

UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA
PROJETO DE INICIAÇÃO CIENTÍFICA EM ENGENHARIA MECATRÔNICA

ALEXANDRE MASCARENHAS

**Sistematização e organização de dados para aplicações em Gêmeo Digital e
Inteligência Artificial em ativos de transporte público sobre trilhos**

São Paulo

2022

ALEXANDRE MASCARENHAS

**Sistematização e organização de dados para aplicações em Gêmeo Digital e
Inteligência Artificial em ativos de transporte público sobre trilhos**

Projeto de Iniciação Científica apresentado
na Escola Politécnica da Universidade de São
Paulo

Área de concentração: Controle e Au-
tomação, Ciência de dados

Orientador: Prof. Dr. Paulo Eigi Miyagi

Coorientador: Prof. Dr. Fabrício Junqueira

São Paulo

2022

Resumo

Hoje os sistemas de transporte público sobre trilhos tem real importância nas grandes cidades, seja pela rapidez com que ele transporta um grande número de pessoas de um lado ao outro da cidade, pelo custo, quando comparado com outros meios de transporte, como ônibus ou carros particulares, ou ainda pela forma de energia da qual se utiliza, sendo a energia elétrica, uma energia renovável e com menores impactos ambientais quando comparada a queima de combustíveis fósseis. Os sistemas de transporte público sobre trilhos apresentam grande complexidade, desde a concepção de seu projeto, até a sua manutenção, uma vez que se trata de sistemas com um grande número de ativos que se inter relacionam ao longo de toda a malha ferroviária. Assim, surge uma necessidade de haver uma abordagem sistemática, para a grande quantidade de dados que esses sistemas podem fornecer, a fim de tornar possível uma análise minuciosa e com resultados práticos na gestão da manutenção destes ativos. Neste contexto, surge o conceito de Gêmeo Digital, uma representação virtual de um ativo real que, associado com técnicas de Big Data, permite que sejam simulados diversos cenários para este ativo, tanto na fase de projeto de sistemas e produtos, quanto na fase de operação, com o objetivo de minimizar possíveis erros de projeto e prolongar o ciclo de vida do ativo quando já em operação. Para que sejam possíveis simulações e previsões do comportamento de um ativo, é necessário que todos os dados referentes ao funcionamento, e consequentemente de seu Gêmeo Digital, estejam devidamente organizados, sendo estes dados referentes tanto ao projeto, quanto dados de monitoramento em tempo real provenientes de sensores instalados nos ativos. O objetivo deste trabalho é, portanto, estudar formas de sistematizar o projeto de sistemas de transporte público sobre trilhos e a organização de dados provenientes destes, com foco na utilização dos mesmos em aplicações de Gêmeo Digital e Inteligência Artificial. Como exemplo de aplicação serão usados os dados provenientes do sistema de Portas de Plataforma utilizado pelo Metrô de São Paulo, por se tratar de um sistema de grande importância na segurança dos usuários e ser um sistema que hoje fornece dados em tempo real, referentes ao seu funcionamento.

Palavras-chaves: Gêmeo Digital, Inteligência Artificial, Transporte público sobre trilhos, Ferrovia, Metrô, Big Data.

Abstract

Today, rail public transport systems are of real importance in large cities, be it for the speed with which they can transport a large number of people from one side of the city to the other, be it for the cost, when compared to other means of transport, such as buses or private cars, or be it for due to the form of energy used, being electric energy a renewable energy with less environmental impacts when compared to the burning of fossil fuels. Rail public transport systems represent systems of certain complexity, from the conception of their project to their maintenance, since they are systems with a large number of assets that are interrelated throughout the entire railway network. Thus, there is a need for a systematic approach to the large amount of data that these systems can provide, in order to make a thorough analysis possible and with practical results in managing the maintenance of these assets. In this context, the concept of Digital Twin emerges, a virtual representation of a real asset, which, associated with Artificial Intelligence and Big Data techniques, allows different scenarios to be simulated for this asset, both in the design phase of systems and products, and in the operational phase, with the objective of minimizing possible design errors and prolonging the asset's life cycle when already in operation. In order for these simulations and predictions of the behavior of an asset to be possible, it is necessary that all data relating to the operation of the asset, and consequently its Digital Twin, are properly organized, with these data referring to both the project and monitoring data in real-time coming from sensors installed in the assets. The objective of this work is, therefore, to study ways to systematize the design of public transport systems on rails and the organization of data from these, with a focus on their use in Digital Twin and Artificial Intelligence applications. As a practical example of application, data from the Platform Screen Doors system used by the Sao Paulo Metro will be used, as it is a system of great importance in the safety of users and is a system that today provides real-time data, referring to its functioning.

Keywords: Digital Twin, Artificial Intelligence, Rail public Transport, Railway, Subway, Big Data.

Lista de figuras

Figura 1 – Etapas do processo de preparação de dados	19
Figura 2 – Carro 5 do trem 10 da frota E do Metrô de São Paulo	24
Figura 3 – Códigos de velocidade ao longo da via (os códigos representam a velocidade que o trem deve estar no bloco em km/h)	26
Figura 4 – Blocos dinâmicos no CBTC	27
Figura 5 – Módulo PSD	29
Figura 6 – Curva de Corrente do Motor de um módulo PSD	30
Figura 7 – Diagrama do Banco de Dados	32

Lista de abreviaturas e siglas

AI	<i>Artificial Intelligence</i>
ATM	<i>Automatic Train Control</i>
ATO	<i>Automatic Train Operation</i>
CBTC	<i>Communication-Based Train Control</i>
CA	Corrente Alternada
CC	Corrente Contínua
CBTC	<i>Communication-Based Train Control</i>
DT	<i>Digital Twin</i>
PSD	<i>Platform Screen Door</i>
PWM	<i>Pulse Width Modulation</i>
TI	Tecnologia da Informação

Sumário

1	Introdução	8
1.1	<i>Contextualização da Indústria 4.0</i>	8
1.1.1	Cyber-Physical Systems (CPS)	9
1.1.2	Internet of Things (IoT)	9
1.1.3	Internet of Services (IoS)	10
1.2	<i>Inteligência Artificial</i>	11
1.3	<i>Digital Twin (DT)</i>	13
1.4	<i>Objetivo</i>	14
2	Revisão bibliográfica	15
2.1	<i>Inteligência Artificial Aplicada a I4.0</i>	15
2.1.1	Big Data	15
2.1.2	Banco de dados	19
3	Proposta	21
4	Exemplo de aplicação	22
4.1	<i>Composição básica dos sistemas</i>	23
4.1.1	Trem	23
4.1.2	ATC (<i>Automatic Train Control</i>) e ATO (<i>Automatic Train Operation</i>)	24
4.1.3	CBTC (<i>Communication-Based Train Control</i>)	25
4.1.4	Sistema PSD (<i>Platform Screen Door</i>)	28
4.1.5	Motor Elétrico da PSD	28
4.2	<i>Tratamento dos dados</i>	30
4.2.1	Aquisição dos dados	31
4.2.2	Pré-processamento	31
4.2.3	Armazenamento	32
5	Resultados	35
6	Conclusões	36

Referências ¹	37
------------------------------------	----

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

1 Introdução

Como uma consequência das revoluções que a precederam, a quarta revolução industrial tem grande relevância na sociedade hoje, e abrange um grande leque de tecnologias e métodos que permitem os avanços tecnológicos da atualidade. A seguir serão introduzidos os principais termos dos quais este trabalho se refere e qual o seu papel no que é chamado de Indústria 4.0.

Nos capítulos que se seguem, é feita uma revisão bibliográfica sobre Inteligência Artificial aplicada na I4.0 focando principalmente na etapa de pré-processamento de dados e em seguida, é demonstrado alguns dos conceitos vistos no exemplo de aplicação, onde algumas das técnicas vistas será aplicada a um sistema de monitoramento que está instalado nas Portas de Plataforma de uma estação da Companhia do Metropolitano de São Paulo (Metrô - SP).

1.1 Contextualização da Indústria 4.0

Com o passar do tempo os equipamentos eletrônicos tornaram-se cada vez mais potentes e flexíveis, novos *softwares* foram desenvolvidos e os preços caíram, viabilizando, assim, o antigo sonho de integração das operações de manufatura à fornecedores, clientes e sua operação remota.

Em 2011, o governo da Alemanha lançou um projeto durante a Feira de Hannover, denominado Plataforma Indústria 4.0 (Plattform Industrie 4.0), com o objetivo de desenvolver alta tecnologia de modo a fazer com que os sistemas automatizados que controlam os equipamentos industriais pudessem se comunicar trocando, assim, informações/dados entre máquinas e seres humanos, de forma a otimizar todo o processo de produção (SACOMANO et al., 2018).

Em 2013, a Plataforma Indústria 4.0 passou a ser divulgada por associações, empresas e academias e, em 2015, foi relançada, agora como programa do governo alemão (HERMANN; PENTEK; OTTO, 2016).

A Indústria 4.0 engloba sistemas e conceitos em várias áreas de conhecimentos, como a automação industrial integrada e inteligente, cadeias produtiva e logística globalmente integradas, fornecimento de matéria-prima e energia, gestão de ativos e plantas industriais,

combinando domínios que vão das mais variadas áreas da engenharia e da tecnologia da informação (CESÁRIO, 2017). São considerados elementos base para a Indústria 4.0, sistemas ciber físicos (CPS, do inglês *Cyber-Physical Systems*), internet das coisas (IoT, do inglês *Internet of Things*) e internet de serviços (IoS, do inglês *Internet of Services*). Será feita a seguir uma breve descrição de cada um destes elementos.

