNN, LSTM, SVR	92	Ontology
43	ARIMA, DBSCAN, KM, LOF, LSTM, MV, OCSVM	93	SNN
44	NN	94	AR
45	SVM	96	GDN, LSTM-NDT, LSTM-VAE, MTAD-GAT, STGAT-MAD, USAD
46	IDEAaS	97	ARIMA, CNN, DNN, LSTM, MLP, RF, SN, WN, ZO
47	CxDBNet, DBNet	98	DNN, HMM, PCA
48	ANN, SMOTE	99	ALSTM, BPNN, BR, DNN, GBDT, GR, SVM, TSMC-CNN
49	DBSCAN, LOF, LSTM, MAD, RNN, SMOTE, SVM	100	LSTM, RNN
50	AB, DT, NN, PCA, RF, SMOTE, SVM, XGB	101	Ontology
51	VA	103	BLSTM, LR, RF, SVM

Fonte: Arruda et al. (2023a).

ocorrências. O método *Random Forest* (RF) apareceu 14 vezes. Os anos de 2019, 2020 e 2021 apresentaram a maior concentração de métodos de ciência de dados.



Figura 8 – Métodos de ciência de dados da revisão sistemática agrupados por ano.



Fonte: Arruda et al. (2023a).

### 3.3.3 GQ3. Quais são as ferramentas de software utilizadas nos estudos?

A implementação de métodos de ciência de dados requer ferramentas de software adequadas, como linguagens de programação, bancos de dados e kits de ferramentas. A Tabela 7 mostra a abreviatura das ferramentas de software utilizadas em cada artigo do corpus, e a Tabela 21 do Apêndice C, contém os nomes completos das as ferramentas. *Python* foi a ferramenta de software mais utilizada, aparecendo em 20 artigos, seguido por *Keras*, em 15 artigos, e *Tensorflow* em 13. *MATLAB* apareceu em oito trabalhos e a linguagem *R* apareceu em seis. *Hadoop* e *SKLEARN* apareceram em cinco estudos cada. *Kafka* e *MongoDB* apareceram em quatro artigos cada.

*Spark* apareceu em três estudos. *doParallel*, *fastcluster*, *foreach*, *InfluxDB*, *JavaScript*, *Jupyter*, *Knime*, *MES*, *MSSQL*, *PyTorch*, *rpud*, *SQL*, *Storm* e *SWRL* apareceram em dois artigos cada. As demais ferramentas de software apareceram apenas uma vez no corpus.

Tabela 7 – Ferramentas de software usadas nos artigos da revisão sistemática.

ID	Ferramenta(s)	ID	Ferramenta(s)
2	Python, PyWavelets	56	Python, PyTorch
4	Keras, Python, SKLEARN, Tensorflow	57	OpenCV
5	MATLAB	60	Elasticsearch, Flatform, Hadoop, Jupyter, Kafka, Kibana, NiFi, Parquet, Python, Spark, Zeppelin
6	Keras, Python, SKLEARN, Spark, Tensorflow	61	Imblearn
7	CouchDB, freedom, QlikView, R, XGBoost	63	Cassandra, EYE, Hadoop, R, Spark
10	R	65	HealthMon, MATLAB
11	PyOD	66	SQL
13	Keras, Pandas, Python	67	Jupyter, Python, SKLEARN
14	Python, Tensorflow	68	Hadoop, MySQL, Python
15	GAI, GPyOpt, Keras, Tensorflow	69	Colab
18	MATLAB	70	MATLAB
20	Keras, Tensorflow	72	iSTEP, MLlib
21	Keras, Tensorflow	73	Python
23	Keras	74	Knime, RPropMLP
25	InfluxDB, Kafka, RAMI4.0, Storm, UPTIME	76	doParallel, fastcluster, foreach, R, rpud
27	ExtruOnt, Neo4j, RDFox, SPARQL, Stardog, SWRL, Virtuoso	77	BURLAP, ERP, Kafka, Keras, MES, Tensorflow
28	Hadoop, MongoDB	78	doParallel, fastcluster, foreach, kohonen, R, rpud
29	PyTorch, SKLEARN	79	Keras, Tensorflow
32	Python, R	81	Hadoop
34	MATLAB, MES, MSSQL, QlikSense, SSDT, SSIS	84	JavaScript, Kafka, MongoDB, Python, Storm
35	AquaONT, Fuseki, Hermit, OWL, Pallet, Protege, SWRL	85	PlanningVis
37	AnoML-IoT, Python	86	Ruptures
38	Keras, Python, Tensorflow	87	Keras, Tensorflow
39	Python	88	Azure
42	MATLAB	89	Flask, Keras, Python, SKLEARN
43	AngularJS, ChartJS, D3JS, Docker, JavaScript, MongoDB, NodeJS, Python	90	AzureML
46	MongoDB	92	SemML
47	SCADA	94	ARHoloLens, C#, C++, Direct3D, MSSQL
49	MATLAB	95	MATLAB, SPHM
51	MUVTIME	97	Keras, Python, Tensorflow, ThunderML
52	t-SNE, Tensorflow	98	InfluxDB, KafkaStreams, Keras, Tensorflow
53	AZAP	101	SQL
54	Python	102	GADPL
55	Knime, Weka	103	Keras, Python, Theano

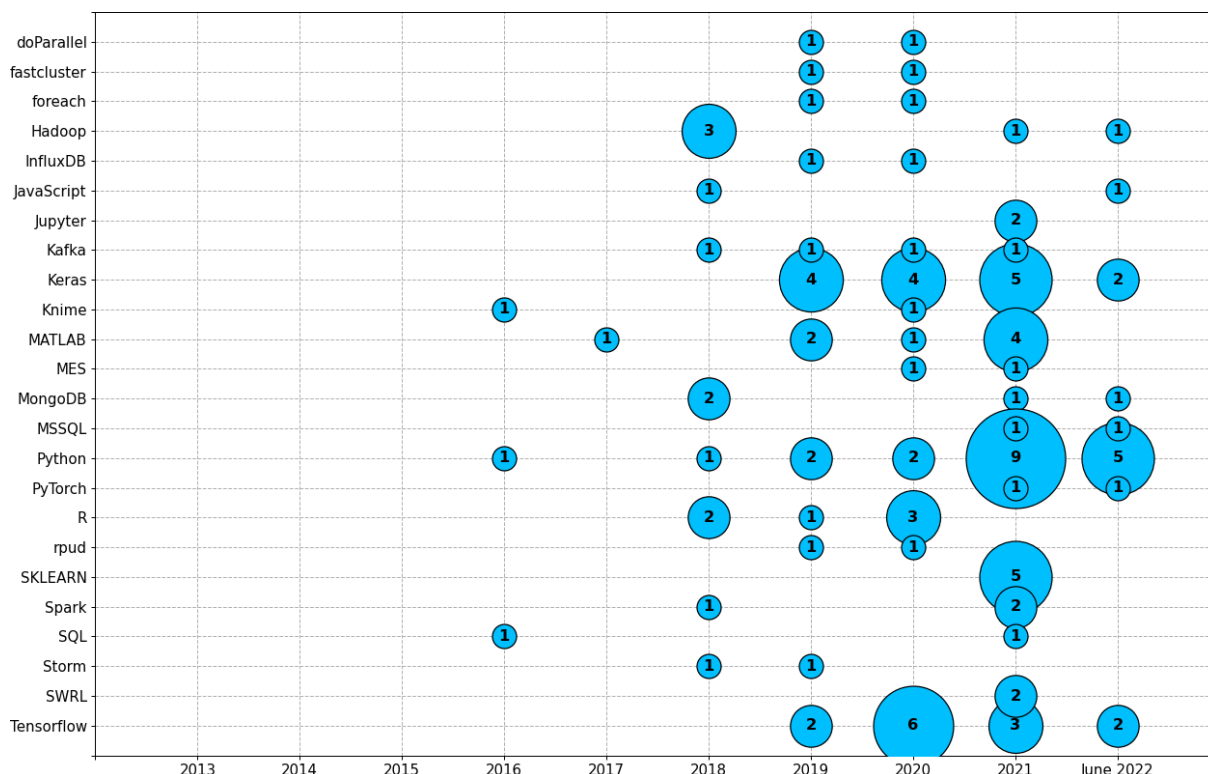
Fonte: Arruda et al. (2023a).

Além disso, a Figura 9 mostra as ferramentas de software agrupadas por anos. A linguagem de programação *Python* foi a ferramenta mais utilizada, aparecendo em 20 artigos, seguida por *Keras*, que apareceu em 15 artigos, e *Tensorflow* que apareceu em 13 artigos.

### 3.3.4 FQ1. Como os estudos empregam séries temporais contextuais?

Onze artigos utilizaram o conceito de contexto de alguma forma. Os trabalhos abordaram ontologias, análise visual, redes bayesianas dinâmicas, sistemas ciberfísi-

Figura 9 – Ferramentas de software da revisão sistemática agrupadas por ano.



Fonte: Arruda et al. (2023a).

cos sensíveis ao contexto, redes neurais convolucionais, redes neurais recorrentes e redes de memória de longo e curto prazo.

Wu et al. (2018) usaram informações de contexto para desenvolver um sistema de análise visual interativo para uma usina petroquímica. O sistema funcionou na fase de operação, utilizando dados de séries temporais de 791 sensores que forneceram o status de diferentes partes da fábrica. Tripathi e Baruah (2020) propuseram um método para identificar anomalias contextuais em um método de rede bayesiana dinâmica (DBN) modificadora de série temporal para suportar informações de contexto, denominado DBN contextual. Os testes de eficácia do novo método ocorreram em dados de perfuração de poços de petróleo. Majdani, Petrovski e Doolan (2016) desenvolveram uma estrutura para sistemas ciberfísicos usando aprendizado de máquina e inteligência computacional. A estrutura utilizou dados de contexto de 25 sensores de diferentes partes de uma turbina a gás. Canizo et al. (2019) propuseram uma arquitetura de rede neural convolucional - rede neural recorrente (CNN-RNN) para extrair recursos e aprender os padrões temporais de dados de séries temporais específicos do contexto de 20 sensores instalados em um elevador de serviço.

Jiang, Lee e Zeng (2020) usaram dois métodos de aprendizado profundo para prever a vida útil restante (RUL) de rolamentos. Os métodos empregaram vetores de contexto em redes de múltiplos canais de séries temporais para redes neurais

convolucionais (TSMC-CNN) e estenderam o método para redes de memória de longo curto prazo baseadas em atenção (TSMC-CNN-ALSTM).

Ståhl et al. (2019) apresentaram um caso de detecção de falhas em chapas de aço usando redes neurais recorrentes bidirecionais (RNN) com mecanismo de atenção. O método utilizou vetores de contexto para representar cada estado do processo. Ma et al. (2020) propuseram uma arquitetura de planejamento de produção preditiva baseada em big data para uma empresa fabricante de cerâmica. A arquitetura usou modelos baseados em cubo para lidar com dados históricos sensíveis ao contexto usando redes LSTM. Yasaei, Hernandez e Faruque (2020) desenvolveram um modelo adaptativo baseado no contexto e baseado em dados usando medidas de 62 sensores heterogêneos de uma estação de águas residuais. O modelo usou redes LSTM para detectar anomalias em dispositivos sensores e anomalias ambientais.

Abbasi, Martinez e Ahmad (2021) desenvolveram uma ontologia para sistemas aquapônicos chamada AquaONT, usando a abordagem metodológica para formular e avaliar o modelo. A ontologia utilizou dados contextuais de uma fazenda padrão para fornecer informações sobre a operação ideal de dispositivos IoT. Bagozi, Bianchini e Antonellis (2021) propuseram uma abordagem focada em sistemas de produção ciberfísicos resilientes (R-CPPS), explorando big data e a perspectiva humana no circuito. O estudo utilizou particionamento de fluxo de dados sensível ao contexto, processando fluxos de dados coletados no mesmo contexto, o que significa a mesma máquina inteligente e o mesmo tipo de processo para produzir o mesmo tipo de produto. Kim, Kim e Ahn (2021) conduziram um experimento para observar a atenção dos participantes em riscos repetidos no local de trabalho, usando a realidade virtual para evitar o risco de lesões. O experimento utilizou uma tarefa de construção para medir os bio sinais dos participantes por meio de sensores de rastreamento ocular e um dispositivo vestível para medir a atividade eletrodérmica, juntamente com recursos contextuais.

### 3.3.5 FQ2. Como está a qualidade dos dados utilizados nos estudos ao longo do tempo ?

A qualidade dos dados é primordial para todos os tipos de segmentos industriais, inclusive as linhas de montagem das indústrias. Conhecer a quantidade de dados ao longo do tempo utilizados em um experimento é fundamental para uma melhor compreensão e análise dos dados. Dos cento e três artigos do corpus, o equivalente a 39,81% (41 artigos) mencionou a quantidade de dados utilizados em um determinado período de tempo. A Tabela 8 apresenta essas informações junto com a identificação do artigo. Apesar de mencionar a quantidade de dados, as unidades de medida apareceram em diferentes formas. Os anos representam a quantidade de dados em 14 estudos, os meses em 17 trabalhos, os dias expressam dados em 7 artigos e as horas

em 3 trabalhos.

Tabela 8 – Quantidade de dados ao longo do tempo dos artigos da revisão sistemática.

ID	Quantidade	ID	Quantidade	ID	Quantidade
6	2 dias	36	1 ano	71	8 dias
7	61 dias	37	2 dias	72	3 anos
8	7 dias	38	2 anos e 6 meses	74	4 anos e 5 meses
10	3655 horas	39	1 ano	76	2 anos
11	5 meses	42	1 mês	77	3 anos
13	3 meses	44	1 ano	78	6 meses
14	1 ano	45	1 ano	84	8 meses
15	4 meses	46	1 ano	85	30 dias
16	2 anos	51	8 meses	87	1 ano
19	3 meses	53	7 anos	88	242 dias
27	1 ano	55	3 meses	98	50 horas
28	3 meses	59	2 meses	102	7 anos
33	3 meses	66	50 horas	103	6 meses
34	6 meses	68	1 ano e 7 meses		

Fonte: Arruda et al. (2023a).

Outro ponto crucial em relação à qualidade dos dados é a origem dos conjuntos de dados utilizados nos experimentos. A Tabela 9 mostra dez artigos do corpus que disponibilizaram seus conjuntos de dados ao público. Três artigos usaram o mesmo repositório, embora dois deles focassem na degradação do motor Turbofan (Lu, Hsu e Huang (2020) e Wu, Ding e Huang (2020)), e o outro em rolamentos (Ding, Yang e Yang (2019)). Shenfield e Howarth (2020) e Kancharla et al. (2022), que trabalharam com dois conjuntos de dados, também usaram rolamentos, mas de repositórios diferentes. Além disso, Apiletti et al. (2018) usaram dados de discos rígidos, Mohsen, Elkaseer e Scholz (2021) trabalharam em um conjunto de dados de atividades humanas, Žvirblis et al. (2022) usaram dados de correias transportadoras, Wahid, Breslin e Intizar (2022) trabalharam com um conjunto de dados de falhas de componentes, e Zhan et al. (2022) usaram dados de turbinas eólicas.

Tabela 9 – Artigos em que os conjuntos de dados estão disponíveis ao público

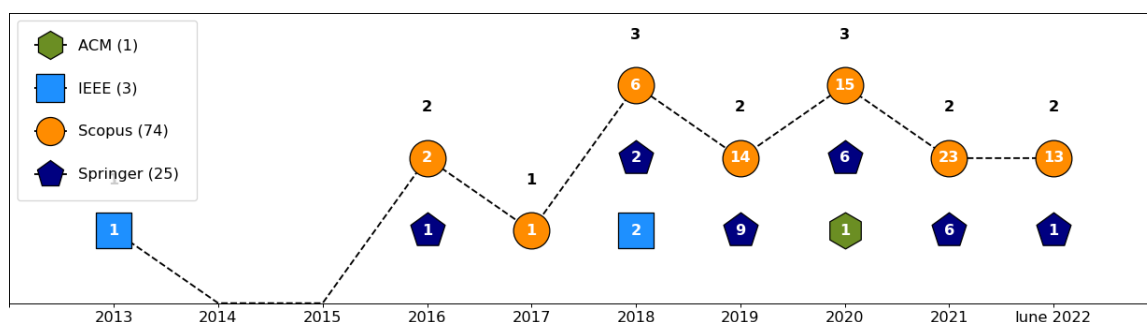
ID	Autor	URL
21	Shenfield e Howarth (2020)	<a href="https://engineering.case.edu/bearingdatacenter/download-data-file">https://engineering.case.edu/bearingdatacenter/download-data-file</a>
23	Ding, Yang e Yang (2019)	<a href="https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing">https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#bearing</a>
31	Lu, Hsu e Huang (2020)	<a href="https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan">https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan</a>
40	Wu, Ding e Huang (2020)	<a href="https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan">https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/#turbofan</a>
67	Mohsen, Elkaseer e Scholz (2021)	<a href="https://www.kaggle.com/datasets/drsaeedmohsen/wisdmdataset2021">https://www.kaggle.com/datasets/drsaeedmohsen/wisdmdataset2021</a>
69	Žvirblis et al. (2022)	<a href="https://github.com/TadasZvirblis/CORBEL">https://github.com/TadasZvirblis/CORBEL</a>
72	Apiletti et al. (2018)	<a href="https://www.backblaze.com/b2/hard-drive-test-data.html">https://www.backblaze.com/b2/hard-drive-test-data.html</a>
73	Kancharla et al. (2022)	<a href="https://engineering.case.edu/bearingdatacenter/download-data-file">https://engineering.case.edu/bearingdatacenter/download-data-file</a> <a href="https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter">https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter</a>
87	Wahid, Breslin e Intizar (2022)	<a href="https://github.com/ashishpatel26/Predictive_Maintenance_using_Machine-Learning_Microsoft_Casestudy">https://github.com/ashishpatel26/Predictive_Maintenance_using_Machine-Learning_Microsoft_Casestudy</a>
96	Zhan et al. (2022)	<a href="https://github.com/zhanjun717/STGAT">https://github.com/zhanjun717/STGAT</a>

Fonte: Arruda et al. (2023a).

### 3.3.6 SQ1. Em quais bases de dados os estudos estão publicados

A revisão aplicou as buscas em cinco bases de dados: ACM, IEEE, Scopus, Springer e Wiley. No entanto, apenas quatro bases de dados tiveram estudos selecionados para o corpus, conforme mostrado na Figura 10. A Scopus teve a grande maioria dos artigos (71,84%), seguida pela Springer (24,27%), IEEE (2,91%) e ACM (0,97%).

Figura 10 – Quantidade de artigos por ano e banco dados da revisão sistemática.

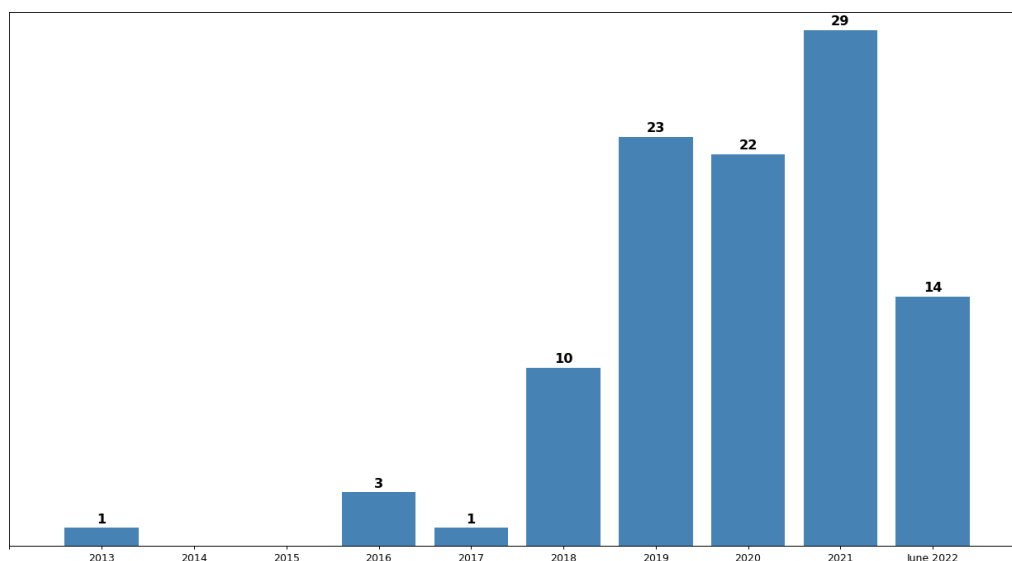


Fonte: Arruda et al. (2023a).

### 3.3.7 SQ2. Qual é o número de publicações por ano?

Nos últimos cinco anos, as publicações relacionadas a este estudo aumentaram, duplicando de 2018 (10 artigos) para 2019 (23 artigos). A figura 11 mostra o andamento anual das publicações, levando em consideração a data de publicação. A primeira publicação que se enquadrava nos critérios de seleção foi em 2013 e a última em 2022. Apenas quatorze trabalhos surgiram até o final de junho de 2022 por ser esta a data em que as buscas foram realizadas.

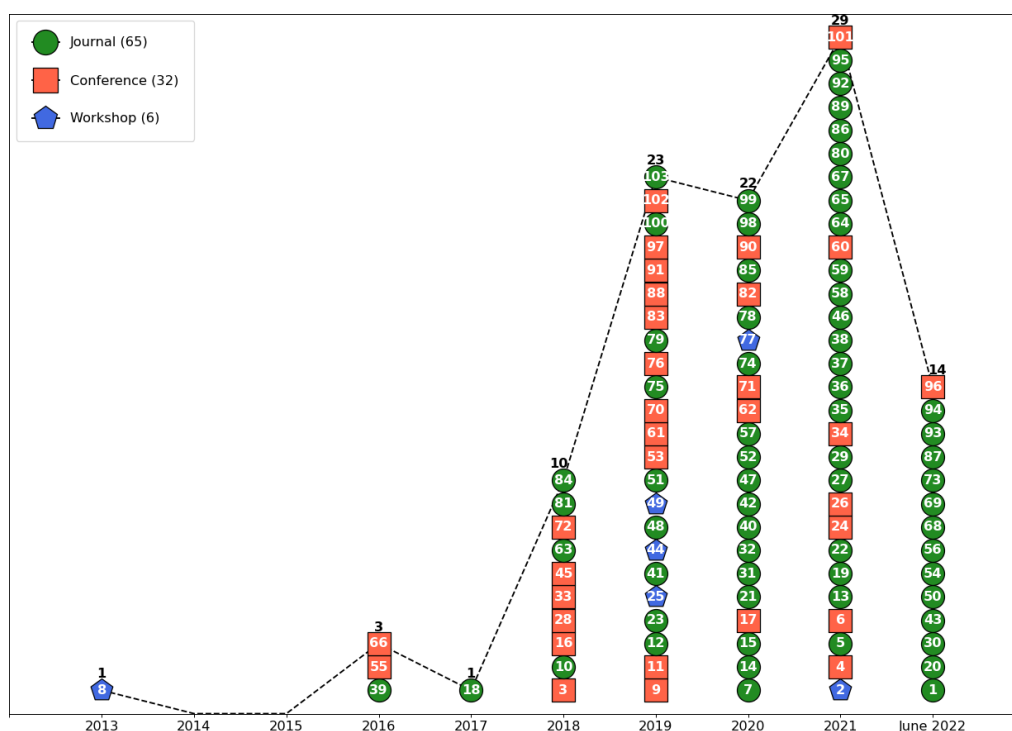
Figura 11 – Número de publicações presentes na revisão sistemática por ano.



Fonte: Arruda et al. (2023a).

Em relação aos tipos de publicações, a Figura 12 mostra o código de identificação do artigo dentro de uma forma geométrica. Os trabalhos de conferências usam um símbolo quadrado, os artigos de periódicos usam um círculo e os trabalhos de workshops usam um símbolo de diamante. Os periódicos tiveram o maior número de artigos (63,11%), seguidos por conferências (31,07%) e workshops (5,83%).

Figura 12 – Tipos de publicação da revisão sistemática por ano.

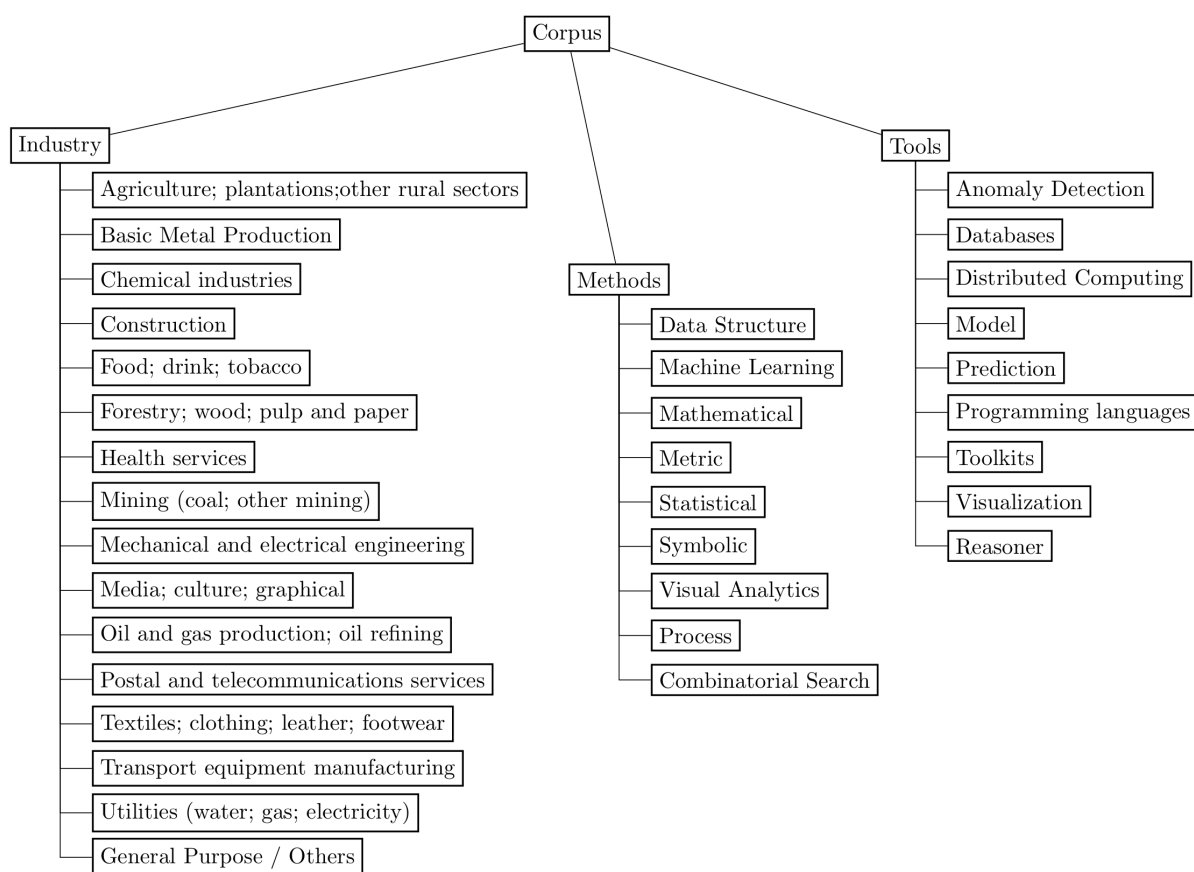


Fonte: Arruda et al. (2023a).

