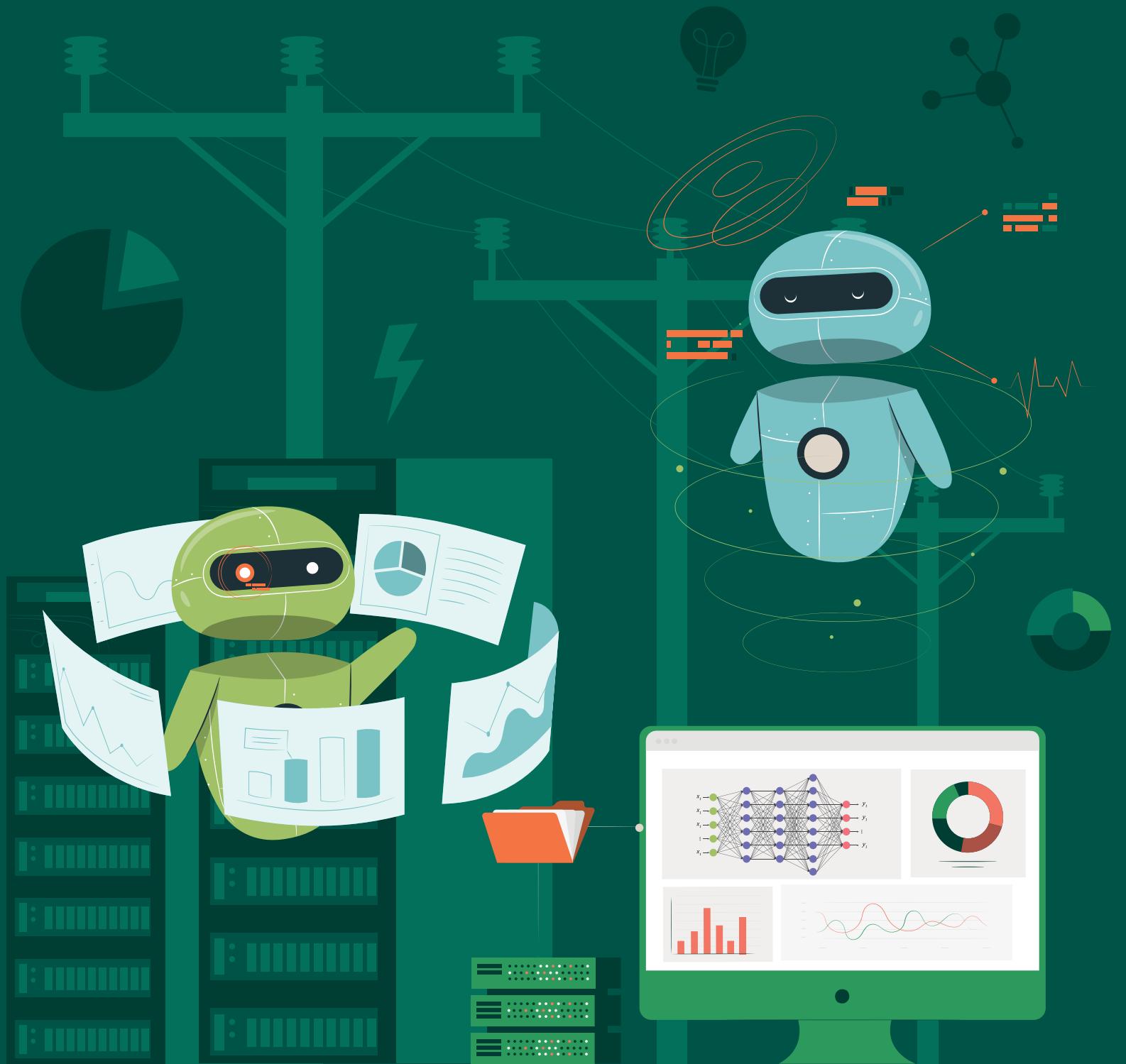


Machine Learning e Big Data no Setor Elétrico

Guia de conteúdo para docentes





Publicado por

Profissionais do Futuro: Competências para a Economia Verde

**Ministério da Educação
(MEC)**

Ministro
Camilo Santana

**Secretário de Educação Profissional e
Tecnológica**
Getúlio Marques

**Deutsche Gesellschaft für Internationale
Zusammenarbeit (GIZ) GmbH**

Diretor Nacional
Michael Rosenauer

**Diretor de Energias Renováveis e Eficiência
Energética**
Johannes Kissel

Diretora do Projeto Profissionais do Futuro
Julia Giebelner Santos

Coordenação e revisão técnica da publicação

Caroline Luciane Broering Dutra – GIZ
Roberta Hessmann Knopki – GIZ

Autor

Gabriel Costa
Marcos Izumida (coordenação geral dos autores)

Coordenação da diagramação

Lucas Tolentino – GIZ

Projeto gráfico e diagramação
João Bosco Gouvea Ramos

Revisão e edição de texto

Ismael Danilo Lima Freitas – GIZ
Luis Felipe Jaber – GIZ

AVISO LEGAL O conteúdo deste ebook reflete apenas a opinião do autor. A GIZ não é responsável pelo uso que possa ser feito das informações nele contidas. Ele foi elaborado apenas para fins didáticos, distribuído de maneira gratuita, sendo expressamente proibida sua comercialização. É vedada a reprodução total ou parcial deste material, por qualquer meio ou processo, sem autorização expressa da GIZ. Conteúdos visuais e textuais, quando de terceiros estão devidamente creditados e mencionados citando fontes e créditos. A violação de direitos autorais constitui crime (Código Penal, art. 184 e §§, e Lei nº 10.695, de 1º/07/2003), sujeitando-se a busca e apreensão e indenizações diversas (Lei nº 9.610/98).

Brasil, janeiro de 2023

Sumário

Abertura	6	
Objetivos	8	
Capítulo 1: Evolução do Setor Elétrico: <i>smart grids</i> e suas tecnologias	10	
1.1. Introdução aos conceitos de digitalização do setor elétrico	10	
1.2. O conceito de <i>smart grid</i>	16	
1.3. As principais tecnologias voltadas para as <i>smart grids</i>	17	
1.3.1. Medidores Inteligentes	17	
1.3.2. Dispositivos IoT	20	
1.3.3. Big data	21	
1.3.4. Machine learning	24	
1.3.5. Computação em nuvem	26	
1.3.6. Digital twin	28	
1.4. Especificação de requisitos de sistemas de monitoramento	31	
1.5. Recapitulando	32	
Capítulo 2: Armazenamento, processamento e gestão de dados	34	
2.1. Aspectos regulatórios de dados relacionados às <i>smart grids</i>	34	
2.2. Aspectos éticos do uso e manipulação dos dados pessoais	37	
2.3. Certificação de segurança para hardware	39	
2.4. Armazenamento de dados	41	
2.5. Arquiteturas processamento, otimização e recuperação de banco de dados	45	
2.6. Externalização de serviços de armazenamento	52	
2.7. Recapitulando	54	
Capítulo 3: Tratamento de dados e <i>machine learning</i>	56	
3.1. Tratamento de dados	56	
3.2. <i>Machine learning</i> : Modelos de redes neurais artificiais	61	
3.3. <i>Machine learning</i> : Modelos de redes neurais profundas	64	
3.4. <i>Machine learning</i> : Classificação de base de dados	66	
3.5. Recapitulando	74	
Capítulo 4: <i>Machine learning</i> e <i>big data</i> aplicados ao setor elétrico	76	
4.1. <i>Big data</i> no setor elétrico: Manipulação de dados para tomadas de decisão	76	
4.2. <i>Machine learning</i> no setor elétrico: Manutenção preditiva e sistemas de previsão	81	
4.3. <i>Machine learning</i> aplicado à modelagem da geração de energia	84	
4.4. Recapitulando	90	
Casos de sucesso	92	
Entidades a consultar sobre o tema <i>machine learning</i> e <i>big data</i> no setor elétrico	96	
Fechamento	97	
Glossário	98	
Referências	100	
Respostas das avaliações	102	



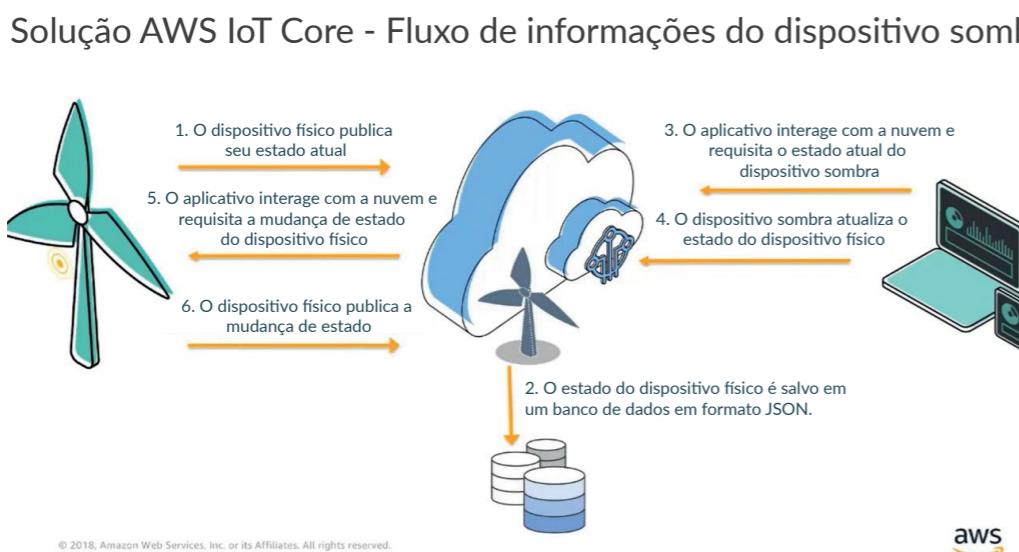
Abertura

A onda de *machine learning* e *big data* veio para perdurar. Em todas as áreas e setores, é possível usufruir de bons resultados de projetos que lidam com estes conceitos. Este ebook traz uma visão global destas tecnologias e onde elas podem ser empregadas no setor elétrico.

Existem muitos esforços voltados para criar e aprimorar tecnologias que possam ser aplicadas no setor elétrico, visando o aumento do controle e da supervisão de seus processos e ativos, a fim de melhorar a disponibilidade, confiabilidade e eficiência. Como exemplo, pode-se citar os produtos da Amazon Web Services (AWS), famosa empresa por prestar serviços de *Cloud Computing*, a qual investe constantemente em novas soluções visando modularizar e facilitar a integração de serviços de *IoT* (*Internet of Things* - internet das coisas).

No Tech Talk “How to Easily and Securely Connect Devices to AWS IoT” (Como conectar fácil e seguramente dispositivos em AWS IoT) da Amazon (link disponível na nota de rodapé), é apresentada uma solução IoT que pode ser aplicada à geração renovável de energia, conforme Figura 1.

Figura 1. Solução da AWS para o setor elétrico¹



Os processos que acontecem na Figura 1, podem ser entendidos como:

1. O dispositivo de campo notifica o seu estado atual para o sistema em nuvem. Este procedimento acontece de forma automática e serve para gerar informações periódicas e em tempo real sobre as condições operacionais da planta.
2. Os dados são armazenados em formato JSON (arquitetura simples que armazena os dados na estrutura {chave:valor}) em um banco de dados. O formato JSON é o mais utilizado pois, comumente, o mapa de medições de soluções *IoT* são de diferentes tamanhos e apresentam inúmeras características. Por isso, prefere-se adotar soluções de armazenamento de informação com estrutura volátil.

Exemplo de um JSON:

```
{  
    "dados_turbina": {  
        "velocidade": "38rpm",  
        "tensão_saída": "380V",  
        "corrente_saída": "92A",  
        "data": "08/12/2021"  
        "hora": "07:54"  
    }  
}
```

3. O aplicativo, que faz interface com o usuário, requisita os dados atualizados dos dispositivos em campo e os disponibiliza de forma automática para a visualização.
4. O aplicativo, então, requisita que comandos sejam enviados para os dispositivos em campo. Esta etapa é atrelada à possibilidade de controlar remotamente os parâmetros de operação da planta ou modificar configurações dos dispositivos *IoT*.
5. O modelo computacional em nuvem, o qual faz o mapeamento dos parâmetros operacionais reais para o mundo digital, requisita a atualização dos dados. Este modelo atua como ponte de comunicação entre todos os processos, ou seja, ele é capaz de se comunicar com o aplicativo e com o banco de dados, de interpretar os dados de medição e de enviar comandos para a planta.
6. O dispositivo de campo notifica o seu estado atual para o sistema em nuvem devido à requisição do usuário. Esta função é importante porque permite ao usuário ver os dados mais atuais que se tem sobre a planta, uma vez que as atualizações são feitas de forma periódica, evitando atrasos entre os valores medidos e os mostrados na interface.