1.1.1 Cyber-Physical Systems (CPS)

Um sistema ciber-físico é composto por elementos computacionais em estreita relação com o ambiente físico, com o intuito de monitorar e controlar entidades físicas em tempo real, bem como testar e simular processos físicos, a partir do ambiente virtual.

Todas as aplicações que utilizam as arquiteturas CPS são formadas por duas camadas: camadas de tecnologia operacional (física); e camada virtual, de aplicações de tecnologia da informação (cyber). No entanto, os protocolos de comunicação nessas arquiteturas podem diferir de modelos de informações usados na tecnologia da informação (GIVEHCHI et al., 2016), sendo mais próximas dos protocolos de automação.

Nos CPSs, os elementos computacionais são interligados aos elementos físicos por meio de sensores e atuadores, de forma que o monitoramento e controle do ambiente físico possa ser realizado a partir do virtual. Equipamentos “inteligentes” permitem a tomada de decisões descentralizada e cooperação com humanos em tempo real.

1.1.2 Internet of Things (IoT)

Os principais conceitos que servem de base para explorar as oportunidades e os desafios de IoT incluem cenários em que a capacidade de conectividade e computação de rede se estende a objetos, sensores e itens comuns não normalmente considerados computadores, permitindo que esses dispositivos gerem, troquem e consumam dados com uma mínima intervenção humana. Não há, no entanto, nenhuma definição única e universal. O IoT é um novo paradigma combinando aspectos e tecnologias provenientes de diferentes abordagens como comunicação, protocolo de internet e etc (PETRONI; GONÇALVES; JÚNIOR, 2017).

Essa rede de comunicação entre os dispositivos que o IoT permite é um recurso fortemente abordado atualmente na I4.0, pois muitas de suas implementações são rápidas e não exigem necessariamente um sistema ou uma arquitetura com alta complexidade. Diversos sensores e atuadores que são hoje utilizados para compor o IoT são de baixo custo e permitem que se tenha um retorno de dados/informações em um curto prazo de tempo.

O IoT é um conceito tão importante da Indústria 4.0 que muitas vezes é inclusive, associado de forma errada como sendo o próprio conceito de I4.0.

1.1.3 Internet of Services (IoS)

O termo *Internet of Services* refere-se a uma infraestrutura na qual a *internet* é o meio do qual se faz uso permitindo a prestação de diversos tipos de serviços de forma rápida e prática. Assim, a IoS prove uma base técnica e de negócio para que seja utilizada nos modelos onde existe a realização de algum serviço a ser prestado.

Segundo Reis e Gonçalves (2018), o termo IoS surge basicamente da junção de dois conceitos: Web 2.0 e SOA (do inglês *Service-oriented architecture*).

O primeiro conceito, Web 2.0 é caracterizado por quatro aspectos (REIS; GONÇALVES, 2018):

- **Interatividade:** Este ganho vem de duas tecnologias: AJAX (*Asynchronous JavaScript and XML*) que permite a comunicação e a manipulação dinâmica de dados entre um servidor e o navegador WEB;
- **Rede social:** As redes sociais surgem a partir de interesses comuns e disponibilizam as informações de cada rede de diferentes formas;
- **Tagging:** os usuários podem adicionar uma palavra-chave como *tag* a um determinado conteúdo da Web, tornando essa *tag* facilmente acessível quando pesquisada por outros usuários;
- **Web Services:** permitem que outros *softwares* façam uso das funcionalidades oferecidas pela aplicação Web, estando disponíveis não apenas para pessoas, mas também para máquinas.

O segundo conceito que forma a IoS é o SOA, que é um jeito de projetar e construir um conjunto de aplicação de TI onde componentes de aplicações e *Web Services*

disponibilizam suas funções no mesmo canal de acesso para uso mútuo. Para satisfazer esses requisitos, os serviços devem ser (REFERÊNCIA):

- **Neutro em termos de tecnologia:** Devem ser invocados através de tecnologias padronizadas de menor denominador comum que estão disponíveis para quase todos os ambientes de Tecnologia da Informação (TI). Isto implica que os mecanismos de invocação (protocolos, descrições e mecanismos de descoberta) devem estar em conformidade com normas amplamente aceites;
- **Fracamente acopladas:** Não devem exigir conhecimentos ou quaisquer estruturas ou convenções internas (contexto) do lado do cliente ou do serviço;
- **Apoio à transparência de localização:** Os serviços devem ter as suas definições e informações de localização armazenadas num repositório como a UDDI e ser acessíveis por uma variedade de clientes que possam invocar os serviços independentemente da sua localização.

1.2 *Inteligência Artificial*

Definir Inteligência Artificial não é algo trivial, uma vez que a própria definição de Inteligência não é muito bem estabelecida. Porém, quando se fala em Inteligência Artificial, ao longo do tempo ela seguiu, à princípio, quatro linhas de pensamento:

1. **Sistemas que pensam como seres humanos:** "O novo e interessante esforço para fazer os computadores pensarem... máquinas com mentes, no sentido total e literal". (HAUGELAND, 1985);
2. **Sistemas que atuam como seres humanos:** "A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas". (KURZWEIL, 1990);
3. **Sistemas que pensam racionalmente:** "O estudo das faculdades mentais pelo seu uso de modelos computacionais". (CHARNIAK; MCDERMOTT, 1985);
4. **Sistemas que atual racionalmente:** "A inteligência computacional é o estudo do projeto de agentes inteligentes". (POOLE; MACKWORTH; GOEBEL, 1998).

No geral, as linhas de pensamento 1 e 3 referem-se ao processo de pensamento e raciocínio, enquanto as 2 e 4 ao comportamento. Além disso, as linhas de pensamento 1 e

2 medem o sucesso em termos de fidelidade ao desempenho humano, enquanto na 3 e 4 medem o sucesso comparando-o a um conceito ideal que de inteligência, que se chamará de racionalidade. Um sistema é racional se “faz tudo certo”, com os dados que tem (RUSSEL; NORVIG, 2004).

A IA permite que os sistemas técnicos percebam o ambiente que os rodeia, lidem com o que percebem e resolvam problemas, agindo no sentido de alcançar um objetivo específico. O computador recebe dados (já preparados ou recolhidos através dos seus próprios sensores, por exemplo, com o uso de uma câmera), processa-os, e responde (REFERÊNCIA).

Os sistemas de IA são capazes de adaptar o seu comportamento, até certo ponto, através de uma análise dos efeitos das ações anteriores e de um trabalho autônomo. Algumas tecnologias de IA existem há mais de 50 anos, mas o melhor desenvolvimento da capacidade de processamento, a disponibilidade de quantidades elevadas de dados e novos algoritmos permitiram grandes progressos da IA nos últimos anos (ARTIFICIAL, 2021) (VER ESSA REFERÊNCIA).

Segundo Artificial (2021) a IA tem forte presença no cotidiano atual da sociedade, elencando algumas aplicações em que a IA esta fortemente relacionada mas que, possivelmente, pode passar despercebida:

- **Compras e publicidade online:** A inteligência artificial é amplamente utilizada para fornecer recomendações personalizadas às pessoas, com base, por exemplo, nas suas pesquisas e compras anteriores ou outro comportamento *online*. A IA é extremamente importante na área comercial - para otimização de produtos, planeamento de inventários, logística, etc.;
- **Pesquisa na web:** Os motores de busca aprendem com a vasta entrada de dados, providenciados pelos seus usuários, para fornecer resultados de pesquisa relevantes;
- **Assistentes pessoais digitais:** Os *smartphones* usam IA para fornecer um produto o mais relevante e personalizado possível. Os assistentes virtuais que respondem a perguntas, fornecem recomendações e ajudam a organizar rotinas diárias tornaram-se onnipresentes;
- **Traduções automáticas:** O *software* de tradução linguística, baseado em texto escrito ou falado, confia na inteligência artificial para fornecer e melhorar traduções. O mesmo se aplica a funções como a legendagem automática;

- **Casas, cidades e infraestruturas inteligentes:** Termostatos inteligentes aprendem com os comportamentos diários dos indivíduos para economizar energia, enquanto os desenvolvedores de cidades inteligentes pretendem controlar o tráfego para melhorar a conectividade e reduzir os engarrafamentos;
- **Carros:** Enquanto os veículos autônomos ainda não são uma realidade, os carros já usam funções de segurança alimentadas por IA como a navegação;
- **Cibersegurança:** Os sistemas de IA podem ajudar a reconhecer e combater os ataques cibernéticos e outras ameaças cibernéticas com base na introdução contínua de dados, no reconhecimento de padrões e no retrocesso dos ataques;
- **Inteligência artificial contra a COVID-19:** No caso da COVID-19, a IA tem sido usada no sistema de controlo térmico em aeroportos e em outros lugares. Na área da medicina, a IA pode ajudar a reconhecer a infeção por meio da tomografia dos pulmões. Também já tem sido usada para fornecer dados para localizar a propagação do vírus;
- **Combate à desinformação:** Certas aplicações de IA podem detetar notícias falsas e desinformação por meio do controle das informações nas redes sociais, da procura de palavras sensacionalistas ou alarmantes e da identificação das fontes fidedignas.

Com estes exemplos é possível ter uma ideia do grande leque de aplicações e tarefas que a IA pode ajudar a tornar mais eficiente ou até mesmo possibilitar a execução de atividades que não seriam possíveis anteriormente ao surgimento da IA.

