



ERCEMAPI

Escola Regional de Computação
do Ceará, Maranhão e Piauí

LIVRO DE MINICURSOS

XII EDIÇÃO - 2024

Organizadores

EDUILSON LIVIO NEVES DA COSTA CARNEIRO
RODRIGO AUGUSTO ROCHA SOUZA BALUZ
ROMUERE RODRIGUES VELOSO E SILVA

REALIZAÇÃO



Sociedade Brasileira
de Computação



ORGANIZAÇÃO



FAPEPI

APOIO



Iniciativa da FIEPI - Federação
das Indústrias do Estado do Piauí



FUNDAÇÃO CULTURAL E DE FOMENTO À PESQUISA,
ENSINO, EXTENSÃO E INOVAÇÃO



ERCEMAPI

Escola Regional de Computação
do Ceará, Maranhão e Piauí

XII Escola Regional de Computação do Ceará, Maranhão e Piauí

11 a 13 de setembro de 2024
Parnaíba - Piauí

LIVRO DE MINICURSOS ERCEMAPI 2024

Organização do Livro

EDUILSON LIVIO NEVES DA COSTA CARNEIRO
RODRIGO AUGUSTO ROCHA SOUZA BALUZ
ROMUERE RODRIGUES VELOSO E SILVA

Sociedade Brasileira de Computação
Porto Alegre
2024

Dados Internacionais de Catalogação na Publicação (CIP)

E65 Escola Regional de Computação do Ceará, Maranhão e Piauí (12. : 11 – 13 setembro 2024 : Parnaíba)
Minicursos da ERCEMAPI 2024 [recurso eletrônico] / organização: Edilson Livio Neves da Costa Carneiro ; Rodrigo Augusto Rocha Souza Baluz ; Romuere Rodrigues Veloso e Silva. Dados eletrônicos. – Porto Alegre: Sociedade Brasileira de Computação, 2023.
139 p. : il. : PDF ; 45 MB

Modo de acesso: World Wide Web. Inclui bibliografia
ISBN 978-85-7669-602-5 (e-book)

1. Computação – Brasil – Evento. 2. Segurança da informação. 3. Análise de dados. I. Carneiro, Edilson Livio Neves da Costa. II. Baluz, Rodrigo Augusto Rocha Souza. III. Silva, Romuere Rodrigues Veloso e. IV. Sociedade Brasileira de Computação. V. Título.

CDU 004(063)

Ficha catalográfica elaborada por Annie Casali – CRB-10/2339

Biblioteca Digital da SBC – SBC OpenLib

Índices para catálogo sistemático:

1. Ciência e tecnologia dos computadores : Informática – Publicação de conferências, congressos e simpósios etc. ... 004(063)

Secretarias Regionais

Anselmo Cardoso de Paiva (Secretário SBC Maranhão - UFMA)

Atslands Rego da Rocha (Secretária SBC Ceará - UFC)

Eduilson Livio Neves da Costa Carneiro (Secretário SBC Piauí – UFDPAr)

Coordenação Geral do Evento

Rodrigo Augusto Rocha Souza Baluz

Coordenação do Comitê do Programa de Minicursos

Romuere Rodrigues Veloso e Silva

Comitê do Programa de Minicursos

Alcemir Rodrigues Santos

Eduilson Livio Neves aa Costa Carneiro

Rodrigo de Melo Souza Veras

Mensagem da Coordenação Geral da ERCEMAPI 2024

As Escolas Regionais são eventos voltados aos estudantes de graduação e pós-graduação cujo objetivo principal é aproximar e facilitar a presença dos estudantes em eventos da área de computação. **A Escola Regional de Computação Ceará, Maranhão, Piauí (ERCEMAPI)** tem como objetivo disseminar o conhecimento técnico e científico sobre temas e assuntos de vanguarda na área de Computação. As Escolas Regionais são eventos promovidos pela Sociedade Brasileira de Computação (SBC). A ERCEMAPI visa se consolidar como um evento de referência nos Estados do Ceará, Maranhão e Piauí, além de ter uma importante representação na promoção da ciência e pesquisa na área da Computação dentro da região Nordeste. Além disso, o evento colabora e contribui com a consolidação dos Programas de Pós-Graduação da região.

Paralelamente, como um evento satélite, teremos a realização do Encontro Unificado de Computação do Piauí (ENUCOMPI). O ENUCOMPI vem unindo as instituições de ensino na área de Informática/Computação no estado do Piauí, com o objetivo de contribuir para a troca de experiências, buscando a união dos acadêmicos e no incentivo à produção de trabalhos científicos ligados à área de tecnologia. Em edições anteriores pode proporcionar aos seus participantes, publicações em revistas indexadas como na Revista IEEE América Latina, Revista Brasileira de Computação Aplicada, Revista de Sistemas e Computação, Revista de Informática Aplicada e Revista Learning & Nonlinear Models.

A primeira edição da Escola Regional de Computação dos Estados do Ceará, Maranhão e Piauí (ERCEMAPI) ocorreu em 2007 e foi sediada em Fortaleza, CE. A segunda edição da ERCEMAPI foi realizada na cidade de São Luís, MA. Em 2009, foi realizada a terceira edição da ERCEMAPI na cidade de Parnaíba, PI. Em 2010, a ERCEMAPI foi realizada na cidade de Sobral, CE. O evento recebeu 258 participantes de diversas IES do país e mais 55 estudantes de ensino médio complementando a plateia dos 41 trabalhos apresentados na ERCEMAPI 2010. No ano de 2011, a ERCEMAPI foi realizada pela UFPI em parceria com IFPI, no estado do Piauí. Foram convidados professores e pesquisadores de diversas instituições do país, entre eles: UFRGS, UFSC, UnB, UFRN, UERN, UFBA, UFMA, UNIVALI, UFCG, UNICAMP, IFPI e da própria UFPI. No ano de 2012, a VI Escola Regional de Computação dos Estados do Ceará, Maranhão e Piauí, ocorreu em São Luís do Maranhão em conjunto com a IV Jornada de Informática do Maranhão (JIM 2012). Os palestrantes convidados abordaram temas de Interação Humano Computador (Profa. Dra. Raquel Prates – UFMG) e Imagens Térmicas e seus usos na Biomédica (Profa. Dra. Aura Conci – UFF). Retornando em 2019, a VII Escola Regional de Computação Ceará, Maranhão, Piauí (ERCEMAPI 2019) foi sediada no Centro de Ciências Exatas da Universidade Federal do Maranhão (UFMA). Tiveram presentes estudantes e/ou professores do Maranhão, Piauí, Ceará, Tocantins, Paraíba, Rio de Janeiro e Rio Grande do Norte. Em 2020, a escola foi organizada pela comunidade do Piauí, na forma de conferência virtual por conta da pandemia. Pesquisa e Empreendedorismo foram temas centrais da edição. A ERCEMAPI 2021 foi realizada, novamente em formato online, pela UFC em Quixadá-CE. Em 2022, o evento foi organizado pelo Instituto Federal do Maranhão (IFMA), UEMA, Universidade Federal do Maranhão (UFMA) e pelo Centro Universitário UNDB. No último ano, a ERCEMAPI foi sediada em Aracati (CE).

Coordenação Geral da ERCEMAPI 2024

Rodrigo Augusto Rocha Souza Baluz

Sinopse em Português

Este livro apresenta uma abordagem detalhada e inovadora sobre a Indústria 5.0, aprendizado de máquina, e o impacto do desbalanceamento de dados no desempenho dos modelos preditivos. Explorando a integração entre humanos e máquinas, os autores discutem como a personalização e a sustentabilidade se tornam elementos centrais na nova era da indústria. Além disso, é abordado o papel crucial dos sistemas de decisão orientados por dados, desde a preparação e tratamento de grandes volumes de informações até sua integração em ambientes industriais. O objetivo é evidenciar como esses sistemas contribuem para a eficiência e a personalização das operações, destacando a relevância de práticas éticas e sustentáveis.

Ao longo de seus capítulos, o livro aprofunda-se nos desafios técnicos da implementação de soluções baseadas em inteligência artificial e aprendizado de máquina, com foco em problemas como o desbalanceamento de dados, que afeta diretamente a acurácia dos modelos. São discutidas técnicas de pré-processamento, geração de dados sintéticos e metodologias avançadas, como o aprendizado por reforço, que permitem melhorar a performance em cenários de dados desiguais. Além disso, o livro apresenta o uso de aplicações baseadas em blockchain, e também a criação, comercialização, coleta e análise de dados na rede blockchain. A obra oferece, portanto, uma visão completa e prática, essencial para profissionais e pesquisadores que buscam compreender as tendências e os desafios da Indústria 5.0 e do aprendizado de máquina em ambientes desbalanceados.

Sinopse em Inglês

This book presents a detailed and innovative approach to Industry 5.0, machine learning, and the impact of data imbalance on the performance of predictive models. Exploring the integration between humans and machines, the authors discuss how personalization and sustainability become central elements in the new industrial era. Additionally, it highlights the crucial role of data-driven decision systems, from the preparation and processing of large volumes of information to their integration into industrial environments. The objective is to demonstrate how these systems contribute to operational efficiency and personalization, emphasizing the importance of ethical and sustainable practices.

Throughout its chapters, the book delves into the technical challenges of implementing artificial intelligence and machine learning-based solutions, focusing on issues such as data imbalance, which directly affects model accuracy. It discusses preprocessing techniques, synthetic data generation, and advanced methodologies, such as reinforcement learning, which improve performance in unbalanced data scenarios. Furthermore, the book presents the use of blockchain-based applications, including the creation, commercialization, collection, and analysis of data in the blockchain network. Thus, the work offers a comprehensive and practical perspective, essential for professionals and researchers looking to understand the trends and challenges of Industry 5.0 and machine learning in imbalanced environments.

Sumário

Capítulo 1 - Construindo aplicações baseadas em <i>blockchain</i> com o <i>Hyperledger FireFly</i>	8
Capítulo 2 - Tiny ML: Introdução ao Aprendizado de Máquina em Sistemas Embarcados	33
Capítulo 3 - Indústria 5.0: Implementando Sistemas de Decisão Orientados por Dados	58
Capítulo 4 - Estratégias para Lidar com Desbalanceamento de Dados em Aprendizado de Máquina	79
Capítulo 5 - <i>Design Web</i> Intuitivo: Potencializando Projetos com Material Design e BeerCSS	96
Capítulo 6 - Tokens não Fungíveis (NFTs): Criação, Comercialização, Coleta e Análise de Dados na rede <i>Blockchain</i> Ethereum e na Plataforma de Vendas OpenSea	121

Capítulo

1

Construindo aplicações baseadas em *blockchain* com o *Hyperledger FireFly*

Maria do Rosário de Fátima Martins Ferreira, Bernardo Ferreira de Moura Ribeiro, Alcemir Rodrigues Santos, Ricardo de Andrade Lira Rabêlo

Abstract

Blockchain technology is no longer just about cryptocurrencies. Companies are increasingly exploring the use cases available to increase the added value of their products and services, as well as improving the security and strengthening the trust of their systems. However, developing and maintaining applications based on this technology is still a major challenge, especially when it comes to integrating them with various web applications, what are now known as Web3 applications. Hyperledger FireFly has emerged as a solution with the potential to ease new developers onto the path of developing such applications. Considering the scarcity of teaching material, as well as learning opportunities for beginners in this technology, this chapter proposes an introductory experience for this target audience to get started in the development of Web3 applications.

Resumo

Blockchain não é uma tecnologia apenas sobre criptomoedas. As empresas estão explorando cada vez mais os casos de uso disponíveis para aumentar o valor agregado de seus produtos e serviços, além de melhorar a segurança e fortalecer a confiança de seus sistemas. No entanto, o desenvolvimento e a manutenção de aplicativos baseados nessa tecnologia ainda são um grande desafio, especialmente quando se trata de integrá-los a vários aplicativos da Web, o que hoje é conhecido como Web3. O Hyperledger FireFly surgiu como uma solução com potencial para facilitar a entrada de novos desenvolvedores nesse caminho de desenvolvimento. Considerando a escassez de material didático, bem como de oportunidades de aprendizado para iniciantes nessa tecnologia, este capítulo propõe uma experiência introdutória para que esse público-alvo inicie o desenvolvimento de aplicativos Web3.

1.1. Introdução

Em 2008, o mundo foi testemunha do nascimento da primeira e ainda mais relevante e valorizada criptomoeda até o momento: o *Bitcoin* (BTC) [Nakamoto 2008]. Sob o título de “Um sistema de dinheiro eletrônico ponto a ponto”, a invenção de Satoshi Nakamoto deu início a um mercado de criptomoedas que viria a ser uma das grandes revoluções financeiras e culturais do século. Até a data de publicação deste capítulo, calcula-se que existam mais de 9 mil criptomoedas no mundo. As principais possuem valor de mercado que variam de algumas centenas de milhões de dólares, como *Zcash*, a 1.2 trilhões¹ do próprio *Bitcoin*. Cada uma delas com seu próprio modelo de negócio, as razões para o sucesso das criptomoedas são diversas, mas dependem da confiança na tecnologia que baseia a implementação e operação destes ativos digitais: a *Blockchain*.

Embora tentativas anteriores como *Ecash* [Chaum 1983] e *Bit Gold* [Szabo 2008] tenham falhado, elas exemplificam como a ideia de usar moedas exclusivamente digitais começou muito antes do surgimento do BTC. O mesmo também pode ser dito quanto à outros conceitos por trás da tecnologia *Blockchain*. O texto “*Bitcoin’s Academic Pedigree*”, [Narayanan and Clark 2017], destaca como o trabalho de Nakamoto é na verdade construído a partir de ideias “esquecidas” na literatura. Por exemplo, a assinatura digital e a marcação temporal da criação e última modificação de um documento foi proposta em 1991 por [Haber and Stornetta 1991]. Além disso, a noção de redes distribuídas, como a *peer-to-peer* (P2P), remete aos anos 60 [Baran 1964] e o algoritmo *Proof-of-Work* (PoW), aos anos 90 [Dwork and Naor 1993]. Essa observação não pretende minimizar a conquista de Nakamoto, mas reconhecer que sua verdadeira inovação foi combinar esses componentes subjacentes para criar sua criptomoeda.

Nakamoto uniu a descentralização característica do modelo P2P com uma robusta estratégia de registro criptografado de transações. Este arranjo, integrado por meio de um protocolo de consenso — PoW —, assegurou a imutabilidade das informações inseridas no sistema, estabelecendo uma base sólida para a realização de transações monetárias seguras sem a necessidade de uma terceira parte confiável ou autoridade central. Assim, ele conseguiu adicionar uma forte capacidade de confiança para seu sistema de dinheiro eletrônico, construindo as bases do que seria, cerca de oito anos depois, chamado de “Máquina de Confiança” [The Economist 2015]. Graças a essa inovação, o BTC e sua *blockchain* ganharam impulso nos anos seguintes, representando uma revolução não apenas na maneira como o dinheiro é tratado digitalmente, mas também na transferência de dados no que viria a ser uma nova era da Internet.

Com essa popularização, a tecnologia *Blockchain* evoluiu e o interesse de organizações privadas, governos e empresas de diversos portes pelo uso corporativo dessa tecnologia aumentou substancialmente, expandindo-se os casos de uso para muito além das criptomoedas. Essas instituições vislumbraram na tecnologia uma oportunidade de reduzir a dependência de intermediários em transações online, o que poderia diminuir atritos, atrasos e custos operacionais, além de potencialmente mitigar os riscos de ataques cibernéticos. Atualmente, já existem aplicações em áreas como finanças, saúde, imobiliária e gestão de dados, e a adoção segue em crescimento, acompanhada pela criação de ferramentas que facilitam o desenvolvimento de novas soluções. Nesse cenário,

¹Fonte: <https://investnews.com.br/cotacao-criptomoeda/>. Acesso em 13/08/2024.

o presente capítulo tem como objetivo introduzir os principais conceitos de *Blockchain* e demonstrar como a ferramenta *Hyperledger FireFly* pode ser uma porta de entrada de novos desenvolvedores ao universo da *Web3*.

As próximas seções estão organizadas como segue:

Seção 1.2: nesta seção, serão introduzidos os conceitos básicos associados à tecnologia *blockchain*, com o objetivo de fundamentar o conhecimento.

Seção 1.3: nesta seção, abordamos brevemente como identificar se a *blockchain* é adequada para as necessidades do seu projeto. Esta é uma consideração importante antes da construção de aplicações *Web3*.

Seção 1.4: esta seção introduz a ferramenta *Hyperledger FireFly*, seu ecossistema “guarda-chuva” e outros projetos relacionados. Além disso, são descritos sua arquitetura, funcionamento e componentes.

Seção 1.5: nesta seção, apresentar-se-á o desenvolvimento de uma aplicação exemplo com o *Hyperledger FireFly*, buscando explorar ao máximo as possibilidades.

Seção 1.6: esta seção apresenta os próximos passos para o(a) leitor(a) na construção de suas próprias aplicações *Web3*;

Seção 1.7: nesta seção de encerramento do capítulo, concluiremos apontando algumas possíveis referências de leitura futura.

1.2. A tecnologia *blockchain*

Em referente a nomenclatura, é interessante apontar que, embora a rede em que o BTC opera seja reconhecida como a primeira rede *blockchain* a ser implementada na história, o termo surgiu somente com a posterior popularização da tecnologia. No *white-paper* original, Nakamoto não utiliza-o sequer uma vez. Na verdade, ele se refere a estrutura de dados como “cadeia de blocos” (“*chain of blocks*”), que apesar de simples resume relativamente bem a essência da tecnologia. Além disso, por ser uma combinação de diversos paradigmas, a definição de *Blockchain* pode ser abordada a partir de diferentes perspectivas [Kakavand et al. 2017], o que pode gerar confusões no entendimento. Aqui, buscamos demonstrar como esses conceitos distintos, na realidade, convergem para definir a rica e multifacetada composição que é a *Blockchain*.

A norma ISO 22739:2024 [ISO Central Secretary 2024] define formalmente um padrão internacional para o uso de vocabulário relativo a tecnologias de livro-razão (*ledger*)² distribuído (DLTs), incluindo a *Blockchain*, com o objetivo de definir e esclarecer conceitos e termos técnicos. No documento, o termo “*blockchain*” é definido como sendo uma estrutura de dados, enquanto “*sistema blockchain*” é definido como um sistema DLT que implementa uma *blockchain*, e finalmente, “*tecnologia blockchain*” é definido como

²Na contabilidade, um livro-razão, uma espécie de registro de escrituração que funciona como uma ferramenta de controle e análise de finanças em empresas, permitindo o registro e a conferência de transações financeiras.

o conjunto de todas as tecnologias necessárias para operação e uso de um sistema *blockchain*. Neste capítulo, adotaremos as definições desta norma ISO, com a ressalva de que quando nos referirmos à tecnologia - *Blockchain*-, usaremos letra maiúscula para diferenciá-la da estrutura de dados e das redes -*blockchain*.

Em geral, uma *blockchain* é uma plataforma digital que armazena e verifica todo o histórico de transações entre usuários da rede de forma inviolável e à prova de edições criminosas. Ela também pode ser vista como uma estrutura de banco de dados subjacente para transações digitais seguras [Kakavand et al. 2017]. Greve *et al.* [Greve et al. 2018] apresentou os elementos fundamentais da *blockchain*: **o livro-razão distribuído, o mecanismo de consenso e as técnicas de criptografia**. A Tabela 1.1 resume as características que permitem que a tecnologia tenha o potencial de substituir transações baseadas em pura confiança por transações baseadas em regras definidas matematicamente e aplicadas mecanicamente. Está fora do escopo desta capítulo detalhar cada um desses atributos, porém mais detalhes podem ser encontrados no trabalho referenciado.

Tabela 1.1. Elementos fundamentais e propriedades. Fonte: [Greve et al. 2018]

Livro-Razão	Transparência, Auditabilidade, Imutabilidade e Irrefutabilidade
Distribuição	Descentralização, Replicação e Disponibilidade
Consenso	Desintermediação, Atualidade e Integridade
Criptografia	Identidade, Assinaturas Digitais, Segurança e Privacidade

1.2.1. Afinal, o que é uma *blockchain*?

Uma *blockchain* é um tipo de sistema descentralizado, o que significa que, diferente da maior parte das aplicações modernas, ela não é controlada por um servidor central, ou seja, não possui uma autoridade como uma grande empresa ou governo controlando-a. Essa plataforma se destaca pela capacidade de, mesmo sem uma autoridade central, verificar a validade desses envios e ainda armazenar todo o histórico de transações em um livro-razão digital distribuído e inviolável, garantindo segurança e auditabilidade para esse registro. Nesta sub-seção buscamos definir com mais precisão técnica o que de fato é uma *blockchain*, explorando algumas de suas definições e seus principais componentes.

No coração de uma *blockchain*, o livro-razão é um elemento chave na definição e no funcionamento da tecnologia. Este elemento é composto por dois componentes interligados: uma base de dados que contém o estado atual da rede, conhecida como estado mundial (*World State*), e um registro de todas as transações que levaram a esse estado, organizadas em uma estrutura de dados baseada em uma cadeia de blocos, a que se nomeia *blockchain*. A cadeia registra as mudanças de estado e “alimenta” o estado mundial, que por sua vez reflete os dados atuais do sistema, sendo implementado como um banco de dados. Isso é vantajoso pois assim ele é capaz de oferecer um conjunto eficiente de operações de armazenamento e recuperação de estados, evitando que um programa precise percorrer toda a cadeia de blocos para calcular um valor específico; em vez disso, ele pode ser obtido diretamente a partir do estado mundial. Enquanto isso, a *blockchain* mantém seguras as informações de evolução deste banco de dados, possibilitando transparência para conferências futuras.

Uma *blockchain*, portanto, é uma estrutura de dados que por razões didáticas pode ser entendida como uma lista ligada de blocos de transações (T_x), ligados entre si em forma de cadeia (Figura 1.1) e que registra todas as transações que ocorrerem entre os nós na forma de arquivo. Neste arranjo, cada bloco é criptograficamente encadeado ao bloco anterior através de um código *hash*³, que resume as informações do antecessor, também chamado bloco-pai. Dessa forma, cada novo bloco criado possui uma marcação de tempo de criação (*Timestamp*) e um número escalar arbitrário chamado *nonce*, que ajuda a validar o bloco e, em seguida, adicioná-lo à *blockchain*. Além de ajudar a preservar a integridade do histórico, esses elementos também conferem à estrutura a característica de ser imutável e apenas anexável, ou seja, é praticamente impossível alterar ou excluir um bloco previamente inserido sem danificar toda a cadeia.

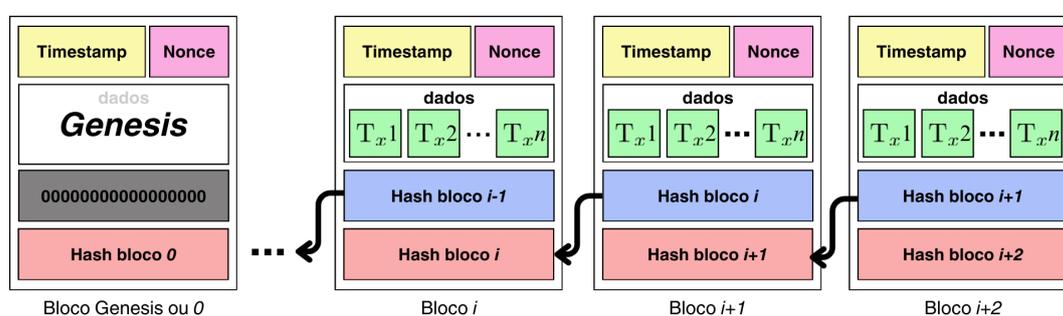


Figura 1.1. Representação de uma *blockchain*. (Fonte: autor)

O livro-razão da *blockchain* também tem a característica de ser distribuído através de uma rede. Com isso, surge o conceito de *blockchain* como sendo também uma rede distribuída do tipo peer-to-peer (P2P), ou seja, par-a-par. Cada par, também chamado nó (*Node*), é um *software* sendo executado em uma máquina diferente, que mantém uma cópia do livro-razão e que pode submeter e receber transações a ele através de uma conexão. O nó é um elemento fundamental da rede, enquanto essa é responsável por conectar as máquinas uma a uma, disseminar informações e sincronizar os livros-razão. Um determinado tipo de nó é capaz de sugerir novos blocos e a cada intervalo de tempo um nó é escolhido para anexar seu bloco proposto ao registro compartilhado. Para evitar que essa escolha acabe sendo partidária ou centralizada, a mesma é tomada em conjunto pelos nós, a partir de um processo chamado consenso. Uma vez escolhido, o bloco “vencedor” é transmitido através da rede por protocolos P2P.

O processo de consenso é crucial para garantir que todas as transações sejam registradas de forma precisa e segura na *blockchain*. Na rede *Bitcoin*, por exemplo, esse consenso é alcançado por meio de um mecanismo chamado *Proof-of-Work* (PoW). Nele, os nós mineradores competem para resolver complexos problemas matemáticos e o primeiro a encontrar uma solução válida adiciona um novo bloco à *blockchain*. Este processo é custoso e exige que a solução seja verificada por outros mineradores, garantindo que todas as transações sejam legítimas e que o novo bloco esteja em conformidade com as regras da rede. Por exemplo, tal como uma poupança, uma carteira de BTC não pode gastar mais

³Resumo das informações do bloco criptografadas por algum algoritmo como SHA-256

do que o disponível como saldo de suas transações anteriores. Uma vez que a maioria dos mineradores aceita um novo bloco como válido, ele é adicionado à *blockchain* e registrado em todas as cópias do livro-razão, assegurando que todos os participantes tenham uma visão consistente e confiável do histórico de transações, ou seja, do livro-razão.

Apesar dessas definições, ressaltamos que, conforme [Kakavand et al. 2017] em “A Revolução *Blockchain*”, essa tecnologia não tem uma única definição universalmente aceita, por se tratar de uma composição complexa de várias dimensões, incluindo tecnológica, operacional e regulatória. Ainda segundo os autores, a tecnologia já foi descrita como sendo, ao mesmo tempo, uma rede distribuída e um banco de dados equipado com segurança integrada. Entretanto, a chegada da rede Ethereum [Buterin 2013] e a concomitante formalização da tecnologia [Wood 2019] endossaram a visão de *Blockchain* também como uma máquina de estados. Por esta razão, e levando em conta a importância desse conceito para o desenvolvimento da *Web3*, consideramos essencial adicionar mais essa definição, o que será feito na seção a seguir.

1.2.2. O surgimento da *Ethereum*, Contratos Inteligentes e *Web3*

No mesmo ano em que o jornal britânico *The Economist* publicou a matéria que chamava a tecnologia *Blockchain* de “Máquina de confiança” [The Economist 2015], uma nova rede pública foi lançada: a Ethereum [Buterin 2013]. Essa revolucionária rede, apesar de funcionar de forma semelhante à Bitcoin, foi concebida com a proposta de fornecer uma *blockchain* equipada com um computador *built-in*, além de uma linguagem de programação completa e integrada, capaz de criar programas executáveis complexos. Conhecidos como contratos inteligentes (*smart contracts*), esses programas são acordos autoexecutáveis baseados em código, projetados para automatizar processos e permitir a criação das mais diversas aplicações descentralizadas (*DApps*), com suas próprias regras de negócio.

Isso é possível pois na rede distribuída da Ethereum, há um único computador canônico, chamado de Máquina Virtual Ethereum (EVM), um computador virtual global cujo estado é aceito e copiado por todos os nós da rede. Qualquer participante pode transmitir uma solicitação, também chamada de transação, para que a EVM execute um cálculo arbitrário. Sempre que uma solicitação desse tipo é transmitida, outros nós da rede verificam, validam e executam o cálculo. Essa execução causa uma alteração de estado na EVM, que é confirmada e propagada por toda a rede. O registro de todas as transações e o estado atual da EVM são então armazenados na *blockchain*.

Por esse motivo, autores como Greve *et al.*, definem que a *blockchain* implementa uma “máquina de estados replicada para a manutenção consistente de um estado global compartilhado por um conjunto de pares distribuídos numa rede P2P” [Greve et al. 2018]. Como reconhece [Buterin 2013], tanto o modelo de *blockchain* da Bitcoin quanto da Ethereum podem ser vistos como máquinas de estado. Entretanto, enquanto o estado na primeira se refere basicamente a contas e saldos, na segunda ele é uma grande estrutura de dados que contém também um estado de máquina, que pode mudar de acordo com um conjunto predefinido de regras definidas pela EVM. É a definição de estados mais complexos, através dos contratos inteligentes, que possibilita a criação de interações mais complexas com a *blockchain*, o que forma a base para execução das *DApps*.

As *DApps* são aplicações que rodam na Internet e permitem execução de contra-

tos inteligentes em redes *blockchain*. Qualquer desenvolvedor pode criar um contrato inteligente personalizado para seu negócio e torná-lo público na rede, usando a *blockchain* como sua camada de dados. Assim, qualquer usuário pode então usar uma *DApp* para chamar o contrato e executar seu código. Essas aplicações são componentes-chave na construção da *Web3*, a emergente nova geração da internet, onde aplicações são executadas de forma descentralizada, visando mais segurança, privacidade e controle dos usuários sobre seus dados e identidades digitais. Essa revolução popularizou ainda mais a tecnologia, levando empresas e governos a adotarem-na para fins além das criptomoedas.

1.2.3. *Blockchain* pública vs privada

A norma ISO 22739, entre outras definições, categoriza as tecnologias DLT como permissionadas e não-permissionadas, bem como públicas e privadas. Na prática, contudo, os termos frequentemente são utilizados como sinônimos, sendo uma *blockchain* não-permissionada compreendida como pública, onde não são necessárias permissões e, portanto, qualquer usuário pode acessar e inserir informações no livro-razão. Em contrapartida, uma *blockchain* permissionada é frequentemente entendida como privada, pois apenas usuários com permissão podem acessá-la e modificá-la.

As redes *Bitcoin* e *Ethereum* são exemplos de redes públicas, ou seja, qualquer usuário pode executar funções de leitura e escrita. Em uma *blockchain* não-permissionada, não há uma autoridade central que controle o acesso ou que possa banir usuários ilegítimos. Essa característica de forte abertura proporciona um alto nível de transparência, permitindo que todas as transações sejam visíveis para todos os participantes da rede. Além disso, os usuários desse modelo não precisam confiar uns nos outros para realizarem transações seguras entre si. Apesar disso, algumas desvantagens das redes públicas é a baixa taxa de transações por segundo (TPS), que pode aumentar a latência e diminuir o *throughput* das transações. De fato, a rede *Bitcoin* leva, em média, por volta de 10 minutos para validar uma transação [Zheng et al. 2018]. Ademais, os elevados gastos com taxas de transação, juntamente com os altos custos computacionais e energéticos destacam-se como desvantagens do modelo não-permissionado.

Por outro lado, redes privadas oferecem a capacidade de restringir o acesso a um grupo específico de participantes, garantindo maior privacidade e controle sobre o livro-razão. Nesse modelo, uma entidade conhecida como Âncora de Confiança é responsável por conceder e revogar permissões de leitura e gravação no sistema. Algumas redes *blockchain* permissionadas limitam o acesso a usuários pré-verificados que já comprovaram sua identidade. Outras permitem a visualização para qualquer pessoa, mas apenas identidades confiáveis podem verificar transações. Embora esse modelo possa parecer contradizer o conceito original de *blockchain*, por depender da confiança em uma entidade específica, é preciso considerar que em ambientes empresariais isso é frequentemente aceitável devido ao nível pré-existente de confiança entre os participantes. Dessa forma, corporações podem aproveitar os benefícios da *Blockchain* enquanto ainda mantém um nível de controle sobre seus negócios. As redes *Hyperledger Fabric* e *R3 Corda* são exemplos notáveis de blockchains permissionadas.

1.3. Quando usar (e quando não usar) *Blockchain*?

Com a popularização da *Blockchain*, a tecnologia se tornou amplamente difundida em diversos setores, sendo vista como uma solução revolucionária para inúmeras aplicações além de seu uso inicial em finanças. Sua ascensão ao destaque levou a implementações bem-sucedidas em áreas como saúde, transporte, indústria e gestão da cadeia de suprimentos, exemplos que serão abordados mais adiante neste capítulo. No entanto, como observa [Zile and Strazdiņa 2018], o “hype” em torno da tecnologia cria a ilusão de que a *Blockchain* é a solução para qualquer problema, o que leva algumas organizações a adotarem a tecnologia sem avaliar sua adequação às necessidades específicas de projetos, resultando em sistemas complexos, ineficientes ou de difícil manutenção.

Este capítulo já discutiu as definições e características da tecnologia *Blockchain*, incluindo sua promessa de segurança, privacidade e operações descentralizadas. Entretanto, essas características, por si só, nem sempre justificam a adoção da tecnologia. Segundo [Wust and Gervais 2018], uma análise das propriedades da *Blockchain*, considerando os requisitos e o contexto específicos de cada aplicação é necessária para avaliar se sua aplicação adequada. Esta seção pretende explorar esses fatores, ajudando a identificar quando o uso da tecnologia é realmente justificado.

Primeiro, alguns autores como [Wust and Gervais 2018], [Hassija et al. 2021] e [Enoch et al. 2023] que discutem o uso adequado de um livro-razão distribuído descrevem que a análise deve considerar: a necessidade de armazenamento de dados, o envolvimento de múltiplas partes e as dinâmicas de confiança entre elas. A *Blockchain* é mais vantajosa quando diversas entidades que não confiam totalmente umas nas outras precisam interagir e modificar o estado de um sistema sem depender de uma terceira parte confiável (TTP). Nesses casos, a *Blockchain* pode viabilizar transações e trocas de dados seguras, assegurando que nenhuma parte tenha controle unilateral sobre o sistema. No entanto, quando uma TTP está disponível e é confiável, um banco de dados tradicional pode ser uma solução mais eficiente, oferecendo melhor desempenho em termos de capacidade de processamento e latência.

Além disso, é importante discutir as diferenças entre redes *blockchain* permissionadas e não-permissionadas, e quando a aplicação de cada uma delas é mais adequada. Como mencionado anteriormente neste capítulo, redes não-permissionadas, como o *Bitcoin*, permitem que qualquer usuário ingresse na rede sem a necessidade de aprovação de uma autoridade central. Em contraste, redes permissionadas, como o *Hyperledger Fabric*, limitam o acesso a um grupo restrito de leitores e escritores, com uma autoridade central responsável por gerenciar a adesão e conceder ou revogar acessos.

Assim, uma rede permissionada é ideal para cenários onde há necessidade de interação entre um número restrito de entidades que são conhecidas entre si e há a necessidade de manter a privacidade, enquanto ainda se beneficiam de um livro-razão descentralizado. Por outro lado, as redes não-permissionadas são projetadas para situações em que o conjunto de escritores não é fixo, os participantes são desconhecidos uns dos outros, e o sistema precisa permanecer aberto a qualquer pessoa que deseje participar. Ele é particularmente vantajoso em ambientes onde a descentralização é fundamental e a verificabilidade pública deve ser garantida sem depender de uma entidade central confiável.

Em conclusão, embora a *Blockchain* tenha o potencial de revolucionar certos negócios, ela não é uma solução universal. Sua necessidade depende de fatores específicos, como a necessidade de armazenar dados, o envolvimento de várias partes e o nível de confiança entre elas. Logo, em situações onde a confiança não é uma questão ou onde sistemas centralizados podem operar com mais eficiência, o uso da tecnologia pode ser desnecessário e até contraproducente. Portanto, é fundamental realizar uma avaliação cuidadosa desses fatores antes de decidir implementar a tecnologia *Blockchain*.

1.4. O *Hyperledger FireFly*

Reconhecendo as vantagens da tecnologia *Blockchain*, várias empresas interessadas no uso corporativo dela perceberam que poderiam obter resultados mais significativos ao colaborar entre si, unindo esforços para desenvolver soluções de código aberto acessíveis a qualquer usuário. Nesta seção, pretendemos demonstrar como um desses projetos colaborativos evoluiu para se tornar uma referência no desenvolvimento de redes *blockchain*, abrangendo não apenas o uso corporativo, mas também acadêmico e científico. Analisaremos o Ecossistema *Hyperledger*, sua visão, os membros que o compõem e os projetos em andamento, especialmente o *Hyperledger FireFly*, que constitui o tema deste capítulo.

1.4.1. Surgimento: a Fundação *Hyperledger*

No contexto de contribuição coletiva, em 2015, sob a guarda da Fundação Linux, a Fundação *Hyperledger* teve início. Originalmente conhecida como Projeto *Hyperledger*, hoje fundação, tinha como objetivo ser um esforço colaborativo para construir uma plataforma de livros-razão distribuídos e código aberto, para uso em demandas industriais específicas [Foundation 2016]. Na atualidade, a Fundação *Hyperledger* é um ecossistema global aberto com foco no uso empresarial da tecnologia. A fundação não se considera uma rede de blockchains, mas como uma rede de consórcios onde empresas e outras organizações podem implementar suas soluções de DLT [Foundation 2024].

A *Hyperledger* propõe nutrir uma comunidade que combina as possibilidades empresariais do uso da *blockchain* com as vantagens do desenvolvimento de código aberto, criando uma plataforma de criação e uso coletivo de DLTs. Todo o código do portfólio da Fundação *Hyperledger* é desenvolvido publicamente e está disponível sob a licença Apache. As organizações se associam à Fundação *Hyperledger* para demonstrar liderança técnica, colaborar e interagir com outras empresas, além de aumentar a conscientização sobre seus esforços na comunidade de *blockchain* empresarial. Os membros da *Hyperledger* incluem organizações líderes em diversos setores: finanças e bancos, como American Express, Visa e BNDES; saúde; cadeias de suprimentos, como Walmart; e tecnologia, com IBM, Oracle e Huawei, entre outros. Em seguida, apresentamos sobre os projetos desenvolvidos no ecossistema *Hyperledger*.

1.4.2. Ecossistema: Projetos *Hyperledger*

A lista de projetos mantidos pela fundação é dinâmica⁴, com a inclusão ou remoção de projetos à medida que novos projetos embrionários vão ganhando maturidade ou proje-

⁴É possível encontrar detalhes dos projetos mantidos atualmente na página Web principal da Fundação: <https://www.hyperledger.org/>

tos maduros vão perdendo espaço na comunidade. Atualmente⁵, existem três categorias de projetos orientados à *blockchain* dentro do ecossistema Hyperledger. A saber, (i) os DLTs, que podem ter propósito geral ou específico de identidade digital; (ii) as bibliotecas; e (iii) as ferramentas. A Tabela 1.2 enumera projetos de cada uma dessas categorias. Embora muitos dos projetos tenham aplicações no contexto de *Web3*, não é do escopo deste capítulo detalhar e/ou apresentar todos os projetos disponíveis para este propósito.

Tabela 1.2. Projetos Hyperledger.

DLTs	Besu, Fabric, Iroha, Inddy, Identus
Bibliotecas	Aries, Web3J
Ferramentas	AnonCredits, Bevel, Cacti, Caliper, Cello, FireFly, Solang

Dentre os projetos graduados, destacam-se as DLTs *Fabric* e *Besu*. O *Hyperledger Fabric* é uma plataforma para criar soluções *blockchain* com uma arquitetura modular e que oferece um alto grau de confidencialidade, flexibilidade, resiliência e escalabilidade. Isso permite que soluções desenvolvidas com a *Fabric* sejam aplicáveis a qualquer setor. Na rede *Fabric* ainda não existe um token nativo (como o *Bitcoin* ou *ETH*), porém tokens podem ser criados a nível de chaincode de aplicação, emulando tokens fungíveis (ERC-20) ou não fungíveis (*NFT*) (ERC-721). Por outro lado, o *Hyperledger Besu*, diferente do *Fabric* que implementa uma rede própria, é um cliente *Ethereum* de fácil utilização empresarial. Ele foi projetado para uso em redes públicas e privadas, além de redes de teste (e.g. Sepolia, Görli). O *Besu* tem suporte nativo para tokens fungíveis, *NFTs* e outros padrões, assim como uma gama bastante variada de opções de consenso (*Proof of Stake*, *Proof of Work* e *Proof of Authority*).

Dentre as ferramentas, o *Hyperledger FireFly*, objeto deste capítulo, e o *Hyperledger Cacti* se destacam. O projeto do *Hyperledger FireFly* será melhor detalhado na sub-seção seguinte (1.4.3). Enquanto isso, o *Hyperledger Cacti* é uma plataforma de integração de *blockchain* que permite aos usuários realizar transferências de ativos entre *blockchains* direfentes (*cross-chain*). Em outras palavras, ele permite trocar, por exemplo, um carro representado por um *token* em uma rede *Fabric* por *ETH* na rede principal *Ethereum*. Por fim, dentre as bibliotecas, destacamos o *Hyperledger Aries*, um kit de ferramentas completo para soluções de identidade descentralizada e confiança digital. Ele permite emitir, armazenar e apresentar credenciais verificáveis com o máximo de preservação da privacidade e estabelecer canais de comunicação confidenciais e contínuos.

1.4.3. Conhecendo melhor o FireFly

O *Hyperledger FireFly* é o projeto *Hyperleger* que tem por objetivo principal encapsular diversas camadas de configuração para construção de aplicações *Web3* escaláveis e seguras. A documentação do projeto⁶ o define como um *supernó*, i.e. uma pilha de aplicações completa. De fato, é possível entender o *FireFly* como uma caixa de ferramentas, na qual os desenvolvedores podem tanto acoplar aplicações já existentes ao mundo *Web3*, quanto

⁵Refere-se ao tempo de escrita do capítulo, Julho de 2024.

⁶Disponível em: <https://hyperledger.github.io/FireFly>

construir aplicações *Web3* do zero. A partir de agora, iremos dar detalhes sobre como o projeto foi concebido e como utilizá-lo para alcançar o seu propósito.

Em outras palavras, o *Hyperledger FireFly* pode ser visto como um *gateway* para um mundo *Web3*, o que inclui todo o ecossistema que sua aplicação deverá participar, *i.e.* múltiplas redes *blockchain*, múltiplas aplicações de finanças descentralizadas, ou mesmo múltiplas redes de negócio. O *Hyperledger FireFly* não é uma nova implementação de uma *blockchain*, na verdade, ele é uma forma de orquestrar todas estas conexões mencionadas, bem como uma camada de dados. De fato, redes *blockchain* públicas, aplicações descentralizadas de armazenamento, padrões e ecossistemas de *tokens*, contratos inteligentes e *frameworks* de identidade são todos exemplos de tecnologias descentralizadas que o *Hyperledger FireFly* permite a integração.

Sem o *Hyperledger FireFly*, o time de desenvolvimento de aplicações teria que construir além da aplicação do negócio, todo o controle de *tokens*, identidade, mensagens, privacidade, gerenciamento de documentos e dados, além da orquestração de dados, do nó da *blockchain* e as conexões com a *Web3*. A proposta do *Hyperledger FireFly* é encapsular tudo isso, para permitir que o time de desenvolvimento se concentre na camada de negócio, deixando que a comunidade possa desenvolver o *Hyperledger FireFly* por meio de um esforço coletivo. De fato, este conceito de *supernó* evoluiu ao longo da última década, com os desenvolvedores percebendo a necessidade de muito mais que a criação de um nó de *blockchain* para que seus projetos fossem bem sucedidos, o que consumiria uma parcela enorme do orçamento disponível para o projeto.

Arquitetura do Hyperledger FireFly

De acordo com os idealizadores do projeto, o *supernó FireFly* foi projetado para ser extensível, com *runtimes* de execução modulares e separadas, orquestradas em uma API intuitiva para desenvolvedores. Uma das principais características do *FireFly* é a modularidade e alta configurabilidade de seus componentes. Para isso, adota-se uma abordagem baseada em microsserviços, que combina pontos de conexão (*plug-points*) no *runtime* principal com extensibilidade de API para *runtimes* remotos, implementados em diversas linguagens de programação.

Figura 1.2 mostra a arquitetura central de um nó *FireFly*. Ela é composta por três áreas principais: os diferentes *runtimes* que encapsulam o nó, as responsabilidades do *runtime* principal e seus elementos plugáveis, e o código em execução dentro do nó. Um nó é, essencialmente, uma coleção de múltiplos *runtimes* unificados por uma única API HTTPS/WebSocket, exposta pelo *Core*, com um banco de dados privado para armazenar dados locais e conectividade para interagir com outros nós da rede por meio de suas respectivas *runtimes*.

Funcionamento do Hyperledger FireFly

O *Hyperledger FireFly* opera em dois modos principais: *Web3 Gateway* e *Multiparty*. Um único *runtime* pode funcionar em ambos os modos por meio do uso de *namespa-*

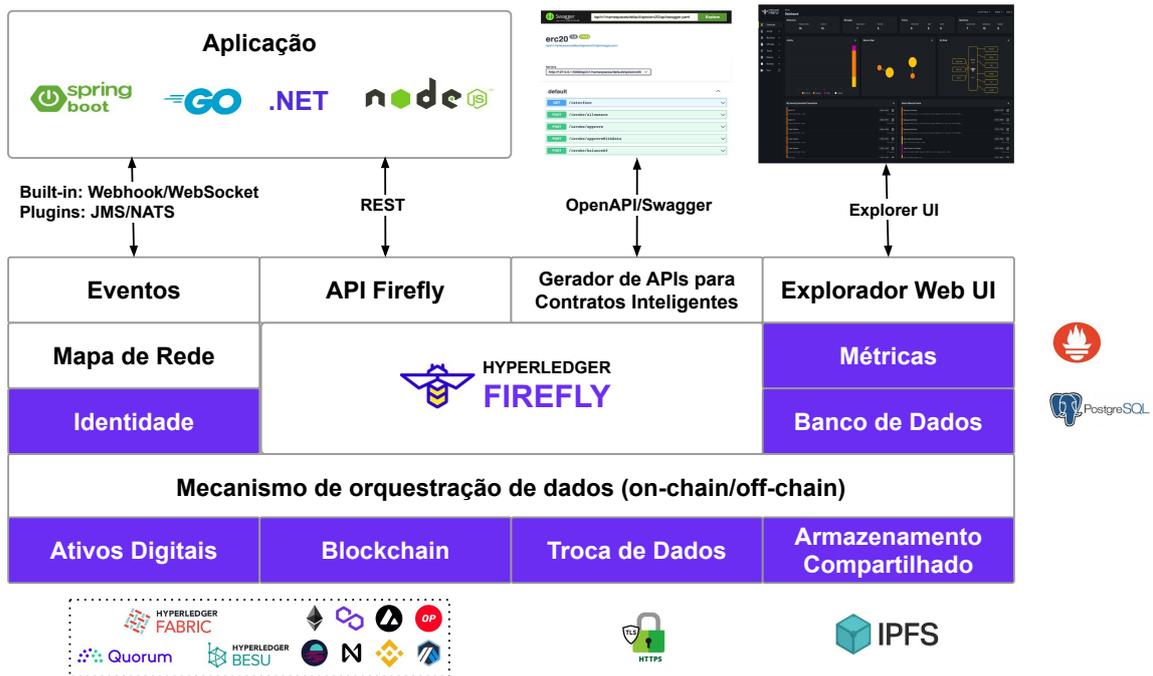


Figura 1.2. Arquitetura do *Hyperledger FireFly*. (Fonte: Adaptado da documentação)

ces distintos. O modo *Web3 Gateway* facilita a interação com qualquer aplicação *Web3*, independentemente de outros membros da rede de negócios utilizarem o *FireFly*. As principais funcionalidades incluem transferências de valores *tokenizados*, invocação de contratos inteligentes, indexação de dados da *blockchain*, disparo de eventos para aplicações e sistemas centrais e gerenciamento de dados descentralizados (como *NFTs*).

O modo *Multiparty*, por outro lado, é projetado para a construção de sistemas com várias partes (organizações) onde um *runtime* de aplicação comum é implantado por cada participante empresarial. Esse modo possibilita o desenvolvimento de aplicações sofisticadas que utilizam as APIs modulares do *Hyperledger FireFly* para oferecer soluções empresariais completas. Além de todas as funcionalidades disponíveis no modo *Web3 Gateway*, esse modo também permite o compartilhamento de formatos de dados comuns, troca privada de dados por um barramento criptografado, coordenação de dados *on-chain* e *off-chain*, além de mascaramento de atividades *on-chain* via *hashes* e uso de um livro de endereços compartilhado para gerenciar identidades

Principais componentes principais do *Hyperledger FireFly*

A Figura 1.3 apresenta um esquema das funcionalidades disponibilizadas pelo *Hyperledger FireFly*. Está fora do escopo deste capítulo explorar cada um dos componentes do *Hyperledger FireFly*. Em vez disso, focaremos nos componentes que serão essenciais para o tutorial deste trabalho: as ferramentas que integram o ambiente de desenvolvimento. Mais especificamente, estudaremos o que são e como utilizar: a Interface de Linha de Comando (*FireFly CLI*), o Explorador (*Web UI*), a OpenAPI (*Swagger API UI*) e a *Sandbox UI*. Além dessas ferramentas, a própria API do *FireFly* e o kit de desenvolvimento (SDK)

serão essenciais para a criação de aplicações *Web3*.