As etapas supracitadas são utilizadas atualmente em algumas áreas do setor elétrico, porém com tecnologias que não permitem que o centro de supervisão e controle fique afastado da planta em operação. Com o uso da tecnologia *IoT* e *Cloud Computing*, é possível instalar centros de supervisão e controle fisicamente muito distantes das plantas de operação, permitindo até mesmo que várias plantas sejam operadas de um mesmo lugar.

À luz desse exemplo, este ebook traz conteúdos relevantes aos/as docentes de disciplinas sobre *machine learning* e *big data* no setor elétrico. São abordados temas como processamento de dados, otimização de busca, programação dinâmica e gestão eficiente de dados, com o intuito elencar algumas soluções importantes utilizadas na atualidade e quais são suas principais aplicações no setor elétrico. Além disso, este ebook aborda também, conceitos importantes como a Lei Geral de Proteção de Dados (LGPD) do Brasil e a Regulação Geral de Proteção de Dados da Europa (*Europe General Data Protection Regulation - EUGDPR*), leis que visam garantir a proteção de dados de pessoas físicas e jurídicas. Este assunto se tornou bastante relevante devido à digitalização dos sistemas e ao aumento da acessibilidade à internet. Cada tópico disponibilizado ao longo dos capítulos aborda os conceitos principais de cada assunto, bem como apresenta sugestões de referências, artigos e notícias que podem auxiliar o/a professor/a que ministra disciplinas nessa temática a se aprofundar um pouco mais sobre cada conteúdo.

1. <https://www.youtube.com/watch?v=FADeK-X8DnQ>



Objetivos

Entende-se a importância de digitalizar e evoluir as tecnologias empregadas no setor elétrico, uma vez que este é a base das atividades econômicas de uma nação. Portanto, este ebook tem como finalidade auxiliar e orientar o/a professor/a acerca dos conteúdos relacionados à *machine learning* e *big data* no setor elétrico. Para isso, o ebook traz conteúdos que servirão de insumo para que os/as professores/as possam, junto com seus/suas alunos/as, alcançar os objetivos de aprendizagem a seguir:

- **Abordar o conceito de digitalização do setor elétrico:** para que os/as alunos/as compreendam como a digitalização pode ajudar na redução dos custos de operação e manutenção, na redução de riscos na atuação de profissionais, na prevenção de falhas e na extensão da vida útil de equipamentos.
- **Abordar a regulação nacional relacionada à digitalização:** para que os/as discentes entendam as principais abordagens da regulação, de forma que possam compreender os desafios e os impasses que a regulação brasileira coloca sobre o desenvolvimento e a aplicação de novas tecnologias no setor, principalmente quando relacionadas ao consumidor cativo.
- **Abordar a LGPD e a EUGDPR:** conscientizando os/as discentes sobre os principais conceitos de segurança cibernética abordados pelas duas leis (brasileira e europeia), que visam resguardar os dados de pessoas físicas e jurídicas sob controle de terceiros.
- **Abordar os conceitos de machine learning e digital twin:** para que os/as alunos/as entendam as principais características dos algoritmos de *machine learning* e os principais campos de aplicação para cada modelo e compreendam também as vantagens de se possuir um modelo *digital twin* de um processo ou equipamento.
- **Abordar os principais algoritmos utilizados na classificação de base de dados:** para que os/as discentes compreendam as principais características dos algoritmos de *machine learning* que são utilizados para classificar preliminarmente uma base de dados extensa, visando facilitar o processo de extração de informação útil.
- **Abordar conceitos de processamento e otimização de busca em base de dados:** evidenciando para os/as alunos/as quais são os principais conceitos de um banco de dados, os tipos mais comuns e os seus mecanismos de funcionamento.
- **Abordar conceitos de execução distribuída e paralela em base de dados:** para que os/as discentes compreendam vantagens e desvantagens (*complexidade versus tempo de execução*) de se processar dados de maneira descentralizada.
- **Abordar as principais técnicas de tratamentos de dados:** dessa forma, os/as alunos/as podem compreender as principais técnicas para tratamento de *outliers* (também conhecidos como pontos fora da curva, aqueles que prejudicam o desempenho do modelo), dados faltantes de qualquer natureza e recuperação de dados.
- **Abordar o conceito de externalização de serviços em nuvem:** os/as alunos/as poderão conhecer os principais serviços e vantagens da computação em nuvem e serão capazes de justificar a importância destes serviços e de apontar a necessidade de seu uso no setor elétrico.
- **Abordar o conceito de hardware certificado em concordância com cibersegurança:** apresentar para os/as discentes um conceito sobre cibersegurança focado na segurança nativa em *hardware* e não somente a nível de *software*, como é amplamente difundida.



Capítulo 1: Evolução do Setor Elétrico: *smart grids* e suas tecnologias

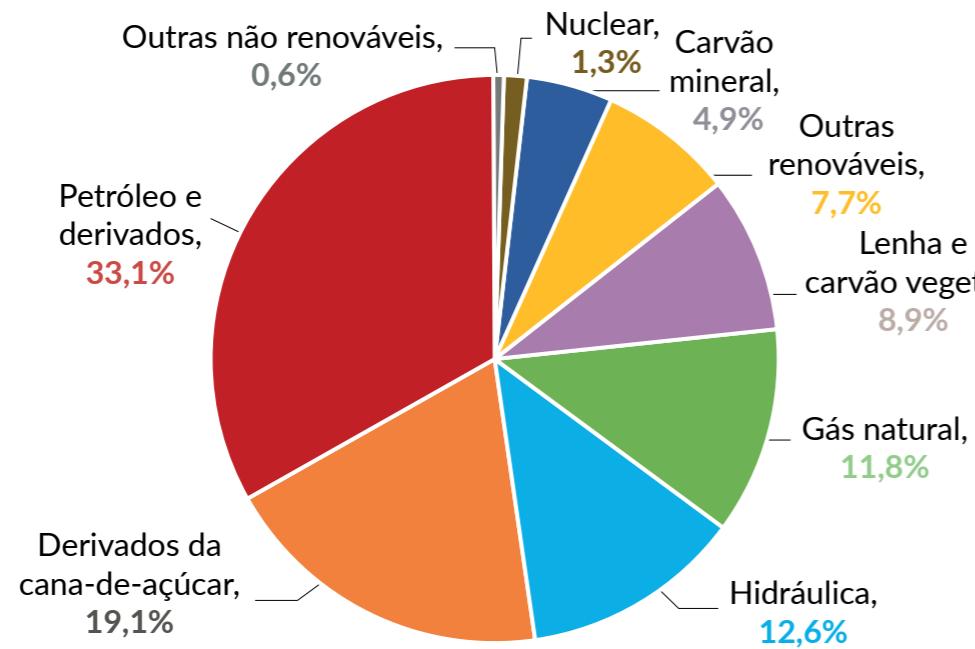


Nota para o/a professor/a – Este capítulo deve servir como base teórica para que os/as alunos/as compreendam os principais conceitos e tecnologias que regem o avanço da digitalização do setor elétrico. Desta forma, os/as discentes serão capazes de compreender a importância da digitalização do setor elétrico, as soluções e os grandes desafios que a implementação desta abordagem digital traz.

1.1. Introdução aos conceitos de digitalização do setor elétrico

A energia consumida por um país é gerada por todas as fontes de energia disponíveis em sua matriz energética. A matriz energética leva em consideração todos os tipos de fontes de energia, como combustíveis (fosseis ou não) que podem ser utilizados para movimentar automóveis e aeronaves, insumos como gás e carvão que podem ser utilizados em processos de aquecimento e energia elétrica que é utilizada em uma vasta quantidade de produtos e processos. A Figura 2 representa a matriz energética brasileira.

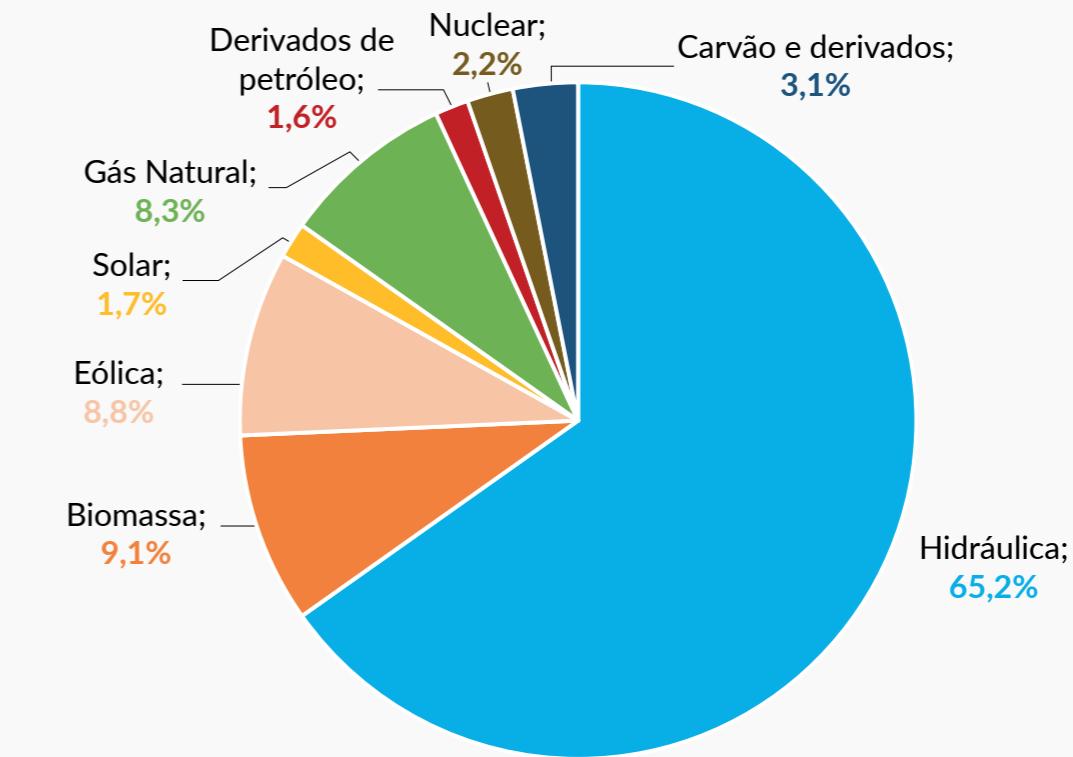
Figura 2. Matriz energética brasileira em 2020²



2. <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>

É importante ressaltar que o setor elétrico faz parte do setor energético, mas não são a mesma coisa. Enquanto o setor energético engloba todos os tipos de fontes de energia, o setor elétrico abrange apenas as fontes estão relacionadas com a energia elétrica. Neste sentido, a Figura 3 mostra a composição da matriz elétrica do Brasil.

Figura 3. Matriz elétrica brasileira em 2020³



Uma outra característica do setor elétrico é sua relação forte com a saúde econômica de um país: se a economia está saudável e crescendo, mais energia elétrica é necessária. Caso contrário, menos energia elétrica é necessária.



Saiba mais: Professor/a, existe uma estreita relação entre o crescimento econômico e o consumo de energia elétrica. O consumo de energia elétrica é um importante indicador de desenvolvimento de uma nação, podendo além do crescimento econômico, indicar a melhoria da qualidade de vida da população.