1.3 *Digital Twin (DT)*

Segundo Errandonea, Beltrán e Arrizabalaga (2020) o conceito de *Digital Twin* (DT) tem evoluído desde sua primeira aparição em 2002. Dada a complexidade do conceito, uma variedade de definições pode ser encontrada na literatura. Por exemplo, a definição varia bastante (AUTIOSALO et al., 2019; KRITZINGER et al., 2018; TAO et al., 2018), mesmo que as vezes seja incorreta.

A primeira definição do conceito de DT foi feita em 2012 por (ARRUMAR REFERÊNCIA)Grievés. Anos depois, ele definiu da seguinte maneira: "Produto físico no espaço real, produto virtual no espaço virtual e a ligação de dados e informações que ligam os dois espaços.". Grievés salientou que se referia a um conjunto de informações que descreve

completamente um ativo, desde a sua geometria mais geral, até ao comportamento mais concreto (GRIEVES; VICKERS, 2016).

Rosen apresenta o conceito como dois espaços idênticos, físico e virtual, o que permite que o espelhamento entre eles analise as condições que ocorrem em todas as fases do ciclo de vida do objecto (ROSEN et al., 2015). Pouco tempo depois, Boschert e Rosen (2016) detalham que um DT cobre toda a informação física e funcional que pode ser útil a partir de um componente, produto ou sistema. Ambos os autores concordam que os DT não são apenas dados, mas também algoritmos que descrevem o comportamento e vêm a decidir sobre ações em produção.

1.4 Objetivo

O objetivo deste trabalho é apresentar alguns conceitos relacionados com a I4.0, tendo um enfoque maior no conceito de inteligência artificial aplicada a manutenção de ativos. O principal será mostrar algumas técnicas de pré-processamento de dados e posterior armazenamento dos mesmos, deixando-os prontos para que seja possível a utilização tanto de técnicas de IA como insumo para a criação de um DT do ativo. Será demonstrado em um exemplo de aplicação, como se dá a utilização, por exemplo, de métodos de limpeza de dados, que é uma fase muito importante na preparação dos dados. O exemplo de aplicação consiste em utilizar dados que são provenientes do sistema de monitoramento das Portas de Plataforma da estação Vila Matilde da Linha 3 - Vermelha do Metrô de São Paulo. Por fim, será proposto um modelo de banco de dados, onde os dados já pré-processados serão armazenados.

2 Revisão bibliográfica

ADICIONAR ALGUM TEXTO AQUI POIS ESTA DESELEGANTE.

2.1 *Inteligência Artificial Aplicada a I4.0*

Como visto anteriormente, Inteligência Artificial não se trata de uma técnica específica ou algum algoritmo em particular. Na verdade, o conceito de IA é muito amplo e abrange diversos campos, técnicas, métodos e abordagens.

Big Data é focado no gerenciamento de grandes quantidades de dados de uma variedade de fontes. Os grandes dados entram em jogo quando o volume de dados é muito grande para que as práticas tradicionais de gerenciamento de dados sejam eficazes. As empresas há muito tempo coletam grandes quantidades de informações sobre consumidores, preços, transações e segurança de produtos, mas eventualmente o volume de dados coletados provou ser muito grande para que os seres humanos possam analisar manualmente (DATA, 2020).

Neste trabalho, as técnicas de pré-processamento de dados, assim como os métodos de armazenamentos, fazem parte do campo de *big data* e portanto, algumas destas técnicas serão vistas a seguir neste capítulo.

2.1.1 Big Data

Em sistemas de *Big Data* é razoável esperar que uma certa quantidade de dados estejam em desacordo com o que se espera, uma vez que já era conhecido a priori os possíveis valores (*range*) que este dado poderia assumir. Segundo Castro e Ferrari (2016) podem ocorrer basicamente três tipos de problemas com os dados:

- **Incompletude:** a incompletude de uma base de dados pode acontecer por exemplo, na ausência de valor de um dado atributo, na ausência deste próprio, ou ainda na ausência do objeto como um todo ao qual o atributo está atrelado;
- **Inconsistência:** ocorre quando há diferentes e conflitantes versões do mesmo dado aparecendo em locais variados. Por exemplo, a idade de uma pessoa que está pedindo para aumentar o limite do cartão de crédito ser menor que do que 15 anos;

- **Ruído:** ruído pode ter diversos significados, a depender do contexto. Entretanto, em tratamento e análise de dados, o conceito de ruído assemelha-se ao utilizado em estatística (variações inexplicáveis em uma amostra) e processamento de sinais (variações indesejadas e normalmente inexplicáveis em um sinal).

Conhecer e preparar de forma adequada os dados para análise é a etapa de chamada de pré-processamento de dados, e é nesta etapa que este trabalho, bem como o exemplo de aplicação é baseado.

Para melhor entender como esses dados brutos podem chegar até a fase de preparação, será visto algumas propriedades dos dados. Com relação a estrutura dos dados, eles podem ser:

- **Estruturados:** quando a base de dados possui campos fixos e estruturados - por exemplo, uma tabela, um arquivo .csv, uma planilha ou um banco de dados relacional (CASTRO; FERRARI, 2016);
- **Semiestruturados:** o dado semiestruturado não possui a estrutura completa de um modelo de dados, mas também não é totalmente desestruturado, utilizam geralmente marcadores (por exemplo, *tags*) para identificar certos elementos dos dados. Exemplo conhecido de dados semiestruturados são arquivos XML;
- **Não estruturados:** não possui nenhum modelo de dados, ou seja, não está organizado de uma maneira predefinida. Exemplo de dados não estruturados são imagens, textos livres, arquivos PDF, entre outros. Esses são os dados que costumam ser de difícil indexação, acesso e análise (CASTRO; FERRARI, 2016).

Neste trabalho será dado um enfoque aos dados estruturados. Neste caso cada linha corresponde a um objeto ou entidade e cada coluna a um atributo, que são propriedades (características) que identificam os objetos. Os atributos podem ser simples, composto, multivalorado ou determinante (LUIS, 2020):

- **Atributo Simples:** Não possui qualquer característica especial. A maioria dos atributos serão simples. Quando um atributo não é composto, recebe um valor único como nome, por exemplo e não é um atributo chave, então ele será atributo simples. A maioria dos atributos são considerados simples.

Em uma entidade cliente, por exemplo, podem ser considerados como atributos simples: nome, sexo, data de nascimento, dentre outros;

- **Atributo Composto:** O seu conteúdo é formado por vários itens menores. Exemplo: Endereço. Seu conteúdo poderá ser dividido em vários outros atributos, como: Rua, Número, Complemento, Bairro, Cep e Cidade. Este tipo de atributo é chamado de atributo composto. Será visto mais de sua aplicação no subcapítulo sobre normalização de dados.

É importante considerar que na aplicação do banco de dados um atributo composto geralmente é desmembrado, por exemplo, para o caso do endereço, pode-se substituí-lo por vários atributos simples, como: Rua, número, complemento, bairro, cidade e cep. Conceitualmente é aceito o endereço como um único atributo, mas na prática geralmente é feito este desmembramento para permitir a organização dos dados inseridos e facilitar a busca e indexação dos mesmos;

- **Atributo Multivalorado:** O seu conteúdo é formado por mais de um valor. Exemplo: Telefone. Uma pessoa poderá ter mais de um número de telefone. É indicado colocando-se um asterisco precedendo o nome do atributo. Os atributos multivalorados serão tratados com mais detalhes na normalização de dados.

Este tipo de atributo é aceito conceitualmente, mas ele pode ser um problema no banco de dados, portanto, há duas principais possibilidades no tratamento do mesmo. A primeira é mantê-lo como multivalorado e permitir que mais de um dado seja inserido no mesmo campo, como por exemplo: dois números de telefone. A segunda alternativa é aplicar o processo de normalização de dados e o transformar em uma entidade à parte ou uma tabela no banco de dados e a relacionar com a tabela principal.

A primeira alternativa é mais simples, mas existe o problema da consulta de dados, caso se faça necessário realizar uma consulta pelo número de um dos telefones apenas. A segunda é mais trabalhosa, porém é mais eficaz;

- **Atributo Determinante:** Identifica de forma única uma entidade, ou seja, não pode haver dados repetidos.

É indicado sublinhando-se o nome do atributo. Exemplo: CNPJ, CPF, Código do fornecedor, Número da matrícula, etc. Os atributos determinantes serão as chaves primárias no banco de dados e seu uso tem implicações na normalização de dados. Deve ser considerado que toda tabela no banco de dados precisa ter um atributo determinante, que também pode ser chamada de chave primária. Desta forma, se a entidade não oferecer por padrão uma sugestão de atributo determinante, o mesmo

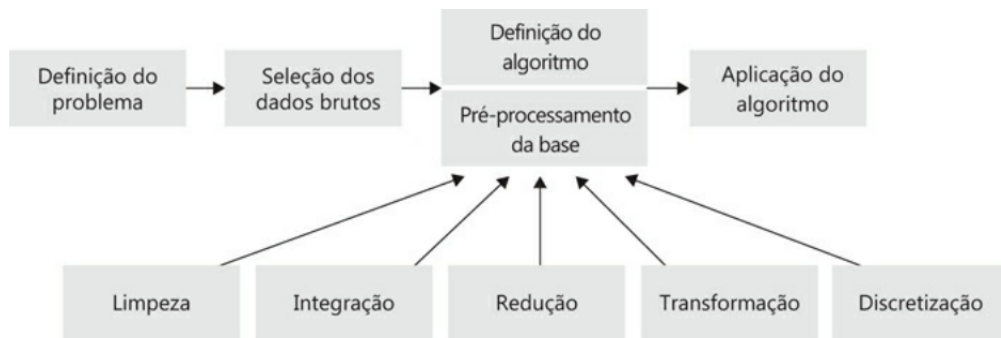
deve ser criado. Este é um princípio bastante básico da análise e modelagem de dados.