### 3.4 Taxonomia

Esta seção resume as respostas às três questões gerais de pesquisa, apresentadas anteriormente na Tabela 2, usando uma abordagem taxonômica para melhor visualizar e compreender os resultados. A Figura 13 descreve uma taxonomia que organiza, classifica e sintetiza hierarquicamente os segmentos industriais (GQ1), métodos de ciência de dados (GQ2) e ferramentas de software (GQ3) encontrados no corpus com os nós *Industry* (ESTEVE-GONZALEZ et al., 2023), *Methods* (PICCIALLI et al., 2020; SOUSA LIMA et al., 2018; SCHRÖER; KRUSE; GÓMEZ, 2021) e *Tools* (ORDONEZ, 2020; BARLAS; LANNING; HEAVEY, 2015), respectivamente. Os segmentos industriais apresentaram dezesseis classes, os métodos de ciência de dados organizaram algoritmos e técnicas em nove ramos e as ferramentas de software apresentaram aplicações e bibliotecas organizadas em nove componentes.

Figura 13 – Os eixos principais da taxonomia da revisão sistemática.



Fonte: Arruda et al. (2023a).

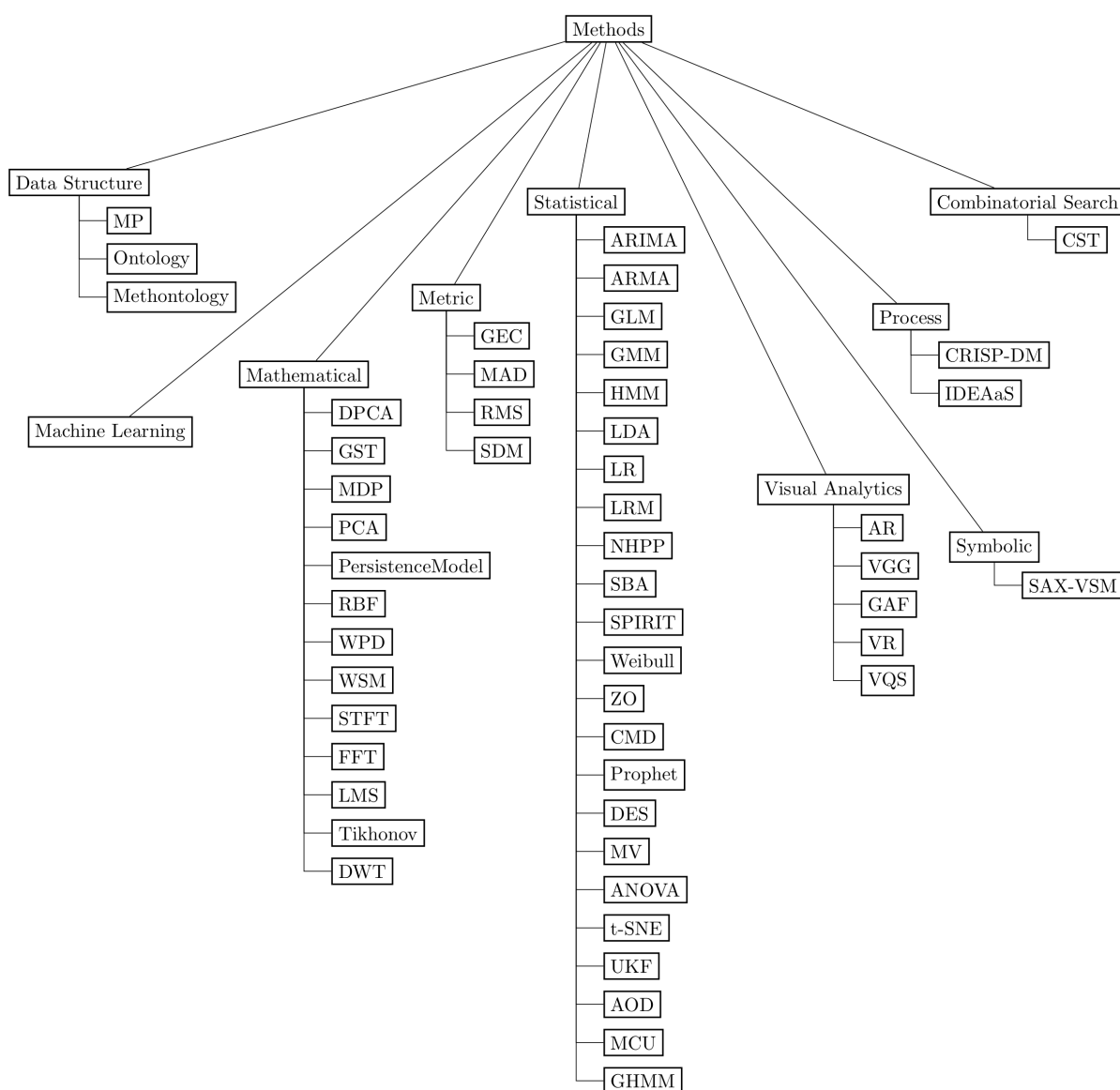
Os segmentos industriais utilizados neste trabalho originaram-se da Organização Internacional do Trabalho (ILO) (<https://www.ilo.org/global/industries-and-sectors>; acessado em 22 de maio de 2023), uma agência das Nações Unidas, que classifica indústrias e setores em 22 segmentos. Os 103 artigos resultantes da revisão sistemática



enquadraram-se em 15 dos 22 segmentos propostos pela ILO: *agricultura, produção de metais básicos, indústrias químicas, construção, alimentação, silvicultura, serviços de saúde, mineração, engenharia mecânica e elétrica, mídia, petróleo e gás, postal e serviços de telecomunicações, têxteis, fabricação de equipamentos de transporte e utilidades*. Esses diferentes segmentos complementam essas indústrias com *uso geral*.

Os métodos de ciência de dados encontrados incluem *estrutura de dados, aprendizado de máquina, matemático, métrico, estatístico, simbólico, análise visual, processo e busca combinatória*, conforme mostrado na taxonomia e detalhado na Figura 14.

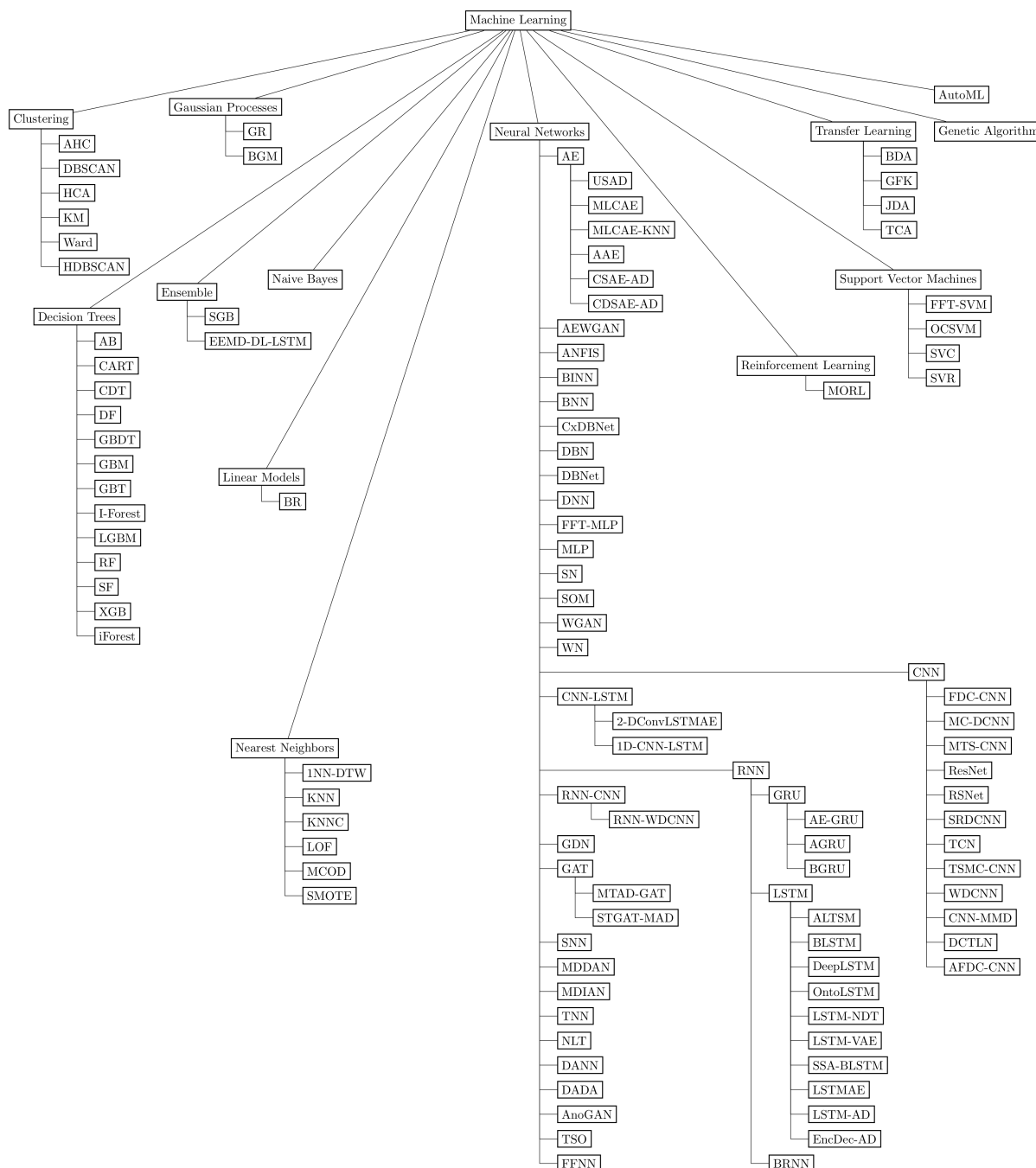
Figura 14 – O eixo Métodos da taxonomia da revisão sistemática.



Fonte: Arruda et al. (2023a).

Devido ao número significativo de métodos e suas variações, o ramo *aprendizado de máquina* teve uma taxonomia separada mostrada na Figura 15.

Figura 15 – O eixo Aprendizado de Máquina da taxonomia da revisão sistemática.



Fonte: Arruda et al. (2023a).

As redes do método de aprendizado de máquina *long short-term memory* (LSTM) representaram o método mais utilizado, com 22 ocorrências. Além disso, havia dez variações de LSTM: *memória de longo prazo baseada em atenção* (ALSTM), que usa um vetor de contexto para inferir diferentes graus de atenção de características de dados distintas em pontos de tempo específicos (JIANG; LEE; ZENG, 2020); *memória bidirecional de longo e curto prazo* (BLSTM), que processa dados tanto em ordem cronológica, do início ao fim, quanto na direção oposta, a ordem inversa (CANIZO

et al., 2019; STÅHL et al., 2019); *deep long short term memory* (DeepLSTM), uma rede LSTM com camadas empilhadas conectadas a uma camada densa distribuída ao longo do tempo (ESSIEN; GIANNETTI, 2020); *memória longa de curto prazo com limiar dinâmico não paramétrico* (LSTM-NDT) (ZHAN et al., 2022); *autoencoder variacional de memória de longo e curto prazo* (LSTM-VAE) (ZHAN et al., 2022); *análise de espectro singular, memória bidirecional de longo e curto prazo* (SSA-BLSTM) (MAHMOOD; LUO; REHMAN, 2022); *autoencoder de memória de longo e curto prazo* (LSTM-MAE) (BAMPOULA et al., 2021); *detecção de anomalias de memória de longo e curto prazo* (LSTM-AD) (TCHATCHOUA et al., 2021). *detecção de anomalia do codificador-decodificador* (EncDec-AD) (TCHATCHOUA et al., 2021); e a *rede neural LSTM baseada em ontologia* (OntoLSTM), que implementa conceitos semânticos usando uma ontologia para aprender a representação de uma linha de produção, juntamente com uma rede LSTM para aprendizagem de dependências temporais (HUANG; ZANNIMERK; CRÉMILLEUX, 2019).

O segundo método de ciência de dados mais utilizado foi o método *support vector machine* (SVM), representando 19 ocorrências. Além disso, o método tinha quatro variações: *máquinas de vetores de suporte baseadas na transformada rápida de Fourier* (FFT-SVM), uma versão do SVM que usa uma transformada rápida de Fourier para extrair recursos (SHENFIELD; HOWARTH, 2020); *one-class SVM* (OCSVM), uma versão não supervisionada de SVM que usa uma única classe para identificar dados semelhantes ou diferentes (CARLETTI et al., 2019); *classificação vetorial de suporte* (SVC), uma variação usada para tarefas de classificação (APILETTI et al., 2018); e a variação *regressão vetorial de suporte* (SVR), que implementa uma função de regressão linear aos dados mapeados (VRIES et al., 2016).

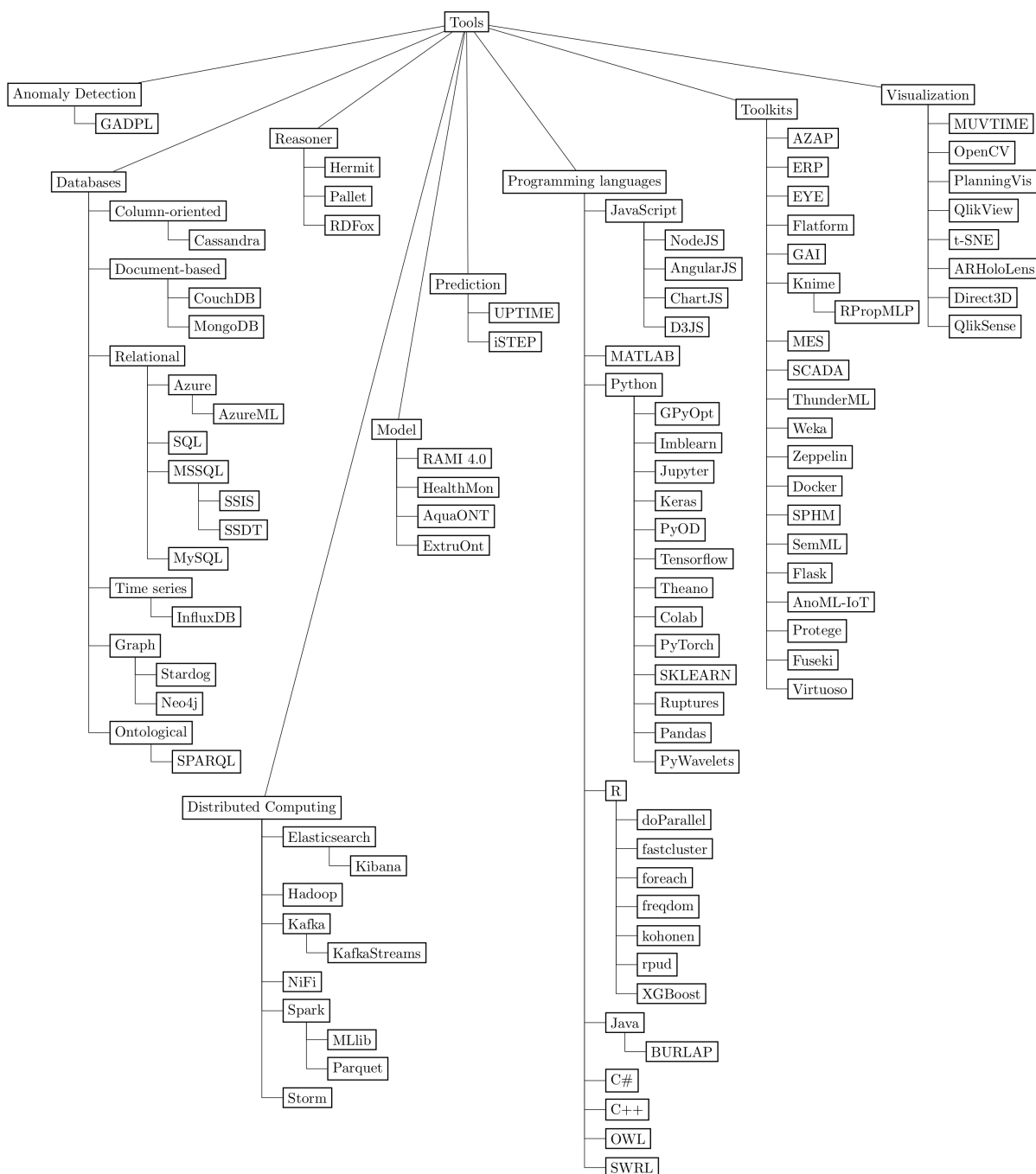
O método de ciência de dados que foi o terceiro mais utilizado foi o método de árvore de decisão *Random Forest* (RF), acumulando 14 ocorrências, seguido por *Convolutional Neural Network* (CNN), com 11 ocorrências, e *Recurrent Neural Network* (RNN), com 9 ocorrências. Doze variações da CNN se destacaram como ramos: *Fault Detection and Classification Convolutional Neural Network* (FDC-CNN), projetada para detectar falhas em sinais de sensores multivariados ao longo de um eixo de tempo, extraíndo características de falha; *Multichannel Deep Convolutional Neural Networks* (MC-DCNN), cujo objetivo é lidar com múltiplos sensores que geram dados com diferentes comprimentos; *Multiple-Time-Series Convolution Neural Network* (MTS-CNN), projetada para diagnóstico e detecção de falhas de séries temporais, usa uma CNN

multicanal para extrair recursos de dados importantes (HSU; LIU, 2021); *Temporal Convolutional Network* (TCN), que funciona resumindo sinais em intervalos de tempo, usando um valor máximo e mínimo por passo (VAN HERREWEGHE et al., 2020); *Residual Neural Networks* (ResNet) (VILLALOBOS; SUYKENS; ILLARRAMENDI, 2020); *Residual-Squeeze Net* (RSNet) (ESSIEN; GIANNETTI, 2020); *Stacked Residual Dilated Convolutional Neural Network* (SRDCNN) (SHENFIELD; HOWARTH, 2020); *Wide First Kernel and Deep Convolutional Neural Network* (WDCNN) (SHENFIELD; HOWARTH, 2020; JANG; CHO, 2021); *Convolutional Neural Network Maximum Mean Discrepancy* (CNN-MMD) (KANCHARLA et al., 2022); *Deep Convolutional Transfer Learning Network* (DCTLN) (JANG; CHO, 2021); *Attention Fault Detection and Classification Convolutional Neural Network* (AFDC-CNN) (TCHATCHOUA et al., 2021); e a *Time-Series Multiple-Channel Convolutional Neural Network* (TSMC-CNN), que usa como entradas séries temporais com N variáveis divididas em segmentos, suavizando a extração de pontos de dados (JIANG; LEE; ZENG, 2020). RNN representou três ramos: *Gated Recurrent Unit* (GRU), *Long Short-Term Memory* (LSTM) e *Bidirectional Recurrent Neural Network* (BRNN).

Em relação às ferramentas de software, nove classes principais apareceram na taxonomia: *Anomaly Detection*, *Databases*, *Distributed Computing*, *Model*, *Prediction*, *Programming Languages*, *Toolkits*, *Visualization* e *Reasoner*, conforme representado na Figura 16. A linguagem *Python* foi a ferramenta de software mais utilizada, com 20 ocorrências, seguida por *Keras* (15 ocorrências) e *Tensorflow* (13 ocorrências). *Keras* é uma estrutura de aprendizado profundo e *Tensorflow* é um back-end de aprendizado de máquina (SHENFIELD; HOWARTH, 2020), e ambos são ramos de *Python* na hierarquia da taxonomia.

Apesar de cobrir hierarquicamente segmentos industriais, métodos de ciência de dados e ferramentas de software, a taxonomia não os vinculou horizontalmente. Essas relações estão na Tabela 5, representando segmentos industriais, Tabela 6 mostrando métodos de ciência de dados, e Tabela 7 fornecendo ferramentas de software.

Figura 16 – O eixo Ferramentas da taxonomia da revisão sistemática.



Fonte: Arruda et al. (2023a).

### 3.5 Discussão

Os resultados apresentados neste capítulo originaram-se de um processo de revisão sistemática focado na Indústria 4.0, ciência de dados e séries temporais. Não houve restrição quanto ao ano de publicação para fornecer todo um espectro de literatura nessas áreas mencionadas. Com isso, a revisão mostrou aplicações do segmento

industrial tanto a partir de casos reais quanto de ambientes simulados, além de identificar métodos de ciência de dados, ferramentas de software e a qualidade dos dados utilizados pelos experimentos.

Diversos segmentos industriais apresentaram interesse em analisar dados, e cada vez mais a análise de dados é crucial para as empresas. Isso contribui para a tomada de decisões em função dos dados históricos gerados por cada indústria. Além disso, estes processos analíticos de dados contribuem para as necessidades específicas das empresas, uma vez que as experiências anteriores são substanciais para melhorar os resultados futuros.

Os segmentos industriais explorados pela literatura foram classificados e agrupados segundo o padrão da Organização Internacional do Trabalho. Isso proporcionou uma melhor forma de visualização na taxonomia (Figura 13). O segmento industrial *uso geral/outros* apareceu em 25 artigos, sendo o mais presente no corpus. O segmento industrial *engenharia mecânica e elétrica* foi o segundo mais publicado (20 artigos). O segmento inclui indústrias estritamente ligadas à tecnologia, como semicondutores, informática e eletrônicos, o que explica por que foi o segmento mais frequente no estudo, depois do *uso geral/outros*. Além disso, essa indústria costuma contar com ambientes controlados e funcionários treinados para trabalhar com tecnologia, simplificando a coleta de dados. Isso favorece a execução de estudos porque esses ambientes industriais já estão preparados para produzir combinações de dados para a tomada de decisões de alto nível.

A maioria dos estudos utilizou instalações industriais reais nos experimentos (81 artigos). Porém, alguns artigos empregaram ambientes simulados (23 trabalhos). O trabalho de Luo, Liu e Qi (2019) apareceu duas vezes nos casos simulados devido à presença de dois segmentos industriais no artigo. O uso de dados reais na maioria dos artigos fornece evidências da evolução das aplicações de ciência de dados na linha de produção da indústria. Isso ocorre porque os sensores e as ferramentas de banco de dados evoluíram e se tornaram mais acessíveis nos últimos anos. Além disso, a qualidade dos conjuntos de dados reais é um ponto positivo para o treinamento de algoritmos de aprendizado de máquina, pois pode melhorar a precisão dos modelos preditivos e fundamentar aplicações futuras que utilizem o mesmo tipo de dados. Isto também é positivo porque reflete cenários industriais reais e potencialmente fornece tecnologia para problemas do mundo real.

Além disso, a literatura apresenta uma ampla utilização de diferentes tecnologias, o que pode dificultar a escolha correta de um método adequado, uma vez que existe a possibilidade de emprego empírico dos métodos. Além dos métodos, escolher a ferramenta certa é outro desafio devido às diferentes implementações do mesmo método em ferramentas distintas, por exemplo, linguagens de programação que apresentam valores alternativos para inicializar os pesos de uma rede neural. Algumas ferramentas

dependem de métodos específicos, como a ferramenta Keras, que trata de aplicações de aprendizagem profunda empregando métodos LSTM e GRU. Além disso, é comum ver as ferramentas Keras e Tensorflow usadas juntas (VILLALOBOS; SUYKENS; ILLARRAMENDI, 2020; SHENFIELD; HOWARTH, 2020; LEPENIOTI et al., 2020; CANIZO et al., 2019; SHRIVASTAVA et al., 2019; CHEN et al., 2020). Tanto Keras quanto Tensorflow suportam a linguagem Python, que é amplamente utilizada para fins científicos, aparecendo em 20 artigos do corpus, conforme apresentado na Tabela 7. Por outro lado, em relação ao uso da combinação de dados para criar informações de alto nível, o corpus incluiu 11 artigos que mencionaram dados contextuais (WU et al., 2018; MA et al., 2020; TRIPATHI; BARUAH, 2020; MAJDANI; PETROVSKI; DOOLAN, 2016; YASAEI; HERNANDEZ; FARUQUE, 2020; CANIZO et al., 2019; JIANG; LEE; ZENG, 2020; STÅHL et al., 2019; ABBASI; MARTINEZ; AHMAD, 2021; BAGOZI; BIANCHINI; ANTONELLIS, 2021; KIM; KIM; AHN, 2021).

Além das tecnologias citadas, as redes neurais estavam entre as 13 variações de métodos de aprendizado de máquina de acordo com a taxonomia. Por outro lado, as próprias redes neurais apresentaram 31 subvariações. Com o aprimoramento desse método de aprendizado de máquina, três abordagens se destacaram: redes baseadas em atenção, bidirecionais e autoencoder. O mecanismo baseado na atenção atua como o comportamento da atenção visual humana, usando um vetor de contexto e focando na importância de diferentes recursos em intervalos de tempo distintos para melhorar a precisão da previsão. Os estudos que se concentraram neste mecanismo baseado na atenção exploraram o uso de, por exemplo, ALSTM e AGRU. Os modelos bidirecionais funcionam como duas redes neurais diferentes percorrendo uma sequência de dados em ambas as direções para evitar o esquecimento de dados. Uma rede vai do início ao fim da sequência e a outra vem na direção oposta. Nesse sentido, os estudos abrangeram o uso de BLSTM, BGRU e BRNN. Um autoencoder é uma rede neural feed-forward não supervisionada comumente usada para extração de recursos e redução de dimensionalidade, composta por um codificador e um decodificador. O codificador compacta os dados em uma camada oculta e o decodificador os remonta aos dados de entrada originais. Em particular, os estudos utilizaram 2-DConvLSTMAE, AEWGAN, AE-GRU e AE. Consequentemente, essas técnicas focaram em novas combinações e variações de redes neurais, que fornecem métodos versáteis para explorar problemas e questões no âmbito da ciência de dados nas indústrias.

Mais especificamente, a análise da qualidade dos dados é fundamental para garantir o bom funcionamento dos métodos de ciência de dados acima mencionados. A falta de detalhes na composição dos dados pode dificultar a compreensão do artigo e a reprodutibilidade do experimento. A quantidade de dados ao longo do tempo não é suficiente para fornecer todas as informações necessárias, pois a frequência pode

variar durante o mesmo período. Por exemplo, é possível medir a temperatura do ar a cada hora ou a cada minuto do dia. Se a medição ocorrer a cada hora, resultará em 24 linhas. Por outro lado, se a medição ocorrer a cada minuto, resultará em 1.440 linhas. Portanto, essas medições fornecem granularidade de dados diferente, o que consequentemente afeta a forma como os resultados são descritos. Mais importante ainda, estes casos requerem uma exposição adequada às metodologias e discussões considerando a especificidade do método.

Em relação às estruturas de dados encontradas nos métodos, as ontologias fornecem uma forma avançada de recuperar informações. Classes e relações organizam os dados como uma taxonomia, mas com a possibilidade de consultar e raciocinar. O SPARQL é a linguagem usada para recuperar informações e Hermit, Pallet e RDFox são exemplos de raciocinadores encontrados na revisão. Um aspecto importante das ontologias é que elas são extensíveis e reutilizáveis (ABBASI; MARTINEZ; AHMAD, 2021; HUANG; ZANNI-MERK; CRÉMILLEUX, 2019; ZHOU et al., 2021a).