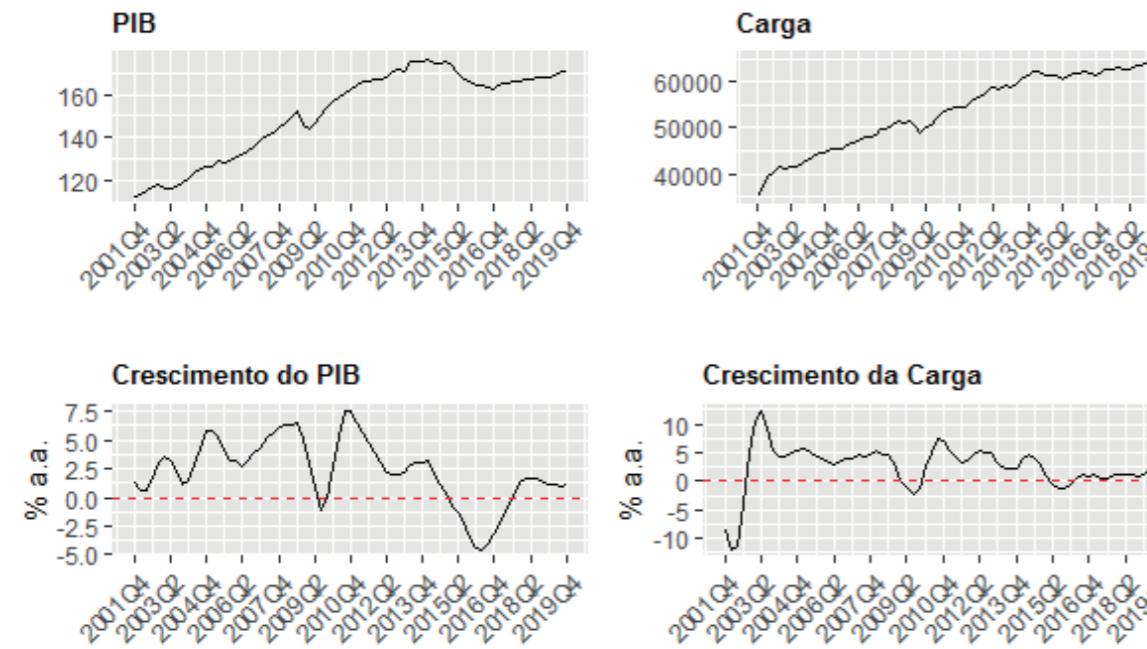
<https://www.publicacoeseventos.unijui.edu.br/index.php/salaconhecimento/article/download/7237/6007>

O comportamento supracitado pode ser observado no gráfico da Figura 4 que faz um comparativo entre os valores históricos da carga elétrica e do PIB no período de 2001 a 2019.

3. <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>



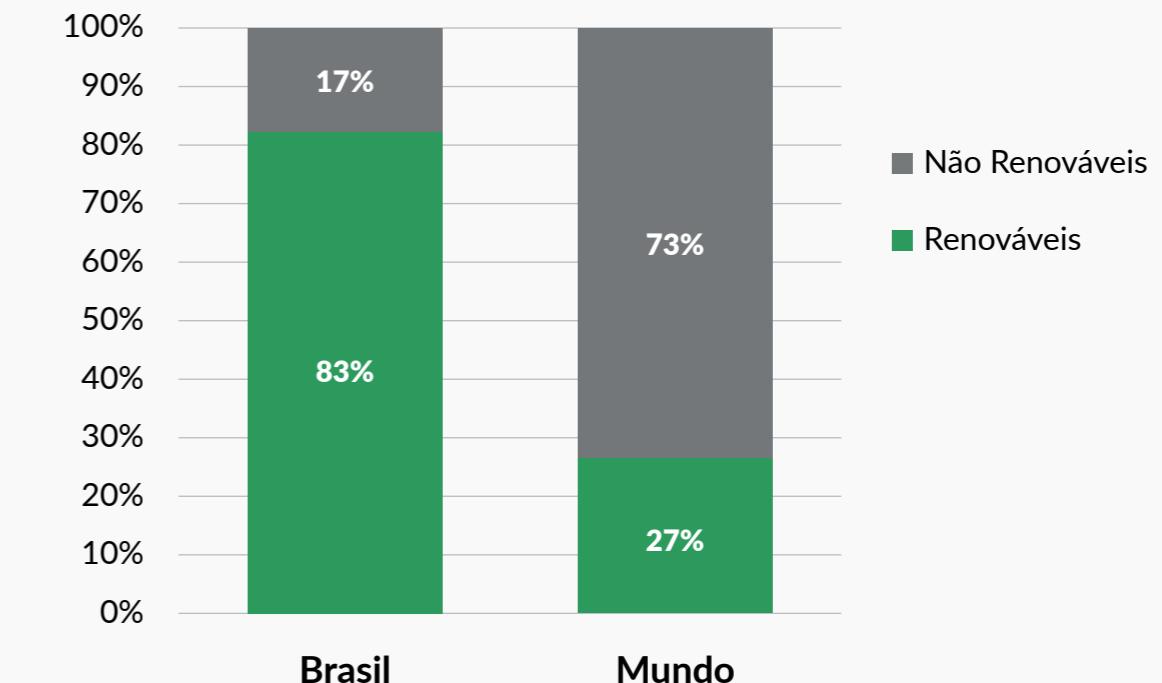
Figura 4. Comparativo da Variação do PIB com a Variação da Carga⁴



Ao fazer um comparativo entre os gráficos PIB e Carga da Figura 4, pode-se observar que no período analisado de 2001-Q4 a 2019-Q4 existe uma tendência crescente em ambos, o que indica uma relação forte entre a economia e o consumo de energia elétrica. Esta relação é comprovada, ao se analisar os gráficos de crescimentos anuais em porcentuais (Crescimento do PIB e Crescimento da Carga). Pode-se observar que no início de 2001-Q4 ambos os gráficos apresentam a mesma característica de queda, seguida de um crescimento que se estende até meados de 2003-Q2. Como ambos os gráficos estão acima da linha de 0.0%, entende-se que houve uma desaceleração seguida de uma aceleração no crescimento do PIB e da carga. Um outro comportamento que pode ser observado é o pico negativo que acontece entre 2015-Q2 e 2016-Q4. Neste caso, ocorreu uma regressão do PIB e da Carga.

Neste sentido, a evolução do setor elétrico está centrada em três pilares: a digitalização, a descentralização e a descarbonização. O setor elétrico brasileiro tem evoluído bastante no sentido de digitalização e descentralização. Em relação a descarbonização, a matriz elétrica brasileira sempre foi uma das mais limpas do mundo, como pode ser observado na Figura 5.

Figura 5. Comparativo da matriz elétrica brasileira e mundial⁵



Saiba mais: Apesar da matriz elétrica brasileira ser uma das mais verdes do mundo, isso não impede que ações de incentivo de implantação de fontes de energia renováveis aconteçam. Com a crescente preocupação sobre o controle da emissão de gases de efeito estufa, a implantação de fontes de energia renováveis (como hidráulica, solar e eólica) tem sido incentivada no Brasil.

Professor/a, o projeto de Lei 254/21 estabelece que as pequenas centrais de energia que usam fontes renováveis terão desconto mínimo de 50% ou 70% (vai variar conforme o porte) das tarifas de uso das redes de transmissão e distribuição. É importante voltar a atenção dos/das alunos/as para este ponto.

<https://hsm.utimaco.com/blog/what-are-the-common-criteria-for-hardware-security-modules-hsms/>

Uma das principais preocupações do setor elétrico durante os processos de digitalização, descentralização e descarbonização é prezar pela disponibilidade, qualidade e procedência da energia. A disponibilidade está ligada ao fato de que a falta de energia pode afetar negativamente a economia, uma vez que afeta a cadeia produtiva e o comércio. A qualidade precisa ser garantida para evitar depreciação precoce de equipamentos no setor elétrico e na indústria, além de assegurar a eficiência de processos de geração, transmissão e distribuição de energia. A procedência está relacionada ao tipo de fonte da qual a energia foi gerada (renovável ou não).

4. <https://analisemacro.com.br/economia/comentario-de-conjuntura/investigando-co-integracao-entre-consumo-de-energia-e-crescimento-do-pib/>

5. <https://www.epe.gov.br/pt/abcdenergia/matriz-energetica-e-eletrica>



A digitalização é, na realidade, um conjunto de sistemas e tecnologias desenvolvidas para facilitar a gestão de dados do setor elétrico visando simplificar a complexidade de operar o sistema, garantindo a minimização de falhas e gestão eficiente de equipamentos e processos. A digitalização busca habilitar o aumento da vida útil de equipamentos através da adoção de tecnologias digitais de sensoriamento em tempo real, o que permite antecipação de diagnósticos de ocorrência de falhas remotamente. Desta forma, os indicadores de disponibilidade de uma rede elétrica podem ser melhorados.



Saiba mais: A manutenção preditiva é a principal aliada no processo de digitalização do setor elétrico, pois é responsável por indicar as condições reais de funcionamento das máquinas com base em dados que informam seu desgaste ou processo de degradação.

Professor/a, conscientize os/as alunos/as sobre a importância da manutenção preditiva no setor elétrico, pois é um assunto importante que deve ser abordado.

<https://hsm.utimaco.com/blog/what-are-the-common-criteria-for-hardware-security-modules-hsms/>

A descentralização é pautada na segmentação dos processos de produção e comercialização de energia, e está ligada à inserção de Recursos Energéticos Distribuídos (REDs) no setor elétrico. Descentralizar a produção e a comercialização tende a incentivar a geração de energia renovável através de painéis fotovoltaicos, turbinas eólicas e unidades de microgeração hidráulica por pequenos produtores, o que promove indiretamente a descarbonização, uma vez que a matriz energética tende a se tornar mais verde.



Dica: O termo verde (ou economia verde) no contexto do setor elétrico muitas vezes é utilizado para mencionar a evolução do setor, resguardando os direitos sociais e o meio ambiente. É importante explicar este termo para os/as alunos/as.

Seguem referências para o assunto:

<https://mundoeducacao.uol.com.br/geografia/economia-verde.htm>
<https://www.osetoreletrico.com.br/economia-verde/>

Uma outra motivação para a descentralização é o aumento da disponibilidade da energia elétrica. A alta disponibilidade da energia elétrica é um ponto indispensável, pois grande parte dos processos produtivos e econômicos dependem de eletricidade para funcionar. A descentralização e diversificação da matriz energética brasileira ganhou muita visibilidade após os blecautes programados que ocorreram em 2001, sendo que uma das motivações foi a escassez de chuvas que afetou diretamente a geração hidráulica. Ainda assim, a diversificação da matriz elétrica do Brasil é realizada de forma cadenciada, como indica o Plano Decenal de Expansão de Energia (PDE) para o ano de 2031, realizado pela Empresa de Pesquisa Energética (EPE). Segundo o PDE, estima-se que apenas 20% das unidades consumidoras possuirão tecnologias de geração distribuída até 2031.



Saiba mais: O racionamento de energia que aconteceu em 2001 afetou vários estados brasileiros. O governo estipulou metas a serem cumpridas que previam a redução do consumo e o encarecimento das contas de energia dos consumidores. Desta forma, foi possível garantir que os serviços essenciais não ficassem sem energia elétrica.

Professor/a, este evento foi muito discutido na época e, até hoje, o setor elétrico extrai conhecimento dele. Levar esta curiosidade para a aula pode ser interessante para o conhecimento dos/das alunos/as.

<https://economia.uol.com.br/faq/o-que-foi-o-apagao-de-2001-risco-racionamento-energia-eletrica.htm>

O principal desafio para a que a descentralização ocorra no Brasil está relacionado ao modelo de tarifação da energia elétrica atualmente vigente. Isto porque as tarifas são baseadas no consumo total de kWh e não possuem granularidade temporal e nem sinal locacional de geração (no sistema de distribuição). Isto dificulta a inserção adequada de REDs no setor elétrico, pois os micro e minigeneradores continuarão dependendo da infraestrutura existente e ainda não existe uma forma concreta de tarifação para este tipo de geração localizada no sistema de distribuição.

Existe também uma grande preocupação relacionada ao planejamento da expansão do setor elétrico, que é uma etapa onde estudos de crescimento de carga e geração são feitos com o intuito de preparar a infraestrutura da rede elétrica para comportar a demanda futura. A descentralização adiciona muita incerteza neste processo, e parte dessa incerteza está relacionada ao dinamismo e evolução dos preços dos equipamentos utilizados em geração distribuída.

Um outro grande desafio é coordenar os planejamentos de geração, transmissão e distribuição para que atendam os requisitos de segurança e confiabilidade com o menor custo possível. A coordenação tende a se tornar mais complexa quando existem REDs no sistema, pois a geração e consumo de energia vai depender de muitos fatores aleatórios e desconhecidos, como condições climáticas e a forma como os proprietários dos REDs controlam seus equipamentos.



Saiba mais: Os recursos energéticos distribuídos adicionam uma alta complexidade em várias etapas de planejamento e operações do setor elétrico. A EPE tem constantemente voltado esforços para compreender os impactos de REDs no planejamento energético.

Professor/a, a EPE é uma boa referência e traz muitos assuntos sobre REDs e o futuro do setor elétrico brasileiro. Notifique os/as alunos/as sobre esta referência.