O pré-processamento, que é também conhecido como a etapa de preparação de dados, é essencial para que seja possível obter os melhores resultados dos dados. É nesta etapa em que se manipula e transforma os dados brutos de forma que as informações e os conhecimentos neles contido possam ser obtidos corretamente. Segundo Castro e Ferrari (2016) a melhor maneira de pré-processar os dados depende de três fatores centrais: os problemas (incompletude, inconsistência e ruído) existentes na base bruta; quais respostas pretendem-se obter das bases, ou seja, qual problema deve ser resolvido; e como operam as técnicas de mineração de dados que serão empregadas. Esses três fatores quase sempre estão inter-relacionados.

As principais tarefas de pré-processamento são descritos a seguir e podem ser vistos na Figura 1 (CASTRO; FERRARI, 2016):

- **Limpeza:** para imputação de valores ausentes, remoção de ruídos e correção de inconsistências;
- **Integração:** para unir dados de múltiplas fontes em um único local, como um armazém de dados (*data warehouse*);
- **Redução:** para reduzir a dimensão da base de dados, por exemplo, agrupando ou eliminando atributos redundantes, ou para reduzir a quantidade de objetos da base, resumizando os dados;
- **Transformação:** para padronizar e deixar os dados em um formato passível de aplicação das diferentes técnicas de mineração;
- **Discretização:** para permitir que métodos que trabalham apenas com atributos nominais possam ser empregados a um conjunto maior de problemas. Também faz com que a quantidade de valores para um dado atributo (contínuo) seja reduzida.

Figura 1 – Etapas do processo de preparação de dados



Fonte: Castro e Ferrari (2016)

2.1.2 Banco de dados

O banco de dados é uma parte crucial hoje em dia em qualquer sistema digital. Um banco de dados é uma coleção organizada de informações - ou dados - estruturadas, normalmente armazenadas eletronicamente em um sistema de computador. Um banco de dados é geralmente controlado por um sistema de gerenciamento de banco de dados (DBMS). Juntos, os dados e o DBMS, juntamente com os aplicativos associados a eles, são chamados de sistema de banco de dados, geralmente abreviados para apenas banco de dados.

Os dados nos tipos mais comuns de bancos de dados em operação atualmente são modelados em linhas e colunas em uma série de tabelas para tornar o processamento e a consulta de dados eficientes. Os dados podem ser facilmente acessados, gerenciados, modificados, atualizados, controlados e organizados. A maioria dos bancos de dados usa a linguagem de consulta estruturada (SQL) para escrever e consultar dados (ORACLE, 2020).

Existem muitos tipos diferentes de bancos de dados. O melhor banco de dados para uma organização específica depende de como a organização pretende usar os dados. A seguir alguns dos tipos mais conhecidos de banco de dados, segundo (ORACLE, 2020):

- **Bancos de dados relacionais:** Os bancos de dados relacionais se tornaram dominantes na década de 1980. Os itens em um banco de dados relacional são organizados como um conjunto de tabelas com colunas e linhas. A tecnologia de banco de da-

dos relacional fornece a maneira mais eficiente e flexível de acessar informações estruturadas;

- **Bancos de dados orientados a objetos:** As informações em um banco de dados orientado a objetos são representadas na forma de objetos, como na programação orientada a objetos;
- **Bancos de dados distribuídos:** Um banco de dados distribuído consiste em dois ou mais arquivos localizados em diferentes *sites*. O banco de dados pode ser armazenado em vários computadores, localizados no mesmo local físico ou espalhados por diferentes redes;
- **Data warehouses:** Um repositório central de dados, um *data warehouse* é um tipo de banco de dados projetado especificamente para consultas e análises rápidas;
- **Bancos de Dados NoSQL:** Um NoSQL, ou banco de dados não relacional, permite que dados não estruturados e semiestruturados sejam armazenados e manipulados (em contraste com um banco de dados relacional, que define como todos os dados inseridos no banco de dados devem ser compostos). Os bancos de dados NoSQL se tornaram populares à medida que os aplicativos web se tornaram mais comuns e mais complexos;
- **Bancos de dados gráficos:** Um banco de dados gráfico armazena dados em termos de entidades e os relacionamentos entre entidades.

Esses são apenas alguns dos vários tipos de bancos de dados em uso atualmente. Outros bancos de dados menos comuns são adaptados para funções científicas, financeiras ou outras muito específicas. Para o exemplo de aplicação deste trabalho, o tipo de banco de dados que será utilizado é o relacional, principalmente devido a sua facilidade de implementação.

3 Proposta

A proposta deste trabalho é a de se utilizar dos conceitos base da I4.0 vistos nos capítulos anteriores e, dando um maior enfoque no tratamento de dados, desenvolver uma arquitetura para posterior aplicação de IA e DT. Mais especificamente na etapa de pré-processamento de dados, aplicando algumas das técnicas já consolidadas na literatura para tratar os dados provenientes de fontes ruidosas a um exemplo prático de aplicação real. Será proposto um modelo de pré-processamento, onde analisadas as características do ativo e do sistema de monitoramento, as devidas técnicas de limpeza de dados serão empregadas, assim como um modelo de armazenamento. Com o modelo proposto, visa-se a futura utilização dos dados em aplicações como a construção de um DT e predição de falhas por meio de técnicas de IA.

4 Exemplo de aplicação

Para demonstrar a utilização dos conceitos de pré-processamento de dados vistos no capítulo 2, serão utilizados dados reais coletados em campo, provenientes do sistema de Portas de Plataformas da Estação Vila Matilde da Linha 3-Vermelha do Metrô de São Paulo.

O sistema de Portas de Plataforma, também chamado de PSD (do inglês *Platform Screen Door*) é um sistema relativamente novo e composto por fachadas, portas automáticas e portas de saída de emergência instaladas nas plataformas de estações, principalmente de metropolitanos (usualmente chamado de metrô), monotrilhos e Veículo Leve sobre Trilhos (VLTs).

São diversos os motivos que justificam a utilização deste sistema nas estações, entre eles é possível destacar como sendo os principais:

- Maior segurança para os passageiros, uma vez que previne a queda acidental ou voluntária do mesmo na via, quaisquer que sejam os motivos;
- Impedir a entrada indevida de usuários nos túneis;
- Maior preservação da via e maior disponibilidade de trens, por impedir que usuários joguem lixo e/ou objetos que possam vir a comprometer a circulação dos trens;
- Melhorar a climatização da plataforma, por impedir que chuva e ventos fortes cheguem a ela.

Dada a importância deste sistema, principalmente no que diz respeito à segurança dos usuários, é de grande preocupação que este seja um sistema com número reduzidos de falhas. Assim como diversos outros sistemas críticos para o devido funcionamento da malha metroferroviária, um dos grandes objetivos das equipes de manutenção, é o de evitar ao máximo ter que fazer qualquer tipo de atuação corretiva durante o período em que o serviço de transporte está sendo prestado (período chamado de "comercial"). Para tanto, é necessário que exista um sistema de monitoramento contínuo capaz de prover dados sobre as reais condições dos equipamentos de forma que, com o devido pré-processamento, tratamento e análise destes dados, seja possível obter informações assertivas sobre a saúde do equipamento em questão. Estas informações, quando tiradas a partir de dados confiáveis e adequadamente interpretados, podem ser de grande valia, pois permitem que as equipes

de manutenção saibam de antemão que determinado equipamento pode estar prestes a falhar.

Para melhor entender as possíveis falhas que o sistema PSD pode vir a apresentar e se ambientar à área de aplicação, faz-se necessário compreender o sistema metroferroviário como um todo, isso será feito nas próximas subseções com a descrição de alguns dos principais sistemas responsáveis pela operação dos trens.

4.1 *Composição básica dos sistemas*

Será feita uma descrição dos componentes básicos que compõem um trem, em seguida, dos principais sistemas de controle e operação dos trens, e posteriormente, o sistema PSD, focando principalmente no sistema de movimentação das portas de plataforma.

4.1.1 Trem

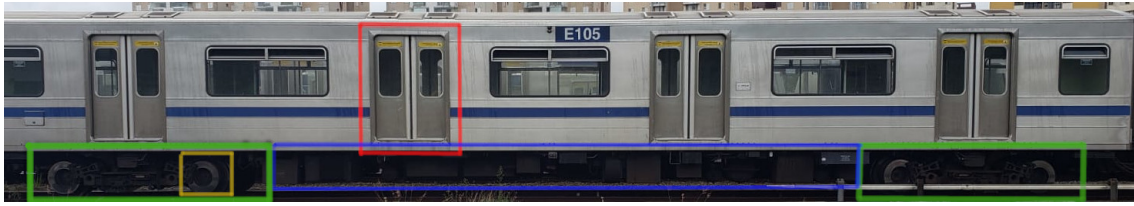
Um trem do Metrô de São Paulo é composto por 6 carros, sendo que a princípio, apenas os carros das extremidades possuem cabines. O enfoque será no funcionamento como um todo de um trem, não vindo a mencionar diversos elementos de controle e potência do mesmo. Assim sendo, simplificadaamente, cada carro do trem é composto por:

- 8 portas, sendo 4 de cada lado do carro;
- 4 Rodeiros, que é o conjunto montado de duas rodas e seus rolamentos em um eixo fixo;
- 2 *Trucks*, formado por dois rodeiros, dois motores elétricos de tração, pinças de freio, molas, suspensões entre outros componentes mecânicos;
- 4 Motores elétricos de tração¹, sendo um responsável pela tração em cada eixo (rodeiro) do carro.