Além disso, outra informação crucial que os estudos devem fornecer claramente é a porcentagem de dados utilizados para treinar e testar o modelo, porque esta estratégia de divisão de dados afeta diretamente os resultados. Além disso, para garantir a reprodutibilidade do experimento, alguns detalhes específicos dos métodos são de significativa importância, por exemplo, o número de camadas ocultas de uma rede neural, ou o tipo de kernel utilizado por uma máquina de vetores de suporte, ou mesmo o número de interações usado por uma floresta aleatória. Nesse sentido, há necessidade de estudos que apresentem mais sobre a organização dos dados e como os métodos de ciência de dados foram empregados. Os artigos devem incluir todos os detalhes da implementação, como a arquitetura e os parâmetros dos métodos de aprendizado de máquina e toda a composição dos vetores de atributos. Com isso, os pesquisadores encontrarão as metodologias mais claras para serem compreendidas e reproduzidas em seus estudos. Assim, isso beneficiará a comunidade, garantindo potenciais situações comuns entre os diferentes segmentos para evitar aspectos técnicos e gerenciais.

### **3.6 Conclusão da Revisão Sistemática**

Este capítulo apresentou uma revisão sistemática da literatura focada na Indústria 4.0, ciência de dados e séries temporais. Investigou a utilização de métodos de ciência de dados e ferramentas de software em diversos segmentos industriais, levando em consideração a implementação de séries temporais e a qualidade dos dados empregada pelos autores. Além disso, uma taxonomia organizou os segmentos industriais, os métodos de ciência de dados e as ferramentas de software de forma hierárquica e sintetizada, o que facilitou a leitura de como os estudos da Indústria 4.0 têm empre-



gado essas tecnologias.

A literatura apresentou vários métodos maduros que cobriram vastas possibilidades para análise industrial. Isto fortalece tanto o mercado como a academia porque quanto mais as empresas empregam as tecnologias, mais pesquisadores e profissionais se tornam especialistas nesses métodos e ferramentas. Nesse sentido, o investimento industrial nessas análises é benéfico porque fornece resultados empíricos para a comunidade sobre casos de uso aplicáveis em diversos segmentos. Além disso, contribui para o amadurecimento e evolução dos métodos e ferramentas tecnológicas empregadas no processo de análise de dados industriais.

Mesmo com esforços para reduzir vieses, esta revisão apresenta limitações como qualquer outra revisão sistemática. A string de busca foi aplicada a cinco bases de dados de pesquisa com o intuito de utilizar diferentes fontes acadêmicas, o que potencialmente diminuiu o viés da fonte. A concepção da string de busca utilizou três eixos empregando as respectivas palavras-chave conhecidas e sinônimos para cada eixo, com foco na redução de vieses de palavras-chave. Além disso, seis critérios de exclusão filtraram os artigos resultantes, gerando o corpus. Assim, esses critérios de exclusão e o restante do processo de filtragem seguiram as diretrizes de Petersen, Vakkalanka e Kuzniarz (2015) para reduzir o viés do processo.

A taxonomia representa uma importante contribuição para futuras pesquisas, uma vez que a organização de métodos de ciência de dados e ferramentas de software auxilia na busca visual em categorias, auxiliando na descoberta de lacunas de pesquisa. Além disso, a variação de um método ou ferramenta específica em um nó aponta tendências no uso daquela tecnologia, o que é importante na hora de escolher qual técnica utilizar. Portanto, a faculdade da taxonomia de organizar e classificar os resultados em classes hierárquicas constitui uma conquista relevante deste trabalho. Além disso, a classe *Industry* foi uma tentativa de padronizar os segmentos de acordo com a Organização Internacional do Trabalho. Assim, a visualização dos resultados na forma de uma taxonomia aumenta as possibilidades de novas pesquisas.

Por fim, esta pesquisa não se concentrou em como os trabalhos lidavam com o tratamento de dados antes de aplicar métodos de ciência de dados aos conjuntos de dados. Esta situação constitui uma limitação adicional e, portanto, é sugerida como trabalho futuro. Além disso, a forma como as ferramentas de software estão ligadas aos métodos de ciência de dados é outro potencial trabalho futuro. Além disso, o último tópico sugerido para trabalhos futuros é correlacionar especificamente os métodos e ferramentas mais utilizados com cada segmento industrial.

### 3.7 Trabalhos Seleccionados e Comparativo

A partir do mapeamento sistemático realizado, com a seleção de 103 trabalhos que respondem as questões de pesquisa elaboradas, foram seleccionados os trabalhos que apresentassem referências a ideia de contexto ou a utilização de ontologias, ficando um total de 10 trabalhos. Entre os trabalhos seleccionados, dois utilizam ontologias e nove aplicam a ideia de contexto. Apenas um trabalho apresentou a aplicação de ontologia e contexto juntos.

O trabalho desenvolvido por Wu et al. (2018) adota técnicas de análise visual iterativa com base nos dados de 791 sensores, que fornecem informações sobre o *status* de diferentes partes de uma indústria petroquímica. O sistema utilizou séries temporais de contexto da área de operação.

A ontologia proposta por Abbasi, Martinez e Ahmad (2021) é voltada para fazendas aquapônicas, que utilizam um método de plantação vertical combinando hidroponia e aquicultura. Utiliza seis classes para representar os domínios de conhecimento focados em produtos de consumo, sistemas de produção, instalações de produção, ambiente, qualidade do produto e dados contextuais. A ontologia, intitulada AquaONT, foi desenvolvida com a ferramenta Protégé e apresenta como uma de suas funcionalidades o acompanhamento das condições ambientais internas da fazenda, informando aos usuários alterações no ambiente que demandem algum tipo de ação.

A arquitetura proposta por Ma et al. (2020) realiza previsões para o planejamento de produção, baseado em *big data*, para uma manufatura de cerâmicas. A arquitetura usa modelos baseados em cubo com redes neurais LSTM para lidar com dados históricos sensíveis ao contexto.

O artigo usa Redes Bayesianas Dinâmicas modificadas, intituladas pelos autores de *Context DBN* para identificar anomalias de contexto em séries temporais. Os testes de eficácia do método proposto ocorreram com dados de poços de perfuração de petróleo (TRIPATHI; BARUAH, 2020).

O trabalho de Majdani, Petrovski e Doolan (2016) utiliza aprendizado de máquina e inteligência computacional aplicados a sistemas ciber físicos. O framework desenvol-

vido usou dados de contexto de 25 sensores instalados em diferentes partes de uma turbina a gás.

O modelo proposto por Huang, Zanni-Merk e Crémilleux (2019) utiliza uma ontologia em conjunto com redes neurais profundas para trabalhar com séries temporais de uma linha de montagem industrial. A ontologia é responsável por gerar as camadas densas da rede neural que tem o objetivo de aprender a representação da linha de montagem e das máquinas de um processo de manufatura.

O artigo de Yasaei, Hernandez e Faruque (2020) propõe um modelo direcionado a dados, adaptativo e sensível ao contexto, usando medidas de diferentes tipos de sensores de uma estação de águas residuais. O modelo usa redes neurais LSTM para detectar anomalias tanto nos dispositivos quanto no ambiente.

O artigo usa Redes Neurais Convolucionais em conjunto com Redes Neurais Recorrentes (CNN-RNN) para extrair características e encontrar padrões em séries temporais de um contexto específico. A arquitetura usa dados de 20 sensores instalados em um elevador de serviço (CANIZO et al., 2019).

O trabalho de Jiang, Lee e Zeng (2020) aplicou duas técnicas de aprendizado profundo para estimar a vida útil de rolamentos usando vetores de contexto. Os métodos aplicados implementaram variações de redes convolucionais com múltiplos canais e redes LSTM com mecanismo de atenção.

O artigo de Ståhl et al. (2019) usa detecção de falhas em chapas de aço como estudo de caso. São utilizadas redes neurais recorrentes bidirecionais com mecanismo de atenção em conjunto com vetores de contexto, que representam cada um dos estados do processo.

A Tabela 10 apresenta um comparativo entre os trabalhos selecionados e o modelo proposto, indicando os critérios que foram utilizados na comparação, juntamente com o seu identificador no Corpus e os autores. O primeiro item, após as referências, indica se a ideia de contexto segundo Dey, Abowd e Salber (2001) foi implementada no trabalho, o segundo e o terceiro itens apontam se os métodos e ferramentas foram

citados de forma explícita pelos autores, e o último item indica se houve uso de algum tipo de ontologia no trabalho.

Tabela 10 – Comparativo dos trabalhos relacionados e modelo proposto.

ID	Referências	Contexto	Métodos	Ferramentas	Ontologia
28	Wu et al. (2018)	Sim	Sim	Sim	Não
35	Abbasi, Martinez e Ahmad (2021)	Sim	Sim	Sim	Sim
42	Ma et al. (2020)	Sim	Sim	Sim	Não
47	Tripathi e Baruah (2020)	Sim	Sim	Sim	Não
55	Majdani, Petrovski e Doolan (2016)	Sim	Sim	Sim	Não
61	Huang, Zanni-Merk e Crémilleux (2019)	Não	Sim	Sim	Sim
71	Yasaei, Hernandez e Faruque (2020)	Sim	Sim	Não	Não
79	Canizo et al. (2019)	Sim	Sim	Sim	Não
99	Jiang, Lee e Zeng (2020)	Sim	Sim	Não	Não
103	Ståhl et al. (2019)	Sim	Sim	Sim	Não
	Modelo Odisseu	Sim	Sim	Sim	Sim

Fonte: Elaborado pelo autor.

Em comparação com os trabalhos selecionados, o modelo Odisseu suporta não somente a ideia de contextos mas a utilização de históricos de contextos. Além disso, diferentes métodos e ferramentas de ciência dados podem ser utilizados para o desenvolvimento dos serviços inteligentes suportados pelo modelo. Por fim, através do uso de uma ontologia o Odisseu estabelece um elo de comunicação entre os dados dos históricos de contextos e os serviços inteligentes.

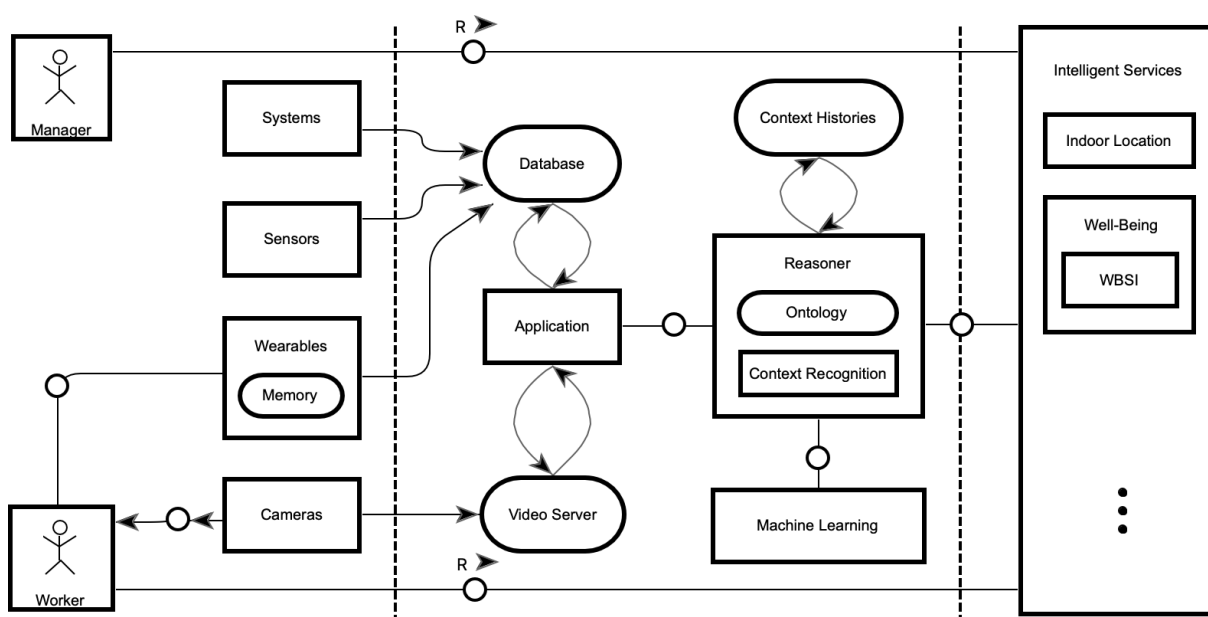
## 4 MODELO ODISSEU

Este capítulo apresenta o modelo ODISSEU (*Ontology Driven Industrial Smart Services for Enterprise Use*), cujo foco é o suporte ao gerenciamento de serviços inteligentes para a Indústria 4.0 usando como base históricos de contextos. Inicialmente é apresentada a arquitetura do modelo, em seguida as estruturas de dados utilizadas e, por fim, a ontologia denominada Odisseia.

### 4.1 Arquitetura

A Figura 17 apresenta a arquitetura do modelo, tendo como base a notação TAM (*Technical Architecture Modeling*), proposta pela empresa SAP. A arquitetura está organizada em diagramas de blocos para representar os agentes, repositórios e conexões envolvidos no processo (GROENE, 2021).

Figura 17 – Arquitetura do modelo Odisseu.



Fonte: Elaborado pelo autor.

O objetivo do modelo é fornecer suporte ao gerenciamento de serviços inteligentes a partir de dados gerados por sensores, imagens, dispositivos móveis ou vestíveis, e sistemas já utilizados na indústria em questão, usando o conceito de históricos de contextos (RENTZ; HECKLER; BARBOSA, 2023). A seguir são descritos os componentes e suas funcionalidades:

- **Worker:** representa os trabalhadores de uma indústria capazes de gerar dados por diversos meios, tais como dispositivos vestíveis e câmeras, além de poder interagir com os serviços inteligentes suportados pelo modelo.

- *Systems*: sistemas já existentes na indústria que podem fornecer dados relevantes para o modelo, por exemplo, sistemas ERP.
- *Sensors*: sensores de ambiente instalados na indústria capazes de gerar dados numéricos para coletas automáticas, como umidade e temperatura do ar.
- *Wearables*: dispositivos vestíveis com a capacidade de armazenar os dados localmente (repositório *Memory*) para posterior sincronização com o banco de dados, por exemplo, pulseiras que medem frequência cardíaca e atividade eletrodérmica.
  - *Memory*: repositório capaz de armazenar os dados coletados pelos sensores do dispositivo mesmo sem conexão com uma rede local ou a Internet.
- *Cameras*: câmeras capazes de detectar movimentação de pessoas na indústria e armazenar as imagens em um servidor de vídeos.
- *Database*: repositório que recebe dados dos agentes *Systems*, *Sensors* e *Wearables*, mantendo comunicação com a camada de aplicação.
- *Video Server*: repositório de vídeos que recebe as imagens do agente *Cameras*, mantendo comunicação com a camada de aplicação.
- *Application*: camada de aplicação responsável por organizar e gerenciar os dados armazenados no repositório *Database* e as imagens armazenadas no repositório *Video Server*, mantendo um canal de comunicação com a ontologia.
- *Reasoner*: agente responsável pelo reconhecimento e organização de informações de contexto (repositório interno *Ontology* e agente *Context Recognition*), mantendo comunicação com o repositório de históricos de contextos, bem como estabelecendo canais de comunicação com a camada de aplicação, a camada de aprendizado de máquina e os serviços inteligentes.
  - *Ontology*: repositório interno que representa a ontologia responsável por armazenar a própria estrutura, além das configurações adicionais dos agentes conectados, como parâmetros dos algoritmos de aprendizado de máquina.
  - *Context Recognition*: agente responsável pela identificação e organização de informações de contexto para o armazenamento de seus históricos.
- *Context Histories*: repositório de históricos de contextos em formato de séries temporais, que além de armazenar os dados de forma estruturada, fornecem dados já tratados para a ontologia.

- *Machine Learning*: agente responsável pelos algoritmos de aprendizado de máquina utilizados por alguns dos serviços inteligentes suportados pelo modelo, mantendo um canal de comunicação com o agente *Reasoner*. Exemplos de algoritmos são Redes Neurais Artificiais (MLP) e Árvores de Decisão (RF).
- *Manager*: gerente ou pessoa responsável pela indústria que faz requisições aos serviços inteligentes disponíveis no modelo.
- *Intelligent Services*: agentes responsáveis pelos serviços inteligentes que venham a ser criados com base em históricos de contextos.
  - *Indoor Location*: agente responsável por indicar a localização interna dos indivíduos (*Indoor Location*).
  - *Well-Being*: agente responsável por estimar o nível do bem-estar subjetivo dos indivíduos, composto pelo Índice de Bem-Estar e Satisfação (*Well-Being and Satisfaction Index*), representado pelo agente *WBSI* que é descrito no Capítulo 5.

Os repositórios *Database* e *Context Histories* estão organizados em estruturas de dados interligadas, cuja principal função é otimizar a consulta por parte dos serviços inteligentes. A Seção 4.2 mostra a organização das estruturas de dados que envolvem os dois repositórios em forma de um diagrama de banco de dados relacional.

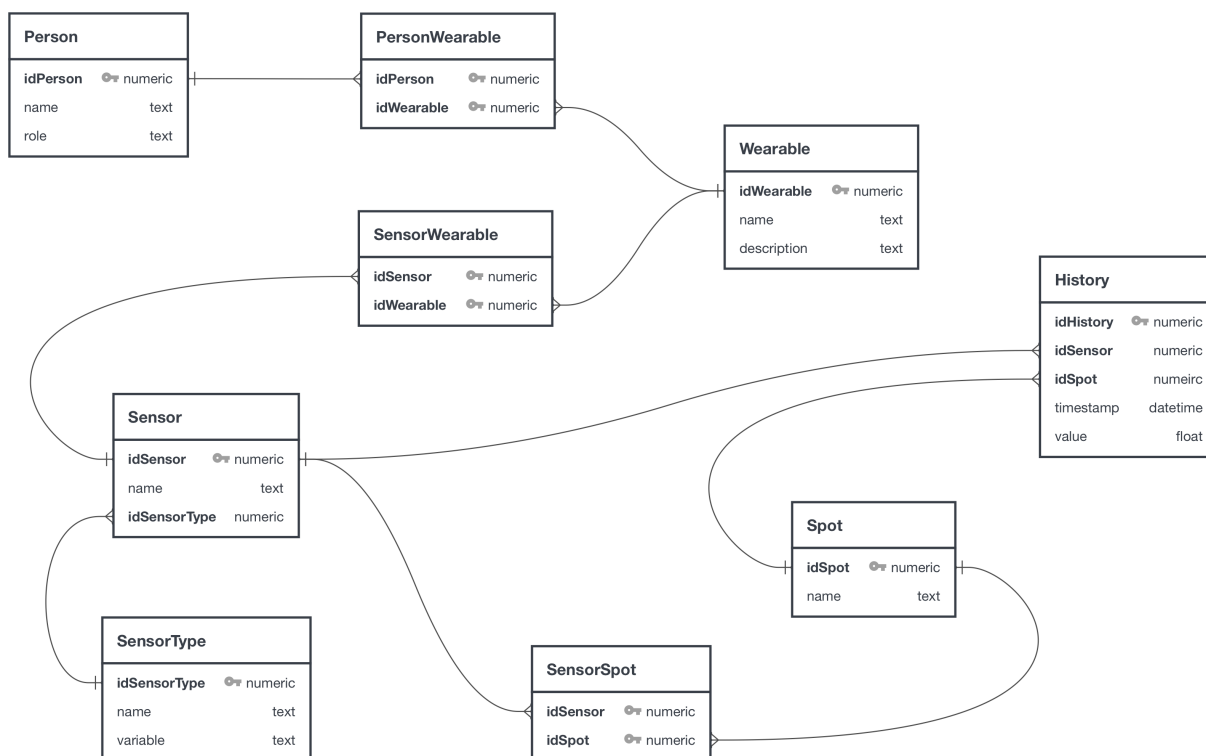
O ponto central do modelo é o agente *Reasoner*, responsável pela ontologia, que atua como elo entre os dados gerados pelos sensores e demais entradas de dados (armazenados nas estruturas de dados) e os serviços inteligentes, sendo o *Indoor Location* e o *Well-Being* os serviços implementados neste trabalho. A Seção 4.3 apresenta a ontologia em detalhes e explica o seu funcionamento.

## 4.2 Estruturas de Dados

Esta seção explica as estruturas de dados que representam os repositórios *Database* e *Context Histories*. Os demais repositórios presentes no modelo são o *Video Server*, responsável por armazenar arquivos gerados por câmeras de vídeo, o repositório *Memory*, que representa a memória interna de dispositivos móveis ou vestíveis, e o repositório *Ontology*, que representa a estrutura da ontologia, a ser apresentada e descrita na próxima seção.

A Figura 18 mostra um diagrama de banco de dados com nove tabelas. Por ordem de descrição, as oito primeiras são referentes ao repositório *Database* e a última se refere ao repositório *Context Histories*. A seguir são descritas cada uma das tabelas e suas funcionalidades:

Figura 18 – Diagrama do banco de dados dos repositórios do modelo Odisseu.



Fonte: Elaborado pelo autor.

- *Person*: responsável por armazenar informações das pessoas que interagem com o modelo, como nome (*name*) e função (*role*).
- *SensorType*: armazena os tipos de sensores instalados em uma determinada indústria (*name*) e a variável medida pelo sensor (*variable*).
- *Sensor*: representa cada um dos sensores presentes em uma indústria (*name*) e os seus respectivos tipos (*idSensorType*).
- *Spot*: organiza os locais ou salas de uma indústria (*name*) onde podem estar presentes um ou mais sensores.
- *SensorSpot*: agrupa os sensores (*idSensor*) instalados em uma determinada sala ou local (*idSpot*).
- *Wearables*: identifica equipamentos vestíveis (*name*) e descreve suas funcionalidades (*description*), podendo agregar sensores e ser usado por pessoas.
- *PersonWearable*: agrupa uma determinada pessoa (*idPerson*) e os equipamentos vestíveis que esteja usando (*idWearable*).
- *SensorWearable*: responsável por agrupar os sensores (*idSensor*) presentes em um determinado equipamento vestível (*idWearable*).



- *History*: armazena cada histórico de contextos gerado por um determinado sensor (*idSensor*) em um local específico (*idSpot*) com o momento de geração (*timestamp*) e o valor do dado (*value*).

As tabelas *Person*, *Sensor*, *SensorType*, *Wearable*, *Spot* e *History* possuem identificadores únicos (chaves primárias), representados pelo prefixo *id* e o nome da tabela. As tabelas *PersonWearable*, *SensorWearable* e *SensorSpot* representam agrupamentos entre outras tabelas, usando seus identificadores para compor a chave.

Os campos da tabela *History* representam a ideia de contexto como proposto por Dey, Abowd e Salber (2001), onde uma informação de contexto é composta pelo tempo (*timestamp*), local (*idSpot*), identidade (*idSensor*) e status (*value*). O histórico de contextos armazena várias informações de contexto que ocorreram em sequência.

### 4.3 Ontologia Odisseia

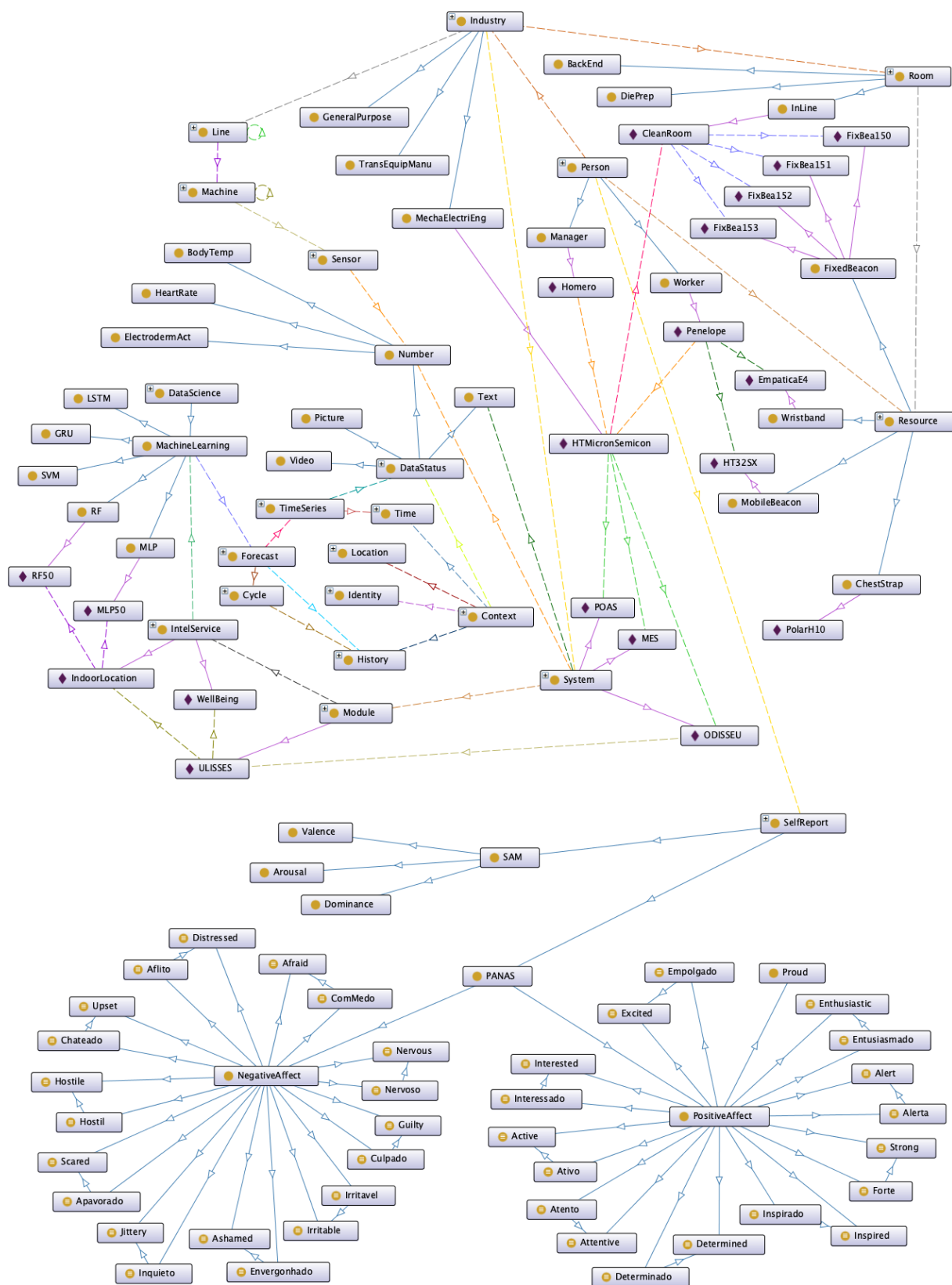
Esta seção descreve a ontologia utilizada no modelo Odisseu, a qual recebeu o nome de Odisseia. A ontologia tem como principal objetivo agir como elo entre a camada de aplicação e os serviços inteligentes, reconhecendo contextos e consultando dados de seus históricos, bem como interagindo com algoritmos de aprendizado de máquina. A Figura 19 apresenta a ontologia com suas classes, instâncias e relações.

A biblioteca *Owlready* foi utilizada para a criação das classes, relações e indivíduos da ontologia, usando o conceito de Programação Orientada a Ontologias. Esse conceito possibilita criar classes e relações como se fossem classes da Programação Orientada a Objetos, e indivíduos como se fossem objetos de uma classe, a partir de instâncias de classes. Um script desenvolvido na linguagem de programação *Python* armazena a ontologia em um arquivo no formato *OWL*, que pode ser editado por outras ferramentas (LAMY, 2017, 2021).