<https://www.epe.gov.br/sites-pt/sala-de-imprensa/noticias/Documents/ND%20-%20Recursos%20Energ%C3%A9ticos%20Distribu%C3%ADdos.pdf>

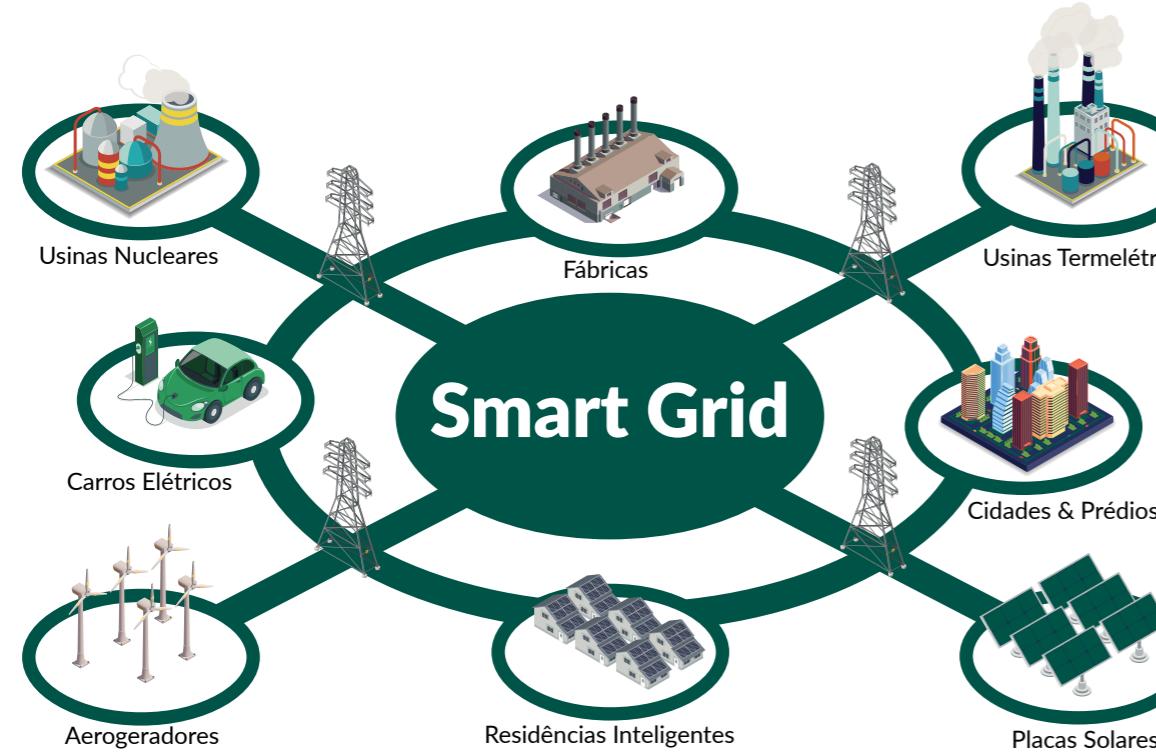


1.2. O conceito de *smart grid*

O Instituto Americano de Tecnologia e Padrões, NIST (*National Institute of Standards and Technology*), define *smart grid* (*SG*) como: “uma rede moderna que permite o fluxo bidirecional de energia, usando comunicação nos dois sentidos e técnicas de controle, que possibilitará novas funcionalidades e novas aplicações”. Neste sentido, uma *SG* deve ser interpretada mais como um conceito do que como uma tecnologia específica.

A *SG* deve possuir grau elevado de automação, supervisão e autonomia. As tecnologias aplicadas em uma *SG* visam assegurar um controle inteligente e otimizado de seus recursos energéticos distribuídos, buscando aumentar a disponibilidade e a qualidade da energia. Muitas vezes as *SGs* são atreladas aos conceitos de consumo, geração e gestão verde de energia, visando minimizar o impacto do consumo de energia no meio ambiente. A Figura 6 exemplifica o conceito de uma *smart grid*.

Figura 6. *smart grid*⁶



Atualmente existem poucas plantas no Brasil que se aproximam do que é apresentado na Figura 6. O que constantemente acontece é a presença de alguns elementos de uma *SG* dentro de uma rede de distribuição de energia elétrica, sendo os mais comuns as estações de recarga de veículos elétricos e sistemas de geração distribuída. Instalar centrais de armazenamento em bancos de baterias ainda é um grande desafio, tendo em vista o elevado custo de obtenção e manutenção dos equipamentos.

6. https://www.gta.ufrj.br/ensino/eel878/redes1-2016-1/16_1/smartgrid/



Saiba mais: Professor/a, o Brasil ainda está um pouco defasado do ponto de vista científico, intelectual e industrial, quando comparado com alguns outros países. Apesar disso, é importante alertar os/as alunos/as que o país caminha para a digitalização do setor elétrico.

<http://www.revistaespacios.com/a21v42n02/a21v42n02p13.pdf>



Futuro: Professor/a, existe uma tendência muito grande para implantação de novas tecnologias no setor elétrico, pois cada vez mais existem incentivos públicos e privados, visando aumentar a disponibilidade e confiabilidade do sistema elétrico, melhorando assim o seu desempenho. Além disso, a implantação de cargas não lineares e de microgeração distribuída tem inserido desafios para o setor, uma vez que ele foi projetado para atender unidades geradoras centralizadas de grande porte sem muitas interferências na rede. Aborde com os/as alunos/as assuntos que dizem respeito à *smart grids*, pois é importante eles terem conhecimento do poder de transformação que estas redes trarão para o setor elétrico.

Seguem referências para o assunto:

<https://osetoreletrico.com.br/cidades-inteligentes-o-futuro-do-smart-grid-no-brasil/>

<https://revistapesquisa.fapesp.br/cidades-do-futuro/>

1.3. As principais tecnologias voltadas para as *smart grids*

Para que exista uma interconexão entre os vários agentes dentro de uma *SG*, é necessário investimentos em arquiteturas que tornem viável a comunicação automática e bidirecional entre o consumidor de energia e a concessionária. Algumas tecnologias, como os medidores inteligentes e os dispositivos IoT, foram criados para cumprir este papel.

Além disso, o grau de automação exigido para que uma *SG* funcione requer que muitos dispositivos de medição e ferramentas de monitoramento sejam implantados, o que tende a gerar uma grande quantidade de dados que precisam ser processados e que posteriormente podem ser utilizados para gerar modelos de *digital twin* ou até mesmo alimentar algoritmos de *machine learning* com o objetivo de auxiliar em diagnósticos de manutenção preditiva.

1.3.1. Medidores Inteligentes

Os medidores inteligentes (Figura 7) são dispositivos com alto desempenho em processamento e transferência online de dados, que são instalados em pontos de medição de energia. Diferentemente dos medidores analógicos convencionais, os medidores inteligentes permitem o monitoramento remoto e em tempo real do ponto medido, e disponibilizam um conjunto de medições que vai além do consumo mensal de energia elétrica.



Com os medidores inteligentes, é possível capturar, armazenar e processar informações de composição harmônica do sinal, corrente, tensão e potência. As medições ficam disponíveis em valores instantâneos e valores eficazes (que é a raiz do valor quadrático médio ou RMS – root mean square em inglês). Ter acesso a este tipo de informação em tempo real permite que muitas técnicas de *machine learning*, otimização e modelos de previsão comportamental da rede elétrica sejam desenvolvidos. Apesar desse tipo de medidor já ser utilizado em aplicações específicas (principalmente em ambiente fabril), a utilização em larga escala apresenta muitas vantagens e desafios.

Figura 7. Medidor Inteligente⁷



Dentre as principais vantagens da utilização de medidores inteligentes, pode-se citar a eliminação da necessidade de leituras mensais presenciais, que atualmente são feitas por profissionais responsáveis por visitar e registrar a medição dos medidores dispostos em instalações de consumidores. Isto será possível uma vez que os dados de consumo estarão disponíveis em tempo real via protocolo de comunicação, o que possibilitará o acesso remoto às informações de todos os consumidores que possuírem este tipo de medidor.

A medição remota em tempo real possibilitará, também, a redução da incidência de erros de leitura, uma vez que poderá ser feita de forma automatizada e agilizará detecção da falta de energia. Além disso, ao passo que as regulamentações tarifárias forem evoluindo e o preço dinâmico de energia estiver disponível em uma granulação horária (ou até menor), será possível adotar técnicas que visem reduzir o gasto com energia.



Saiba mais: A concessionária deve estar preparada para lidar com a quantidade de dados que será gerada, não só pelos medidores inteligentes, mas também por toda a *smart grid*.

Professor/a, segundo pesquisa realizada pela Oracle, as implementações de SG têm aumentado exponencialmente o volume de dados. Neste sentido, alerte os/as alunos/as sobre a importância de saber tratar e manipular dados, pois estas serão habilidades cada vez mais requisitadas.

<https://ipnews.com.br/smart-grids-aumentam-em-180-vezes-a-coleta-de-dados-das-utilities/>

Apesar da adoção de medidores inteligentes trazer muitas vantagens, existem muitos desafios que precisam ser superados para que estes medidores possam ser adotados em larga escala, principalmente por consumidores residenciais. Um dos principais fatores que mais prejudicam a aceitação dos medidores inteligentes é a sensação de segurança ameaçada, uma vez que todos os dados de uma unidade consumidora são transmitidos via protocolo de comunicação. Um outro grande desafio é o custo de implementação para o consumidor final, quando o medidor não é cedido pela concessionária, e mesmo quando a concessionária arca com os custos do medidor, o tempo de *payback* é longo. O fato de que os medidores inteligentes necessitam de uma infraestrutura de comunicação para funcionar adequadamente também é um grande desafio a ser superado. Os investimentos em infraestrutura e conectividade tendem a ser altos e, por este motivo, devem ser adotados de forma cadenciada.



Futuro: Professor/a, os medidores inteligentes de energia prometem revolucionar o mercado de energia elétrica cativo. A tendência é que, após a implantação em massa, seja mais fácil prever o consumo de energia, o que tende a facilitar tanto a gestão interna da distribuidora quanto dar mais liberdade aos consumidores, permitindo que eles possam decidir como e quando consumir mais ou menos energia.

Alerte os/as discentes sobre a tendência de aumento da adesão deste tipo de medidor nos próximos anos.

Seguem referências para o assunto:

<https://atech.com.br/saiba-por-que-os-medidores-inteligentes-vao-reduzir-custos-na-distribuicao-de-energia-2/>

<https://www.yesenergysolutions.co.uk/advice/benefits-of-smart-meters>

7. https://static.weg.net/medias/h4b/h8c/MKT_WDC_METERS_SMW_1200Wx1200H.jpg



1.3.2. Dispositivos IoT

O conceito de internet das coisas (*Internet of Things*, do inglês - *IoT*) se refere ao processo de conectar objetos que já são utilizados comumente no dia a dia das pessoas à internet. Um termo bastante citado quando o assunto *IoT* entra em pauta é a comunicação máquina a máquina (do inglês *machine to machine* - *M2M*), que está relacionada com a troca de informações entre dispositivos de campo, que podem tomar decisões de forma autônoma. Neste sentido, a *IoT* se refere a qualquer sistema de dispositivos físicos que recebem e transferem dados por redes sem fio e que necessitam de pouca intervenção humana.



Dica: Existem diversos tipos de protocolos de comunicação para os dispositivos IoT. Cada protocolo apresenta vantagens e desvantagens em relação à custo, área de cobertura e latência. Sugere-se que o/a professor/a realize um trabalho de pesquisa com os/as alunos/as.

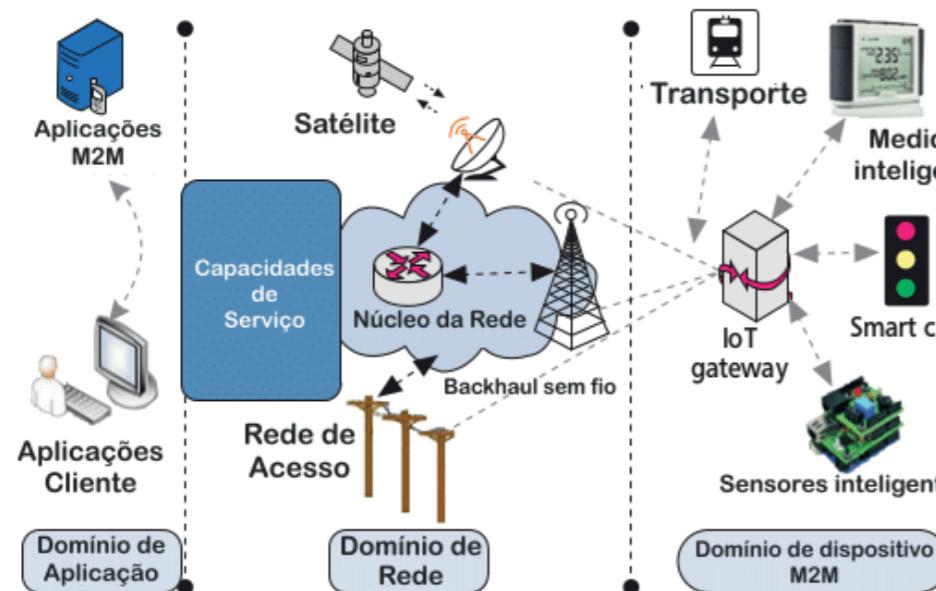
Seguem referências para o assunto:

<https://www.profissionaisti.com.br/10-protocolos-de-iot-que-voce-deveria-conhecer/>

<https://www.avsystem.com/blog/iot-protocols-and-standards/>

Os dispositivos *IoT*, no âmbito do setor elétrico, fazem parte de soluções de automação e supervisão, que integram de forma simples os sensores e atuadores de um processo às plataformas de supervisão e controle. Normalmente estas soluções fazem uso de computação robusta e de baixo custo em recursos físicos e concentram sua inteligência e processamento de dados na computação em nuvem. Uma estrutura básica de uma solução *IoT* é apresentada na Figura 8.