A Figura 3 mostra o carro 5 do trem 10 da frota E, aonde estão destacados os itens descritos. Demarcado por verde estão os *Trucks*, em vermelho uma porta, em amarelo um rodeiro e em azul os demais elementos responsáveis pelo controle, alimentação e tração do trêm.

¹ Atualmente a maioria dos trens do Metrô de São Paulo possuem motores de tração de corrente alternada (CA). Apenas os trens da frota E (a mais antiga ainda em operação) utilizam motores de tração de corrente contínua (CC).

Figura 2 – Carro 5 do trem 10 da frota E do Metrô de São Paulo



Fonte: Autor, 2022

4.1.2 ATC (*Automatic Train Control*) e ATO (*Automatic Train Operation*)

ATC e ATO referem-se respectivamente aos sistemas de Controle Automático de Trens (do inglês *Automatic Train Control*) e Operação Automática de Trens (do inglês *Automatic Train Operation*).

O ATC é o sistema de controle dos trens. Seu funcionamento se dá por meio de blocos fixos, que são trechos ao longo da via com aproximadamente o tamanho de um trem e é responsável por diversas funções, algumas delas são:

- Ler a velocidade do trem, assim como a velocidade máxima permitida naquele bloco;
- Comandar a atuação dos freios de forma segura;
- Comandar a atuação das Máquinas de Mudança de Via (também chamadas de Máquinas de Chave), responsáveis por comutar as rotas dos trens;
- Comunicar com o ATO;
- Controlar a velocidade do trem de acordo com o código de velocidade recebido pela via.

De modo geral a tarefa do ATC é a de controlar o movimento dos trens através dos trilhos de forma automática. Cada bloco fixo é composto por pelo menos uma antena de transmissão (TX) e uma de recepção (RX). Essas antenas formam um "*loop*" no trilho e são responsáveis por fornecer os códigos de velocidade que permitem ao trem saber a velocidade que ele deve estar naquele bloco específico. Ou seja, é possível dizer que no sistema ATC o trem não possui uma "inteligência", no sentido de que o controle de seus movimentos é feito pelos sinais que a via fornece, e o trem apenas age de acordo. Na próxima subseção, será apresentado o sistema CBTC, que é o Controle de Trens Baseado em Comunicação (do inglês *Communication-Based Train Control*), nesse sistema os trens se comunicam entre si, e não há a necessidade de blocos fixos.

Em cada bloco fixo é fornecida apenas a informação da velocidade com a qual o trem deve estar naquele bloco. Assim, o trem só terá a informação se ele deve acelerar, desacelerar ou até mesmo parar, quando ele chegar no próximo bloco e ler a informação proveniente do mesmo. O código de velocidade chega ao trem por meio de um sinal que é induzido em uma antena que fica nos carros em suas extremidades. Essa indução é feita pelo sinal transmitido pela antena TX no *loop* da via e recebido pela antena RX. Esse *loop* é composto por um cabo que corre paralelo ao trilho, por onde passa um sinal com frequência e amplitude bem definidas. Sem entrar em detalhes da forma e protocolo utilizado por este sinal e sua interpretação pelo ATC no trem, a codificação do valor de velocidade é feita pela modulação da frequência do sinal transmitido, formando um código de 6 bits, onde um valor específico de frequência é interpretado como "1" e outro valor específico como "0".

Como forma de garantir uma distância mínima entre os trens, o ATC faz com que os blocos imediatamente antes e depois do bloco no qual o trem está tenham código de velocidade 0. Assim, é garantida uma distância mínima de pelo menos um bloco fixo entre um trem e outro, conforme mostra a Figura 3. Também faz parte do ATC a parte de sinalização das vias.

Já o ATO é responsável por toda a parte de operação automática dos trens, como abertura e fechamento das portas nas estações, o controle das paradas dos trens, restrições de velocidade na via, etc. Portanto, é a integração do ATC com o ATO que permite o funcionamento automático dos trens², tendo implementado em cada um deles diversos mecanismos de segurança, como redundâncias, super dimensionamento, intertravamentos, entre outros, que garantem a operação segura e automática dos trens.

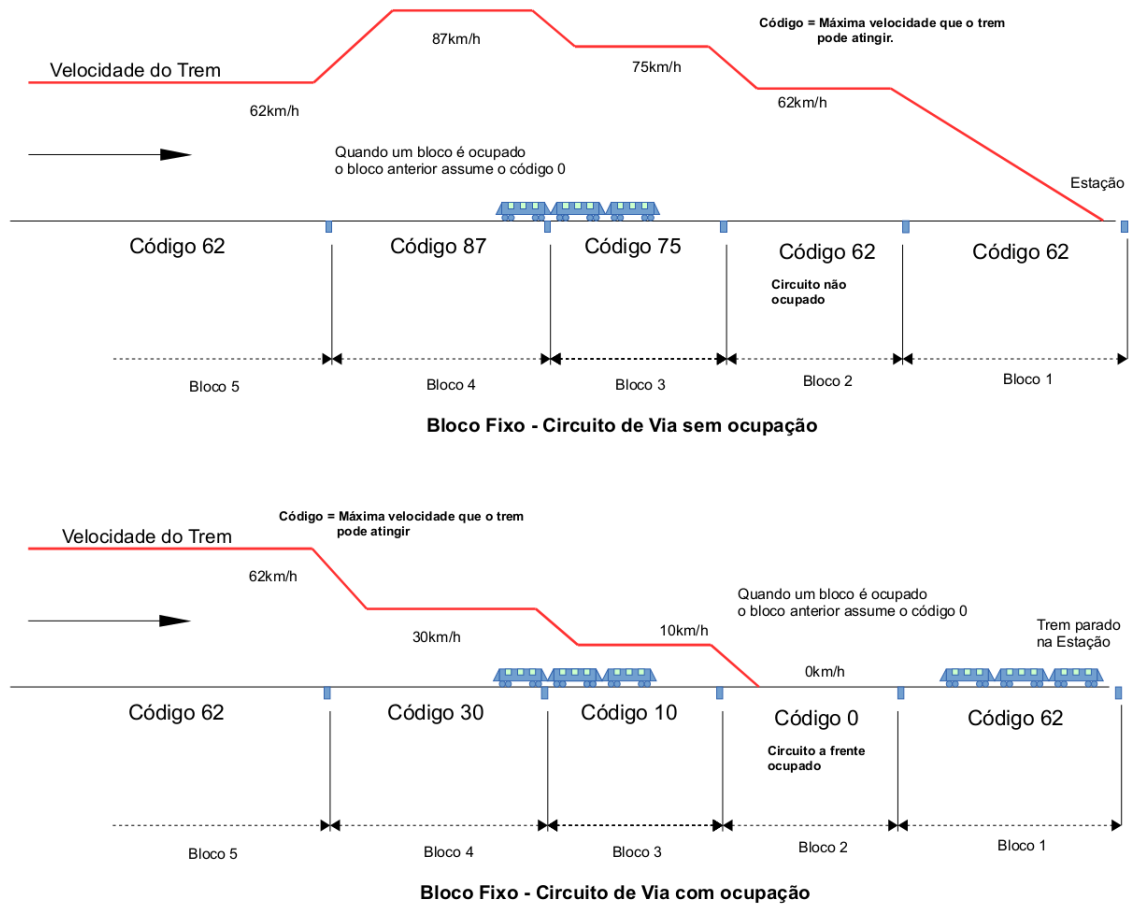
Atualmente apenas as linhas 1-Azul e 3-Vermelha do Metrô de São Paulo, funcionam exclusivamente através dos sistemas ATC e ATO, as demais linhas utilizam o CBTC.

4.1.3 CBTC (*Communication-Based Train Control*)

O CBTC é um sistema de controle trens mais moderno que ATC/ATO. Diferentemente do ATC onde o trem não toma decisão alguma, apenas recebe as informações da

² É importante lembrar que embora todo este controle do tráfego dos trens seja feito de forma automática, é comum que em sistemas de transporte sobre trilhos que utilizem o ATC e o ATO, contem ainda com um operador humano nos trens, a fim de mitigar qualquer situação de emergência não prevista que possa vir a ocorrer

Figura 3 – Códigos de velocidade ao longo da via (os códigos representam a velocidade que o trem deve estar no bloco em km/h)