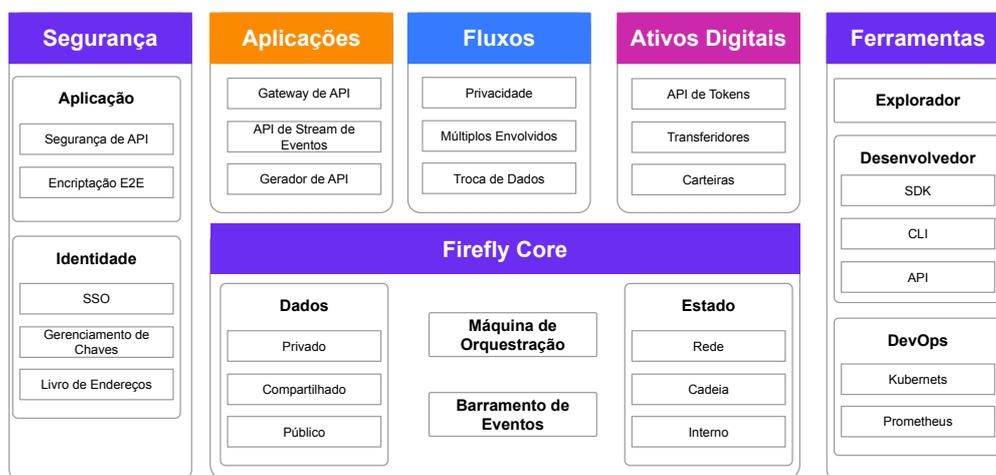


Figura 1.3. Funcionalidades do Hyperledger FireFly. (Fonte: Adaptado da documentação)

Dado que a maioria das plataformas de *blockchain* oferece a capacidade de executar contratos inteligentes, o *Hyperledger FireFly* suporta a interação com esses contratos por meio de APIs *REST*, além de permitir a escuta de eventos via *WebSocket*. A API garante uma experiência de uso uniforme, independentemente da implementação específica da *blockchain* subjacente. O *Hyperledger FireFly* também possui uma *SDK* cliente para *Node.js*, que permite que aplicações sejam criadas com facilidade.

A CLI do *FireFly* é a ferramenta de desenvolvimento fundamental, a partir da qual as demais serão acessadas. Com a CLI, nós locais do *FireFly* podem ser criados e executados para facilitar o desenvolvimento *offline* de aplicações baseadas em *blockchain*. Isso permite que os desenvolvedores iterem rapidamente em suas ideias sem a necessidade de configurar uma grande infraestrutura antes de escrever a primeira linha de código. Já o Explorador faz parte do próprio *FireFly Core* e oferece uma visão do sistema que permite monitorar o estado atual da rede e investigar transações, mensagens e eventos específicos. Além disso, é uma ferramenta útil para que desenvolvedores visualizem os resultados de suas implementações ao utilizar a API do *FireFly*.

Enquanto isso, o *Swagger* é uma interface interativa gerada pela especificação OpenAPI do *FireFly*, que serve como uma documentação visual e também como ferramenta de teste para a API. Ela permite que desenvolvedores explorem e façam chamadas diretamente na API do *FireFly* de forma simples e intuitiva. A *Sandbox* do *Hyperledger FireFly* não faz parte do super-nó. Na verdade, é um aplicativo Web de usuário final que usa a API do *FireFly*. Assim como o *Swagger*, a *Sandbox* também serve ao propósito de orientar os desenvolvedores a usar mais facilmente a API, mas diferente dele, ela tem um backend e um frontend projetados para tal. A *Sandbox* também fornece trechos de código como exemplos de como criar esses recursos em seu próprio aplicativo.

1.5. Desenvolvendo soluções com Hyperledger FireFly

1.5.1. Instalação da FireFly CLI

Para começar a desenvolver aplicações utilizando o *Hyperledger FireFly*, o primeiro passo é a instalação das dependências e da interface de linha de comando (CLI) do FireFly. Há várias maneiras de instalar a CLI do FireFly, porém destacamos apenas a maneira mais fácil, que consiste em fazer o download de um binário pré-compilado (no formato compactado `gzip`) da versão mais recente, cujo download pode ser feito a partir do repositório oficial⁷. Até o momento, estão disponíveis somente versões do binário para os sistemas operacionais Linux e MacOS, mas usuários do sistema Windows podem executá-lo através do módulo Subsistema Windows para Linux (WSL)⁸ ou através de máquinas virtuais. A Tabela 1.3 destaca os pré-requisitos necessários para o funcionamento da CLI.

Tabela 1.3. Pré-requisitos FireFly CLI.

Docker	Tecnologia de containerização e virtualização.
Docker Compose	A pilha do FireFly será executada em um projeto <code>docker-compose</code> .
OpenSSL	Para funções básicas de criptografia e autenticação

Após a instalação dos pré-requisitos e o download do pacote apropriado para seu sistema operacional e arquitetura de CPU, o passo seguinte compreende a extração do binário. Supondo que o pacote do *GitHub* tenha sido baixado em seu diretório *Downloads*, execute o comando presente no Código 1.1 via terminal *bash* para extrair o binário e movê-lo para `/usr/bin/local`. Caso o pacote baixado se encontre em um diretório diferente, será necessário alterar o comando abaixo para incluir o local correto onde o arquivo `FireFly-cli_*.tar.gz` estiver localizado.

```
1 sudo tar -zxf ~/Downloads/FireFly-cli_*.tar.gz -C /usr/local/bin  
ff && rm ~/Downloads/FireFly-cli_*.tar.gz
```

Código 1.1. Extração do binário da FireFly CLI

Um método de instalação alternativo é instalar a CLI via *Go*, caso a máquina possua um ambiente de desenvolvimento local apropriado e a variável `PATH` esteja configurada corretamente. Para ambos os métodos, um tutorial mais detalhado pode ser encontrado na documentação oficial do *Hyperledger FireFly*⁹.

1.5.2. Inicialização do ambiente

Uma vez configurada a CLI, a próxima etapa é criar e iniciar uma pilha (*stack*) do *FireFly* para executar um ou vários super-nós. Uma pilha é um conjunto de super-nós com rede e configuração projetados para trabalhar juntos em uma única máquina de desenvolvimento. Dessa forma, no modo *multiparty*, uma pilha tem vários membros (também chamados de

⁷<https://github.com/hyperledger/FireFly-cli/releases/tag/v1.3.0>

⁸<https://learn.microsoft.com/pt-br/windows/wsl/>

⁹https://hyperledger.github.io/FireFly/latest/gettingstarted/FireFly_cli

organizações) e cada membro tem seu próprio super-nó dentro da pilha. Isso permite que os desenvolvedores criem e testem fluxos de dados com uma combinação de dados públicos e privados entre várias partes, tudo em um único ambiente de desenvolvimento. Cada membro da pilha possui uma instância de cada ferramenta do *FireFly* (e.g. *Sandbox*, *Explorador*, *Swagger*), através das quais a API pode ser chamada.

Como mencionado anteriormente, um super-nó é uma estrutura complexa que integra diversos módulos e controladores. Uma das principais vantagens do *Hyperledger FireFly* é trazer facilidade e rapidez na criação desses elementos complexos utilizando apenas uma linha de código. O comando presente no Código 1.2 permite a configuração básica de uma nova pilha, sendo possível ainda, durante este processo, selecionar a *blockchain* que servirá de base (e.g. Ethereum, Fabric ou Tezos), garantindo a integração adequada com a infraestrutura desejada. Ao executar esse trecho, serão solicitados o nome da nova pilha e o número de membros desejados.

```
1 ff init < ethereum | fabric | tezos >
```

Código 1.2. Criando uma stack FireFly

```
1 ff init ethereum dev 3 --node-name=pessoa1_node --org-name=
  pessoa1_org --node-name=pessoa2_node --org-name=pessoa2_org
  --node-name=pessoa3_node --org-name=pessoa3_org
```

Código 1.3. Criando uma stack baseada em ethereum com 3 membros

A escolha da rede base depende no geral dos objetivos do projeto. É importante destacar, porém, que a rede Fabric não possui suporte nativo para criação, gerenciamento e transação de *tokens*, embora seja possível anexar módulos específicos para isso, como o Fabric Token SDK¹⁰. Por razões de praticidade, neste tutorial criaremos, via Docker Compose, uma pilha baseada em Ethereum com 3 membros. Para isso, é crucial garantir a disponibilidade de pelo menos 3 GB de memória *RAM* na máquina. Para sistemas que executam contêineres Docker dentro de uma máquina virtual, como o macOS, é necessário certificar-se de ter alocado memória suficiente.

Ao inicializar uma pilha no *FireFly*, há diversas opções de customização disponíveis, que permitem ajustar as configurações de acordo com as necessidades do projeto. Por enquanto, utilizaremos apenas argumentos para especificar os nomes dos nós e organizações, para evitar a geração randômica e facilitar a visualização. Nessa inicialização, definimos também automaticamente o nome da pilha como *{dev}* e a quantidade de membros como três (3). O Código 1.3 exibe o comando com os argumentos apropriados. No mais, manteremos os padrões. Para ver a lista completa de opções do *FireFly*, basta executar `ff-help` na interface de linha de comando.

Após a criação da pilha, um arquivo Docker Compose correspondente será gerado, mas ainda será necessário executar mais um comando para iniciar a pilha. Para isso, basta rodar o comando mostrado no Código 1.4. Na primeira vez que o comando é executado, esse processo pode levar alguns minutos, durante os quais a CLI realiza várias etapas automaticamente: faz o *download* das imagens Docker dos componentes do

¹⁰<https://github.com/hyperledger-labs/fabric-token-sdk>

super-nó; inicializa e configura uma nova *blockchain* dentro de um contêiner; estabelece a comunicação entre os componentes; implanta os contratos inteligentes necessários (e.g. *BatchPin* e *token* ERC-1155); e registra as identidades dos membros e nós da rede.

```
1 ff start dev
```

Código 1.4. Executando a stack *dev*

Depois que sua pilha terminar de iniciar, ela imprimirá os *links* de acesso para o Explorador Web (Web UI), a interface de interação com a API (Swagger API UI) e a Sandbox de cada membro. Caso tudo funcione corretamente, deve ser retornado algo semelhante a 1.5. Atenção: Para os usuários de macOS, a porta padrão (5000) já está em uso pelo serviço *ControlCe* (*AirPlay Receiver*). É possível desativar esse serviço ou, alternativamente, criar a pilha usando uma porta diferente, adicionando o argumento `{-p 8000}` ao comando de criação (1.2 ou 1.3).

```
1 Web UI for member '0': http://127.0.0.1:5000/ui
2 Swagger API UI for member '0': http://127.0.0.1:5000/api
3 Sandbox UI for member '0': http://127.0.0.1:5109
4
5 Web UI for member '1': http://127.0.0.1:5001/ui
6 Swagger API UI for member '1': http://127.0.0.1:5001/api
7 Sandbox UI for member '1': http://127.0.0.1:5209
8
9 Web UI for member '2': http://127.0.0.1:5002/ui
10 Swagger API UI for member '2': http://127.0.0.1:5002/api
11 Sandbox UI for member '2': http://127.0.0.1:5309
```

Código 1.5. Links e portas correspondentes a execução local da stack *dev*

1.5.3. Usando a Sandbox

A Sandbox (Figura 1.4) é dividida em três colunas, cada uma com uma função. Na esquerda da página, destacado em vermelho, há três guias que alternam entre os três conjuntos principais de funcionalidade do Sandbox. Nessas guias, você pode criar mensagens, *tokens* e contratos simplesmente preenchendo campos de formulário.

Por razões práticas, esta seção explorará como exemplo apenas a função de enviar mensagens, ou seja, a guia MENSAGENS. É nela que podemos enviar transmissões *broadcasts* e mensagens privadas através da *Blockchain*. Algumas ações a serem exploradas incluem: enviar uma mensagem de *broadcast* e visualizar o conteúdo em todos os membros no Explorador; enviar uma mensagem privada a um único membro e verificar que o conteúdo não está visível para o terceiro membro; ou enviar um arquivo de imagem e realizar o *download* a partir do Explorador de outro membro.

Na coluna do meio, em azul, você poderá visualizar o código que será enviado ao servidor para executar aquela tarefa e, abaixo, ver a resposta. Ao digitar os dados no formulário no lado esquerdo da página, você poderá notar que o código-fonte na parte superior central da página é atualizado automaticamente. Se você estiver criando uma aplicação de *backend*, esse é um exemplo de código que seu aplicativo poderia usar para

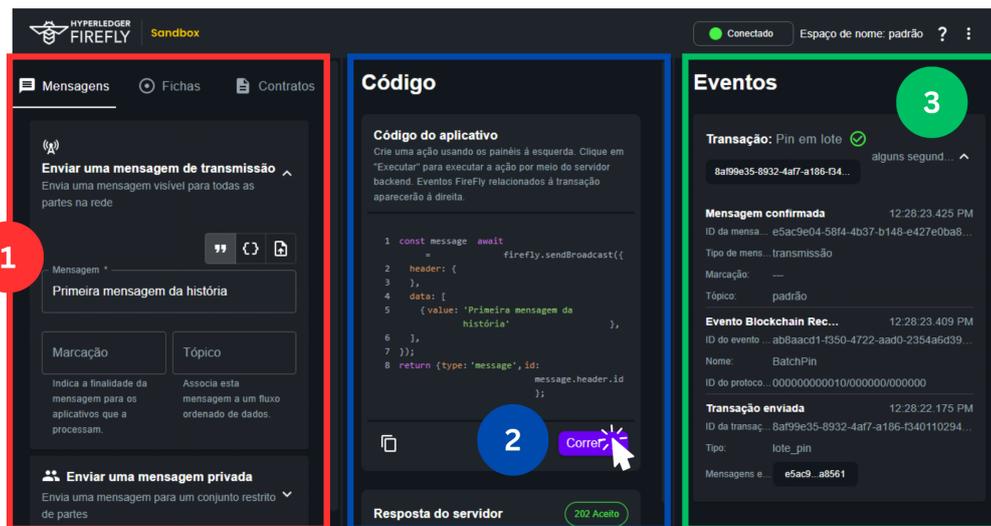


Figura 1.4. Sandbox *Hyperledger FireFly* correspondente ao membro 1 (Fonte: autor)

chamar o FireFly SDK. A coluna do meio também contém um botão *RUN* (na imagem, correr) para realmente enviar a solicitação de transação. Na imagem acima, é possível notar que a mensagem em questão é enviada através da função como um JSON no campo data. Ao construir sua própria aplicação, esse recurso pode ajudar na essencial tarefa de se atentar aos dados e formatos adequados a cada função da API.

Finalmente, no lado direito da página, é possível acompanhar o fluxo de eventos recebidos pela conexão *WebSocket* que o *backend* estabeleceu com o *FireFly*. Por exemplo, ao realizar solicitações para envio de mensagens, é possível visualizar a confirmação dessas mensagens em tempo real. Esse mesmo histórico de eventos pode ser visualizado integralmente a partir do Explorador Web. Outros exemplos de uso como a criação e mineração de *tokens* podem ser encontrados na documentação oficial do *FireFly* ¹¹.

1.5.4. Desenvolvimento de um contrato inteligente simples

Nesta sub-seção, descreveremos em etapas como implementar e fazer *deploy* de um contrato inteligente simples em uma *blockchain* Ethereum. Em seguida, exploraremos como o *FireFly* pode ser usado para interagir com o contrato a fim de enviar transações, consultar estados e ouvir eventos. Para esta parte do guia, é requerido que esteja em execução na máquina tanto o Docker quanto uma pilha *FireFly* local com pelo menos dois membros. Detalhamos como uma pilha pode ser executada na sub-seção 1.5.2, Código 1.4.

Criação, compilação e *deploy*

O contrato inteligente que usaremos como exemplo (Código 1.6) é chamado *SimpleStorage* e foi extraído da documentação oficial do *Hyperledger FireFly*. Como o nome indica, é um contrato muito simples que armazena um número inteiro sem sinal (*unsigned*) de

¹¹<https://hyperledger.github.io/FireFly/latest/gettingstarted/sandbox/#things-to-try-out>

256 bits e que emite um evento na *Blockchain* sempre que o valor armazenado é atualizado, permitindo que o valor atual seja recuperado. Este contrato foi escrito na linguagem Solidity e deve ser compilado usando o compilador *solc* ou alguma outra ferramenta. Recomendamos que a compilação seja feita com ajuda de uma IDE online chamada *Remix*, que permite uma rápida configuração de versão do compilador e dispensa mais instalações na máquina, facilitando o processo. Para ambos os casos, após compilar o contrato, um arquivo JSON será retornado e o utilizaremos para solicitar o *deploy*.

```
1 // SPDX-License-Identifier: Apache-2.0
2 pragma solidity ^0.8.10;
3
4 // Declara o contrato
5 contract SimpleStorage {
6     // Define a variavel que armazenara um valor inteiro sem
7     // sinal
8     uint256 x;
9
10    // Funcao que altera o valor da variavel
11    function set(uint256 newValue) public {
12        x = newValue;
13        emit Changed(msg.sender, newValue);
14    }
15
16    // Retorna o valor armazenado na variavel x
17    function get() public view returns (uint256) {
18        return x;
19    }
20
21    event Changed(address indexed from, uint256 value);}
```

Código 1.6. Exemplo de contrato inteligente em Solidity (Fonte: adaptado da documentação)

É crucial destacar que as versões mais recentes do compilador *solc* estão apresentando problemas de incompatibilidade com algumas versões da EVM, impossibilitando o *deploy* de alguns contratos por razões que fogem do escopo deste capítulo. Entretanto, por segurança, recomendamos o uso da versão 0.8.19 do compilador e, se possível, da versão paris da EVM. Essas configurações podem ser facilmente ajustadas através do *Remix*, na aba de configurações avançadas.

Uma vez compilado, o próximo passo é fazer *deploy* do contrato utilizando a API de implantação de contrato integrada do *Hyperledger FireFly*. Existem duas formas de fazer isso, a primeira delas utilizando o comando apresentado no Código 1.7 diretamente na CLI do *FireFly*. Assim, caso tudo ocorra corretamente, o endereço de *blockchain* do seu contrato será impresso. Esse endereço será essencial em seguida para registrar a API do contrato e é ele que nos referiremos nas sub-seções seguintes.

```
1 ff deploy ethereum dev simple_storage.json
```

Código 1.7. Deploy de contrato inteligente através da CLI

A segunda abordagem, que consideramos mais simples, envolve o uso da interface *Swagger*, especificamente a rota `/postContractDeploy` (Figura 1.5). Nesse método, enviamos ao servidor uma requisição com base em campos específicos do JSON, que incluem: *i*) o *bytecode* compilado do contrato inteligente, inserido no campo “*contract*” como uma cadeia de caracteres codificada em hexadecimal ou Base64; *ii*) o *array* ABI completo, que deve ser inteiramente copiado para o campo “*definition*”. Além disso, se aplicável, é necessário fornecer uma lista ordenada dos argumentos do contato. Devido a simplicidade do exemplo, basta definir o corpo da solicitação substituindo as seções apropriadas pela saída resultante da compilação do *Remix*, de acordo com o Código 1.8.

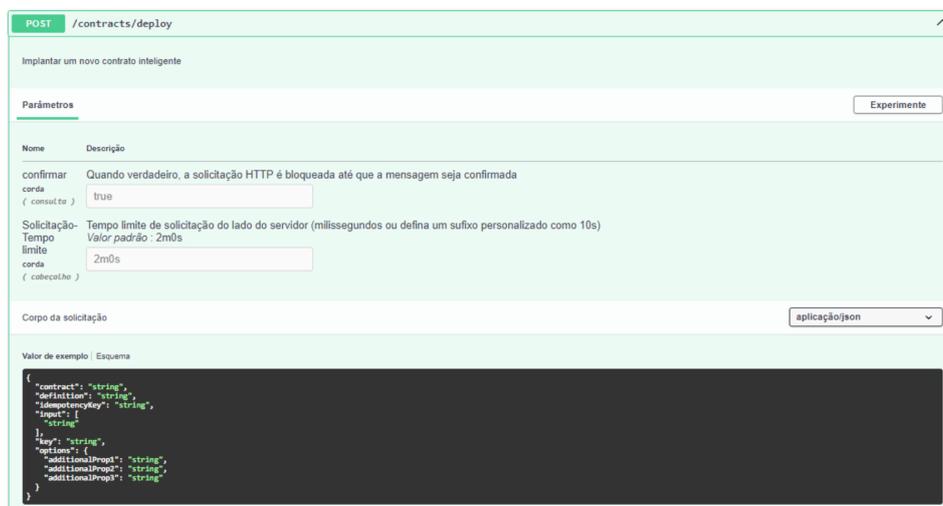


Figura 1.5. Rota para *deploy* de contrato com o *Swagger* do *FireFly*. (Fonte: autor)

```

1 {
2 "contract": "INSIRA_BYTECODE_AQUI",
3   "definition": INSIRA_ABI_AQUI,
4   "input": []
5 }

```

Código 1.8. Esquema de requisição para *deploy* de contrato

Se a requisição for bem-sucedida, estará confirmada a transação e uma resposta detalhada será gerada. Ao rolar para baixo, na parte inferior, o endereço do contrato implantado estará disponível no campo “*address*”. Um evento correspondente ao *deploy* também será registrado na *blockchain* e será possível examiná-lo através da interface do *Explorer*. Para obter informações detalhadas sobre o bloco, que serão necessárias em uma etapa posterior deste guia, basta acessar o evento e verificá-las.

Gerando uma interface e uma API HTTP para o contrato inteligente

Fazer *deploy* de um contrato armazena-o na *blockchain*, mas não disponibiliza nenhuma interação com ele aos demais membros da rede. Assim, para que esses possam referenciá-lo, é preciso criar e compartilhar uma descrição do contrato, de suas funções, tipos de

entrada e retornos, ou seja, uma Interface de Contrato (FFI). A API do *Hyperledger FireFly* possui uma rota específica para isso, cujo uso é preferível à escrita manual da FFI. A partir deste passo, todos os passos seguintes que configuram chamadas à API podem ser executados tanto através do Swagger quanto da Sandbox. Novamente, por questões de praticidade, descreveremos apenas a segunda maneira, deixando como recomendação a leitura da documentação para demais instruções.

Na Sandbox, acesse a guia “contratos” à esquerda da página e navegue até o formulário “Definir Interface de Contrato”, selecione a opção Solidity ABI, dê um nome e uma versão à sua interface preenchendo os campos apropriados. Cole no campo Esquema (*schema*) a mesma ABI do contrato que usamos no passo anterior e, por fim, garanta que a *checkbox* “Publish” esteja ativada, para que a interface seja automaticamente publicada à rede. Execute o código e visualize a confirmação do evento na aba à direita. De volta à guia de contratos, navegue até o formulário logo a baixo do anterior para enfim gerar uma API HTTP para interagir com o contrato. Neste passo, será necessário selecionar a interface desejada, nomear o contrato e indicar o endereço dele na *blockchain* (vide subseção anterior). Por fim, indique que a API deve ser publicada automaticamente na rede, execute o comando e verifique o registro do evento.

Agora, o contrato deve estar acessível a todos os membros da pilha e portanto é possível verificar a API e obter sua interface OpenAPI interativa por meio do *Explorador*. Utilizando as rotas `post/invoke`, já é possível interagir com o contrato.

1.5.5. Integração do contrato inteligente na aplicação Web3

Finalmente, poderemos criar nossas próprias aplicações *Web3* utilizando a API que criamos para interagir com o contrato desenvolvido. Para isso, o *Hyperledger FireFly* disponibiliza um Kit de Desenvolvimento de Software (SDK) em `Node.js` para criação de projetos utilizando a linguagem *typescript*. A Tabela 1.4 resume algumas das principais funções disponíveis para manipulação da API do FireFly, que permitem que o desenvolvedor realize todas as instruções anteriores deste guia diretamente de sua própria aplicação *backend*, com maior grau de controle e personalização.

Tabela 1.4. Principais funções da FireFly Node.js SDK. Fonte: <https://github.com/hyperledger/FireFly-sdk-nodejs/tree/main>

Função	Uso
<code>sendBroadcast</code>	Envia mensagens em transmissão <i>Broadcast</i>
<code>createTokenPool</code>	Cria um pool de <i>token</i> personalizado
<code>deployContract</code>	Possibilita o <i>deploy</i> de contrato inteligente
<code>invokeContractAPI</code>	Invoca a API de um contrato inteligente

Além disso, a função `invokeContractAPI` possibilita a interação com contratos inteligentes customizados. A partir daqui, o leitor já tem as ferramentas para começar a construir sua primeira aplicação *Web3* completa, desde o *setup* da infraestrutura, o *deploy* de contratos inteligentes e a criação de APIs para interação com seu *backend*, tudo isso facilitado pelo *Hyperledger FireFly* e suas ferramentas. A próxima seção indicará os

possíveis próximos passos para escolha ou construção de um estudo de caso que inspire novos desenvolvedores a construir soluções baseadas em *blockchain* utilizando o FireFly.

1.6. Próximos passos

1.6.1. Construindo o seu caso de uso

Antes de tudo, a equipe de desenvolvimento precisa (*Passo 0*) decidir se o uso de *blockchain* é mesmo necessário. Como apresentado, decidir usar *blockchain* sem a real necessidade pode implicar no aumento desnecessário de complexidade e custo para o seu projeto. Em caso afirmativo, o próximo passo seria (*Passo 1*) definir o tipo de rede *blockchain* a ser utilizada, seja permissionada ou não, pública ou privada. Após esta decisão, é necessário (*Passo 2*) selecionar entre as opções que estiverem disponíveis a rede *blockchain* a ser utilizada. No momento de escrita deste capítulo, a rede privada permissionada mais utilizada é sem dúvidas a *Hyperledger Fabric*, enquanto a pública não-permissionada é *Ethereum*. É possível usar ambas na construção de *aplicações descentralizadas* com o *Hyperledger FireFly*, como visto anteriormente. Por fim, o time deve (*Passo 3*) projetar os casos de uso de sua aplicação que terão dados persistidos em *blockchain – on-chain* – e aqueles que não – *off-chain*.

1.6.2. Casos de uso existentes para inspiração

A expansão da tecnologia *Blockchain* para além de suas aplicações iniciais no setor financeiro criou uma ampla gama de oportunidades em diversos setores. Embora aproximadamente 30% das soluções que utilizam *blockchain* atualmente no mercado continuem focadas em finanças [Zile and Strazdiņa 2018], o potencial da tecnologia vai muito além desse setor. Zile e Strazdiņa, em um survey de 2018, exploram esse cenário mais amplo, apresentando uma variedade de soluções propostas e avaliando sua viabilidade. Isso inclui o uso da *blockchain* para aumentar a transparência e a segurança de ativos, produtos agroalimentares, dados e outros, demonstrando como a *Blockchain* está influenciando os setores ao melhorar a segurança, a eficiência e a confiança. Nesta seção, exploraremos alguns desses casos de uso da tecnologia

Transferência e rastreabilidade de propriedade

Por suas características de segurança e auditabilidade, a tecnologia *Blockchain* tem encontrado bom uso na transferência e rastreabilidade de propriedades de artigos de luxo. Nesse contexto, a Everledger teve início com a proposta de rastrear diamantes para garantir que não sejam provenientes de zonas de conflito. Construída sobre o *Hyperledger Fabric*, a solução da empresa combina *blockchain*, IA, IoT e nanotecnologia para criar um gêmeo digital seguro de cada diamante, detalhando sua procedência, caracterização e quem o possui. Além disso, a Everledger busca eliminar conflitos na produção dos diamantes e melhorar os direitos dos trabalhadores, promovendo a sustentabilidade, melhorando a rastreabilidade e a transparência, e a aumentando a confiança no setor de pedras preciosas por meio de uma avançada tecnologia de rastreamento de ativos.

Enquanto isso, Santos [Santos 2021] apresentou alternativas voltadas à compensação justa de pesquisadores por seus trabalhos, além de propostas para ajudar na adaptação

do processo de avaliação de artigos científicos por meio dos princípios de ciência aberta. Por outro lado, a Orvium está integrando tecnologia *Blockchain*, contratos inteligentes, computação em nuvem, análises de *big data* e aprendizado de máquina na busca por transformar o cenário da publicação científica e acadêmica. O uso do registro imutável da *blockchain* garante que todas as transações, como a publicação ou compra de uma publicação, sejam registradas de forma segura e não possam ser alteradas, proporcionando proteção de direitos autorais e reduzindo problemas de pirataria digital. À medida que a publicação digital evolui, a Orvium é um exemplo de como a *blockchain* pode aprimorar a segurança, a verificação de propriedade e a colaboração dentro da indústria, abrindo caminho para novos padrões na área.

Transferência e gerenciamento de dados

Já quanto a transferência e armazenamento de dados, tanto a Storj quanto a Filecoin são plataformas que utilizam a tecnologia *Blockchain* para melhorar o armazenamento de dados, oferecendo alternativas descentralizadas, seguras e eficientes aos serviços tradicionais de nuvem. Ambas as plataformas permitem que os usuários monetizem o espaço de armazenamento não utilizado em seus computadores focando em garantir a integridade e disponibilidade dos dados por meio de criptografia. Ao oferecer maior privacidade e redução de custos, essas plataformas propõem ampliar as possibilidades de armazenamento de dados e também democratizar o acesso e o controle sobre a informação.

Além disso, na área de gestão de dados de identidade, o *IBM Digital Credentials* é um sistema de identidade descentralizado e projetado para melhorar a troca segura de informações e dar a indivíduos e organizações controle sobre sua identidade e dados. Por meio da colaboração com líderes da indústria, como a *Decentralized Identity Foundation* e as *W3C Verifiable Credentials*, a IBM propõe garantir um alto nível de confiança e privacidade nas interações digitais. A plataforma, baseada em tecnologia Hyperledger, é construída em código aberto com foco em escalabilidade, segurança e interoperabilidade, e permite que as organizações adotem rapidamente processos de identidade e credenciamento descentralizados.

Rastreabilidade de cadeia de suprimentos agrícolas

O Walmart, gigante do varejo, se deparou com o desafio de rastrear rapidamente a origem de alimentos durante surtos de doenças transmitidas por alimentos, um processo que poderia levar semanas. Em busca de uma solução, a empresa identificou o potencial da tecnologia *blockchain* para melhorar a rastreabilidade no complexo ecossistema da cadeia de fornecimento de alimentos. Em parceria com a IBM, o Walmart desenvolveu um sistema de rastreabilidade baseado em Hyperledger Fabric e testou sua eficácia com dois projetos-piloto: um para rastrear mangas nos EUA e outro para rastrear carne suína na China. O sistema provou ser eficiente, reduzindo o tempo de rastreamento das mangas de 7 dias para apenas 2,2 segundos e possibilitando o registro de certificados de autenticidade para carne suína, aumentando a confiança no processo. Atualmente, o Walmart rastreia mais de 25 produtos de diferentes fornecedores e planeja expandir a aplicação da solução

para outras categorias, exigindo que fornecedores de verduras frescas também adotem o sistema.

Outros

No contexto de sistemas de saúde, a *Blockchain* pode ajudar a contornar o desafio de manter históricos médicos confiáveis e, ao mesmo tempo, proteger dados sensíveis dos pacientes e prestadores de serviço. Nesse âmbito, o *Synaptic Health Alliance* é uma coalizão de líderes de empresas fornecedoras de planos de saúde dos EUA (incluindo importantes representantes da Humana, Multiplan, UnitedHealth, entre outros), que utiliza o *Hyperledger FireFly* integrado ao Quorum para manter de forma segura uma base de dados descentralizada baseada em *blockchain*, visando abordar o desafio de manter dados precisos com segurança, privacidade e eficiência.

Além disso, a tecnologia *Blockchain* oferece uma série de benefícios para a Internet das Coisas (IoT), principalmente em termos de segurança, transparência e descentralização. [Chowdhury et al. 2020] revisaram alguns dos benefícios da tecnologia à área e apresentaram alguns casos de uso em que a adoção tecnologia é uma solução promissora. Segundo eles, essa união permitiria comunicação segura e direta entre dispositivos (M2M). Além disso, com contratos inteligentes, há mais eficiência e redução de custos. Casos de uso incluem a gestão de cadeias de suprimentos, onde a rastreabilidade dos produtos é melhorada; a automação em casas inteligentes; e a segurança em cidades inteligentes, com monitoramento descentralizado e em tempo real.

No âmbito industrial, a *TradeGo* é um consórcio internacional de transporte de mercadorias que através de um produto chamado *Digital Presentation* permite o compartilhamento eficiente e transparente de documentos comerciais originais e confidenciais em formato digital, agilizando operações comerciais. A adoção de *Blockchain* no intermédio dessas transações reduz e pode até eliminar o tempo gasto na transferência de documentos em papel, além de eliminar custos e desburocratizar todo o processo.

1.7. Considerações finais

Neste capítulo buscamos levar o leitor em uma jornada desde os conceitos fundamentais da tecnologia *blockchain* até a construção de suas primeiras aplicações *Web3* com o *Hyperledger FireFly*. Entendemos que a tecnologia é complexa e que existe um caminho longo até a compreensão de muitos dos detalhes aqui apresentados. Além disso, consideramos que a escassez de materiais didáticos escritos na língua portuguesa constitui uma barreira para o ingresso de desenvolvedores e pesquisadores brasileiros à área. Por esses motivos, a intenção deste capítulo foi diminuir tanto a barreira linguística, quanto da dispersão de conteúdo na Internet, que também pode afastar muitos dos interessados. Esperamos que este conteúdo ajude e motive estudantes nesta caminhada em direção a uma carreira de desenvolvedor de aplicações *Blockchain*.

O(a) leitor(a) pode dar continuidade aos estudos utilizando-se dos trabalhos aqui referenciados e, claro, da documentação do *Hyperledger FireFly* e das redes *Blockchain* escolhidas para sua aplicação. Entendemos que a documentação do *Hyperledger FireFly* é uma fonte de informações bastante rica, que contém detalhes sobre o seu funcionamento,

os quais não foram abordados aqui por limitação de espaço. Nesta mesma documentação, também existe bastante conteúdo sobre as redes *blockchain* mencionadas neste capítulo (*Hyperledger Fabric* e *Ethereum*), assim como exercícios interessantes. Além disso, acreditamos que existem diversas oportunidades de aprendizado com outras redes não trabalhadas aqui, tais como *Internet Computer*, *Cardano*, *Polygon*, entre outras. Independente do caminho que você escolher traçar, desejamos sucesso.

Referências

- [Baran 1964] Baran, P. (1964). On distributed communications networks. *IEEE Transactions on Communications Systems*, 12(1):1–9.
- [Buterin 2013] Buterin, V. (2013). Ethereum white paper: A next generation smart contract & decentralized application platform.
- [Chaum 1983] Chaum, D. (1983). Blind signatures for untraceable payments. In Chaum, D., Rivest, R. L., e Sherman, A. T., editors, *Advances in Cryptology*, pages 199–203, Boston, MA. Springer US.
- [Chowdhury et al. 2020] Chowdhury, M. J. M., Ferdous, M. S., Biswas, K., Chowdhury, N., e Muthukkumarasamy, V. (2020). A survey on blockchain-based platforms for iot use-cases. *The Knowledge Engineering Review*, 35:e19.
- [Dwork and Naor 1993] Dwork, C. e Naor, M. (1993). Pricing via processing or combatting junk mail. In Brickell, E. F., editor, *Advances in Cryptology — CRYPTO’ 92*, pages 139–147, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Enoch et al. 2023] Enoch, J. D., Omijeh, B. O., e Okeke, R. O. (2023). Comparative analysis of blockchain and database. *European Journal of Science, Innovation and Technology*, 3(3):454–463.
- [Foundation 2016] Foundation, H. (2016). Open source blockchain effort for the enterprise elects leadership positions and gains new investments. <https://www.hyperledger.org/announcements/2016/03/29/open-source-blockchain-effort-for-the-enterprise-elects-leadership-positions-and-gains-new-investments>. Accessed June 2024.
- [Foundation 2024] Foundation, H. (2024). An overview of hyperledger foundation. https://8112310.fs1.hubspotusercontent-na1.net/hubfs/8112310/Hyperledger/Offers/HL_Whitepaper_IntroductiontoHyperledger.pdf. Accessed: 18 May 2024.
- [Greve et al. 2018] Greve, F., Sampaio, L., Abijaude, J., Coutinho, A. A., Brito, I., e Queiroz, S. (2018). *Blockchain e a Revolução do Consenso sob Demanda*, page v. 30. SBC.
- [Haber and Stornetta 1991] Haber, S. e Stornetta, W. S. (1991). How to time-stamp a digital document. *J. Cryptology*, 3:99–111.

- [Hassija et al. 2021] Hassija, V., Zeadally, S., Jain, I., Tahiliani, A., Chamola, V., e Gupta, S. (2021). Framework for determining the suitability of blockchain: Criteria and issues to consider. *Transactions on Emerging Telecommunications Technologies*, 32(10):e4334.
- [ISO Central Secretary 2024] ISO Central Secretary (2024). Blockchain and distributed ledger technologies — vocabulary. Standard ISO 22739:2024, International Organization for Standardization, Geneva, CH.
- [Kakavand et al. 2017] Kakavand, H., Sevres, N. K. D., e Chilton, B. (2017). The blockchain revolution: An analysis of regulation and technology related to distributed ledger technologies. *SSRN*. Disponível em SSRN: <https://ssrn.com/abstract=2849251> ou <http://dx.doi.org/10.2139/ssrn.2849251>.
- [Nakamoto 2008] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. <https://bitcoin.org/bitcoin.pdf>. Accessed: 2021-Jun-09.
- [Narayanan and Clark 2017] Narayanan, A. e Clark, J. (2017). Bitcoin’s academic pedigree: The concept of cryptocurrencies is built from forgotten ideas in research literature. *Queue*, 15(4):20–49.
- [Santos 2021] Santos, A. R. (2021). Open scientist in the wonderland: advocating for blockchain-based decentralized applications for science. In *Proceedings of the 1st Workshop on Open Science Practices for Software Engineering*, pages 1–3, Porto Alegre, RS, BRA. SBC OpenLib.
- [Szabo 2008] Szabo, N. (2008). Bit gold. <https://unenumerated.blogspot.com/2005/12/bit-gold.html>. Acesso em: 02 jul. 2024.
- [The Economist 2015] The Economist (2015). The trust machine. <https://www.economist.com/leaders/2015/10/31/the-trust-machine>. Accessed: 2024-05-14.
- [Wood 2019] Wood, G. (2019). Ethereum: A secure decentralised generalised transaction ledger.
- [Wust and Gervais 2018] Wust, K. e Gervais, A. (2018). Do you need a blockchain? In *2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*, pages 45–54. IEEE.
- [Zheng et al. 2018] Zheng, Z., Xie, S., Dai, H.-N., Chen, X., e Wang, H. (2018). Blockchain challenges and opportunities: a survey. *International Journal of Web and Grid Services*, 14(4):352–375.
- [Zīle and Strazdiņa 2018] Zīle, K. e Strazdiņa, R. (2018). Blockchain use cases and their feasibility. *Applied Computer Systems*, 23(1):12–20.

Capítulo

2

Tiny ML: Introdução ao Aprendizado de Máquina em Sistemas Embarcados

Ramon Santos Nepomuceno, Luís Fagner de Carvalho da Silva, Dorgival Pereira da Silva Netto, Rafael Perazzo Barbosa Mota Carlos, Carlos Vinicius Gomes Costa Lima e Carlos Julian Menezes Araujo

Abstract

This chapter provides a practical introduction to Machine Learning in Embedded Systems, with a focus on the Edge Impulse platform. It covers fundamental concepts and techniques for developing machine learning models for embedded devices, including practical applications using smartphones and TinyML kits. The practical sessions address data collection, model training, and deployment on real devices, showcasing the technology's applicability. Additionally, the chapter explores theoretical aspects of Tiny Machine Learning, such as Impulse configuration, spectral feature extraction, and classifier training for resource-constrained systems. By the end, readers will be equipped to implement and test machine learning models on microcontrollers and low-power devices, leveraging Edge Impulse to effectively bridge theory and practice.

Resumo

Este capítulo oferece uma introdução prática ao Aprendizado de Máquina em Sistemas Embarcados com ênfase na plataforma Edge Impulse. O texto explora conceitos fundamentais e técnicas para o desenvolvimento de modelos de aprendizado de máquina em dispositivos embarcados, incluindo o uso de smartphones e kits TinyML. A abordagem prática inclui a coleta de dados, treinamento de modelos e implantação em dispositivos reais, com exemplos que demonstram a aplicabilidade da tecnologia. Além disso, o capítulo aborda aspectos teóricos do Tiny Machine Learning, como a configuração do Impulse, extração de features espectrais e treinamento de classificadores para sistemas com recursos limitados. Ao final, os leitores estarão aptos a implementar e testar modelos de aprendizado de máquina em microcontroladores e dispositivos de baixo consumo de energia, utilizando o Edge Impulse para integrar teoria e prática de forma eficaz.

2.1. Introdução

A crescente interconexão entre dispositivos e sistemas está revolucionando setores como saúde e transporte, onde a capacidade de processar e analisar dados de forma autônoma se tornou essencial. Neste contexto, a computação na borda, ou edge computing, se destaca como uma solução para atender à crescente demanda por análise em tempo real, reduzindo a dependência de conexões de rede instáveis. Este paradigma permite o processamento local de dados nos dispositivos, diminuindo a latência e ampliando as capacidades da Inteligência Artificial (IA) [3].

A aplicação de Inteligência Artificial (AI) em sistemas embarcados, ou edge AI, viabiliza uma ampla gama de aplicações, desde diagnósticos médicos imediatos até a otimização de processos industriais. O processamento local possibilita a realização de inferências e a tomada de decisões em tempo real, melhorando a eficiência, a velocidade e a segurança dos sistemas. Adicionalmente, processar dados localmente contribui para a preservação da privacidade e da segurança, pois informações sensíveis não são transmitidas a servidores externos [18].

O conceito de Tiny Machine Learning (Tiny ML) refere-se à execução de modelos de aprendizado de máquina em dispositivos com recursos limitados, como microcontroladores e sensores. Tiny ML viabiliza a implementação de funcionalidades de IA em hardware com baixo consumo de energia e pouca memória, expandindo significativamente as capacidades da computação na borda [13].

Este capítulo apresenta o desenvolvimento de projetos de IA em sistemas embarcados, introduzindo os fundamentos do aprendizado de máquina e explorando a computação na borda. A plataforma Edge Impulse é utilizada para fornecer um guia prático para a criação e implantação de modelos de aprendizado de máquina em dispositivos reais [6]. Este processo visa demonstrar o potencial da Inteligência Artificial nesse cenário, esclarecendo as vantagens e os desafios dessa área emergente.

2.2. Projetos Práticos com a Plataforma Edge Impulse

A Edge Impulse é uma plataforma projetada para facilitar o desenvolvimento e a implementação de modelos de aprendizado de máquina em dispositivos embarcados [6]. Ela oferece ferramentas que permitem que desenvolvedores e engenheiros criem soluções rapidamente, integrando coleta de dados, processamento de sinais e aprendizado de máquina em um único ambiente.

A plataforma possibilita projetos que utilizam diversos sensores, desde projetos mais simples como os que envolvem gestos ou medição de temperatura, até mais complexos, como projetos com áudio e visão computacional. O restante do capítulo será focado no reconhecimento de movimentos usando a plataforma até a implementação em dispositivos mobile e no kit TinyML.

2.2.1. Projeto de Reconhecimento de Movimento

Nas seções seguintes, será apresentado um projeto de reconhecimento contínuo de movimentos (Continuous Motion Recognition). O objetivo é identificar e classificar diferentes tipos de movimentos em tempo real. Utilizando sensores como acelerômetros ou

giroscópios para capturar dados de movimento, o sistema detectará padrões específicos e os classificará em categorias predefinidas. Durante o desenvolvimento deste projeto, os principais conceitos relacionados à inteligência artificial em dispositivos embarcados (Edge AI) serão abordados de maneira prática e didática.

2.2.1.1. Criando uma Conta e Iniciando um Projeto

Para começar, acesse <https://www.edgeimpulse.com> e crie uma conta. Uma vez logado, você pode iniciar um novo projeto clicando em *Create New Project*. Como o projeto envolve a detecção de movimento, sugere-se escolher um nome que reflita essa finalidade, como *MotionTrack*. Isso não só ajudará a identificar facilmente o projeto mais tarde, mas também dará um contexto claro sobre seu objetivo. Após nomear o projeto, o *dashboard* será aberto, permitindo gerenciar todas as fases do projeto, desde a coleta de dados até o treinamento e implantação do modelo.

A Figura 2.1 mostra a tela inicial de um projeto na plataforma *Edge Impulse*. O *dashboard* oferece uma visão das funcionalidades e ferramentas disponíveis, integrando os passos necessários para a criação e implantação dos modelos.

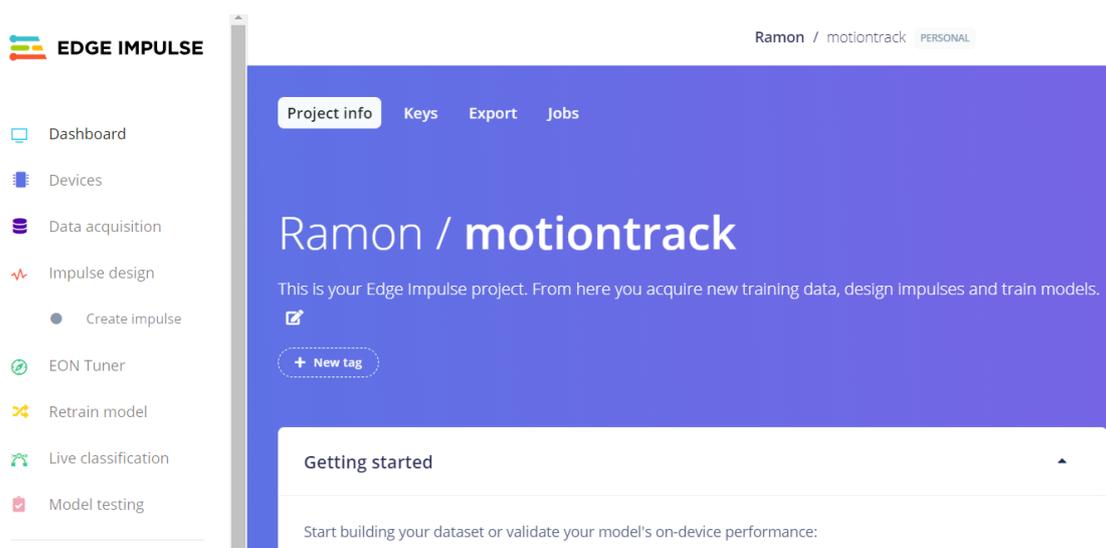


Figura 2.1. Dashboard da plataforma Edge Impulse.

2.2.1.2. Conectando Dispositivos e Coletando Dados

Após criar o projeto, clique em *Devices* e selecione *Connect a New Device* para escolher o dispositivo a ser utilizado. Nesta etapa inicial, será utilizado um smartphone como dispositivo alvo. Conecte-o escaneando o código QR exibido na tela, como mostra a Figura 2.2.

A criação da base de dados é um dos primeiros passos no desenvolvimento dos modelos, pois fornece os exemplos necessários para que o modelo aprenda a identificar padrões e tomar decisões. Em termos simples, um modelo de aprendizado de máquina

analisa dados de entrada, reconhece padrões e faz previsões ou decisões com base nesses padrões. Para funcionar corretamente, o modelo deve ser treinado com um conjunto de dados que represente as variáveis e condições que ele encontrará na aplicação real. Durante o treinamento, o modelo ajusta seus parâmetros internos para minimizar o erro entre suas previsões e os resultados reais, melhorando sua precisão ao longo do tempo.

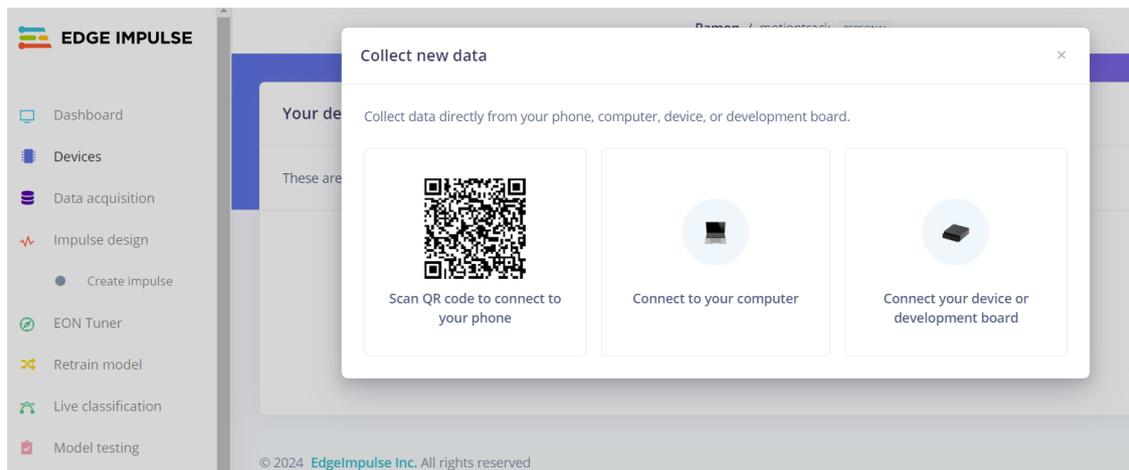


Figura 2.2. Adicionando dispositivo no Edge Impulse.

Após conectar o dispositivo, vá para *Data Acquisition*. No campo de seleção do tipo de sensor, escolha *Acelerômetro*. No campo *Label*, insira o tipo de movimento que deseja registrar; comece com *left_right*. Clique em *Start Sampling* e mova o celular da esquerda para a direita por 10 segundos. Repita esse processo 20 vezes para coletar as amostras necessárias.

Depois de completar as coletas para *left_right*, repita o procedimento para os movimentos *up_down* (para cima e para baixo), *circle* (em círculos) e *idle* (com o celular em repouso). Esse processo resultará em um total de 80 amostras, com 20 para cada classe de movimento. A Figura 2.3 exemplifica a tela de coleta de dados.