Figura 8. Arquitetura de Solução IoT⁸



8. https://www.researchgate.net/figure/Figura-23-Arquitetura-IoT-Adaptado-de-Hakiri-et-al-2015_fig2_324780227

Basicamente, a implantação de um sistema *IoT* pode ser dividida em 3 domínios:

- **Domínio de dispositivo M2M:** Neste domínio estão presentes dispositivos de medição e atuadores, responsáveis por coletar dados do processo e manipulá-lo fisicamente, de acordo com as informações trocadas entre os dispositivos, sem a necessidade de intervenção humana. Normalmente, os dispositivos de campo se comunicam através de um *gateway IoT*, o qual é responsável, também, por transferir os dados do processo físico para o domínio de rede.
- **Domínio de rede:** Neste domínio está presente toda a infraestrutura de comunicação. Esta camada é responsável por estabelecer a conectividade de diversas tecnologias e dar acesso à rede central de comunicação do processo.
- **Domínio de aplicação:** Este domínio inclui toda a inteligência e a infraestrutura de servidor/nuvem que é utilizada. É nesta camada que ocorre a análise de dados e que alguns modelos de *machine learning* são gerados para o processo. Aqui também estão presentes as aplicações que fazem interface com o usuário final, como sistemas de supervisão e controle manual.



Saiba mais: Professor/a, os dispositivos IoT prometem proporcionar mudanças disruptivas no setor elétrico, que vão desde ganho em eficiência operacional até novos modelos de negócio. Atente os/as alunos/as para este assunto.

<https://www.cpqd.com.br/internet-das-coisas-iot-backup/iot-habilitando-novas-formas-de-gerar-valor-no-setor-eletrico/>



Futuro: Professor/a, explique aos/as discentes que a adesão dos equipamentos IoT no sistema elétrico já é realidade. Muitas distribuidoras de energia estão utilizando dispositivos IoT para melhorar o monitoramento dos seus ativos de forma inteligente e eficaz. Com a chegada do 5G e o barateamento das novas tecnologias, a tendência é que os dispositivos IoT assumam mais tarefas de controle que dependem de performance e exatidão.

Seguem referências para o assunto:

<https://sealtelecom.com.br/integracao-tecnologia-5g-e-iot/>

<https://nae.global/introducao-ao-5g-o-grande-salto-para-a-internet-das-coisas/>

1.3.3. Big data

O aumento do volume de dados gerados devido à adesão de dispositivos inteligentes no setor elétrico permite que estudos de *Business Intelligence* (*BI*), modelos de manutenção preditiva e *machine learning* sejam construídos. Do ponto de vista de *BI*, uma base de dados rica pode ser utilizada para extrair informações úteis de consumidores, produtos, mercado e processos operacionais internos de uma indústria (como tempo de atendimento de reclamações dos clientes). O *BI* é muito comum em ambientes onde existe relacionamento direto com cliente, possibilitando encontrar possíveis gargalos no fornecimento de serviços e produtos ou até mesmo identificar novas tendências de mercado.



No ambiente fabril, apesar do *Business Intelligence* ser muito utilizado para gerar indicadores chave de desempenho (*Key Performance Indicators - KPIs*) do processo produtivo, é mais comum se deparar com outros conceitos como manutenção preditiva e *machine learning*. A manutenção preditiva é uma subárea da manutenção que visa, através de dados históricos, medições em tempo real e análise de sinais, prever defeitos em equipamentos que podem prejudicar o funcionamento do sistema no qual estão inseridos. *Machine learning* é um termo utilizado para se referenciar a qualquer algoritmo que seja capaz de se auto aperfeiçoar. Este tipo de algoritmo também necessita de um grande volume de dados para ser treinado e é comumente utilizado para prever comportamentos de um sistema.

O que os modelos e estudos supracitados têm em comum é a quantidade de dados necessária para que sejam assertivos. Normalmente, quanto mais rica em características e quanto mais registros estiverem disponíveis em uma base de dados, mais assertivos e detalhados serão os resultados. Para lidar com a vasta quantidade de dados, algumas técnicas de *big data* começam a aparecer.

Big data é um termo em Tecnologia da Informação que trata sobre grandes conjuntos de dados que precisam ser processados e armazenados, sendo embasado em 3 principais pilares, conhecidos como 3Vs (Figura 9).

Figura 9. Os 3Vs do big data⁹



- **Velocidade:** Relaciona-se com a frequência em que novos dados são gerados.
- **Volume:** Relaciona-se com número de dados efetivamente disponível, para consulta e processamento.
- **Variedade:** Relaciona-se com a estrutura e as características de cada dado gerado.

9. <https://solvimm.com/blog/o-que-e-big-data/>



Dica: O conceito de *big data* é relativamente novo (o termo surgiu em 1997) e está em constante evolução. Algumas referências abordam os pilares de *big data* como sendo os 5Vs: Velocidade, Volume, Variedade, Veracidade e Valor, ou até mesmo os 7Vs: 5Vs + Visualização e Volatilidade. Recomenda-se trabalhar também esta abordagem com os/as discentes.

Seguem referências para o assunto:

<https://www.cortex-intelligence.com/blog/inteligencia-de-mercado/os-5-vs-do-big-data>

<https://fia.com.br/blog/big-data/>

Normalmente, os dados coletados passam por um processo de tratamento e estruturação automatizado de dados que é conhecido como *pipeline* de dados (Figura 10). Um *pipeline* de dados é composto por uma série de softwares e ferramentas que são responsáveis por executar as etapas de engenharia, preparação e análise de dados.

Figura 10. Pipeline de dados¹⁰



Saiba mais: Professor/a, existem muitas ferramentas para lidar com *big data*. Estas tecnologias vão de banco de dados a ferramentas de processamento e tratamento de dados. Algumas ferramentas, como Apache Hadoop, Apache Spark e Apache Storm, são muito utilizadas. É importante que os/as alunos/as tomem conhecimento da existência destes tipos de ferramentas.

<https://tecnit.com.br/as-10-melhores-ferramentas-de-big-data-de-codigo-aberto-para-2020/>

As etapas de um *pipeline* de dados podem ser divididas em 3 macro etapas: Engenharia de dados, preparação de dados e análise de dados.

10. <http://blog.sudoers.com.br/o-que-e-pipeline-de-dados/>



- **Engenharia de dados:** é responsável pela coleta, remoção de dados inconsistentes e ingestão de dados em um *data lake*, que é um repositório de dados desestruturados (2.4).
- **Preparação de dados:** é onde ocorre o processamento dos dados em que se tem interesse. Nesta etapa é muito utilizado processamento de dados em paralelo (orquestração de vários clusters de processamento) para que o tempo de enriquecimento e preparação seja reduzido. Normalmente os dados “purificados” são injetados em um *data warehouse* (2.4).
- **Análise dos dados:** é onde são realizados estudos em que modelos de *machine learning* são alimentados com os dados estruturados e enriquecidos.



Futuro: Atualmente, no ambiente tecnológico pode-se dizer que só otimiza seus processos quem mantém um sistema de dados sólido e eficiente. A tendência é cada vez mais os bancos de dados (estruturados ou não) estarem presentes na indústria, permitindo estudos que gerem novas ideias sobre como melhorar um processo ou produto.

Professor/a, conscientize os/as alunos/as que a área de armazenamento e processamento de dados já é uma tendência e ajuda a entender e moldar o futuro de produtos e processos.

Seguem referências para o assunto:

<https://blog.indicium.tech/10-principais-tendencias-big-data-2021/>

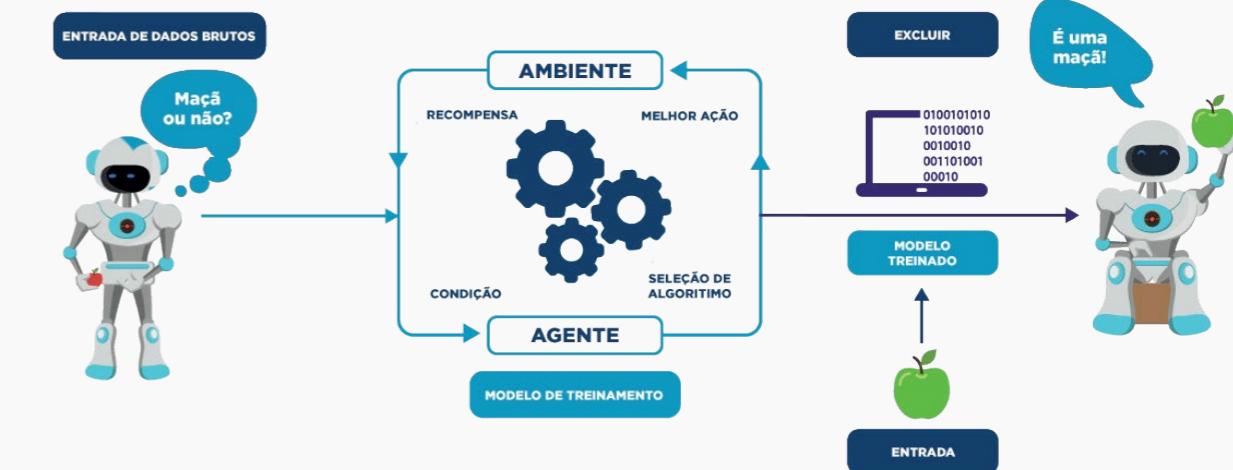
<https://venturebeat.com/2021/12/22/top-6-trends-in-data-and-analytics-for-2022/>

1.3.4. Machine learning

Machine learning (ML) é uma área da ciência da computação, sendo uma subárea da inteligência artificial, que tem como objetivo principal encontrar e entender padrões em dados utilizando algoritmos que são capazes de se automelhorarem com o tempo e com volume de dados. A grande vantagem de utilizar modelos de ML é que, muitas vezes, eles são capazes de identificar e descrever padrões que seres humanos não seriam. Uma arquitetura geral de um algoritmo de ML pode ser observada na Figura 11.

Os modelos de ML podem ser aplicados em diversas áreas e podem automatizar tarefas importantes dentro de um setor. É muito comum encontrar estes modelos sendo responsáveis por gerar análises em marketing, sistemas financeiros, sistemas de vendas e, principalmente, na indústria. Apesar de serem modelos eficientes e robustos, existe um grande desafio em dispor dados de qualidade para treiná-los, sendo esta a principal diferença entre os casos de sucesso e fracasso na aplicação do *machine learning* em algum setor.

Figura 11. Arquitetura genérica de um algoritmo de ML¹¹



Saiba mais: Professor/a, é importante explicar aos/as alunos/as que nem sempre ter um grande volume de dados indica que os modelos de machine learning serão bons. O importante é ter dados de qualidade, que sejam capazes de representar muito bem o problema de forma geral. Desta forma, as chances são de que o nível de acurácia de uma análise ou de um modelo aumente.

<https://www.supero.com.br/blog/qualidade-de-dados/>

A expansão e modernização do setor elétrico naturalmente traz o aumento do número de variáveis que podem ser medidas e controladas. Desta forma, os processos e equipamentos podem ser ajustados para trabalhar de forma otimizada visando melhores resultados e preservação do tempo de vida útil dos ativos. Deste modo, os algoritmos de ML têm ganhado cada vez mais espaço no setor elétrico para gerar modelos responsáveis por encontrar padrões de operação otimizados em ativos, supervisionar e identificar potenciais riscos de falhas nos equipamentos e até mesmo encontrar lacunas em processos, que antes, sem o uso do *machine learning*, eram muito complexas de se identificar.