A ferramenta *Protégé* foi escolhida para a criação do gráfico da ontologia, pois oferece a possibilidade de visualização de todos os componentes, indicando classes com um círculo amarelo e indivíduos com um losango roxo, posicionados à esquerda do nome do componente. Apesar desta tese ter usado o *Protégé* somente para visualização, a ferramenta também permite a edição da ontologia (MUSEN, 2015).

A classe *Industry* é responsável por classificar a indústria em questão, com subclasses representando os segmentos industriais propostos pela Organização Internacional do Trabalho, uma agência das Nações Unidas, que classificou os segmentos industriais em 22 tipos (Tabela 22 do Apêndice). A figura mostra os três segmentos com maior número de ocorrências na revisão sistemática. A subclasse *GeneralPurpose* representa o segmento *Propósito geral / outros*, a subclasse *TransEquipManu* se refere ao segmento *Fabricação de equipamentos de transporte* e a subclasse *MechaElectriEng*, usada como estudo de caso, indica o segmento *Engenharia mecânica e*

Figura 19 – Ontologia do modelo Odisseu.



Fonte: Elaborado pelo autor.

*elétrica* e possui o indivíduo *HTMicronSemicon*, representando a HT Micron Semicondutores, onde foi conduzido o experimento de coleta de dados.

O ontologia Odisseia estendeu a ontologia *OntoLearnWell* (ARRUDA et al., 2023b), baseada em contextos e orientada ao bem-estar emocional em um ambiente educacional. Foram utilizadas as classes *SelfReport*, *Person*, *Resource*, *Wristband*, *Context*, *History*, *Room*, *FixedBeacon*, *MobileBeacon*, *Status* (renomeada para *DataStatus*), *BodyTemp*, *HeartRate*, *ElectrodermAct*, *Time*, *Location* e *Identity*. As classes foram adaptadas ao ambiente industrial e serão descritas individualmente de acordo com suas funcionalidades na ontologia.

As classes *Line*, *Machine* e *Sensor* fazem parte da ontologia *OntoLSTM* (HUANG; ZANNI-MERK; CRÉMILLEUX, 2019), estendida pela ontologia proposta. *Line* se refere a uma linha de montagem, *Machine* representa as máquinas que compõem essa linha e *Sensor* indica os sensores que monitoram as máquinas. Portanto, a ontologia indica que uma determinada indústria possui uma linha de montagem com máquinas monitoradas por sensores.

A indústria em questão também possui sistemas, representados pela classe *System*, que apresenta como exemplos as instâncias *POAS*, *MES* e *ODISSEU*. A instância *POAS* se refere aos sistemas de captura de dados em tempo real, *MES* representa os sistemas de chão de fábrica e *ODISSEU* representa o modelo proposto. Assim como os sensores, os sistemas também produzem dados que serão utilizados pelos serviços inteligentes.

A classe *DataStatus* representa os tipos de dados tratados pela ontologia e possui quatro subclasses: *Number*, *Text*, *Video* e *Picture*. A instância *Number* se refere aos valores numéricos gerados tanto pelos sensores como pelos sistemas, *Text* corresponde aos dados alfanuméricos originados pelos sistemas, *Video* representa as imagens produzidas por câmeras de vídeo e *Picture* indica imagens estáticas.

Em conjunto com as classes *Time*, *Location* e *Identity*, a classe *DataStatus* compõe uma informação de contexto, representada pela classe *Context*. Conforme mencionado nos capítulos anteriores, informações de contexto são formadas por uma identificação única (*Identity*), o local em que a informação foi coletada (*Location*), o momento na linha do tempo em que ocorreu o evento (*Time*) e o valor do dado propriamente dito, também referenciado como status da informação (*DataStatus*).

A classe *History* representa os dados históricos das informações de contexto produzidas pela classe *Context*. Os históricos de contextos são necessários para que os serviços inteligentes, representados pela classe *IntelService*, possam funcionar. O índice de bem-estar que será apresentado no próximo capítulo (*WBSI*), faz parte do serviço inteligente de bem-estar subjetivo, aqui representado pela instância *WellBeing*. A instância *IndoorLocation* é referente ao serviço de localização de funcionários no interior da fábrica. A classe *Module* representa um módulo de software e possui a

instância *ULISSES*, referente ao protótipo evolutivo que usa os serviços inteligentes *IndoorLocation* e *WellBeing*.

Os algoritmos de aprendizado de máquina fazem parte dos serviços inteligentes, representados pela classe *MachineLearning*, que é uma subclasse de *DataScience*. Cinco subclasses exemplificam os algoritmos usados pelos serviços inteligentes: *LSTM*, *GRU*, *SVM*, *RF* e *MLP*. Todos os algoritmos citados constam na revisão sistemática apresentada no Capítulo 3 e estão na Tabela 20 do Apêndice B. Os indivíduos *RF50* e *MLP50* representam instâncias dos algoritmos de aprendizado de máquina utilizados pelo serviço *IndoorLocation*.

A classe *Forecast* representa a capacidade dos algoritmos de aprendizado de máquina de realizar predição de valores futuros baseado em dados do passado. Para isso, as predições podem utilizar partes de um histórico de contextos, representadas pela classe *Cycle*, que utiliza a quantidade mínima de dados capaz de fazer a predição. Essa quantidade mínima de dados pode variar de acordo com os tipos de dados utilizados e deve ser obtida por meio de testes com diferentes tamanhos de janelas.

É importante ressaltar que apesar de funcionar em conjunto com os serviços inteligentes usando históricos de contextos, os algoritmos de aprendizado de máquina também são capazes de realizar predições usando séries temporais, representadas pela classe *TimeSeries*. Séries temporais são compostas por valores de dados (*DataStatus*) ordenados na linha do tempo (*Time*).

A Tabela 11 mostra as relações entre as classes da ontologia, que são utilizadas em consultas escritas na linguagem *SPARQL* para retornar conjuntos de dados armazenados como indivíduos da ontologia.

A classe *SelfReport* representa os dois tipos de autorrelato suportados pela ontologia. O primeiro tipo é representado pela classe *SAM*, que se refere a *Self Assessment Manikin* e leva em consideração as dimensões valência (prazer-desprazer), representada pela classe *Valence*, excitação (agitação-relaxamento), classe *Arousal*, e dominância (submissão-controle), classe *Dominance*. A técnica parte do uso de figuras (manequins), ou seja, é uma escala pictórica simples, que utiliza cinco imagens para representar cada uma das três dimensões (BRADLEY; LANG, 1994).

O segundo tipo de autorrelato suportado pela ontologia é a Escala de Afetos Positivos e Negativos (*Positive and Negative Affect Schedule*), representada pela classe *PANAS*, que foi utilizada no experimento com o objetivo de estimar o bem-estar subjetivo dos participantes. A classe *PositiveAffect* representa os afetos positivos e a classe *NegativeAffect*, os afetos negativos (WATSON; ANNA; TELLEGEN, 1988). A escala foi validada para português e a Tabela 18 do Capítulo 5 apresenta os afetos equivalentes representados como classes da ontologia (CARVALHO et al., 2013).

As classes *ElectrodermAct*, *HeartRate* e *BodyTemp* representam os sinais vitais atividade eletrodérmica, frequência cardíaca e temperatura da pele, respectivamente.

Tabela 11 – Relações da ontologia Odisseia.

Classe	Relação	Classe	Classe	Relação	Classe
Context	hasDataStatus	DataStatus	Machine	precedesMachine	Machine
Context	produceHistory	History	Machine	hasSensor	Sensor
Context	hasIdentity	Identity	MachineLearning	makesForecast	Forecast
Context	hasLocation	Location	Module	usesIntelService	IntelService
Context	hasTime	Time	Person	attendsIndustry	Industry
Cycle	composesHistory	History	Person	usesResource	Resource
Forecast	usesCycle	Cycle	Person	answersSelfReport	SelfReport
Forecast	usesHistory	History	Room	hasResource	Resource
Forecast	needsTimeSeries	TimeSeries	Sensor	producesNumber	Number
Industry	hasLine	Line	System	hasModule	Module
Industry	hasRoom	Room	System	producesNumber	Number
Industry	hasSystem	System	System	producesText	Text
IntelService	usesMachineLearning	MachineLearning	TimeSeries	needsDataStatus	DataStatus
Line	precedesLine	Line	TimeSeries	needsTime	Time
Line	hasMachine	Machine			

Fonte: Elaborado pelo autor.

São subclasses da classe *Number* pois os valores das medidas dos sensores são registrados em formato numérico.

A classe *Person* representa as pessoas que trabalham na indústria e possui as subclasses *Manager* (Gerente) e *Worker* (Funcionário). O gerente é representado pelo indivíduo *Homero* e o funcionário representado pelo indivíduo *Penelope*. Os indivíduos *Homero* e *Penelope* são codinomes utilizados para manter a privacidade dos participantes do experimento (TESSON, 2019).

A classe *Room* indica os tipos de salas presentes na indústria, representadas pelas subclasses *BackEnd*, *DiePrep* e *InLine*, que possui o indivíduo *CleanRoom*. No experimento, o indivíduo *CleanRoom* representa a sala utilizada para a coleta de dados relacionada com o indivíduo *HTMicronSemicon*, que por sua vez representa a própria HT Micron Semicondutores como um indivíduo da classe *MechaElectriEng*.

A classe *Resource* representa os recursos utilizados na indústria, subdividindo-se em beacons fixos e móveis (subclasses *FixedBeacon* e *MobileBeacon*), pulseiras (subclasse *Wristband*) e cintas peitorais (subclasse *ChestStrap*). O indivíduo *PolarH10* é um exemplo de cinta peitoral com sensor de frequência cardíaca que foi cogitado para o experimento, mas que não chegou a ser utilizado em função das pulseiras serem dispositivos mais práticos para utilização (SCHAFFARCZYK et al., 2022). O indivíduo *EmpaticaE4* representa a pulseira utilizada no experimento, responsável por coletar os sinais vitais dos participantes, como atividade eletrodérmica,

frequência cardíaca e temperatura da pele (EMPATICA, 2021). O indivíduo *HT32SX* representa um dos beacon móveis utilizados no experimento e melhor detalhado no próximo capítulo (HTMICRON, 2022). O indivíduo *Penelope* usa os recursos *EmpaticaE4* e *HT32SX*. Os indivíduos *FixBea150*, *FixBea151*, *FixBea152* e *FixBea153* representam os beacons fixos instalados na sala limpa (indivíduo *CleanRoom*).

A Tabela 12 apresenta as métricas da ontologia geradas pelo software *Protégé*, incluindo o número de axiomas, classes, propriedades e indivíduos. São mostrados os valores iniciais imediatamente após a criação da ontologia e os valores com os dados dos históricos de contextos, importados posteriormente e detalhados no próximo capítulo juntamente com os outros dados utilizados no experimento.

Tabela 12 – Métricas da ontologia Odisseia.

<b>Métrica</b>	<b>Valor Inicial</b>	<b>Valor com Históricos</b>
Axiomas	385	1015113
Axiomas lógicos	245	870009
Axiomas de declaração	140	145104
Classes	93	93
Propriedades de objetos	28	28
Propriedades de dados	1	6
Indivíduos	19	144978

Fonte: Elaborado pelo autor.

## 5 AVALIAÇÃO E RESULTADOS

Esta tese propôs um modelo computacional para serviços inteligentes voltados para a Indústria 4.0, com foco na localização de funcionários nas instalações de uma indústria de semicondutores e no seu bem-estar, utilizando históricos de contextos. A questão de pesquisa foi definida na Seção 1.1: “Como seria um modelo computacional para gerenciamento de serviços inteligentes em ambientes industriais baseado na análise de históricos de contextos?”

Neste sentido, a proposta para avaliação do modelo consistiu na coleta e tratamento dos dados obtidos, além do treinamento e testes dos algoritmos antes da execução dos serviços inteligentes implementados em um protótipo funcional. Na análise de resultados, os históricos de contextos são consultados para indicar a posição de uma pessoa na sala limpa de uma indústria e para prever seu bem-estar subjetivo por meio de afetos positivo e negativos, usando um índice proposto.

O modelo Odisseu propõe a utilização de dados coletados de sensores e de formulários de autorrelato armazenados em formato de séries temporais de contexto para uso em ambientes industriais. Os dados são organizados pela ontologia intitulada Odisseia, responsável por integrar as diferentes fontes de dados até a execução dos serviços inteligentes por meio de um módulo de software de um protótipo evolutivo denominado Ulisses.

Dessa forma, a avaliação buscou responder às seguintes questões específicas de pesquisa, definidas na Seção 1.1: a) “Como deve ser um modelo computacional para suportar de forma genérica serviços inteligentes baseados em históricos de contextos gerados por diferentes fontes de dados em ambientes industriais?” b) “De que forma seria aplicada uma ontologia como base para um modelo computacional baseado em históricos de contextos voltado para a Indústria 4.0?” c) “Como seria um serviço inteligente voltado para a localização de funcionários em ambientes industriais *indoor* com base em históricos de contextos?” d) “É possível criar um serviço inteligente baseado em históricos de contextos para acompanhamento do bem-estar, usando um índice que mede o bem-estar subjetivo no ambiente de trabalho por meio de questões de autorrelato e dos sinais vitais dos funcionários de uma indústria?”

O desenvolvimento de dois serviços inteligentes respondeu a primeira questão de pesquisa, onde diversas fontes de dados foram utilizadas para armazenar históricos de contextos utilizados pelos serviços. A utilização de uma ontologia como elemento de ligação entre os dados dos históricos de contextos e os serviços inteligentes, respondeu a segunda questão de pesquisa. A terceira e quarta questões de pesquisa foram respondidas após a análise dos dados de localização interna e da estimativa do bem-estar subjetivo dos voluntários da HT Micron Semicondutores S.A. durante a coleta de dados do experimento.

A avaliação ocorreu em três etapas. A primeira envolveu a coleta de dados de diferentes tipos de sensores (Seção 5.1). A segunda ocorreu com a coleta de dados de autorrelato (Seção 5.2). A terceira consistiu na implementação dos serviços inteligentes propostos e na integração dos serviços em um protótipo evolutivo (Seção 5.3).

Para a realização da coleta de dados, um projeto foi submetido ao Comitê de Ética e Pesquisa da Universidade do Vale do Rio dos Sinos e aprovado com o Certificado de Apresentação para Apreciação Ética (CAAE) n. 70918323.7.0000.5344. Os voluntários que participaram da coleta assinaram um Termo de Consentimento Livre e Esclarecido (TCLE) disponível no Apêndice E.

## 5.1 Coleta de Dados de Sensores

Os sensores utilizados no experimento coletaram dados de localização dos funcionários da fábrica por meio de beacons fixos e móveis, seus sinais vitais através de uma pulseira, além de informações do ambiente geradas por sensores previamente instalados na sala limpa onde ocorreu o experimento.

Foram coletados dados de localização de cinco funcionários da HT Micron Semicondutores do time de produção, convidados pela empresa e que participaram por adesão, que corresponde ao número de beacons móveis disponíveis para o experimento. Os voluntários tiveram os dados coletados durante cinco dias durante um turno de trabalho de oito horas.

Para a coleta dos dados correspondentes aos sinais vitais, ocorreu um revezamento da pulseira utilizada em função de somente haver um equipamento disponível. Cada funcionário utilizou a pulseira durante um turno de trabalho.

Os dados dos sensores de ambiente foram fornecidos pela HT Micron como parte do projeto *Intelligent Factory*, que possibilitou a condução do experimento na sala limpa da fábrica (SILVEIRA et al., 2020).

### 5.1.1 Localização

Antes da coleta dos dados de localização no ambiente de produção foram definidos dez pontos de interesse na sala limpa, que correspondem a locais estratégicos para identificação do posicionamento dos funcionários. Além disso, foram definidos quatro pontos para a instalação dos beacons fixos, responsáveis por medir a intensidade do sinal dos beacons móveis.

A partir da definição dos pontos de interesse e dos locais de instalação dos beacons fixos, foi realizada uma coleta inicial para o treinamento dos algoritmos de aprendizado de máquina utilizados pelos serviços inteligentes. O processo ocorreu com o posicionamento de um beacon móvel no local de cada um dos pontos de interesse.



Em seguida, foram criados vetores de atributos com as combinações dos valores de cada beacon fixo em relação aos pontos de interesse. A Tabela 13 mostra o número de leituras da intensidade do sinal RSSI para cada ponto de interesse, totalizando 1066 registros, o número de leituras após a exclusão das leituras com valores repetidos, totalizando 190 registros únicos, e o número de vetores de atributos gerados, totalizando 39841 registros.

Tabela 13 – Número de leituras do sinal RSSI de cada ponto de interesse.

Ponto de interesse	Registros	Registros únicos	Vetores de atributos
0	144	22	1440
1	84	15	144
2	150	36	33660
3	106	17	448
4	112	10	300
5	100	23	1080
6	94	21	1800
7	106	16	240
8	89	12	81
9	81	18	648

Fonte: Elaborado pelo autor.

Durante o experimento foram coletados 36823 registros de RSSI dos quatro beacons fixos, sendo que 18 estavam corrompidos, restando então 36805 registros. A Tabela 14 mostra o total de registros de cada voluntário ao longo dos cinco dias de coleta e o número de registros diários.

Tabela 14 – Número de leituras do sinal RSSI dos voluntários durante a coleta de dados.

Voluntário	Registros	Dia 1	Dia 2	Dia 3	Dia 4	Dia 5
1	8365	1622	1644	1815	1614	1670
2	4628	0	1265	1686	0	1677
3	7012	1611	1829	1823	1742	7
4	8395	1657	1509	1702	1821	1706
5	8405	1572	1697	1718	1748	1670

Fonte: Elaborado pelo autor.

Os beacons fixos e móveis foram desenvolvidos pela HT Micron utilizando o iMCP HT32SX, um Circuito Integrado Multicomponentes voltado para aplicações com foco em Internet das Coisas (HTMICRON, 2022).

### 5.1.2 Sinais Vitais

A pulseira Empatica E4 foi utilizada para a coleta dos sinais vitais dos funcionários da HT Micron Semicondutores. O equipamento já foi usado em outros estudos para detecção de estresse em crianças (COŞKUN et al., 2023), para medir interações diádicas (MILSTEIN; GORDON, 2020), para detecção precoce de ataques de enxaqueca (KOSKIMÄKI et al., 2017), entre outros tipos de pesquisa que envolvem dados provenientes de sinais vitais humanos.

O dispositivo conta com sensores que coletam os sinais vitais e armazenam os dados já em formato de séries temporais no padrão CSV, onde cada registro de informação é separado por vírgulas, o que facilita o processo de análise dos dados. Além da coleta de dados de sensores, a pulseira tem um botão que pode registrar um evento ocorrido na linha do tempo (EMPATICA, 2021). A Tabela 15 mostra cada um dos arquivos gerados e a frequência de coleta, quando se aplicar.

Tabela 15 – Arquivos gerados pela Empatica E4.

Nome do arquivo	Descrição	Frequência (Hz)
ACC.CSV	Acelerômetro de 3 eixos	32
BVP.CSV	Fotopletismógrafo (pulso de volume sanguíneo)	64
EDA.CSV	Atividade eletrodérmica	4
HR.CSV	Frequência cardíaca média (do BVP)	1
IBI.CSV	Tempo entre batimentos cardíacos (do BVP)	N/A
TAGS.CSV	Eventos do botão	N/A
TEMP.CSV	Temperatura da pele	4

Fonte: Elaborado pelo autor.

Cada voluntário usou a pulseira uma vez durante o turno diário de trabalho. A Tabela 16 mostra o número de registros coletados por dia, o número de registros de cada arquivo CSV e o número de registros agrupados pela menor frequência.

Tabela 16 – Registros coletados por dia com a Empatica E4.

Dia	Registros	ACC	BVP	EDA	HR	IBI	TEMP	Agrupados
1	2988488	910260	1820445	113772	28435	1800	113776	28435
2	3144908	956136	1912251	119496	29869	7644	119512	29869
3	2964956	902094	1804110	112734	28180	5086	112752	28180
4	3145264	957990	1915947	119736	29927	1920	119744	29927
5	3002839	914262	1828343	114294	28548	3120	114272	28548

Fonte: Elaborado pelo autor.

### 5.1.3 Ambiente

Sensores de ambiente instalados na indústria são capazes de gerar dados numéricos para coletas automáticas, como concentração de partículas, umidade relativa e temperatura do ar. Esses dados foram fornecidos pela HT Micron Semicondutores e fazem parte do projeto *Intelligent Factory*, conduzido em suas instalações. A concentração de partículas é medida em pés cúbicos (particles/ft<sup>3</sup>), a umidade relativa em percentual (%RH) e a temperatura em graus Celsius (°C) (SILVEIRA et al., 2020).

Os dados são armazenados pela indústria em um banco de dados específico para séries temporais chamado *InfluxDB* (NAQVI; YFANTIDOU; ZIMÁNYI, 2017), porém foram exportados como arquivos CSV para uso neste trabalho. A Tabela 17 mostra o total diário de quatro sensores de umidade, quatro sensores de temperatura e quatro sensores de partículas, instalados na sala do experimento. Os sensores de umidade e temperatura coletam um registro por minuto (0,0167 Hz) e os sensores de partículas coletam um registro a cada quinze minutos (0,0011 Hz).

Tabela 17 – Registros de sensores de ambiente coletados por dia.

Dia	Umidade	Temperatura	Partículas
1	2160	2160	280
2	2160	2160	384
3	2160	2160	384
4	2160	2160	384
5	2156	2156	248

Fonte: Elaborado pelo autor.

## 5.2 Coleta de Dados de Autorrelato

Com o objetivo de medir o bem-estar subjetivo dos participantes, foi utilizada a escala PANAS, que analisa os afetos positivos e negativos por meio de vinte tipos de afetos, sendo dez positivos e dez negativos (WATSON; ANNA; TELLEGEN, 1988).

A escala PANAS foi inicialmente concebida na língua inglesa e posteriormente validada para o português, sendo que o tipo de afeto *proud* foi desconsiderado no processo de validação por apresentar dupla conotação. A Tabela 18 apresenta os tipos de afetos da escala e suas versões original e traduzida (CARVALHO et al., 2013).

Para a aplicação da escala PANAS foi utilizado um formulário *online* onde cada um dos afetos foi indicado pelo funcionário ao final do expediente de trabalho usando uma escala Likert, que utiliza valores entre 1, para menor intensidade do afeto, e 5, para a maior intensidade (PIRES et al., 2013). Ao todo foram coletados 24 registros

Tabela 18 – Escala de Afetos Positivos e Negativos (PANAS).

<b>Tipo de Afeto</b>	<b>Original em Inglês</b>	<b>Traduzido para Português</b>
Positivo	<i>Active</i>	Ativo
Positivo	<i>Alert</i>	Alerta
Positivo	<i>Attentive</i>	Atento
Positivo	<i>Determined</i>	Determinado
Positivo	<i>Enthusiastic</i>	Entusiasmado
Positivo	<i>Excited</i>	Empolgado
Positivo	<i>Inspired</i>	Inspirado
Positivo	<i>Interested</i>	Interessado
Positivo	<i>Proud</i>	N/A
Positivo	<i>Strong</i>	Forte
Negativo	<i>Afraid</i>	Com medo
Negativo	<i>Ashamed</i>	Envergonhado
Negativo	<i>Distressed</i>	Aflito
Negativo	<i>Guilty</i>	Culpado
Negativo	<i>Hostile</i>	Hostil
Negativo	<i>Irritable</i>	Irritável
Negativo	<i>Jittery</i>	Inquieto
Negativo	<i>Nervous</i>	Nervoso
Negativo	<i>Scared</i>	Apavorado
Negativo	<i>Upset</i>	Chateado

Fonte: Elaborado pelo autor.

de autorrelato, pois um dos voluntários precisou se ausentar durante o turno. Os itens da escala são: “Nem um pouco”, “Um pouco”, “Moderadamente”, “Bastante” e “Extremamente”. O Apêndice F mostra o formulário completo de aplicação da escala.

### 5.3 Serviços Inteligentes

Esta seção tem como objetivo apresentar os serviços inteligentes *Indoor Location*, responsável por indicar a posição de um funcionário no interior da fábrica, e *Well-Being*, que é responsável pela mensuração do bem-estar subjetivo no ambiente de trabalho e composto pelo *WBSI* (Índice de Bem-Estar e Satisfação ou *Well-Being and Satisfaction Index*). Também é apresentado o protótipo Ulisses, que integra os serviços inteligentes.

### 5.3.1 Indoor Location

O experimento que validou o serviço inteligente de localização interna foi conduzido na Sala Limpa da HT Micron Semicondutores. Foram instalados quatro beacons fixos e escolhidos dez pontos de interesse para o treinamento dos algoritmos de aprendizado de máquina. O objetivo dos pontos de interesse é apontar onde é mais importante localizar alguém para o responsável pela indústria. A Figura 20 mostra os beacons fixos como retângulos azuis e os pontos de interesse como círculos verdes.

Figura 20 – Planta baixa da sala limpa da HT Micron Semicondutores, onde foram instalados os beacons fixos (retângulos azuis) e definidos os pontos de interesse (círculos verdes).



Fonte: Elaborado pelo autor.

Em cada ponto de interesse foi utilizado um beacon móvel que recebeu sinais RSSI dos beacons fixos instalados no ambiente. Devido ao grande número de equipamentos de rádio que utilizam a frequência 2.4 GHz, optou-se por usar uma frequência mais baixa, de 915 MHz, para evitar interferência e perda do sinal. Os beacons fixos foram instalados próximos ao teto para evitar obstáculos e ligados a rede elétrica local. Os beacons móveis foram energizados por pilhas e encapsulados em forma de braceletes vestíveis. Diferentes intervalos de captura dos sinais RSSI foram testados a fim de otimizar a precisão da localização e o consumo das pilhas.

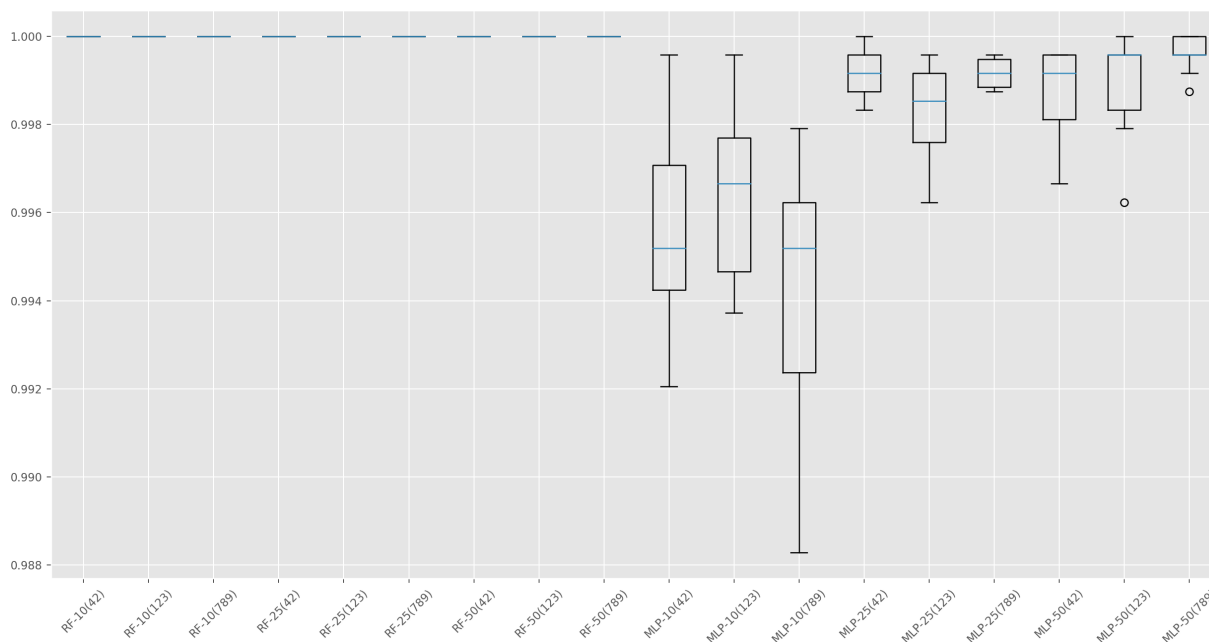
Para o desenvolvimento e implantação do software, que inclui os algoritmos de aprendizado de máquina, foi utilizada a linguagem de programação Python (VAN ROSSUM; DRAKE, 2009) e os conjuntos de bibliotecas de aprendizado de máquina Scikit-learn (PEDREGOSA et al., 2011) e Keras (CHOLLET et al., 2015), amplamente usados pela comunidade científica, visando otimizar o tempo de desenvolvimento e implantação do sistema na indústria.