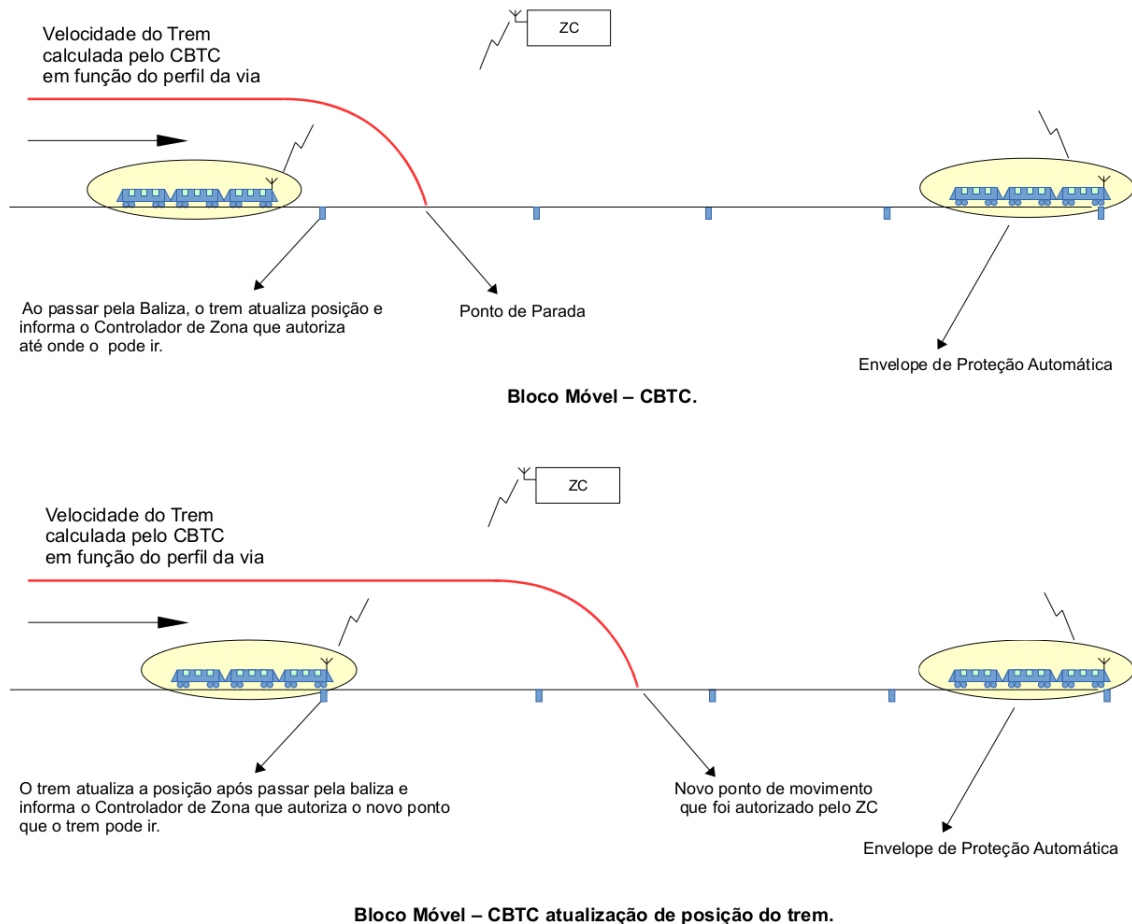
Fonte: Menezes e Alves (2015)

via e trafega de acordo, no CBTC os trens possuem antenas e módulos que permitem a comunicação entre si. Daí o conceito de controle baseado em comunicação. Dessa forma, toda a operação automática dos trens se dá em essência pela comunicação de um trem com os demais e conta ainda com um sistema secundário de segurança que faz uso de balizas ao longo da via, que são responsáveis por atualizar a posição do trem, onde, em casos de falha de comunicação entre os trens, passa a garantir a distância mínima entre eles.

Entre as vantagens de haver uma comunicação entre os trens, está a de não mais se fazer necessário os blocos fixos ao longo da via, como no ATC. A informação sobre qual velocidade o trem deve estar em determinado trecho é obtida por ele próprio com base na distância que ele se encontra dos demais trens na via. Ao invés de um bloco fixo, o que existe no CBTC é um bloco dinâmico, chamado também de "bolha", "bloco móvel", ou ainda "envelope de proteção automática", que envolve o trem e assegura uma distância mínima entre eles. O tamanho desse bloco dinâmico é recalculado a todo instante,

e aumenta proporcionalmente com a velocidade que o trem está. Ou seja, quanto mais alta a sua velocidade, maior é o bloco dinâmico que o envolve e, conseqüentemente, maior a distância mínima entre ele e os demais trens. Na Figura 4 é possível ver o funcionamento dos blocos dinâmicos.

Figura 4 – Blocos dinâmicos no CBTC



Fonte: Menezes e Alves (2015)

Esse sistema garante uma maior disposição dos trens ao longo das linhas, pois permite que os mesmos cheguem mais perto uns dos outros, e aumenta a velocidade média no percurso, diminuindo assim, o intervalo de tempo entre os trens nas estações (também chamado de *headway*).

4.1.4 Sistema PSD (*Platform Screen Door*)

O sistema PSD se comunica tanto com o ATC quanto com o CBTC.³ O sistema será descrito sem entrar em detalhes da parte em que se dá a comunicação com o sistema de controle de tráfego dos trens.

Em cada plataforma é instalada uma fachada de portas de plataforma. Cada fachada é constituída de 24 módulos PSD, e cada módulo possui os componentes que são mostrados na Figura 5 e descritos a seguir:

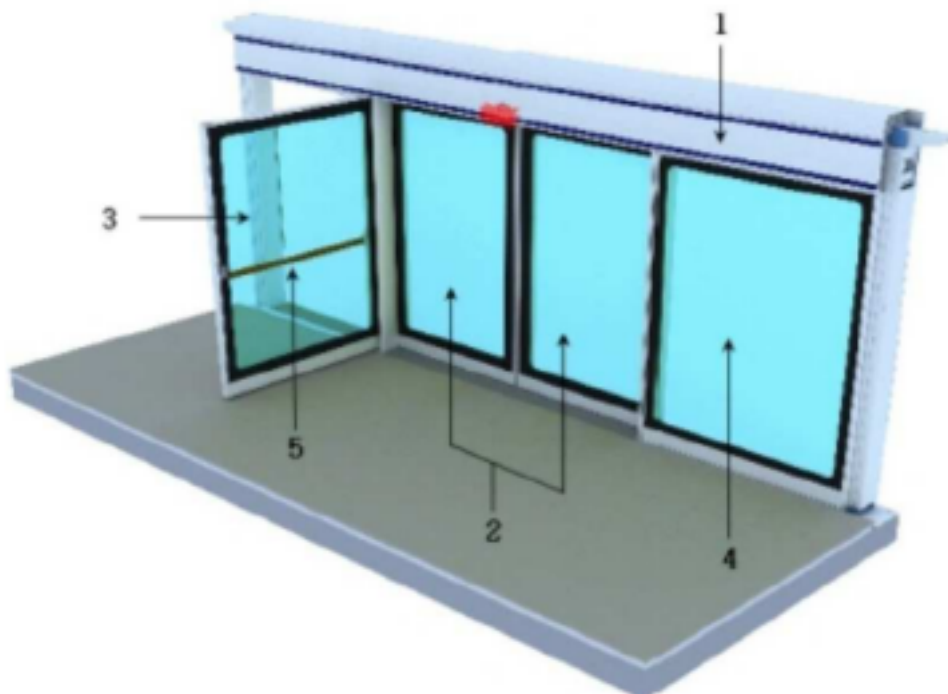
- 1 - *Header Box*: Local em que ficam os componentes responsáveis pelo controle, movimentação, segurança e sinalização do módulo PSD. É no *Header Box* que está acomodado o motor elétrico, componente principal, juntamente com a correia, pela movimentação das portas;
- 2 - Porta Deslizante Motorizada (PDM): Composta de duas folhas deslizantes, que são as partes móveis, responsáveis por abrir e fechar e permitir a passagem dos usuários entre o trem e a plataforma;
- 3 - Porta de Emergência (PEE): É utilizada em caso de emergência, permitindo que o usuário, por meio da Barra Anti-Pânico, por exemplo, consiga abri-la apenas apoiando-se na barra, permitindo ao usuário entrar na plataforma. A barra Anti-Pânico fica do lado da via na PEE;
- 4 - Painel Fixo (PFX): complementa o módulo, sendo uma folha fixa de vidro especial de alta resistência, assegurando o completo fechamento de toda a fachada da plataforma;
- 5 - Barra Anti-Pânico: Barra responsável pela fácil abertura da PEE.

4.1.5 Motor Elétrico da PSD

O principal componente a ser objeto de estudo deste projeto é o motor elétrico, localizado no *Header Box*, responsável pela movimentação (abertura e fechamento) das portas.

³ A depender do tipo de sistema que é utilizado no controle dos trens da linha em questão. No caso da linha 3-Vermelha do Metrô de São Paulo, a comunicação é feita com o ATC, pois é o sistema de controle de trens da mesma.

Figura 5 – Módulo PSD



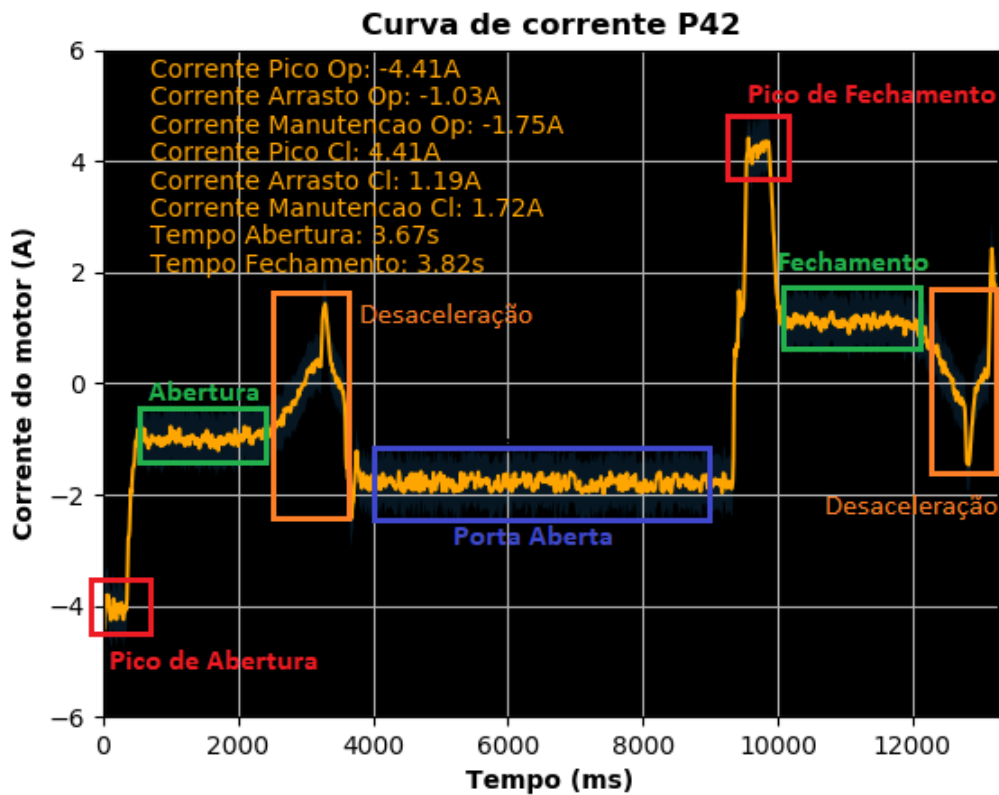
Fonte: Diniz et al. (2014)