Dica: Existem muitas aplicações para os algoritmos de machine learning no setor elétrico. Uma rápida pesquisa na internet possibilita encontrar estudos apontando o uso destes algoritmos na previsão da demanda, geração fotovoltaica, eficiência energética, perdas não técnicas e na manutenção preditiva de ativos críticos (como transformadores de subestação e geradores de usinas). Sugere-se ao/a professor/a realizar um trabalho de pesquisa com os/as alunos/as sobre estes temas.

Seguem referências para o assunto:

<https://www.foxiot.info/post/impactos-da-intelig%C3%A1ncia-artificial-no-setor-el%C3%A9trico>

<https://www.apd.pt/a-inteligencia-artificial-no-setor-energetico-e-as-suas-aplicacoes/>

<http://dspace.sti.ufcg.edu.br:8080/jspui/handle/riufcg/20761>

11. <https://solvimm.com/blog/o-que-e-machine-learning/>



Algoritmos de *machine learning* são muito utilizados para melhorar a eficiência energética de uma instalação. A Google (empresa conhecida mundialmente por seu sistema de busca na internet), tem voltado esforços para melhorar a eficiência energética de seus *data centers*, o que, além de colaborar com a preservação do meio ambiente, tem se mostrado uma grande estratégia de *marketing*, uma vez que ao contratar os serviços da Google, o usuário está automaticamente utilizando um serviço rotulado como “verde” (serviços que visam otimizar os recursos que utilizam da natureza, minimizando o impacto no meio ambiente).



Futuro: Professor/a, é importante notificar os/as alunos/as sobre como os algoritmos de machine learning podem ser utilizados para solucionar problemas em diversas áreas. A tendência é que a sua utilização seja cada vez maior nos setores estratégicos.

Em relação ao setor elétrico mundial, o World Economic Forum (fórum sueco que trata de assuntos globais, regionais e da indústria nos tópicos políticos, empresariais e culturais) realizou um estudo que aponta potencial de agregação de 1,3 trilhão de dólares em valor para o setor elétrico até 2025.

Seguem referências para o assunto:

<https://www.informa.com.br/transformacao-digital-e-industria-4-0-no-setor-elettrico-2/>

<http://reports.weforum.org/digital-transformation/electricity-an-industry-ready-for-digitization/>

1.3.5. Computação em nuvem

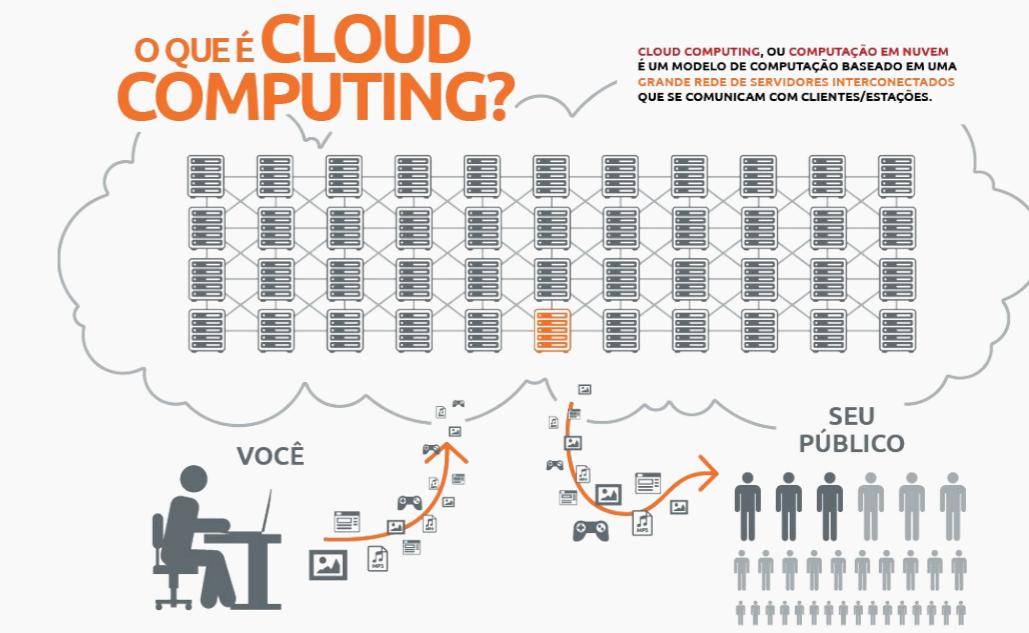
O avanço da tecnologia aumentou exponencialmente a disponibilidade de dados e, com isso, a capacidade de processá-los e gerar informação útil tornou-se uma tarefa complexa e custosa. Neste sentido, surgiram as soluções de computação em nuvem (ou *cloud computing* em inglês), visando mitigar a complexidade e o custo de lidar com o processamento de dados, além de disponibilizarem soluções que visam evitar que o cliente precise comprar hardware de alta performance, montar uma estrutura de rede física e dar manutenção em seus equipamentos.



Dica: Professor/a, segundo a Microsoft (empresa conhecida mundialmente por ter desenvolvido o sistema operacional de computadores Windows), a computação em nuvem é definida como sendo o fornecimento de serviços de computação, incluindo servidores, armazenamento, banco de dados, rede, software, análise e inteligência pela Internet (“a nuvem”) para oferecer inovações mais rápidas, recursos flexíveis e economias de escala.

A Figura 12 mostra como realmente funciona uma solução de computação em nuvem.

Figura 12. O que é Computação em nuvem?¹²



As empresas que fornecem os serviços de computação em nuvem, na verdade estão alugando seus servidores e softwares para que clientes de todo o mundo que possuam acesso à internet possam utilizar. Existem muitos fornecedores deste tipo de serviço, cita-se: Amazon, Google, Microsoft, VMware, Salesforce, Verizon, Cisco, AT&T e Citrix. A diferença entre cada fornecedor pode variar de valor à disponibilidade dos servidores em diversos países.



Saiba mais: Os principais fornecedores globais das tecnologias de computação em nuvem são: Amazon, Google e Microsoft. Cada fornecedor é especializado em algum tipo de tecnologia e utiliza estratégia de especificação diferente.

Professor/a, como o assunto de *cloud computing* é uma tendência, é importante deixar o conhecimento dos principais provedores do serviço com os/as alunos/as.

Seguem referências para o assunto:

<https://skyone.solutions/hub/nuvem/guerra-na-nuvem-aws-x-google-x-azure/>

<https://blog.saninternet.com/aws-vs-azure-vs-google>

A notícia publicada no site valor.globo.com, intitulada “Nuvem é o segmento com maior avanço previsto para 2021”, mostrou que os gastos com infraestrutura e plataformas de *cloud pública* teriam o potencial de atingir 3 bilhões de dólares em 2021. A Figura 13 exemplifica e elenca os principais objetivos e iniciativas das corporações nacionais em se utilizar a computação em nuvem.

12. <https://img.mandic.com.br/mandic-cloud/cloud-o-que-e-cloud-computing-infografico.gif>



Figura 13. Principais objetivos e iniciativas de TI para 2021¹³



A adoção em larga escala de dispositivos *IoT* e medidores inteligentes, fará com que o setor elétrico tenda cada vez mais a utilizar serviços de computação em nuvem. Desta forma, o armazenamento e processamento de informações pode ser mais bem gerenciado, e estará resguardado em um ambiente especializado e altamente disponível.



Futuro: Professor/a, a grande vantagem de utilizar cloud computing é usufruir de uma tecnologia de ponta sem precisar gastar uma elevada quantia para adquirir os equipamentos e nem precisar dar manutenção periódica no hardware. Um outro ponto que está sendo muito visado é que a computação em nuvem está cumprindo um papel importante na redução do consumo de energia e está ajudando os setores a se tornarem mais ecossustentáveis. Neste sentido, a adoção destes serviços é cada vez mais visada.

Seguem referências para o assunto:

<https://skyone.solutions/hub/como-a-cloud-diminui-a-demanda-de-energia-eletrica/>

<https://www.accenture.com/br-pt/insights/strategy/green-behind-cloud>

1.3.6. Digital twin

Digital twin (ou Gêmeo Digital, Figura 14) é uma representação tridimensional de algum objeto físico ou processo em um meio digital, utilizando computação gráfica e visão computacional, de forma que esta réplica seja completamente fiel e tenha o potencial de fornecer todas as perspectivas e dados importantes. Desta forma, é possível gerar diagnósticos sem que se precise ir à campo. Existem vários tipos de gêmeos digitais, dependendo do nível de ampliação do objeto. A maior diferença entre eles é a área de aplicação. É comum que diferentes tipos coexistam em um sistema ou processo.

13. <https://valor.globo.com/publicacoes/suplementos/noticia/2021/04/29/nuvem-e-segmento-com-maior-avanco-previsto-para-2021.ghtml>

- **Gêmeos componentes**— são a unidade básica dos *digital twins*, ou seja, o menor exemplo de um componente funcional.
- **Gêmeos ativos**— quando duas ou mais partes trabalham juntas, formam o que é conhecido como um ativo. Os *asset twins* possibilitam que você estude a interação desses componentes e uma grande variedade de dados de desempenho. Estes podem ser processados e, em seguida, transformados em percepções açãoáveis.
- **Gêmeos do sistema**— o próximo nível de ampliação envolve unidades gêmeas que permitem a visualização de diferentes ativos, que se juntam para formar um sistema maior em funcionamento. Os *unit twins* fornecem visibilidade sobre a interação de ativos e podem sugerir melhorias de desempenho.
- **Gêmeos de processo**— é o nível macro, que mostra como os sistemas podem trabalhar juntos para criar uma unidade de produção inteira. Os *process twins* podem ajudar a determinar os esquemas de tempo exatos que precisam ser sincronizados e influenciam a eficácia geral.

Figura 14. Aplicação de um digital twin¹⁴



No âmbito de uma SG, os seus ativos podem ter *digital twins* para facilitar o entendimento das suas condições operacionais, possibilitando prever ou encontrar falhas em tempo real. No que diz respeito às redes de transmissão e distribuição de energia, *digital twins* podem trazer uma série de ganhos financeiros para as operadoras do setor elétrico.



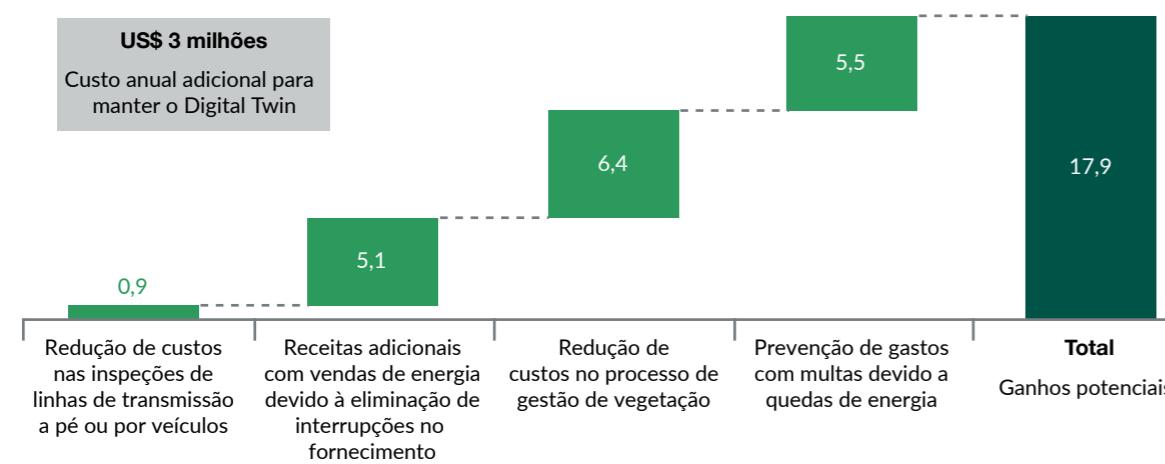
Saiba mais: A tecnologia do *digital twin* vem transformando o setor elétrico. É importante que o/a professor/a introduza aos/às discentes os papéis que as soluções atuais têm na gestão inteligente do inventário de ativos, no plano de manutenção e até mesmo na vegetação que pode entrar em contato com as linhas de transmissão de energia.

<https://www.pwc.com.br/pt/estudos/setores-atividades/energia/2019/digital-twin-19.pdf>



Com base em um estudo publicado pela PwC (empresa que se intitula como uma comunidade de solvers, responsável por solucionar problemas diversificados no ambiente empresarial), o setor elétrico norte-americano, especializado em linhas de transmissão/distribuição, tem um potencial de economia de 10% nas despesas operacionais após a implementação do *digital twin*. A Figura 15 resume o resultado do estudo e elenca os pontos encontrados em que a tecnologia foi responsável por reduzir os custos e gerar receitas adicionais.