Após a aquisição das amostras, é necessário dividir os dados em conjuntos de treinamento e teste. O conjunto de treinamento é utilizado para ajustar os parâmetros do modelo, enquanto o conjunto de teste é empregado para avaliar a capacidade do modelo de generalizar para novos dados. Esse processo permite verificar a eficácia do modelo em identificar e classificar amostras não vistas anteriormente e detectar padrões que podem representar maior dificuldade para a inferência correta. Recomenda-se uma divisão de 80% para o conjunto de treinamento e 20% para o conjunto de teste. O Edge Impulse oferece ferramentas automáticas para realizar essa divisão, mas também é possível fazê-la manualmente, clicando nos três pontos de cada amostra e movendo-a para o conjunto de teste.

2.2.1.3. Projetando o Impulso

Após a coleta de dados, é possível criar o Impulse clicando em *Impulse Design*. Um Impulse completo consiste em um pipeline de aprendizado de máquina composto por

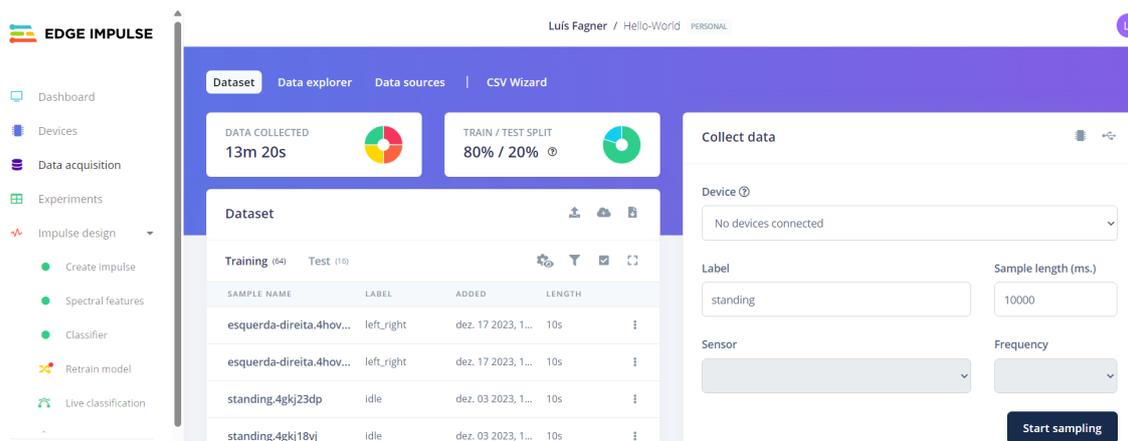


Figura 2.3. Coletando dados na plataforma Edge Impulse.

três principais componentes: **Bloco de Entrada**, **Bloco de Processamento** e **Bloco de Aprendizado**, conforme exemplificado na Figura 2.4.

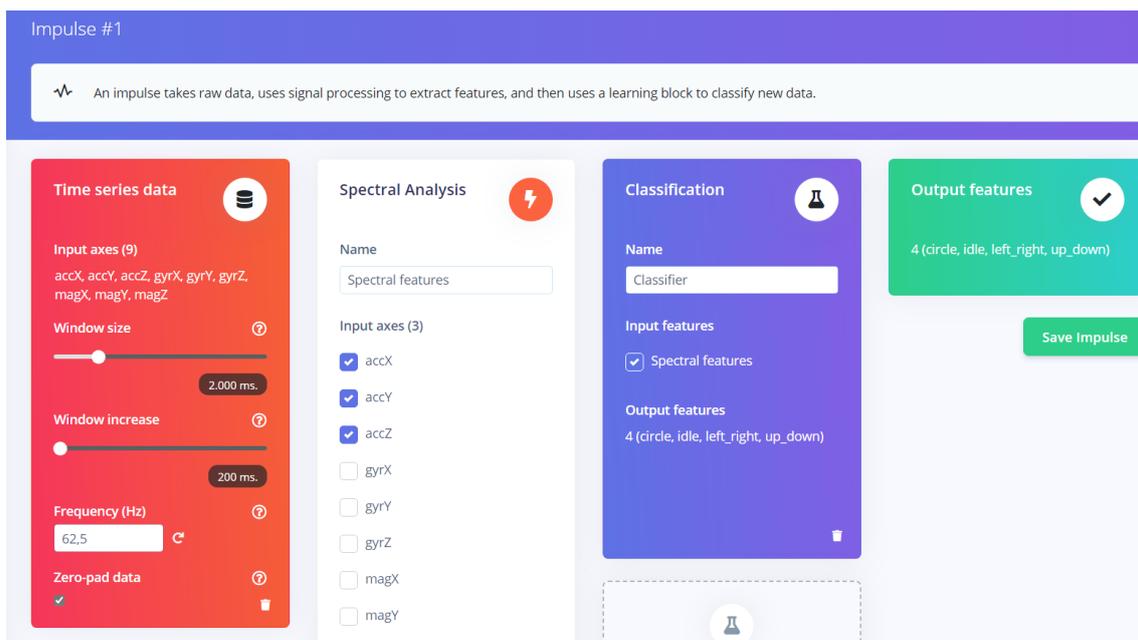


Figura 2.4. Projeto do Impulse.

- **Bloco de Entrada (Input Block):** Define como os dados são coletados e estruturados para o treinamento do modelo. Este bloco especifica o tipo de dados que será utilizado, como séries temporais (áudio, vibração, movimentos) ou imagens. No caso de séries temporais, o bloco indica quais eixos são referenciados a partir do conjunto de dados de treinamento, além de configurar o tamanho da janela, que determina o tamanho das características brutas usadas para o treinamento. O aumento da janela é utilizado para criar artificialmente mais características e fornecer mais informações ao bloco de aprendizado. A frequência dos dados é calculada automaticamente com base nas amostras de treinamento, embora possa ser ajustada, desde

que não seja menor que 0,000016 (menos de 1 amostra a cada 60 segundos). Além disso, o bloco pode adicionar valores zero quando características brutas estão ausentes, garantindo que o modelo receba um fluxo contínuo de dados. Para o projeto em questão, deixe as configurações padrões.

- **Bloco de Processamento (Processing Block):** Realiza o pré-processamento dos dados para torná-los adequados para a etapa de aprendizado. Este bloco atua como um extrator de características, utilizando operações de Processamento de Sinais Digitais (DSP) para extrair informações relevantes dos dados brutos. Clique em *Add a Processing Block*, e selecione *Spectral Analysis*. Este bloco extrai informações sobre frequência, potência e outras características de um sinal. Filtros passa-baixa e passa-alta podem ser aplicados para remover frequências indesejadas, sendo especialmente úteis para examinar padrões repetitivos como movimentos ou vibrações captados pelo acelerômetro. A escolha do bloco de processamento recomendado é facilitada pela interface do Edge Impulse, que utiliza uma estrela para indicar a opção mais adequada com base nos dados de entrada.
- **Bloco de Aprendizado (Learning Block):** Treina o modelo de aprendizado de máquina utilizando os dados processados. Este bloco é essencial para ajustar os parâmetros do modelo e permitir que ele faça previsões precisas. Após adicionar o bloco de processamento, o próximo passo é adicionar um bloco de aprendizado. Clique em *Add Learning Block* e selecione *Classification* para adicionar um classificador de rede neural. O bloco de aprendizado pode variar conforme o objetivo do modelo e o tipo de dados no conjunto de treinamento, incluindo algoritmos para classificação, regressão, detecção de anomalias, transferência de aprendizado em imagens, reconhecimento de palavras-chave ou detecção de objetos. Após selecionar o tipo de aprendizado, clique em *Save Impulse* para criar o impulso. O bloco de aprendizado selecionado será configurado para aprender com os dados extraídos no bloco de processamento, permitindo a criação de um modelo que atenda às necessidades específicas do projeto.

2.2.1.4. Gerando *Features Espectrais*

Após configurar o bloco de aprendizado, o próximo passo no projeto de detecção de movimento é gerar as *features espectrais*. Isso é feito clicando em *Spectral Features*, seguido de *Save Parameters* e, finalmente, *Generate Features*. As *features espectrais* são características extraídas de um sinal no domínio da frequência, que incluem informações sobre frequência, potência e outras propriedades. Esse processo é essencial para identificar padrões repetitivos, como movimentos ou vibrações captados por um acelerômetro, pois permite a identificação de componentes frequenciais significativos que podem não ser evidentes na análise temporal.

A análise espectral, ao converter os dados do domínio do tempo para o domínio da frequência, permite eliminar frequências indesejadas por meio de filtros passa-baixa e passa-alta, aprimorando a qualidade dos dados de entrada para o modelo de aprendizado de máquina [8]. O bloco de *features espectrais* é particularmente útil para sinais que apre-

sentam padrões irregulares ou transitórios permitindo que o modelo detecte e classifique movimentos com maior precisão.

O bloco de análise espectral oferece várias opções de configuração, como mostrado na Figura 2.5:

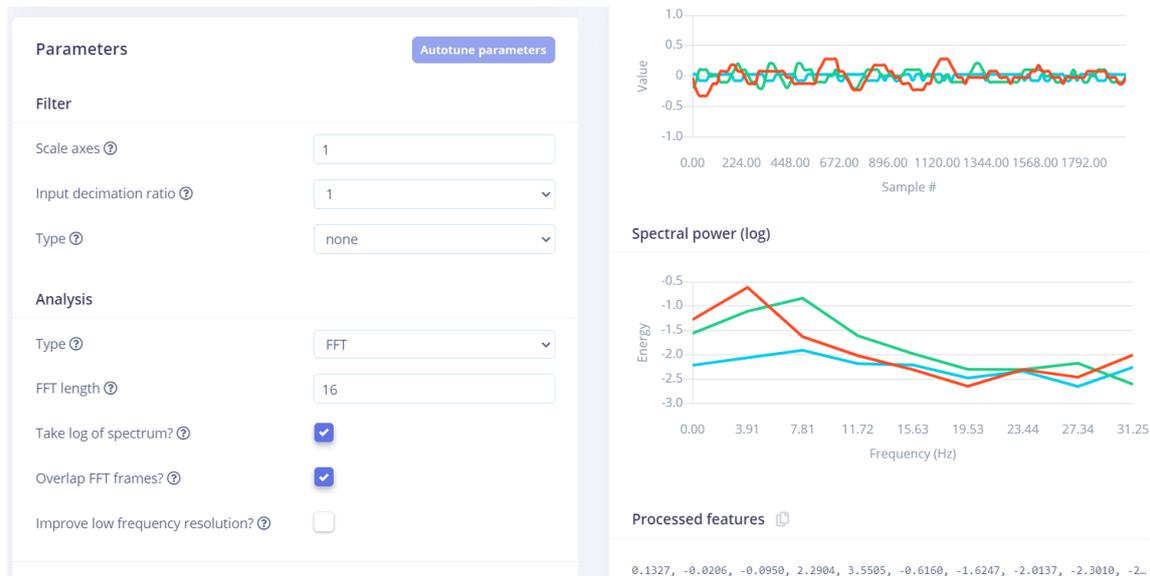


Figura 2.5. Análise de Features na Edge Impulse.

- **Filtro:** Antes de calcular a Transformada Rápida de Fourier (FFT) [23], os dados de séries temporais podem ser filtrados para suavizar o sinal ou remover artefatos indesejados. A janela de amostragem utilizada no filtro é ajustada com base nos parâmetros do bloco de entrada.
- **Escala de eixos:** Multiplica todos os valores de entrada brutos por um número específico, ajustando a amplitude dos dados.
- **Razão de dizimação de entrada:** Reduz a taxa de amostragem do sinal para diminuir a quantidade de features, melhorando a resolução de frequência sem aumentar o uso de recursos.
- **Tipo de filtro:** Escolha entre passa-baixa, passa-alta ou sem filtro para aplicar ao dado bruto.
- **Frequência de corte:** Define a frequência de corte em hertz, removendo frequências indesejadas e otimizando o tamanho do modelo.
- **Ordem:** A ordem do filtro Butterworth, determinando a precisão do corte de frequências. Um valor mais alto resulta em um corte mais acentuado, mas aumenta a latência.

A análise de potência espectral é uma abordagem baseada na FFT, e a configuração dos parâmetros, como o tamanho da FFT e a sobreposição dos frames, permite a extração precisa das características desejadas.

- **Tamanho da FFT:** Define a resolução de frequência e a quantidade de bins da FFT. Um tamanho menor reduz a quantidade de features e o tamanho do modelo, enquanto um tamanho maior permite separar mais sinais.
- **Logaritmo do espectro:** Aplica logaritmo aos bins da FFT, expandindo o alcance de sinais de baixa intensidade.
- **Sobreposição de frames da FFT:** Utiliza uma abordagem de *sliding frame* para evitar a perda de eventos transitórios entre frames, sendo ativado por padrão.

Para cada bin da FFT, o valor máximo é mantido como feature, enquanto frequências indesejadas são descartadas, otimizando o modelo para treinamento com menos dados.

2.2.1.5. Treinando o Classificador

Com as features espectrais geradas, o próximo passo é treinar o modelo. Isso é feito clicando em *Classifier* e depois em *Start Training*, como mostrado na Figura 2.6. É nessa etapa que o modelo aprende a diferenciar entre as classes de movimento com base nos dados de entrada processados. Um classificador recebe dados de entrada e fornece uma pontuação de probabilidade que indica a probabilidade de os dados pertencerem a uma classe específica. No contexto do projeto de detecção de movimento, isso significa identificar padrões de movimento a partir dos dados do acelerômetro.

A Edge Impulse oferece vários blocos de treinamento predefinidos, como por exemplo:

- **Classificação (Keras):** Ideal para diferenciar entre várias classes, como diferentes tipos de movimentos [4].
- **Regressão (Keras):** Para prever valores contínuos [4].
- **Detecção de Anomalias (K-means, GMM):** Para identificar padrões incomuns nos dados [5].
- **Classificação de Imagem (Transfer Learning, NVIDIA TAO):** Para reconhecimento de imagens [12].
- **Detecção de Objetos (MobileNetV2, SSD, FPN, FOMO):** Para localizar objetos em imagens [17] [25].
- **Reconhecimento de Palavras-chave:** Para identificar palavras específicas em áudio.

Para a tarefa de detecção de movimento, o bloco de *Classifier* é o mais indicado.

O classificador utilizado é uma rede neural, composta por múltiplas camadas de neurônios. Cada neurônio está interconectado com os neurônios da camada subsequente,

The screenshot displays the 'Neural Network settings' page in the Edge Impulse web interface. On the left is a navigation sidebar with the following items: Dashboard, Devices, Data acquisition, Impulse design (with sub-items: Create impulse, Spectral features, Classifier), EON Tuner, Retrain model, Live classification, Model testing, Versioning, and Deployment. Below the sidebar is an 'Upgrade Plan' section with a star icon and text: 'Get access to higher job limits, collaborators and a full commercial license.'

The main content area is titled 'Neural Network settings' and is divided into several sections:

- Training settings:**
 - Number of training cycles: 30
 - Use learned optimizer:
 - Learning rate: 0.0005
 - Training processor: CPU
- Advanced training settings:** (Collapsible section, currently expanded)
- Neural network architecture:**
 - Input layer (39 features)
 - Dense layer (20 neurons)
 - Dense layer (10 neurons)
 - Add an extra layer (dashed box)
 - Output layer (2 classes)

At the bottom center of the main area is a green 'Start training' button.

Figura 2.6. Etapa de Treinamento na Edge Impulse.

e os pesos dessas conexões são ajustados durante o treinamento. Inicialmente, os pesos são definidos aleatoriamente. Durante o processo de treinamento, a rede é alimentada com um conjunto de dados de treinamento, e a saída gerada pela rede é comparada com as respostas corretas. Com base nessa comparação, os pesos das conexões são ajustados por meio de um processo chamado *backpropagation* [15].

Essa arquitetura permite que a rede aprenda a mapear os dados de entrada para as classes corretas, aprimorando sua capacidade preditiva a cada iteração.

Na página de configuração do classificador, diversas opções estão disponíveis para otimizar o processo de treinamento, por exemplo:

- **Número de Ciclos de Treinamento:** Define quantas vezes o algoritmo percorrerá todos os dados de treinamento para ajustar os parâmetros do modelo.
- **Taxa de Aprendizado:** Controla a quantidade de ajuste dos parâmetros internos do modelo em cada passo. Uma taxa de aprendizado muito alta pode levar a um *overfitting* rápido.
- **Tamanho do Conjunto de Validação:** Percentual do conjunto de treinamento reservado para validação, geralmente 20%.
- **Tamanho do Lote (Batch Size):** Número de amostras processadas antes de atualizar o modelo.
- **Autoajuste de Classes:** Foca mais em amostras de classes sub-representadas durante o treinamento para evitar *overfitting*.

A arquitetura da rede neural recebe as features extraídas como entrada e as processa através de suas camadas. No caso da classificação, a última camada usada é uma camada *softmax* [10], que fornece a probabilidade de pertencimento a uma das classes. Em modo visual, é possível adicionar diferentes tipos de camadas para ajustar a arquitetura conforme necessário.

Para usuários avançados, a Edge Impulse oferece um modo expert, permitindo acessar a API completa do Keras para customizar a arquitetura, ajustar a função de perda, o otimizador e até mesmo implementar *callbacks* para evitar o *overfitting*.

Overfitting ocorre quando o modelo se ajusta excessivamente aos dados de treinamento, capturando também o ruído e os detalhes irrelevantes, o que resulta em um desempenho ruim em novos dados. Prevenir o *overfitting* garante que o modelo seja mais generalizável e eficaz em diferentes conjuntos de dados [24].

Durante o treinamento, o painel de saída exibe os logs, permitindo acompanhar o progresso do modelo. Após o treinamento, a performance do modelo pode ser avaliada através de algumas métricas, como mostrado nas Figuras 2.7 e 2.8, são elas:

- **Acurácia de Validação e Função de Perda:** A acurácia de validação e a função de perda são métricas para avaliar o desempenho do modelo. A acurácia de validação mede a proporção de previsões corretas realizadas pelo modelo sobre o conjunto

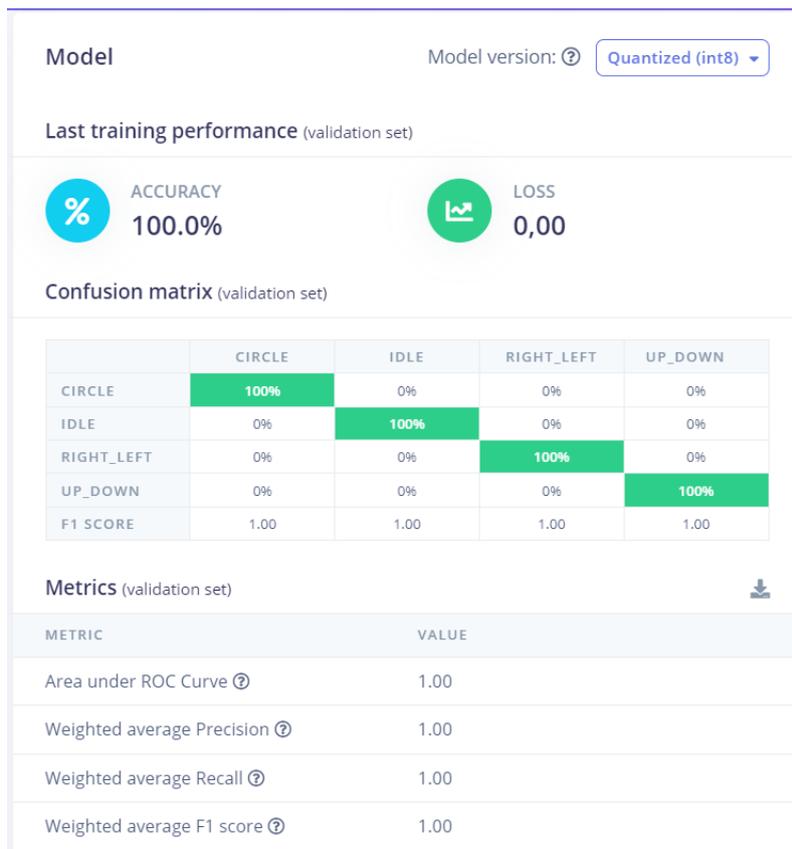


Figura 2.7. Resultado do treinamento do Impulso.

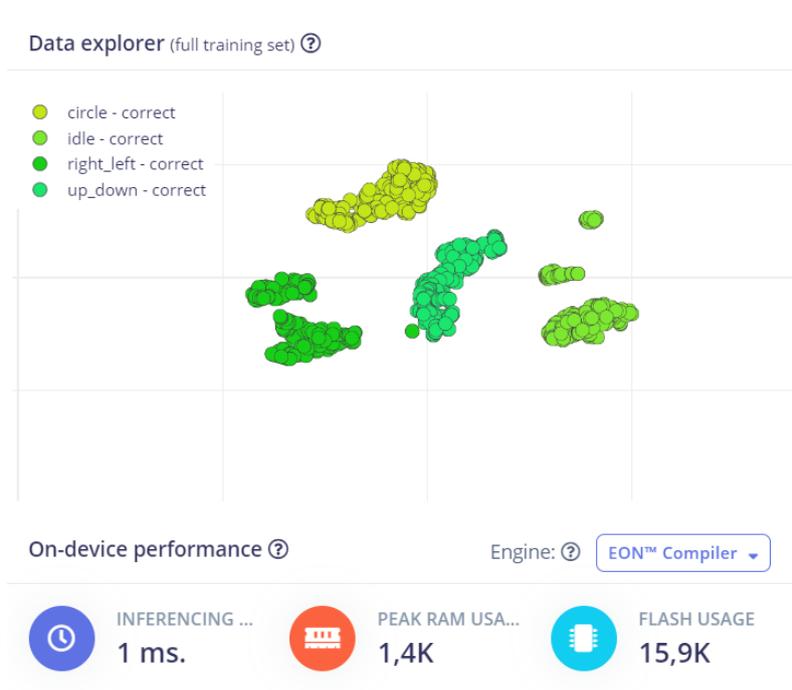


Figura 2.8. Métricas de consumo no dispositivo.

de validação, enquanto a função de perda quantifica o erro de previsão, refletindo a diferença entre as previsões do modelo e os valores reais.

- **Matriz de Confusão:** A matriz de confusão é uma ferramenta útil para a avaliação do desempenho do modelo, fornecendo uma representação detalhada das previsões corretas e incorretas. Ela permite a análise das taxas de verdadeiro positivo, falso positivo, verdadeiro negativo e falso negativo, facilitando a compreensão das forças e fraquezas do modelo.
- **Explorador de Características:** O explorador de características permite a visualização da distribuição das variáveis de entrada no espaço de características. Ele mostra quais variáveis foram corretamente classificadas e quais apresentaram dificuldades, oferecendo insights sobre a eficácia do modelo em distinguir entre diferentes classes e a qualidade das informações extraídas.

Com base no dispositivo alvo selecionado na página do Dashboard, estimativas de tempo de inferência, uso máximo de RAM e uso de flash são fornecidas, ajudando a validar se o modelo pode ser executado no dispositivo de destino dentro de suas limitações, como mostrado na Figura [2.8](#).

Este processo de treinamento do classificador é uma etapa importante no desenvolvimento do modelo de detecção de movimento, garantindo que o modelo possa aprender eficazmente a partir dos dados de treinamento e generalizar para novos dados de forma precisa e eficiente.

2.2.1.6. Implantação do Modelo no Smartphone

Após o treinamento do modelo de aprendizado de máquina, é possível implantá-lo em um *smartphone* para realização de teste. Para isso, acesse a seção *Deployment* no painel do Edge Impulse e selecione *WebAssembly (browser, SIMD)* como o formato de implantação. Em seguida, use a câmera de um *smartphone* para escanear o QR code, que o redirecionará para uma página onde o modelo estará ativo para teste imediato, como mostrado na Figura [2.9](#). O uso do WebAssembly oferece algumas vantagens:

1. compatibilidade com todos os principais navegadores, permitindo execução sem instalação adicional;
2. desempenho otimizado para operar de forma rápida e eficiente;
3. portabilidade, eliminando a dependência de plataformas específicas e aumentando a flexibilidade da solução.

Com a implantação no *smartphone*, o dispositivo se torna um sensor de movimento, demonstrando a aplicabilidade prática do aprendizado de máquina em dispositivos móveis e destacando o potencial para soluções para o mundo real.

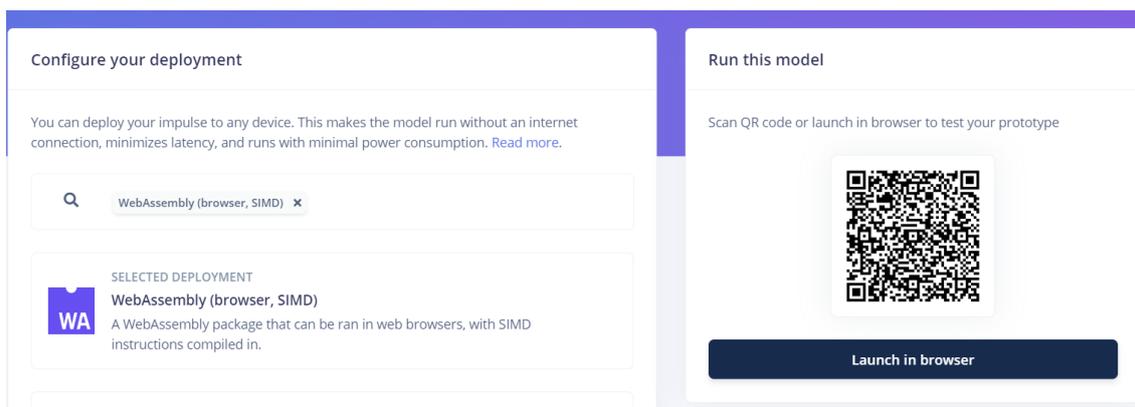


Figura 2.9. Tela de deploy no smartphone.

2.2.2. Reconhecimento de Movimento com o Kit TinyML

Nessa seção, o processo de detecção de movimento descrito anteriormente será repetido, mas agora utilizando o kit TinyML da Edge Impulse [14]. TinyML refere-se à implementação de modelos de aprendizado de máquina em dispositivos com recursos computacionais limitados, como microcontroladores e sensores integrados [21]. Essa tecnologia permite que algoritmos complexos sejam executados diretamente em dispositivos de borda, possibilitando a coleta e o processamento de dados em tempo real sem a necessidade de conexão com a nuvem.

O kit Arduino Tiny Machine Learning da Edge Impulse, ilustrado na Figura 2.10, é um exemplo dessa tecnologia. Ele inclui um microcontrolador otimizado para machine learning, sensores integrados e um ambiente de desenvolvimento que facilita a criação e a implantação de modelos de aprendizado de máquina diretamente no dispositivo. Este kit é usado em aplicações que requerem baixo consumo de energia, resposta rápida e operação autônoma, como na Internet das Coisas (IoT), automação industrial e dispositivos vestíveis.



Figura 2.10. Kit Arduino Tiny Machine Learning Edge Impulse.

Esta seção aborda configuração do kit TinyML, coleta de dados, treinamento do modelo e implantação da solução para reconhecimento de movimento.

2.2.2.1. Conectando o Kit TinyML e Coletando Dados

Para iniciar a coleta de dados com o kit TinyML da Edge Impulse, o primeiro passo é criar um novo projeto na plataforma Edge Impulse. Recomenda-se nomear o projeto de forma descritiva, como “Detecção de Movimento com Arduino Nano 33 BLE Sense”.

Com o projeto criado, siga os passos abaixo para conectar o dispositivo e coletar os dados:

1. **Conectando o Dispositivo:** Para conectar o Arduino Nano 33 BLE Sense ao computador, use um cabo micro-USB. Em seguida, pressione o botão RESET duas vezes rapidamente para iniciar o bootloader. O LED da placa deve começar a piscar, indicando que o dispositivo está pronto para a conexão.
2. **Atualizando o Firmware:** A placa não vem com o firmware adequado pré-instalado. Para atualizá-lo, baixe o firmware mais recente da Edge Impulse e descompacte o arquivo. Execute o script de flash apropriado para seu sistema operacional (flash_windows.bat, flash_mac.command ou flash_linux.sh) para atualizar o firmware da placa. Após a conclusão do processo, pressione o botão RESET uma vez para iniciar o novo firmware.
3. **Configurando as Chaves:** Abra um prompt de comando ou terminal e execute o comando edge-impulse-daemon. Isso iniciará um assistente que solicitará o login e a seleção do projeto Edge Impulse. Caso deseje mudar de projeto, adicione a opção –clean ao comando. Em versões recentes dos navegadores Google Chrome e Microsoft Edge, também é possível coletar dados diretamente do seu dispositivo sem a necessidade do CLI da Edge Impulse.

Para a coleta de dados, siga os mesmos passos descritos na Seção [2.2.1.2](#). No Edge Impulse, clique em *Data Acquisition*. No campo de seleção do tipo de sensor, escolha *Acelerômetro*. No campo *Label*, insira o tipo de movimento que deseja registrar e colete dados para todos os movimentos desejados. Repita o processo para cada tipo de movimento a ser registrado.

2.2.3. Projetando Impulso, Gerando as *Features Espectrais* e Treinando o Classificador para o Kit TinyML

Após a coleta de dados para o projeto, os seguintes passos envolvem a configuração do Impulse, a extração das features espectrais e o treinamento do classificador. Para realizar esses procedimentos, consulte as Seções [2.2.1.3](#), [2.2.1.4](#), e [2.2.1.5](#).

Inicialmente, crie o Impulse acessando a opção *Impulse Design*. Adicione um bloco de processamento selecionando *Spectral Analysis*, e em seguida, insira um bloco de aprendizado escolhendo *Classification* para incluir um classificador baseado em rede neural. Após a configuração dos blocos, finalize o processo clicando em *Save Impulse*.

Com o Impulse configurado, prossiga para a geração das features espectrais clicando em *Spectral Features*, seguido de *Save Parameters* e, finalmente, *Generate Features*. Este processo extrai características frequenciais e outras propriedades relevantes do sinal, essenciais para a análise dos dados.

Finalmente, para treinar o modelo, selecione *Classifier* e clique em *Start Training*. O treinamento ajustará o classificador para identificar padrões de movimento com base nas *features espectrais* geradas, conforme ilustrado na Figura 2.6.

2.2.3.1. Implantação no Kit Arduino TinyML

Para implantar o modelo treinado no kit Arduino TinyML, siga os seguintes passos:

1. **Implantação do Modelo:** No Edge Impulse, vá para a seção *Deployment* e selecione o dispositivo *Arduino Nano 33 BLE Sense*. Clique em *Build Model* para compilar o modelo treinado para o dispositivo.
2. **Atualização do Dispositivo:** Após a compilação, baixe o arquivo de firmware gerado e execute o script de flash correspondente ao seu sistema operacional (por exemplo, *flash_windows.bat* para Windows, *flash_mac.command* para macOS, ou *flash_linux.sh* para Linux) para atualizar o firmware no Arduino Nano 33 BLE Sense.
3. **Execução do Modelo:** Após a atualização do firmware, execute o comando *edge-impulse-run-impulse* no terminal ou prompt de comando do computador ao qual o dispositivo está conectado. O Arduino Nano 33 BLE Sense começará a realizar inferências em tempo real e exibirá os resultados no console.

Com o modelo treinado implantado no kit TinyML, o dispositivo está pronto para realizar inferências em tempo real, demonstrando a capacidade de detecção de movimento diretamente na borda. Este processo ilustra como tecnologias de aprendizado de máquina podem ser integradas em sistemas embarcados para aplicações eficientes e autônomas. No entanto, para entender plenamente as possibilidades e os desafios do TinyML, é essencial aprofundar-se nos conceitos fundamentais e nas abordagens que tornam essa tecnologia viável. A Seção 2.3 explora esses conceitos em detalhes, proporcionando uma base sólida para o desenvolvimento e a aplicação de soluções TinyML.

2.3. Conceitos Fundamentais e Abordagens em TinyML

No campo da tecnologia, termos como *Edge AI* e *TinyML* tornaram-se comuns. Para compreender plenamente essas inovações e suas aplicações, é essencial explorar os conceitos fundamentais que as sustentam. Esta seção, fundamentada no trabalho apresentado em [19], aborda os seguintes temas centrais: Sistemas Embarcados, Computação na Borda e Internet das Coisas, Inteligência Artificial, Aprendizado de Máquina, Inteligência Artificial na Borda, Aprendizado de Máquina Embarcado e Tiny Machine Learning, e Processamento Digital de Sinais. Ao entender esses conceitos, será possível apreciar melhor a complexidade e o potencial transformador do TinyML.

2.3.1. Sistemas Embarcados

Os sistemas embarcados são computadores projetados especificamente para executar funções dedicadas em diversos dispositivos físicos, como fones de ouvido Bluetooth e unidades de controle de motores em veículos modernos [11]. O software embarcado, que opera nestes sistemas, pode variar desde simples códigos em microcontroladores de relógios digitais até sistemas operacionais mais sofisticados, como o Linux integrado em smart TVs. Ao contrário dos computadores de uso geral, como laptops e smartphones, os sistemas embarcados são desenvolvidos para realizar tarefas específicas e dedicadas.

A relevância dos sistemas embarcados na tecnologia moderna é evidente. Em 2020, estima-se que mais de 28 bilhões de microcontroladores foram distribuídos globalmente, destacando a onipresença desses sistemas em nossa vida cotidiana, desde residências e veículos até fábricas e cidades [9]. Esta ubiquidade sugere que é raro estarmos a mais de alguns metros de um sistema embarcado.

Estes sistemas frequentemente enfrentam limitações impostas pelos ambientes em que operam, como a necessidade de eficiência energética e restrições de memória ou velocidade de *clock*. O desenvolvimento de software para sistemas embarcados envolve a criação de soluções que maximizam a funcionalidade dentro dessas limitações, ressaltando a complexidade e a importância do trabalho dos engenheiros especializados na área.

Neste capítulo, a discussão sobre sistemas embarcados se conecta diretamente ao desenvolvimento e à aplicação de técnicas de aprendizado de máquina, particularmente no contexto do TinyML. A capacidade de implementar modelos de aprendizado de máquina em sistemas embarcados permite a realização de análises avançadas e a execução de tarefas inteligentes diretamente em dispositivos com recursos limitados. A compreensão das restrições e capacidades dos sistemas embarcados é essencial para o desenvolvimento eficaz de soluções que aproveitam o processamento local para oferecer desempenho eficiente e respostas em tempo real.

2.3.2. Computação na Borda e Internet das Coisas

A evolução das redes de computadores tem alternado entre centralização e descentralização do processamento ao longo do tempo. Nos primórdios, o processamento era centralizado em grandes mainframes, com terminais periféricos assumindo funções específicas. A introdução dos computadores pessoais levou a uma maior descentralização, permitindo que dispositivos individuais realizassem tarefas de forma independente de um servidor central.

Com o surgimento da internet e das soluções baseadas na nuvem, houve um retorno à centralização dos processos. No entanto, a expansão da Internet das Coisas (IoT) trouxe uma nova perspectiva, com uma vasta rede de dispositivos interconectados [7].

Esses dispositivos situados na borda da rede utilizam a computação na borda (Edge Computing), uma abordagem que permite o processamento local dos dados. Esse método possibilita a coleta e análise de dados diretamente nos dispositivos, a tomada de decisões em tempo real e a comunicação entre eles, minimizando a necessidade de processamento e armazenamento na nuvem [2].

Os dispositivos de borda desempenham um papel fundamental na conexão entre o

mundo físico e a internet. Dispositivos móveis, como smartphones e tablets, são exemplos de tecnologias de borda que incorporam inteligência artificial para oferecer funcionalidades avançadas, como reconhecimento de voz e fotografia inteligente.

2.3.3. Inteligência Artificial

A Inteligência Artificial (IA) refere-se ao campo da ciência da computação que busca desenvolver sistemas capazes de executar tarefas que normalmente requerem inteligência humana. Este campo é vasto e complexo, abrangendo desde a resolução de problemas até a otimização de processos por meio de máquinas e algoritmos [16].

O conceito de inteligência é multifacetado e difícil de definir de maneira universal. Por exemplo, organismos como o mofo do limo demonstram capacidades de resolução de problemas complexos, como percorrer labirintos, apesar de não possuírem um sistema nervoso central. Uma definição prática de inteligência poderia ser a habilidade de escolher e executar a ação apropriada em situações específicas.

Atividades cotidianas, como acionar uma torneira para lavar as mãos ou frear um veículo para evitar uma colisão, podem parecer simples, mas exigem um nível de inteligência para serem realizadas de maneira eficaz. Enquanto a inteligência humana geral é adaptativa e pode lidar com uma ampla gama de tarefas, a IA geralmente é projetada para tarefas específicas.

Desenvolver uma inteligência artificial com a mesma amplitude e flexibilidade da inteligência humana é uma tarefa complexa e desafiadora. No entanto, a criação de sistemas de IA especializados para tarefas concretas é não apenas viável, mas amplamente realizada. Assim, a IA pode ser vista como a implementação de sistemas artificiais que tomam decisões baseadas em dados e utilizam técnicas de aprendizado de máquina para melhorar seu desempenho ao longo do tempo.

2.3.4. Aprendizado de Máquina

O aprendizado de máquina (ML) é uma abordagem que permite descobrir padrões e tendências em dados, automatizando esse processo por meio de algoritmos. Embora frequentemente associado à inteligência artificial (IA), o aprendizado de máquina (ML) é, na verdade, uma subárea da IA. Enquanto a IA engloba uma variedade de técnicas e abordagens para simular a inteligência humana, o ML é especificamente voltado para a criação de algoritmos que aprendem e fazem previsões com base em dados. Dessa forma, o ML se insere dentro do escopo mais amplo da IA, complementando e expandindo suas capacidades [1].

Para ilustrar o conceito de aprendizado de máquina, considere o exemplo de um sistema de reconhecimento de movimento contínuo descrito neste Capítulo. Neste cenário, um dispositivo, como um smartphone ou um wearable, é equipado com sensores de movimento, como acelerômetros e giroscópios. O objetivo é identificar e classificar diferentes tipos de movimentos, como caminhada, corrida ou gestos específicos, com base nos dados capturados pelos sensores.

Primeiramente, dados são coletados a partir de um conjunto de movimentos realizados por indivíduos. Esses dados incluem medidas contínuas de aceleração e rotação em

diferentes eixos. O objetivo é extrair padrões que permitam distinguir entre os diversos tipos de movimentos. A análise manual desses dados pode ser desafiadora e exigir técnicas avançadas para identificar os padrões relevantes. É aqui que o aprendizado de máquina se mostra valioso.

Com o uso de ML, os dados coletados são processados por um algoritmo de treinamento, que busca mapear a relação entre os sinais dos sensores e os tipos de movimentos correspondentes, criando um modelo. Se o treinamento for bem-sucedido, o modelo será capaz de identificar e classificar novos movimentos com base nas informações capturadas pelos sensores.

A capacidade do modelo de aplicar seu conhecimento a novos dados é chamada de generalização. Durante a fase de treinamento, o modelo aprende a identificar características específicas de cada atividade, permitindo que o rastreador interprete novos dados com precisão. Embora o ML utilize lógica semelhante à lógica condicional, ele oferece uma abordagem mais precisa e eficiente para a análise de dados complexos.

Existem diversos algoritmos de aprendizado de máquina, cada um com suas próprias vantagens e desvantagens. O aprendizado de máquina é particularmente valioso quando se trata de dados que são difíceis de analisar manualmente.

2.3.5. Inteligência Artificial na Borda

A inteligência artificial na borda, ou Edge AI, representa a integração de capacidades de IA em dispositivos localizados na borda da rede. Estes sistemas embarcados, equipados com sensores, capturam dados ambientais e permitem a análise desses dados sem necessidade de transmitir grandes volumes para servidores centrais. Embora a coleta de dados seja intensa, a informação isolada de cada sensor pode ser insuficiente por si só [20].

No passado, os dispositivos da Internet das Coisas (IoT) atuavam principalmente como coletores de dados, enviando informações para processamento centralizado. No entanto, a transmissão contínua de grandes quantidades de dados pode ser cara e energeticamente ineficiente, especialmente para dispositivos IoT movidos a bateria. Isso frequentemente resulta no descarte de dados valiosos devido à incapacidade de transmitir e processar tudo de maneira eficaz.

A abordagem Edge AI enfrenta essas limitações processando dados diretamente no dispositivo onde são gerados. Esse método permite a tomada de decisões localmente, minimizando a necessidade de envio de dados para servidores distantes e reduzindo assim custos e consumo de energia.

A inteligência artificial aplicada na borda pode variar desde a implementação de regras simples até o uso de técnicas avançadas de aprendizado profundo. Essa estratégia não apenas melhora a eficiência ao tomar decisões próximas à origem dos dados, mas também se encaixa bem em arquiteturas de computação distribuída, onde as operações são distribuídas entre dispositivos na borda, gateways locais e servidores na nuvem.

2.3.6. Aprendizado de Máquina Embarcado e Tiny Machine Learning

O conceito de aprendizado de máquina embarcado refere-se à implementação de modelos de aprendizado de máquina em dispositivos projetados para operar com recursos limita-

dos. A abordagem conhecida como Tiny Machine Learning (TinyML) expande essa ideia para ambientes com hardware extremamente restrito, como microcontroladores, processadores digitais de sinal (DSPs) e pequenas FPGAs (Field Programmable Gate Arrays) [22].

Em geral, o aprendizado de máquina embarcado foca na fase de inferência, que consiste em processar dados de entrada e gerar previsões ou classificações. Por exemplo, pode-se usar um modelo para reconhecer atividades físicas com base nas leituras de um acelerômetro. A fase de treinamento, que exige consideráveis recursos computacionais, ainda é realizada em máquinas mais potentes, como computadores pessoais ou servidores.

Os sistemas embarcados enfrentam desafios notáveis, especialmente em termos de memória. A execução de modelos de aprendizado de máquina pode ser limitada pela quantidade de memória somente leitura (ROM) disponível para armazenar o modelo e pela memória de acesso aleatório (RAM) necessária para processar dados temporários durante a inferência. Além disso, o poder computacional restrito desses sistemas pode dificultar a implementação de modelos complexos.

Recentemente, avanços em técnicas de otimização têm permitido a execução eficiente de modelos de aprendizado de máquina em hardware com recursos limitados e baixo consumo de energia. Esses avanços possibilitam a operação de modelos complexos nesses dispositivos, tornando a inteligência artificial mais acessível em ambientes com restrições de hardware.

2.4. Quando Utilizar Edge AI

A relevância do Edge AI torna-se evidente ao analisar o ambiente tecnológico atual, onde a conectividade é amplamente disponível, mesmo em locais remotos. Imagine, por exemplo, uma manhã em um parque nacional distante, onde a música é transmitida pelo smartphone e as fotos são enviadas sem dificuldades. Dada a ubiquidade das conexões de dados, qual seria a vantagem de aplicar inteligência artificial diretamente em dispositivos de borda, se servidores poderosos estão acessíveis?

Edge AI oferece soluções práticas que podem melhorar significativamente a eficiência e a funcionalidade dos sistemas tecnológicos. Para compreender melhor os benefícios do Edge AI, podemos recorrer ao acrônimo BLERP, que representa: *Bandwidth, Latency, Economics, Reliability e Privacy*. Este acrônimo foi descrito em detalhes na referência [19], destacando as principais razões para implementar inteligência na borda.

2.4.1. Largura de Banda

A largura de banda limitada é um desafio significativo para dispositivos IoT, que frequentemente geram mais dados do que podem transmitir devido a restrições na capacidade de rede. Por exemplo, considere um sistema de monitoramento ambiental que utiliza sensores para medir a qualidade do ar em uma área remota. Esses sensores podem coletar dados sobre poluentes atmosféricos, mas a capacidade de transmitir todos esses dados para um centro de análise pode ser restrita.

Se a largura de banda for insuficiente, a maior parte dos dados coletados pode ser descartada, mesmo que contenham informações importantes sobre tendências ou padrões. Embora o envio dos dados para um servidor na nuvem permita análises aprofundadas,

essa abordagem não é sempre viável devido às limitações de conectividade e consumo de energia.

Em muitos cenários, o consumo de energia associado à transmissão de dados é elevado, o que reduz a duração da bateria dos dispositivos IoT. Por exemplo, a comunicação contínua com servidores pode esgotar rapidamente a bateria de um sensor ambiental que coleta dados em tempo real. Alguns algoritmos de aprendizado de máquina podem ser intensivos em processamento, mas frequentemente consomem menos energia do que a transmissão constante de grandes volumes de dados.

A solução para esses desafios é a inteligência artificial na borda (Edge AI). Com a Edge AI, é possível realizar a análise dos dados diretamente no dispositivo, sem a necessidade de enviar todas as informações para a nuvem. Isso permite que o dispositivo identifique e reaja a padrões importantes, como picos de poluição, com a largura de banda disponível de maneira mais eficiente. Dessa forma, o sistema pode emitir alertas locais com base na análise realizada no próprio dispositivo.

Além disso, para dispositivos que operam em locais sem conectividade de rede, a Edge AI oferece uma solução crucial. Esses dispositivos podem realizar análise e tomada de decisão no local, tornando viável a implementação de aplicações avançadas em áreas anteriormente inacessíveis.

2.4.2. Latência

O tempo necessário para transmitir dados entre um dispositivo e um servidor pode ser considerável, mesmo em redes com alta largura de banda. A latência, que é o intervalo entre o envio e o recebimento dos dados, pode variar de milissegundos a minutos ou até horas, dependendo da distância e da infraestrutura envolvida. Em sistemas de comunicação via satélite, por exemplo, a latência pode ser significativamente maior, complicando a interação em tempo real.

Para muitas aplicações, especialmente aquelas que demandam respostas rápidas, a latência elevada pode ser um problema crítico. Por exemplo, em sistemas de controle remoto de drones, a capacidade de responder rapidamente a comandos é essencial para a navegação segura e precisa. Se o tempo de resposta for muito longo, o controle sobre o drone pode ser comprometido, afetando a eficácia da operação.

A Edge AI surge como uma solução eficaz para esses desafios de latência, ao processar e analisar dados localmente no próprio dispositivo. Um exemplo notável é o uso de Edge AI em veículos autônomos, onde sistemas de IA a bordo permitem uma resposta imediata a eventos, como a detecção repentina de obstáculos. Isso garante que o veículo possa ajustar sua trajetória instantaneamente, sem depender da comunicação com servidores distantes.

2.4.3. Economia

O custo de conectar dispositivos e transmitir dados pode ser elevado, especialmente quando se considera a infraestrutura necessária para suportar essa conectividade. Dispositivos que transmitem grandes quantidades de dados, particularmente em áreas remotas ou de difícil acesso, podem gerar despesas significativas. O aumento da largura de banda necessária

para transmitir dados também contribui para esses custos, sendo um desafio crucial para dispositivos que operam em locais isolados e dependem de conexões caras como satélites.

Uma solução eficaz para mitigar esses custos é a implementação de Edge AI, que processa dados diretamente no local de captura. Ao realizar o processamento localmente, em vez de enviar constantemente grandes volumes de dados para servidores externos ou para a nuvem, os dispositivos podem reduzir significativamente os custos associados à transmissão e ao armazenamento de dados.

Considere o exemplo de uma rede de sensores usados para monitorar a qualidade da água em um sistema de distribuição. Esses sensores podem detectar variações importantes na composição da água e gerar grandes volumes de dados continuamente. Tradicionalmente, enviar todos esses dados para análise em tempo real poderia ser caro e ineficiente, especialmente se a rede se estender por grandes distâncias.

Com Edge AI, os sensores podem realizar análises preliminares dos dados localmente e apenas relatar informações críticas ou anômalas. Isso reduz a necessidade de transmissão contínua de grandes quantidades de dados e minimiza o consumo de largura de banda. Como resultado, os custos operacionais são reduzidos, e a eficiência do sistema de monitoramento é aprimorada.

Essa abordagem também é vantajosa em situações em que a conectividade é limitada ou intermitente, permitindo que o sistema opere de forma independente e apenas compartilhe informações essenciais quando necessário.

2.4.4. Confiabilidade

A confiabilidade dos sistemas que utilizam inteligência artificial pode ser significativamente afetada pela dependência de conexões contínuas com a nuvem. Quando um dispositivo depende de uma rede para operar, ele está sujeito a uma série de variáveis externas, incluindo a integridade da comunicação sem fio e a estabilidade dos servidores na internet. Cada um desses elementos representa uma possível fonte de falha.

Quando um dispositivo se baseia em uma conexão constante para realizar suas funções, o sistema pode enfrentar problemas se a conectividade falhar. Por exemplo, um dispositivo inteligente de controle de temperatura em um edifício que depende da nuvem para ajustar as configurações pode falhar em manter a temperatura adequada se a conexão for interrompida, resultando em desconforto e possivelmente danos.

Por outro lado, um sistema embarcado que incorpora inteligência artificial localmente é menos vulnerável a essas falhas de comunicação. Um bom exemplo é um dispositivo de segurança em uma casa, que monitora e reage a movimentos suspeitos. Se esse dispositivo for totalmente autônomo e não depender de comunicação constante com um servidor remoto, ele continuará a operar efetivamente mesmo se houver uma interrupção na conexão com a internet. Isso é crucial em situações em que a continuidade do serviço é essencial para a segurança.

A confiança em sistemas que operam localmente elimina muitas das incertezas associadas à dependência de uma infraestrutura de rede externa. Isso garante que o sistema possa manter suas funções críticas, independentemente de problemas de conectividade, aumentando a segurança e a confiabilidade geral.