Trata-se de um motor elétrico de corrente contínua, com giro nos dois sentidos e com seu controle sendo feito por PWM⁴ (do inglês *Pulse Width Modulation*). Ao ser acionado, o rotor realiza giros controlados. Este rotor está acoplado a um redutor e este a uma correia, que ao girar, arrasta consigo as folhas da PDM. Tanto o controle do giro do rotor, como o torque que ele deve ter, é feito pelo MCP (do inglês *Motor Controller Pro*) que fica localizado no *Header Box* da porta. As características desse controle são parametrizadas no próprio *firmware* do MCP, e fazem com que a curva de corrente do motor na abertura e no fechamento da porta tenha uma forma padrão a ser seguida, que pode ser vista na Figura 6.

Essa curva pode vir a ter leves ou fortes distorções em sua forma, na ocorrência de alguma excepcionalidade, como por exemplo, um usuário que segura a porta durante seu movimento. Porém, o próprio desgaste e desajuste dos componentes que compõem o módulo PSD pode vir a serem responsáveis por variações na forma da curva. E são justamente

⁴ É uma das principais formas de controlar um atuador de corrente contínua de forma digital. Controlando o tempo em que o sinal fica em HIGH e LOW, é possível através da média, controlar o giro e torque do motor.

Figura 6 – Curva de Corrente do Motor de um módulo PSD



Fonte: Autor, 2021

essas variações na curva de corrente que permitem que as equipes de manutenção tirem informações e percebam estes sinais de desgaste precocemente, podendo prever, com certa precisão, a ocorrência de uma falha em determinada porta.

4.2 Tratamento dos dados

Para o presente exemplo de aplicação, a grandeza física que será coletada do motor a fim de se obter um monitoramento contínuo de seu estado, é a corrente elétrica. Uma vez que a mesma possui uma relação direta com o estado de conservação das bobinas e contatos do motor, pois um mal contato ou mesmo uma diminuição da resistência das bobinas, ocasionado por exemplo, por um aumento na temperatura, irá refletir no valor da corrente elétrica, assim como o possível desgaste e desajuste de outros componentes mecânicos que compõem o módulo PSD também poderiam vir a influenciar nesse valor, uma vez que exigiriam maior torque do motor para realizar o ciclo de abertura e fechamento da porta. Serão agora descritas as etapas realizadas desde a aquisição até o armazenamento dos

dados. Por fim, os dados estarão devidamente pré-processados e armazenados de uma base de dados, de forma a estarem prontos para o uso em aplicações de IA e *Digital Twin*.

4.2.1 Aquisição dos dados

A primeira etapa no processo de monitoramento contínuo do motor é a etapa de aquisição de dados. É nesta etapa em que, por meio de sensores e um aquisitor de dados, o valor da grandeza elétrica a ser acompanhada é obtido. Esse valor é uma representação em valor digital do sinal analógico da corrente elétrica, condicionado e convertido para tensão através de um sensor de corrente⁵. Essa tensão passa por um conversor Analógico/Digital responsável por fornecer o valor equivalente em binário que representa a tensão, tornando possível, que o dado seja tratado por meio de algoritmos e armazenado em um banco de dados.

Porém, antes de obter as informações destes dados e tirar quaisquer conclusões a respeito do motor elétrico, faz-se necessário que este dado bruto passe por uma etapa de pré-processamento. Nesta etapa é onde serão retirados dados com valores que por algum motivo, seja ele externo ou interno ao aquisitor, não represente um valor real da grandeza física medida. Essa variação no valor da medida, que pode ser causada, por exemplo, por interferências eletromagnéticas no local da aquisição, mal funcionamento da fonte de alimentação, componente danificado, entre outros, é também chamada de ruído. Para impedir que esses ruídos surjam na etapa de aquisição, permitindo que sejam feitas inferências erradas a respeito do sinal medido, e conseqüentemente, do real estado do motor, será realizada a etapa de pré-processamento.

4.2.2 Pré-processamento

Das diversas técnicas de pré-processamento que foram vistas no capítulo 2, para os dados de corrente do motor, será utilizado o processo de limpeza de dados, que tem como principal objetivo solucionar o problema com dados faltantes e dados ruidosos.

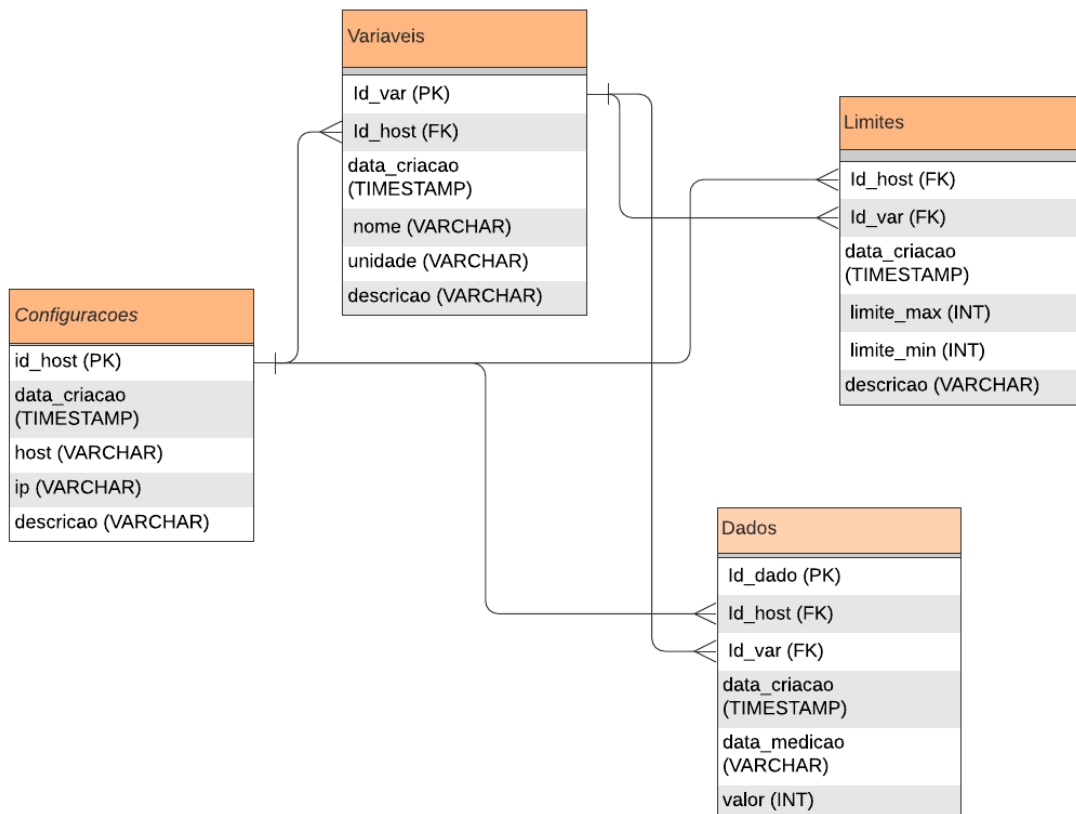
⁵ Existem diversos tipos de sensores de corrente elétrica. Entre os mais comuns, está o que se utiliza do efeito *Hall*. Independente do tipo de sensor de corrente utilizado, a função principal dele é a de fornecer uma saída em tensão proporcional à corrente que está sendo medida.

A aquisição dos dados de corrente do motor é feita de forma contínua, durante toda a operação comercial do Metrô, o que proporciona um grande número de dados ao longo do dia. Dessa forma, a ação a ser tomada a respeito dos dados faltantes e ruidosos será a exclusão dos mesmos. Esta é uma abordagem que só pode ser considerada em casos onde se tem um grande número de dados, como é o caso dos valores de corrente do motor.

4.2.3 Armazenamento

Uma vez que os dados forem limpos pelo método da exclusão, os mesmos serão armazenados em um banco de dados. A proposta de banco de dados para esta aplicação é a de separar os dados em 4 tabelas que se relacionam. A Figura 7 mostra o diagrama do banco de dados que será utilizado para armazenar os dados de corrente do motor já pré-processados.

Figura 7 – Diagrama do Banco de Dados



Fonte: Autor, 2022

Nesta arquitetura de armazenamento, cada uma das 4 tabelas é responsável por armazenar uma categoria de informação a respeito do dado. Sendo que apenas a tabela

Dados possui de fato o valor do dado aquisitado e pré-processado. A seguir será feita uma descrição das tabelas e de alguns de seus atributos, não entrando em detalhes nas relações entre as chaves primárias e estrangeiras.