Figura 15. Possíveis cortes de custos e ganhos na implementação de um digital twin¹⁵



Além da redução de gastos com multas por quedas de energia, é possível prever uma queda nos custos de inspeção de linhas de distribuição aéreas, uma vez que a tecnologia permite às empresas um controle melhor sobre sua infraestrutura, fazendo uso reduzido do quadro de funcionários. Uma outra vantagem de se utilizar um *digital twin* é o aumento da confiabilidade da rede, pois as falhas e defeitos podem ser antecipados. Um outro ponto positivo a ser citado é o aumento da segurança para os funcionários, pois estes serão expostos à ambientes perigosos com menos frequência.



Futuro: Professor/a, a revolução do digital twin promete reprogramar o mundo ao nosso redor, permitindo mais interatividade na relação com pessoas, produtos e empresas. É importante abordar o assunto com os/as alunos/as, pois prevê-se que na indústria, o uso de gêmeos digitais tem o potencial de melhorar 10% de sua eficiência, tornando-se uma tendência em investimentos.

Seguem referências para o assunto:

<https://mittechreview.com.br/a-revolucao-do-digital-twin/>

<https://www.gartner.com/smarterwithgartner/prepare-for-the-impact-of-digital-twins>

15. <https://www.pwc.com.br/pt/estudos/setores-atividades/energia/2019/digital-twin-19.pdf>

1.4. Especificação de requisitos de sistemas de monitoramento

Uma grande preocupação que aparece quando se pensa sobre a modernização do setor elétrico é a fraude de dados. A implementação das tecnologias citadas nos capítulos anteriores e a digitalização facilitam alguns aspectos operacionais do setor elétrico, mas abrem brechas para manipulação e acesso indevido de dados caso as devidas medidas de segurança não forem tomadas.



Dica: Professor/a, a ANEEL estabelece requisitos mínimos para sistemas de medição, sendo eles inteligentes ou não.

https://www2.aneel.gov.br/cedoc/aren2021956_2_4.pdf

O Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro) vislumbra sua atuação no desenvolvimento de *smart grids* na área de segurança do sistema, provendo requisitos e mecanismos que adicionem confiança e credibilidade a dados metrologicamente relevantes ou críticos que trafegam pela rede. Este processo de prover segurança engloba alguns aspectos, tais como:

- Definição do processo de validação dos softwares embarcados nos sistemas de medição que utilizem medidores inteligentes de energia elétrica;
- Definir técnicas de proteção de software para uma infraestrutura avançada de medição;
- Definir um programa de certificação dos equipamentos utilizados em *smart grids*;
- Definir um programa de certificação das redes de telecomunicações utilizadas pelas *smart grids*.

O Inmetro centraliza a especificação dos requisitos de segurança de todos os equipamentos responsáveis pela medição no setor elétrico. Neste sentido, nenhuma atualização de hardware e de firmware pode ir ao mercado sem ser pré-discutida com o Inmetro. Desta forma, o Inmetro garante a padronização da segurança das tecnologias e minimiza a geração de dados fraudulentos.



Saiba mais: É muito importante dizer aos/as discentes que, com a modificação do perfil das cargas e o crescimento no uso de energia eólica e fotovoltaica no país, o Inmetro apresentou um projeto sobre avaliação do desempenho de medidores de energia elétrica em um ambiente de geração distribuída.

<https://www.gov.br/inmetro/pt-br/centrais-de-conteudo/noticias/inmetro-desenvolve-projetos-para-setor-de-energia>



Saiba mais: Professor/a, mesmo com todas as camadas de proteção de dados e criptografia, sem a devida supervisão é muito complexo detectar fraude em dados. Atualmente, o Inmetro está desenvolvendo pesquisas relacionadas à análise de código, proteção de software, algoritmos criptográficos e protocolos de segurança, sempre com foco na aplicação em medidores inteligentes e outros dispositivos dotados de software embarcado.

<https://www.poli.usp.br/noticias/3203-gaes-e-inmetro-lancam-laboratorio-de-simulacao-antifraude-e-compliance.html>



1.5. Recapitulando

Ao final deste capítulo, o/a aluno/a deverá ser capaz de:

- Compreender o que são *smart grids* e qual a sua importância para a sociedade e o setor elétrico;
- Compreender a importância que os medidores inteligentes e os dispositivos *IoT* têm no processo de digitalização de uma rede;
- Compreender o conceito de *big data* e as etapas necessárias para lidar com dados que se enquadram nos critérios mínimos 3Vs;
- Compreender a importância que o *machine learning* desempenha no setor elétrico e o potencial que esta tecnologia tem para modificar a maneira que se lida com os processos e a gestão de ativos;
- Compreender o que e para que é utilizada a computação em nuvem e o motivo que este serviço tem se popularizado cada vez mais no setor elétrico;
- Compreender a importância de um *digital twin* e os benefícios que estes modelos têm agregado ao setor elétrico.

Atividades Sugeridas:

- Apresentações expositivas em slides com os temas: Digitalização do setor elétrico, medidores inteligentes e dispositivos *IoT*, soluções e desafios em *big data*, a importância do *machine learning*, computação em nuvem e *digital twin* para o setor elétrico.
- Apresentação de vídeos para consolidação do conceito e novidades relacionadas aos temas abordados em aula:
 - <https://www.youtube.com/watch?v=9Wq0Nm45YHU>
 - https://www.youtube.com/watch?v=CVGHE_9yJFA
 - <https://www.youtube.com/watch?v=2TXOZFmhGGo>
 - https://www.youtube.com/watch?v=77Y2mJX_xzA
 - <https://www.youtube.com/watch?v=97l0Ahu2efE>
- Promover atividades em grupo ou individuais, tais como seminários para avaliação do aprendizado. Sugestão de temas:
 - A razão da certificação do Inmetro ser tão importante no cenário nacional de energia elétrica.
 - O que o *machine learning* traz para o setor elétrico: Potenciais avanços e desafios.
 - Computação em nuvem: eficiência energética e meio ambiente.
- Convidar os/as alunos/as para participarem de webinar e outros eventos disponibilizados pela instituição ou fora dela, como forma de intensificar a sua capacitação.

Algumas fontes interessantes:

- [The Digitalization of Distribution Systems](#)
- [How digitalization brings new value to aging power infrastructure](#)
- [Digital Transformation of power systems](#)
- [Digital applications in implementation of smart grid](#)
- [The Effects of the smart grid System on the National Grids](#)
- [Smart metering in smart grid framework](#)
- [smart grid Research: Connected Homes](#)
- [The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0](#)
- [Research and Application of smart grid Early Warning Decision Platform Based on big data Analysis](#)
- [Ensuring future clean electrical energy supply through cloud computing](#)
- [The Design Concept of digital twin](#)

Avaliação:

De acordo com todo o conhecimento adquirido neste capítulo, fale sobre as principais características de uma *smart grid* e como que elas se relacionam com os conceitos de *IoT*, *cloud computing*, *big data*, *machine learning*, *digital twin* e manutenção preditiva.



Capítulo 2: Armazenamento, processamento e gestão de dados



Nota para o/a professor/a – Este capítulo deve servir como base teórica para que os/as alunos/as compreendam os principais conceitos de armazenamento, processamento e gestão de dados. Desta forma, os/as discentes serão capazes de compreender aspectos básicos das regulamentações voltadas para a gestão consciente de dados sensíveis, como e onde armazenar dados de operação e como tratar e manipular os dados para extrair conhecimento relevante deles.

2.1. Aspectos regulatórios de dados relacionados às *smart grids*

As *smart grids* prometem promover grandes melhorias para o setor elétrico nacional, mas existem grandes desafios para que todos os seus benefícios possam ser usufruídos pois a regulação atual possui baixa flexibilidade. Apesar disso, a Agência Nacional de Energia Elétrica (ANEEL) já prometeu prover mudanças para que as soluções mais arrojadas possam ser implantadas no sistema elétrico.



Saiba mais: Professor/a, em outubro de 2010 a ANEEL deu um grande passo em direção à digitalização do setor elétrico, obrigando as distribuidoras a substituir medidores eletromecânicos defeituosos por novos medidores eletrônicos. É importante notificar os/as alunos/as sobre esta decisão.

<http://g1.globo.com/economia-e-negocios/noticia/2010/09/aneel-defende-aparelho-eletronico-para-medir-energia.html>

Internacionalmente falando, existe um interesse maior de outros países em implantar novas tecnologias para modernizar o setor elétrico. Porém, mesmo com a legislação mais flexível, eles ainda enfrentam problemas com a falta de padrão regulatório, falta de incentivos, falta de engajamento da população e, principalmente, dificuldade de integração entre as tecnologias apresentadas pelo mercado.

O Operador Nacional do Sistema Elétrico (ONS) e a ANEEL são dois dos mais importantes órgãos reguladores presentes no setor energético brasileiro. Neste sentido, eles se preocupam em balizar a gestão de dados dos agentes do setor elétrico, buscando garantir integridade, transparência e consonância com as leis brasileiras de privacidade e segurança cibernética.

O ONS estabelece um conjunto de regras, as quais precisam ser aprovadas pela ANEEL, para atividades de coordenação e controle da operação da geração e transmissão de energia elétrica do Sistema Interligado Nacional (SIN). Este conjunto de regras é conhecido como Procedimentos de Rede. Todo agente do setor elétrico, interligado ao SIN, é responsável por cumprir todas as especificações ditadas em cada módulo dos Procedimentos de Rede.



Dica: Professor/a, em 8 de dezembro de 2020, a Agência Nacional de Energia Elétrica - ANEEL publicou a Resolução Normativa nº 903 que aprova a nova estrutura dos Procedimentos de Rede.

http://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/relacao_SM_reestruturados_antiga_estrutura.pdf

A estrutura dos procedimentos de rede é dividida em 9 partes:

1. **Relacionamento com Agentes:** Estabelece as responsabilidades do ONS e dos agentes envolvidos em relação aos produtos, prazos e etapas do processo de criação, identificação e tratamento das não conformidades.
2. **Critérios e Requisitos:** Estabelece as premissas, critérios e metodologias para estudos elétricos. Elenca os requisitos mínimos para os ativos e sistemas de medição, supervisão, teleproteção e comunicação.
3. **Planejamento da Operação:** Estabelece produtos, responsabilidades, prazos e etapas do processo de elaboração das propostas de ampliações, reforços e melhorias para o planejamento da operação elétrica de médio prazo. Uniformiza o tratamento, organização e atualização de dados que devem ser utilizados em modelos computacionais para subsidiar as atividades de estudos elétricos.
4. **Programação da Operação:** Estabelece os produtos, as responsabilidades, os prazos e as etapas dos processos relativos ao planejamento da operação elétrica com horizonte mensal e diário. Regulamenta a análise e tratamento dos dados hidroenergéticos e estipula a disponibilização de informações sobre restrições hidráulicas de usinas, acompanhamento e previsão climática.
5. **Operação do Sistema:** Estabelece as regras para o controle e operação da transmissão e geração, visando manter um funcionamento eficiente mesmo sob cenários de contingência. Elenca o procedimento para a recomposição da rede de operação e as regras para a operação teleassistida e a manutenção de serviços de comunicação.
6. **Avaliação da Operação:** Estabelece a coleta e apuração de dados da operação para a confecção de relatórios e estudos do setor elétrico. Elenca responsabilidades, prazos e etapas para a manutenção de ativos.
7. **Integração de Instalações:** Estabelece premissas e requisitos que as instalações que se conectam ao SIN devem possuir. Dita a implantação dos sistemas de supervisão e controle para a operação, como também os sistemas de faturamento.
8. **Administração dos Contratos e Contabilização Financeira:** Estabelece produtos, responsabilidades, prazos e etapas dos processos relativos à disponibilização de dados para o cálculo das Tarifas de Uso do Sistema de Transmissão. Dita como é feita a apuração de serviços e encargos de transmissão.
9. **Indicadores:** Estabelece como devem ser realizados os cálculos dos principais indicadores de desempenho e confiabilidade da Rede Básica e seus sistemas de supervisão e controle.



Saiba mais: É importante dizer aos/as alunos/as que os procedimentos de rede da ONS são documentos de domínio público e podem ser acessados por qualquer pessoa que se interesse. A consulta de qualquer módulo supracitado pode ser realizada no próprio site do Operador.

<http://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes>

Os submódulos que tratam de apuração de dados (conteúdo presente em maioria nos submódulos 6 e 9) definem responsabilidades tanto para o ONS quanto para os agentes que se interligam ao SIN. A Tabela 1 apresenta algumas responsabilidades definidas de modo geral pelos submódulos em relação aos dados de operação.