2.4.5. Privacidade

A crescente integração de tecnologias inteligentes em nosso cotidiano frequentemente implica uma troca entre a conveniência e a proteção dos dados pessoais. Muitos usuários aceitam compartilhar suas informações para beneficiar-se de funcionalidades avançadas oferecidas por dispositivos conectados, aceitando que suas informações sejam processadas por servidores remotos na nuvem.

Para alguns dispositivos, como um rastreador de exercícios que envia dados de desempenho para análise, a preocupação com a privacidade pode ser mínima. No entanto, para outros casos, a proteção da privacidade é crucial. Um exemplo seria um assistente doméstico inteligente que coleta dados de áudio contínuos. Mesmo com a promessa de conveniência, muitos usuários podem se sentir desconfortáveis com a ideia de ter gravações de suas conversas pessoais transmitidas e armazenadas em servidores externos.

A tecnologia Edge AI fornece uma solução eficaz para mitigar essas preocupações. Em vez de enviar continuamente dados sensíveis para a nuvem, dispositivos equipados com inteligência local podem realizar análises diretamente no próprio dispositivo. Por exemplo, um dispositivo de monitoramento de saúde pode analisar sinais vitais e detectar anomalias sem enviar informações para fora do dispositivo. Se uma anomalia for detectada, o sistema pode alertar o usuário ou o profissional de saúde de forma segura, sem comprometer dados pessoais.

Esse modelo de processamento local protege a privacidade ao evitar a transmissão de dados sensíveis, limitando o risco de exposição e possíveis abusos. Ao manter as informações dentro do dispositivo, a Edge AI garante que dados pessoais sejam utilizados de maneira mais segura e controlada, respeitando a privacidade dos usuários.

2.5. Considerações Finais

Neste capítulo, foram explorados os princípios fundamentais e as abordagens técnicas que sustentam a tecnologia TinyML, destacando como a integração da inteligência artificial (IA) em sistemas embarcados pode transformar diversos setores, oferecendo soluções mais autônomas e eficientes. Inicialmente, discutimos a aplicação prática de TinyML, através de projetos com smartphones e kits TinyML, demonstrando como esses dispositivos podem ser utilizados para tarefas como a detecção de movimento. Estes exemplos práticos ilustram a viabilidade e o impacto real da tecnologia em cenários do dia a dia.

O capítulo avançou para uma análise aprofundada dos desafios e benefícios associados à Edge AI. Foi abordado como a tecnologia pode superar limitações críticas de largura de banda, latência e custos. A capacidade de realizar processamento local é essencial para minimizar a transmissão de dados, economizando largura de banda e reduzindo os custos operacionais. Isso é particularmente importante em ambientes com conectividade limitada, onde a análise local permite a continuidade das operações e aumenta a resiliência dos sistemas.

A redução da latência foi outro ponto crucial discutido, destacando como a Edge AI permite respostas quase instantâneas ao processar dados diretamente no dispositivo. Isso é fundamental para aplicações que requerem decisões rápidas e precisas, como em veículos autônomos e sistemas de controle industrial, onde a rapidez e a eficácia das

respostas são essenciais para a segurança e a eficiência.

Além dos desafios econômicos, a integração de Edge AI foi analisada sob a perspectiva da redução dos custos de infraestrutura e processamento na nuvem. A análise local não só reduz a necessidade de servidores e largura de banda, mas também torna a tecnologia mais acessível e econômica. Em contextos de conectividade limitada ou inexistente, essa abordagem é crucial para a operação contínua e para a resiliência a falhas de comunicação.

A confiabilidade dos sistemas embarcados também foi abordada, ressaltando que a análise local reduz a vulnerabilidade a falhas de conectividade e problemas de rede. A Edge AI melhora a robustez das aplicações, proporcionando maior estabilidade e segurança para sistemas críticos.

No tocante à privacidade, a Edge AI oferece uma solução ao minimizar a transmissão de dados pessoais para a nuvem. Isso reduz o risco de exposição e abusos, especialmente em áreas sensíveis como segurança, saúde e educação, onde a proteção de dados é essencial.

A integração dos conceitos e práticas discutidos neste capítulo oferece uma base para a compreensão e aplicação da tecnologia TinyML. Os projetos práticos com smartphones e kits TinyML demonstram a aplicabilidade da tecnologia, enquanto as análises dos desafios e benefícios ressaltam a importância da Edge AI na evolução das tecnologias embarcadas.

À medida que a tecnologia avança, espera-se que o TinyML continue a expandir as fronteiras da inovação tecnológica, oferecendo novas oportunidades para desenvolvimento e aplicação em diversos setores. Compreender as nuances da Edge AI é essencial para profissionais e entusiastas que buscam se destacar nas inovações tecnológicas. Espera-se que o conhecimento adquirido inspire novas ideias e soluções, contribuindo para um futuro em que a inteligência artificial se integre de maneira eficiente e segura em no cotidiano.

Referências

- [1] Yoshua Bengio. *Deep Learning*. Adaptive Computation and Machine Learning series. MIT Press, London, England, November 2016.
- [2] Jie Cao, Quan Zhang, and Weisong Shi. *Edge Computing: A Primer*. Springer International Publishing, Cham, 2018.
- [3] Gonçalo Carvalho, Bruno Cabral, Vasco Pereira, and Jorge Bernardino. Edge computing: current trends, research challenges and future directions. *Computing*, 103(5):993–1023, May 2021.
- [4] François Chollet et al. Keras. <https://keras.io>, 2015.
- [5] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *JSTOR: Applied Statistics*, 28(1):100–108, 1979.
- [6] Shawn Hymel, Colby R. Banbury, Daniel Situnayake, Alex Elium, Carl Ward, Matthew Kelcey, Mathijs Baaijens, Mateusz Majchrzycki, Jenny Plunkett, David

- Tischler, Alessandro Grande, Louis Moreau, Dmitry Maslov, Arthur Beavis, Jan Jongboom, and Vijay Janapa Reddi. Edge impulse: An ml ops platform for tiny machine learning. *ArXiv*, abs/2212.03332, 2022.
- [7] F Khodadadi, A V Dastjerdi, and R Buyya. Internet of things: An overview. In *Internet of Things*, pages 3–27. Elsevier, 2016.
- [8] Richard G Lyons. *Understanding digital signal processing*. Prentice Hall, Philadelphia, PA, 3 edition, November 2010.
- [9] MarketsandMarkets. Global microcontroller market: Analysis and forecast. Technical report, MarketsandMarkets, 2021.
- [10] A. Martins and R. Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *International Conf. on Machine Learning - ICML*, volume 48, pages 1614–1623, June 2016.
- [11] K C S Murti. *Design principles for embedded systems*. Transactions on Computer Systems and Networks. Springer, Singapore, Singapore, 1 edition, September 2021.
- [12] S.J. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- [13] Visal Rajapakse, Ishan Karunanayake, and Nadeem Ahmed. Intelligence at the extreme edge: A survey on reformable tinyml. *ACM Comput. Surv.*, 55(13s), jul 2023.
- [14] Partha Pratim Ray. A review on TinyML: State-of-the-art and prospects. *J. King Saud Univ. - Comput. Inf. Sci.*, 34(4):1595–1623, April 2022.
- [15] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [16] Stuart Russell and Peter Norvig. *Artificial intelligence*. Pearson, Upper Saddle River, NJ, 4 edition, November 2020.
- [17] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2018. cite arxiv:1801.04381.
- [18] Raghubir Singh and Sukhpal Singh Gill. Edge AI: A survey. *Internet of Things and Cyber-Physical Systems*, 3:71–92, 2023.
- [19] Weixing Su, Linfeng Li, Fang Liu, Maowei He, and Xiaodan Liang. Ai on the edge: a comprehensive review. *Artif. Intell. Rev.*, 55(8):6125–6183, dec 2022.
- [20] Xiaofei Wang, Han Yu, Chunyan Miao, and Qiang Yang. *Edge AI: Convergence of Edge Computing and Artificial Intelligence*. Springer, 2020.
- [21] Pete Warden. *Tiny ML*. O’Reilly Media, Sebastopol, CA, January 2020.
- [22] Pete Warden. *Tiny ML*. O’Reilly Media, Sebastopol, CA, January 2020.

- [23] William H. Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes 3rd edition*. Cambridge University Press, Cambridge, England, 3 edition, September 2007.
- [24] Xue Ying. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 1168(2):022022, feb 2019.
- [25] Orr Zohar, Kuan-Chieh Wang, and Serena Yeung. Prob: Probabilistic objectness for open world object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11444–11453, June 2023.

Capítulo

3

Indústria 5.0: Implementando Sistemas de Decisão Orientados por Dados.

Júlio V. M. Marques, Clésio A. Gonçalves, Armando L. Borges, Viviane B. Leal Dias, Willians S. Santos e Romuere R. V. Silva

Abstract

The transition from Industry 4.0 to Industry 5.0 highlights the importance of reintegrating humans into production processes, fostering collaboration between humans and machines. Personalization and sustainability become essential, with data-driven decision systems playing a key role. This work explores the implementation of these systems, from data preparation to integration in industrial environments, demonstrating their capacity to transform operations and maximize efficiency. It also discusses the challenges and opportunities of Industry 5.0, emphasizing the need for continuous innovation, ethics, and organizational adaptation to ensure success in the new industrial era.

Resumo

A transição da Indústria 4.0 para a Indústria 5.0 destaca a importância de reintegrar o ser humano nos processos produtivos, promovendo colaboração entre humanos e máquinas. A personalização e a sustentabilidade tornam-se essenciais, com sistemas de decisão baseados em dados desempenhando um papel importante. Este trabalho explora a implementação desses sistemas, desde a preparação de dados até a integração em ambientes industriais, evidenciando sua capacidade de transformar operações e maximizar a eficiência. Também discute os desafios e as oportunidades da Indústria 5.0, enfatizando a necessidade de inovação contínua, ética e adaptação organizacional para garantir o sucesso na nova era industrial.

1.1. Introdução

A Revolução Industrial, em cada uma de suas fases, trouxe mudanças profundas na forma como as sociedades produzem e consomem bens e serviços. A Indústria 4.0, com seus avanços tecnológicos centrados na automação, conectividade e digitalização, abriu um

novo capítulo nessa evolução [1]. Esse período foi marcado por uma automação em larga escala, resultando em ganhos significativos de produtividade e redução de custos. Tecnologias que permitiram a produção em massa de forma flexível foram desenvolvidas, possibilitando que máquinas operassem de maneira autônoma, sem a constante necessidade de intervenção humana [2]. Contudo, esse progresso também suscitou preocupações acerca da desumanização dos processos produtivos e da potencial perda de empregos qualificados, à medida que a dependência de máquinas e algoritmos aumentava. Além disso, surgiram questões relacionadas à sustentabilidade ambiental, destacando a importância de equilibrar o uso da tecnologia com a preservação dos valores humanos e do meio ambiente.

Em resposta aos desafios apresentados pela Indústria 4.0, a Indústria 5.0 emerge com o propósito de reintegrar o ser humano nos processos produtivos, promovendo um ambiente de trabalho mais colaborativo, sustentável e personalizado [3]. Diferentemente da Indústria 4.0, que enfatiza a automação e a eficiência, a Indústria 5.0 valoriza a colaboração entre humanos e máquinas [4]. Essa nova fase industrial visa combinar o poder da tecnologia com a criatividade e a adaptabilidade humanas, criando produtos e serviços que atendem às necessidades específicas dos clientes, com uma ênfase especial em sustentabilidade e ética [5]. A coleta e análise de grandes volumes de dados permitem a personalização dos processos produtivos, enquanto a inteligência artificial apoia os trabalhadores na realização de tarefas complexas. Ademais, a Indústria 5.0 promove a economia circular e o uso de materiais reciclados, reduzindo assim o impacto ambiental.

A transição da Indústria 4.0 para a Indústria 5.0 oferece às empresas não apenas a oportunidade de aprimorar sua competitividade e resiliência, mas também impõe uma reestruturação significativa que abrange tanto aspectos tecnológicos quanto culturais. Este capítulo se propõe a explorar profundamente os conceitos centrais da Indústria 5.0, destacando suas diferenças em relação à Indústria 4.0, além dos principais desafios e oportunidades que surgem com sua adoção. A análise busca fornecer *insights* valiosos para que as empresas adaptem suas operações a essa nova era, aproveitando os avanços tecnológicos sem negligenciar a importância do toque humano. O objetivo final é demonstrar como as organizações podem integrar harmoniosamente inovação tecnológica e personalização, garantindo que a evolução industrial se alinhe com a sustentabilidade e as necessidades humanas.

Este trabalho está estruturado da seguinte forma: a Seção 1.2 explora os fundamentos dos sistemas de decisão baseados em dados; a Seção 1.3 discute a implementação dos sistemas de decisão na indústria 5.0; a Seção 1.4 aborda as vantagens da decisão orientada por dados; a Seção 1.5 apresenta os desafios e barreiras na implementação desses sistemas; a Seção 1.6 explora o futuro dos sistemas de decisão na indústria 5.0; a Seção 1.7 demonstra um estudo de caso e suas principais etapas, por fim, a Seção 1.8 oferece os principais *insights* e conclusões.

1.2. Fundamentos dos Sistemas de Decisão Baseados em Dados

Com o advento da Indústria 5.0, uma nova dimensão se estabelece na interação entre humanos e máquinas, onde a personalização e a tomada de decisões em tempo real tornam-se elementos centrais dos processos produtivos [6]. Nesse cenário, os sistemas de decisão

baseados em dados emergem como ferramentas indispensáveis para alinhar a capacidade analítica das máquinas com a intuição e o conhecimento humano. Esses sistemas processam grandes volumes de dados, fornecendo informações precisas que permitem às organizações responder rapidamente a mudanças, minimizar riscos potenciais, maximizar oportunidades e otimizar processos internos.

Os sistemas de decisão baseados em dados são uma forma de Sistemas de Apoio à Decisão (SAD). Esses SADs são modelos computacionais projetados para analisar grandes volumes de dados, auxiliando seus usuários na tomada de decisões dentro de um contexto específico [7]. Esses sistemas são particularmente valiosos no ambiente empresarial, onde a precisão na análise de variáveis voláteis e em cenários dinâmicos é crucial. A análise correta e precisa dessas variáveis pode ser o fator decisivo para o sucesso das ações implementadas.

Dada a sua relevância, é importante destacar que os SADs são capazes de resolver problemas pouco ou não-estruturados. Em outras palavras, eles são capazes de lidar com situações onde as variáveis e os critérios para a tomada de decisão não estão claramente definidos, são incertos ou até mesmo implícitos. A função dos SADs é auxiliar os tomadores de decisão na exploração de diferentes alternativas e na avaliação dos possíveis impactos de suas escolhas. Devido à sua grande capacidade, esses sistemas são amplamente utilizados no nível estratégico da pirâmide organizacional [8]. A habilidade dos SADs de compilar grandes volumes de dados e de sintetizar essa informação de maneira organizada e acessível para gestores e analistas os torna eficazes na execução de atividades como o planejamento estratégico, o controle gerencial e operacional. A Figura 1.1 ilustra a abrangência dos SADs na pirâmide organizacional.



Figura 1.1: A pirâmide organizacional demonstra que, embora os SADs sejam amplamente aplicados no nível estratégico, eles também podem ser utilizados para apoiar os outros níveis da organização.

Como ilustrado na Figura 1.1, os SADs podem ser aplicados em todos os três ní-

veis da pirâmide organizacional. No nível operacional, onde atividades rotineiras garantem o funcionamento contínuo da organização, os SADs são empregados para automatizar e otimizar processos, como controle de produção, gestão de inventário e monitoramento de desempenho, garantindo agilidade na resolução de questões ou problemas decisórios [9]. No nível tático, que se concentra em planos e decisões de médio prazo, os SADs fornecem análises detalhadas e previsões de cenários futuros [10], auxiliando gestores a alinhar as operações diárias com os objetivos estratégicos da empresa e a otimizar o uso de recursos. Finalmente, no nível estratégico, onde são tomadas decisões de longo prazo que impactam toda a organização, os SADs oferecem suporte essencial aos executivos, proporcionando análises aprofundadas, simulações de cenários e ferramentas de previsão que permitem aos líderes avaliar riscos e identificar oportunidades potenciais [11].

Nesse contexto, os sistemas de decisão baseados em dados são uma forma de SADs que se concentram na utilização de grandes volumes de dados e técnicas analíticas avançadas para apoiar a tomada de decisões. Esses sistemas priorizam o uso de quantidades significativas de dados em conjunto com técnicas de aprendizado de máquina e outros métodos analíticos para descobrir padrões e tendências. Além disso, eles são capazes de automatizar partes do processo decisório, fornecendo recomendações ou ações sugeridas com base na análise de dados em tempo real.

1.2.1. Componentes principais

De acordo com Power [12], em sua obra intitulada *Decision Support Systems: Concepts and Resources for Managers*, os SADs são compostos por quatro componentes principais: banco de dados, modelos, comunicação e interface com o usuário. A Figura 1.2 apresenta um diagrama que ilustra como esses componentes se interligam e formam um Sistema de Apoio à Decisão.

O componente de banco de dados é responsável por armazenar e gerenciar todos os dados necessários para fundamentar a tomada de decisões. Em essência, sua função principal é alocar todos os dados, documentos e informações que servirão como base para os modelos analíticos [12]. Esse componente pode incluir bancos de dados relacionais, *data warehouses*, *data marts*, entre outras formas de armazenamento de dados. Nesse contexto, a qualidade e a acessibilidade dos dados são cruciais para o sucesso das análises.

O componente de modelo abrange os modelos analíticos e matemáticos responsáveis pelo processamento dos dados [12]. Nesse componente estão incluídos algoritmos para análises estatísticas, simulações, aprendizado de máquina e inteligência artificial. Esses algoritmos transformam os dados brutos em informações úteis para apoiar o processo decisório, identificando até mesmo pequenas nuances que podem ter um impacto significativo na decisão final. Por meio de simulações e análises, os modelos podem destacar a melhor solução dentre as disponíveis, com base em parâmetros pré-estabelecidos, e até realizar previsões sobre cenários futuros.

O componente de comunicação, como o próprio nome indica, é responsável por mediar e assegurar a troca de dados e informações entre os diversos componentes internos do SAD, bem como com sistemas externos [12]. Sua estrutura inclui protocolos, redes de comunicação, APIs e ferramentas que facilitam um processo de comunicação eficiente. Esse componente atua como uma ponte, conectando todos os elementos do SAD e permi-

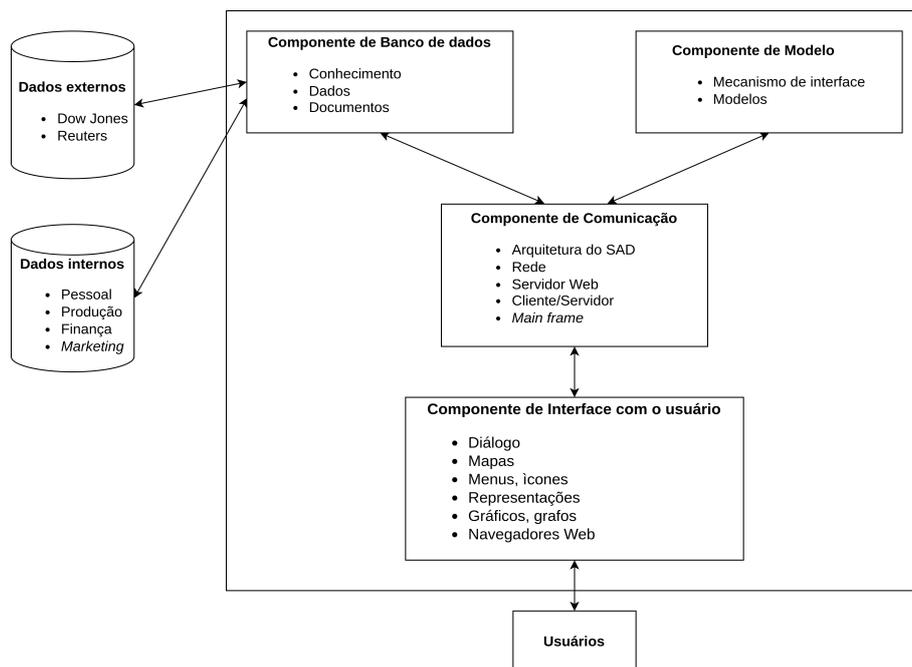


Figura 1.2: Diagrama dos componentes de um SAD, conforme descrito por Power [12]. Imagem adaptada do livro *Decision Support Systems: Concepts and Resources for Managers*.

tindo que interajam de forma contínua. Além disso, ele possibilita que diferentes usuários e partes interessadas participem do processo decisório e de análise.

O quarto componente é a interface com o usuário, que serve como o ponto de contato entre os indivíduos e o SAD. Através desse componente, os usuários podem visualizar os resultados das análises processadas pelos modelo [12]. Ele define como os usuários interagem com o sistema, por meio de uma interface bem projetada e adaptada às suas necessidades. Com essa interface, os usuários podem se comunicar com o sistema, visualizando informações processadas, realizando consultas, explorando dados e ajustando parâmetros. Essas funcionalidades permitem investigar possíveis cenários e soluções alternativas para questões ou problemas específicos, levando a conclusões fundamentadas no processo decisório.

1.2.2. Técnicas Analíticas

Conforme mencionado anteriormente, uma variedade de algoritmos e técnicas pode ser utilizada para processar dados brutos, com o objetivo de extrair *insights* relevantes para a tomada de decisões. A análise estatística é uma das abordagens fundamentais para a obtenção de informações a partir de um conjunto de dados [13]. Esse processo envolve uma série de cálculos e métodos descritivos, como testes de hipóteses, análise de regressão, análise de séries temporais e análise de variância, que permitem identificar padrões, tendências, correlações e outras características importantes nos dados.

Outra técnica amplamente utilizada devido à sua capacidade de avançar rapidamente e produzir resultados precisos é o uso de algoritmos de aprendizado de máquina e

inteligência artificial. Esses algoritmos permitem que computadores identifiquem e interpretem padrões a partir de conjuntos de dados [14]. Nesse contexto, é importante destacar que existem algoritmos supervisionados, onde o conjunto de dados deve ser rotulado para que o modelo aprenda com base na diferença entre suas inferências e os rótulos fornecidos. Por outro lado, os algoritmos não supervisionados não requerem rótulos, sendo capazes de identificar padrões sem orientação explícita sobre o resultado esperado. Além disso, existem os algoritmos de aprendizado por reforço, nos quais um agente aprende a tomar decisões ao interagir com um ambiente, recebendo recompensas ou penalidades com base em suas ações. Em geral, esses algoritmos podem melhorar suas previsões ou decisões ao longo do tempo, sem a necessidade de serem programados explicitamente para cada tarefa.

Nesse contexto, também é relevante mencionar a análise de simulação. Essa técnica envolve a criação de modelos que replicam o comportamento de um sistema real para prever o impacto de diferentes variáveis ou cenários. Isso permite que os tomadores de decisão testem diversas hipóteses e compreendam os possíveis resultados sem a necessidade de implementar mudanças no mundo real. Por meio desse método, é possível detalhar aspectos do modelo com base nos fatores que se deseja considerar [15]. Algumas técnicas comuns de simulação incluem a Simulação de Monte Carlo, que utiliza métodos estatísticos para modelar a probabilidade de diferentes resultados em processos que podem ser repetidos várias vezes, sendo frequentemente aplicada para prever incertezas em modelos financeiros ou de negócios. Além disso, é importante destacar a Simulação Discreta de Eventos, onde o valor ou estado de um ou mais fenômenos é alterado em pontos discretos no tempo [16]. Essa técnica é amplamente utilizada para modelar e otimizar linhas de produção, simulando o fluxo de materiais, tempos de processamento e gargalos.

1.3. Implementação de Sistemas de Decisão na Indústria 5.0

A implementação de sistemas de decisão na Indústria 5.0 representa um avanço significativo ao combinar inteligência artificial com a criatividade humana, estabelecendo um novo paradigma de colaboração [17]. Nesse contexto, a parceria entre humanos e máquinas inteligentes resulta em um aumento substancial na eficiência e na personalização dos processos industriais. A fusão da precisão, velocidade e automação das máquinas com a intuição e o discernimento humano gera um sistema altamente adaptável e ágil. Esses sistemas facilitam a análise de dados em tempo real, permitindo que os processos de produção se adaptem às demandas do mercado e às necessidades individuais dos consumidores [18]. O desenvolvimento desses sistemas segue etapas definidas e utiliza tecnologias avançadas, sendo projetado para ser eficiente, sustentável e alinhado com os princípios de responsabilidade social e ambiental [19]. Cada sistema de decisão contribui de maneira distinta para a otimização dos processos industriais, oferecendo soluções específicas para diferentes desafios e promovendo uma gestão mais eficaz e flexível.

Os SADs são projetados para auxiliar gestores na análise de grandes volumes de dados provenientes de diversas fontes. Eles integram informações operacionais, financeiras e de mercado para criar relatórios e análises que apoiam decisões estratégicas e táticas. Por meio da modelagem matemática e da análise estatística, os SADs fornecem dados detalhados sobre diferentes cenários, facilitando a tomada de decisões mais informadas. A implementação de um SAD exige a integração de dados de sistemas existentes,

como os de gestão de recursos (ERP) e de relacionamento com clientes (CRM). Além disso, são desenvolvidos *dashboards* interativos que permitem explorar e entender os dados de forma visual e intuitiva. Esses *dashboards* asseguram a qualidade das informações e oferecem flexibilidade para atender às necessidades analíticas. Pesquisas indicam que os SADs melhoram a eficiência na tomada de decisões ao utilizarem ferramentas avançadas de análise e visualização de dados, permitindo uma interpretação mais clara de conjuntos complexos de informações [20]. Sua capacidade de adaptação a diferentes setores demonstra sua versatilidade e contribui para melhores resultados operacionais [21]. Existem diferentes categorias de sistemas de decisão, cada uma com características e finalidades específicas para otimizar processos em suas respectivas áreas de aplicação. Nos parágrafos seguintes, serão apresentados os principais sistemas de decisão, explorando suas funcionalidades básicas e como contribuem para a evolução da indústria moderna.

Os Sistemas de Informação Gerencial (SIG) são desenvolvidos para auxiliar as empresas na gestão e monitoramento de suas atividades diárias. Esses sistemas coletam e organizam informações detalhadas sobre o desempenho das operações da empresa, como produtividade e eficiência. A partir desses dados, os SIGs geram relatórios automatizados que ajudam os gestores a compreender o andamento das operações e a tomar decisões baseadas em dados. Para implementar um SIG de forma eficaz, é necessário configurar sistemas que possam coletar dados em tempo real, definir Indicadores-Chave de Desempenho (KPIs) que medem aspectos críticos das operações e desenvolver relatórios automatizados que apresentem essas informações de maneira clara e acessível. Além disso, o SIG deve integrar dados de diferentes departamentos para fornecer uma visão completa e integrada do desempenho da empresa. Em outras palavras, o SIG transforma dados operacionais em informações úteis e acionáveis, facilitando a gestão e a tomada de decisões, aumentando o valor comercial e a lucratividade por meio da utilização eficaz dos dados [22]. A integração eficiente dos dados e a clareza na apresentação das informações são essenciais para o sucesso desses sistemas.

Os Sistemas Especialistas são projetados para replicar o conhecimento e a experiência de especialistas humanos em áreas específicas, oferecendo recomendações fundamentadas em regras e conhecimento técnico especializado. Esses sistemas podem aumentar a eficiência operacional em até 42% e contribuem positivamente para o crescimento organizacional [23]. Os Sistemas Especialistas são particularmente valiosos na resolução de problemas que exigem um elevado nível de expertise técnica. Eles operam por meio de bases de conhecimento e inferência lógica, simulando o raciocínio dos especialistas. Esses sistemas aplicam regras de decisão codificadas para fornecer soluções a questões complexas. A implementação de Sistemas Especialistas envolve várias etapas, incluindo a captura e modelagem do conhecimento especializado, a codificação das regras de decisão e o treinamento do sistema com dados relevantes. Além disso, é essencial que esses sistemas sejam continuamente atualizados com novos conhecimentos e permitam interações com os usuários, a fim de aprimorar a precisão e a relevância das respostas fornecidas. Esse processo contínuo de atualização e interação garante que o sistema permaneça eficaz e alinhado com as necessidades dos usuários.

Sistemas Baseados em Inteligência Artificial (IA) utilizam algoritmos avançados para analisar grandes volumes de dados e gerar previsões ou recomendações. Esses sistemas têm a capacidade de aprender com os dados, aprimorando seu desempenho ao longo

do tempo, identificando padrões complexos e realizando previsões precisas com base em informações históricas e em tempo real, demonstrando sua capacidade de adaptação e melhoria contínua [24]. A implementação desses sistemas envolve etapas vitais. Primeiramente, é necessário desenvolver e treinar modelos utilizando dados históricos e atuais. Em seguida, esses modelos devem ser integrados aos processos existentes e disponibilizados por meio de interfaces de usuário adequadas. Para garantir que esses sistemas permaneçam precisos e relevantes, é essencial que os modelos sejam ajustados e atualizados regularmente. Esse processo contínuo de ajuste e monitoramento assegura que os sistemas de IA permaneçam eficazes e alinhados com as necessidades dos usuários.

Os Sistemas de Análise Preditiva utilizam técnicas de análise estatística e algoritmos de aprendizado de máquina para prever eventos futuros e identificar tendências com base em dados históricos e em tempo real. Esses sistemas permitem antecipar falhas, variações na demanda e outras variáveis críticas, possibilitando a adoção de ações preventivas e a otimização de processos [25]. A implementação desses sistemas envolve etapas essenciais, começando pela coleta e preparação dos dados históricos. Em seguida, é necessário desenvolver modelos preditivos que serão integrados aos sistemas de decisão, permitindo que as previsões sejam aplicadas de maneira prática. É fundamental que o sistema tenha a capacidade de lidar com variáveis em tempo real e ajustar suas previsões conforme necessário, garantindo a precisão e a relevância das análises. A adaptação e a atualização contínuas são vitais para assegurar que os sistemas de análise preditiva permaneçam eficazes e alinhados com as condições dinâmicas do ambiente operacional.

Os Sistemas de Gestão de Conhecimento (SGC) são projetados para capturar, organizar e compartilhar o conhecimento dentro de uma organização. Eles auxiliam na utilização tanto do conhecimento explícito quanto do tácito, aumentando a eficiência e promovendo a inovação. Os SGC criam bases de dados onde o conhecimento pode ser armazenado e acessado, facilitando a colaboração entre funcionários e permitindo o uso de informações acumuladas para resolver problemas e desenvolver novas ideias. A implementação de um SGC exige a construção de uma plataforma onde as informações possam ser armazenadas e compartilhadas de maneira fácil e intuitiva. Isso envolve o desenvolvimento de ferramentas de busca eficientes e o incentivo à participação dos colaboradores na contribuição de conhecimento. A facilidade de uso e o estímulo à colaboração são fundamentais para garantir que todos na organização possam usufruir ao máximo do sistema e contribuir para o compartilhamento de informações. Esses sistemas são vitais para aumentar a competitividade organizacional, simplificando a coleta, o processamento e a propagação do conhecimento, apoiando o aprendizado e desenvolvimento contínuos [26].

1.3.1. Estratégias para Implementação de Sistemas de Decisão na Indústria 5.0

A implementação de sistemas de decisão na Indústria 5.0 exige um conjunto cuidadosamente planejado de estratégias organizadas em fases distintas, cada uma desempenhando um papel importante para o sucesso do sistema. Essas estratégias são fundamentais para assegurar que o sistema atenda às necessidades da organização e contribua para a eficiência e a inovação dos processos industriais. A Figura 1.3 apresenta visualmente as principais etapas necessárias para a implementação eficaz de um sistema de decisão.

O planejamento é a fase inicial e essencial na implementação de sistemas de de-



Figura 1.3: Etapas da implementação de um sistema de decisão.

ção. Nesta etapa, são definidos os objetivos do sistema, os requisitos de dados e as metas de desempenho. Um planejamento eficaz garante que o sistema de decisão esteja alinhado com as estratégias da organização e atenda tanto às necessidades operacionais quanto às estratégicas [27]. Durante essa fase, são identificadas as necessidades da organização, estabelecido o escopo do projeto, e definidos cronogramas e orçamentos. Esse planejamento fornece uma base sólida para todas as fases subsequentes [28], garantindo que o sistema esteja preparado para cumprir as expectativas e objetivos da organização.

A fase de desenvolvimento abrange a criação ou personalização do software de decisão, a configuração dos modelos de decisão e a integração com os sistemas existentes. É vital que o desenvolvimento leve em conta a flexibilidade e a escalabilidade do sistema, permitindo adaptações conforme as necessidades da organização evoluem. Nesta etapa, são realizadas atividades como a codificação do sistema, a integração de dados provenientes de diversas fontes e o desenvolvimento de interfaces de usuário que sejam intuitivas e funcionais. O alinhamento do desenvolvimento com as metas organizacionais influencia de forma positiva os resultados obtidos [29]. Um desenvolvimento bem-sucedido resulta em um sistema que atende aos requisitos atuais e é capaz de se adaptar a futuras mudanças e expansões.

Os testes são fundamentais para garantir que o sistema funcione conforme o esperado e atenda às especificações estabelecidas. Esta fase inclui a realização de simulações e

validações para avaliar o desempenho do sistema em situações reais. Os testes verificam a funcionalidade geral do sistema, seu desempenho em diferentes cenários e a necessidade de ajustes baseados nos resultados obtidos. A condução de testes rigorosos assegura que o sistema esteja pronto para operar de maneira eficiente no ambiente de produção [30].

O treinamento é uma etapa essencial para garantir que os usuários possam operar o sistema de decisão de maneira eficiente. Durante essa fase, são desenvolvidos manuais e materiais de apoio, realizados *workshops* e sessões práticas, além de ser oferecido suporte contínuo aos usuários. O objetivo é assegurar que os operadores se sintam confortáveis e capacitados ao utilizar o novo sistema [30]. Um treinamento eficaz não apenas facilita a adoção do sistema, mas também maximiza seu potencial, capacitando os usuários a explorar e compreender todas as funcionalidades com excelência.

A implementação envolve a introdução gradual do sistema no ambiente de produção. Essa fase deve ser planejada com cuidado para minimizar riscos e permitir ajustes conforme necessário. A implementação inclui a integração do sistema com os processos existentes, a adaptação dos operadores às novas ferramentas e a realização de ajustes durante a transição. Uma implementação bem planejada e executada garante uma integração suave e eficiente do sistema, minimizando a interrupção das operações diárias [31].

Após a implementação, se inicia a fase de monitoramento contínuo e manutenção. Essa etapa é vital para garantir que o sistema continue operando de maneira eficiente e se adapte às mudanças nas condições operacionais. O monitoramento envolve a observação constante do desempenho do sistema, a atualização dos modelos de decisão conforme necessário e a realização de ajustes para resolver quaisquer problemas que possam surgir [32]. A manutenção contínua assegura que o sistema permaneça relevante e eficaz ao longo do tempo, atendendo às necessidades em constante evolução da organização [28].

Essas estratégias são essenciais para a implementação bem-sucedida de sistemas de decisão. Elas fornecem uma base sólida para a tomada de decisões eficazes e para a otimização dos processos industriais, estabelecendo um novo padrão de eficiência e inovação no ambiente industrial moderno. A integração harmoniosa entre a inteligência artificial e a criatividade humana é o que define o sucesso dessas estratégias, permitindo que as organizações aproveitem ao máximo as vantagens oferecidas pela Indústria 5.0.

1.4. Vantagens da Decisão Orientada por Dados

A tomada de decisão baseada em dados é o processo de utilizar dados para orientar as decisões e validar o curso das ações. De acordo com uma pesquisa realizada pela PwC com mais de 1.000 executivos seniores, organizações altamente orientadas por dados têm três vezes mais probabilidade de relatar melhorias significativas na tomada de decisões em comparação àquelas que dependem menos de dados [33].

Decisões fundamentadas em dados permitem a otimização de processos e recursos, eliminando ineficiências. Isso resulta em maior produtividade ao automatizar tarefas repetitivas e concentrar esforços em áreas que geram maior valor. Ao utilizar dados coletados de clientes e mercados, as empresas podem desenvolver produtos e serviços altamente personalizados, atendendo às necessidades e preferências específicas dos clientes. Esse nível de personalização aumenta a satisfação e a fidelização do cliente.

Além disso, decisões baseadas em dados podem contribuir para monitorar e reduzir o desperdício, otimizar o uso de recursos e melhorar a sustentabilidade ambiental das operações, alinhando-se com as metas de responsabilidade corporativa.

1.5. Desafios e Barreiras

Com a chegada da Indústria 5.0, é fácil subestimar os desafios que ela traz. No entanto, para que a Indústria 5.0 se traduza em sucesso para os negócios, é essencial que esses desafios sejam identificados e resolvidos de maneira eficaz.

A Indústria 5.0 requer transformações culturais significativas nas empresas, incluindo a adoção de novos modelos de gestão, a promoção de uma colaboração mais estreita entre equipes multidisciplinares, o aumento da demanda por profissionais altamente qualificados e a possível redução de postos de trabalho em determinadas áreas.

A coleta e o uso de grandes volumes de dados levantam preocupações substanciais em relação à segurança. Proteger esses dados contra violações e ataques cibernéticos, além de garantir a conformidade com regulamentos de privacidade, torna-se fundamental. Além disso, a integração e sincronização dos sistemas existentes com novos sistemas de coleta de dados apresentam um desafio considerável para o funcionamento eficiente da Indústria 5.0.

A transição para uma cultura orientada por dados pode enfrentar resistência dentro das organizações, seja devido ao medo de mudanças, à falta de habilidades digitais, ou a preocupações com a substituição do trabalho humano pela automação. A adoção de tecnologias avançadas exige um compromisso adicional de tempo e esforço por parte dos trabalhadores. Fábricas inteligentes, conectadas por software, inteligência artificial, informações em tempo real e a Internet das Coisas (IoT), são componentes essenciais para o sucesso da Indústria 5.0.

A implementação de sistemas de decisão orientados por dados pode demandar investimentos substanciais em tecnologia, infraestrutura e capacitação. Para muitas organizações, especialmente pequenas e médias empresas, o custo inicial pode representar uma barreira significativa.

1.6. O Futuro dos Sistemas de Decisão na Indústria 5.0

A Indústria 5.0 representa uma mudança de foco, passando de uma abordagem puramente econômica para uma ênfase no valor social e no bem-estar. Ela abrange diversos processos, não se limitando apenas à fabricação, e é viabilizada por avanços tecnológicos, incluindo inteligência artificial, automação, dados, Internet das Coisas e aprendizado de máquina [34].

Conforme apontado por Barata e Kayser [35], as principais correntes na literatura sobre a Indústria 5.0 são orientadas para o futuro e intersetoriais. Enquanto a Indústria 4.0 utiliza tecnologias digitais para resolver problemas específicos de fabricação e melhorar a produtividade, a Indústria 5.0 busca um propósito mais amplo, não apenas com uma visão mais centrada no cliente, mas também com um foco maior no ser humano. Pode-se, portanto, argumentar que a Indústria 5.0 não é guiada pela tecnologia, mas sim pelo valor. A Indústria 5.0 enxerga o ser humano, por um lado, como um consumidor que espera

produtos personalizados em larga escala e, por outro, como um trabalhador qualificado que interage com robôs para viabilizar essa personalização.

Na Indústria 5.0, a ênfase estará na colaboração entre humanos e máquinas, onde as decisões orientadas por dados serão combinadas com a intuição e a criatividade humana para gerar soluções mais inovadoras e personalizadas. Os sistemas de decisão se tornarão mais autônomos e capazes de aprender com os dados em tempo real, adaptando-se rapidamente às mudanças no ambiente de negócios e nas demandas do mercado.

Com o aumento do uso de decisões automatizadas, surgirá a necessidade de garantir que essas decisões sejam tomadas de maneira ética e transparente. A explicabilidade dos algoritmos de decisão será crucial para conquistar a confiança dos *stakeholders*. O futuro dos sistemas de decisão também estará alinhado com práticas de sustentabilidade, onde as decisões não apenas maximizam o lucro, mas também consideram os impactos ambientais e sociais, promovendo uma indústria mais responsável.

1.7. Estudo de Caso

A transformação digital na indústria tem impulsionado a adoção de tecnologias avançadas, como a inteligência artificial (IA) e a Internet das Coisas (IoT), no monitoramento e na manutenção de equipamentos industriais. No contexto da Indústria 5.0, que enfatiza a colaboração entre humanos e máquinas inteligentes, a manutenção preditiva emerge como uma prática essencial para maximizar a eficiência operacional e reduzir custos associados a falhas inesperadas. Este estudo de caso teórico foca na implementação de um Sistema de Recomendação para Manutenção Preditiva em uma fábrica moderna, alinhada aos princípios da Indústria 5.0.

A empresa analisada, referida como Indústria XYZ, opera em um setor onde a continuidade dos processos produtivos é importante para a competitividade no mercado. Com o objetivo de minimizar o tempo de inatividade e otimizar a utilização dos recursos, a empresa enfrenta o desafio de integrar um sistema que não apenas preveja falhas iminentes, mas também recomende ações corretivas de forma proativa. Para isso, é necessário desenvolver um sistema inteligente capaz de analisar dados sensoriais em tempo real e fornecer recomendações precisas para manutenção, utilizando técnicas de aprendizado de máquina e análise preditiva [18].

Este estudo de caso explora teoricamente cada etapa do desenvolvimento do sistema, desde a preparação dos dados, passando pelo treinamento dos modelos de aprendizado de máquina, até a implementação e validação em um ambiente de produção real. O objetivo principal é demonstrar como a integração de tecnologias avançadas e o conhecimento especializado dos operadores podem transformar processos industriais, tornando-os mais resilientes, personalizados e eficientes. A capacidade de prever falhas e tomar decisões rápidas, fundamentadas em dados robustos e em tempo real, representa uma vantagem competitiva significativa para empresas que buscam se adaptar às exigências da Indústria 5.0 [19].

1.7.1. Fundamentos e Preparação de Dados

A Figura 1.4 ilustra o fluxo completo de preparação de dados, desde a coleta inicial até a divisão final dos dados em conjuntos de treino, validação e teste. Esta imagem proporciona uma visão clara das interconexões entre as diferentes etapas e como elas se complementam para formar um pipeline de dados eficiente e robusto.



Figura 1.4: Fluxo de preparação de dados.

A primeira etapa na construção de um sistema de recomendação para manutenção preditiva é a preparação dos dados. A coleta de dados é importante e deve ser realizada de maneira sistemática e contínua, utilizando sensores instalados nas máquinas e equipamentos ao longo da linha de produção. Esses sensores devem monitorar variáveis críticas, como temperatura, vibração e pressão, que são indicativas do estado de funcionamento das máquinas e da qualidade do produto final [36].

Após a coleta, os dados brutos frequentemente contêm inconsistências, valores ausentes e ruídos que podem comprometer a eficácia do modelo. Por isso, a limpeza dos dados é uma etapa fundamental. Ela envolve a remoção de entradas duplicadas, o tratamento de valores ausentes por meio de imputação e a padronização dos dados. A limpeza dos dados assegura que o conjunto de dados seja robusto e confiável, o que é essencial para o sucesso das etapas subsequentes de modelagem [37].

Em seguida, deve-se realizar a engenharia de *features*. Esta etapa é vital para identificar e selecionar as variáveis que realmente impactam o desempenho do modelo. Em um cenário típico, como o da Indústria XYZ, onde há uma grande quantidade de recursos disponíveis, a seleção de *features* ajuda a reduzir a complexidade do modelo, concentrando-se nas variáveis mais relevantes. A engenharia de *features* é particularmente crítica em sistemas de manutenção preditiva, como demonstrado na indústria automotiva, onde a seleção e transformação adequadas de *features* podem melhorar significativamente a precisão das previsões de falhas e aumentar a longevidade dos equipamentos [38].

Técnicas como a relação sinal-ruído, testes estatísticos e métodos de seleção como *Relief* e *Pearson* são comumente aplicadas para determinar as variáveis que devem ser mantidas no modelo. Esses métodos são utilizados para a seleção de *features* em vários contextos, incluindo a classificação do câncer de próstata, onde essas técnicas ajudam a identificar as variáveis mais significativas para a construção de modelos preditivos precisos [39].

A normalização e padronização dos dados são igualmente importantes para garantir que diferentes tipos de dados possam ser comparados e utilizados de maneira eficaz pelos modelos de aprendizado de máquina. A normalização ajusta os dados para uma escala comum, enquanto a padronização assegura que as variáveis tenham distribuições semelhantes, facilitando o processo de aprendizado dos algoritmos e garantindo que nenhuma variável domine o modelo por estar em uma escala maior [40].

Finalmente, a divisão dos dados em conjuntos de treino, validação e teste é necessária para garantir a eficácia do modelo. A validação cruzada [41] é uma técnica comum utilizada para avaliar o desempenho do modelo, permitindo que ele seja testado em diferentes segmentos do conjunto de dados. Essa abordagem ajuda a evitar o sobreajuste, garantindo que o modelo possa generalizar bem para novos dados, um requisito crucial para a implementação em um ambiente de produção real [42].

1.7.2. Desenvolvimento de Modelos de *Machine Learning*

O desenvolvimento de um sistema de recomendação para manutenção preditiva envolve um ciclo iterativo contínuo, conforme ilustrado na Figura 1.5. Este ciclo começa com a seleção dos algoritmos de aprendizado de máquina mais adequados, seguido pelo treinamento dos modelos com os dados disponíveis. Após o treinamento, o modelo passa por ajustes de hiperparâmetros e validação cruzada para otimizar seu desempenho. A etapa final do ciclo envolve a avaliação do modelo. Dependendo dos resultados obtidos, o ciclo pode recomeçar, permitindo ajustes contínuos e refinamentos no modelo, garantindo sua robustez e eficácia em um ambiente de produção. Esse processo iterativo é essencial para o desenvolvimento eficaz de modelos de aprendizado de máquina [43]. Além disso, métricas como F1 Score são cruciais para fornecer uma avaliação equilibrada do desempenho do modelo, especialmente em cenários com desequilíbrio entre as classes, onde é necessário um equilíbrio entre precisão e recall [44].

Esse ciclo iterativo é essencial para assegurar que o modelo final atenda às exigências de precisão e adaptabilidade, características fundamentais na Indústria 5.0. Nos parágrafos a seguir, são detalhadas as principais etapas desse fluxo de desenvolvimento, desde a seleção dos algoritmos até a avaliação do modelo, com ênfase nos métodos e técnicas aplicadas em cada fase.

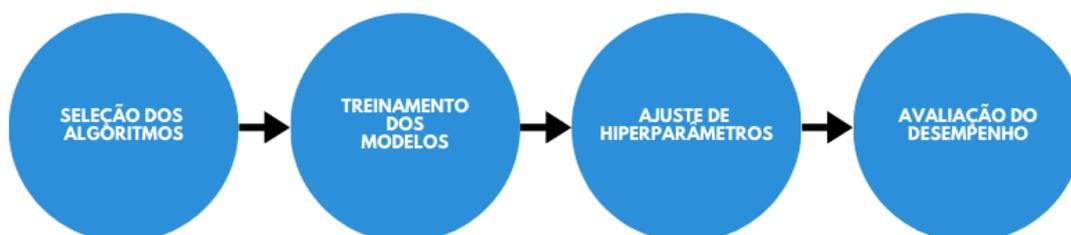


Figura 1.5: Ciclo iterativo no desenvolvimento de um sistema de recomendação para manutenção preditiva, ilustrando as etapas de seleção de algoritmos, treinamento dos modelos, ajuste de hiperparâmetros e avaliação do desempenho.

A primeira etapa no desenvolvimento de um sistema de recomendação para manutenção preditiva é a seleção dos algoritmos mais adequados para o problema em questão. No contexto da Indústria 5.0, onde a precisão e a capacidade de adaptação são essenciais, algoritmos como regressão logística [45], árvores de decisão [46], florestas aleatórias [47] e redes neurais profundas [48] são frequentemente considerados.

Cada um desses algoritmos possui características únicas que os tornam mais ou menos adequados para diferentes tipos de problemas. Por exemplo, redes neurais pro-

fundas são extremamente eficazes em capturar padrões complexos em grandes volumes de dados, mas podem ser mais desafiadoras para treinar e interpretar. Por outro lado, as árvores de decisão são mais simples e interpretáveis, mas podem sofrer de sobreajuste se não forem reguladas adequadamente. Portanto, a escolha do algoritmo deve equilibrar a precisão, interpretabilidade e a capacidade de generalização.

O treinamento dos modelos é a fase em que os algoritmos são ajustados aos dados de treinamento. Durante o treinamento, os algoritmos aprendem a partir dos padrões presentes nos dados, ajustando seus parâmetros para minimizar o erro nas previsões. Técnicas como o ajuste de hiperparâmetros e a validação cruzada são comumente utilizadas nesta fase para otimizar o desempenho do modelo [44]. O objetivo é desenvolver um modelo que não apenas tenha um bom desempenho no conjunto de treinamento, mas que também seja capaz de generalizar para novos dados, garantindo sua eficácia em um ambiente de produção.

A avaliação do desempenho dos modelos é realizada utilizando o conjunto de validação e, posteriormente, o conjunto de teste. Métricas como acurácia, precisão, *recall* e a métrica F1 são empregadas para avaliar a eficácia do modelo. A escolha das métricas depende dos objetivos específicos do sistema de manutenção preditiva. Por exemplo, em um ambiente onde a detecção de falhas é crítica, o *recall* pode ser mais importante que a precisão, pois assegura que a maioria das falhas seja detectada, mesmo que isso resulte em alguns falsos positivos.