Os dados foram organizados em formato de vetor de atributos para o treinamento dos algoritmos. No pré-processamento, os dados coletados mais de uma vez com o mesmo valor foram excluídos. O vetor é composto pelo produto cartesiano dos valores de RSSI de cada um dos pontos fixos, ou seja, todas as combinações dos valores dos quatro pontos fixos em relação a um determinado ponto móvel.

Do total de vetores de atributos (39841), 60% (23904) foi usado para treinamento, 20% (7969) para testes e 20% (7968) para validação. Os algoritmos de aprendizado de máquina utilizados foram o *Random Forest* (RF) e um *Bagging* do *Multilayer Perceptron* (MLP). Os algoritmos escolhidos estão entre os métodos de ciência de dados mais constantes na revisão sistemática do Capítulo 3 (RF apareceu 14 vezes e MLP apareceu 8 vezes). O algoritmo RF é uma coleção de árvores de decisão que escolhe a melhor opção entre as classes envolvidas. As árvores de decisão são inicializadas de forma independente e tem um voto único para a melhor classe. O algoritmo MLP é uma rede neural artificial que usa neurônios interconectados entre as camadas, divididas entre camada de entrada, camadas ocultas e camada de saída. O *Bagging* é uma técnica que combina diferentes versões de um modelo com o objetivo de melhorar a precisão, usando subconjuntos dos dados originais de forma aleatória e agregando os resultados individuais na previsão final. Como o RF usa essa técnica de forma nativa, foi incorporada ao MLP para tornar a comparação mais justa.

A Figura 21 mostra os resultados da comparação em formato de dezoito *boxplots*, sendo nove para cada algoritmo com variações de inicialização e configuração. Todas as medianas de RF chegaram a 100% e as de MLP ficaram entre 99,5% e 100%. Tanto RF como MLP apresentaram acurácia de 100% e correlação (MCC) de +1 na validação. Foi realizada validação cruzada com 10 grupos e utilizadas três variações de *seed* para a inicialização dos modelos (42, 123, 789) e três variações para o número de árvores de decisão de RF e para os neurônios da camada oculta de MLP (10, 25, 50). Após o treinamento, os modelos foram validados *in loco* pelo time da HT Micron Semicondutores nas instalações da fábrica.

Figura 21 – Resultados da comparação entre os modelos RF e MLP para o Indoor Location.



Fonte: Elaborado pelo autor.

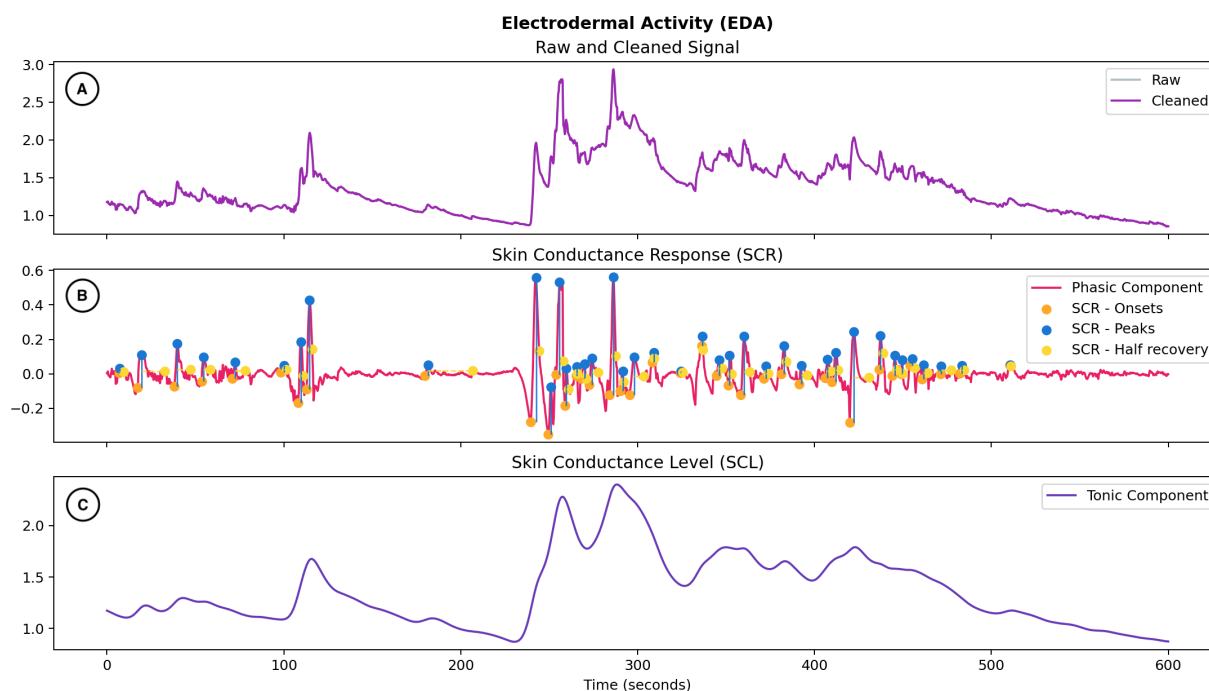
Dessa forma, o algoritmo RF foi escolhido para uso no protótipo evolutivo responsável por indicar a localização, usando dados dos históricos de contextos com os valores do sinal RSSI dos beacons fixos e retornando a posição do beacon móvel.

### 5.3.2 Well-Being (WBSI)

Para propor o Índice de Bem-Estar e Satisfação (*Well-Being and Satisfaction Index*), foram utilizados dados de sensores que medem os sinais vitais e dados de formulários de autorrelato voltados para o bem-estar subjetivo. Os sinais vitais utilizados para a concepção do índice foram a atividade eletrodérmica e o pulso de volume sanguíneo em conjunto com os dados obtidos pela aplicação da escala PANAS. O valor do índice representa a qualidade do bem-estar subjetivo de um determinado dia.

A Figura 22 mostra dados de atividade eletrodérmica (EDA) coletados com a pulseira Empatica E4 e tratados com a biblioteca *NeuroKit2* usando a linguagem *Python*. A biblioteca foi desenvolvida para o processamento de dados neurofisiológicos, como atividade cardíaca e respiratória, além da atividade eletrodérmica. Na parte superior da figura (A) aparece o sinal de EDA bruto e limpo. Na parte do meio (B) estão os dados de resposta de condutância da pele. Na parte inferior inferior (C) está o nível de condutância da pele (MAKOWSKI et al., 2021).

Figura 22 – Amostra de dados da atividade eletrodérmica coletados durante dez minutos.



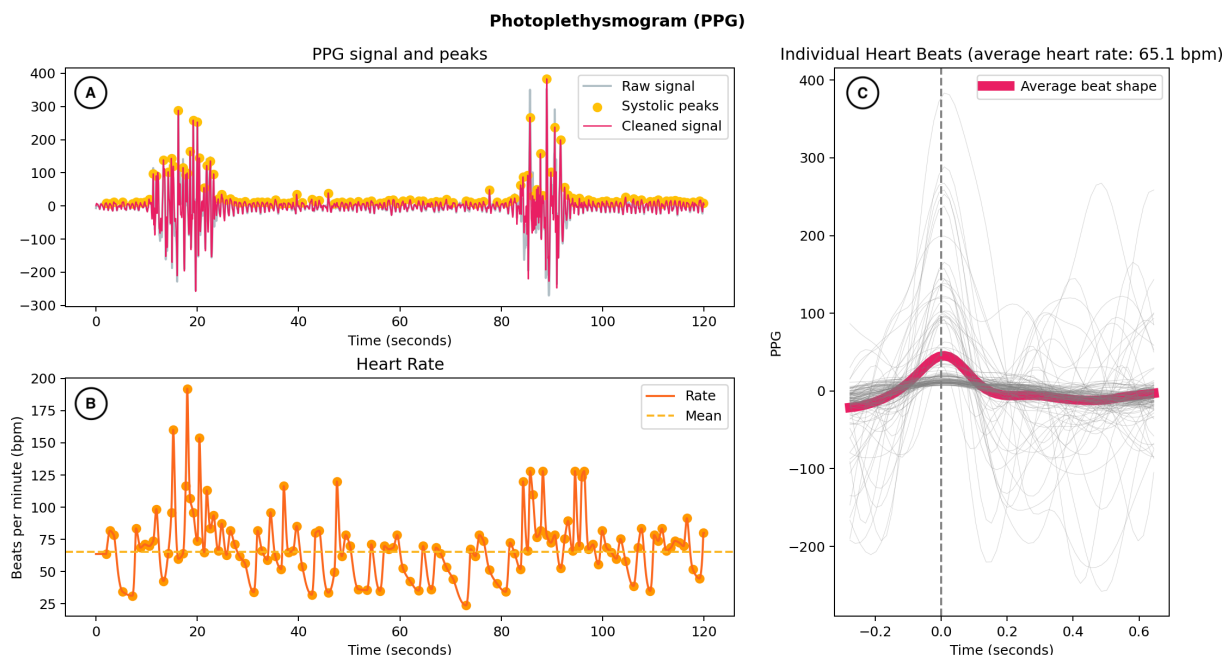
Fonte: Elaborado pelo autor.

A atividade eletrodérmica é um sinal vital importante a ser considerado para análise do bem-estar subjetivo, pois está relacionada com o estresse no ambiente de trabalho. Estresse é um conceito que se refere a aspectos psicológicos e biológicos, podendo ocorrer durante diferentes situações cognitivas e emocionais (KOLDIJK; NEERINCX; KRAAIJ, 2018). Variações em parâmetros fisiológicos, como atividade eletrodérmica, podem ser atribuídas a situações de estresse (CAVALLO et al., 2021).

Além da atividade eletrodérmica, o pulso de volume sanguíneo também é um sinal vital importante para a análise do bem-estar subjetivo. A Figura 23 mostra dados de uma fotopletismografia de um dos voluntários do experimento em um momento aleatório do tempo, que mede o pulso de volume sanguíneo, com o objetivo de identificar variações no sinal vital que indiquem alterações de estresse. A parte superior esquerda da figura (A) mostra o sinal bruto e limpo e os picos de atividade sistólica. A parte inferior esquerda (B) mostra a frequência cardíaca. Na parte direita da figura (C) está a média do formato de batidas do coração (LISOWSKA; WILK; PELEG, 2021).



Figura 23 – Amostra de dados do fotopletismógrafo coletados durante dois minutos.



Fonte: Elaborado pelo autor.

Para compor o índice de bem-estar e satisfação, foram definidas as variáveis *EDA*, correspondente a dados de atividade eletrodérmica, e *BVP*, correspondente a dados de pulso de volume sanguíneo (Equação 5.1).

$$EDA = peaks + maxamp + meanamp + stdamp + meanris + stdris; BVP = peaks \quad (5.1)$$

A variável *EDA* é composta pelas variáveis *peaks* que corresponde ao número de picos de atividade ocorridas em um determinado intervalo de tempo, *maxamp* corresponde a amplitude máxima alcançada pelos picos, *meanamp* é a média da amplitude dos picos, *stdamp* é o desvio padrão da amplitude, *meanris* corresponde a média do tempo de subida dos picos e *stdris* é o desvio padrão da média do tempo de subida dos picos. A variável *BVP* é composta pela variável *peaks* que corresponde ao número de picos sistólicos ocorridos em um determinado intervalo de tempo, utilizada para desambiguação dos picos de EDA e BVP.

Em conjunto com os sinais vitais, os afetos coletados por meio da escala PANAS também compõem o índice. Afetos positivos podem ser definidos como vivências de emoções como alegria, felicidade, contentamento ou excitação. As medidas de afetos positivos avaliam como um determinado indivíduo se sente em um período de tempo específico, como parte de um dia ou durante um expediente de trabalho. Situações de estresse podem ser amortecidas pelos sentimentos de afetos positivos (OKELY; WEISS; GALE, 2017).

A variável *WBSI* representa a fórmula do índice, sendo composta pela diferença dos afetos positivos (*pos*) e negativos (*neg*) dividido por dois, somado com a raiz sétima do somatório da variável *EDA* mais o logaritmo da variável *BVP*. A raiz sétima foi utilizada levando em consideração as seis variáveis de *EDA* mais a variável de *BVP*. O logaritmo foi utilizado para reduzir a escala de *BVP* (Equação 5.2).

$$WBSI = \frac{(pos - neg)}{2} + \sqrt[7]{\sum EDA + \log BVP} \quad (5.2)$$

As variáveis *pos* e *neg* correspondem, respectivamente, aos percentuais de afetos positivos e negativos da escala PANAS. Para o cálculo do percentual é atribuído um valor numérico para cada elemento da escala, sendo 1 para “Nem um pouco”, 2 para “Um pouco”, 3 para “Moderadamente”, 4 para “Bastante” e 5 para “Extremamente”. Assim, o máximo de pontos a serem atribuídos para afetos positivos é 45, devido à exclusão do afeto *proud* na validação brasileira da escala, e 50 para afetos negativos (CARVALHO et al., 2013).

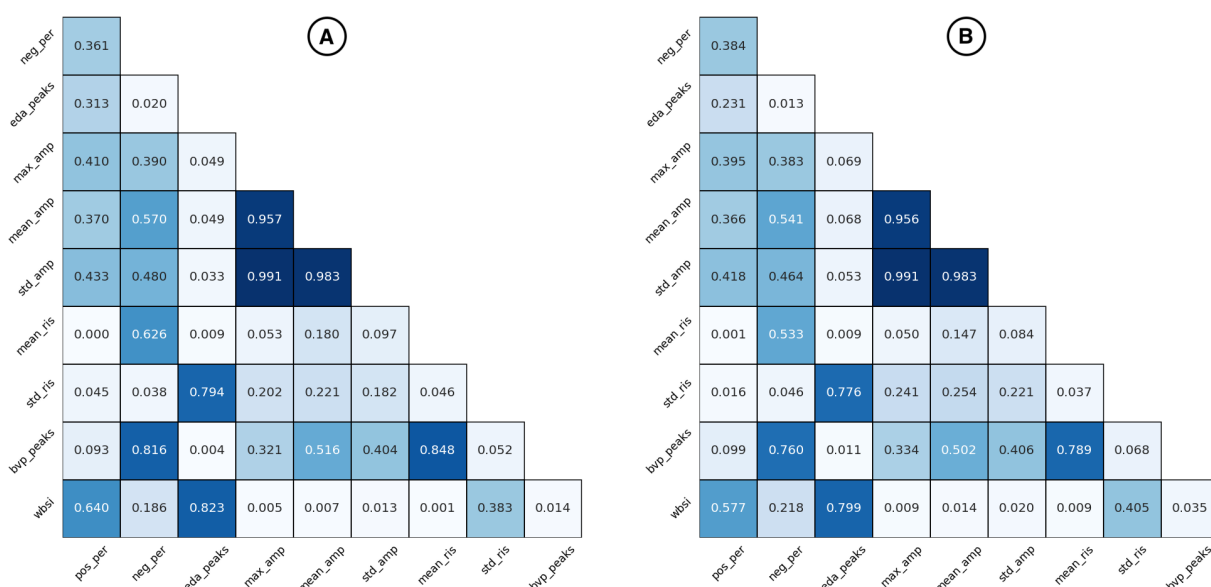
A partir dos percentuais obtidos e levando em consideração a média e o desvio padrão do trabalho original de Watson, Anna e Tellegen (1988), que leva em consideração os afetos do dia da coleta, foram atribuídos valores para cada faixa de percentual. Os afetos positivos abaixo de 42% foram considerados baixos (*low*) e tiveram o valor 1 atribuído, entre 42% e 75% foram considerados médios (*average*) e receberam o valor 2, e os afetos acima de 75% foram considerados altos (*high*) e classificados com o valor 3. Os afetos negativos abaixo de 20% foram considerados baixos (valor 1), entre 20% e 45% considerados médios (valor 2) e acima de 45% considerados altos (valor 3).

Como a coleta de dados dos sinais vitais ocorreu apenas uma vez para cada voluntário, foi necessário gerar dados sintéticos para validar o índice proposto. Para isso foi utilizada a biblioteca *SDV* para a linguagem *Python*, capaz de gerar dados sintéticos mantendo as propriedades estatísticas. A biblioteca implementa o método estatístico Cópula Gaussiana para geração dos dados, recebendo como parâmetros os registros originais, os metadados com os nomes e tipos de cada variável e o número dos registros sintéticos desejados (PATKI; WEDGE; VEERAMACHANENI, 2016).

Com os cinco dias de dados obtidos foram gerados mais trezentos e sessenta. Para confirmar a validade estatística foram criadas duas matrizes de covariância com os dados originais e com o total de dados usados. A Figura 24 mostra as matrizes de covariância compostas pelas variáveis do índice de bem-estar e satisfação e pelo próprio índice, (A) antes da geração dos dados sintéticos e (B) após a geração dos dados. Cada elemento da matriz apresenta o coeficiente de determinação, que indica maior covariância entre as variáveis de valor mais alto. Isto significa que essas variáveis tendem a variar juntas. As variáveis *pos\_per* e *neg\_per* representam, respectivamente, os percentuais de afetos positivos e negativos. A variável *wbsi* representa o valor calculado do índice de bem-estar e satisfação. As outras variáveis representam os de-

mais componentes do índice: picos de EDA (*eda\_peaks*), máxima amplitude de EDA (*max\_amp*), média da amplitude de EDA (*mean\_amp*), desvio padrão da amplitude de EDA (*std\_amp*), média do tempo de subida dos picos de EDA (*mean\_ris*), desvio padrão dos picos de subida de EDA (*std\_ris*) e picos de BVP (*bvp\_peaks*).

Figura 24 – Matrizes de covariância com as variáveis do índice de bem-estar e satisfação.



Fonte: Elaborado pelo autor.

A comparação das matrizes usando como método estatístico a Distância de Mahalanobis indicou o valor de 0,245. Como as matrizes utilizaram o coeficiente de determinação, que varia entre 0 e 1, este valor pode ser interpretado como baixo, indicando similaridade entre as matrizes (LEYS et al., 2018).

Partindo do princípio que afetos positivos podem estar relacionados ao bem-estar subjetivo e atenuação do estresse (VAN STEENBERGEN et al., 2021) e levando em consideração que sinais vitais como EDA (POSADA-QUINTERO; CHON, 2020; WICKRAMASURIYA; QI; FAGHIH, 2018) e BVP (MEDEIROS et al., 2022; BELLANTE et al., 2021) também possuem relação com estresse, é proposta a utilização dos sinais vitais para prever os níveis dos estados afetivos positivos e negativos.

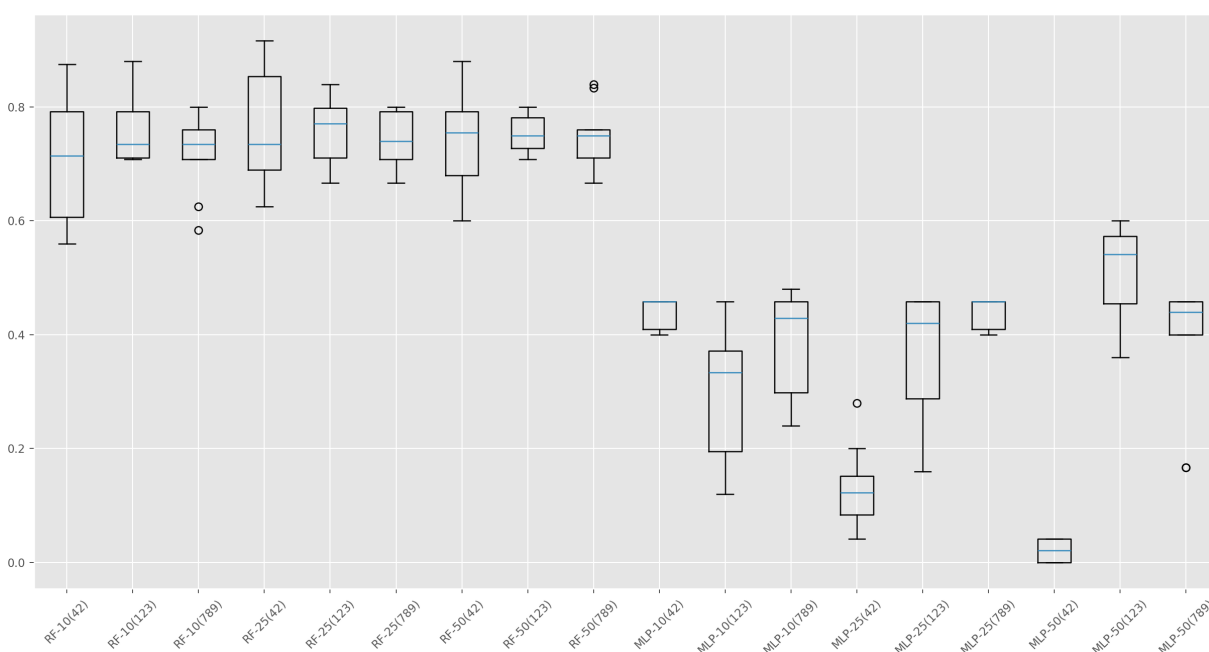
Para isso, os níveis dos estados afetivos diários foram combinados numericamente de modo a servir como rótulo para os algoritmos. Ou seja, se um determinado voluntário apresentou 80% de pontuação nos afetos positivos e 20% nos afetos negativos, seus afetos positivos foram considerados altos (valor 3) e seus afetos negativos foram considerados médios (valor 2). Logo, o rótulo para os afetos desse voluntário é definido como 32, que representa a concatenação de 3 e 2.

Os dados dos sinais vitais e níveis de afetos foram organizados em formato de vetores de atributos (365), com 67% (244) dos vetores utilizados para treinamento e 33% (121) reservados para os testes. Como a grande maioria dos dados foi ge-

rada sinteticamente, a própria parcela dos dados de teste foi utilizada para validação. Foram utilizados novamente os algoritmos RF e MLP (*Bagging*) com as mesmas configurações de inicialização de *seed* (42, 123, 789) e de árvores de decisão para RF e neurônios da camada oculta para MLP (10, 25, 50).

A Figura 25 mostra os resultados da comparação em formato de dezoito boxplots, sendo nove para cada algoritmo com variações de inicialização e configuração. As medianas de RF ficaram entre 71,4% e 77,1%, e as de MLP ficaram entre 2,1% e 54,2%. O algoritmo RF apresentou acurácia de 74% nos dados de teste e o algoritmo MLP obteve 47%. A melhor configuração de parâmetros de RF foi com 25 árvores de decisão e inicialização com *seed* 42. O algoritmo MLP teve sua melhor configuração com 50 neurônios na camada oculta e inicialização com *seed* 123. Os algoritmos foram executados com validação cruzada de 10 grupos.

Figura 25 – Resultados da comparação entre os modelos RF e MLP para o serviço *Well-Being*.



Fonte: Elaborado pelo autor.

Com a melhor acurácia, o algoritmo RF foi escolhido para uso no protótipo evolutivo que estima o bem-estar subjetivo por meio de sinais vitais, usando dados dos afetos positivos e negativos armazenados em formato de históricos de contextos. O índice proposto agrega os dados dos afetos com os sinais vitais.

### 5.3.3 Protótipo Ulisses

O protótipo evolutivo ULISSES (*User Locator for Indoor aSSEs*) é um módulo do Odisseu e foi desenvolvido com a linguagem *Python* e a biblioteca *PyQt* (WILL-MAN, 2020) como *frontend*, também utilizando as bibliotecas *Matplotlib* e *Seaborn*

para geração de gráficos (BISONG; BISONG, 2019). O aplicativo integra os serviços inteligentes *Indoor Location* e *Well-Being (WBSI)* em uma tela de consulta que mostra o posicionamento de um funcionário no interior da fábrica, seus sinais vitais e o valor do índice de bem-estar e satisfação. A consulta é feita por meio do nome do funcionário, aqui representado por um codinome para preservar a identidade e manter a privacidade, e um horário ocorrido durante o expediente de trabalho.

A Figura 26 mostra a tela do protótipo com o nome do funcionário indicado pelo campo *User* e o horário indicado pelo campo *Time*, no canto superior esquerdo. Abaixo desses dois campos encontra-se o botão *Find* usado para executar a consulta. A localização do funcionário é mostrada pelo campo *SPOT*, que indica a proximidade do ponto de interesse, e visualmente em uma planta baixa no lado direito da tela. Os campos *HRV*, *EDA*, *BVP* e *TEMP* são referentes aos sinais vitais do funcionário e correspondem respectivamente a frequência cardíaca média, atividade eletrodérmica, pulso de volume sanguíneo e temperatura da pele. O campo *WBSI* representa o índice. O botão *Clear* limpa a tela e o botão *Close* fecha o protótipo.

Figura 26 – Tela do protótipo evolutivo Ulisses.

Field	Value
SPOT	2
HRV	87.82
EDA	0.295578
BVP	-7.18640625
TEMP	33.81
WBSI	2.72299

Fonte: Elaborado pelo autor.

A Figura 27 mostra a sentença SQL que consulta os históricos de contextos (*history*, na linha 6) e retorna o ponto de interesse (*idspot*, na linha 1) e os valores de RSSI de cada um dos beacons fixos (*pi\_150*, *pi\_151*, *pi\_152* e *pi\_153*, respectivamente nas linhas 2, 3, 4 e 5) em um determinado momento do tempo (*timestamp*, nas linhas 1 e

7). A sentença condiciona o agrupamento de quatro momentos do tempo com valores idênticos para garantir a precisão do serviço de localização (linhas 8 e 9).

Figura 27 – Sentença SQL usada no protótipo.

```

1  select timestamp, idspot,
2  ( select value from history where timestamp = h.timestamp and idspot = h.idspot and idsensor = 150 ) pi_150,
3  ( select value from history where timestamp = h.timestamp and idspot = h.idspot and idsensor = 151 ) pi_151,
4  ( select value from history where timestamp = h.timestamp and idspot = h.idspot and idsensor = 152 ) pi_152,
5  ( select value from history where timestamp = h.timestamp and idspot = h.idspot and idsensor = 153 ) pi_153
6  from history h
7  where timestamp = ?
8  group by timestamp, idspot
9  having count( * ) = 4;

```

Fonte: Elaborado pelo autor.