- **Configuracoes:** Responsável por guardar as informações a respeito do ativo.
 - host: Nome do ativo na rede.
 - ip: IP do ativo na rede.
 - descricao: Descreve as principais características do ativo, como fabricante, parâmetros de configuração, modelo, etc.
- **Variaveis:** Responsável por guardar as informações sobre as variáveis que estão sendo aquisitadas.
 - nome: Nome da variável que esta sendo medida (no exemplo de aplicação: Corrente elétrica do motor).
 - unidade: Unidade da variável (no exemplo de aplicação: Amperes, ou apenas A).
 - descricao: Descreve a variável que esta sendo armazenada.
- **Limites:** Responsável por guardar as informações sobre os limites que a variável pode atingir, a fim de disparar alarmes e que devem ser aplicados nos dados.
 - limite_max: Valor maximo que a variável pode atingir.
 - limite_min: Valor minimo que a variável pode atingir.
 - descricao: Descreve os limites, podendo dar maiores detalhes dos valores empregados.
- **Dados:** Responsável por guardar as informações a respeito do dado.
 - data_medicao: Data em que foi feita a medição, contendo inclusive a hora e o minuto⁶.
 - valor: Valor do dado medido.

Uma vez que os dados já passaram por uma etapa de pré-processamento, retirando possíveis ruídos e espúrios provenientes da etapa de aquisição, e estão devidamente armazenados em um banco de dados, toda a parte de inteligência pode ser aplicada a essa base de dados. É importante lembrar a importância de todas estas etapas vistas até aqui

⁶ A depender do caso, pode ser necessário inclusive a precisão em milissegundos ou até mesmo em microssegundos do momento em que foi feita a medida.

em um projeto como esse, em que se pretende posteriormente, aplicar algoritmos de IA, a fim de prever comportamentos de componentes físicos, como é o caso do motor elétrico do Módulo PSD, bem como a construção de uma base para o DT do mesmo. Vale ressaltar que, dados incorretos que não correspondem a valores físicos reais da grandeza medida podem permitir que sejam tiradas informações erradas a respeito do ativo, invalidando todas as outras etapas do monitoramento e ocasionando ações por parte das equipes de manutenção que além de não trazerem resultados positivos, acabam descredibilizando o sistema de monitoramento, assim como uma base de dados mal sistematizada ou projetada, pode inviabilizar o uso dos dados em determinadas aplicações, por não seguir uma estrutura que permita a injeção desses dados nos algoritmos e técnicas.

5 Resultados

Com a implementação da técnica de limpeza de dados, o posterior trabalho com os dados foi muito facilitado, uma vez que o *software* que é utilizado para pegar os pontos chave da curva de corrente do motor necessita que os dados estejam com seus valores não muito discrepantes do esperado, pois isso acaba fazendo com que o *software* perca as suas referências e não ache os pontos críticos da curva. O pré-processamento, portanto, tornou o uso do *software* muito mais eficiente e assertivo. O modelo de banco de dados proposto não chegou a ser implementado até o momento em que este trabalho foi escrito, porém ele contém diversos pontos positivos, pensados na posterior aplicação de técnica e modelos de IA assim como a criação de um DT do ativo. Por se tratar de um exemplo de aplicação bastante específico, não foi encontrado na literatura outros exemplos a fim de obter alguma comparação de eficiência dos modelos propostos.

6 Conclusões

Este trabalho teve como intuito, fazer uma breve apresentação do que é a Indústria 4.0, assim como alguns de seus conceitos mais importantes como o de IoT, IoS, *Digital Twin* e a Inteligência Artificial aplicada a casos de manutenção. O foco foi abordar algumas técnicas de pré-processamento de dados, etapa essa que é muito importante quando se tem os dados vindo de fontes que podem sofrer diversos tipos de interferências e acabar deixando os dados ruidosos, o que pode refletir em informações dúbias ou até mesmo incorretas. Foi feita também uma revisão do que é um banco de dados e seus tipos e, por fim, foi utilizado um exemplo de aplicação de um caso real, onde um ativo que esta sendo monitorado no Metrô de São Paulo fornece os dados em tempo real para o sistema de monitoramento. Com estes dados foi proposta uma abordagem de pré-processamento para a limpeza de ruídos e um modelo de armazenamento de dados para os mesmos. Essa proposta foi feita com o objetivo de deixar os dados prontos para uma possível criação de um DT do ativo e utilização de técnicas de IA, que permitam análises e previsões.

Os próximos passos deste trabalho seria a implementação do modelo de banco de dados proposto, e analisar a eficiência que traria ao *software*, uma vez que o acesso aos dados já seria feito de forma muito mais objetiva.

Outros trabalhos que poderiam ser realizados são:

- Utilização de outras técnicas de limpeza de dados;
- A possibilidade de retirar os ruídos dos dados por *hardware* e não só por *software*;
- Utilizar um banco de dados não relacional, permitindo utilizar outros paradigmas de análise de dados;

Referências¹

- ARTIFICIAL, I. *O que é a inteligência artificial e como funciona?* 2021. Disponível em: <<https://www.europarl.europa.eu/news/pt/headlines/society/20200827STO85804/o-que-e-a-inteligencia-artificial-e-como-funciona>>. Citado na página 12.
- AUTIOSALO, J. et al. A feature-based framework for structuring industrial digital twins. *IEEE Access*, IEEE, 2019. Citado na página 13.
- BOSCHERT, S.; ROSEN, R. *Digital Twin—The Simulation Aspect*. [S.l.]: Springer, 2016. Citado na página 14.
- CASTRO, L. N. de; FERRARI, D. G. Introdução à mineração de dados : conceitos básicos, algoritmos e aplicações. Saraiva, 2016. Citado 4 vezes nas páginas 15, 16, 18 e 19.
- CHARNIAK, E.; MCDERMOTT, D. Bayesian model of plan recognition. *Massachusetts: Addison-Wesley*, 1985. Citado na página 11.
- DATA, B. *Big data and artificial intelligence: a quick comparison*. 2020. Disponível em: <<https://www.algorithmia.com/blog/big-data-and-artificial-intelligence-a-quick-comparison>>. Citado na página 15.
- DINIZ, M. A. et al. Implementações para melhoria do desempenho do sistema de portas de plataforma do metro de são paulo. *20ª Semana de Tecnologia Metroferroviária*, AEAMESP, 2014. Citado na página 29.
- ERRANDONEA, I.; BELTRÁN, S.; ARRIZABALAGA, S. *Digital Twin for maintenance: A literature review*. [S.l.]: Elsevier, 2020. Citado na página 13.
- GIVEHCHI, O. et al. Interoperability for industrial cyber-physical systems: An approach for legacy systems. *IEEE Transactions on Industrial Informatics PP(99):1-1*, IEEE, 2016. Citado na página 9.
- GRIEVES, M.; VICKERS, J. *Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems*. [S.l.]: Springer, 2016. Citado na página 14.
- HAUGELAND, J. Artificial intelligence: The very idea. *Massachusetts: The MIT Press*, 1985. Citado na página 11.
- HERMANN, M.; PENTEK, T.; OTTO, B. Design principles for industrie 4.0 scenarios. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, IEEE, 2016. Citado na página 8.
- KRITZINGER, W. et al. Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, Elsevier, v. 51, p. 1016–1022, 2018. Citado na página 13.
- KURZWEIL, R. The age of spiritual machines. *Massachusetts: The MIT Press*, 1990. Citado na página 11.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

- LUIS, B. *Entidade: Atributos simples, compostos e multivalorados*. 2020. Disponível em: <<https://www.luis.blog.br/analise-de-entidade-atributos-simples-compostos-multivalorados.html>>. Citado na página 16.
- MENEZES, W. L.; ALVES, M. R. Modernização dos atcs l/o para os trens das frota k e l do metro. *21^a Semana de Tecnologia Metroferroviária*, AEAMESP, 2015. Citado 2 vezes nas páginas 26 e 27.
- ORACLE. *O que é um Banco de Dados?* 2020. Disponível em: <<https://www.oracle.com/br/database/what-is-database/>>. Citado na página 19.
- PETRONI, B.; GONÇALVES, R.; JÚNIOR, I. G. Impacto da internet das coisas na indústria 4.0: uma revisão sistemática da literatura. *IV World Congress on Systems Engineering and Information Technology*, WCSEIT, 2017. Citado na página 9.
- POOLE, D.; MACKWORTH, A.; GOEBEL, R. Computational intelligence: A logical approach. *Oxford: Oxford University*, 1998. Citado na página 11.
- REIS, J. Z.; GONÇALVES, R. F. The role of internet of services (ios) on industry 4.0 through the service oriented architecture (soa): Ifip wg 5.7. *International Conference, APMS 2018, Seoul, Korea, August 26-30, 2018, Proceedings, Part II*, 2018. Citado na página 10.
- ROSEN, R. et al. About the importance of autonomy and digital twins for the future of manufacturing. *IFAC-PapersOnLine*, Elsevier, v. 48, p. 567–572, 2015. Citado na página 14.
- RUSSEL, S.; NORVIG, P. Inteligência artificial. 2. ed. *Rio de Janeiro: Campos*, 2004. Citado na página 12.
- SACOMANO, J. B. et al. *Indústria 4.0: Conceitos e fundamentos*. [S.l.]: Edgar Blucher Ltda., 2018. Citado na página 8.
- TAO, F. et al. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 94, p. 3563–3576, 2018. Citado na página 13.