1.7.3. Implementação e Testes do Sistema de Decisão

Após o desenvolvimento e a avaliação dos modelos de aprendizado de máquina, a próxima etapa é a implementação do sistema de decisão em um ambiente de produção. A Figura 1.6 ilustra as etapas envolvidas nesse processo, desde a integração inicial do modelo até a manutenção contínua após a implementação. Esse fluxo de trabalho cíclico garante que o sistema permaneça eficaz e adaptável às necessidades em constante evolução da Indústria XYZ.



Figura 1.6: Fluxo das etapas de implementação de um sistema de decisão para manutenção preditiva, começando com a integração do modelo ao sistema de monitoramento, seguido pelo desenvolvimento da interface do usuário, testes em ambiente controlado, e finalizando com a manutenção contínua e monitoramento pós-implementação.

Essa fase inicia com a integração do modelo ao sistema de monitoramento existente na Indústria XYZ. Isso inclui a configuração de interfaces para a coleta contínua de

dados sensoriais, o processamento em tempo real e a geração automatizada de recomendações de manutenção.

A interface do usuário desempenha um papel crucial no sistema, pois é por meio dela que os operadores interagem com as recomendações geradas pelo modelo. Uma interface bem projetada deve ser intuitiva, facilitando a compreensão das previsões pelos operadores e permitindo que eles tomem decisões informadas de forma ágil. A visualização clara dos dados e das recomendações é essencial para garantir que os operadores possam interpretar rapidamente as informações e tomar ações corretivas imediatas. A importância de uma interface de usuário eficaz é destacada na literatura como um fator essencial para aumentar a confiança dos operadores nas decisões automatizadas e melhorar a adoção da tecnologia [49].

Antes da implantação completa, o sistema deve ser submetido a testes rigorosos em um ambiente controlado. Esses testes envolvem a simulação de diferentes cenários de falhas e a avaliação da resposta do sistema a essas situações. Esse processo é fundamental para identificar quaisquer ajustes necessários no modelo ou na interface do usuário, garantindo que o sistema funcione conforme o esperado em condições reais de operação.

A manutenção contínua do sistema após a implementação é igualmente essencial. O sistema deve ser monitorado constantemente para garantir sua eficácia a longo prazo. Atualizações regulares nos modelos e ajustes baseados em novos dados são necessários para manter o sistema relevante e eficaz. A capacidade de adaptação do sistema, aliada à sua robustez, permite que a Indústria XYZ opere de forma eficiente e segura, minimizando o tempo de inatividade e os custos associados a falhas inesperadas. A importância dessa abordagem é destacada na literatura, que demonstra como a manutenção contínua e as atualizações são cruciais para reduzir o tempo de inatividade e evitar custos adicionais causados por falhas inesperadas [50].

1.8. Conclusões

A evolução da Indústria 4.0 para a Indústria 5.0 representa uma mudança paradigmática na forma como as empresas abordam a produção e a gestão de processos. Enquanto a Indústria 4.0 se concentrou na automação e na eficiência através da digitalização e da conectividade, a Indústria 5.0 vai além, buscando reintegrar o ser humano no centro dos processos produtivos. Esta nova fase industrial não apenas adota tecnologias avançadas como a inteligência artificial, aprendizado de máquina, e Internet das Coisas (IoT), mas também valoriza a colaboração entre humanos e máquinas, promovendo um ambiente de trabalho mais personalizado, sustentável e ético.

Neste trabalho, exploramos a implementação de sistemas de decisão baseados em dados como uma aplicação prática dos princípios da Indústria 5.0. O estudo de caso da Indústria XYZ, que desenvolveu um sistema de recomendação para manutenção preditiva, exemplifica como a integração de tecnologia avançada com o conhecimento humano pode otimizar processos industriais. Através de uma análise detalhada das etapas de preparação de dados, desenvolvimento de modelos de aprendizado de máquina, e implementação prática do sistema, foi demonstrado que essas tecnologias podem transformar a maneira como as empresas operam, tornando-as mais resilientes, eficientes e adaptáveis às demandas do mercado.

A Indústria 5.0 oferece inúmeras vantagens, incluindo a capacidade de personalizar produtos e serviços de acordo com as necessidades individuais dos clientes, reduzir o impacto ambiental através da economia circular e promover um ambiente de trabalho mais colaborativo. No entanto, também apresenta desafios significativos, como a necessidade de mudanças culturais nas empresas, o aumento da demanda por profissionais qualificados, e a complexidade da integração de sistemas tecnológicos avançados. Este trabalho destaca a importância de abordar esses desafios de forma proativa, garantindo que as empresas possam tirar o máximo proveito das oportunidades oferecidas pela Indústria 5.0.

Trabalhos futuros devem se concentrar em várias áreas para expandir o impacto da Indústria 5.0. Primeiramente, é essencial desenvolver novas metodologias e ferramentas que facilitem a integração harmoniosa entre humanos e máquinas, assegurando que a tecnologia realmente amplifique as capacidades humanas. Além disso, pesquisas adicionais são necessárias para explorar como os sistemas de decisão baseados em dados podem ser aplicados em diferentes setores industriais, ampliando seu alcance e eficácia. Outra área promissora é o desenvolvimento de algoritmos de inteligência artificial que não apenas aprendam e adaptem-se em tempo real, mas também garantam transparência e ética em suas decisões, ganhando a confiança dos usuários. Por fim, estudos devem investigar como as empresas podem superar as barreiras culturais e tecnológicas para adotar plenamente as práticas da Indústria 5.0, assegurando que essa evolução industrial seja inclusiva, sustentável e benéfica para todos os envolvidos.

Referências

- [1] Edvânia da Glória Silva. *O efeito da indústria 5.0 na produção*. PhD thesis, 2023.
- [2] Henning Kagermann, Johannes Helbig, Ariane Hellinger, and Wolfgang Wahlster. *Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group*. Forschungsunion, 2013.
- [3] Carolina Narvaez Rojas, Gustavo Adolfo Alomia Peñafiel, Diego Fernando Loaiza Buitrago, and Carlos Andrés Tavera Romero. Society 5.0: A japanese concept for a superintelligent society. *Sustainability*, 13(12):6567, 2021.
- [4] John Andrew van der Poll. Problematizing the adoption of formal methods in the 4ir–5ir transition. *Applied System Innovation*, 5(6):127, 2022.
- [5] Ray Y Zhong, Xun Xu, Eberhard Klotz, and Stephen T Newman. Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, 3(5):616–630, 2017.
- [6] Jiewu Leng, Weinan Sha, Baicun Wang, Pai Zheng, Cunbo Zhuang, Qiang Liu, Thorsten Wuest, Dimitris Mourtzis, and Lihui Wang. Industry 5.0: Prospect and retrospect. *Journal of Manufacturing Systems*, 65:279–295, 2022.
- [7] James A O'brien and George M Marakas. *Administração de sistemas de informação*. AMGH, 2013.

- [8] Tharcylla Rebecca Negreiros Clemente Pedro Henrique Cavalcanti Lins Andre Gardi Koury Marcelo Hazin Alencar Cristiano Alexandre Virginio Cavalcante. Sistema de apoio a decisão para definição de estratégias de manutenção a partir da abordagem mcc. In *XXXI ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, Belo Horizonte, Minas Gerais, oct 2011.
- [9] George Anthony Gorry and Michael S Scott Morton. A framework for management information systems. 1971.
- [10] R. Bloemen and J. Maes. A dss for optimizing the aggregate production planning at monsanto antwerp. *European Journal of Operational Research*, 61(1):30–40, 1992. IFORS-SPC Conference on Decision Support Systems.
- [11] M Bruccoleri, G Lo Nigro, S La Noto Diega, P Renna, and G Perrone. A dss for strategic planning. *Design of Advanced Manufacturing Systems: Models for Capacity Planning in Advanced Manufacturing Systems*, pages 37–71, 2005.
- [12] Daniel J Power. *Decision support systems: concepts and resources for managers*, volume 13. Quorum Books Westport, 2002.
- [13] Claus Weihs and Katja Ickstadt. Data science: the impact of statistics. *International Journal of Data Science and Analytics*, 6:189–194, 2018.
- [14] Muhammet Atalay and Enes Çelik. Büyük veri analizinde yapay zeka ve makine öğrenmesi uygulamaları - artificial intelligence and machine learning applications in big data analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22):155–172, 2017.
- [15] Mayra dos Santos Guidorizzi Ana Paula Bispo dos Santos André Barbosa de Oliveira Fabrizio Leonardi. Simulação como apoio a tomada de decisão para a solução de problemas causados pelos gargalos formados no trânsito. In *XXIX ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO*, Salvador, Bahia, oct 2009.
- [16] George S Fishman. *Discrete-event simulation: modeling, programming, and analysis*, volume 537. Springer, 2001.
- [17] Praveen Kumar Reddy Maddikunta, Quoc-Viet Pham, Prabadevi B, N Deepa, Kapal Dev, Thippa Reddy Gadekallu, Rukhsana Ruby, and Madhusanka Liyanage. Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, 26:100257, 2022.
- [18] Ashok Vajravelu, Yamunarani Thanikachalam, Mohd Helmy Bin Abd Wahab, Muhammad Mahadi Abdul Jamil, and S. Sivaranjani. Human-machine collaboration and emotional intelligence in industry 5.0. *Advances in computational intelligence and robotics book series*, pages 220–232, 2024.
- [19] Andreas Eschbach. Postagem do conselho: Como o setor 5.0 transformará a manufatura de processos como a conhecemos, 2021. Acessado em: 12 de agosto de 2024.

- [20] Bahar Asgarova, Elvin Jafarov Elvin Jafarov, Nicat Babayev, Allahshukur Ahmadzada, Abdullayev V.H, and Triwiyanto Triwiyanto. Development process of decision support systems using data mining technology. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(1):703–703, 2024.
- [21] Shima Maniyan, Rouzbeh Ghousi, and Abdorrahman Haeri. Data mining-based decision support system for educational decision makers: Extracting rules to enhance academic efficiency. 2024.
- [22] Gerald Van Amburg. Management information system (mis), 2022.
- [23] Arihant Chhajer. Expert systems for emulating the decision making ability of a human expert. *TechnoareteTransactions on Intelligent Data Mining and Knowledge Discovery*, 1(1):15–18, 2022.
- [24] Ewerton Alex Avelar and Ricardo Vinícius Dias Jordão. The role of artificial intelligence in the decision-making process: a study on the financial analysis and movement forecasting of the world's largest stock exchanges. *Management Decision*, 2024.
- [25] Ahmad Talal Ali. Systems and methods for predicting environmental conditions, 2021.
- [26] Mariusz Żytniewski. Autopoietic knowledge management systems. pages 364–379. IGI Global, 2020.
- [27] Dr. Ram Bajaj. (6) "mastering the art of strategic planning: A blueprint for success". *Journal for ReAttach therapy and developmental diversities*, 2023.
- [28] Andriy Zaverbnyj and Vitalii Ilnytskyi. Phases of the life cycle of projects as a key factor in the success of their implementation. *Infrastruktura rinku*, (78), 2023.
- [29] Sri Wulandari and Hapzi Ali. Determination of management decision quality: Integration, effectiveness, development of management information systems. *Dinasti International Journal of Management Science (DIJMS)*, 2023.
- [30] Anaïs Barbier, C. Rousselière, Laurine Robert, E Cousein, and Bertrand Décaudin. Elaboration d'un guide méthodologique sur l'implantation d'un système d'aide à la décision pharmaceutique: retour d'expérience d'un centre hospitalier universitaire français. *Annales pharmaceutiques françaises*, 81(1):163–172, 2022.
- [31] Oleksandr V. Darushyn, Sergiy Bodenчук, and Ivan Buzian. Development and implementation technology of managerial decisions in enterprise management. *Naukovij poglâd: ekonomika ta upravlinnâ*, (2(86)), 2023.
- [32] Rosário Macário, Vasco Reis, and Juan Pablo Antún. Assessment at different stages of implementation. pages 233–243. Elsevier BV, 2023.
- [33] Tim Stobierski. The advantages of data-driven decision-making, 2019. Disponível em: <https://online.hbs.edu/blog/post/data-driven-decision-making>. Acesso em: 25 ago. 2024.

- [34] Ecanorea. Industry 5.0: The new business revolution, 2023. Disponível em: <https://www.plainconcepts.com/industry-50/>. Acesso em: 25 ago. 2024.
- [35] João Barata and Ina Kayser. Industry 5.0 – past, present, and near future. *Procedia Computer Science*, 219:778–788, 2023. CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2022.
- [36] Mudita Uppal, Deepali Gupta, Nitin Goyal, Agbotiname Lucky Imoize, Arun Kumar, Stephen Ojo, Subhendu Kumar Pani, Yongsung Kim, and Jaeun Choi. A real-time data monitoring framework for predictive maintenance based on the internet of things. *Complexity*, 2023(1):9991029, 2023.
- [37] Idrus Assagaf, Agus Sukandi, and Abdul Azis Abdillah. Machine failure detection using deep learning. *Recent in Engineering Science and Technology*, 1(03):26–31, Jul. 2023.
- [38] Venkata Sushma Chinta, Sowmya Kethi Reddi, and Nagini Yarramsetty. Optimal feature selection on serial cascaded deep learning for predictive maintenance system in automotive industry with fused optimization algorithm. *Advanced Engineering Informatics*, 57:102105, 2023.
- [39] Sunil Kumar Prabhakar and Seong-Whan Lee. Transformation based tri-level feature selection approach using wavelets and swarm computing for prostate cancer classification. *IEEE Access*, 8:127462–127476, 2020.
- [40] Marcellino Bonamutial and Simeon Yuda Prasetyo. Exploring the impact of feature data normalization and standardization on regression models for smartphone price prediction. In *2023 International Conference on Information Management and Technology (ICIMTech)*, pages 294–298, 2023.
- [41] Payam Refaeilzadeh, Lei Tang, and Huan Liu. *Cross-Validation*, pages 532–538. Springer US, Boston, MA, 2009.
- [42] İbrahim Yazıcı and Emre Gures. A novel approach for machine learning-based load balancing in high-speed train system using nested cross validation. In *2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pages 1–6, 2023.
- [43] Ramona Leenings, Nils Ralf Winter, Lucas Plagwitz, Vincent Holstein, Jan Ernsing, Kelvin Sarink, Lukas Fisch, Jakob Steenweg, Leon Kleine-Vennekate, Julian Gebker, Daniel Emden, Dominik Grotegerd, Nils Opel, Benjamin Risse, Xiaoyi Jiang, Udo Dannlowski, and Tim Hahn. Photonai—a python api for rapid machine learning model development. *PLOS ONE*, 16(7):1–19, 07 2021.
- [44] Sina Shafieezadeh, Gian Marco Duma, Giovanni Mento, Alberto Danieli, Lisa Antoniazzi, Fiorella Del Popolo Cristaldi, Paolo Bonanni, and Alberto Testolin. Methodological issues in evaluating machine learning models for eeg seizure prediction:

Good cross-validation accuracy does not guarantee generalization to new patients. *Applied Sciences*, 13(7), 2023.

- [45] Michael P. LaValley. Logistic regression. *Circulation*, 117(18):2395–2399, 2008.
- [46] Barry de Ville. Decision trees. *WIREs Computational Statistics*, 5(6):448–455, 2013.
- [47] Adele Cutler, D. Richard Cutler, and John R. Stevens. *Random Forests*, pages 157–175. Springer New York, New York, NY, 2012.
- [48] Daniel Durstewitz, Georgia Koppe, and Andreas Meyer-Lindenberg. Deep neural networks in psychiatry. *Molecular Psychiatry*, 24(11):1583–1598, Nov 2019.
- [49] Md Abdul Kadir, Abdulrahman Mohamed Selim, Michael Barz, and Daniel Sonntag. A user interface for explaining machine learning model explanations. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23 Companion, page 59–63, New York, NY, USA, 2023. Association for Computing Machinery.
- [50] Mounia Achouch, Mariya Dimitrova, Rizck Dhouib, Hussein Ibrahim, Mehdi Adda, Sasan Sattarpanah Karganroudi, Khaled Ziane, and Ahmad Aminzadeh. Predictive maintenance and fault monitoring enabled by machine learning: Experimental analysis of a ta-48 multistage centrifugal plant compressor. *Applied Sciences*, 13(3), 2023.

Capítulo

4

Estratégias para Lidar com Desbalanceamento de Dados em Aprendizado de Máquina

Hector Batista Ribeiro, Leandro Oliveira da Silva e Ricardo de Andrade Lira Rabêlo

Abstract

Data imbalance is a critical issue in machine learning, impacting model performance due to uneven class representation. This chapter explores the causes and effects of imbalance, discussing preprocessing techniques such as resampling and synthetic data generation. It also covers modeling methods and weight adjustments to enhance performance. Furthermore, the chapter reviews specific evaluation metrics and advanced methodologies, including ensemble methods and reinforcement learning, providing practical insights for addressing imbalance and developing more balanced, accurate models.

Resumo

O desbalanceamento de dados é um desafio crucial em aprendizado de máquina, onde a desigualdade na representação das classes afeta a eficácia dos modelos preditivos. Este capítulo explora as causas e consequências do desbalanceamento, destacando técnicas de pré-processamento como a reamostragem e geração de dados sintéticos. Também aborda métodos de modelagem e ajustes de pesos para melhorar o desempenho. Além disso, examina métricas de avaliação específicas e metodologias avançadas, como ensemble methods e aprendizado por reforço, oferecendo uma visão prática para lidar com o problema e desenvolver modelos mais equilibrados e precisos.

4.1. Introdução

O desbalanceamento de dados é um desafio intrínseco e amplamente reconhecido na prática de aprendizado de máquina, caracterizado pela disparidade na distribuição das classes em um

conjunto de dados [Chaudhary 2023]. Em muitos cenários do mundo real, os dados disponíveis para treinamento de modelos preditivos não são distribuídos uniformemente entre as diferentes categorias ou classes, resultando em um desequilíbrio significativo [Azank 2020]. Este fenômeno não é meramente uma peculiaridade estatística, mas uma questão que pode afetar a eficácia e a precisão dos algoritmos de aprendizado de máquina de maneira substancial.

O desbalanceamento de dados refere-se a uma situação em que uma ou mais classes em um conjunto de dados são representadas com muito mais frequência do que outras. Em contextos binários, isso se traduz em uma classe majoritária que é muito mais abundante do que a classe minoritária [Chaudhary 2023]. Em cenários multiclasse, o problema pode se manifestar como uma discrepância significativa na frequência de ocorrência entre as classes. Esse desbalanceamento pode levar a uma série de desafios, como a degradação da performance dos modelos preditivos e a dificuldade em generalizar para dados não vistos.

A relevância do desbalanceamento de dados se manifesta em diversas áreas e aplicações práticas. Por exemplo, na detecção de fraudes financeiras, os eventos fraudulentos são relativamente raros em comparação com transações legítimas. Sem um tratamento adequado, um modelo treinado em um conjunto de dados desbalanceado pode apresentar um viés sistemático em favor da classe majoritária, comprometendo sua capacidade de identificar fraudes de forma eficaz. Similarmente, em diagnósticos médicos, a identificação de doenças raras pode ser prejudicada se a classe de interesse estiver sub-representada no conjunto de dados de treinamento [Azank 2020].

O impacto do desbalanceamento de dados nos modelos de aprendizado de máquina é multifacetado e pode afetar significativamente a performance dos sistemas preditivos [Hasib et al. 2020]. Quando um modelo é treinado em um conjunto de dados desbalanceado, ele tende a aprender a reconhecer predominantemente a classe majoritária, muitas vezes à custa da capacidade de identificar a classe minoritária. Esse viés pode se manifestar em métricas de desempenho, como precisão, revocação e pontuação F1, que podem ser enganosas se avaliadas apenas com base na classe majoritária [L. A. Jeni et al. 2013].

Além disso, o desbalanceamento pode levar a uma avaliação inadequada do modelo. Por exemplo, um modelo que apresenta uma alta taxa de precisão pode ainda ter um desempenho fraco na detecção da classe minoritária, resultando em uma alta taxa de falsos negativos [Hasib et al. 2020]. Em contextos críticos, como o monitoramento de segurança e o diagnóstico de doenças, isso pode ter consequências graves, como a falha em identificar eventos adversos ou condições de saúde que requerem atenção urgente.

Em diversas aplicações reais, o desbalanceamento de dados é a norma e não a exceção. Na área da saúde, por exemplo, conjuntos de dados usados para diagnosticar doenças raras muitas vezes contêm muito mais exemplos de indivíduos saudáveis do que de indivíduos doentes [A. K. Ilavarasi 2020]. Da mesma forma, em sistemas de detecção de fraudes financeiras, as transações legítimas são vastamente mais numerosas do que as fraudulentas [Rubaidi Z. 2022]. Em ambos os casos, a incapacidade de detectar as classes minoritárias pode ter consequências significativas, seja na forma de diagnósticos incorretos ou perdas financeiras substanciais.

A necessidade de tratar o desbalanceamento de dados se torna ainda mais crucial à medida que os sistemas de aprendizado de máquina são cada vez mais integrados em processos críticos de tomada de decisão. Sem técnicas adequadas para lidar com o desbalanceamento, os modelos podem apresentar uma performance ilusória de alta precisão, enquanto na verdade falham em detectar ou prever os casos mais críticos e raros. Portanto, é imperativo que cientistas de dados e engenheiros de aprendizado de máquina estejam equipados com estratégias eficazes para mitigar os efeitos do desbalanceamento de dados.

A abordagem do desbalanceamento envolve uma combinação de técnicas de pré-processamento de dados, ajustes de modelagem e métricas de avaliação específicas. A utilização dessas técnicas não apenas melhora a performance dos modelos, mas também contribui para a criação de sistemas preditivos mais justos e equilibrados [Thabtah F. 2019]. As seções seguintes deste capítulo irão explorar em detalhes as diferentes abordagens para lidar com o desbalanceamento de dados, incluindo técnicas de reamostragem, geração de dados sintéticos, ajustes de pesos e métodos avançados como o aprendizado por reforço.

Neste capítulo, exploraremos uma variedade de técnicas e abordagens que foram desenvolvidas para enfrentar o problema do desbalanceamento de dados. Estas técnicas podem ser categorizadas em três grandes áreas:

Técnicas de Pré-processamento de Dados: Incluem métodos como reamostragem (oversampling e undersampling), aumento de dados e seleção de características que ajudam a equilibrar as classes antes da modelagem.

Técnicas de Modelagem: Envolvem a modificação dos algoritmos de aprendizado de máquina ou a introdução de novos algoritmos que são mais robustos a dados desbalanceados, incluindo métodos baseados em custo e o uso de métricas de avaliação específicas.

Métodos Avançados: Abrangem abordagens mais recentes e sofisticadas, como aprendizado semi-supervisionado, deep learning e a utilização de arquiteturas especializadas para lidar com o desbalanceamento.

O capítulo será organizado para proporcionar uma compreensão abrangente e prática sobre como lidar com o desbalanceamento de dados. Começaremos com uma discussão sobre a natureza do desbalanceamento, como ele pode ser quantificado e os problemas que ele causa. Em seguida, detalharemos as técnicas de pré-processamento de dados, como reamostragem, gerenciamento de dados e aumento de dados. Exploraremos algoritmos e métodos específicos para dados desbalanceados, bem como as métricas apropriadas para avaliar modelos treinados com esses dados. Analisaremos técnicas avançadas, incluindo aprendizado profundo e semi-supervisionado. Apresentaremos estudos de caso com exemplos reais e lições aprendidas. Introduziremos ferramentas e bibliotecas populares usadas para lidar com desbalanceamento de dados, incluindo exemplos de código. Finalmente, concluiremos com um resumo das principais técnicas discutidas e uma análise dos desafios e tendências futuras na área.

4.2. Tipos de Desbalanceamento

O desbalanceamento de dados é um fenômeno que pode assumir diversas formas, cada uma com suas implicações e desafios específicos. Neste capítulo, exploraremos os diferentes tipos de desbalanceamento, fornecendo uma compreensão mais profunda das suas características e como eles afetam o desempenho dos modelos de aprendizado de máquina. A compreensão desses tipos é fundamental para selecionar as técnicas apropriadas para lidar com o desbalanceamento em diferentes cenários.

4.2.1. Desbalanceamento de Classes

O tipo mais comum de desbalanceamento de dados é o desbalanceamento de classes. Este ocorre quando as diferentes classes em um conjunto de dados têm quantidades significativamente variadas de exemplos [Longadge R. 2013]. Por exemplo, em um conjunto de dados de diagnóstico médico, pode haver milhares de casos de pacientes saudáveis e apenas alguns casos de uma doença rara. Esse desbalanceamento pode levar os modelos a se tornarem altamente enviesados em favor da classe majoritária, resultando em baixa capacidade de detecção da classe minoritária. É crucial reconhecer que o desbalanceamento de classes pode afetar diversas métricas de desempenho do modelo, como a precisão, recall e F1-score, de maneira desproporcional.

4.2.2. Desbalanceamento em Séries Temporais

O desbalanceamento em séries temporais refere-se a situações onde os dados desbalanceados não são apenas desiguais em quantidade, mas também distribuídos de maneira não uniforme ao longo do tempo [H. Cao et al. 2013]. Em cenários de séries temporais, como a detecção de anomalias em dados financeiros ou sistemas de monitoramento industrial, eventos raros podem ocorrer esporadicamente. Por exemplo, uma falha em um sistema de monitoramento pode ser um evento raro que ocorre apenas ocasionalmente, enquanto o comportamento normal é muito mais frequente. O desafio aqui é lidar com a variação temporal e garantir que o modelo possa detectar esses eventos raros sem se perder nos dados normais.

4.2.3. Desbalanceamento em Dados Multiclasse

Em problemas de classificação multiclasse, o desbalanceamento pode ocorrer quando algumas classes estão muito mais representadas do que outras [Mathew R. and Gunasundari R. 2021]. Por exemplo, em um conjunto de dados para classificação de espécies de plantas, algumas espécies podem ter centenas de amostras, enquanto outras podem ter apenas algumas. Esse tipo de desbalanceamento pode tornar a tarefa de classificação mais complexa, pois o modelo pode ter dificuldade em aprender a distinguir as classes minoritárias corretamente. A abordagem para lidar com o desbalanceamento em problemas multiclasse pode exigir técnicas adaptadas, como reamostragem específica para cada classe ou o uso de métricas de avaliação que considerem a importância relativa de cada classe.

4.2.4. Desbalanceamento de Rótulos em Dados Multimodais

No contexto de dados multimodais, onde diferentes tipos de dados (como texto, imagem e áudio) são usados, o desbalanceamento pode ocorrer em diferentes modalidades de forma desigual [S. Pouyanfar et al. 2019]. Por exemplo, em um sistema de reconhecimento de

emoção que utiliza tanto imagens quanto texto, pode haver muito mais dados disponíveis para uma modalidade (como imagens) do que para outra (como texto). Esse desbalanceamento pode complicar a integração e a fusão de informações de diferentes fontes e exigir técnicas específicas para equilibrar a influência de cada modalidade.

Cada tipo de desbalanceamento apresenta seus próprios desafios e requer abordagens específicas para garantir que os modelos de aprendizado de máquina possam realizar previsões precisas e confiáveis. A compreensão desses diferentes tipos de desbalanceamento é essencial para a escolha e implementação das técnicas de tratamento adequadas. No próximo capítulo, abordaremos as técnicas de pré-processamento de dados que podem ser aplicadas para lidar com esses tipos de desbalanceamento, proporcionando estratégias para equilibrar os conjuntos de dados e melhorar o desempenho dos modelos.

4.3. Técnicas de Pré-processamento de Dados

No pré-processamento de dados, a reamostragem é uma técnica essencial para lidar com o desbalanceamento de classes. Esta técnica pode ser dividida em abordagens de oversampling, undersampling e métodos combinados. Além disso, a geração de dados sintéticos é uma estratégia avançada que inclui técnicas como SMOTE e ADASYN. Cada uma dessas abordagens tem suas próprias características e aplicações, que serão discutidas a seguir.

4.3.1. Reamostragem

A reamostragem busca equilibrar o número de exemplos entre classes majoritárias e minoritárias para melhorar a performance do modelo. Existem três abordagens principais para a reamostragem: *oversampling*, *undersampling* e métodos combinados.

4.3.1.1. Oversampling: Aumentando a Quantidade da Classe Minoritária

De acordo com [Thabtah 2019] e [Hasib et al. 2020], *oversampling* é uma técnica que aumenta a quantidade de exemplos da classe minoritária. O objetivo é criar um conjunto de dados mais equilibrado ao adicionar exemplos adicionais que ajudam o modelo a aprender melhor sobre a classe minoritária. Uma abordagem comum de oversampling é a geração de novos exemplos sintéticos para a classe minoritária, o que pode ajudar a melhorar a capacidade do modelo de identificar a classe minoritária sem alterar os exemplos da classe majoritária [Hasib et al. 2020].

4.3.1.2. Undersampling: Reduzindo a Quantidade da Classe Majoritária

De acordo com [Thabtah 2019] e [Hasib et al. 2020], *undersampling* envolve a redução da quantidade de exemplos da classe majoritária para equilibrar o conjunto de dados. Isso é feito removendo exemplos da classe majoritária até que a proporção de exemplos entre as classes seja mais equilibrada. Embora o undersampling possa reduzir o viés em favor da classe majoritária, é importante notar que essa abordagem pode levar à perda de informações valiosas e diminuir a capacidade do modelo de aprender sobre a variabilidade da classe majoritária [Thabtah 2019].

4.3.1.3. Métodos Combinados: Oversampling e Undersampling Combinados

Métodos combinados utilizam uma combinação de oversampling e undersampling para criar um conjunto de dados mais equilibrado, como visto em [Junsomboon N. and Phienthrakul T. 2017] e [H. Shamsudin et al. 2020]. Uma abordagem comum é aplicar oversampling na classe minoritária e undersampling na classe majoritária simultaneamente. Isso pode ajudar a mitigar os problemas de desbalanceamento enquanto mantém uma quantidade razoável de dados para ambas as classes. Métodos combinados podem oferecer uma solução mais equilibrada do que usar apenas uma das técnicas isoladamente.

4.3.1.4. Geração de Dados Sintéticos

A geração de dados sintéticos é uma abordagem avançada para lidar com o desbalanceamento de dados, especialmente quando é difícil ou impossível obter mais dados reais da classe minoritária [Hasib et al. 2020]. As técnicas de geração de dados sintéticos criam novos exemplos a partir dos existentes para aumentar a representação da classe minoritária.

4.3.1.5. SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE (Synthetic Minority Over-sampling Technique) é uma técnica popular para gerar exemplos sintéticos da classe minoritária. Em vez de simplesmente duplicar exemplos existentes, o SMOTE cria novos exemplos interpolando entre exemplos da classe minoritária [Chawla 2002]. Para cada exemplo na classe minoritária, o SMOTE seleciona seus vizinhos mais próximos e cria novos exemplos ao longo das linhas que conectam o exemplo original com seus vizinhos. Isso ajuda a criar uma fronteira de decisão mais suave e generalizável para a classe minoritária.

4.3.1.6. ADASYN (Adaptive Synthetic Sampling)

ADASYN (Adaptive Synthetic Sampling) é uma variação do SMOTE que se concentra em gerar mais exemplos sintéticos em regiões onde a densidade da classe minoritária é baixa. O ADASYN adapta a quantidade de dados sintéticos gerados com base na dificuldade de aprendizado em diferentes regiões do espaço de características [He 2008]. Ao gerar mais exemplos em regiões problemáticas, o ADASYN pode ajudar a melhorar a capacidade do modelo de aprender a distinguir a classe minoritária em áreas onde há maior confusão.

4.3.1.7. Exemplos e Aplicações

A reamostragem e a geração de dados sintéticos são amplamente utilizadas em várias áreas para enfrentar o desbalanceamento de dados.

- **Detecção de Fraude:** No setor financeiro, onde as fraudes são raras comparadas às transações legítimas, técnicas de oversampling como SMOTE são usadas para criar exemplos sintéticos de fraudes e melhorar a capacidade do modelo de detectar atividades fraudulentas, como visto em [Mqadi 2021].
- **Diagnóstico Médico:** Em diagnósticos de doenças raras, onde há poucos casos positivos em relação aos negativos, oversampling e técnicas de geração de dados sintéticos ajudam a equilibrar o conjunto de dados para melhorar a acurácia dos modelos de diagnóstico, como visto em [Kosolwattana T. et al. 2022].

- **Reconhecimento de Anomalias:** Em sistemas de monitoramento industrial, onde anomalias podem ocorrer raramente, undersampling da classe majoritária e oversampling da classe minoritária ajudam a criar um conjunto de dados equilibrado para treinar modelos que detectam falhas com mais eficácia, como visto em [J. -R. Jiang and Y. -T. Chen 2022].

Cada técnica tem suas vantagens e desvantagens, e a escolha da abordagem apropriada pode depender das características específicas do conjunto de dados e dos objetivos do projeto. No próximo capítulo, exploraremos como essas técnicas de pré-processamento se integram com os algoritmos de modelagem para enfrentar o desbalanceamento de dados de forma eficaz.

4.4. Técnicas de modelagem

Após o pré-processamento dos dados desbalanceados, o próximo passo crucial é selecionar e ajustar algoritmos de modelagem que possam lidar eficazmente com o desbalanceamento. Este capítulo explora as principais técnicas de modelagem que podem ser aplicadas para enfrentar o desbalanceamento de dados. Serão discutidos os algoritmos robustecidos para desbalanceamento, ajustes específicos em métodos baseados em árvores e redes neurais, e a importância do ajuste de pesos durante o treinamento.

4.4.1. Algoritmos Robustecidos para Desbalanceamento

Algoritmos robustecidos para desbalanceamento são projetados ou ajustados para operar eficientemente em cenários onde há uma disparidade significativa entre as classes. Esses algoritmos incorporam técnicas internas para mitigar o viés em favor da classe majoritária, tornando-os adequados para problemas de classificação com classes desbalanceadas.

4.4.1.1. Métodos Baseados em Árvores: XGBoost, Random Forest

Métodos baseados em árvores, como XGBoost e Random Forest, são particularmente populares devido à sua flexibilidade e capacidade de lidar com dados desbalanceados de forma eficaz. Na tabela abaixo é apresentada uma descrição de cada um deles.

Tabela 4.1. Métodos baseados em árvores.

Método	Definição
XGBoost (Extreme Gradient Boosting)	Algoritmo de boosting que combina várias árvores de decisão fracas para formar um modelo poderoso [Pristyanto Y. et al. 2023]. Para lidar com o desbalanceamento, o XGBoost permite o ajuste do parâmetro <code>scale_pos_weight</code> , que pondera o impacto das classes minoritárias no erro de treinamento. Além disso, o XGBoost suporta a amostragem de dados de forma que as classes minoritárias possam ser mais bem representadas durante o treinamento, mitigando o viés em favor da classe majoritária.

Random Forest	Algoritmo de ensemble que combina várias árvores de decisão independentes. Para lidar com o desbalanceamento, uma abordagem comum é ajustar a frequência de amostragem de cada árvore, de modo que as classes minoritárias sejam amostradas com mais frequência [M. Bader-El-Den et al. 2019]. Isso é feito ajustando o parâmetro <code>class_weight</code> durante o treinamento, garantindo que a importância das classes minoritárias seja aumentada, e o modelo seja mais sensível a elas.
---------------	--

4.4.1.2. Métodos Baseados em Redes Neurais: Arquiteturas e Ajustes Específicos

Redes neurais oferecem flexibilidade e poder para modelagem em cenários complexos, mas podem exigir ajustes específicos para lidar com desbalanceamento de dados.

Arquiteturas Profundas: Redes neurais profundas, como CNNs (Redes Neurais Convolucionais) e RNNs (Redes Neurais Recorrentes), podem ser adaptadas para lidar com desbalanceamento de dados [J. Justin and K. Taghi 2019]. Uma estratégia comum é adicionar camadas específicas que ajustam a sensibilidade do modelo para a classe minoritária. Por exemplo, em uma CNN para reconhecimento de imagens, pode-se ajustar os filtros convolucionais ou as camadas de pooling para destacar características mais representativas da classe minoritária.

Ajustes de Hiperparâmetros: Ajustar hiperparâmetros como a taxa de aprendizado e a regularização também pode ajudar a melhorar a performance em dados desbalanceados [Z. Fan et al. 2022]. A regularização L2, por exemplo, pode ser ajustada para evitar que o modelo se torne excessivamente confiante em previsões para a classe majoritária. O dropout também pode ser usado para reduzir o overfitting e melhorar a generalização do modelo em relação à classe minoritária.

4.4.2. Ajuste de Pesos

O ajuste de pesos é uma técnica eficaz para lidar com desbalanceamento ao influenciar diretamente o processo de aprendizado do modelo, garantindo que as classes minoritárias sejam tratadas de forma justa durante o treinamento.

4.4.2.1. Ponderação de Classes: Ajuste de Pesos Durante o Treinamento

A ponderação de classes envolve ajustar os pesos atribuídos a cada classe durante o processo de treinamento, de modo que o erro nas classes minoritárias tenha um impacto maior na atualização dos parâmetros do modelo. Isso pode ser implementado de várias maneiras:

Loss Function (Função de Perda): a função de perda pode ser ajustada para incluir pesos para cada classe [J. Justin and K. Taghi 2019]. Por exemplo, na entropia cruzada ponderada, cada classe recebe um peso inversamente proporcional à sua frequência no conjunto de dados. Isso força o modelo a penalizar mais os erros nas classes minoritárias, promovendo um aprendizado mais equilibrado [M. R. Rezaei-Dastjerdehei et al. 2020].

Custom Loss Functions: em casos mais complexos, funções de perda personalizadas podem ser criadas para incorporar diferentes penalidades para erros em classes minoritárias [J. Justin and K. Taghi 2019]. Por exemplo, em um cenário de detecção de fraudes, a perda pode ser ajustada para penalizar falsos negativos (fraudes não detectadas) muito mais do que falsos positivos, refletindo a gravidade do erro.

4.5. Técnicas de Avaliação

A avaliação de modelos treinados em conjuntos de dados desbalanceados requer a utilização de métricas e métodos que sejam sensíveis à distribuição desigual entre as classes. Este capítulo explora as principais técnicas de avaliação utilizadas para medir a performance de modelos em cenários de desbalanceamento de dados. Discutiremos métricas como a Curva ROC e AUC, a matriz de confusão, precisão, revocação, F1-Score, balanced accuracy, e a importância da validação cruzada estratificada.

4.5.1. Métricas de Avaliação para Dados Desbalanceados

Em um cenário de desbalanceamento de dados, métricas tradicionais como a acurácia podem ser enganosas, pois um modelo que simplesmente prevê a classe majoritária pode obter uma alta acurácia sem ser efetivo em identificar a classe minoritária [L. A. Jeni et al. 2013]. Portanto, é essencial utilizar métricas que ofereçam uma visão mais equilibrada do desempenho do modelo.

4.5.1.1. Curva ROC e AUC

A Curva ROC (*Receiver Operating Characteristic*) é uma ferramenta poderosa para avaliar a capacidade de um modelo de distinguir entre as classes. A curva ROC é um gráfico que plota a taxa de verdadeiros positivos contra a taxa de falsos positivos para diferentes limiares de classificação [Rodrigues 2018]. A principal vantagem da curva ROC é que ela não é afetada pelo desbalanceamento de classes, uma vez que mede a capacidade de discriminação do modelo.

A AUC (*Area Under the ROC Curve*) representa a área sob a curva ROC e varia de 0 a 1 [Rodrigues 2018]. Um AUC de 0,5 indica que o modelo não tem capacidade de discriminação (equivalente a um palpite aleatório), enquanto um AUC próximo de 1 indica excelente discriminação entre as classes. Em cenários de desbalanceamento, a AUC é uma métrica robusta, pois não é influenciada pela distribuição das classes.

4.5.1.2. Matriz de Confusão

A matriz de confusão é uma representação tabular que mostra o desempenho do modelo em termos de verdadeiros positivos (TP), falsos positivos (FP), verdadeiros negativos (TN) e falsos negativos (FN) [De Souza 2019]. A matriz de confusão permite calcular diversas métricas importantes para avaliação em cenários de desbalanceamento. A matriz é detalhada na tabela abaixo:

Tabela 2. Matriz de confusão.

Caso	Definição
Verdadeiros Positivos	Casos em que a classe positiva é corretamente identificada pelo modelo.
Falsos Positivos	Casos em que a classe negativa é incorretamente identificada como positiva.
Verdadeiros Negativos	Casos em que a classe negativa é corretamente identificada pelo modelo
Falsos Negativos	Casos em que a classe positiva é incorretamente identificada como negativa.

A matriz de confusão é a base para calcular métricas como precisão, revocação, e F1-Score, que são essenciais para a avaliação em contextos de desbalanceamento.

4.5.1.3. Precisão, Revocação e F1-Score

Precisão (Precision): é a proporção de verdadeiros positivos sobre todos os exemplos que foram classificados como positivos [De Souza 2019]. A precisão é importante em cenários onde o custo de um falso positivo é alto.

Revocação (Recall): também conhecida como sensibilidade ou taxa de verdadeiros positivos, a revocação mede a proporção de verdadeiros positivos sobre todos os exemplos que realmente são positivos [De Souza 2019]. Revocação é crucial quando o custo de um falso negativo é alto.

F1-Score: o F1-Score é a média harmônica entre a precisão e a revocação [De Souza 2019]. É uma métrica útil quando há necessidade de um equilíbrio entre precisão e revocação, especialmente em casos onde as classes estão desbalanceadas.

Essas métricas oferecem uma visão mais completa e equilibrada do desempenho do modelo em relação às classes minoritárias e majoritárias.

4.5.1.4. Balanced Accuracy

A balanced accuracy é uma métrica que corrige o viés presente na acurácia tradicional em cenários de desbalanceamento. Ela é definida como a média das taxas de verdadeiros positivos para cada classe, proporcionando uma avaliação mais justa do modelo em relação às classes minoritárias [Olugbenga 2023].

A balanced accuracy é especialmente útil quando a distribuição das classes é altamente desbalanceada, garantindo que o modelo seja avaliado com base em sua capacidade de prever corretamente tanto as classes majoritárias quanto as minoritárias.

4.5.2. Validação Cruzada Estratificada

A validação cruzada estratificada é uma técnica que assegura que cada subdivisão (fold) dos dados em um processo de validação cruzada mantém a mesma proporção de classes que o conjunto de dados original [Muralidhar 2021]. Isso é particularmente importante em cenários

de desbalanceamento, pois garante que cada fold usado no treinamento e teste do modelo represente adequadamente a distribuição das classes.

Em um cenário de validação cruzada tradicional, as subdivisões dos dados podem, por acaso, resultar em distribuições desbalanceadas, o que pode levar a uma avaliação enganosa da performance do modelo. A validação cruzada estratificada previne essa situação ao preservar as proporções de classes em cada fold, resultando em uma avaliação mais confiável e representativa.

4.5.3. Importância das Técnicas de Avaliação

As técnicas de avaliação discutidas neste capítulo são fundamentais para medir a eficácia de modelos treinados em conjuntos de dados desbalanceados. O uso de métricas apropriadas, como a curva ROC e AUC, matriz de confusão, precisão, revocação, F1-Score, e balanced accuracy, combinado com métodos robustos como a validação cruzada estratificada, garante uma avaliação mais justa e precisa. Essas práticas permitem que os desenvolvedores e pesquisadores compreendam melhor o desempenho de seus modelos em cenários desafiadores, levando a soluções mais eficazes e generalizáveis para problemas de classificação desbalanceada.

4.6. Estratégias Avançadas

Nesta seção, serão abordadas estratégias avançadas que vão além das técnicas tradicionais de pré-processamento e modelagem, oferecendo soluções mais sofisticadas para lidar com o desbalanceamento de dados. Serão explorados métodos de ensemble, como bagging e boosting, modelos de votação e voto ponderado, além de técnicas de aprendizado por transferência (transfer learning). Também discutiremos outros métodos avançados, como aprendizado semi-supervisionado e aprendizado por reforço, que têm se mostrado promissores em cenários de desbalanceamento.

4.6.1. Ensemble Methods

Os métodos de ensemble combinam múltiplos modelos para melhorar a precisão e a robustez das previsões [Galar M. 2012]. Essas técnicas são particularmente eficazes em cenários de dados desbalanceados, pois diferentes modelos podem capturar diferentes aspectos das classes minoritárias.

O bagging é uma técnica que cria múltiplos modelos treinados em diferentes subconjuntos dos dados, gerados por amostragem com reposição (bootstrap) [Galar M. 2012]. Um exemplo clássico de bagging é o Random Forest. No contexto de dados desbalanceados, o bagging pode ser ajustado para amostrar a classe minoritária com mais frequência, aumentando sua representação nos modelos individuais.

O boosting é uma técnica de ensemble que cria modelos sequencialmente, onde cada novo modelo tenta corrigir os erros dos modelos anteriores [Galar M. 2012]. XGBoost e AdaBoost são exemplos de algoritmos de boosting. Para lidar com o desbalanceamento, o boosting pode ajustar os pesos das observações com base em seu erro de classificação, dando mais importância às instâncias da classe minoritária.

4.6.1.2. Modelos de Votação

Os modelos de votação combinam as previsões de vários modelos para tomar uma decisão final [A. Dogan and D. Birant 2019]. Em cenários de desbalanceamento, é comum utilizar *weighted voting* (votação ponderada), onde as previsões de modelos que se saem melhor na detecção da classe minoritária recebem maior peso na decisão final.

No modelo de votação mais simples, chamado de *majority voting*, a classe predita pela maioria dos modelos é selecionada [A. Dogan and D. Birant 2019]. Essa abordagem pode ser insuficiente em cenários desbalanceados, pois a classe majoritária tende a dominar.

No *weighted voting*, cada modelo recebe um peso proporcional à sua performance em detectar a classe minoritária [A. Dogan and D. Birant 2019]. Por exemplo, se um modelo tem alta precisão para a classe minoritária, suas previsões têm maior influência no voto final.

4.6.2. Transfer Learning

O *transfer learning* é uma técnica onde um modelo pré-treinado em um grande conjunto de dados (geralmente equilibrado) é ajustado para uma nova tarefa com dados potencialmente desbalanceados [Al-Stouhi S. and Reddy C. 2015]. Essa técnica tem se mostrado eficaz em cenários onde há pouca disponibilidade de dados para a classe minoritária.

4.6.2.1. Aplicação em Cenários de Dados Desbalanceados

Fine-tuning de modelos pré-treinados: em *transfer learning*, um modelo, como uma rede neural profunda pré-treinada em um grande conjunto de dados, pode ser ajustado (fine-tuned) em um novo conjunto de dados desbalanceado. Esse processo permite que o modelo retenha o conhecimento geral adquirido no treinamento inicial e o aplique para melhorar o desempenho em classes minoritárias.

Feature extraction: outra aplicação de *transfer learning* é a extração de características (features) de modelos pré-treinados. As características extraídas podem ser usadas como entrada para outro modelo, como uma SVM ou um random forest, treinado especificamente para a tarefa com dados desbalanceados.

4.6.3. Outros Métodos Avançados

Além das técnicas mencionadas previamente, outros métodos avançados têm sido desenvolvidos para lidar com o desbalanceamento de dados, aproveitando técnicas mais recentes de aprendizado de máquina.

4.6.3.1. Aprendizado Semi-Supervisionado

O aprendizado semi-supervisionado utiliza uma combinação de dados rotulados e não rotulados para treinar modelos [Gui Q. et al. 2024]. Em cenários de desbalanceamento, onde a classe minoritária pode ser sub-representada nos dados rotulados, o aprendizado semi-supervisionado pode ajudar a expandir o conjunto de treinamento.

Label Propagation: técnicas de propagação de rótulos podem ser usadas para inferir rótulos para o conjunto de dados não rotulado, potencialmente aumentando a representação da classe minoritária [Gui Q. et al. 2024].

Pseudo-Labeling: em *pseudo-labeling*, o modelo inicial faz previsões sobre os dados não rotulados, e essas previsões são então tratadas como rótulos para o próximo ciclo de treinamento, ajudando a reforçar a detecção da classe minoritária [Gui Q. et al. 2024].

4.6.3.2. Aprendizado por Reforço

O aprendizado por reforço envolve treinar um agente para tomar decisões em um ambiente, maximizando uma recompensa ao longo do tempo [Lin E. et al. 2020]. Essa abordagem pode ser adaptada para lidar com desbalanceamento de dados, especialmente em contextos onde as decisões precisam ser adaptativas e dinâmicas.

Reward Shaping: no contexto de dados desbalanceados, a função de recompensa pode ser ajustada para penalizar fortemente os erros na detecção da classe minoritária, incentivando o agente a focar mais nessa classe [Lin E. et al. 2020].

Q-Learning e Deep Q-Networks (DQN): técnicas de aprendizado por reforço, como Q-Learning e DQN, podem ser aplicadas em cenários onde as instâncias da classe minoritária são raras, treinando o agente para reconhecer e priorizar a classe minoritária em suas decisões [Lin E. et al. 2020].

4.7. Desafios e Considerações Finais

As técnicas discutidas ao longo deste material oferecem o conhecimento necessário para lidar com o desbalanceamento de dados. No entanto, essas técnicas vêm com suas próprias limitações e desafios, que devem ser cuidadosamente consideradas ao selecionar e implementar uma estratégia. Nesta última seção serão discutidas as limitações das técnicas existentes, os aspectos a serem considerados ao escolher uma abordagem, e as tendências futuras que podem moldar o campo nos próximos anos.