A Figura 28 mostra a sentença SPARQL que consulta os históricos de contextos da ontologia (*:History*, na linha 5) e retorna a frequência cardíaca média (*?hrv*, nas linhas 2 e 8), a atividade eletrodérmica (*?eda*, nas linhas 2 e 9), o pulso de volume sanguíneo (*?bvp*, nas linhas 2 e 10) e a temperatura da pele (*?temp*, nas linhas 2 e 11) do usuário em um determinado momento do tempo (*?ut*, nas linhas 2 e 12).

Figura 28 – Sentença SPARQL usada no protótipo.

```

1  prefix : <http://odisseia.owl#>
2  select ?class ?individual ?ut ?hrv ?eda ?bvp ?temp
3  where {
4  ?class rdf:type owl:Class .
5  filter( ?class = :History ) .
6  ?individual rdf:type ?class .
7  ?individual :hasUnixTimestamp ?ut .
8  ?individual :hasHRV ?hrv .
9  ?individual :hasEDA ?eda .
10 ?individual :hasBVP ?bvp .
11 ?individual :hasTEMP ?temp .
12 filter( ?ut = ?? ) . }

```

Fonte: Elaborado pelo autor.

O valor do índice de bem-estar e satisfação é mostrado no campo *WBSI* e calculado a partir da equação proposta. No exemplo, o valor 2,72299 é obtido por meio dos percentuais de afetos positivos e negativos aplicados na fórmula com as variações dos dados de sinais vitais. A ideia é que quanto maior o percentual de afetos positivos e menor o percentual de afetos negativos, o valor do índice aumente indicando boa qualidade do bem-estar subjetivo no dia em que os dados foram coletados (Equação 5.3).

$$\frac{(0,8 - 0,2)}{2} + \sqrt[7]{(478 + 0,558 + 0,062 + 0,058 + 0,782 + 0,512) + \log 30381} = 2,72299 \quad (5.3)$$

Além do *frontend*, o módulo Ulisses também contempla os scripts SQL de importação dos arquivos CSV e de pré-processamento dos dados. No conceito de prototipagem evolutiva, os módulos são desenvolvidos e testados por demanda até entrarem em produção (PRESSMAN, 2005).

## 6 CONCLUSÕES

Esta tese apresentou um modelo computacional que suporta serviços inteligentes na Indústria 4.0 com base em históricos de contextos, usando como estudo de caso para validação do modelo a empresa HT Micron Semicondutores S.A., onde ocorreu o trabalho de campo com a coleta de dados para realização dos experimentos.

Para validar o modelo foram utilizados dois serviços inteligentes. Um deles é o serviço de localização de pessoas em ambientes internos (*Indoor Location*), que usa dados de beacons fixos e móveis para indicar o posicionamento de um indivíduo. O outro é o serviço voltado para o bem-estar (*Well-Being*), que tem como objetivo apresentar um índice de bem-estar e satisfação baseado em estimativas do nível de bem-estar subjetivo dos funcionários, associadas a questionários de autorrelato. As medidas do nível de bem-estar foram estimadas por meio de sinais vitais dos funcionários da indústria, como o pulso de volume sanguíneo e a atividade eletrodérmica.

A revisão sistemática realizada contribuiu para a identificação de métodos e ferramentas de ciência de dados utilizadas no meio industrial, ajudando na escolha das técnicas mais adequadas para implementação dos serviços inteligentes com base em históricos de contextos. Além disso, a taxonomia proposta ajudou na criação da ontologia implementada como parte do modelo computacional.

### 6.1 Contribuições

A partir das evidências positivas geradas em relação à resposta da questão de pesquisa, são descritas a seguir as principais contribuições desta tese para a área da computação aplicada com foco no ambiente industrial:

- **Modelo computacional Odisseu:** o modelo proposto possibilita o acompanhamento dos dados desde a entrada, seja por meio de sensores ou de autorrelato, até a saída em formato de serviços inteligentes;
- **Ontologia para gerenciamento do modelo:** a ontologia representa o domínio de conhecimento e atua como elo do modelo com os serviços inteligentes, além de funcionar também como repositório de dados organizados em formato de históricos de contextos;
- **Serviço inteligente para localização de funcionários:** o serviço de localização de funcionários no interior da fábrica se mostrou eficaz nos testes e na validação realizados em ambiente de produção, com potencial de utilização tanto consultando dados históricos como em tempo real;
- **Serviço inteligente para estimativa do bem-estar:** o serviço para estimativa

do bem-estar funciona em conjunto com a proposta do índice de bem-estar e satisfação, possibilitando um melhor acompanhamento do bem-estar subjetivo dos funcionários.

## 6.2 Limitações

O reconhecimento do bem-estar subjetivo representa um grande desafio para o modelo. Além disso, a utilização de sinais de rádio para localização interna também apresenta questões referentes à precisão. Dessa forma, algumas escolhas podem representar limitações e são passíveis de discussão, tais como:

- A quantidade limitada de dados fisiológicos coletados durante o experimento pode afetar diretamente tanto na identificação dos afetos como nos valores do índice proposto, apesar da utilização de dados sintéticos;
- O tipo de ferramenta de autorrelato com frequência diária é capaz de identificar o estado geral do funcionário, porém é incapaz de indicar um momento específico do dia em que possa ter ocorrido uma alteração nos afetos positivos e negativos;
- Os algoritmos utilizados no serviço de bem-estar e a variação dos parâmetros de configuração, como o número de árvores de decisão de RF ou a quantidade de neurônios da camada oculta de MLP, podem melhorar a acurácia da estimativa;
- Situações ocorridas durante o expediente de trabalho podem afetar os sinais vitais, como o ato de subir uma escada ou carregar um equipamento pesado;
- A distância entre os pontos de interesse e o número de pontos escolhidos para localização interna em relação ao tamanho da sala utilizada podem afetar a precisão do serviço de localização;
- Campos eletromagnéticos gerados por alguns tipos de máquinas da linha de produção podem gerar interferência no sinal RSSI, ocasionando perda ou atenuação do sinal, dependendo da intensidade do campo.

## 6.3 Trabalhos Futuros

A partir da avaliação do Odisseu e com base nos resultados obtidos, novas oportunidades para trabalhos futuros podem ser exploradas:

- Usar a ontologia como base de dados única, concentrando todos os dados em um mesmo repositório de modo a explorar todo o potencial da linguagem *OWL*;



- Implementar o reconhecimento de contextos específicos usando como base a frequência dos históricos obtidos, com o objetivo de identificar ciclos de dados que possam ser utilizados na predição do bem-estar;
- Variar os tipos de sinais vitais e de ferramentas de autorrelato com o objetivo de melhorar a estimativa do nível de bem-estar e identificar com maior precisão, também, situações de ansiedade e estresse;
- Desenvolver um aplicativo de gerenciamento de funcionários em tempo real, utilizando os algoritmos já treinados e implementados tanto no serviço de localização interna quanto no serviço de bem-estar;
- Incluir no modelo um serviço inteligente voltado para identificação do trabalho em equipe utilizando o serviço de localização interna;
- Incluir no modelo um serviço inteligente voltado para o aprendizado colaborativo utilizando o serviço de identificação do trabalho em equipe e o bem-estar subjetivo dos funcionários envolvidos;
- Implementar tolerância a falhas no serviço de localização interna, treinando os algoritmos com cada um dos beacons fixos desligados, de modo a prevenir erros na localização de funcionários em caso de pane em algum dos beacons fixos.

#### **6.4 Publicações**

Dois artigos foram publicados como parte desta tese. A revisão sistemática sobre métodos e ferramentas de ciência de dados, que é parte do Capítulo 3, foi publicada na revista *Sensors* (ARRUDA et al., 2023a). Parte da ontologia apresentada no Capítulo 4, baseada em contextos e orientada ao bem-estar emocional, foi publicada na *Revista Brasileira de Computação Aplicada* (ARRUDA et al., 2023b).

Ademais, a coautoria em um mapeamento sistemático sobre Internet das Coisas e Bem-estar Ocupacional na Indústria 4.0, ajudou na elaboração e desenvolvimento da revisão sistemática supracitada, além de contribuir com o melhor entendimento de conceitos sobre o bem-estar no ambiente de trabalho. O mapeamento sistemático foi publicado na revista *Computers & Industrial Engineering* (BAVARESCO et al., 2021).

## REFERÊNCIAS

- ABBASI, R.; MARTINEZ, P.; AHMAD, R. An ontology model to represent aquaponics 4.0 system's knowledge. **Information Processing in Agriculture**, [S.l.], p. S2214317321000937, Dec. 2021.
- AKIN, Ö. et al. Enabling Big Data Analytics at Manufacturing Fields of Farplas Automotive. In: INTELLIGENT AND Fuzzy Techniques: Smart AND Innovative Solutions, 2021, Cham. **Anais...** Springer International Publishing, 2021. p. 817–824. (Advances in Intelligent Systems and Computing).
- ALEXOPOULOS, T.; PACKIANATHER, M. A monitoring and data analysis system to achieve zero-defects manufacturing in highly regulated industries. **Smart Innovation, Systems and Technologies**, [S.l.], v. 68, p. 303–313, 2017.
- ALVES, A. P. et al. Uncovering Lying Detection using Blood Volume Pulse and Electrodermal Activity. In: Conference on Telecommunications (ConfTele), 9., 2013. **Proceedings...** [S.l.: s.n.], 2013.
- STAAB, S.; STUDER, R. (Ed.). Web ontology language: owl. In: \_\_\_\_\_. **Handbook on ontologies**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 67–92.
- APILETTI, D. et al. iSTEP, an Integrated Self-Tuning Engine for Predictive Maintenance in Industry 4.0. In: IEEE Intl Conf ON Parallel Distributed Processing WITH Applications, Ubiquitous Computing Communications, Big Data Cloud Computing, Social Computing Networking, Sustainable Computing Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 924–931.
- ARANDA, J. A. S. et al. A computational model for adaptive recording of vital signs through context histories. **Journal of Ambient Intelligence and Humanized Computing**, [S.l.], 2021.
- ARRUDA, H. M. et al. Data Science Methods and Tools for Industry 4.0: A Systematic Literature Review and Taxonomy. **Sensors**, [S.l.], v. 23, n. 11, p. 5010, May 2023.
- ARRUDA, H. M. et al. OntoLearnWell: uma ontologia baseada em contextos orientada ao bem-estar emocional no ambiente educacional. **Revista Brasileira de Computação Aplicada**, [S.l.], v. 15, n. 1, p. 59–69, Apr. 2023.
- AYVAZ, S.; ALPAY, K. Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time. **Expert Systems with Applications**, [S.l.], v. 173, p. 114598, July 2021.
- AZAMFAR, M.; LI, X.; LEE, J. Deep Learning-Based Domain Adaptation Method for Fault Diagnosis in Semiconductor Manufacturing. **IEEE Transactions on Semiconductor Manufacturing**, [S.l.], v. 33, n. 3, p. 445–453, 2020.
- BAGOZI, A.; BIANCHINI, D.; ANTONELLIS, V. D. Context-Based Resilience in Cyber-Physical Production System. **Data Science and Engineering**, [S.l.], v. 6, n. 4, p. 434–454, Dec. 2021.

BAMPOULA, X. et al. A Deep Learning Model for Predictive Maintenance in Cyber-Physical Production Systems Using LSTM Autoencoders. **Sensors**, [S.l.], v. 21, n. 3, p. 972, Feb. 2021.

BAQUERIZO, J. et al. Siamese Neural Networks for Damage Detection and Diagnosis of Jacket-Type Offshore Wind Turbine Platforms. **Mathematics**, [S.l.], v. 10, n. 7, p. 1131, Apr. 2022.

BARLAS, P.; LANNING, I.; HEAVEY, C. A survey of open source data science tools. **International Journal of Intelligent Computing and Cybernetics**, [S.l.], v. 8, p. 232–261, 2015.

BAVARESCO, R. et al. Design and evaluation of a context-aware model based on psychophysiology. **Computer Methods and Programs in Biomedicine**, [S.l.], v. 189, p. 105299, 2020.

BAVARESCO, R. et al. Internet of Things and occupational well-being in industry 4.0: A systematic mapping study and taxonomy. **Computers & Industrial Engineering**, [S.l.], v. 161, p. 107670, Nov. 2021.

BECHER, M. et al. Situated Visual Analysis and Live Monitoring for Manufacturing. **IEEE Computer Graphics and Applications**, [S.l.], v. 42, n. 2, p. 33–44, Mar. 2022.

BELHADI, A. et al. Understanding Big Data Analytics for Manufacturing Processes: Insights from Literature Review and Multiple Case Studies. **Computers & Industrial Engineering**, [S.l.], v. 137, p. 106099, Nov. 2019.

BELLANTE, A. et al. EMoCy: Towards Physiological Signals-Based Stress Detection. In: IEEE EMBS International Conference ON Biomedical AND Health Informatics (BHI), 2021., 2021, Athens, Greece. **Anais...** IEEE, 2021. p. 1–4.

BENGIO, Y.; LECUN, Y.; HINTON, G. Deep learning for AI. **Communications of the ACM**, [S.l.], v. 64, n. 7, p. 58–65, July 2021.

BERGES, I.; RAMÍREZ-DURÁN, V. J.; ILLARRAMENDI, A. A Semantic Approach for Big Data Exploration in Industry 4.0. **Big Data Research**, [S.l.], v. 25, p. 100222, July 2021.

BIBAUD-ALVES, J.; THOMAS, P.; EL HAOUZI, H. Demand forecasting using artificial neuronal networks and time series: Application to a French furniture manufacturer case study. In: IJCCI 2019 - PROCEEDINGS OF THE 11TH INTERNATIONAL JOINT CONFERENCE ON COMPUTATIONAL INTELLIGENCE, 2019. **Anais...** [S.l.: s.n.], 2019. p. 502–507.

BISONG, E.; BISONG, E. Matplotlib and seaborn. **Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners**, [S.l.], p. 151–165, 2019. Publisher: Springer.

BOUSDEKIS, A. et al. A RAMI 4.0 View of Predictive Maintenance: Software Architecture, Platform and Case Study in Steel Industry. In: ADVANCED Information Systems Engineering Workshops, 2019. **Anais...** Springer International Publishing, 2019. p. 95–106. (Lecture Notes in Business Information Processing).

- BRADLEY, M. M.; LANG, P. J. Measuring emotion: The self-assessment manikin and the semantic differential. **Journal of Behavior Therapy and Experimental Psychiatry**, [S.l.], v. 25, n. 1, p. 49–59, Mar. 1994.
- BRZYCHCZY, E.; TRZCIONKOWSKA, A. Process-Oriented Approach for Analysis of Sensor Data from Longwall Monitoring System. In: INTELLIGENT Systems IN Production Engineering AND Maintenance, 2019. **Anais...** Springer International Publishing, 2019. p. 611–621. (Advances in Intelligent Systems and Computing).
- CANDANEDO, I. S. et al. Machine Learning Predictive Model for Industry 4.0. In: UDEN, L.; HADZIMA, B.; TING, I.-H. (Ed.). **Knowledge Management in Organizations**. Cham: Springer International Publishing, 2018. v. 877, p. 501–510. Series Title: Communications in Computer and Information Science.
- CANIZO, M. et al. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. **Neurocomputing**, [S.l.], v. 363, p. 246–260, 2019.
- CARLETTI, M. et al. A deep learning approach for anomaly detection with industrial time series data: A refrigerators manufacturing case study. In: PROCEDIA MANUFACTURING, 2019. **Anais...** [S.l.: s.n.], 2019. v. 38, p. 233–240.
- CARVALHO, E. C. et al. Towards a Smart Fault Tolerant Indoor Localization System Through Recurrent Neural Networks. In: International Joint Conference ON Neural Networks (IJCNN), 2019., 2019, Budapest, Hungary. **Anais...** IEEE, 2019. p. 1–7.
- CARVALHO, H. W. d. et al. Structural validity and reliability of the Positive and Negative Affect Schedule (PANAS): Evidence from a large Brazilian community sample. **Revista Brasileira de Psiquiatria**, [S.l.], v. 35, n. 2, p. 169–172, June 2013.
- CAVALLO, F. et al. Mood classification through physiological parameters. **Journal of Ambient Intelligence and Humanized Computing**, [S.l.], v. 12, n. 4, p. 4471–4484, Apr. 2021.
- CHEN, B. et al. Time Series Data for Equipment Reliability Analysis with Deep Learning. **IEEE Access**, [S.l.], v. 8, p. 105484–105493, 2020.
- CHEN, Y.-J.; LEE, Y.-H.; CHIU, M.-C. Construct an Intelligent Yield Alert and Diagnostic Analysis System via Data Analysis: Empirical Study of a Semiconductor Foundry. In: ADVANCES IN Production Management Systems. Smart Manufacturing FOR Industry 4.0, 2018. **Anais...** Springer International Publishing, 2018. p. 394–401. (IFIP Advances in Information and Communication Technology).
- CHOLLET, F. et al. **Keras**. Disponível em: <<https://github.com/fchollet/keras>>. Acesso em: jul. 2021.
- CORDONI, F. et al. A multi-modal unsupervised fault detection system based on power signals and thermal imaging via deep AutoEncoder neural network. **Engineering Applications of Artificial Intelligence**, [S.l.], v. 110, p. 104729, Apr. 2022.
- COŞKUN, B. et al. A physiological signal database of children with different special needs for stress recognition. **Scientific Data**, [S.l.], v. 10, n. 1, p. 382, June 2023.

CUI, Y.; KARA, S.; CHAN, K. C. Manufacturing big data ecosystem: A systematic literature review. **Robotics and Computer-Integrated Manufacturing**, [S.l.], v. 62, p. 101861, Apr. 2020.

DAVENPORT, T. H.; PATIL, D. J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard business review**, [S.l.], v. 90, n. 5, p. 70–76, 2012.

DEY, A. K.; ABOWD, G. D.; SALBER, D. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-aware Applications. **Hum.-Comput. Interact.**, [S.l.], v. 16, n. 2, p. 97–166, Dec. 2001.

DIENER, E.; OISHI, S.; TAY, L. Advances in subjective well-being research. **Nature Human Behaviour**, [S.l.], v. 2, n. 4, p. 253–260, Feb. 2018.

DING, H.; YANG, L.; YANG, Z. A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data. **IEEE Access**, [S.l.], v. 7, p. 108684–108702, 2019.

DING, M. et al. A Data Analytic Engine Towards Self-Management of Cyber-Physical Systems. In: IEEE 33RD International Conference ON Distributed Computing Systems Workshops, 2013., 2013. **Anais...** [S.l.: s.n.], 2013. p. 303–308. ISSN: 1545-0678, 2332-5666.

DOBRE, C.; XHAFA, F. Intelligent services for Big Data science. **Future Generation Computer Systems**, [S.l.], v. 37, p. 267–281, July 2014.

DZEDZICKIS, A.; KAKLAUSKAS, A.; BUCINSKAS, V. Human Emotion Recognition: Review of Sensors and Methods. **Sensors**, [S.l.], v. 20, n. 3, p. 592, Jan. 2020.

EL WAHABI, A. et al. Detection and Control System for Automotive Products Applications by Artificial Vision Using Deep Learning. **Advances in Intelligent Systems and Computing**, [S.l.], v. 1104 AISC, p. 224–241, 2020.

EMPATICA. **E4 wristband - real-time physiological data streaming and visualization**. Disponível em: <<http://https://www.empatica.com/research/e4>>. Acesso em: set. 2021.

ESSIEN, A.; GIANNETTI, C. A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders. **IEEE Transactions on Industrial Informatics**, [S.l.], v. 16, n. 9, p. 6069–6078, 2020.

ESTEVE-GONZALEZ, P. et al. Cybersecurity implications of changing patterns of office, home, and hybrid work: an exploratory global survey. **Home, and Hybrid Work: An Exploratory Global Survey (January 11, 2023)**, [S.l.], 2023.

EZE, E.; HALSE, S.; AJMAL, T. Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm. **Water**, [S.l.], v. 13, n. 13, p. 1782, June 2021.

FERREIRA, B. V. et al. Exploiting the Use of Convolutional Neural Networks for Localization in Indoor Environments. **Applied Artificial Intelligence**, [S.l.], p. 1–9, May 2017.

FILIPPETTO, A. S.; LIMA, R.; BARBOSA, J. L. V. A risk prediction model for software project management based on similarity analysis of context histories. **Information and Software Technology**, [S.l.], v. 131, p. 106497, 2021.

FU, W.; CHIEN, C.-F.; LIN, Z.-H. A Hybrid Forecasting Framework with Neural Network and Time-Series Method for Intermittent Demand in Semiconductor Supply Chain. In: ADVANCES IN Production Management Systems. Smart Manufacturing FOR Industry 4.0, 2018. **Anais...** Springer International Publishing, 2018. p. 65–72. (IFIP Advances in Information and Communication Technology).

GARMAROODI, M. S. S. et al. Detection of Anomalies in Industrial IoT Systems by Data Mining: Study of CHRIST Osmotron Water Purification System. **IEEE Internet of Things Journal**, [S.l.], v. 8, n. 13, p. 10280–10287, July 2021.

GHOSH, N.; BANERJEE, I. IoT-based seismic hazard detection in coal mines using grey systems theory. In: INTERNATIONAL WIRELESS COMMUNICATIONS AND MOBILE COMPUTING CONFERENCE, IWCMC 2019, 2019., 2019. **Anais...** [S.l.: s.n.], 2019. p. 871–876.

GIL, D. et al. Internet of Things: A Review of Surveys Based on Context Aware Intelligent Services. **Sensors**, [S.l.], v. 16, n. 7, p. 1069, July 2016.

GIUSTOZZI, F.; SAUNIER, J.; ZANNI-MERK, C. Context modeling for industry 4.0: an ontology-based proposal. **Procedia Computer Science**, [S.l.], v. 126, p. 675 – 684, 2018. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia.

GRASS, A.; BEECKS, C.; SOTO, J. A. C. Unsupervised Anomaly Detection in Production Lines. In: MACHINE Learning FOR Cyber Physical Systems, 2019. **Anais...** Springer Berlin Heidelberg, 2019. p. 18–25. (Technologien für die intelligente Automation).

GROENE, B. **Technical architecture modeling (tam) - the sap way combining fmc and uml**. Disponível em: <<http://www.fmc-modeling.org/fmc-and-tam>>. Acesso em: jul. 2021.

HALL, S. A mental-health crisis is gripping science — toxic research culture is to blame. **Nature**, [S.l.], v. 617, n. 7962, p. 666–668, May 2023.

HOPPENSTEDT, B. et al. CONSENSORS: A Neural Network Framework for Sensor Data Analysis. In: ON THE Move TO Meaningful Internet Systems: OTM 2018 Workshops, 2019. **Anais...** Springer International Publishing, 2019. p. 196–200. (Lecture Notes in Computer Science).

HSU, C.-Y.; LIU, W.-C. Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. **Journal of Intelligent Manufacturing**, [S.l.], v. 32, n. 3, p. 823–836, Mar. 2021.

HTMICRON, S. **iMCP HT32SX V2.2 – SiP Sigfox**. 2022.

HUANG, X.; ZANNI-MERK, C.; CRÉMILLEUX, B. Enhancing deep learning with semantics: An application to manufacturing time series analysis. In: PROCEDIA COMPUTER SCIENCE, 2019. **Anais...** [S.l.: s.n.], 2019. v. 159, p. 437–446.

- HUH, J.-H.; SEO, K. An Indoor Location-Based Control System Using Bluetooth Beacons for IoT Systems. **Sensors**, [S.I.], v. 17, n. 12, p. 2917, Dec. 2017.
- IGUAL, L.; SEGUÍ, S. **Introduction to Data Science**. Cham: Springer International Publishing, 2017. (Undergraduate Topics in Computer Science).
- JANG, G.-B.; CHO, S.-B. Feature Space Transformation for Fault Diagnosis of Rotating Machinery under Different Working Conditions. **Sensors**, [S.I.], v. 21, n. 4, p. 1417, Feb. 2021.
- JIANG, J.-R.; LEE, J.-E.; ZENG, Y.-M. Time series multiple channel convolutional neural network with attention-based long short- term memory for predicting bearing remaining useful life. **Sensors (Switzerland)**, [S.I.], v. 20, n. 1, 2020.
- KAGERMANN, H.; WAHLSTER, W.; HELBIG, J. **Recommendations for implementing the strategic initiative INDUSTRIE 4.0**. [S.I.]: acatech – National Academy of Science and Engineering, 2013.
- KANCHARLA, C. R. et al. Latent Dimensions of Auto-Encoder as Robust Features for Inter-Conditional Bearing Fault Diagnosis. **Applied Sciences**, [S.I.], v. 12, n. 3, p. 965, Jan. 2022.
- KAYAN, H. et al. AnoML-IoT: An end to end re-configurable multi-protocol anomaly detection pipeline for Internet of Things. **Internet of Things**, [S.I.], v. 16, p. 100437, Dec. 2021.
- KESHAV, S. How to Read a Paper. **SIGCOMM Comput. Commun. Rev.**, [S.I.], v. 37, n. 3, p. 83–84, July 2007.
- KHODABAKHSH, A. et al. Multivariate Sensor Data Analysis for Oil Refineries and Multi-mode Identification of System Behavior in Real-time. **IEEE Access**, [S.I.], v. 6, p. 63489–64405, 2018.
- KIANGALA, K.; WANG, Z. An Effective Predictive Maintenance Framework for Conveyor Motors Using Dual Time-Series Imaging and Convolutional Neural Network in an Industry 4.0 Environment. **IEEE Access**, [S.I.], v. 8, p. 121033–121049, 2020.
- KIM, J.; LEE, J. Y. Data-analytics-based factory operation strategies for die-casting quality enhancement. **The International Journal of Advanced Manufacturing Technology**, [S.I.], v. 119, n. 5, p. 3865–3890, Mar. 2022.
- KIM, N.; KIM, J.; AHN, C. R. Predicting workers' inattentiveness to struck-by hazards by monitoring biosignals during a construction task: A virtual reality experiment. **Advanced Engineering Informatics**, [S.I.], v. 49, p. 101359, Aug. 2021.
- KOLDIJK, S.; NEERINCX, M. A.; KRAAIJ, W. Detecting Work Stress in Offices by Combining Unobtrusive Sensors. **IEEE Transactions on Affective Computing**, [S.I.], v. 9, n. 2, p. 227–239, Apr. 2018.
- KOSKIMÄKI, H. et al. Early detection of migraine attacks based on wearable sensors: experiences of data collection using Empatica E4. In: ACM International Joint Conference ON Pervasive AND Ubiquitous Computing AND Proceedings OF THE 2017 ACM International Symposium ON Wearable Computers, 2017., 2017, Maui Hawaii. **Proceedings...** ACM, 2017. p. 506–511.

KOVÁCS, T.; KŐ, A. Machine Learning Based Monitoring of the Pneumatic Actuators' Behavior Through Signal Processing Using Real-World Data Set. In: **FUTURE Data AND Security Engineering**, 2019, Cham. **Anais...** Springer International Publishing, 2019. p. 33–44. (Lecture Notes in Computer Science).

KOVACS, T.; KO, A. Monitoring Pneumatic Actuators' Behavior Using Real-World Data Set. **SN Computer Science**, [S.l.], v. 1, n. 4, p. 196, June 2020.

KURPANIK, J. et al. EYE: Big data system supporting preventive and predictive maintenance of robotic production lines. **Communications in Computer and Information Science**, [S.l.], v. 928, p. 47–60, 2018.

LAMY, J.-B. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. **Artificial Intelligence in Medicine**, [S.l.], v. 80, p. 11–28, July 2017.

LAMY, J.-B. **Ontologies with Python - Programming OWL 2.0 Ontologies with Python and Owlready2**. [S.l.]: Apress, 2021.

LEPENIOTI, K. et al. Machine Learning for Predictive and Prescriptive Analytics of Operational Data in Smart Manufacturing. In: **ADVANCED Information Systems Engineering Workshops**, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 5–16. (Lecture Notes in Business Information Processing).

LEYS, C. et al. Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance. **Journal of Experimental Social Psychology**, [S.l.], v. 74, p. 150–156, Jan. 2018.

LI, Z.; WANG, Y.; WANG, K. A deep learning driven method for fault classification and degradation assessment in mechanical equipment. **Computers in Industry**, [S.l.], v. 104, p. 1–10, 2019.

LIAO, Y. et al. Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal. **International Journal of Production Research**, [S.l.], v. 55, n. 12, p. 3609–3629, June 2017.

LIMA, M. J. de et al. HealthMon: An approach for monitoring machines degradation using time-series decomposition, clustering, and metaheuristics. **Computers & Industrial Engineering**, [S.l.], v. 162, p. 107709, Dec. 2021.

LISOWSKA, A.; WILK, S.; PELEG, M. Catching Patient's Attention at the Right Time to Help Them Undergo Behavioural Change: Stress Classification Experiment from Blood Volume Pulse. In: TUCKER, A. et al. (Ed.). **Artificial Intelligence in Medicine**. Cham: Springer International Publishing, 2021. v. 12721, p. 72–82. Series Title: Lecture Notes in Computer Science.

LU, Y.-W.; HSU, C.-Y.; HUANG, K.-C. An autoencoder gated recurrent unit for remaining useful life prediction. **Processes**, [S.l.], v. 8, n. 9, 2020.

LUO, S.; LIU, H.; QI, E. Big data analytics – enabled cyber-physical system: model and applications. **Industrial Management and Data Systems**, [S.l.], v. 119, n. 5, p. 1072–1088, 2019.



MA, S. et al. Big data driven predictive production planning for energy-intensive manufacturing industries. **Energy**, [S.I.], v. 211, 2020.

MACHADO, S. D. et al. Ambient intelligence based on iot for assisting people with alzheimer's disease through context histories. **Electronics**, [S.I.], v. 10, n. 11, 2021.

MAHMOOD, J.; LUO, M.; REHMAN, M. An accurate detection of tool wear type in drilling process by applying PCA and one-hot encoding to SSA-BLSTM model. **The International Journal of Advanced Manufacturing Technology**, [S.I.], v. 118, n. 11-12, p. 3897–3916, Feb. 2022.

MAJDANI, F.; PETROVSKI, A.; DOOLAN, D. Designing a Context-Aware Cyber Physical System for Smart Conditional Monitoring of Platform Equipment. In: **ENGINEERING Applications OF Neural Networks**, 2016. **Anais...** Springer International Publishing, 2016. p. 198–210. (Communications in Computer and Information Science).

MAJUMDER, S.; MONDAL, T.; DEEN, M. Wearable Sensors for Remote Health Monitoring. **Sensors**, [S.I.], v. 17, n. 12, p. 130, Jan. 2017.

MAKOWSKI, D. et al. NeuroKit2: A Python toolbox for neurophysiological signal processing. **Behavior Research Methods**, [S.I.], v. 53, n. 4, p. 1689–1696, Aug. 2021.

MARIN, C. A. et al. A Conceptual Architecture Based on Intelligent Services for Manufacturing Support Systems. In: **IEEE International Conference ON Systems, Man, AND Cybernetics**, 2013., 2013, Manchester. **Anais...** IEEE, 2013. p. 4749–4754.

MARTINI, B. G. et al. Indoorplant: a model for intelligent services in indoor agriculture based on context histories. **Sensors**, [S.I.], v. 21, n. 5, 2021.

MATEUS, B. C. et al. Anticipating Future Behavior of an Industrial Press Using LSTM Networks. **Applied Sciences**, [S.I.], v. 11, n. 13, p. 6101, June 2021.

MAZZEI, D.; RAMJATTAN, R. Machine learning for industry 4.0: a systematic review using deep learning-based topic modelling. **Sensors**, [S.I.], v. 22, n. 22, 2022.

MEDEIROS, V. P. et al. Um Modelo de Aprendizado Profundo Multimodal para Classificação de Estresse Utilizando Sinais Obtidos por Dispositivos Vestíveis de Pulso. In: **XXII Simpósio Brasileiro DE Computação Aplicada À Saúde (SBCAS 2022)**, 2022, Brasil. **Anais...** Sociedade Brasileira de Computação - SBC, 2022. p. 370–380.

MILSTEIN, N.; GORDON, I. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. **Frontiers in Behavioral Neuroscience**, [S.I.], v. 14, p. 148, Aug. 2020.

MIN, Q. et al. Machine Learning based Digital Twin Framework for Production Optimization in Petrochemical Industry. **International Journal of Information Management**, [S.I.], v. 49, p. 502–519, 2019.

- MOHSEN, S.; ELKASEER, A.; SCHOLZ, S. G. Industry 4.0-Oriented Deep Learning Models for Human Activity Recognition. **IEEE Access**, [S.l.], v. 9, p. 150508–150521, 2021.
- MOSAVI, N. S. et al. Intelligent energy management using data mining techniques at Bosch Car Multimedia Portugal facilities. **Procedia Computer Science**, [S.l.], v. 201, p. 503–510, 2022.
- MULRENNAN, K. et al. A data science approach to modelling a manufacturing facility's electrical energy profile from plant production data. In: IEEE 5TH INTERNATIONAL CONFERENCE ON DATA SCIENCE AND ADVANCED ANALYTICS, DSAA 2018, 2018., 2019. **Proceedings...** [S.l.: s.n.], 2019. p. 387–391.
- MUSEN, M. A. The protégé project: a look back and a look forward. **AI Matters**, [S.l.], v. 1, n. 4, p. 4–12, June 2015.
- NAQVI, S. N. Z.; YFANTIDOU, S.; ZIMÁNYI, E. Time Series Databases and InfluxDB. **Université Libre de Bruxelles**, [S.l.], 2017.
- NASKOS, A. et al. Event-Based Predictive Maintenance on Top of Sensor Data in a Real Industry 4.0 Case Study. In: MACHINE Learning AND Knowledge Discovery IN Databases, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 345–356. (Communications in Computer and Information Science).
- NASUTION, M. K. M.; SITOMPUL, O. S.; NABABAN, E. B. Data Science. **Journal of Physics: Conference Series**, [S.l.], v. 1566, n. 1, p. 012034, June 2020.
- NEWMAN, N. Apple iBeacon technology briefing. **Journal of Direct, Data and Digital Marketing Practice**, [S.l.], v. 15, n. 3, p. 222–225, Jan. 2014.
- NIEVES AVENDANO, D. et al. Anomaly detection and event mining in cold forming manufacturing processes. **The International Journal of Advanced Manufacturing Technology**, [S.l.], v. 115, n. 3, p. 837–852, July 2021.
- OKELY, J. A.; WEISS, A.; GALE, C. R. The interaction between stress and positive affect in predicting mortality. **Journal of Psychosomatic Research**, [S.l.], v. 100, p. 53–60, Sept. 2017.
- ONUS, U. et al. A Case Study on Challenges of Applying Machine Learning for Predictive Drill Bit Sharpness Estimation. In: IEEE International Workshop ON Metrology FOR Industry 4.0 & IoT (MetroInd4.0&IoT), 2021., 2021, Rome, Italy. **Anais...** IEEE, 2021. p. 275–280.
- OOI, B. et al. Operation status tracking for legacy manufacturing systems via vibration analysis. In: CONFERENCE RECORD - IEEE INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE, 2019. **Anais...** [S.l.: s.n.], 2019. v. 2019-May.
- ORDONEZ, C. A comparison of data science systems. In: BIG DATA ANALYTICS, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 3–11.
- ORRÙ, P. et al. Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry. **Sustainability (Switzerland)**, [S.l.], v. 12, n. 11, 2020.

- PARK, K. T. et al. Cyber Physical Energy System for Saving Energy of the Dyeing Process with Industrial Internet of Things and Manufacturing Big Data. **International Journal of Precision Engineering and Manufacturing-Green Technology**, [S.l.], Feb. 2019.
- PATKI, N.; WEDGE, R.; VEERAMACHANENI, K. The Synthetic Data Vault. In: IEEE International Conference ON Data Science AND Advanced Analytics (DSAA), 2016., 2016, Montreal, QC, Canada. **Anais...** IEEE, 2016. p. 399–410.
- PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, [S.l.], v. 12, p. 2825–2830, 2011.
- PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology**, [S.l.], v. 64, p. 1–18, Aug. 2015.
- PICCIALLI, F. et al. Data science for the internet of things. **IEEE Internet of Things Journal**, [S.l.], v. 7, n. 5, p. 4342–4346, 2020.
- PIRES, P. et al. Positive and Negative Affect Schedule: Psychometric Properties for the Brazilian Portuguese Version. **The Spanish Journal of Psychology**, [S.l.], v. 16, p. E58, 2013.
- POSADA-QUINTERO, H. F.; CHON, K. H. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. **Sensors**, [S.l.], v. 20, n. 2, p. 479, Jan. 2020.
- PRESSMAN, R. S. **Software engineering**: a practitioner's approach. [S.l.]: Palgrave macmillan, 2005.
- PROVOST, F.; FAWCETT, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. **Big Data**, [S.l.], v. 1, n. 1, p. 51–59, Feb. 2013.
- QUATRINI, E. et al. Predictive model for the degradation state of a hydraulic system with dimensionality reduction. In: PROCEDIA MANUFACTURING, 2020. **Anais...** [S.l.: s.n.], 2020. v. 42, p. 516–523.
- RAJPUT, S.; SINGH, S. P. Industry 4.0 Model for circular economy and cleaner production. **Journal of Cleaner Production**, [S.l.], v. 277, p. 123853, Dec. 2020.
- REHSE, J.-R.; MEHDIYEV, N.; FETTKE, P. Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory. **KI - Künstliche Intelligenz**, [S.l.], v. 33, n. 2, p. 181–187, June 2019.
- RENTZ, D. M.; HECKLER, W. F.; BARBOSA, J. L. V. A computational model for assisting individuals with suicidal ideation based on context histories. **Universal Access in the Information Society**, [S.l.], Mar. 2023.
- REZENDE, J. et al. A case study on the analysis of an injection moulding machine energy data sets for improving energy and production management. In: ECEEE INDUSTRIAL SUMMER STUDY PROCEEDINGS, 2018. **Anais...** [S.l.: s.n.], 2018. v. 2018-June, p. 231–238.

RIBEIRO, D. et al. A Comparison of Anomaly Detection Methods for Industrial Screw Tightening. In: COMPUTATIONAL Science AND Its Applications – ICCSA 2021, 2021, Cham. **Anais...** Springer International Publishing, 2021. p. 485–500. (Lecture Notes in Computer Science).

ROUSOPOULOU, V. et al. Data Analytics Towards Predictive Maintenance for Industrial Ovens. In: ADVANCED Information Systems Engineering Workshops, 2019, Cham. **Anais...** Springer International Publishing, 2019. p. 83–94. (Lecture Notes in Business Information Processing).

ROUSOPOULOU, V. et al. Cognitive analytics platform with AI solutions for anomaly detection. **Computers in Industry**, [S.l.], v. 134, p. 103555, Jan. 2022.

SANTOS, R. C.; MARTINHO, J. L. An Industry 4.0 maturity model proposal. **Journal of Manufacturing Technology Management**, [S.l.], v. 31, n. 5, p. 1023–1043, Dec. 2019.

SARDA, K. et al. A Multi-Step Anomaly Detection Strategy Based on Robust Distances for the Steel Industry. **IEEE Access**, [S.l.], v. 9, p. 53827–53837, 2021.

SCHAFFARCZYK, M. et al. Validity of the Polar H10 Sensor for Heart Rate Variability Analysis during Resting State and Incremental Exercise in Recreational Men and Women. **Sensors**, [S.l.], v. 22, n. 17, p. 6536, Aug. 2022.

SCHMIDT, P. et al. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In: ACM International Conference ON Multimodal Interaction, 20., 2018, Boulder CO USA. **Proceedings...** ACM, 2018. p. 400–408.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A systematic literature review on applying crisp-dm process model. **Procedia Computer Science**, [S.l.], v. 181, p. 526–534, 2021. CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020.

SHENFIELD, A.; HOWARTH, M. A novel deep learning model for the detection and identification of rolling element-bearing faults. **Sensors (Switzerland)**, [S.l.], v. 20, n. 18, p. 1–24, 2020.

SHRIVASTAVA, S. et al. ThunderML: A Toolkit for Enabling AI/ML Models on Cloud for Industry 4.0. In: WEB Services – ICWS 2019, 2019. **Anais...** Springer International Publishing, 2019. p. 163–180. (Lecture Notes in Computer Science).

SILVA ARANTES, J. da et al. A novel unsupervised method for anomaly detection in time series based on statistical features for industrial predictive maintenance. **International Journal of Data Science and Analytics**, [S.l.], v. 12, n. 4, p. 383–404, Oct. 2021.

SILVEIRA, G. D. N. et al. I4.0 Pilot Project on a Semiconductor Industry: Implementation and Lessons Learned. **Sensors**, [S.l.], v. 20, n. 20, p. 5752, Oct. 2020.

SOLTANALI, H. et al. A comparative study of statistical and soft computing techniques for reliability prediction of automotive manufacturing. **Applied Soft Computing**, [S.l.], v. 98, 2021.

SONG, S.; BAEK, J.-G. New anomaly detection in semiconductor manufacturing process using oversampling method. In: ICAART 2020 - PROCEEDINGS OF THE 12TH INTERNATIONAL CONFERENCE ON AGENTS AND ARTIFICIAL INTELLIGENCE, 2020. **Anais...** [S.l.: s.n.], 2020. v. 2, p. 926–932.

SONTHITED, P.; KOOLPIRUK, D.; SONGKASIRI, W. Prediction of production performance for tapioca industry using LSTM neural network. In: INTERNATIONAL CONFERENCE ON ELECTRICAL ENGINEERING/ELECTRONICS, COMPUTER, TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY, ECTI-CON 2019, 16., 2019. **Proceedings...** [S.l.: s.n.], 2019. p. 147–150.

SOUSA LIMA, W. et al. Human activity recognition based on symbolic representation algorithms for inertial sensors. **Sensors**, [S.l.], v. 18, n. 11, 2018.

STÅHL, N. et al. Using recurrent neural networks with attention for detecting problematic slab shapes in steel rolling. **Applied Mathematical Modelling**, [S.l.], v. 70, p. 365–377, 2019.

SUBRAMANIYAN, M. et al. A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines. **Computers and Industrial Engineering**, [S.l.], v. 125, p. 533–544, 2018.

SUN, D. et al. PlanningVis: A Visual Analytics Approach to Production Planning in Smart Factories. **IEEE Transactions on Visualization and Computer Graphics**, [S.l.], v. 26, n. 1, p. 579–589, 2020.

SUNDARAM, S.; ZEID, A. Smart Prognostics and Health Management (SPHM) in Smart Manufacturing: An Interoperable Framework. **Sensors**, [S.l.], v. 21, n. 18, p. 5994, Sept. 2021.

SYAFRUDIN, M. et al. Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. **Sensors (Switzerland)**, [S.l.], v. 18, n. 9, 2018.

TAGAWA, Y.; MASKELIŪNAS, R.; DAMAŠEVIČIUS, R. Acoustic Anomaly Detection of Mechanical Failures in Noisy Real-Life Factory Environments. **Electronics**, [S.l.], v. 10, n. 19, p. 2329, Sept. 2021.

TCHATCHOUA, P. et al. A Comparative Evaluation of Deep Learning Anomaly Detection Techniques on Semiconductor Multivariate Time Series Data. In: IEEE 17TH International Conference ON Automation Science AND Engineering (CASE), 2021., 2021, Lyon, France. **Anais...** IEEE, 2021. p. 1613–1620.

TEDESCO, S. et al. A Scalable Deep Learning-Based Approach for Anomaly Detection in Semiconductor Manufacturing. In: Winter Simulation Conference (WSC), 2021., 2021, Phoenix, AZ, USA. **Anais...** IEEE, 2021. p. 1–12.

TESSON, S. **Um Verão com Homero**. [S.l.]: L&PM, 2019.

TOMA, R. N. et al. A Bearing Fault Classification Framework Based on Image Encoding Techniques and a Convolutional Neural Network under Different Operating Conditions. **Sensors**, [S.I.], v. 22, n. 13, p. 4881, June 2022.

TORRES, J. F. et al. Deep learning for time series forecasting: a survey. **Big Data**, [S.I.], v. 9, n. 1, p. 3–21, feb 2021.

TRIPATHI, A.; BARUAH, R. Contextual Anomaly Detection in Time Series Using Dynamic Bayesian Network. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, [S.I.], v. 12034 LNAI, p. 333–342, 2020.

VAN HERREWEGHE, M. et al. A Machine Learning-Based Approach for Predicting Tool Wear in Industrial Milling Processes. In: MACHINE Learning AND Knowledge Discovery IN Databases, 2020, Cham. **Anais...** Springer International Publishing, 2020. p. 414–425. (Communications in Computer and Information Science).

VAN ROSSUM, G.; DRAKE, F. L. **Python 3 reference manual**. Scotts Valley, CA: CreateSpace, 2009.

VAN STEENBERGEN, H. et al. How positive affect buffers stress responses. **Current Opinion in Behavioral Sciences**, [S.I.], v. 39, p. 153–160, June 2021.

VANHECKE, T. E. Zotero. **Journal of the Medical Library Association : JMLA**, [S.I.], v. 96, n. 3, p. 275–276, July 2008.

VARELA, L. et al. Decision support visualization approach in textile manufacturing a case study from operational control in textile industry. **International Journal for Quality Research**, [S.I.], v. 13, n. 4, p. 987–1004, 2019.

VICÊNCIO, D. et al. An Intelligent Predictive Maintenance Approach Based on End-of-Line Test Logfiles in the Automotive Industry. In: INDUSTRIAL IoT Technologies AND Applications, 2021, Cham. **Anais...** Springer International Publishing, 2021. p. 121–140. (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering).

VILLALOBOS, K.; SUYKENS, J.; ILLARRAMENDI, A. A flexible alarm prediction system for smart manufacturing scenarios following a forecaster–analyzer approach. **Journal of Intelligent Manufacturing**, [S.I.], July 2020.

VRIES, D. et al. Application of machine learning techniques to predict anomalies in water supply networks. **Water Science and Technology: Water Supply**, [S.I.], v. 16, n. 6, p. 1528–1535, 2016.

WAHID, A.; BRESLIN, J. G.; INTIZAR, M. A. Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework. **Applied Sciences**, [S.I.], v. 12, n. 9, p. 4221, Apr. 2022.

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**, [S.I.], v. 34, n. 2, p. 77–84, 2013.

WANG, M. et al. Design of PM2.5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA Mode. **PLOS ONE**, [S.l.], v. 17, n. 5, p. e0267440, May 2022.

WANG, X. H. et al. Ontology based context modeling and reasoning using owl. In: IEEE ANNUAL CONFERENCE ON PERVASIVE COMPUTING AND COMMUNICATIONS WORKSHOPS, 2004. PROCEEDINGS OF THE SECOND, 2004. **Anais...** [S.l.: s.n.], 2004. p. 18–22.

WANG, Y. et al. Detecting anomalies in time series data from a manufacturing system using recurrent neural networks. **Journal of Manufacturing Systems**, [S.l.], v. 62, p. 823–834, Jan. 2022.

WATSON, D.; ANNA, L.; TELLEGEN, A. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. **Journal of Personality and Social Psychology**, [S.l.], 1988.

WICKRAMASURIYA, D. S.; QI, C.; FAGHIH, R. T. A State-Space Approach for Detecting Stress from Electrodermal Activity. In: Annual International Conference OF THE IEEE Engineering IN Medicine AND Biology Society (EMBC), 2018., 2018, Honolulu, HI. **Anais...** IEEE, 2018. p. 3562–3567.

WIEDEMANN, T. et al. Simcop: a framework for similarity analysis of context histories. **International Journal of Business Information Systems**, [S.l.], v. 1, p. 1, 01 2020.

WIJNGAARDS, I. et al. Worker Well-Being: What it Is, and how it Should Be Measured. **Applied Research in Quality of Life**, [S.l.], v. 17, n. 2, p. 795–832, Apr. 2022.

WILLMAN, J. M. **Beginning PyQt**. [S.l.]: Springer, 2020.

WOLF, H. et al. Bringing Advanced Analytics to Manufacturing: A Systematic Mapping. In: ADVANCES IN Production Management Systems. Production Management FOR THE Factory OF THE Future, 2019, Cham. **Anais...** Springer International Publishing, 2019. p. 333–340. (IFIP Advances in Information and Communication Technology).

WU, Q.; DING, K.; HUANG, B. Approach for fault prognosis using recurrent neural network. **Journal of Intelligent Manufacturing**, [S.l.], v. 31, n. 7, p. 1621–1633, Oct. 2020.

WU, W. et al. A Visual Analytics Approach for Equipment Condition Monitoring in Smart Factories of Process Industry. In: IEEE Pacific Visualization Symposium (PacificVis), 2018., 2018. **Anais...** [S.l.: s.n.], 2018. p. 140–149. ISSN: 2165-8773.

YASAEI, R.; HERNANDEZ, F.; FARUQUE, M. A. A. IoT-CAD: context-aware adaptive anomaly detection in IoT systems through sensor association. In: International Conference ON Computer-Aided Design, 39., 2020, New York, NY, USA. **Proceedings...** Association for Computing Machinery, 2020. p. 1–9. (ICCAD '20).

YUE, G.; PING, G.; LANXIN, L. An End-to-End model based on CNN-LSTM for Industrial Fault Diagnosis and Prognosis. In: IEEE INTERNATIONAL CONFERENCE

ON NETWORK INFRASTRUCTURE AND DIGITAL CONTENT, IC-NIDC 2018, 2018., 2018. **Proceedings...** [S.l.: s.n.], 2018. p. 274–278.

ZHAN, J. et al. Stgat-Mad : Spatial-Temporal Graph Attention Network For Multivariate Time Series Anomaly Detection. In: ICASSP 2022 - 2022 IEEE International Conference ON Acoustics, Speech AND Signal Processing (ICASSP), 2022, Singapore, Singapore. **Anais...** IEEE, 2022. p. 3568–3572.

ZHANG, Y. et al. A CPPS based on GBDT for predicting failure events in milling. **International Journal of Advanced Manufacturing Technology**, [S.l.], v. 111, n. 1-2, p. 341–357, 2020.

ZHOU, B. et al. SemML: Facilitating development of ML models for condition monitoring with semantics. **Journal of Web Semantics**, [S.l.], v. 71, p. 100664, Nov. 2021.

ZHOU, D. et al. Towards Ontology Reshaping for KG Generation with User-in-the-Loop: Applied to Bosch Welding. In: THE 10TH International Joint Conference ON Knowledge Graphs, 2021, Virtual Event Thailand. **Anais...** ACM, 2021. p. 145–150.

ZUFLE, M. et al. A Predictive Maintenance Methodology: Predicting the Time-to-Failure of Machines in Industry 4.0. In: IEEE 19TH International Conference ON Industrial Informatics (INDIN), 2021., 2021, Palma de Mallorca, Spain. **Anais...** IEEE, 2021. p. 1–8.

ZURITA, D. et al. Industrial process monitoring by means of recurrent neural networks and Self Organizing Maps. In: IEEE INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES AND FACTORY AUTOMATION, ETFA, 2016. **Anais...** [S.l.: s.n.], 2016. v. 2016-November.

ŽVIRBLIS, T. et al. Investigation of deep learning models on identification of minimum signal length for precise classification of conveyor rubber belt loads. **Advances in Mechanical Engineering**, [S.l.], v. 14, n. 6, p. 168781322211027, June 2022.