4.7.1. Limitações das Técnicas

Apesar dos avanços significativos na área de desbalanceamento de dados, as técnicas atuais não são universais e apresentam limitações que podem impactar a eficácia em diferentes cenários.

4.7.1.1. Reamostragem

Técnicas de reamostragem, como oversampling e undersampling, são simples e eficazes, mas podem introduzir novos desafios. O oversampling pode levar a overfitting, especialmente se forem gerados muitos exemplos sintéticos da classe minoritária. Já o undersampling pode resultar na perda de informações valiosas, comprometendo o desempenho geral do modelo.

4.7.1.2. Algoritmos de Modelagem

Embora algoritmos robustecidos, como os métodos baseados em árvores (XGBoost, Random Forest), sejam eficazes em muitos casos, eles podem se tornar complexos e difíceis de

interpretar. Além disso, o ajuste de hiperparâmetros para lidar com desbalanceamento pode ser um processo longo e não garantir uma solução ideal. Modelos de redes neurais, por sua vez, podem ser altamente sensíveis à configuração dos pesos e exigem uma quantidade significativa de dados para evitar underfitting.

4.7.1.3. Métodos Avançados

Técnicas avançadas, como aprendizado semi-supervisionado e por reforço, requerem grandes volumes de dados e processamento, tornando-se impraticáveis em cenários onde os recursos são limitados. Além disso, essas abordagens podem ser difíceis de implementar e interpretar, especialmente para quem não tem familiaridade com os conceitos subjacentes.

4.7.2. Aspectos a Serem Considerados ao Escolher uma Técnica

A escolha da técnica apropriada para lidar com o desbalanceamento de dados depende de uma série de fatores, que devem ser ponderados com cuidado.

4.7.2.1. Natureza dos Dados

A natureza dos dados, incluindo o grau de desbalanceamento, o número de características e a complexidade das relações entre as variáveis, é um fator crítico na escolha da técnica. Para conjuntos de dados com alto grau de desbalanceamento, técnicas que ajustam os pesos ou que combinam oversampling e undersampling podem ser mais eficazes.

4.7.2.2. Objetivo do Modelo

O objetivo do modelo, seja ele focado em maximizar precisão, revocação, ou uma combinação das duas, também orienta a escolha da técnica. Modelos que exigem alta precisão podem se beneficiar de técnicas de reamostragem ou ajuste de pesos, enquanto aqueles que priorizam a revocação podem requerer algoritmos mais complexos ou métodos de ensemble.

4.7.2.3. Recursos Disponíveis

Os recursos computacionais e o tempo disponível para o desenvolvimento do modelo são considerações importantes. Técnicas mais simples, como reamostragem ou ajuste de pesos, podem ser implementadas rapidamente, mas podem não oferecer a melhor performance. Métodos avançados, como deep learning ou aprendizado por reforço, requerem mais tempo e poder computacional, mas podem oferecer ganhos significativos em precisão e generalização.

4.7.3. Tendências Futuras

O campo de aprendizado de máquina continua a evoluir rapidamente, e novas abordagens para lidar com o desbalanceamento de dados estão emergindo. Nesta seção, discutimos as tendências e tecnologias promissoras que podem definir como tratamos dados desbalanceados futuramente.

4.7.3.1. Aprendizado Federado

O aprendizado federado é uma abordagem emergente que permite que modelos sejam treinados em dados distribuídos entre várias fontes, preservando a privacidade dos dados [C. Nguyen et al. 2021]. Essa técnica tem o potencial de melhorar o desempenho em cenários de desbalanceamento ao permitir que dados minoritários de várias fontes sejam combinados para criar um modelo mais robusto, sem a necessidade de centralizar os dados.

4.7.3.2. Modelos Auto-supervisionados

Modelos auto-supervisionados, que podem aprender representações úteis dos dados sem a necessidade de rótulos extensivos, estão ganhando popularidade [Bergmann 2023]. Essas técnicas podem ajudar a superar o problema do desbalanceamento, permitindo que o modelo aprenda com grandes volumes de dados não rotulados e aplicando esse conhecimento para melhorar a detecção da classe minoritária.

4.8. Referências

- [Abdi and Hashemi 2015] Abdi, L., Hashemi, S.: To combat multi-class imbalanced problems by means of over-sampling techniques. *IEEE transactions on Knowledge and Data Engineering* 28(1), 238–251 (2015).
- [Coelho et al. 2022] Coelho, G.M.C., Ramos, A.C., de Sousa, J., Cavaliere, M., de Lima, M.J., Mangeth, A., Frajhof, I.Z., Cury, C., Casanova, M.A.: Text classification in the brazilian legal domain. In: *Intern. Conference on Enterprise Information Systems*. pp. 355–363 (2022).
- [Feng et al. 2019] Feng, W., Dauphin, G., Huang, W., Quan, Y., Bao, W., Wu, M., Li, Q.: Dynamic synthetic minority over-sampling technique-based rotation forest for the classification of imbalanced hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12(7), 2159–2169 (2019).
- [Chaudhary 2023] Chaudhary, K.: How to Deal with Imbalanced Data in Classification. Medium. Disponível em: <https://medium.com/game-of-bits/how-to-deal-with-imbalanced-data-in-classification-bd03cfc66066>. Acesso em: 23 ago. 2024.
- [Azank 2020] Azank F.: Dados Desbalanceados — O que são e como lidar com eles. Disponível em: <https://medium.com/turing-talks/dados-desbalanceados-o-que-s%C3%A3o-e-como-evit%C3%A1-los-43df4f49732b>. Acesso em: 23 ago. 2024.
- [Hasib et al. 2020] Hasib K., Iqbal Md., Shah F., Mahmud J., Popel M., Showrov Md. I. H., Ahmed S., Rahman O. (2020). A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem. *Journal of Computer Science*. 16. 1546 - 1557. 10.3844/jcssp.2020.1546.1557.
- [A. K. Ilavarasi 2020] A. K. Ilavarasi (2020). Class imbalance learning for Identity Management in Healthcare. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, 2020, pp. 995-1000, doi: 10.1109/I-SMAC49090.2020.9243420.
- [Rubaidi Z. 2022] Rubaidi Zainab, Ben Ammar Boulbaba, Ben Aouicha (2022). Fraud Detection Using Large-scale Imbalance Dataset. In: *International Journal on Artificial Intelligence Tools*. 31. 10.1142/S0218213022500373.
- [L. A. Jeni et al. 2013] L. A. Jeni, J. F. Cohn, F. De La Torre (2013). Facing Imbalanced Data--Recommendations for the Use of Performance Metrics. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, Geneva, Switzerland, 2013, pp. 245-251, doi: 10.1109/ACII.2013.47.
- [Thabtah F. 2019] Thabtah F., Hammoud S., Kamalov F., Gonsalvesv A. (2019). Data Imbalance in Classification: Experimental Evaluation. In: *Information Sciences*. 513. 10.1016/j.ins.2019.11.004.
- [Longadge R. and Dongre S. 2013] Longadge Rushi, Dongre Snehlata. (2013). Class Imbalance Problem in Data Mining Review. In: *Int. J. Comput. Sci. Netw.*. 2.

- [H. Cao et al. 2013] H. Cao, X. -L. Li, D. Y. -K. Woon and S. -K. Ng (2013). Integrated Oversampling for Imbalanced Time Series Classification. In: IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 12, pp. 2809-2822, Dec. 2013, doi: 10.1109/TKDE.2013.37.
- [Mathew R. and Gunasundari R. 2021] R. Mary Mathew, R. Gunasundari (2021). A Review on Handling Multiclass Imbalanced Data Classification In Education Domain. In: 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021, pp. 752-755, doi: 10.1109/ICACITE51222.2021.9404626.
- [S. Pouyanfar et al. 2019] S. Pouyanfar, T. Wang and S. -C. Chen (2019). A Multi-label Multimodal Deep Learning Framework for Imbalanced Data Classification. In: 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 2019, pp. 199-204, doi: 10.1109/MIPR.2019.00043.
- [Junsomboon N. and Phienthrakul T. 2017] Junsomboon Nutthaporn and Phienthrakul Tanasanee (2017). Combining Over-Sampling and Under-Sampling Techniques for Imbalance Dataset. In: Proceedings of the 9th International Conference on Machine Learning and Computing, 2017, pp. 243-247, doi: 10.1145/3055635.3056643.
- [H. Shamsudin et al. 2020] H. Shamsudin, U. K. Yusof, A. Jayalakshmi and M. N. Akmal Khalid (2020). Combining oversampling and undersampling techniques for imbalanced classification: A comparative study using credit card fraudulent transaction dataset. In: 2020 IEEE 16th International Conference on Control & Automation (ICCA), Singapore, 2020, pp. 803-808, doi: 10.1109.
- [Chawla 2002] Chawla Nitesh, Bowyer Kevin, Hall Lawrence, Kegelmeyer W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. In: J. Artif. Intell. Res. (JAIR). 16. 321-357. 10.1613/jair.953.
- [He 2008] Haibo He, Yang Bai, E. A. Garcia, Shutao Li (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.
- [Mqadi 2021] Mqadi Nhlakanipho, Naicker Nalindren, Adeliyi Timothy (2021). A SMOTE based Oversampling Data-Point Approach to Solving the Credit Card Data Imbalance Problem in Financial Fraud Detection. In: International Journal of Computing and Digital Systems. 10. 277 - 286. 10.12785/ijcds/100128.
- [Kosolwattana T. et al. 2022] Kosolwattana Tanapol, Liu Chenang, Hu Renjie, Han Shizhong, Chen Hua, Lin Ying (2022). A Self-inspected Adaptive SMOTE Algorithm (SASMOTE) for Highly Imbalanced Data Classification in Healthcare. 10.21203/rs.3.rs-1647776/v1.
- [J. -R. Jiang and Y. -T. Chen 2022] J. -R. Jiang and Y. -T. Chen (2022). Industrial Control System Anomaly Detection and Classification Based on Network Traffic. In IEEE Access, vol. 10, pp. 41874-41888, 2022, doi: 10.1109/
- [Zhang P. 2022] Zhang Ping, Jia Yiqiao, Shang Youlin (2022). Research and application of XGBoost in imbalanced data. In: International Journal of Distributed Sensor Networks. 18. 155013292211069. 10.1177/15501329221106935.
- [Pristyanto Y. et al. 2023] Pristyanto Yoga, Mukarabiman Zulfikar, Nugraha Anggit (2023). Extreme Gradient Boosting Algorithm to Improve Machine Learning Model Performance on Multiclass Imbalanced Dataset. In: JOIV: International Journal on Informatics Visualization. 7. 710-715. 10.30630/joiv.7.3.1102.
- [M. Bader-El-Den et al. 2019] M. Bader-El-Den, E. Teitei and T. Perry (2019). Biased Random Forest For Dealing With the Class Imbalance Problem. In: IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 7, pp. 2163-2172, July 2019, doi: 10.1109/TNNLS.2018.2878400.
- [J. Justin and K. Taghi 2019] Johnson Justin and Khoshgoftaar Taghi (2019). Survey on deep learning with class imbalance. In: Journal of Big Data. 6. 27. 10.1186/s40537-019-0192-5.

- [Z. Fan et al. 2022] Zhang Fan, Petersen Melissa, Johnson Leigh, Hall James and O'Bryant Sid (2022). Hyperparameter Tuning with High Performance Computing Machine Learning for Imbalanced Alzheimer's Disease Data. In: Applied Sciences. 12. 6670. 10.3390/app12136670.
- [M. R. Rezaei-Dastjerdehei et al. 2020] M. R. Rezaei-Dastjerdehei, A. Mijani and E. Fatemizadeh (2020). Addressing Imbalance in Multi-Label Classification Using Weighted Cross Entropy Loss Function. In: 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 2020, pp. 333-338, doi: 10.1109/ICBME51989.2020.9319440.
- [Rodrigues 2018] Rodrigues V.: Entenda o que é AUC e ROC nos modelos de Machine Learning. Disponível em: <https://medium.com/bio-data-blog/entenda-o-que-%C3%A9-auc-e-roc-nos-modelos-de-machine-learning-8191fb4df772>. Acesso em: 26 ago 2024.
- [De Souza 2019] De Souza Emanuel (2019). Entendendo o que é Matriz de Confusão com Python. Disponível em: <https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509>. Acesso em 26 ago 2024.
- [Olugbenga 2023] Olugbenga Motunrayo (2023). Balanced Accuracy: When Should You Use It? Disponível em: <https://neptune.ai/blog/balanced-accuracy>. Acesso em 26 ago. 2024.
- [Muralidhar 2021] Muralidhar KSV. What is Stratified Cross-Validation in Machine Learning? Disponível em: <https://towardsdatascience.com/what-is-stratified-cross-validation-in-machine-learning-8844f3e7ae8e>. Acesso em: 26 ago. 2024.
- [Galar M. 2012] Galar Mikel, Fernández Alberto, Barrenechea Edurne, Sola Humberto and Herrera, Francisco (2012). A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. In: Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on. 42. 463 - 484. 10.1109/TSMCC.2011.2161285.
- [A. Dogan and D. Birant 2019] A. Dogan and D. Birant (2019). A Weighted Majority Voting Ensemble Approach for Classification. In: 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, 2019, pp. 1-6, doi: 10.1109/UBMK.2019.8907028.
- [Al-Stouhi S. and Reddy C. 2015] Al-Stouhi Samir and Reddy Chandan (2015). Transfer Learning for Class Imbalance Problems with Inadequate Data. Knowledge and Information Systems. 48. 10.1007/s10115-015-0870-3.
- [Gui Q. et al. 2024] Gui Q., Zhou H., Guo N. *et al.* (2024). A survey of class-imbalanced semi-supervised learning. In: Mach Learn **113**, 5057–5086 (2024). <https://doi.org/10.1007/s10994-023-06344-7>.
- [Lin E. et al. 2020] Lin E., Chen Q. and Qi X (2020). Deep reinforcement learning for imbalanced classification. In: Appl Intell **50**, 2488–2502 (2020). <https://doi.org/10.1007/s10489-020-01637-z>.
- [C. Nguyen et al. 2021] C. Nguyen Dinh, Ding Ming, Pathirana Pubudu, Seneviratne Aruna, Li Jun and Poor H. Vincent. (2021). In: Federated Learning for Internet of Things: A Comprehensive Survey.
- [Bergmann 2023] Bergmann Dave (2023). What is self-supervised learning? Disponível em: <https://www.ibm.com/topics/self-supervised-learning>. Acesso em 26 ago. 2024.

Capítulo

5

Design Web Intuitivo: Potencializando Projetos com Material Design e BeerCSS

Matusalen Costa Alves, Maria Steffany da Silva Viana, Iallen Gábio de Sousa Santos

Abstract

This chapter delves into the development and significance of Human-Computer Interaction (HCI), detailing the evolution of graphical interfaces and the impact of Material Design on creating effective and visually appealing user interfaces. It introduces BeerCSS, a lightweight framework that integrates Material Design principles to streamline interface development. The chapter provides guidance on installing BeerCSS via CDN and NPM, and implementing light and dark themes with JavaScript. Additionally, it illustrates the practical application of these technologies through the creation of a developer portfolio, demonstrating how to build compelling and functional user experiences.

Resumo

Este capítulo explora o desenvolvimento e a importância da Interface Humano-Computador (IHC), abordando a evolução das interfaces gráficas e a aplicação dos princípios do Material Design na criação de interfaces estéticas e funcionais. O estudo destaca o framework BeerCSS, que oferece uma solução leve e eficiente para o desenvolvimento rápido de interfaces ao incorporar os princípios do Material Design. Inclui instruções para a instalação do BeerCSS via CDN e NPM e para a implementação de temas claros e escuros com JavaScript. Além disso, demonstra a aplicação prática dessas tecnologias por meio da criação de um portfólio de desenvolvedor, evidenciando a capacidade de criar experiências de usuário impactantes e funcionais.

5.1. Introdução

Em 1990, o cientista cognitivo Don Norman se juntou à Apple, empresa de tecnologia, e desenvolveu o que é conhecido hoje como User Experience (UX) [Norman 2002]. Na era digital atual, a UX tornou-se um dos aspectos prioritários e críticos no design de

interfaces. A forma como os usuários percebem e interagem com um site ou aplicativo pode determinar não apenas sua satisfação, mas também a eficácia da comunicação e o sucesso geral do produto.

Na maior parte dos casos, os usuários geralmente decidem consumir ou não um produto no prazo de 15 a 20 segundos após o primeiro acesso [Nielsen et al. 2014]. Nesse cenário, as percepções visuais e sensoriais desempenham um papel fundamental, moldando profundamente a forma como os indivíduos respondem a diferentes elementos de design.

As cores, tipografia, disposição de elementos e harmonia visual são essenciais para capturar a atenção do usuário e criar uma experiência mais agradável [Norman 2002]. As cores evocam emoções e estabelecem uma identidade visual; a tipografia contribui para a legibilidade e percepção de credibilidade; a disposição dos elementos afeta a facilidade de navegação e a aparência geral. Esses aspectos são mais do que simples escolhas estéticas, eles impactam diretamente na eficiência da interação do usuário com o produto.

O design tem um impacto significativo na forma de interação e a comunicação visual [Norman 2002]. Além das telas de dispositivos digitais, isto também é percebido em mídias impressas e ambientes físicos, como outdoors e exposições. A maneira como os elementos são dispostos e apresentados pode afetar a forma como as pessoas percebem e entendem a informação.

Com o mercado cada vez mais competitivo, onde a oferta de produtos e serviços é vasta e diversificada, se destacar é um desafio. A capacidade de prender a atenção do usuário e criar uma impressão memorável é essencial para o sucesso. Uma apresentação visual impactante, combinada com uma funcionalidade intuitiva, pode fazer a diferença entre conquistar a lealdade do cliente ou ser esquecido [Eyal and Hoover 2014].

Em ambientes digitais, oferecer interfaces que sejam visualmente atraentes e fáceis de usar é imprescindível, visto que o usuário realiza um julgamento rápido, enquanto em contextos físicos, como publicidade e embalagens, um design inovador pode influenciar a decisão de compra [Krug 2000]. A capacidade de criar experiências que se destacam pelo apelo visual é, portanto, uma vantagem competitiva significativa no mercado atual.

Além disso, o design deve considerar não apenas a aparência, mas também a funcionalidade [Kao and Archer 1997]. A integração harmoniosa desses elementos cria uma experiência aprimorada, que aumenta a satisfação do usuário e a eficácia do produto.

Sob esta óptica, destaca-se a Interação Humano-Computador (IHC), um campo que estuda a interação entre usuários e sistemas computacionais, focando em como os humanos interagem com as tecnologias e como essas interações podem ser aprimoradas [Waldeilson and Paixão 2016]. A IHC abrange aspectos como a usabilidade, acessibilidade e a UX, com o objetivo de criar interfaces mais eficazes. Este campo se preocupa em entender as necessidades dos usuários, suas capacidades cognitivas e comportamentais e como o design pode facilitar a realização de tarefas de maneira satisfatória [Franz and Kauer-Franz 2024]. A aplicação dos princípios de IHC no design de interfaces assegura que as soluções tecnológicas sejam funcionalmente robustas e ofereçam uma experiência produtiva para os usuários.

Neste contexto, o Material Design [Google 2014], criado pela Google [Inc. 2024], surge como uma abordagem inovadora que vai além das interfaces digitais, inspirando-se na física dos materiais e na forma como eles interagem com a luz e o espaço. Lançado em 2014, o Material Design foi desenvolvido para proporcionar uma experiência visual coesa e funcional que se aplica a diversos contextos e mídias. A ideia central é que, assim como os materiais físicos têm propriedades específicas, o design digital pode ser orientado por princípios que priorizam a hierarquia visual e clareza.

O planejamento dessa ideologia de design foi cuidadosamente pensado para atender às necessidades da Interação Humano-Computador e proporcionar uma abordagem estruturada ao design de interfaces. Seus princípios incluem o uso de camadas e sombras para criar profundidade, tipografia legível e uma paleta de cores bem definida para promover a comunicação visual. Esses elementos vão além de escolhas estéticas, são projetados visando garantir uma navegação otimizada para uma experiência de usuário satisfatória. Isto traz maior conforto para o usuário, o que por sua vez, o torna suscetível a consumir mais daquele conteúdo [Dulsrud and Bygstad 2022].

Outro aspecto do Material Design é o foco na acessibilidade. As diretrizes possibilitam a criação de interfaces que possam ser usadas por todos, independentemente de suas capacidades físicas ou cognitivas. Isso se reflete na escolha cuidadosa de cores, na adaptação de tamanhos de texto e no suporte a tecnologias assistivas, garantindo que a navegação seja intuitiva e que a informação esteja sempre ao alcance do usuário.

A consistência na navegação também participa dos padrões. O Material Design estabelece padrões claros para a organização de elementos, como barras de ferramentas, menus e botões. Esses componentes são projetados para oferecer uma experiência uniforme em todas as plataformas, ajudando os usuários a se orientarem e realizarem tarefas sem confusão ou frustração, mesmo quando mudam de dispositivo.

O Material Design adota uma abordagem responsiva, garantindo que as interfaces se adaptem automaticamente a diferentes tamanhos de tela e dispositivos. Isso significa que, seja em um smartphone, tablet ou desktop, o usuário terá uma experiência igualmente consistente. A responsividade não só melhora a estética do design, mas também assegura que a usabilidade seja mantida, independentemente de como ou onde o sistema é acessado [Marcotte 2011].

Neste minicurso, será abordado o Material Design 3.0 (M3), esta é a versão mais atual do sistema de design da Google. O M3 traz inovações para o aprimoramento da UX, além disso, enfatiza a personalização e várias novas adições significativas. Lançado como uma evolução do Material Design original, o 3.0 enfatiza a personalização e a adaptabilidade, permitindo que designers criem interfaces ajustadas às necessidades específicas de seus usuários [Google 2022].

A estilização de páginas da web é geralmente feita através de CSS. Os desenvolvedores podem descrever aspectos estéticos de cada parte da página utilizando esta tecnologia. Entretanto, existem ferramentas de CSS pré-implementado que facilitam o trabalho ao estilos prontos. Essas ferramentas permitem que os desenvolvedores apliquem padrões visuais consistentes de maneira mais rápida e eficiente, sem a necessidade de criar todo o código CSS manualmente, acelerando o processo de desenvolvimento e

garantindo um design mais uniforme. Estas são denominadas frameworks de design.

Para criar aplicativos e páginas na web, os desenvolvedores utilizam diversos frameworks baseados em diferentes tecnologias de design. Exemplos dessas incluem o Human Interface Guidelines da Apple [Apple 2024] e o Fluent Design System da Microsoft [Microsoft 2024]. Cada um desses frameworks oferece um conjunto único de diretrizes e componentes para criar interfaces correspondentes aos critérios de design das marcas.

FRAMEWORK	FOCO	CUSTO	ESTADO DE ATUALIZAÇÃO
Bootstrap for MDB	Material Design for Bootstrap	Pago	Constante
Bootstrap	Bootstrap	Gratuito	Constante
BeerCSS	Material Design 3.0	Gratuito	Constante
Tailwind CSS	Human Interface Guidelines	Pago	Constante
Office UI	Fluent Design Style	Gratuito	Descontinuado
Fluent UI	Fluent Design Style	Gratuito	Constante
Materialize	Material Design 2.0	Gratuito	Em hiato

Tabela 5.1. Comparação entre diferentes frameworks de CSS

A lista dessas ferramentas é extensa e algumas delas são de acesso gratuito e outras pago, como apresentado na Tabela 5.1. Neste minicurso, concentramos nossa atenção no BeerCSS [BeerCSS 2022], devido a sua gratuidade, design atual e integração com o ecossistema da Google, a fim de explorar como essa ferramenta pode nos ajudar a concluir o nosso objetivo de criar de maneira ágil páginas responsivas e visualmente agradáveis.

O restante deste capítulo está organizado da seguinte maneira: a Seção 5.2 explora a história do IHC, UX e interface de usuário (IU), a Seção 5.3 apresenta os princípios fundamentais do Material Design, enquanto a Seção 5.4 o BeerCSS e suas boas práticas, a Seção 5.5 o estudo de caso com um portfólio de desenvolvedor e, por fim, a Seção 5.6 a conclusão deste trabalho.

5.2. Interface Humano-Computador, IU e UX

A Interface Humano-Computador (IHC) surge com campo fundamental da computação devido a necessidade de aprimoramento da interação dos seres humanos com os sistemas computacionais. A IHC é a base sobre a qual se desenvolvem conceitos como interface do usuário (IU) e experiência do usuário (UX) [Franz and Kauer-Franz 2024]. No design de interfaces, IU e UX são termos frequentemente usados para descrever aspectos específicos da interação, mas cada um tem seu próprio foco e importância.

Enquanto a IU se concentra na criação de interfaces visuais eficientes e intuitivas [Krug 2000], a UX abrange uma visão mais ampla, englobando as emoções, percepções e respostas do usuário ao interagir com o produto [Garrett 2010]. Esta seção introduz os principais conceitos de IHC, IU e UX, explorando como esses elementos se inter-relacionam para criar experiências de usuário mais eficazes e satisfatórias.

5.2.1. História e Evolução da Interface Humano-Computador

A Interface Humano-Computador (IHC) começou a se formar como um campo de estudo na década de 1960, quando os primeiros computadores começaram a se tornar mais acessíveis para empresas e instituições de pesquisa. Naquela época, a interação com os

computadores era restrita a comandos de texto complexos, exigindo conhecimentos técnicos aprofundados. A evolução da IHC foi impulsionada pela necessidade de tornar a tecnologia mais acessível para um público mais amplo e a popularização dos computadores pessoais [Raskin 2000].

Nos anos 1970 e 1980, os avanços em hardware e software, como o desenvolvimento das primeiras interfaces gráficas de usuário (GUIs), revolucionaram a forma como os humanos interagem com os computadores. A introdução do mouse, janelas, ícones e menus, liderada por empresas como Xerox e Apple, marcou um ponto de virada ao permitir que os usuários interagissem com computadores de forma mais natural.

Na década de 1990, o surgimento da web trouxe novos desafios e oportunidades para a IHC. A criação de interfaces eficazes para diferentes dispositivos e contextos tornou-se crucial. Com a expansão da internet, surgiram novas plataformas e dispositivos com características e limitações próprias [Raskin 2000]. Isso exigiu que os designers encontrassem soluções adaptáveis para garantir que as interfaces funcionassem bem em diversos cenários. A web não só ampliou o alcance da IHC, mas também trouxe novos elementos e interações que impactaram profundamente o design de interfaces.

Atualmente, a IHC está em constante evolução, incorporando tecnologias inovadoras que transformam a interação com a tecnologia. Realidade virtual aumentada, inteligência artificial e interfaces naturais, como reconhecimento de voz e gestos, moldam o futuro da interação [Zaphiris 2008]. Esses avanços buscam criar experiências eficientes para permitir uma comunicação mais natural entre humanos e sistemas. À medida que essas tecnologias se tornam parte do cotidiano, elas redefinem a experiência do usuário e ampliam as possibilidades para o design de interfaces.

5.2.2. Princípios Fundamentais da IHC

Os princípios fundamentais da IHC guiam o design de interfaces que promovem uma interação fluida entre o usuário e a tecnologia [Zaphiris 2008]. Esses princípios são desenvolvidos com base no entendimento das capacidades cognitivas e comportamentais dos usuários, com o objetivo de criar sistemas que não apenas atendam às suas necessidades, mas também ofereçam uma experiência de fácil compreensão. Entre esses princípios, destacam-se a simplicidade, consistência, visibilidade e feedback. Estes aspectos visam garantir que as interfaces sejam acessíveis e fáceis de navegar, independentemente do nível de habilidade do usuário.

A usabilidade é um dos pilares centrais da IHC, ela consiste na criação de interfaces que sejam eficazes, eficientes e satisfatórias [Krug 2000]. Um sistema com alta usabilidade permite que os usuários realizem suas tarefas com facilidade, sem a necessidade de treinamento extenso ou frustração. Para alcançar isso, o design deve ser intuitivo, permitindo que o usuário aprenda rapidamente a utilizar o sistema e se sinta confortável ao interagir com ele. A simplicidade na apresentação das informações e a redução de complexidade são estratégias fundamentais para melhorar a usabilidade.

O reaproveitamento de componentes de interfaces é uma prática crucial para garantir a consistência nos sistemas. Um exemplo clássico é o botão de fechar representado na Figura 5.1, que geralmente é vermelho. Essa escolha não é acidental; culturalmente,

o vermelho é associado a ações de interrupção ou alerta, como parar, cancelar ou fechar um processo [Garrett 2010]. Ao utilizar consistentemente um botão vermelho para interromper uma ação, os usuários rapidamente associam essa cor com a função, reduzindo o tempo necessário para entender e operar o sistema.

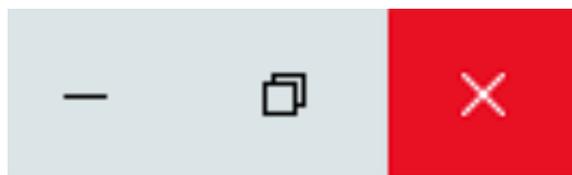


Figura 5.1. Botão Fechar do Windows [Microsoft 2024] em destaque

Essa reutilização de componentes familiares, como ícones, cores e padrões de interação, contribui para uma curva de aprendizado mais suave, já que os usuários podem transferir conhecimentos de uma parte do sistema para outra. Além disso, a padronização desses elementos ajuda a minimizar erros, pois os usuários sabem o que esperar ao interagir com diferentes partes da interface, o que aumenta a satisfação geral com o sistema.

A acessibilidade é outro princípio importante para a IHC, garantindo que os sistemas sejam utilizáveis por todos, independentemente de suas limitações físicas ou cognitivas. Isso inclui a adaptação de interfaces para pessoas com deficiências, como a utilização de textos alternativos para imagens, a implementação de teclas de atalho, e a compatibilidade com tecnologias assistivas. Um design acessível cumpre normas de inclusão e também amplia o alcance do produto, permitindo que mais pessoas possam interagir com ele de maneira eficaz [Johnson 2014].

O feedback ao usuário se concentra em fornecer respostas claras e imediatas às ações do usuário. Isso pode ser feito por meio de mensagens de confirmação, indicações visuais ou sonoras e outros sinais que informam o usuário sobre o estado atual do sistema ou o sucesso de suas ações. O feedback adequado ajuda a evitar confusões, assim, permite que o usuário compreenda facilmente o impacto de suas interações e continue o uso de maneira mais confiável.

5.2.3. Interface do Usuário (IU)

A IU é um dos componentes mais críticos de qualquer sistema, pois é o ponto de interação direta entre o usuário e a tecnologia [Norman 2002]. O design visual da IU é fundamental para gerar uma experiência adequada. Elementos como cores, tipografia, e layout desempenham um papel crucial na comunicação eficaz das funções e na estética geral da interface. Um design visual bem executado melhora a percepção de profissionalismo e facilita a compreensão e o uso do sistema. Por outro lado, um design visual fraco pode proporcionar frustrações e falhas.

A navegação é um aspecto que impacta diretamente a facilidade com que os usuários podem encontrar e acessar diferentes partes do sistema. Uma navegação lógica é essencial para que os usuários possam realizar suas tarefas sem fracasso. Isso inclui a organização clara de menus, botões e links, bem como a implementação de convenções de design reconhecíveis que ajudam os usuários a se orientarem rapidamente.

A consistência é um princípio chave no design da UI herdado diretamente da IHC. Ela visa garantir que todos os elementos da interface mantenham um padrão uniforme. Essa consistência ajuda a criar uma experiência previsível, assim, pode reduzir a carga cognitiva do usuário e facilitar a aprendizagem do uso daquele sistema. Manter uma estética e um comportamento consistentes em diferentes seções e plataformas do sistema contribui com feedbacks indispensáveis como uma navegação fluida e uma interação eficiente.

5.2.4. Experiência do Usuário (UX)

A UX abrange todos os aspectos da interação do usuário com um produto ou sistema, indo além da simples interface visual [Norman 2002]. UX é um campo multidimensional que engloba fatores como satisfação, eficiência e a emoção experimentada durante o uso de produtos digitais. A satisfação do usuário é fundamental, pois está diretamente relacionada à percepção de valor e ao prazer obtido ao usar o sistema. Um design que proporciona uma experiência satisfatória tende a aumentar a lealdade e a recomendação do usuário.

A eficiência é outro componente crucial da UX. Sistemas bem projetados permitem que os usuários realizem suas tarefas com o mínimo de esforço e tempo possível. Isso envolve a otimização de fluxos de trabalho, a redução de etapas desnecessárias e a garantia de que as funções mais usadas sejam facilmente acessíveis.

Além disso, a emoção desempenha um papel significativo na experiência do usuário. As emoções que um usuário experimenta ao interagir com um sistema podem influenciar a forma como ele percebe e se relaciona com o produto. Um design que evoca emoções positivas como prazer e satisfação, pode criar uma conexão mais profunda e duradoura com o usuário. A integração de elementos que proporcionam prazer estético e interações agradáveis é essencial para uma experiência do usuário bem-sucedida.

5.3. Material Design

5.3.1. Conceito e Fundamentos do Material Design

O Material Design, fornece um conjunto de diretrizes que auxilia o usuário a criar interfaces que sejam esteticamente agradáveis, funcionais, interativas e de fácil compreensão. Ele também traz uma inspiração no mundo físico, demonstrando texturas, elevações e diferenças tonais para demonstrar uma certa hierarquia dentro dos elementos.

5.3.2. Princípios de Design: Cores, Tipografia e Formas

Chamadas de tríade do design, cor, tipografia e forma são pilares fundamentais do design. A cor é a linguagem emotiva onde traz significados psicológicos, culturais, harmonia e contraste. A tipografia deve trazer hierarquia, legibilidade e assim ser a voz do design. A forma é a estrutura visual, onde se encaixam as formas geométricas e orgânicas, além dos elementos básicos como ponto, linha, plano e volume. O Material Design se utiliza desses pilares para a criação de boas *interfaces*.

O Material Design enfatiza a necessidade de um contraste entre as cores e ele realiza isso automaticamente, tanto no tema claro como no tema escuro integrado.

O Material Design 3.0 prevê dois sistemas de cores que podem ser utilizados, o sistema de cores estático e o sistema de cores dinâmico. No sistema estático o desenvolvedor escolhe manualmente as cores e criar seu próprio contraste. Já no sistema de cores dinâmico, a paleta de cores é gerada a partir de uma cor básica, isto pode ser utilizado para criar sistemas personalizáveis pelo usuário mantendo a consistência de cores e o contraste do Material Design.

Em ambos os sistemas, o sistema enquadra as cores da paleta em rótulos denominados 'cores-chave complementares', estas são: *Primary*, *Secondary*, *Tertiary*, *Neutral* e *Neutral variant*; A partir destas cores deve ser criada uma escala tonal para cada uma das cores-chave a fim de deixar o layout acessível. O Material Design recomenda um contraste mínimo de 3:1. Além das cores-chave também é apresentada a *Surface* que trás a função de ser a cor utilizada em fundos e grandes áreas com baixa enfase.

A cor-chave *Primary* deve ser utilizada nos componentes mais importantes da UI portanto, deve ter uma cor mais viva e chamativa. Os botões de alta enfase e estados ativos tem a mais alta hierarquia visual dentro da página, portanto devem ter a cor mais chamativa. A cor *Secondary* é utilizada nos componentes menos proeminentes da página, como botões tonais, ele tem uma cor mais neutra e se assemelha mais a cor de *Surface*. A *Secondary* é baseada na cor *Primary* mas com uma alteração no croma neutralizando a mesma. Já a *Tertiary* é usada para trazer contraste e equilíbrio com as cores *Primary* e *Secondary*. Ela é utilizada para trazer atenção a determinado elemento como um campo de entrada. As funções da cor de erro são estáticas mesmo num sistema de cores dinâmicas, ela não muda em nenhuma ocasião caso haja uma alteração das cores do layout.

Para cada cor-chave complementar são geradas outras cores que recebem nomes conforme suas funções dentro do sistema. Existe um prefixo *On* que significa que essa cor será aplicada sobre outra e o sufixo *Container* que significa que a cor será aplicada em um campo de espaço delimitado. Com essas nomeações e funções temos quatro aplicações nessas cores-chaves para cada uma delas: *Primary*, *On Primary*, *Primary Container* e *On Primary Container* como pode ser observado na Figura 1.2 . Há além disso o conceito de *Variant* que é usado para nomear uma opção com menos enfase em comparação a opção não variante.

Primary	Secondary	Tertiary	Error
On Primary	On Secondary	On Tertiary	On Error
Primary Container	Secondary Container	Tertiary Container	Error Container
On Primary Container	On Secondary Container	On Tertiary Container	On Error Container

Figura 5.2. Exemplo de token de cores

Além da cor para o conforto óptico, é recomendado que se alinhe também o tipo de texto, que o Material Design nos trás cinco tipos que são nomeados como, display, headline, title, body e label, a fonte utilizada na escrita, que podem tanto ser criada exclusivamente para determinado fim, como pode ser usada uma já existente que se encaixe nas necessidades e o uso de tokens para uma padronização e melhor coesão entre os assuntos dentro do layout.



Figura 5.3. Exemplo dos tipos de texto

Botões, menus, inputs, cards, etc. São o que o sistema nomeia de *Containers*, portanto todos os elementos de design interagem mutuamente, existem formas no estilo da fonte utilizada assim como tem uma cor vinda da paleta de cores-chave.

5.3.3. Movimento e Interatividade

Elevações e animações fazem com que a UI fique mais interativa e de fácil entendimento e usabilidade ao usuário. As elevações, por exemplo, permitem enfatizar a hierarquia dentro de um layout. Neste caso a elevação acontece no eixo z que é medida em *density-independent pixels* (dps), onde se tem a limitação de 12dps e cinco níveis de descanso, as elevações podem ser classificadas em três tipos de representação: através de alteração de cor entre as superfícies em diferentes alturas, através do sombreado da superfície mais alta e sobreposição de telas.

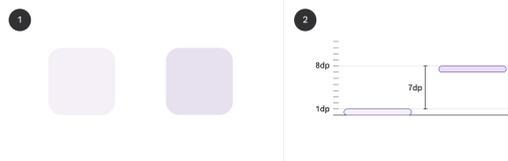


Figura 5.4. Representação de elevação no eixo Z

A versão 3 do Material Design nos trás os *easing tokens*, exemplificados na tabela, agrupados em dois conjuntos, o enfatizado e o padrão, onde o enfatizado é recomendado para a maioria das transições, já o padrão são transições pequenas e focadas e para o sistema iOS e Web é o único porque os mesmos não conseguem suportar o conjunto enfatizado, essas transições devem ser rápidas.

Por inspiração do mundo físico onde as coisas não começam e terminam mecanicamente, os *easing tokens* buscam deixar o movimento mais natural, acelerando e desacelerando assim suavizando o movimento e trazendo o especto de naturalidade.

5.3.4. Componentes

Os componentes do Material Design são os blocos interativos que estruturam a interface do usuário, organizando elementos visuais e ações de forma lógica e intuitiva. Eles estão divididos em categorias funcionais que servem tanto para guiar o design quanto para

EASING	DURAÇÃO	TIPO DE TRANSIÇÃO
Enfaticado	500ms	Iniciar e terminar na tela
Desaceleração enfatizada	400ms	Entrar na tela
Aceleração enfatizada	200ms	Sair da tela
Padrão	300ms	Iniciar e terminar na tela
Desaceleração padrão	250ms	Entrar na tela
Aceleração padrão	200ms	Sair da tela

padronizar a experiência do usuário em diferentes dispositivos. Esses componentes se destacam por sua flexibilidade e integração, facilitando a criação de layouts consistentes, fluídos e que seguem os princípios de uma boa experiência visual.

Entre os principais componentes de ação, encontram-se os botões, ícones de ação e botões flutuantes, que permitem aos usuários realizar funções primárias e secundárias com facilidade. A simplicidade de um botão comum ou a visibilidade de um botão flutuante exemplificam a clareza que esses elementos trazem para a interface, tornando a navegação mais direta.

Elementos de contenção, como cartões e diálogos, são usados para agrupar conteúdo relevante e ações associadas, ajudando a organizar visualmente a interface. Esses componentes permitem que as informações sejam apresentadas de forma acessível, facilitando a compreensão e a interação do usuário com o sistema.

Na comunicação, indicadores de progresso e snackbars desempenham papéis fundamentais ao fornecer feedback imediato sobre o estado de um processo ou tarefa. Esses elementos mantêm o usuário informado de maneira discreta e eficiente, sem sobrecarregar a interface.

A navegação é guiada por barras de ferramentas, menus e trilhos, que garantem que os usuários possam se mover entre diferentes seções do sistema sem dificuldades. A padronização desses componentes facilita a compreensão da interface, mesmo quando o usuário alterna entre diferentes dispositivos ou plataformas.

Os componentes de seleção e entrada de texto, como campos de texto, caixas de seleção e botões de rádio, permitem que o usuário insira informações e faça escolhas dentro do sistema de forma clara e direta. A simplicidade e a consistência desses elementos garantem uma interação mais eficiente e agradável.

Esses componentes, baseados em princípios de interação intuitiva e design coeso, garantem que as interfaces construídas com o Material Design sejam eficientes, responsivas e adequadas às expectativas do usuário moderno.

5.3.5. Acessibilidade e Responsividade

Além do conforto óptico de um bom design, o Material Design tem o objetivo de se adaptar a diferentes tamanhos de tela. Isso acontece pela flexibilidade de seus componentes que, em sua última versão, foram projetados para se adaptar automaticamente, se expandindo ou contraindo, a diferentes tamanhos de tela. O sistema de grade e layout que permite uma reorganização inteligente e automática, e esses componentes são intuitivos em relação as respostas das interações com o usuário. O texto também se adapta a dife-

rentes telas por meio da mudança de tamanho e espessura das letras para que se mantenha o conforto óptico e legibilidade.

Neste contexto, o usuário tem uma fácil navegação em diferentes tipos de tela, pois a interface se adapta de forma a prover uma experiência familiar e intuitiva. Além disso, visando a acessibilidade que gera o contraste de cores certo assim como a responsividade na interação com os componentes do layout.

5.4. BeerCSS

O BeerCSS é um projeto de código aberto com licença MIT em desenvolvimento contínuo por parte do patrocinadores e apoiadores no *GitHub* e *Open Collective*. O *framework* visa combinar simplicidade e eficiência em um pacote que atende às necessidades tanto de desenvolvedores iniciantes quanto experientes.

O encorajamento ao uso dessa ferramenta se deriva de diversas das suas qualidades, entre elas, destacam-se o amadurecimento, a adoção do Material Design em HTML semântico, o seu tamanho pelo menos dez vezes menor que os outros frameworks de CSS baseados no Material Design e a adaptabilidade a utilizar qualquer framework JavaScript (JS). Além disso, não existe a necessidade de compilação, configuração e dependências, uma vez que essa ferramenta é autosuficiente.

5.4.1. Introdução ao BeerCSS: Conceito e Filosofia

O lema do BeerCSS [BeerCSS 2022] é: "Build material design in record time without stress for devs". O BeerCSS é um framework de design leve, pautado pela filosofia minimalista. Criado para gerar interfaces elegantes e funcionais com agilidade, ele permite que desenvolvedores, tanto iniciantes quanto veteranos, trabalhem de forma eficiente sem enfrentar complicações desnecessárias.

O BeerCSS adota uma abordagem em que cada componente é pensado para ser direto e eficaz. O framework elimina a sobrecarga de estilos excessivos, concentrando-se em funcionalidades essenciais. Com isso, o desenvolvedor pode manter o foco na construção de interfaces nítidas e intuitivas sem se preocupar com configurações complexas ou dependências externas.

Sua modularidade destaca-se pela reutilização de componentes alinhada aos princípios do Material Design 3.0 (M3). Ao contrário de frameworks mais robustos que oferecem uma vasta gama de opções e personalizações, o BeerCSS concentra-se em um conjunto enxuto de ferramentas garantindo eficiência no processo de desenvolvimento. Isso favorece a adaptação das interfaces a diversos cenários, isto garante consistência visual clara e coerente.

Em projetos que exigem rapidez como, *Minimum Viable Product* (MVPs) ou sites temporários, o BeerCSS oferece uma solução ágil. Sua arquitetura leve permite ao desenvolvedor entregar interfaces modernas sem necessidade de ajustes extensivos ou longas compilações. Assim, o BeerCSS se consolida como uma ferramenta eficaz para criar projetos funcionais em tempo reduzido, sem sacrificar a qualidade da experiência do usuário.

5.4.2. Settings, Elements e Helpers

Inspirado pela "Lei de Pureza da Cerveja da Alemanha" (*Reinheitsgebot*), que limitava a produção de cerveja a três ingredientes essenciais — água, malte de cevada e lúpulo, o BeerCSS segue um princípio semelhante ao organizar sua documentação em três pilares principais: Settings, Elements e Helpers. Esses pilares simplificam o desenvolvimento, isto proporciona uma experiência ágil e eficiente, sem a complexidade de outras abordagens como *Block, Element, Modifier* (BEM) ou *Scalable and Modular Architecture for CSS* (SMACSS).

As *Settings* controlam aspectos globais da interface e definem as variáveis principais que afetam o comportamento e a aparência do projeto. Entre as funcionalidades mais importantes, estão a escolha entre os temas claro ou escuro e a personalização de cores e elevações dos componentes.

Abaixo estão as principais classes e variáveis de *Settings*:

- **Temas:** O desenvolvedor pode definir o tema padrão como claro ou escuro usando as classes **‘light‘** ou **‘dark‘** diretamente no elemento **‘<body>‘**, como sugere o Código 5.1.

```
<body class="light">...</body>
<body class="dark">...</body>
```

Código 5.1. Definindo o tema claro ou escuro

- **Cores Dinâmicas:** Utilizando **‘material-dynamic-colors‘**, é possível alterar o tema em tempo real com base em cores específicas ou imagens, criando uma experiência altamente personalizável, que, apesar da sua funcionalidade, não será profundamente abordado nesse capítulo. O Código 5.2 demonstra isto.

```
let theme = await ui("theme", "#ffd700");
```

Código 5.2. Alterando o tema dinamicamente com uma cor hexadecimal

- **Variáveis Globais:** O BeerCSS permite a customização completa das cores, animações e fontes usando variáveis CSS. Essas variáveis, como **‘—primary‘**, **—background**, **—shadow**, entre outras, ajustam a paleta de cores e a tipografia, garantindo consistência em todo o projeto. O Código 5.3 exibe um exemplo de como se parece uma paleta, como sugere o *Settings* [BeerCSS 2024], gerada através do Material Theme Builder [Material 2024].

```
:root,
body.light {
  --primary: #6750a4;
  --on-primary: #ffffff;
  --background: #fffbff;
  --on-background: #1c1b1e;
}
body.dark {
  --primary: #cfbcff;
  --on-primary: #381e72;
  --background: #1c1b1e;
  --on-background: #e6e1e6;
```

```
}
```

Código 5.3. Customizando o tema padrão

Os *Elements* são os componentes básicos do BeerCSS, como botões, cards, grids e ícones. Esses componentes seguem os princípios do Material Design, garantindo uma estética moderna e funcionalidade robusta.

- **Botões:** Permitem aos usuários executar ações com um único toque. Podem ser customizados com classes como **‘primary’**, **‘secondary’**, entre outras, para ajustar a cor e a aparência do botão.

- **Cards:** Superfícies que exibem conteúdo e ações relacionadas a um único tópico. São ideais para apresentar informações de forma clara e hierárquica.

Os *Helpers* adicionam flexibilidade e personalização adicional aos *Elements*. Eles são usados para ajustar o alinhamento, margens, sombras, cores e muito mais, sem a necessidade de criar CSS personalizado.

- **Alinhamento:** Helpers como **left-align**, **right-align**, **center-align** controlam o posicionamento de elementos dentro de seus contêineres.

- **Margens e Espaçamentos:** Classes como **small-margin**, **large-margin**, **no-margin** permitem ajustes rápidos e precisos no espaçamento entre os elementos.

- **Cores e Tipografia:** Helpers para cores (**amber-text**, **blue-border**) e tipografia (**bold**, **italic**, **uppercase**) permitem uma rápida customização da aparência textual e cromática.

Esses três pilares — **Settings**, **Elements** e **Helpers** — trabalham juntos para criar uma experiência de desenvolvimento eficiente e flexível, simplificando tanto o processo de design quanto a implementação de interfaces modernas e responsivas.

5.4.3. Layouts e Grid

O sistema de grid do BeerCSS é uma ferramenta essencial para a construção de layouts responsivos e flexíveis. Baseado em uma estrutura de colunas e linhas, esse sistema permite que os desenvolvedores criem layouts que se adaptam dinamicamente a diferentes tamanhos de tela e dispositivos. A abordagem modular do grid simplifica a tarefa de projetar interfaces que funcionam bem tanto em desktops quanto em dispositivos móveis, assim, pode garantir que a experiência do usuário seja consistente em qualquer tela.

O sistema de grid do BeerCSS divide o espaço em 12 colunas, permitindo que o conteúdo ocupe diferentes larguras dependendo do tamanho da tela. Helpers como **s1 — s12**, **m1 — m12** e **l1 — l12** controlam a quantidade de colunas ocupadas em dispositivos pequenos, médios e grandes.

```
<div class="grid">
  <div class="blue s12 m6 l3">
    <h3>PRIMEIRO</h3>
  </div>
  <div class="red s12 m6 l3">
    <h3>SEGUNDO</h3>
  </div>
```

```

<div class="yellow s12 m6 l3">
  <h3>TERCEIRO</h3>
</div>
<div class="purple s12 m6 l3">
  <h3>QUARTO</h3>
</div>
</div>

```

Código 5.4. Exemplo de Grid no BeerCSS

No exemplo apresentado no Código 5.4, cada div ocupa 12 partes em telas pequenas, 6 partes em telas médias e 3 partes em telas grandes. Na Figura 5.5 é possível visualizar como diferentes telas apresentam este conteúdo com disposição diferente dos elementos.