## APÊNDICE A – CORPUS

Tabela 19: Corpus

ID	Autor	Título	Evento
1	Toma et al. (2022)	A Bearing Fault Classification Framework Based on Image Encoding Techniques and a Convolutional Neural Network under Different Operating Conditions	Periódico
2	Onus et al. (2021)	A Case Study on Challenges of Applying Machine Learning for Predictive Drill Bit Sharpness Estimation	Workshop
3	Rezende et al. (2018)	A case study on the analysis of an injection moulding machine energy data sets for improving energy and production management	Conferência
4	Tchatchoua et al. (2021)	A Comparative Evaluation of Deep Learning Anomaly Detection Techniques on Semiconductor Multivariate Time Series Data	Conferência
5	Soltanali et al. (2021)	A comparative study of statistical and soft computing techniques for reliability prediction of automotive manufacturing	Periódico
6	Ribeiro et al. (2021)	A Comparison of Anomaly Detection Methods for Industrial Screw Tightening	Conferência
7	Zhang et al. (2020)	A CPPS based on GBDT for predicting failure events in milling	Periódico
8	Ding et al. (2013)	A Data Analytic Engine Towards Self-Management of Cyber-Physical Systems	Workshop
9	Mulrennan et al. (2019)	A data science approach to modelling a manufacturing facility's electrical energy profile from plant production data	Conferência
10	Subramaniyan et al. (2018)	A data-driven algorithm to predict throughput bottlenecks in a production system based on active periods of the machines	Periódico
11	Carletti et al. (2019)	A deep learning approach for anomaly detection with industrial time series data: A refrigerators manufacturing case study	Conferência
12	Li, Wang e Wang (2019)	A deep learning driven method for fault classification and degradation assessment in mechanical equipment	Periódico
13	Bampoula et al. (2021)	A Deep Learning Model for Predictive Maintenance in Cyber-Physical Production Systems Using LSTM Autoencoders	Periódico
14	Essien e Giannetti (2020)	A Deep Learning Model for Smart Manufacturing Using Convolutional LSTM Neural Network Autoencoders	Periódico
15	Villalobos, Suykens e Illarramendi (2020)	A flexible alarm prediction system for smart manufacturing scenarios following a forecaster–analyzer approach	Periódico
16	Fu, Chien e Lin (2018)	A Hybrid Forecasting Framework with Neural Network and Time-Series Method for Intermittent Demand in Semiconductor Supply Chain	Conferência
17	Van Herreweghe et al. (2020)	A Machine Learning-Based Approach for Predicting Tool Wear in Industrial Milling Processes	Conferência
18	Alexopoulos e Packianather (2017)	A monitoring and data analysis system to achieve zero-defects manufacturing in highly regulated industries	Periódico
19	Sarda et al. (2021)	A Multi-Step Anomaly Detection Strategy Based on Robust Distances for the Steel Industry	Periódico

Tabela 19: (continuação da tabela Corpus)

ID	Autor	Título	Evento
20	Cordoni et al. (2022)	A multi-modal unsupervised fault detection system based on power signals and thermal imaging via deep AutoEncoder neural network	Periódico
21	Shenfield e Howarth (2020)	A novel deep learning model for the detection and identification of rolling element-bearing faults	Periódico
22	Silva Arantes et al. (2021)	A novel unsupervised method for anomaly detection in time series based on statistical features for industrial predictive maintenance	Periódico
23	Ding, Yang e Yang (2019)	A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data	Periódico
24	Zufle et al. (2021)	A Predictive Maintenance Methodology: Predicting the Time-to-Failure of Machines in Industry 4.0	Conferência
25	Bousdekis et al. (2019)	A RAMI 4.0 View of Predictive Maintenance: Software Architecture, Platform and Case Study in Steel Industry	Workshop
26	Tedesco et al. (2021)	A Scalable Deep Learning-Based Approach for Anomaly Detection in Semiconductor Manufacturing	Conferência
27	Berges, Ramírez-Durán e Illarramendi (2021)	A Semantic Approach for Big Data Exploration in Industry 4.0	Periódico
28	Wu et al. (2018)	A Visual Analytics Approach for Equipment Condition Monitoring in Smart Factories of Process Industry	Conferência
29	Tagawa, Maskeliūnas e Damaševičius (2021)	Acoustic Anomaly Detection of Mechanical Failures in Noisy Real-Life Factory Environments	Periódico
30	Mahmood, Luo e Rehman (2022)	An accurate detection of tool wear type in drilling process by applying PCA and one-hot encoding to SSA-BLSTM model	Periódico
31	Lu, Hsu e Huang (2020)	An autoencoder gated recurrent unit for remaining useful life prediction	Periódico
32	Kiangala e Wang (2020)	An Effective Predictive Maintenance Framework for Conveyor Motors Using Dual Time-Series Imaging and Convolutional Neural Network in an Industry 4.0 Environment	Periódico
33	Yue, Ping e Lanxin (2018)	An End-to-End model based on CNN-LSTM for Industrial Fault Diagnosis and Prognosis	Conferência
34	Vicêncio et al. (2021)	An Intelligent Predictive Maintenance Approach Based on End-of-Line Test Logfiles in the Automotive Industry	Conferência
35	Abbasi, Martinez e Ahmad (2021)	An ontology model to represent aquaponics 4.0 system's knowledge	Periódico
36	Nieves Avendano et al. (2021)	Anomaly detection and event mining in cold forming manufacturing processes	Periódico
37	Kayan et al. (2021)	AnoML-IoT: An end to end re-configurable multi-protocol anomaly detection pipeline for Internet of Things	Periódico
38	Mateus et al. (2021)	Anticipating Future Behavior of an Industrial Press Using LSTM Networks	Periódico
39	Vries et al. (2016)	Application of machine learning techniques to predict anomalies in water supply networks	Periódico
40	Wu, Ding e Huang (2020)	Approach for fault prognosis using recurrent neural network	Periódico

Tabela 19: (continuação da tabela Corpus)

ID	Autor	Título	Evento
41	Luo, Liu e Qi (2019)	Big data analytics – enabled cyber-physical system: model and applications	Periódico
42	Ma et al. (2020)	Big data driven predictive production planning for energy-intensive manufacturing industries	Periódico
43	Rousopoulou et al. (2022)	Cognitive analytics platform with AI solutions for anomaly detection	Periódico
44	Hoppenstedt et al. (2019)	CONSENSORS: A Neural Network Framework for Sensor Data Analysis	Workshop
45	Chen, Lee e Chiu (2018)	Construct an Intelligent Yield Alert and Diagnostic Analysis System via Data Analysis: Empirical Study of a Semiconductor Foundry	Conferência
46	Bagozi, Bianchini e Antonellis (2021)	Context-Based Resilience in Cyber-Physical Production System	Periódico
47	Tripathi e Baruah (2020)	Contextual Anomaly Detection in Time Series Using Dynamic Bayesian Network	Periódico
48	Park et al. (2019)	Cyber Physical Energy System for Saving Energy of the Dyeing Process with Industrial Internet of Things and Manufacturing Big Data	Periódico
49	Rousopoulou et al. (2019)	Data Analytics Towards Predictive Maintenance for Industrial Ovens	Workshop
50	Kim e Lee (2022)	Data-analytics-based factory operation strategies for die-casting quality enhancement	Periódico
51	Varela et al. (2019)	Decision support visualization approach in textile manufacturing a case study from operational control in textile industry	Periódico
52	Azamfar, Li e Lee (2020)	Deep Learning-Based Domain Adaptation Method for Fault Diagnosis in Semiconductor Manufacturing	Periódico
53	Bibaud-Alves, Thomas e El Haouzi (2019)	Demand forecasting using artificial neuronal networks and time series: Application to a French furniture manufacturer case study	Conferência
54	Wang et al. (2022a)	Design of PM2.5 monitoring and forecasting system for opencast coal mine road based on internet of things and ARIMA Mode	Periódico
55	Majdani, Petrovski e Doolan (2016)	Designing a Context-Aware Cyber Physical System for Smart Conditional Monitoring of Platform Equipment	Conferência
56	Wang et al. (2022b)	Detecting anomalies in time series data from a manufacturing system using recurrent neural networks	Periódico
57	El Wahabi et al. (2020)	Detection and Control System for Automotive Products Applications by Artificial Vision Using Deep Learning	Periódico
58	Garmaroodi et al. (2021)	Detection of Anomalies in Industrial IoT Systems by Data Mining: Study of CHRIST Osmotron Water Purification System	Periódico
59	Eze, Halse e Ajmal (2021)	Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm	Periódico
60	Akin et al. (2021)	Enabling Big Data Analytics at Manufacturing Fields of Farplas Automotive	Conferência
61	Huang, Zanni-Merk e Crémilleux (2019)	Enhancing deep learning with semantics: An application to manufacturing time series analysis	Conferência

Tabela 19: (continuação da tabela Corpus)

ID	Autor	Título	Evento
62	Naskos et al. (2020)	Event-Based Predictive Maintenance on Top of Sensor Data in a Real Industry 4.0 Case Study	Conferência
63	Kurpanik et al. (2018)	EYE: Big data system supporting preventive and predictive maintenance of robotic production lines	Periódico
64	Jang e Cho (2021)	Feature Space Transformation for Fault Diagnosis of Rotating Machinery under Different Working Conditions	Periódico
65	Lima et al. (2021)	HealthMon: An approach for monitoring machines degradation using time-series decomposition, clustering, and metaheuristics	Periódico
66	Zurita et al. (2016)	Industrial process monitoring by means of recurrent neural networks and Self Organizing Maps	Conferência
67	Mohsen, Elkaseer e Scholz (2021)	Industry 4.0-Oriented Deep Learning Models for Human Activity Recognition	Periódico
68	Mosavi et al. (2022)	Intelligent energy management using data mining techniques at Bosch Car Multimedia Portugal facilities	Periódico
69	Žvirblis et al. (2022)	Investigation of deep learning models on identification of minimum signal length for precise classification of conveyor rubber belt loads	Periódico
70	Ghosh e Banerjee (2019)	IoT-based seismic hazard detection in coal mines using grey systems theory	Conferência
71	Yasaei, Hernandez e Faruque (2020)	IoT-CAD: context-aware adaptive anomaly detection in IoT systems through sensor association	Conferência
72	Apiletti et al. (2018)	iSTEP, an Integrated Self-Tuning Engine for Predictive Maintenance in Industry 4.0	Conferência
73	Kancharla et al. (2022)	Latent Dimensions of Auto-Encoder as Robust Features for Inter-Conditional Bearing Fault Diagnosis	Periódico
74	Orrù et al. (2020)	Machine learning approach using MLP and SVM algorithms for the fault prediction of a centrifugal pump in the oil and gas industry	Periódico
75	Min et al. (2019)	Machine Learning based Digital Twin Framework for Production Optimization in Petrochemical Industry	Periódico
76	Kovács e Kő (2019)	Machine Learning Based Monitoring of the Pneumatic Actuators' Behavior Through Signal Processing Using Real-World Data Set	Conferência
77	Lepeniotti et al. (2020)	Machine Learning for Predictive and Prescriptive Analytics of Operational Data in Smart Manufacturing	Workshop
78	Kovacs e Ko (2020)	Monitoring Pneumatic Actuators' Behavior Using Real-World Data Set	Periódico
79	Canizo et al. (2019)	Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study	Periódico
80	Hsu e Liu (2021)	Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing	Periódico
81	Khodabakhsh et al. (2018)	Multivariate Sensor Data Analysis for Oil Refineries and Multi-mode Identification of System Behavior in Real-time	Periódico
82	Song e Baek (2020)	New anomaly detection in semiconductor manufacturing process using oversampling method	Conferência

Tabela 19: (continuação da tabela Corpus)

ID	Autor	Título	Evento
83	Ooi et al. (2019)	Operation status tracking for legacy manufacturing systems via vibration analysis	Conferência
84	Syafrudin et al. (2018)	Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing	Periódico
85	Sun et al. (2020)	PlanningVis: A Visual Analytics Approach to Production Planning in Smart Factories	Periódico
86	Kim, Kim e Ahn (2021)	Predicting workers' inattentiveness to struck-by hazards by monitoring biosignals during a construction task: A virtual reality experiment	Periódico
87	Wahid, Breslin e Intizar (2022)	Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework	Periódico
88	Sonthited, Koolpiruk e Songkasiri (2019)	Prediction of production performance for tapioca industry using LSTM neural network	Conferência
89	Ayvaz e Alpay (2021)	Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time	Periódico
90	Quatrini et al. (2020)	Predictive model for the degradation state of a hydraulic system with dimensionality reduction	Conferência
91	Brzychczy e Trzcionkowska (2019)	Process-Oriented Approach for Analysis of Sensor Data from Longwall Monitoring System	Conferência
92	Zhou et al. (2021b)	SemML: Facilitating development of ML models for condition monitoring with semantics	Periódico
93	Baquerizo et al. (2022)	Siamese Neural Networks for Damage Detection and Diagnosis of Jacket-Type Offshore Wind Turbine Platforms	Periódico
94	Becher et al. (2022)	Situated Visual Analysis and Live Monitoring for Manufacturing	Periódico
95	Sundaram e Zeid (2021)	Smart Prognostics and Health Management (SPHM) in Smart Manufacturing: An Interoperable Framework	Periódico
96	Zhan et al. (2022)	Stgat-Mad : Spatial-Temporal Graph Attention Network For Multivariate Time Series Anomaly Detection	Conferência
97	Shrivastava et al. (2019)	ThunderML: A Toolkit for Enabling AI/ML Models on Cloud for Industry 4.0	Conferência
98	Chen et al. (2020)	Time Series Data for Equipment Reliability Analysis with Deep Learning	Periódico
99	Jiang, Lee e Zeng (2020)	Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life	Periódico
100	Rehse, Mehdiyev e Fettke (2019)	Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory	Periódico
101	Zhou et al. (2021a)	Towards Ontology Reshaping for KG Generation with User-in-the-Loop: Applied to Bosch Welding	Conferência
102	Graß, Beecks e Soto (2019)	Unsupervised Anomaly Detection in Production Lines	Conferência
103	Stähl et al. (2019)	Using recurrent neural networks with attention for detecting problematic slab shapes in steel rolling	Periódico

## APÊNDICE B – MÉTODOS

Tabela 20: Métodos

Método	Nome
1D-CNN-LSTM	One-dimensional convolutional neural network long short-term memory
1NN-DTW	One-nearest-neighbor with dynamic time warping
2-DConvLSTMAE	Deep convolutional LSTM stacked autoencoder for univariate, multistep machine speed forecasting
AAE	Attentional autoencoder
AB	AdaBoost
AE	Autoencoder
AE-GRU	Autoencoder gated recurrent unit
AEWGAN	Autoencoder Wasserstein generative adversarial networks
AFDC-CNN	Attention fault detection and classification convolutional neural network
AGRU	Attention-based gated recurrent unit
AHC	Agglomerative hierarchical clustering
ALSTM	Attention-based long short-term memory
AML	AutoML
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
AnoGAN	Anomaly detection generative adversarial networks
ANOVA	Analysis of variance
AOD	Anomaly and outlier detector
AR	Augmented reality
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BDA	Balanced distribution adaptation
BGM	Bayesian Gaussian mixture
BGRU	Bidirectional gated recurrent unit
BINN	Bayesianly interpretable neural network
BLSTM	Bidirectional long short-term memory
BNN	Bayesian neural network
BPNN	Back propagation neural network
BR	Bayesian ridge/regularization
BRNN	Bidirectional recurrent neural network
CART	Classification and regression tree
CDSAE-AD	Convolutional denoising sparse autoencoders anomaly detection
CDT	Complex decision tree
CMD	Central mean discrepancy

Tabela 20: (continuação da tabela Métodos)

<b>Método</b>	<b>Nome</b>
CNN	Convolutional neural network
CNN-LSTM	Convolutional neural network–long short-term memory
CNN-MMD	Convolutional neural network maximum mean discrepancy
CRISP-DM	Cross-industry standard process for data mining
CSAE-AD	Convolutional sparse autoencoders anomaly detection
CST	Combinatorial search of two
CxDBNet	Contextual dynamic Bayesian network
DADA	Discriminative adversarial domain adaptation
DANN	Domain-adversarial training of neural networks
DBN	Deep belief network
DBNet	Dynamic Bayesian network
DBSCAN	Density-based spatial clustering of applications with noise
DCTLN	Deep convolutional transfer learning network
DeepLSTM	Deep long short-term memory
DES	Double exponential smoothing method
DF	Decision forest
DNN	Deep neural network
DPCA	Dynamic principal component analysis
DT	Decision tree
DWT	Discrete wavelet transformation
EEMD-DL-LSTM	Ensemble empirical mode decomposition and deep learning long short-term memory
EncDec-AD	Encoder–decoder anomaly detection
FDC-CNN	Fault detection and classification convolutional neural network
FFNN	Feed-forward neural network
FFT	Fast Fourier transformation
FFT-MLP	Fast Fourier transform based multilayer perceptron
FFT-SVM	Fast Fourier transform based support vector machines
GA	Genetic algorithm
GAF	Gramian angular field
GBDT	Gradient boosting decision tree
GBM	Gradient boosting machine
GBT	Gradient-boosted tree
GDN	Graph deviation network
GEC	Gross error classification
GFK	Geodesic flow kernel
GHMM	Gaussian hidden Markov models
GLM	Generalized linear model

Tabela 20: (continuação da tabela Métodos)

<b>Método</b>	<b>Nome</b>
GMM	Gaussian mixture models
GR	Gaussian regression
GRU	Gated recurrent unit
GST	Grey systems theory
HCA	Hierarchical clustering algorithm/analysis
HDBSCAN	Hierarchical density-based spatial clustering of applications with noise
HMM	Hidden Markov models
I-Forest	Isolation forest
IDEAaS	Interactive data exploration as-a-service
iForest	Isolation forest
JDA	Joint distribution adaptation
KM	K-means
KNN	K-nearest neighbors
KNNC	K-nearest-neighbor classification
LDA	Linear discriminant analysis
LGBM	LightGBM
LMS	logMelSpectrogram
LOF	Local outlier factor
LR	Logistic regression
LRM	Linear regression model
LSTM	Long short-term memory
LSTM-AD	Long short-term memory anomaly detection
LSTM-NDT	LSTM with nonparametric dynamic thresholding
LSTM-VAE	Long short-term memory variational autoencoder
LSTMAE	LSTM-Autoencoder
MAD	Mean absolute deviation
MC-DCNN	Multichannel deep convolutional neural networks
MCOD	Streaming distance-based outlier detection algorithm
MCU	Minimum covariance determinant
MDDAN	Multiscale deep domain-adaptive network
MDIAN	Multiscale deep intraclass adaptive network
MDP	Markov decision process
Methontology	Methontology
MLCAE	Multilayer convolutional autoencoder
MLCAE-KNN	Multilayer convolutional autoencoder K-nearest neighbors
MLP	Multilayer perceptron
MORL	Multiobjective reinforcement learning



Tabela 20: (continuação da tabela Métodos)

<b>Método</b>	<b>Nome</b>
MP	Matrix profile
MTAD-GAT	Multivariate time-series anomaly detection via graph attention network
MTS-CNN	Multiple time-series convolution neural network
MV	Majority voting
NB	Naive Bayes
NHPP	Nonhomogeneous Poisson process
NLT	Neural linear transformation
NN	Neural network
OCSVM	One-class SVM
Ontology	Ontology
OntoLSTM	Ontology-based LSTM neural network
PCA	Principal component analysis
PersistenceModel	Operates on the assumption that the predicted value remains unchanged from the previous time lag
Prophet	Prophet
RBF	Radial basis function
ResNet	Residual neural networks
RF	Random forest
RMS	Root mean square
RNN	Recurrent neural network
RNN-WDCNN	Recurrent neural network with a wide first kernel and deep convolutional neural network
RSNet	Residual-squeeze net
SAX-VSM	Symbolic aggregate approximation and vector space model
SBA	Syntetos–Boylan Approximation
SDM	Seismic detection method
SF	Shapelet forests
SGB	Stochastic gradient boosting
SMOTE	Synthetic minority oversampling technique
SN	SeriesNet
SNN	Siamese neural networks
SOM	Self-organizing maps
SPIRIT	Streaming pattern discovery on multiple time series
SRDCNN	Stacked residual dilated convolutional neural network
SSA-BLSTM	Singular spectrum analysis bidirectional long short-term memory
STFT	Short-term Fourier transform
STGAT-MAD	Spatial–temporal graph attention network for multivariate time series anomaly detection
SVC	Support vector classification
SVM	Support vector machine

Tabela 20: (continuação da tabela Métodos)

<b>Método</b>	<b>Nome</b>
SVR	Support vector regression
t-SNE	t-Distributed stochastic neighbor embedding
TCA	Transfer component analysis
TCN	Temporal convolutional network
Tikhonov	Tikhonov
TNN	Transformer neural network
TSMC-CNN	Time-series multiple-channel convolutional neural network
TSO	Tournament search optimization
UKF	Unscented Kalman filter
USAD	Unsupervised anomaly detection for multivariate time series
V A	Visual analytics
VGG	Visual geometry group
VQS	Visual query system
VR	Virtual reality
Ward	Wards method
WDCNN	Wide-first kernel and deep convolutional neural network
Weibull	Weibull Model
WGAN	Wasserstein generative adversarial networks
WN	WaveNet
WPD	Wavelet packet decomposition
WSM	Weighted sum model
XGB	XGBoost
ZO	Zero order

## APÊNDICE C – FERRAMENTAS

Tabela 21: Ferramentas

Ferramenta	Nome
AngularJS	AngularJS
AnoML-IoT	AnoML-IoT
AquaONT	AquaONT
ARHoloLens	AR HoloLens
AZAP	Software suite
Azure	Database
AzureML	Azure Machine Learning Studio
BURLAP	Brown-UMBC Reinforcement Learning and Planning library
C#	Programming language
C++	Programming language
Cassandra	Database
ChartJS	ChartJS
Colab	Google Colaboratory Platform
CouchDB	Amazon CouchDB
D3JS	D3JS
Direct3D	Direct3D
Docker	Docker
doParallel	R library for parallel execution
Elasticsearch	Distributed RESTful search engine built for the cloud
ERP	Enterprise resource planning system
ExtruOnt	ExtruOnt
EYE	Data storage and analysis system
fastcluster	R library for clustering
Flask	Flask
Platform	Big data platform
foreach	R library for parallel execution
freedom	R package freedom
Fuseki	Apache Jena Fuseki (SPARQL server)
GADPL	generic anomaly detection for production lines
GAI	Google AI Platform
GPyOpt	Python open-source library for Bayesian Optimization
Hadoop	Framework for processing of large data sets
HealthMon	HealthMon
Hermit	Hermit

Tabela 21: (continuação da tabela Ferramentas)

<b>Ferramenta</b>	<b>Nome</b>
Imblearn	Python imbalanced-learn API
InfluxDB	Database
iSTEP	Integrated self-tuning engine for predictive maintenance
JavaScript	Programming language
Jupyter	Open-source web application for Python language to create and share documents
Kafka	Streaming platform
KafkaStreams	Kafka Streams
Keras	Neural Network library for Python
Kibana	Browser-based analytics and search dashboard for Elasticsearch
Knime	Data analytics, reporting, and integration platform
kohonen	R package Kohonen self-organizing maps (KSOM)
MATLAB	Programming platform
MES	Manufacturing execution systems
MLlib	Machine learning library
MongoDB	Database
MSSQL	Microsoft SQL
MUVTIME	Desktop application designed to assist in the process of multivariate time series data visual analysis
MySQL	MySQL
Neo4j	NoSQL graph database
NiFi	System to process and distribute data
NodeJS	NodeJS
OpenCV	Open-Source Computer Vision Library
OWL	OWL
Pallet	Pallet
Pandas	Pandas
Parquet	Machine-readable columnar storage format available in the Spark+Hadoop ecosystem
PlanningVis	Visual analytics system
Protege	Protégé
PyOD	Python toolbox
Python	Programming language
PyTorch	PyTorch
PyWavelets	PyWavelets
QlikSense	QlikSense
QlikView	QlikView
R	Programming language
RAMI4.0	Reference architecture model
RDFOx	RDFOx

Tabela 21: (continuação da tabela Ferramentas)

<b>Ferramenta</b>	<b>Nome</b>
RPropMLP	Knime Node
rpud	R library for the dissimilarity matrix calculation
Ruptures	Python library for offline change point detection
SCADA	Supervisory control and data acquisition
SemML	SemML
SKLEARN	Scikit-learn: Machine Learning in Python
Spark	Unified analytics engine
SPARQL	SPARQL
SPHM	Smart prognostics and health management
SQL	Query language for relational databases
SSDT	SQL Server Data Tools
SSIS	SQL Server Integration Services
Stardog	Stardog
Storm	Real-time computation system
SWRL	Semantic Web Rule Language
t-SNE	T-distributed stochastic neighbor embedding
Tensorflow	Machine learning platform
Theano	Python library for mathematical expressions
ThunderML	Machine learning toolkit
UPTIME	Unified predictive maintenance platform
Virtuoso	Virtuoso
Weka	Graphical user interface for machine learning
XGBoost	R package XGBoost
Zeppelin	Web-based notebook that enables data-driven, interactive data analytics and collaborative documents

## APÊNDICE D – SEGMENTOS INDUSTRIAIS

Tabela 22: Segmentos Industriais

ID	Nome
1	<i>Agriculture; plantations; other rural sectors</i>
2	<i>Basic Metal Production</i>
3	<i>Chemical industries</i>
4	<i>Commerce</i>
5	<i>Construction</i>
6	<i>Education</i>
7	<i>Financial services; professional services</i>
8	<i>Food; drink; tobacco</i>
9	<i>Forestry; wood; pulp and paper</i>
10	<i>Health services</i>
11	<i>Hotels; tourism; catering</i>
12	<i>Mining (coal; other mining)</i>
13	<i>Mechanical and electrical engineering</i>
14	<i>Media; culture; graphical</i>
15	<i>Oil and gas production; oil refining</i>
16	<i>Postal and telecommunications services</i>
17	<i>Public service</i>
18	<i>Shipping; ports; fisheries; inland waterways</i>
19	<i>Textiles; clothing; leather; footwear</i>
20	<i>Transport (including civil aviation; railways; road transport)</i>
21	<i>Transport equipment manufacturing</i>
22	<i>Utilities (water; gas; electricity)</i>

## APÊNDICE E – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Você está sendo convidado(a) a participar, como voluntário(a), no projeto de pesquisa intitulado "ODISSEU: Um Modelo para Serviços Inteligentes na Indústria 4.0 baseado em Análise de Históricos de Contextos". O pesquisador responsável pelo projeto é o doutorando Helder Moreira Arruda, do Programa de Pós-Graduação em Computação Aplicada (PPGCA) da Universidade do Vale do Rio dos Sinos (UNISINOS), orientado pelo Prof. Dr. Jorge Luis Victória Barbosa e coorientado pelo Prof. Dr. Rafael Kunst. O modelo ODISSEU tem como objetivos específicos indicar a localização de pessoas em uma indústria e medir o bem-estar subjetivo dessas pessoas.

A sua participação nesta pesquisa consistirá em ter sua localização rastreada dentro das instalações da HT Micron Semicondutores, seus sinais vitais coletados e responder a questionários de autorrelato referentes ao seu bem-estar subjetivo. Os dados coletados serão utilizados apenas para fins de pesquisa, de forma anônima. Esses dados serão guardados pelo pesquisador responsável em local seguro e por um período de 5 anos. Cabe aqui salientar que a sua identidade será preservada, pois não serão divulgados nomes, informações ou imagens que possam identificá-lo(a).

Você receberá, via e-mail, os resultados da pesquisa, tão logo forem obtidos ou publicados em periódicos ou eventos científicos.

Os riscos pela participação na pesquisa são apenas o desconforto de usar o beacon móvel, a pulseira E4 e o tempo disposto para responder o formulário de autorrelato. Os pesquisadores farão o possível para minimizar esses inconvenientes, mas você pode sempre optar, a qualquer momento, a retirar seu consentimento se perceber um desconforto maior. Sua participação não lhe trará nenhum benefício pessoal direto, além da colaboração para que sejam desenvolvidos novos conhecimentos científicos acerca do bem-estar no ambiente de trabalho.

Após ser esclarecido(a) sobre as informações acima, no caso de aceitar fazer parte desta pesquisa, por favor assinale o campo abaixo "Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar". Caso não deseje participar da pesquisa, assinale a opção "Não desejo participar dessa pesquisa". Preencha também seu nome completo e seu endereço de e-mail.

A qualquer momento, você pode retirar o consentimento. Sua recusa não trará nenhum prejuízo na sua relação com o pesquisador, com a UNISINOS e com a HT Micron Semicondutores.

Este termo de consentimento será impresso em duas vias, sendo que uma será arquivada pelo pesquisador responsável, e a outra será fornecida ao participante.

Em caso de dúvida, você pode procurar o pesquisador Helder Moreira Arruda, telefone, (51) 99212-1133 ou pelo e-mail [arruda.helder@gmail.com](mailto:arruda.helder@gmail.com)

( ) Declaro que entendi os objetivos, riscos e benefícios da minha participação na pesquisa e concordo em participar.

( ) Não desejo participar dessa pesquisa.

São Leopoldo, \_\_\_\_\_ de \_\_\_\_\_ de 2023

\_\_\_\_\_  
Helder Moreira Arruda (responsável)

\_\_\_\_\_  
Nome completo do(a) participante

\_\_\_\_\_  
Assinatura

\_\_\_\_\_  
Endereço de e-mail do(a) participante

## APÊNDICE F – FORMULÁRIO DE APLICAÇÃO DA ESCALA PANAS

Esta escala consiste de um número de palavras que descrevem diferentes sentimentos e emoções. Leia cada item e depois selecione a resposta adequada. Indique até que ponto você tem se sentido desta forma no dia de hoje.

Ativo	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Alerta	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Atento	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Determinado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Entusiasmado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Empolgado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Inspirado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Interessado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Forte	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Com medo	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Envergonhado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Aflito	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Culpado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Hostil	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Irritável	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Inquieto	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Nervoso	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Apavorado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente
Chateado	<input type="checkbox"/> Nem um pouco	<input type="checkbox"/> Um pouco	<input type="checkbox"/> Moderadamente	<input type="checkbox"/> Bastante	<input type="checkbox"/> Extremamente