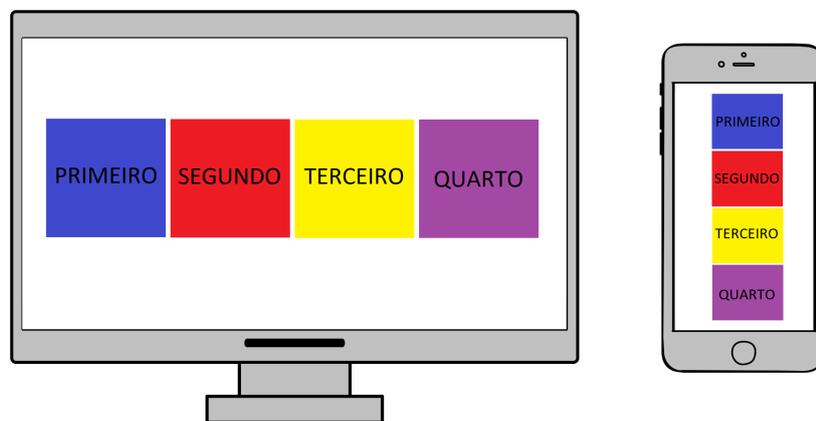


Figura 5.5. Visualização do resultado do código em desktops e dispositivos móveis

O espaço total da tela pode ser dividido como um contêiner para alinhar todo o conteúdo dentro do mesmo. Isto reaproveita o mesmo grid para várias seções ou áreas da mesma página, de forma que otimiza o uso dessa ferramenta.

Os layouts podem ser flexíveis ou fixos, utilizando as classes **absolute** para elementos relativos ao contêiner ou **fixed** para elementos fixos na tela. No Código 5.5 é apresentado o layout absoluto, enquanto no 5.6 um layout fixo.

O comportamento esperado é de que o Código 5.5 gere um contêiner de tamanho pequeno que ocupe toda a área da esquerda até a direita, assim, fazendo jus às suas características **'absolute left bottom right'**. Enquanto o do Código 5.6 deve criar um elemento que permanecerá no canto inferior da janela do navegador, cobrindo toda a largura da página, mesmo que o usuário role-a, graças às classes **'fixed left bottom right'**.

```

<article class="small">
  <div class="absolute left bottom right">
    <h5>Fundo do container</h5>
  </div>

```

```
</article>
```

Código 5.5. Exemplo de Layout Absoluto no BeerCSS

```
<div class="fixed left bottom right">  
  <h5>Fundo do documento</h5>  
</div>
```

Código 5.6. Exemplo de Layout Fixo no BeerCSS

5.4.4. Temas e Personalização

O BeerCSS oferece uma vasta gama de opções de personalização que permitem aos desenvolvedores ajustar tanto a aparência quanto o estilo das interfaces com facilidade. O framework se destaca por sua flexibilidade, permitindo a adaptação rápida de cores, tipografia e outros elementos visuais para que o design se alinhe perfeitamente com a identidade visual de um projeto ou marca.

O sistema de temas do BeerCSS é uma das suas principais características, permitindo a geração automática de paletas de cores completas, como demonstrado no Código 5.7. Isso oferece flexibilidade aos desenvolvedores, possibilitando a criação de funcionalidades de personalização para os usuários de maneira dinâmica. Alternativamente, o desenvolvedor pode optar por utilizar temas estáticos que requerem o trabalho manual de especificar cada cor da paleta do Material Design. Essa abordagem possibilita uma personalização visual sem a necessidade de reescrever o código.

```
let theme = await ui("theme", "#ffd700"); // Alterando pelo amarelo
```

Código 5.7. Alterando o tema dinamicamente com uma cor hexadecimal

Além de temas, o BeerCSS também simplifica a personalização de cores através de variáveis CSS, conforme o exemplo do Código 5.8. Essas variáveis permitem que a paleta de cores da interface seja rapidamente ajustada, facilitando a adaptação do design a preferências estéticas ou necessidades de branding.

```
{  
  --primary: #6750a4;  
  --on-primary: #ffffff;  
  --background: #ffffbf;  
  --on-background: #1c1b1e;  
}
```

Código 5.8. Exemplo de customização de cores no BeerCSS

O uso de variáveis CSS como ‘**—primary**’ e ‘**–background**’ facilita a personalização de cores, permitindo mudanças globais na aparência da interface com apenas algumas linhas de código.

Além das opções básicas de temas e cores, o BeerCSS também oferece uma ampla variedade de classes utilitárias, como as classes utilizadas na div do Código 5.9. Essas classes permitem ajustes rápidos e precisos em aspectos como margens, paddings e alinhamentos, sem a necessidade de escrever CSS adicional. Isso acelera o desenvolvimento e mantém o código mais limpo.

```
<div class="padding small-margin center-align">
  <p>Texto centralizado com margem pequena e padding.</p>
</div>
```

Código 5.9. Exemplo de uso de classes utilitárias no BeerCSS

Por fim, para a importação das fontes, baseado na Seção , deve ser feita como o conteúdo do Código 5.10 apresenta. O Material Theme Builder fará a importação dessa fonte logo após a escolha da paleta. Isso deve ser inserido dentro do arquivo CSS.

```
@import "https://fonts.googleapis.com/css2?family=Roboto:wght@400;500;700&display=swap";
:root {
  --font: "Roboto Flex";
  --font-icon: none;
}
```

Código 5.10. Exemplo de uso de classes utilitárias no BeerCSS

5.4.5. Instalação e Configuração do Ambiente

Nesta seção, abordaremos os preparativos necessários para a criação de um projeto utilizando o BeerCSS. A configuração inicial inclui a instalação do ambiente de desenvolvimento e a adição do framework ao projeto.

O primeiro passo envolve a instalação do BeerCSS no seu projeto. Existem duas maneiras principais de fazer isso: via CDN (Content Delivery Network) ou utilizando o NPM (Node Package Manager). Cada método possui suas vantagens, e você pode escolher de acordo com suas necessidades.

5.4.5.1. Utilizando CDN para Adicionar o BeerCSS

O método CDN oferece uma maneira rápida e simples de adicionar o BeerCSS ao seu projeto. Para utilizar este método, basta incluir as seguintes linhas de código no ‘<head>’ do seu arquivo HTML.

```
<head>
  <link href="https://cdn.jsdelivr.net/npm/beercss@3.6.13/dist/cdn/beer.min.css" rel="stylesheet"/>
  <script type="module" src="https://cdn.jsdelivr.net/npm/beercss@3.6.13/dist/cdn/beer.min.js"></script>
  <script type="module" src="https://cdn.jsdelivr.net/npm/material-dynamic-colors@1.1.2/dist/cdn/material-dynamic-colors.min.js"></script>
</head>
```

Código 5.11. Código de importação via CDN para BeerCSS

Essa abordagem carrega o BeerCSS diretamente da web, permitindo que você comece a utilizá-lo sem precisar instalar nada localmente. A vantagem do CDN é a simplicidade, especialmente útil para projetos menores ou protótipos rápidos, além de ser mais

simples para desenvolvedores iniciantes que não possuem experiência com o Node.js e o NPM.

5.4.5.2. Utilizando NPM para Adicionar o BeerCSS

Caso prefira um controle maior sobre o ambiente de desenvolvimento, a instalação via NPM pode ser a melhor opção. O NPM permite gerenciar dependências localmente e é amplamente utilizado em projetos de maior escala.

Para adicionar o BeerCSS ao seu projeto com NPM, execute os seguintes comandos no terminal:

```
npm i beercss  
npm i material-dynamic-colors
```

Código 5.12. Instalando o BeerCSS via NPM

Após a instalação, você pode importar o BeerCSS e utilizá-lo em seu projeto JavaScript da seguinte maneira:

```
import "beercss";  
import "material-dynamic-colors";
```

Código 5.13. Importando o BeerCSS no JavaScript

Essa abordagem permite que você tenha uma maior flexibilidade e controle sobre a versão e as atualizações do framework.

5.4.5.3. Testando a Configuração

Após a instalação do BeerCSS, é importante garantir que tudo está funcionando conforme o esperado. Para isso, vamos criar um exemplo simples que usa um botão estilizado pelo BeerCSS. O Código 5.14 mostrará um botão "Clique aqui" que utilizará o estilo padrão do framework.

```
<button class="primary">Clique aqui</button>
```

Código 5.14. Testando a configuração do BeerCSS

Ao executar este projeto, você deve ver um botão estilizado pelo BeerCSS. Se o botão aparecer corretamente, isso confirma que a instalação e configuração do BeerCSS foram bem-sucedidas.

5.4.6. Principais Componentes

O BeerCSS oferece uma série de componentes prontos para uso que facilitam a criação de interfaces consistentes e eficientes já apresentados anteriormente. Nesta seção, vamos explorar três dos componentes mais utilizados: *Botões*, *Cartões* e *Navbar*.

5.4.6.1. Botões

Os botões são elementos essenciais em qualquer interface. O BeerCSS simplifica sua criação ao fornecer classes que aplicam automaticamente estilos baseados no Material Design. O código desse componente corresponde ao conteúdo do Código 5.15 e a sua aparência na Figura 5.6.

```
<button class="primary">Clique aqui</button>
<button>
  <i>home</i> <span>Clique aqui</span>
</button>
<button>
   <span>Clique aqui</span>
</button>
```

Código 5.15. Botão primário no BeerCSS



Figura 5.6. Botões implementados em uma página web

5.4.6.2. Cartões

Os cartões são componentes versáteis usados para agrupar informações e ações relacionadas. Eles são frequentemente utilizados em interfaces modernas para exibir conteúdo de forma organizada. O Código 5.16 demonstra o elemento, a Figura 5.7 a sua visualização.

```
<article class="border">
  <h5>Titulo do cartao</h5>
  <p>Conteudo do cartao</p>
</article>
```

Código 5.16. Exemplo de cartão no BeerCSS

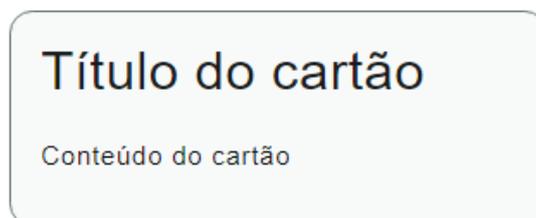


Figura 5.7. Cartão implementado em uma página web

5.4.6.3. Navbar

A barra de navegação (navbar) é um componente importante em muitas interfaces, proporcionando acesso fácil e rápido a diferentes seções de um site ou aplicativo. As linhas

do Código 5.17 exemplificam sua implementação e a Figura 5.17 o seu formato.

```
<nav>
  <button>Botao</button>
  <a class="chip">Chip</a>
  <i>home</i>
  
</nav>
```

Código 5.17. Exemplo de Navbar no BeerCSS

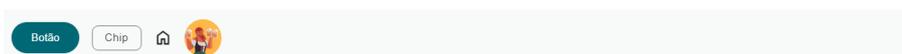


Figura 5.8. Navbar implementada em uma página web

Esse exemplo de navbar inclui um título de página centralizado e ícones de navegação em ambos os lados. A barra de navegação segue os padrões de design responsivo do BeerCSS.

5.5. Estudo de Caso: Aplicando Material Design com BeerCSS

Agora, nesta seção, como estudo de caso, será apresentado uma webpage baseada em um portfólio de desenvolvedor. A página segue as boas práticas da Otimização para Motores de Busca, ou Search Engine Optimization (SEO). Essa métrica serve para aumentar a visibilidade orgânica [Clarke 2023].

5.5.1. Implementação Prática: Criação de um Portfólio

Um portfólio de desenvolvedor deve ter seções definidas e organizadas, essas são: uma área de apresentação para caracterizá-lo ao leitor, uma área para suas produções bibliográficas e outra para os seus projetos pessoais, assim como um meio de contato.

Para implementação do Material Design utilizaremos alguns dos principais componentes do BeerCSS demonstrados na Seção 5.4. Para a página inteira, o grid, que viabiliza responsividade em diferentes tamanhos de tela. O header dessa página deve conter botões que visam facilitar a navegação pelo portfólio, nas informações sobre o desenvolvedor textos que o apresentem para o leitor. Assim como elementos textuais na parte que irá conter as suas produções bibliográficas

Já na área de tecnologias dominadas, deve ser utilizado elementos textuais e ícones gerados a partir de imagens. Na área de projetos, serão utilizados cartões, enquanto na área de contato mais conteúdo sobre o desenvolvedor e uma textarea para enviar mensagem. Por fim, um footer para sinalizar o fim da página.

Com o ambiente de desenvolvimento previamente configurado e a paleta de cores escolhida através do Material Theme Builder, deve-se inserir a linha de código apresentada no Código 5.18 no ‘<head>’ do HTML. Isso importará o arquivo CSS com as cores e fontes.

```
<head>
  <link rel="stylesheet" href="css/style.css">
```

```
</head>
```

Código 5.18. Importando o arquivo CSS para o HTML

Feito isso, o JS deverá ser importado de maneira semelhante no final da página, mas no ‘<body>’, como sugere o Código 5.19. Com isso as partes interativas da página, como o botão de alternar o tema, funcionarão corretamente.

```
<body>
  <script src="js/main.js"></script>
</body>
```

Código 5.19. Importando o arquivo JS para o HTML

Logo em seguida, escolha o tema padrão da página ao inserir ‘light’ ou ‘dark’ como classe no ‘<body>’, exatamente como no Código 5.20. Esta classe fará com que a página inicie sempre com o tema claro ou escuro.

```
<body class="light">
</body>
```

Código 5.20. Escolha do tema claro como padrão do portfólio

Agora, com todas as dependências ajustadas, o header é o primeiro componente a ser inserido na página. Esse componente serve para organizar e exibir conteúdo da página, geralmente relacionado à navegação. Neste deve haver uma navbar, o nome do desenvolvedor em escala reduzida, botões de navegação e um botão com ícone para alternar entre os temas. O Código 5.21 apresenta como ele é o estilo dela.

```
<header class="fixed max primary grid no-space small-padding no-margin"
  >
  <div class="s0 m2 l3"></div>
  <nav class="s12 m8 l6 grid tiny-space">
    <span class="bold s1 m3 l2 primary-container-text">Rafael Alves</span>
    <div class="s0 m0 l1"></div>
    <button class="s2 m2 l2 border primary-container small-round bold"
      >
      <a href="#artigos">Artigos</a>
    </button>
    <button class="s1 m1 l1 transparent square bold absolute middle
      right" id="theme-toggle">
      <i>dark_mode</i>
    </button>
  </nav>
  <div class="s0 m2 l3"></div>
</header>
```

Código 5.21. Exemplo de header no portfólio

As classes dos componentes que o BeerCSS dispõe são autoexplicativas.

• **Header:** A classe ‘fixed’ é referente ao documento, o que faz com que esse header fique sempre aparente na tela, enquanto ‘max’ faz esse elemento ocupar o máximo de

espaço possível da esquerda até a direita. Quanto à **‘primary’**, refere-se à cor, como exemplificado na Seção 5.4.2. O **‘grid no-space’** cria um grid sem espaço entre os elementos. Por fim, **‘small-padding’** faz o header assumir um tamanho menor entre o conteúdo e sua borda.

- **Nav:** Sua classe **‘grid tiny-space’** transforma esse elemento em um grid com um espaço pequeno entre cada elemento.

- **Button:** Cada botão tem uma disposição diferente na tela, dependendo do seu conteúdo interno. Baseado nisso, define-se o espaço que ele ocupa e se ajusta o grid. A classe **border** reforça as bordas do botão, **‘primary-container’** é a sua cor. O **‘small-round’** define o seu arredondamento e **‘bold’** que o seu conteúdo será negrito. A tag **‘<a>** conecta o botão até a parte referida da página, a fim de facilitar a navegação.

- **Toggle Theme:** O último botão é o responsável por alternar entre os temas, trata-se de um ícone. Sua aparência é transparente devido a sua classe **‘transparent’** e seu contorno é quadrado, derivado de **‘square’**. O restante das suas classes seguem a mesma lógica dos outros itens que já foram explanados durante os últimos parágrafos.

A página é separada em seções e todas tem o mesmo espaçamento entre si por uma **‘<div>** com a classe **‘large-space’**. Isto organiza visualmente o conteúdo, de forma que facilita a leitura. Os espaços variam de **‘tiny-space’** até **‘large-space’** e sempre serão utilizados para separar o conteúdo.

Como citado anteriormente, a próxima parte do portfólio é a apresentação do desenvolvedor. Neste exemplo utilizaremos informações fictícias. Assim, ele deve ser apresentado por meio de títulos, como o Código 5.22 demonstra. Todas as seções do texto têm três grid de área livre nas laterais e seis de conteúdo ao centro.

```
<h1 class="bold">Oi, sou o Rafael</h1>
<h6 class="">Desenvolvedor de software com foco em projetos open source
,
pesquisador e graduando Ciencia da Computacao pela UFPI.</h6>
<div class="small-space"></div>
<h6 class="">Neste site voce pode conferir todos os artigos tecnicos
que escrevi, ver alguns projetos pessoais que desenvolvi e as
ferramentas que eu utilizo.</h6>

```

Código 5.22. Exemplo de uma apresentação no portfólio

O sistema de grid que o BeerCSS oferece permite subdividi-lo inúmeras vezes, ou seja, permite que uma porção do espaço dividido em doze possa ser dividido em doze novamente. O grid e as suas funcionalidades tem uma explicação mais aprofundada na Seção 5.4.3.

As imagens podem ter diferentes tamanhos e formatos a depender da sua função. Neste caso, é a foto de perfil do desenvolvedor dessa página, então, tem dimensões pequena e responsivas, visível o suficiente em qualquer dispositivo. A classe **‘center’** e **‘middle’** alinha essa imagem ao centro da **‘<div>** que ela ocupa.

Logo após, devemos criar a seção das principais produções bibliográficas, como

no Código 5.23. Este é o título e um botão que leva o usuário para um repositório com todos os artigos disponíveis. Ele será criado dentro de um grid porque os elementos textuais (títulos e datas) estarão alinhados com o início do título e o fim do botão.

```
<div class="grid no-space" id="artigos">
  <nav class="s10 m8 l6">
    <h3 class="bold">Artigos</h3>
    <div class="max"></div>
    <button class="border small-round tertiary-border">
      <span class="tertiary-text bold">Ver todos</span>
    </button>
  </nav>
</div>
```

Código 5.23. Exemplo de um título de seção do portfólio

Todos os títulos de seções seguirão o mesmo padrão do Código 5.23. Este contém uma `<nav>`, onde o título é inserido através de um `'h3'` e a div com classe `'max'` ocupa todo o espaço vazio disponível naquela mesma linha. O cor do conteúdo dentro do botão é definido pela classe `'tertiary-text'`.

Depois do título, uma div com `'medium-space'` e, por fim, o título do trabalho, a data de lançamento e uma linha para dividir os seus subsequentes, assim como o Código 5.24 apresenta. Nesta exposição, recomenda-se até cinco títulos.

```
<h6 class="s8 m8 l8">BeerCSS: Interfaces Web Responsivas</h6>
<p class="right-align tertiary-text monospace s2 m3 l2">Agosto 2024</p>
<hr class="small s10 m8 l6">
```

Código 5.24. Exemplo do artigo no portfólio

Nessa seção as classes novas são `'tertiary-text'`, para colorir e dar destaque ao trecho e `'monospace'` para aplicar um estilo de fontes, ambas para identificação visual.

A próxima seção exige um título nos mesmos conformes da área de *Artigos* com o nome *Habilidades*, sem o botão ao lado. Por isso, o Código 5.25 nos apresenta somente o modelo de como cada habilidade deve ser disposta.

```

<h6 class="bold">Typescript</h6>
<p class="large-text top-align">Superconjunto de Javascript, com
  tipagem estatica e robustez.</p>
```

Código 5.25. Exemplo da habilidade no portfólio

Essa `<div>` é um elemento composto pelo ícone da tecnologia, o seu nome e uma legenda que explica para quê o desenvolvedor a utiliza. As classes da imagem são divididas entre `'left middle'`, para posicionamento dentro da sua `<div>`, `'round'` para arredondamento dos seus cantos e `'extra'` para torná-la um ícone. Nesta seção é recomendado um número par dessas tecnologias, caso contrário, criar uma `<div>` com a classe `'s12 m6 l6'`.

Após isso, a seção *Projetos* segue a mesma lógica de títulos de seção com um botão, como a de *Artigos*. Dentro de *Projetos*, serão inseridos os cartões apresentados na Seção 5.4.6.2, como exibido no Código 5.26.

```

<article class="medium-width border surface-container s12 m6 l4">
  <time class="">2020</time>
  <h5 class="bold">TakeNote</h5>
  <p class="">Um aplicativo open source de notas para paginas web.</p>
  <nav>
    <button class="border small-round tertiary-border">
      <span class="tertiary-text bold">Documentacao</span>
    </button>
    <button class="border small-round tertiary-border">
      <span class="tertiary-text bold">Demo</span>
    </button>
  </nav>
</article>

```

Código 5.26. Exemplo de projeto na portfólio

Depois de implementar todas as seções de conteúdo na página, o desenvolvedor deve colocar um campo para que o leitor, caso interessado, entre em contato. O Código 5.27 demonstra como criar essa área.

```

<h6 class=" s12 m12 l12">Se quiser entrar em contato, me envie uma
  mensagem na caixa de texto abaixo.</h6>
<div class="field border extra">
  <input type="text">
</div>
<button class="border small-round tertiary-border">
  <span class="tertiary-text bold">Enviar</span>
</button>

```

Código 5.27. Exemplo da área de contato no portfólio

Dessa forma, finaliza-se os elementos do portfólio e as suas disposições.

5.5.2. Habilitando os Temas

Nesta seção será exibido o código JavaScript desse projeto. O Código 5.28 contém as linhas que serão inseridas neste arquivo.

```

const toggleTheme = async () => {
  const currentMode = ui("mode");
  if (currentMode === "light") {
    ui("mode", "dark");
  } else {
    ui("mode", "light");
  }
};
document.getElementById('theme-toggle').addEventListener('click',
  toggleTheme);
document.getElementById("theme-toggle").addEventListener("click",
  function() {
    const icon = this.querySelector("i");
    if (icon.textContent === "light_mode") {
      icon.textContent = "dark_mode";
    } else {
      icon.textContent = "light_mode";
    }
  });

```

```
}  
});
```

Código 5.28. Conteúdo do arquivo JavaScript do projeto

5.6. Conclusão

Este capítulo explorou a evolução das interfaces, destacando a Interface Humano-Computador (IHC) e seu papel fundamental na melhoria da interação entre usuários e sistemas computacionais. A importância do Material Design foi enfatizada, revelando como suas diretrizes ajudam na criação de interfaces visuais que combinam estética e funcionalidade.

O BeerCSS foi apresentado como uma solução prática e eficiente, alinhada com os princípios do Material Design. Seu design leve permite a construção rápida de interfaces, e a configuração via CDN e NPM, além da implementação de temas claros e escuros, mostra sua flexibilidade. Essas ferramentas e técnicas ajudam desenvolvedores a criar experiências de usuário adaptáveis e atraentes, respondendo às demandas modernas de design e usabilidade. Por fim, demonstra a criação de um portfólio de desenvolvedor a partir dessa ferramenta.

O portfólio construído neste processo está disponível em [ERCEMAPI 2024] e o código-fonte pode ser acessado no repositório do GitHub em [Alves 2024].

Referências

- [Alves 2024] Alves, M. C. (2024). Código do beercss. Acesso em: agosto 26, 2024.
- [Apple 2024] Apple (2024). Human interface guidelines. Acesso em: 16 de agosto de 2024.
- [BeerCSS 2022] BeerCSS (2022). Beercss: A lightweight css framework. Disponível em: <https://www.beercss.com/>. Acessado em: 31 de maio de 2024.
- [BeerCSS 2024] BeerCSS (2024). Beercss settings documentation. <https://github.com/beercss/beercss/blob/main/docs/SETTINGS.md>. Acesso em: 24 ago. 2024.
- [Clarke 2023] Clarke, A. (2023). *SEO 2023: Learn Search Engine Optimization with Smart Internet Marketing Strategies*. Simple Effectiveness LLC. Focado em estratégias atualizadas de SEO para 2023.
- [Dulsrud and Bygstad 2022] Dulsrud, A. and Bygstad, B. (2022). Digital ecosystems as social institutions: exploring the role of consumption through four research streams of digital ecosystems. *Consumption and Society*, 1(1):99 – 119.
- [ERCEMAPI 2024] ERCEMAPI (2024). Portfólio ercemapi. Acesso em: agosto 26, 2024.
- [Eyal and Hoover 2014] Eyal, N. and Hoover, R. (2014). *Hooked: How to Build Habit-forming Products*. Penguin Business. Portfolio Penguin.

- [Franz and Kauer-Franz 2024] Franz, B. and Kauer-Franz, M. (2024). *Usability and User Experience Design: The Comprehensive Guide to Data-Driven UX Design*. Rheinwerk Computing. Rheinwerk Publishing.
- [Garrett 2010] Garrett, J. J. (2010). *The Elements of User Experience: User-Centered Design for the Web and Beyond*. Voices That Matter. Pearson Education, 2nd edition.
- [Google 2014] Google (2014). Material design 1.0. Disponível em: <https://material.io/design>. Acessado em: 16 de agosto de 2024.
- [Google 2022] Google (2022). Material design 3.0. Disponível em: <https://material.io/design>. Acessado em: 16 de agosto de 2024.
- [Inc. 2024] Inc., G. (2024). Google. Acesso em: 15 de agosto de 2024.
- [Johnson 2014] Johnson, J. (2014). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Guidelines*. Elsevier, London, 2 edition.
- [Kao and Archer 1997] Kao, D. and Archer, N. P. (1997). Abstraction in conceptual model design. *International Journal of Human-Computer Studies*, 46(1):125–150.
- [Krug 2000] Krug, S. (2000). *Don't Make Me Think!: A Common Sense Approach to Web Usability*. Circle.com Library. New Riders.
- [Marcotte 2011] Marcotte, E. (2011). *Responsive Web Design*. Book Apart. A Book Apart.
- [Material 2024] Material (2024). Material theme builder. <https://material-foundation.github.io/material-theme-builder/>. Acessado em: agosto 20, 2024.
- [Microsoft 2024] Microsoft (2024). Fluent design system. Acesso em: 16 de agosto de 2024. O Fluent Design System foi introduzido em 2017.
- [Nielsen et al. 2014] Nielsen, J., Pernice, K., and Whitenton, K. (2014). *How People Read on the Web: The Eyetracking Evidence*. Nielsen Norman Group, New York.
- [Norman 2002] Norman, D. A. (2002). *The Design of Everyday Things*. Basic Books, New York.
- [Raskin 2000] Raskin, J. (2000). *The Humane Interface: New Directions for Designing Interactive Systems*. ACM Press Series. Addison-Wesley Professional, illustrated edition.
- [Waldeilson and Paixão 2016] Waldeilson, A. L. Z. and Paixão (2016). Influência da estética visual na usabilidade e experiência do usuário em interfaces humano-computador. *Ergodesign HCI*, 4(Especial):93–101.
- [Zaphiris 2008] Zaphiris, P., editor (2008). *Human Computer Interaction: Concepts, Methodologies, Tools, and Applications, Volume 3*. IGI Global, illustrated edition.

Chapter

6

Tokens não Fungíveis (NFTs): Criação, Comercialização, Coleta e Análise de Dados na rede Blockchain Ethereum e na Plataforma de Vendas OpenSea

Samuel de Oliveira Ribeiro, Saul S. da Rocha, Dayan Ramos Gomes, Nara Raquel D. Andrade, Emanuel Aurélio F. de Miranda, Glauber Dias Gonçalves

Abstract

The support for smart contracts on the Ethereum blockchain network has enabled the emergence of a new digital market centered around Non-Fungible Tokens (NFTs), unique digital assets that have transformed the way value is assigned and exchanged in the decentralized web. This chapter provides a detailed exploration of three essential aspects that drive the NFT ecosystem: (1) the technical and operational fundamentals of creating and commercializing NFTs, (2) the categorization and automated analysis of collections using the OpenSea API, and (3) the challenges and future opportunities related to scalability, price prediction, and the impact of communities on the NFT market. We present the principles underpinning NFTs, highlighting OpenSea as the largest trading platform for these assets. Subsequently, we discuss data collection and automated categorization techniques, which are crucial for organizing and analyzing the market. Finally, we explore recent research and developments that will address challenges and enhance the growth of the NFT market in the coming years.

Resumo

O suporte a contratos inteligentes na rede blockchain Ethereum viabilizou o surgimento de um novo mercado digital baseado em Tokens Não Fungíveis (NFTs), ativos digitais únicos que transformaram a forma como o valor é atribuído e trocado na web descentralizada. Este capítulo aborda de forma detalhada três aspectos essenciais que impulsionam o ecossistema dos NFTs: (1) os fundamentos técnicos e operacionais da criação e comercialização de NFTs, (2) a categorização e análise automatizada de coleções utilizando a API da OpenSea, e (3) os desafios e oportunidades futuras relacionados à escalabilidade, previsão de preços e o impacto das comunidades no mercado de NFTs. Apresentamos os

princípios que sustentam os NFTs, com destaque para a OpenSea como a maior plataforma de negociação desses ativos. Em seguida, discutimos as técnicas de coleta de dados e categorização automatizada, fundamentais para a organização e análise do mercado. Finalmente, exploramos as pesquisas e desenvolvimentos mais recentes que enfrentarão os desafios e potencializarão o crescimento do mercado de NFTs nos próximos anos.

6.1. Introdução

Token não fungível ou NFT, do inglês *Non-fungible Token*, é um objeto digital registrado em redes blockchain, tipicamente, associado a conteúdo de texto ou imagem, que o confere características únicas e o torna também um objeto colecionável. Adicionalmente, NFT permite definição de um autor (criador), transferências de propriedade entre usuários, recebimento de royalties pelo autor em transferências, dentre outros recursos do ambiente distribuído de blockchains. Devido a essas características, NFTs vem sendo adotado por artistas para criação e distribuição de conteúdo digital, visando a proteção do direito autoral e o ganho com royalties na revenda de itens.

A credibilidade no comércio de NFTs se baseia nas propriedades da tecnologia blockchain como descentralização, transparência e imutabilidade para o registro transações de compra e venda. Primeiro, a descentralização permite processar e armazenar as transações em computadores organizados em uma rede par-a-par (P2P) sem um controle central, ou seja, todos os computadores participantes gerenciam a rede. Por conseguinte, a transparência decorre de todos os participantes da rede terem a mesma réplica das transações registradas na rede, facilitando o acesso a essas informações. A imutabilidade, por sua vez, ocorre via uma estrutura de dados (o encadeamento de blocos) que garante que uma transação registrada na rede pela maioria dos participantes não possa ser estornada e, gradativamente, todos tenham as mesmas transações. Em suma, a tecnologia blockchain e suas propriedades decorre da unificação de tecnologias computacionais maduras como mecanismos de consenso distribuído, criptografia e comunicação via redes P2P [Greve et al. 2018].

A tecnologia blockchain foi concebida originalmente para garantir segurança às transações da rede P2P de dinheiro eletrônico Bitcoin [Nakamoto 2008]. Contudo, o potencial dessa tecnologia fez com que seu uso se expandisse além do registro de transações para a execução de aplicações na rede blockchain, como é o caso de NFTs atualmente [Casino et al. 2019]. Essas aplicações, conhecidas por *contratos inteligentes* [Szabo 1997], são programas de computador autônomos, desenvolvidos em linguagens de alto nível e suportados desde a segunda geração de redes blockchain como a Ethereum [Wood 2014]. Os contratos inteligentes, uma vez iniciados, executam automaticamente e de acordo com seu código registrado na blockchain.

Vários contratos inteligentes constituem uma aplicação descentralizada ou DApp, do inglês *decentralized application*. Atualmente, DApps ocupam uma posição relevante nas atividades econômicas e financeiras internacionais, e despertam o interesse de governos e empresas [Schär 2021]. Existem uma variedade de DApps que oferecem um intrincado ecossistema de serviços, desde a área de finanças ao entretenimento digital. Tal ecossistema vem propiciando a emergência de uma nova geração da Internet baseada na infraestrutura de blockchain, popularmente intitulada *Web Decentral-*

izada [Murray et al. 2023]. Essa nova geração busca unificar as raízes descentralizadas da primeira geração da Internet, suportadas por conteúdos públicos e protocolos abertos (e.g., TCP/IP e HTTP), com as funcionalidades da geração vigente (dita Web2), baseadas em plataformas centralizadas, a exemplo, serviços em nuvem e redes sociais providas por Amazon, Google, Facebook, etc. A promessa da web descentralizada é que esses serviços tenham versões alternativas em DApps, favorecendo não apenas “Big Techs” mas todo um ecossistema de vários pequenos provedores de serviços tecnológicos na web.

A essência dos negócios em DApps, e por conseguinte NFT, se baseia no conceito de *token*, que é um objeto digital registrado na blockchain. Um *token* é único e está associado a um usuário, que é o seu proprietário, e somente este pode transferir a propriedade do *token* a outro usuário. Isso garante a possibilidade de valor ao *token*, i.e., a sua escassez ou impossibilidade de posses duplicadas (gasto duplo), diferentemente de objetos digitais tradicionais da web. Assim, blockchain atua como a tecnologia base que permite a posse de *tokens* em redes de acesso público e confiável com segurança garantida por criptografia.

A rede blockchain Ethereum é a pioneira na emissão de NFTs e concentra a vasta maioria de coleções desses *tokens* atualmente. Ela oferece padrões de contratos inteligentes específicos para esse tipo de token. A maioria desses padrões estabelecem um único contrato para uma coleção de NFTs. Por sua vez, cada NFT da coleção é vinculado de forma imutável a atributos que o tornam único. Esses atributos incluem em especial um link para um arquivo de mídia temático (e.g., jogos, fotos, vídeos ou desenhos), identificadores do autor e proprietário, valores de venda e royalties para o autor. Toda transferência de propriedade, i.e., venda, de NFT é automaticamente registrada como uma transação na rede blockchain.

Nesse contexto, NFTs abrem um novo caminho para utilização e veiculação de objetos digitais, i.e., tokens, sob a Internet, onde o aspecto mais relevante é a autoria ou propriedade do token. Por exemplo, em 11 de março de 2021, o artista Beeple realizou a venda de sua obra de arte digital em formato de NFT na blockchain Ethereum pelo valor de US\$ 69 milhões. Em 22 de março de 2021, o fundador do Twitter Jack Dorsey vendeu o NFT do seu famoso primeiro tweet pelo valor de US\$ 2,9 milhões. Essas obras podem ser acessadas gratuitamente na Internet e facilmente replicadas. Contudo, quanto mais popular e copiado na Internet é o token, mais benefícios ele pode trazer ao seu proprietário, que em tese possui direitos exclusivos sobre a sua comercialização e imagem [Okonkwo 2021].

Uma plataforma de comercialização de NFTs é um ambiente virtual onde são oferecidos vários serviços e facilidades para divulgação e vendas desses tokens. OpenSea é uma das plataformas de comercialização de NFTs pioneiras, que iniciou suas operações nessa área em 2018 e já negociou milhões de dólares em NFTs. Outras plataformas, contudo, vem crescendo em termos de negociações e comunidades de usuários que as suportam, se destacando as plataformas Blur, SuperRare e Rarible. Elas competem por melhores preços de venda, tarifas e royalties para os criadores de NFT. Esse mercado é extremamente dinâmico, e é difícil prever quais plataformas terão a liderança, atraindo maior volume de negociações futuramente.

Esse capítulo versa sobre os principais tópicos tecnológicos e de pesquisa envol-

vendo NFT. Nesse sentido, iniciamos descrevendo os principais fundamentos técnicos sobre a tecnologia blockchain na Seção 6.2. Como uma sequência desses fundamentos, apresentamos tópicos sobre o desenvolvimento de NFTs na Seção 6.3. A comercialização de NFTs via plataformas de exposição, compra e venda entre usuários são discutidas na Seção 6.4. Na Seção 6.5, são discutidos os desafios e oportunidades de pesquisa e inovação em NFTs. Finalmente, apresentamos nossas considerações finais com um resumo do capítulo na Seção 6.6. Importante mencionar que materiais adicionais, atualizações desse capítulo, códigos fonte, artigos e resultados de pesquisa desenvolvidos pelos autores podem ser obtidos no repositório https://github.com/LABPAAD/blockchain_nft.

6.2. Fundamentos

6.2.1. Blockchain

A tecnologia blockchain é fundamentada em uma estrutura de encadeamento de blocos, como ilustra a Figura 6.1, onde cada bloco contém um conjunto de transações registradas e é ligado ao bloco anterior através de um *hash* criptográfico. Essa estrutura garante que qualquer alteração em um bloco invalide todos os blocos subsequentes, tornando qualquer modificação na blockchain auditável, por conseguinte, imutável e resistente a fraudes. O *hash* de cada bloco é gerado a partir dos dados contidos nele e do *hash* do bloco anterior, criando uma cadeia contínua e segura de registros.

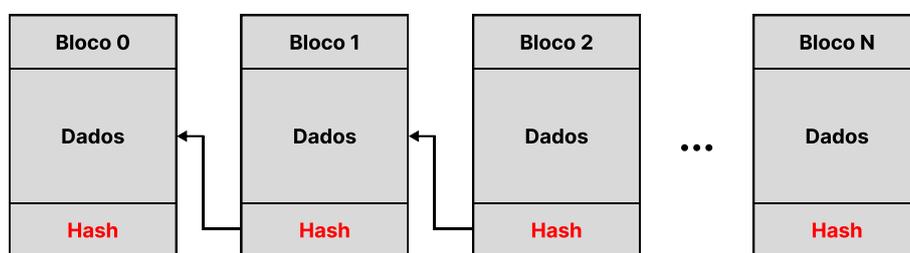


Figure 6.1: Estrutura de uma Blockchain.

Essa estrutura de dados contendo blocos encadeados (i.e., a blockchain) é replicada em vários computadores para ser armazenada de forma descentralizada. A replicação da blockchain é feita por meio de uma rede de computadores par-a-par (P2P), onde cada computador (i.e., nó da rede) possui uma cópia completa da blockchain, como mostra a Figura 6.2. Quando um novo bloco é adicionado, ele é transmitido para todos os nós na rede, e esses o valida de acordo com as regras de um mecanismo de consenso antes de incorporá-lo às suas cópias da blockchain. Esse modelo de replicação descentralizada garante que a blockchain seja resistente a falhas, pois não depende de um único ponto de controle, e qualquer tentativa de ataque requereria o comprometimento simultâneo de muitos nós.

Os principais mecanismos de consenso utilizados para validar as transações e adicionar novos blocos à blockchain incluem o *Proof of Work* (PoW) e o *Proof of Stake* (PoS), ilustrados na Figura 6.3. No PoW, os nós da rede competem para resolver prob-

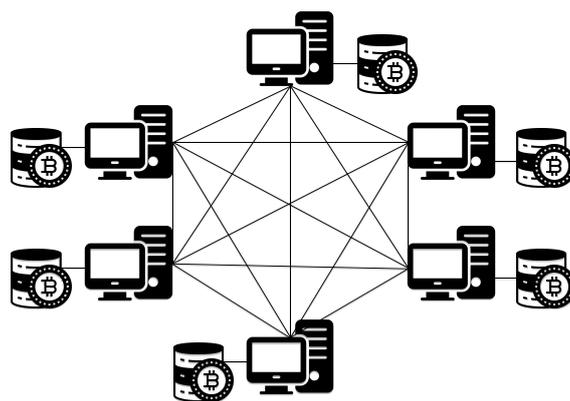


Figure 6.2: Replicação de uma Blockchain.

letras matemáticas complexas, e o primeiro a resolver se torna o proponente do novo bloco à blockchain, sendo recompensado por isso. Já no PoS, a probabilidade de um nó ser o proponente de um novo bloco é proporcional à quantidade de criptomoeda que ele possui e está disposto a "apostar" como garantia. Ambos os mecanismos visam garantir a segurança e integridade da blockchain, mesmo em ambientes descentralizados e potencialmente hostis.

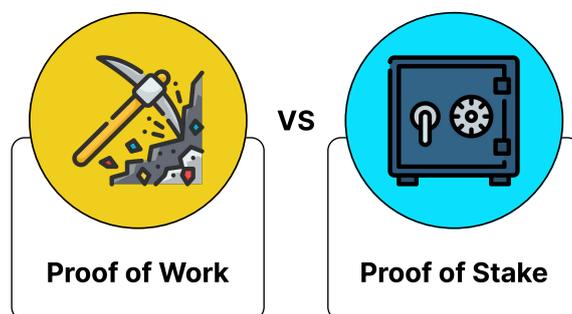


Figure 6.3: Representação de Proof of Work e Proof of Stake.

6.2.2. Contratos Inteligentes

Os contratos inteligentes [Szabo 1997] desempenham um papel essencial no ecossistema de Finanças Descentralizadas (DeFi), permitindo a automação e execução de acordos financeiros sem a necessidade de intermediários. Esses contratos são programas de computador autônomos registrados em uma blockchain, onde estabelecem regras e condições que orientam as interações entre as partes envolvidas. A execução é determinada pela lógica programada, o que resulta em maior previsibilidade e redução de custos nos acordos financeiros.

A rede blockchain Ethereum foi a primeira a suportar contratos inteligentes, possibilitando a criação de aplicações descentralizadas que funcionam sem intermediários centralizados [Wood 2014]. Contratos inteligentes, uma vez implantados na blockchain, tornam-se imutáveis e recebem um endereço exclusivo que permite a interação dos usuários. Esses contratos são escritos em linguagens de programação como Solidity, compilados em

bytecode, e executados por qualquer nó da rede, garantindo a segurança e a eficiência das transações.

Mais especificamente, um contrato inteligente é um programa de computador escrito em uma linguagem de alto nível, como ilustrado no Algoritmo 6.1, adaptado de [Palma et al. 2022], que exemplifica um contrato inteligente em Solidity chamado C. Ele possui uma variável de estado `nome` do tipo `bytes32` e duas funções. A função `get` recebe um parâmetro `_nome` e o atribui à variável `nome`, alterando o estado do contrato. A função `set` retorna o valor atual de `nome` sem modificar o estado, pois é marcada como `view`. O contrato demonstra operações básicas de leitura e escrita na blockchain, e a linha `pragma solidity >=0.7.0 <0.9.0;` especifica a versão do compilador compatível. Note que, após ser desenvolvido e testado, um contrato necessita ser implantado, ou seja, registrado de forma imutável, na blockchain.

```
1 // SPDX-License-Identifier: GPL-3.0
2 pragma solidity >=0.7.0 <0.9.0;
3
4 contract C{
5     bytes32 nome;
6     function get (bytes32 _nome) public{
7         nome = _nome;
8     }
9     function set () public view returns (bytes32) {
10        return nome;
11    }
12 }
```

Algoritmo 6.1: Exemplo de código fonte de contrato inteligente tomando com referência a linguagem Solidity Ethereum, adaptado de [Palma et al. 2022].

O contrato implantado na blockchain, junto com suas respectivas aplicações clientes, que transmitem as requisições dos usuários, constitui o que é conhecido como aplicação descentralizada ou DApp. Para interagir com o DApp, é necessário submeter uma transação à blockchain, especificando o endereço do contrato e a função desejada. Conforme ilustrado na Figura 6.4 proposta por [Palma et al. 2022], uma vez que o contrato inteligente é implantado, como demonstrado no bloco B_i pela transação "deploy C", os usuários podem interagir com ele em blocos subsequentes, como B_{i+1} e B_{i+2} , através de funções específicas do contrato, como `get()` e `set(true)`. Essas interações são realizadas por meio de uma aplicação cliente conectada à rede blockchain, equipada com as interfaces de comunicação necessárias para enviar as transações, garantindo que as operações programadas sejam executadas e registradas de forma imutável na blockchain.

As principais propriedades de segurança dos contratos inteligentes e da blockchain são descentralização, transparência e imutabilidade. A descentralização é garantida pela replicação da blockchain em uma rede de pares (P2P), onde cada nó mantém uma cópia completa do histórico de transações. Essa estrutura descentralizada distribui o poder de decisão entre todos os participantes da rede, eliminando a possibilidade de controle centralizado e tornando o sistema resistente a falhas e ataques. Mesmo que alguns nós sejam comprometidos, a integridade da blockchain permanece intacta, pois a maioria dos nós segue as regras de consenso estabelecidas.

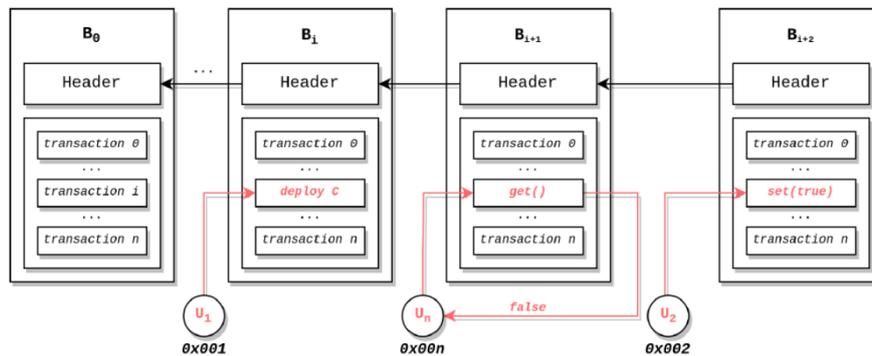


Figure 6.4: Exemplo de execução de uma transação que modifica um estado definido pelas variáveis programadas no contrato [Palma et al. 2022].

A transparência é assegurada pelo fato de que todas as transações e contratos registrados na blockchain são públicos e auditáveis por qualquer pessoa. Isso cria um ambiente em que todas as operações são visíveis e verificáveis, reduzindo significativamente o risco de fraudes ou manipulações. A transparência não apenas promove a confiança entre os usuários, mas também facilita a auditoria e a conformidade regulatória, elementos essenciais para a aceitação da tecnologia blockchain em sistemas mais amplos. A imutabilidade, garantida pelo uso de *hash* criptográficos, assegura que uma vez que um bloco é adicionado à blockchain, ele não pode ser alterado sem modificar todos os blocos subsequentes. Isso torna as transações e os contratos registrados permanentes e inalteráveis, garantindo a integridade dos acordos realizados. A imutabilidade elimina a possibilidade de alterações retroativas nos registros, garantindo que os dados armazenados permaneçam confiáveis ao longo do tempo.

Essas três propriedades — descentralização, transparência e imutabilidade — combinadas fornecem um ambiente seguro e confiável para a execução de contratos inteligentes. Eles não apenas aumentam a eficiência das transações, mas também democratizam o acesso a serviços financeiros e jurídicos, permitindo que indivíduos e organizações em qualquer parte do mundo participem de mercados globais de forma segura e eficiente.

6.2.3. Tokens

No contexto de Finanças Descentralizadas os *tokens* desempenham um papel central, atuando como objetos digitais registrados em uma blockchain. Esses *tokens* podem ser fungíveis ou não fungíveis, cada um com suas características e funcionalidades específicas. De forma simplificada, um *token* pode ser entendido como uma representação digital de um ativo, seja ele tangível ou intangível. Cada *token* é associado a um proprietário, e apenas este pode transferir sua propriedade a outro usuário, o que assegura a escassez e a impossibilidade de duplicações, evitando o problema do "gasto duplo". Essa propriedade exclusiva torna os *tokens* valiosos dentro do ecossistema blockchain.

A ideia geral dos *tokens* é tornar ativos mais acessíveis e facilitar as transferências ou negociações desses ativos de maneira eficiente e segura. Por exemplo, um ativo do mundo real pode ser representado na blockchain por um *token*, e a transação que altera a propriedade desse *token* refletirá a mudança de propriedade do próprio ativo no mundo

real. Esses *tokens* podem ser transferidos globalmente em questão de segundos e podem ser utilizados por diversas aplicações descentralizadas, desde que estejam vinculados a contratos inteligentes que controlam sua emissão e transferência.

Os *tokens* fungíveis foram os primeiros a serem desenvolvidos na plataforma Ethereum, por meio do padrão [ERC-20](#). Esses *tokens* são intercambiáveis, ou seja, cada unidade de um *token* fungível é igual a qualquer outra unidade do mesmo *token*, assim como cada unidade de uma moeda é igual a outra. O padrão ERC-20 estabelece um conjunto de regras e funcionalidades que permitem a criação, emissão e transferência de *tokens* fungíveis de forma consistente e segura. O padrão [ERC-20](#) define funcionalidades essenciais para que um objeto digital seja considerado um *token* fungível, como o nome, símbolo, número de decimais, total de *tokens* emitidos (*total supply*), e eventos como *Transfer* e *Approval*. O *Transfer* permite a transferência de *tokens* entre usuários, enquanto o *Approval* permite que o proprietário de um *token* autorize outra conta a gastar uma quantidade específica de seus *tokens*. Além disso, o padrão também inclui mecanismos de emissão (*minting*) e queima (*burning*) de *tokens*, controlados pelo contrato inteligente que emitiu o *token*.

Os *tokens* não fungíveis, conhecidos como NFTs, são um tipo específico de *token* que representa a propriedade de um ativo único ou colecionável na blockchain. Diferente dos *tokens* fungíveis, cada NFT é único e não pode ser trocado por outro de forma equivalente. O padrão [ERC-721](#), amplamente utilizado na Ethereum, define um conjunto de regras para a criação e interação com esses *tokens* digitais únicos e indivisíveis. Os principais aspectos do padrão ERC-721 incluem a unicidade e indivisibilidade dos *tokens*, permitindo que cada NFT tenha atributos exclusivos, como a propriedade de uma obra de arte digital ou um item de jogo. As transferências de NFTs são realizadas individualmente, com cada *token* possuindo seu histórico de propriedade e rastreabilidade na blockchain. O padrão também define métodos de interface, como *balanceOf* (saldo de *tokens* de um proprietário), *ownerOf* (proprietário de um *token*), e *transferFrom* (transferência de propriedade de um *token*), além de permitir o armazenamento de metadados adicionais que descrevem o ativo representado pelo *token*.

O padrão [ERC-1155](#) foi introduzido como uma evolução dos padrões anteriores, combinando as características dos *tokens* fungíveis e não fungíveis em um único contrato inteligente. Esse padrão permite a criação de múltiplos tipos de ativos digitais dentro de um único contrato, o que é especialmente útil em jogos e outras aplicações que requerem a gestão de diferentes tipos de itens, como moedas, armas, e acessórios. Uma das principais vantagens do [ERC-1155](#) é a eficiência nas transferências e na gestão de ativos, permitindo que várias instâncias de *tokens* sejam criadas e gerenciadas com menos custo e complexidade. Além disso, esse padrão oferece maior flexibilidade aos desenvolvedores, que podem criar e gerenciar diferentes tipos de *tokens* dentro de um único contrato inteligente, e também proporciona maior segurança, pois os ativos digitais são registrados na blockchain, garantindo sua autenticidade e imutabilidade.

Cada um dos padrões ERC-20, ERC-721 e ERC-1155 tem sua importância e aplicabilidade dentro do ecossistema blockchain. Enquanto o ERC-20 é fundamental para a criação de *tokens* fungíveis usados em diversas aplicações financeiras e utilitárias, o ERC-721 é crucial para o crescimento do mercado de NFTs, que permite a criação e comercial-

ização de ativos digitais únicos. O ERC-1155, por sua vez, oferece uma solução híbrida que combina o melhor dos dois mundos, permitindo uma gestão eficiente de múltiplos tipos de ativos digitais em um único contrato. Esses padrões, ao definirem regras claras para a criação, emissão e transferência de *tokens*, desempenham um papel vital na construção de um ecossistema blockchain robusto, seguro e interoperável, que suporta uma ampla variedade de aplicações descentralizadas e casos de uso inovadores.

6.3. Desenvolvimento de NFTs

6.3.1. Principais Conceitos

A concepção inicial das NFTs surgiu em 2012 com as "Moedas Coloridas" do Bitcoin, que consistiam em *tokens* representando diversos ativos físicos, como imóveis, carros e títulos, tendo como principal característica a exclusividade [Rosenfeld et al. 2012]. Este conceito evoluiu para o meio digital como ferramenta de compra e venda de artigos de jogos digitais em 2017, mas atualmente está inserido em uma vasta rede de negociações [Valeonti et al. 2021]. Os NFTs são implementados por meio de contratos inteligentes na rede blockchain Ethereum, que é a plataforma pioneira na criação e disponibilização de NFTs.

Atualmente os NFTs são compreendidos como uma forma de representação e comercialização de ativos únicos e indivisíveis, garantindo autenticidade e propriedade por meio da tecnologia blockchain. As negociações de NFTs acontecem em uma plataforma online onde as transações de compra e venda, bem como leilões e outros tipos de negociações, podem ocorrer. Um NFT pode ser uma versão digital de um item físico ou um registro digital de uma imagem, um vídeo, um som ou qualquer outro formato eletrônico. [Yilmaz et al. 2023]. Ao contrário dos *tokens* fungíveis (por exemplo, criptomoedas ou moedas fiduciárias), um NFT não é intercambiável com outro NFT em uma base um para um.

Aos NFTs são atribuídos códigos de identificação e metadados que os distinguem uns dos outros e garantem sua propriedade de unicidade. Existem duas partes em um NFT [Wlasinsky 2023]:

- Item NFT: O item digital associado a um NFT é descrito em metadados. O item, i.e a mídia ou bem ao qual o NFT está associado, não está armazenada diretamente dentro da blockchain Ethereum.
- Metadados de NFT (*token*): São armazenados em uma blockchain e geralmente incluem informações que identificam o item NFT, sua localização online, propriedade e informações de transações.

De modo a simplificar os conceitos apresentados, considere a icônica obra intitulada Mona Lisa, criada pelo renomado pintor italiano Leonardo da Vinci¹. Como amplamente conhecido, essa pintura é exibida no Museu do Louvre, em Paris [Le Louvre 2024].

¹Visualize e aprenda mais sobre essa famosa obra de arte https://pt.wikipedia.org/wiki/Mona_Lisa

Embora a Mona Lisa seja uma obra de arte física, ela, assim como outras peças artísticas, pode ser transformada em um NFT e armazenada em uma blockchain, possibilitando a digitalização e comercialização de sua representação no ambiente digital. Assim, o quadro seria o correspondente ao que definimos anteriormente como Item NFT. O item continuaria a ser armazenado no museu onde se encontra, porém, os metadados associados a ele seriam armazenados na rede Blockchain. O quadro seria descrito em metadados únicos e armazenados em Blockchain, contendo diversos tipos de informações de identificação do NFT, como localização, proprietário, autor, informações sobre histórico de transações, entre outras informações importantes.

Além das duas partes fundamentais já citadas (Item NFT e Metadados NFT), existe um atributo essencial que contribui para diversidade dos NFTs dentro de uma coleção: Os *Traits* (Traços). Os *Traits* são atributos específicos que definem as características visuais de um NFT e o diferenciam de outros da mesma coleção. Esses atributos podem incluir elementos visuais, como cor de fundo, formas, ou padrões, bem como características de raridade, como é apresentado na Figura 6.5, a seguir.



Figure 6.5: Exemplos de NFTs da coleção Bored Ape Yacht Club, adaptado de [Bored Ape Yacht Club 2024]

Todos os NFTs apresentados fazem parte da mesma coleção, intitulada Bored Ape Yacht Club. Porém, pode-se observar que eles possuem atributos visuais que os diferenciam entre si, como a cor de fundo, vestimentas, pelagem, formato dos olhos e boca, estes atributos são os *traits*. Além disso, os NFTs também possuem características que asseguram a manutenção dos princípios fundamentais que os regem, que são:

- Prova de propriedade e autenticidade: O NFT, por ser armazenado em blockchain, permite que a propriedade seja comprovada por históricos imutáveis armazenados na rede, bem como impede a falsificação.
- Indivisibilidade: Diferentemente das criptomoedas, os NFTs não podem ser divididos ou fracionado em partes menores, de modo que toda transação que envolva um NFT, utiliza a sua totalidade.
- Interoperabilidade: Os NFTs possibilitam interoperabilidade de ativos através de diversas plataformas devido ao padrão de *tokens* utilizados em sua criação.

- Imutabilidade: Uma vez criado e registrado em uma blockchain, os dados associados a um NFT são imutáveis, o que significa que eles não podem ser alterados ou apagados, garantindo a permanência das informações.

6.3.2. Exemplo Prático

A seguir temos a implementação de um código em Solidity que segue o padrão ERC-721 para criar e gerenciar *tokens* não fungíveis. Este contrato, ilustrado no Algoritmo 6.2, é construído utilizando as bibliotecas da *OpenZeppelin*, que oferecem uma base segura e robusta para o desenvolvimento de contratos baseados em Ethereum. O contrato exige que o compilador Solidity seja de versão 0.8.20 ou superior, garantindo que recursos e otimizações específicas desta versão estejam disponíveis. As bibliotecas da OpenZeppelin utilizadas incluem *ERC721URIS storage*, que estende o padrão básico ERC-721 ao permitir que cada *token* tenha um URI único associado, e *Ownable*, que facilita o controle de propriedade sobre o contrato, permitindo que certas funções sejam restritas ao proprietário [Entriiken et al. 2018].

A estrutura principal do contrato é definida como *MyNFT*, que herda as funcionalidades de *ERC721URISStorage* e *Ownable*. Dentro desta estrutura uma variável privada *tokenIds* é declarada, responsável por armazenar o número de *tokens* criados, assegurando que cada novo *token* receba um ID único. O construtor do contrato é executado no momento da implantação e tem duas funções principais. Ele define o nome "*MyNFT*" e o símbolo "*NFT*" da coleção de tokens, e também atribui a propriedade inicial do contrato ao endereço fornecido como argumento *initialOwner*.

Isso estabelece quem terá controle administrativo sobre o contrato desde o início. A função principal do contrato, *mintNFT*, permite a criação de novos NFTs. Essa função pode ser chamada apenas pelo proprietário do contrato, graças ao modificador *onlyOwner*. Dentro da função, o contador *tokenIds* é incrementado para gerar um novo ID de token. Em seguida, o *token* é emitido e atribuído ao endereço especificado como *recipient*. Após isso, a URI de metadados para o *token* recém-criado é configurada, associando informações descritivas ao token. Finalmente, a função retorna o ID do novo token.

```

1 // SPDX-License-Identifier: MIT
2 pragma solidity ^0.8.20;
3
4 import "@openzeppelin/contracts/token/ERC721/extensions/
   ERC721URISStorage.sol";
5 import "@openzeppelin/contracts/access/Ownable.sol";
6
7 contract MyNFT is ERC721URISStorage, Ownable {
8     uint256 private _tokenIds;
9
10    constructor(address initialOwner) ERC721("MyNFT", "NFT")
        Ownable(initialOwner) {}
11
12    function mintNFT(address recipient, string memory tokenURI)
13        public
14        onlyOwner
15        returns (uint256)
16    {
17        _tokenIds++;

```

```

18
19     uint256 newItemId = _tokenIds;
20     _mint(recipient, newItemId);
21     _setTokenURI(newItemId, tokenURI);
22
23     return newItemId;
24 }
25 }

```

Algoritmo 6.2: Exemplo de código fonte de contrato inteligente tomando com referência a linguagem Solidity Ethereum, adaptado de [Alchemy API 2024].

O Algoritmo 6.3 mostra a implementação de uma função em Solidity que é responsável por ampliar a funcionalidade de emissão, permitindo a criação de múltiplos NFTs em uma única transação. Ela recebe como parâmetros dois arrays: um contendo os endereços dos destinatários (*recipients*) e outro contendo os URIs de metadados correspondentes (*tokenURIs*). A função garante que o número de destinatários corresponda ao número de URIs, evitando erros de inconsistência. Dentro de um laço de repetição, a função incrementa o ID do *token* para cada nova emissão, atribui o *token* ao destinatário correspondente e associa o URI apropriado. Após o término do laço, o contador *tokenIds* é atualizado para refletir o número total de tokens cunhados. A função retorna o ID do último *token* criado.

```

1 function mintMultipleNFTs(address[] memory recipients, string[]
  memory tokenURIs)
2     public onlyOwner
3     returns (uint256)
4     {
5         require(recipients.length == tokenURIs.length, "
          Recipients and tokenURIs length mismatch");
6
7         uint256 newItemId = _tokenIds;
8         for (uint256 i = 0; i < recipients.length; i++) {
9             newItemId++;
10            _mint(recipients[i], newItemId);
11            _setTokenURI(newItemId, tokenURIs[i]);
12        }
13        _tokenIds = newItemId; // Atualiza _tokenIds uma vez
          ap s o loop
14        return newItemId;
15    }

```

Algoritmo 6.3: Exemplo de código fonte de contrato inteligente tomando com referência a linguagem Solidity Ethereum, adaptado de [Alchemy API 2024].

6.4. Comercialização de NFTs na plataforma OpenSea

A OpenSea se destaca como a maior e mais influente plataforma de comercialização de NFTs no mercado global. Desde seu lançamento em 2017, a OpenSea tem desempenhado um papel crucial no desenvolvimento e na popularização dos NFTs, oferecendo um marketplace acessível tanto para criadores quanto para colecionadores. A importância da OpenSea reside não apenas em seu tamanho e volume de transações, mas também em sua capacidade de suportar uma ampla variedade de ativos digitais, desde obras de arte e

música até itens de jogos e imóveis virtuais. Essa versatilidade, aliada à facilidade de uso, tornou a OpenSea a escolha preferida de milhões de usuários, consolidando sua posição como líder no ecossistema de NFTs.

Como serviço web centralizado, a OpenSea oferece uma interface intuitiva que simplifica o processo de criação, compra e venda de NFTs. Ao contrário de outras plataformas descentralizadas, a OpenSea centraliza suas operações, o que permite maior controle sobre a experiência do usuário e a implementação de funcionalidades adicionais, como criação automática de contratos inteligentes sem a necessidade de codificação manual. Esse modelo centralizado facilita o gerenciamento de usuários e transações, proporcionando uma experiência mais estável e segura. Além disso, a plataforma garante compatibilidade com uma ampla gama de carteiras digitais, como MetaMask, Coinbase Wallet, Fortmatic, entre outras, o que amplia o acesso e a conveniência para seus usuários, permitindo que operem com a carteira de sua preferência [Bhujel and Rahulamathavan 2022].

Embora a OpenSea seja a plataforma dominante, outras alternativas como Rarible e Foundation também oferecem recursos exclusivos que atraem diferentes públicos. A Rarible, por exemplo, adota uma abordagem mais descentralizada, permitindo que os usuários tenham maior controle sobre o processo de criação e venda de NFTs, além de oferecer um *token* de governança nativo (RARI) que permite aos usuários participar das decisões sobre o futuro da plataforma [Bhujel and Rahulamathavan 2022]. Já a Foundation se destaca por seu foco em obras de arte digital de alta qualidade e curadoria rigorosa, atraindo artistas renomados e colecionadores exigentes [Bhujel and Rahulamathavan 2022]. Essas plataformas, embora menores em volume de transações em comparação com a OpenSea, oferecem características e vantagens únicas que atendem a nichos específicos dentro do mercado de NFTs.

A OpenSea também se destaca pelo robusto suporte para uso de API, o que a torna uma ferramenta poderosa para desenvolvedores e pesquisadores interessados em acessar e analisar dados do mercado de NFTs. A API da OpenSea permite que os usuários obtenham informações detalhadas sobre ativos digitais, como metadados de NFTs, histórico de transações, preços de mercado e muito mais. Essa funcionalidade é particularmente útil para aqueles que desejam integrar os dados da OpenSea em suas próprias aplicações ou realizar análises avançadas sem precisar interagir diretamente com a interface da plataforma. Embora a API seja específica para o marketplace da OpenSea e não ofereça acesso direto à blockchain Ethereum, ela oferece um meio eficiente e acessível de trabalhar com os dados de NFTs em grande escala.

Além disso, a OpenSea simplifica o processo de comercialização de NFTs com um formulário que automatiza a criação de contratos inteligentes, permitindo que mesmo usuários sem conhecimentos técnicos possam criar e listar seus ativos digitais. A plataforma oferece diversas opções de venda, incluindo preço fixo, leilão e ofertas, o que dá flexibilidade aos vendedores para escolherem a estratégia que melhor atende às suas necessidades. No entanto, devido à popularidade dos NFTs, houve um aumento significativo em incidentes de segurança, como ataques a carteiras de criptomoedas e fraudes em marketplaces [Bhujel and Rahulamathavan 2022]. A OpenSea implementa medidas de segurança rigorosas para proteger seus usuários, como a verificação de contratos inteligentes e práticas de segurança cibernética. Contudo, é essencial que os usuários estejam cientes dos riscos,

como fraudes e questões de direitos autorais, e tomem as devidas precauções ao comercializar NFTs na plataforma.

6.4.1. API OpenSea

A API da OpenSea é uma ferramenta poderosa e amplamente utilizada para coletar e analisar dados relacionados a NFTs na plataforma. Como uma das maiores plataformas de comercialização de NFTs, a OpenSea oferece uma *API REST* que permite o acesso a uma vasta gama de informações, como preços, histórico de transações, detalhes sobre propriedade, e dados específicos de coleções. Essa API é particularmente útil para desenvolvedores e pesquisadores que desejam automatizar a coleta de dados e integrá-los em suas próprias aplicações ou estudos. Entre os diversos *endpoints* disponíveis, alguns dos mais utilizados incluem `/api/v1/collections` para obter listas de coleções, `/api/v1/collection/collection-name` para detalhes de coleções específicas, e `/api/v1/collection/collection-name/stats` para estatísticas em tempo real dessas coleções, conforme ilustrado na Tabela 6.1.

Código	Endpoint	Parâmetros
endpoint-1	<code>/api/v2/collections?offset=offset&limit=limit</code>	offset, limit
endpoint-2	<code>/api/v1/collection/collection-name</code>	collection-name
endpoint-3	<code>/api/v1/collection/collection-name/stats</code>	collection-name

Table 6.1: Exemplos de *Endpoints* da *OpenSea* utilizados para coletar os dados.

Para utilizar a API da OpenSea de maneira eficaz, é essencial configurar corretamente o ambiente de desenvolvimento e realizar chamadas HTTP para os *endpoints* desejados. A documentação oficial² da OpenSea fornece instruções detalhadas sobre como configurar a chave da API, que é necessária para autenticação e acesso aos dados. Após a configuração, os desenvolvedores podem realizar requisições programáticas para coletar dados específicos, como metadados de NFTs, histórico de vendas, e preços de mercado. Um exemplo prático seria utilizar o *endpoint* `/api/v1/asset/contract-address/token-id` para obter informações detalhadas sobre uma NFT específica, incluindo seu proprietário atual e o histórico de transações.

Durante o processo de coleta de dados, é comum configurar intervalos de tempo entre as requisições para evitar o bloqueio de chamadas pela API devido a limitações de taxa. Por exemplo, em um estudo típico, as requisições podem ser configuradas para ocorrer a cada segundo para *endpoints* que retornam listas de coleções e a cada 0,5 segundos para *endpoints* que fornecem detalhes específicos de coleções e suas estatísticas. Após a coleta, os dados são geralmente salvos em formatos estruturados, como *JSON*, que facilitam o processamento e a análise posterior. Esses dados podem ser armazenados em drives ou bancos de dados, dependendo da necessidade do projeto.

A estruturação dos dados coletados é crucial para garantir que as informações sejam facilmente acessíveis e processáveis. Ao salvar os dados em arquivos *JSON*, por exemplo, é possível manter a integridade das informações e garantir que todos os atributos

²<https://docs.opensea.io/reference/api-overview>

importantes sejam capturados de forma organizada. Essa abordagem permite que desenvolvedores e pesquisadores possam realizar análises mais profundas posteriormente, seja para estudos acadêmicos ou para o desenvolvimento de aplicações que dependem de informações atualizadas e precisas do mercado de NFTs.

6.4.2. Atributos e Categorias de NFTs

A API da OpenSea permite acessar uma ampla gama de metadados associados às coleções de NFTs, os quais são cruciais para a categorização e análise desses ativos digitais. Esses metadados incluem informações detalhadas sobre o desempenho, a popularidade e as características específicas das coleções de NFTs na plataforma. A partir dos dados coletados, é possível identificar diferentes categorias de NFTs e os atributos que as definem, permitindo uma análise mais profunda do mercado e facilitando a curadoria desses ativos.

Categoria	Descrição
<i>Art</i>	Obra de arte cunhada em Blockchain, podem ser obras de arte físicas digitalizadas ou podem ser criadas nativamente usando ferramentas digitais.
<i>PFPs</i>	<i>Profile pictures</i> são itens digitais que representam a propriedade de uma imagem ou obra de arte única e colecionável que pode ser usada como foto de perfil.
<i>Memberships</i>	Tipo de <i>token</i> não fungível que fornece acesso a uma experiência, utilidade ou comunidade e, em alguns casos, todos os três.
<i>Music</i>	Utilizado como um certificado digital de propriedade que usa blockchain para verificar e proteger a propriedade de um conteúdo relacionado à música.
<i>Photography</i>	Item digital que representa uma fotografia digital ou imagem em movimento.
<i>Gaming</i>	NFTs associados a qualquer item digital do reino dos jogos virtuais e do meta-verso como personagens, <i>skins</i> , personalizações, mapas, itens colecionáveis - qualquer criação digital que alguém usaria em ambientes de jogos virtuais.
<i>Virtual Worlds</i>	Ambientes digitais dentro de mundos virtuais ou metaversos, representando uma parte do espaço virtual que os usuários podem possuir, desenvolver e interagir.
<i>Domain Names</i>	Um <i>domain name</i> NFT é um <i>token</i> que armazena um nome de domínio no blockchain em vez de no tradicional Sistema de Nomes de Domínio, ou DNS.
<i>Sports</i>	Coleções de NFTs relacionados a esportes, como cartões colecionáveis digitais, arte esportiva e momentos históricos em formato digital.

Table 6.2: Categorias de coleções de NFTs utilizados na *OpenSea*, adaptado de [Ribeiro et al. 2024].

A Tabela 6.2 apresenta as principais categorias de coleções de NFTs disponíveis na OpenSea, como arte, música, fotografia, e outras. Cada uma dessas categorias é definida com base em atributos específicos, como o tipo de conteúdo representado e o uso pretendido dos NFTs. Por exemplo, a categoria "PFPs" (*Profile Pictures*) representa itens digitais que podem ser usados como fotos de perfil e possuem características visuais distintas, que são altamente valorizadas dentro da comunidade de colecionadores.

Além das categorias, os atributos das coleções de NFTs, detalhados na Tabela 6.3, desempenham um papel fundamental na análise e categorização dessas coleções. Esses atributos, como o volume total de vendas (*total-volume*), o número de proprietários (*num-owners*), e o preço médio de venda (*average-price*), oferecem *insights* importantes sobre a popularidade e o valor de uma coleção na plataforma. A combinação desses atributos permite uma compreensão mais completa das dinâmicas do mercado de NFTs na OpenSea, facilitando o desenvolvimento de ferramentas automatizadas para a curadoria e recomendação de NFTs. Esses dados são essenciais para construir modelos preditivos e para a implementação de sistemas de categorização que melhorem a experiência do usuário na plataforma.

Atributo	Descrição
<i>total-volume</i>	Total de vendas no <i>OpenSea</i> em ETH.
<i>total-sales</i>	Número total de transações de venda que ocorreram dentro da coleção.
<i>total-supply</i>	Total de NFTs criados para uma coleção.
<i>num-owners</i>	Número de proprietários de NFTs distintos na coleção.
<i>average-price</i>	Preço médio de venda dos itens da coleção no <i>OpenSea</i> .
<i>num-reports</i>	Número de relatórios de abuso no <i>OpenSea</i> sobre NFTs da coleção.
<i>market-cap</i>	Capitalização de mercado da coleção no <i>OpenSea</i> .
<i>qtd-traits</i>	Número de características visuais de todos os NFTs da coleção.
<i>floor-price</i>	Preço mínimo de venda dos NFTs da coleção no <i>OpenSea</i> .
<i>qtd-editors</i>	Quantidade de editores da coleção.
<i>category</i>	Categoria da coleção utilizada no <i>OpenSea</i> .

Table 6.3: Atributos contidos em metadados de coleções de NFTs extraídos da plataforma *OpenSea*, adaptado de [\[Ribeiro et al. 2024\]](#).

6.4.3. Análise de Dados de NFTs

Nesta seção, exploramos duas abordagens fundamentais para a análise de dados, cada uma baseada em diferentes paradigmas de classificação. Inicialmente, discutimos a abordagem de classificação supervisionada, que se baseia nos modelos de aprendizagem de máquina. Essa abordagem requer um conjunto de dados rotulados, onde o objetivo é treinar um modelo para prever a classe de novos dados. Em seguida, exploramos a abordagem de classificação não supervisionada, utilizando o método de classificação *K-Means*. Ao contrário da abordagem supervisionada, esse método permite identificar padrões nos dados para classificá-los utilizando apenas seus atributos, sem a necessidade de dados rotulados para treino.

6.4.3.1. Classificação Supervisionada

Existem diversas abordagens para as quais podem ser utilizadas técnicas de análise de dados de NFTs, em [\[Ribeiro et al. 2024\]](#) são apresentadas técnicas de classificação de NFTs a partir de dados coletados via API *OpenSea*, utilizando técnicas de aprendizado de máquina. O estudo demonstra como os dados de NFTs podem ser processados e analisados utilizando técnicas de aprendizagem de máquina e outras metodologias avançadas. Essas abordagens possibilitam uma compreensão mais profunda das dinâmicas de mercado e dos fatores que influenciam o valor e o comportamento de NFTs.

O artigo explora diversas técnicas que podem ser aplicadas para processar e analisar dados de NFTs. Entre as principais abordagens estão:

- **Filtragem e Limpeza de Dados:** Foram aplicadas técnicas de filtragem para remover dados inconsistentes ou irrelevantes, garantindo a qualidade do conjunto de dados final.
- **Análise Estatística:** Ferramentas de análise estatística foram utilizadas para identificar correlações entre as características dos NFTs e seu valor de mercado. Isso incluiu análise de regressão e cálculo de métricas como média, mediana e variância.
- **Clusterização e Categorização:** Os dados de NFTs foram submetidos ao algoritmo

K-Means e agrupados com base em suas coleções e características estruturais, permitindo uma análise comparativa entre diferentes grupos.

- **Visualização de Dados:** Para facilitar a interpretação dos resultados, foram criadas diversas visualizações gráficas, como histogramas, gráficos de dispersão e mapas de calor, que ilustram a distribuição de valores e a relação entre diferentes características dos NFTs.

Essas técnicas quando combinadas com as ferramentas corretas, permitem a extração de informações importantes, como a identificação de padrões de compra, a previsão de valorização de NFTs e a otimização de estratégias de venda em plataformas de mercado, bem como a classificação dos NFTs de acordo com seus atributos.

Os autores utilizam um conjunto de classificadores para identificar de forma ternária as categorias de coleções, considerando as duas maiores em número de coleções (i.e., *art* e *PFPs*) e todas as demais categorias minoritárias em uma única classe, identificada como *outros*. Os modelos de classificação utilizados foram: K-Vizinhos mais Próximos (KNN), Floresta Aleatória (RF) e Máquinas de Vetor de Suporte (SVM). Adicionalmente, foi realizado o balanceamento de classes com *undersampling*, de modo que foram removidas aleatoriamente amostras da classe majoritária para igualar à minoritária, e *oversampling*, que suplementa as classes minoritárias com instâncias de dados sintéticos que seguem a distribuição das instâncias reais.

Para conduzir a classificação, os dados foram organizados em 75% para treino e 25% para teste, treinando os classificadores com e sem balanceamento de classes. Os resultados obtidos são apresentados na Tabela 6.4, a seguir. Os autores optaram por apresentar apenas os melhores resultados para cada modelo.

Modelos	Art			PFPs			Outros			Acc	F1-macro
	P	R	F1	P	R	F1	P	R	F1		
KNN	0,70	0,68	0,69	0,38	0,20	0,26	0,51	0,61	0,56	0,61	0,50
SVM	0,68	0,87	0,76	0,60	0,09	0,16	0,67	0,51	0,58	0,67	0,50
RF (oversampling)	0,72	0,72	0,72	0,30	0,45	0,36	0,65	0,54	0,69	0,64	0,56

Table 6.4: Desempenho dos modelos avaliados com as métricas Precisão (P), Revocação (R), F1-score (F1), Acurácia (Acc) e F1-macro, adaptado de [Ribeiro et al. 2024].

Os resultados obtidos pelo estudo mostram que o modelo RF com *oversampling* obteve um desempenho equilibrado e estável em todas as métricas, destacando-se na classificação das categorias *art* e *outros*. Comparativamente, os modelos KNN e SVM obtiveram desempenhos inferiores, especialmente na categoria *PFPs*, ao passo que o uso da técnica de *oversampling* melhora o desempenho do classificador RF, tornando-o mais adequado para a classificação ternária.

A Figura 6.6 reporta as matrizes de confusão para o classificador RF, que obteve o melhor desempenho, e o classificador SVM, que obteve o segundo melhor desempenho. Na Figura 6.6a, observa-se que o modelo RF obtém maior número de acertos classificando a classe *Outros*, mas enfrenta maior dificuldade na diferenciação entre as classes *PFPs* e *Art*, bem como entre *Art* e *PFPs*. Já na Figura 6.6b observa-se que o modelo SVM possui maior erro na distinção entre as classes *Art* e *PFPs* e também na distinção da classe *Outros* com *Art* e *PFPs*, evidenciando o menor desempenho em relação a RF.

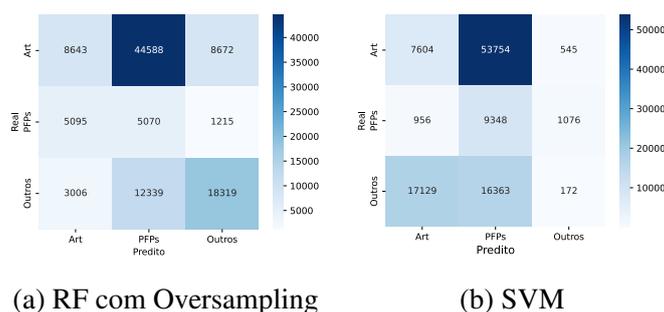


Figure 6.6: Matriz de confusão, adaptado de [\[Ribeiro et al. 2024\]](#)

6.4.3.2. Classificação Não Supervisionada

Em [\[Ribeiro et al. 2023\]](#), os autores realizaram uma análise detalhada das características estruturais dos NFTs no Ethereum. Os autores Utilizaram uma abordagem de classificação não supervisionada para conhecer propriedades estruturais dessas coleções. Isso nos permitiu definir quatro classes de coleções de NFTs que podem ser facilmente compreendidas por usuários para facilitar o comércio e a valoração de seus tokens. Assim como o trabalho citado na subseção anterior, este também faz uso de alguma técnicas que são aplicadas para o processamento e análise de dados, que são apresentadas a seguir:

- Filtragem e Limpeza de Dados: Foram aplicadas técnicas de filtragem para remover dados inconsistentes ou irrelevantes, garantindo a qualidade do dataset final.
- Agrupamento e Categorização: Os NFTs foram agrupados com base em suas coleções e características, permitindo uma análise comparativa entre diferentes grupos.
- Análise Estatística: Ferramentas de análise estatística foram utilizadas para identificar correlações entre as características dos NFTs e seu valor de mercado. Isso incluiu análise de regressão e cálculo de métricas como média, mediana e variância.
- Visualização de Dados: Para facilitar a interpretação dos resultados, foram criadas diversas visualizações gráficas, como histogramas, gráficos de dispersão e mapas de calor, que ilustram a distribuição de valores e a relação entre diferentes características dos NFTs.

Para a seleção das características a serem utilizadas, os autores realizaram o cálculo da correlação de Pearson entre todas as características obtidas na coleta, e em seguida definiram 3 grupos distintos de coleções com base nesses valores:

- Conjunto-1: todas as características.
- Conjunto-2: todas as características, removendo as que apresentaram alta correlação.
- Conjunto-3: Contendo apenas as características não correlacionadas com nenhuma outra característica

Considerando que os dados utilizados pelos autores não são rotulados, o uso do algoritmo de classificação não supervisionada definido requer a determinação do valor K (número de clusters). Para isso foi utilizado o método do cotovelo (também conhecido como *elbow method*), que funciona traçando um gráfico que mostra a variação da soma dos erros quadráticos em relação ao número de *clusters*. O objetivo é encontrar o ponto no gráfico onde a adição de mais um *cluster* não resulta em uma redução significativa na soma dos erros quadráticos. Combinando com esse método outras métricas, como Coeficiente de *Silhouette*, Índice de Calinski-Harabasz, Índice Davies-Bouldin, que são métricas de avaliação da qualidade da clusterização.

A Figura 6.7 mostra o resultado das métricas analisadas para cada conjunto de dados.

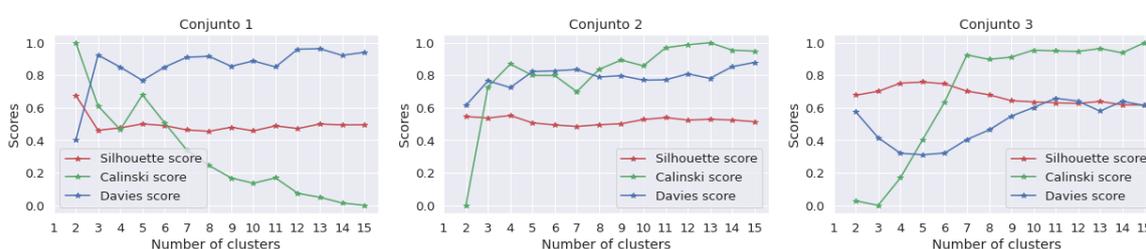


Figure 6.7: Resultado das métricas Coeficiente de *Silhouette*, Índice de Calinski-Harabasz e Índice Davies-Bouldin aplicadas aos 3 conjuntos de dados, adaptado de [Ribeiro et al. 2023].

A melhor escolha de K é aquela que maximiza o Coeficiente de *Silhouette* e o Índice de Calinski-Harabasz e minimiza o Índice Davies-Bouldin, então com base nisso, para prosseguir nas análises os autores optaram por utilizar os dados do conjunto-2, visto que ele foi o grupo que melhor obteve o valor K observando as métricas utilizadas e é o grupo que possui número intermediário de atributos. A partir dessas técnicas de avaliação e clusterização, os autores identificaram 4 clusters distintos e suas características estruturais, que apresentados a seguir:

- *Cluster 0*: Tem um volume de transações relativamente baixo em comparação com os outros *clusters*. No entanto, as vendas totais são bastante significativas (i.e., média de US\$ 246.75). Isso sugere que as transações desse *cluster* envolvem itens de alto valor. O número de relatórios (característica *num-reports*) é zero, o que pode indicar que a comunidade que negocia neste *cluster* é menos ativa em termos de relatar problemas. A quantidade média de características (i.e., característica *qtd-traits*) é intermediária, aproximadamente 5.04, sugerindo que este *cluster* pode ter alguns *traits* distintos, mas não é tão segmentado quanto o *cluster 3*.
- *Cluster 1*: Os membros deste *cluster* possuem uma quantidade total de *tokens* (característica *total-supply*) relativamente baixa de NFTs (média de 580.70 *tokens*), vendas e possuem 0 *traits*. Isso pode indicar que esses NFTs não são tão distintos ou interessantes para os compradores em comparação com os outros *clusters*, e também que a comunidade que negocia neste *cluster* é relativamente pequena e pode não ter uma demanda significativa por NFTs.

- *Cluster 2*: Este *cluster* tem o menor volume total entre os *clusters*, mas possui um número relativamente alto de vendas totais (i.e, 314.37) em comparação com outros *clusters*, o que indica que, apesar do baixo volume, os dados agrupados nesse *cluster* têm um alto valor de transações. A quantidade média de características é baixa (i.e, 0.70), sugerindo que este *cluster* pode não ter características únicas.
- *Cluster 3*: Este *cluster* tem o maior volume total, bem como o maior número de vendas e proprietários em comparação com os outros *clusters*. A quantidade média de *traits* é a mais alta (média de 11.32), sugerindo que este *cluster* possui características mais distintas e únicas em comparação com os outros. Além disso, este *cluster* tem um número significativo de relatórios (média de 0.88), indicando uma forte atividade em torno dele.

<i>Cluster</i>	<i>total-volume</i>	<i>total-sales</i>	<i>total-supply</i>	<i>num-reports</i>	<i>qtd-traits</i>
0	13.40	246.75	1347.60	0.00	5.04
1	9.26	74.42	580.70	0.00	0.00
2	0.00	314.37	644.59	0.00	0.70
3	61885.75	22064.33	13414.64	0.88	11.32

Table 6.5: Valor médio das características para cada *cluster*. Adaptado de [Ribeiro et al. 2023]

Na Tabela 6.5, a seguir, são apresentadas o valor médio das características consideradas para cada *cluster* identificado. Ainda de acordo com os autores, as características médias de cada *cluster* podem ajudar a entender melhor a natureza dos NFTs em cada grupo e informar decisões de investimento ou comercialização. Por exemplo, um investidor pode optar por focar em NFTs do *cluster 2*, que possuem um valor total de vendas relativamente alto, mas baixo volume total, enquanto pode decidir evitar NFTs do *cluster 3* devido ao número médio de relatórios negativos que eles receberam.

6.5. Desafios e Oportunidades Futuras dos NFTs

Os NFTs têm mostrado um crescimento impressionante, mas também enfrentam uma série de desafios que precisam ser abordados para garantir sua sustentabilidade e evolução. Esta seção discutirá alguns dos principais desafios e oportunidades no ecossistema de NFTs, considerando a concorrência à rede Ethereum, a categorização e automação de NFTs, a predição de preços, e as dinâmicas de comunidades e efeitos de rede no comércio de NFTs.

- **Concorrência à rede Ethereum:** A Ethereum é indiscutivelmente a plataforma mais popular para NFTs devido à sua capacidade de executar contratos inteligentes, que são fundamentais para a criação, venda e transferência de NFTs. Os contratos inteligentes da Ethereum permitem que transações sejam realizadas de forma autônoma e sem intermediários, garantindo segurança e transparência. No entanto, a popularidade da Ethereum tem vindo acompanhada de desafios significativos. Um dos maiores problemas enfrentados pela rede é a escalabilidade. Conforme

o número de transações cresce, a rede torna-se congestionada, resultando em altos custos de transação, conhecidos como *gas fees*. Esses custos podem ser proibitivos, especialmente para pequenos criadores e compradores, limitando a acessibilidade e a adoção em massa dos NFTs.

Devido a esses desafios, alternativas à Ethereum têm surgido, oferecendo soluções mais eficientes em termos de escalabilidade e custos. Solana, por exemplo, é uma blockchain de alta performance que promete transações rápidas e taxas baixas³, o que a torna atraente para criadores e colecionadores de NFTs. Outra alternativa, a Tezos, destaca-se por sua governança em cadeia, permitindo que a rede evolua sem a necessidade de hard forks, e por sua abordagem sustentável com um mecanismo de consenso proof-of-stake, que consome menos energia. Essas plataformas emergentes estão ganhando popularidade e podem representar uma concorrência significativa para a Ethereum no futuro próximo, especialmente se continuarem a oferecer vantagens claras em termos de custo e eficiência⁴.

- **Categorização de NFTs e sua automação:** A categorização eficiente de NFTs é um desafio fundamental para melhorar a experiência do usuário nas plataformas de comercialização, como a OpenSea. Com a explosão do número de NFTs disponíveis, a dificuldade em navegar e encontrar itens de interesse aumenta. Para enfrentar esse problema, nossa pesquisa [Ribeiro et al. 2024], propõe um sistema automatizado de categorização de NFTs usando técnicas de *machine learning*. Essa abordagem visa classificar NFTs de acordo com seus metadados e atributos visuais, facilitando a descoberta de coleções e auxiliando na curadoria dos conteúdos disponíveis nas plataformas.

Os resultados preliminares da nossa pesquisa mostram que a automação da categorização não apenas economiza tempo, mas também melhora a precisão na classificação de NFTs em categorias relevantes. Esse avanço é crucial para lidar com o volume crescente de NFTs, permitindo que plataformas como a OpenSea mantenham uma organização eficiente e que os usuários encontrem facilmente os ativos que procuram. Além disso, a categorização automatizada pode ajudar a identificar novas tendências no mercado de NFTs, permitindo que coleções emergentes ganhem visibilidade mais rapidamente e que os criadores adaptem suas estratégias conforme as preferências dos usuários evoluem.

- **Predição de preços de NFTs:** A predição de preços de NFTs é uma tarefa complexa, devido à natureza especulativa e altamente volátil desse mercado. Estudos recentes, como os de [Costa et al. 2023, Nadini et al. 2021]), têm explorado métodos para prever os preços dos NFTs utilizando uma combinação de análise de dados e aprendizado de máquina. Esses estudos focam em variáveis como o histórico de vendas, a popularidade do criador e os atributos específicos dos NFTs, como a raridade, para construir modelos preditivos que podem auxiliar investidores e criadores a tomar decisões mais informadas.

Apesar do progresso significativo, a predição de preços de NFTs continua a enfrentar desafios devido à imprevisibilidade das tendências de mercado e à influên-

³<https://br.cointelegraph.com/news/solana-beats-ethereum-weekly-total-fees-research>

⁴<https://blockzeit.com/pt/rivais-do-ethereum-os-4-maiores-concorrentes>

cia de fatores externos, como o comportamento das comunidades e as flutuações no interesse público. Os modelos desenvolvidos até agora conseguem capturar padrões de curto prazo com razoável precisão, mas a previsão a longo prazo permanece uma área de grande incerteza. Isso indica que, enquanto as ferramentas de predição podem fornecer *insights* valiosos, elas devem ser usadas com cautela e complementadas por outras formas de análise e monitoramento contínuo do mercado [Nadini et al. 2021]

- Comunidades e efeitos de rede no comércio de NFTs: As comunidades formadas em torno dos NFTs e os efeitos de rede que surgem dessas interações são fundamentais para o sucesso e a valorização desses ativos digitais. O trabalho de [Nadini et al. 2021] explora como as redes de compradores e vendedores afetam o valor dos NFTs, mostrando que a dinâmica social dentro dessas comunidades pode impulsionar o crescimento do mercado de maneira exponencial. Efeitos de rede ocorrem quando o valor de um NFT aumenta à medida que mais pessoas se envolvem com ele, criando um ciclo de valorização sustentado pelo engajamento da comunidade.

[Vasan et al. 2022] ampliam essa análise ao investigar como as características das comunidades impactam diretamente o sucesso das coleções de NFTs. Eles observam que as coleções apoiadas por comunidades ativas e engajadas tendem a ter volumes de negociação mais altos, o que não só aumenta a liquidez dos NFTs, mas também fortalece a marca da coleção. O estudo de [Colavizza 2023] vai além, sugerindo que as comunidades em torno de NFTs estão se tornando centros de inovação dentro do metaverso, onde a propriedade e o comércio de ativos digitais são integrados em experiências sociais imersivas. Colavizza argumenta que o futuro dos NFTs não se limita à sua função como colecionáveis digitais, mas se estende à criação de ecossistemas digitais onde a interação social e o comércio se fundem.

Esses estudos mostram que o valor dos NFTs está intrinsecamente ligado à força das comunidades que os sustentam e ao impacto dos efeitos de rede. Para plataformas de NFTs, entender e fomentar essas dinâmicas é crucial para manter o crescimento sustentável e a relevância no mercado. As comunidades não apenas criam valor financeiro para os NFTs, mas também estabelecem laços sociais e culturais que podem definir o futuro dos mercados digitais. Assim, a capacidade de uma plataforma de construir e nutrir comunidades fortes pode ser um diferencial significativo no cenário competitivo dos NFTs [Nadini et al. 2021, Vasan et al. 2022, Colavizza 2023].

6.6. Considerações Finais

Este trabalho explorou os principais aspectos tecnológicos e de pesquisa relacionados aos NFTs, com foco na criação, comercialização, coleta e análise de dados na rede blockchain Ethereum e na plataforma OpenSea. A tecnologia blockchain, ao oferecer propriedades como descentralização, transparência e imutabilidade, tem se mostrado fundamental para o desenvolvimento e sucesso dos NFTs. A Ethereum, como a principal

rede para a emissão e transação desses tokens, continua sendo a escolha dominante, embora enfrente desafios em termos de escalabilidade e custos, que abrem espaço para a concorrência de outras blockchains emergentes.

No decorrer do capítulo, foram discutidos os fundamentos técnicos da blockchain e dos contratos inteligentes, que são essenciais para a operação e segurança dos NFTs. A comercialização de NFTs na plataforma OpenSea foi analisada detalhadamente, destacando-se como a maior e mais influente plataforma nesse mercado, oferecendo uma interface centralizada, suporte para múltiplas carteiras digitais, e diversas opções de comercialização. A seção de coleta e análise de dados demonstrou a importância da API da OpenSea para a obtenção de metadados e informações críticas para a pesquisa e desenvolvimento de ferramentas automatizadas.

Além disso, foram discutidos os desafios e oportunidades futuras no campo dos NFTs, incluindo a necessidade de melhorar a categorização e automação desses tokens, prever seus preços e entender as dinâmicas das comunidades que os suportam. A pesquisa e o desenvolvimento contínuos nesses aspectos são cruciais para o avanço e a sustentabilidade do ecossistema de NFTs, oferecendo novas oportunidades de inovação e exploração dentro do mercado digital global.

Em suma, este trabalho fornece uma visão abrangente dos NFTs, desde seus fundamentos técnicos até as perspectivas futuras, contribuindo para um melhor entendimento e desenvolvimento deste campo emergente na interseção entre tecnologia e economia digital.

Agradecimentos

Os autores agradecem o apoio da Fundação de Amparo à Pesquisa do Piauí (FAPEPI) processo no. 00110.000235/2022-78, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI) da Universidade Federal do Piauí.

References

- [Alchemy API 2024] Alchemy API (2024). How to create an nft on ethereum tutorial. <https://docs.alchemy.com/docs/how-to-create-an-nft> (Accessed on 05/26/2024).
- [Bhujel and Rahulamathavan 2022] Bhujel, S. and Rahulamathavan, Y. (2022). A survey: Security, transparency, and scalability issues of nft's and its marketplaces. *Sensors*, 22(22):8833.
- [Bored Ape Yatch Club 2024] Bored Ape Yatch Club (2024). Bored Ape Yatch Club. <https://opensea.io/collection/boredapeyachtclub>. Accessed: 2024-08-26.
- [Casino et al. 2019] Casino, F., Dasaklis, T. K., and Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and informatics*, 36:55–81.

- [Colavizza 2023] Colavizza, G. (2023). Seller-buyer networks in nft art are driven by preferential ties. *Frontiers in Blockchain*, 5:1073499.
- [Costa et al. 2023] Costa, D., La Cava, L., and Tagarelli, A. (2023). Show me your nft and i tell you how it will perform: Multimodal representation learning for nft selling price prediction. In *Proceedings of the ACM Web Conference 2023*, pages 1875–1885.
- [Entriken et al. 2018] Entriken, W., Shirley, D., Evans, J., and Sachs, N. (2018). Eip-721: Non-fungible token standard," ethereum improvement proposals, no. 721, january 2018. <https://eips.ethereum.org/EIPS/eip-721>.
- [Greve et al. 2018] Greve, F., Sampaio, L., Abijaude, J., Coutinho, A. A., Brito, I., and Queiroz, S. (2018). Blockchain e a Revolução do Consenso sob Demanda. In *Proc. of SBRC Minicursos*.
- [Le Louvre 2024] Le Louvre (2024). From the 'Mona Lisa' to 'The Wedding Feast at Cana'. <https://www.louvre.fr/en/explore/the-palace/from-the-mona-lisa-to-the-wedding-feast-at-cana>. Accessed: 2024-08-26.
- [Murray et al. 2023] Murray, A., Kim, D., and Combs, J. (2023). The promise of a decentralized internet: What is web3 and how can firms prepare? *Business Horizons*, 66(2):191–202.
- [Nadini et al. 2021] Nadini, M., Alessandretti, L., Giacinto, F. D., Martino, M., Aiello, L. M., and Baronchelli, A. (2021). Mapping the nft revolution: market trends, trade networks and visual features.
- [Nakamoto 2008] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- [Okonkwo 2021] Okonkwo, I. E. (2021). Nft, copyright; and intellectual property commercialisation. *SSRN*. <https://ssrn.com/abstract=3856154>.
- [Palma et al. 2022] Palma, L., Martina, J., and Vigil, M. (2022). On and off: Extracting the transaction history of permissioned blockchain networks. Universidade Federal de Santa Catarina. Computing Science PhD Candidate Dissertation.
- [Ribeiro et al. 2024] Ribeiro, S. d. O., Gomes, D. R., Andrade, N. R. D., de Miranda, E. A. F., and Gonçalves, G. D. (2024). Classificação de coleções de nfts explorando metadados e aprendizagem de máquina. In *Colóquio em Blockchain e Web Descentralizada (CBlockchain)*, pages 50–55. SBC.
- [Ribeiro et al. 2023] Ribeiro, S. O., Gomes, D. R., Coutinho, E., and Gonçalves, G. D. (2023). Análise de características estruturais de tokens não fungíveis no ethereum. In *Anais do VI Workshop em Blockchain: Teoria, Tecnologias e Aplicações*, pages 1–14. SBC.
- [Rosenfeld et al. 2012] Rosenfeld, M. et al. (2012). Overview of colored coins. *White paper; bitcoil. co. il*, 41:94.
- [Schär 2021] Schär, F. (2021). Decentralized finance: On blockchain-and smart contract-based financial markets. *FRB of St. Louis Review*.

- [Szabo 1997] Szabo, N. (1997). Formalizing and securing relationships on public networks. *First monday*.
- [Valeonti et al. 2021] Valeonti, F., Bikakis, A., Terras, M., Speed, C., Hudson-Smith, A., and Chalkias, K. (2021). Crypto collectibles, museum funding and openglam: challenges, opportunities and the potential of non-fungible tokens (nfts). *Applied Sciences*, 11(21):9931.
- [Vasan et al. 2022] Vasan, K., Janosov, M., and Barabási, A.-L. (2022). Quantifying nft-driven networks in crypto art. *Scientific reports*, 12(1):2769.
- [Wlasinsky 2023] Wlasinsky, O. (2023). Literature review on the most popular of nfts types. *International Journal of Educational Technology and Artificial Intelligence*, 2(1):8–12.
- [Wood 2014] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. *Ethereum project yellow paper*, 151:1–32.
- [Yilmaz et al. 2023] Yilmaz, T., Sagfossen, S., and Velasco, C. (2023). What makes nfts valuable to consumers? perceived value drivers associated with nfts liking, purchasing, and holding. *Journal of Business Research*, 165:114056.