Tabela 1. Resumo das responsabilidades definidas em relação à dados operacionais e manutenção definidos nos procedimentos de rede.¹⁶

	Atividade	Responsável	Prazo	Periodicidade	Horizonte
1	Disponibilizar Informativo Preliminar Diário da Operação (IPDO) e Boletim Diário da Operação (BDO).	ONS	Definido no Submódulo 5.13	Diária	Dia anterior
2	Disponibilizar Boletim Semanal da Operação (BSO).	ONS	Definido no Submódulo 5.13	Semanal	Semana anterior
3	Disponibilizar Relatório de Comunicações de Ocorrência em Instalações de Geração	ONS	4 dias úteis após início da indisponibilidade	Quando existir ocorrência de evento	Período de indisponibilidade
4	Disponibilização aos agentes envolvidos dos dados apurados de geração e de indisponibilidades	ONS	Mês subsequente ao da apuração	Mensal	-
5	Solicitação ao ONS de alteração despacho térmico para eventos que já tenham sido consistidos	Agentes de geração	Até 90 dias após a data do evento	Sob demanda	-
6	Disponibilização à CCEE dos dados apurados de geração e de indisponibilidades	ONS	Conforme estabelecido no Acordo Operacional ONS/CCEE	Mensal	-
7	Disponibilização dos eventos classificados como ultrapassagem de Montante do uso de transmissão (MUST) aos agentes envolvidos	ONS	Até o 16º dia útil do mês subsequente ao da medição	Mensal	-
8	Consistência dos dados e eventos apurados pelo ONS referentes a ultrapassagens de MUST	Agentes envolvidos	Até 10 dias úteis após a Atividade	Mensal	-

9	Consistência dos eventos classificados como sobrecontratação de MUST	Agentes de distribuição	Até 1º de março	Anual	-
10	Disponibilização, no sistema computacional do ONS, dos planos de manutenção para equipamentos e linhas de transmissão da Rede Básica	Agente de transmissão	Entre agosto e novembro	Anual	-
11	Disponibilização dos indicadores, com a avaliação de desempenho, para os agentes e para a Aneel, no sistema computacional	ONS	10º dia útil	Mensal	Indicadores do mês anterior, conforme Submódulo 9.2
12	Elaboração do Plano anual de manutenção preventiva do sistema de medição (SMF) para faturamento relativo ao ano subsequente e envio aos agentes envolvidos	Agente responsável pelo SMF	Até dia 1º de outubro de cada ano	-	-
13	Envio ao ONS e aos agentes envolvidos do seu Plano anual de manutenção preventiva do SMF	Agente responsável pelo SMF	Até dia 1º de dezembro de cada ano	-	-

A preocupação com a gestão eficiente e segura de dados moveu a ANEEL a instituir o Plano de Dados Abertos, conhecido como PDA. O objetivo do PDA é dar ao público acesso a informações importantes sobre: Geração; Transmissão; Comercialização; Expansão da Oferta/ Consumo; Qualidade; Eficiência Energética; Pesquisa e Desenvolvimento (P&D); Mercado de Energia Elétrica; Tarifas; Políticas Públicas; Interação ANEEL-Sociedade;

Os dados que a ANEEL apresenta em seu PDA são obtidos através do SCDE (Sistema de Coleta de Dados de Energia) e da CCEE (Câmara de Comercialização de Energia Elétrica). Os dados do SCDE podem ser coletados ativamente ou não através de meio remoto, sendo que a CCEE se responsabiliza em coletar e garantir a qualidade e confiabilidade dos dados.

2.2. Aspectos éticos do uso e manipulação dos dados pessoais

Com a popularização da internet e o avanço da tecnologia, muitos dados ficaram disponíveis na internet, inclusive dados pessoais de pessoas físicas e jurídicas. Por isso, muitas organizações começaram a se movimentar, visando proteger os dados pessoais de acessos indevidos. Estas ações culminaram em leis e regras nacionais e internacionais que devem ser seguidas, para garantir a integridade de dados pessoais.

General Data Protection Regulation (GDPR) é um projeto voltado para a proteção de dados e identidade dos cidadãos da União Europeia. Como a União Europeia considera que a proteção dos dados pessoais é um direito dos países de seu bloco, todas as empresas que atuam dentro da região devem seguir as regras de coleta, processamento, compartilhamento e resguardo de dados pessoais.

16. Fonte: <http://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes>



As principais obrigações da GDPR são:

1. O serviço deverá permitir que o usuário escolha como os seus dados serão tratados e autorize ou não o uso.
2. O usuário tem direito de saber quais dados estão sendo coletados e para quais finalidades.
3. Deve haver meios para que o usuário solicite a exclusão de informações pessoais ou interrompa a coleta de dados, com a decisão devendo ser respeitada.
4. O usuário também pode acessar, solicitar cópia ou migrar dados coletados para outros serviços (quando cabível).
5. Uso de linguagem clara, concisa e transparente para que qualquer pessoa possa compreender comunicações sobre seus dados, inclusive termos de privacidade.
6. Em caso de incidentes que resultem em vazamento ou violação de dados que podem ferir direitos e a liberdade das pessoas, a organização deverá notificar autoridades em até 72 horas.
7. Aplicação da privacidade por design: a proteção dos dados deve ser considerada desde o início do projeto de um sistema, como parte imprescindível deste.
8. Recomendação de pseudonimização: quando cabível, é recomendável que a empresa proteja informações sensíveis ocultando-as ou substituindo-as de alguma forma para que a identificação do usuário só seja possível com a adição de outros dados.
9. As empresas terão, em certas circunstâncias, que trabalhar com um *Data Protection Officer* (DPO), executivo que deverá supervisionar o tratamento de dados pessoais, bem como prestar esclarecimentos ou se comunicar com autoridades sobre o assunto.

A LGPD (Lei Geral de Proteção de Dados) chegou ao Brasil oriunda de estudos realizados em legislação já existente na Europa, a chamada GDPR, que tem por objetivo regular de forma efetiva, clara e transparente o processamento de dados pessoais no ambiente corporativo, garantindo sua proteção. A LGPD (lei 13.709/18) entrou em vigor em 2020 e regula a coleta e o armazenamento de dados pessoais pelos setores público e privado. Esta lei visa proteger os direitos de liberdade, privacidade e o livre desenvolvimento da pessoa natural.



Dica: Professor/a, a LGPD é um assunto muito importante, empresas e instituições tiveram que realizar várias etapas de revisão em suas estratégias de gestão de dados pessoais. Disponibilize a referência abaixo para os/as discentes, pois ela expõe as etapas que uma empresa deve seguir no processo de revisão de suas estratégias de gestão de dados. É recomendado que o/a professor/a aborde em suas aulas os temas elencados no site da LGPD Brasil.

<https://www.lgpdbrasil.com.br/>

A LGPD prevê que as empresas que lidam com dados de pessoas físicas devem alinhar e garantir que suas práticas de governança de dados atendam ao mínimo as diretrizes:

1. Demonstrar o comprometimento do controlador em adotar processos e políticas internas que assegurem o cumprimento, de forma abrangente, de normas e boas práticas relativas à proteção de dados pessoais.
2. Ser aplicável a todo o conjunto de dados pessoais que estejam sob seu controle, independentemente do modo como se realizou sua coleta.
3. Ser adaptado à estrutura, à escala e ao volume de suas operações, bem como à sensibilidade dos dados tratados.

4. Estabelecer políticas e salvaguardas adequadas com base em processo de avaliação sistemática de impactos e riscos à privacidade.
5. Ter o objetivo de estabelecer relação de confiança com o titular, por meio de atuação transparente e que assegure mecanismos de participação do titular.
6. Estar integrado a sua estrutura geral de governança, estabelecendo e aplicando mecanismos de supervisão internos e externos.
7. Contar com planos de resposta a incidentes e remediação.
8. Ser atualizado constantemente, com base em informações obtidas a partir de monitoramento contínuo e avaliações periódicas.

No setor elétrico, os dados pessoais são coletados e armazenados principalmente pelas empresas distribuidoras de energia, mas podem ser compartilhados e trocados por toda a cadeia, incluindo os elos de geração, transmissão e comercialização. A aplicação da LGPD no setor elétrico é de extrema importância, visto que nele existem dados sensíveis de quase todos os brasileiros, pois a energia está presente em quase toda extensão territorial. Existe uma grande expectativa que a adequação da LGPD seja efetuada pelo setor elétrico em um efeito cascata, devido a exigência que surge dos órgãos regulamentadores.



Saiba mais: O risco de apropriação e usos indevidos de dados pessoais é real. No final do ano passado, a distribuidora Enel SP também sofreu um ataque hacker, que afetou cerca de 280 mil clientes da companhia, ou 4% do total de 7,2 milhões de consumidores atendidos por ela. Neste sentido, é muito importante conscientizar os/as alunos/as sobre a importância da LGPD no setor elétrico.

<https://www.uol.com.br/tilt/noticias/redacao/2020/11/10/enel-informa-que-dados-de-clientes-foram-vazados-em-osasco.htm>



Futuro: Professor/a, a chegada da LGPD e a evolução do conceito de cibersegurança ao longo dos anos fez o mercado de trabalho criar uma necessidade de profissionais que comprehendam e que sejam capazes de aplicar a LGPD em uma empresa ou instituição. Uma profissão que surgiu na Europa como consequência da GDPR e que deve ser tendência para os próximos anos é o *Data Protection Officer*, profissional responsável por ser guardião dos dados da empresa. Alerta os/as alunos/as sobre esta tendência.

<https://www.maisdados.com.br/solucoes-lgpd/a-lgpd-e-o-futuro-das-profissoes/>

<https://gdpr-info.eu/>

2.3. Certificação de segurança para hardware

A LGPD e a GDPR nasceram com o objetivo de resguardar os dados pessoais dos usuários de serviços na internet. É importante ressaltar que apenas essas leis não são responsáveis para garantir a segurança dos dados, principalmente em relação aos dados que podem comprometer a própria empresa. Neste sentido entram também os conceitos de cibersegurança, que é a prática de proteger ativos de informação (tais como sistemas, computadores, servidores, entre outros) contra ameaças cibernéticas ou ataques maliciosos.



Dica: Professor/a, é muito importante conscientizar os/as discentes que os crimes cibernéticos vêm de todas as formas e tamanhos. Existem muitos tipos de estratégias maliciosas para se apropriar de dados de terceiros. É indicado que o/a professor/a conscientize os/as alunos/as sobre os tipos de ciberataques existentes.

<https://br.claranet.com/blog/tudo-sobre-ciberseguranca>

É muito comum atrelar soluções de cibersegurança com softwares de criptografia e proteção de sistemas, porém o mercado também oferece soluções eficientes e arrojadas a nível de hardware, que podem ser utilizadas em conjunto com os softwares para garantir a segurança dos dados.

Devido à crescente cobrança sobre manter uma proteção adequada para os dados, o uso da criptografia popularizou nos últimos anos. A IBM foi a empresa pioneira no uso de encriptação de informação nativa em servidores, adicionando uma camada de segurança a mais e embarcada no próprio produto. Atualmente existem módulos de segurança de hardware (ou do inglês *Hardware Security Module* - HSM), que fornecem segurança extra para dados confidenciais utilizando a tecnologia de encriptação introduzida pela IBM.

Figura 16. Módulo de segurança em hardware¹⁷



Uma outra boa prática a nível corporativo que permite o uso de hardware especializado em segurança é adotar as *workstations* (computadores de alto desempenho, especificados sob medida e que permitem a utilização de hardware modularizado), pois estas permitem um controle maior na interação com o usuário, permitindo assim que ações maliciosas ou inapropriadas de acordo com a LGPD sejam interceptadas antes que problemas maiores aconteçam.



Futuro: Professor/a, em relação ao futuro pós-pandemia, uma pesquisa da Microsoft identificou que apenas 38% das Pequenas e Médias Empresas (PMEs) pretendem retornar ao local de trabalho físico com políticas flexíveis e 30% delas vão manter o trabalho remoto integralmente. Novos modelos de negócio, novas tecnologias e mudanças na estrutura e processo de trabalho demandam atenção redobrada com a segurança. Neste sentido, é importante conscientizar os/as alunos/as sobre tecnologias que podem ser embarcadas em computadores, permitindo autenticar a procedência do firmware durante a inicialização e criar uma cadeia de segurança com o sistema operacional.

<https://inforchannel.com.br/2021/12/15/pmes-devem-continuar-a-investir-em-tecnologia-e-seguranca-do-hardware-ao-usuario/>

2.4. Armazenamento de dados

A gestão eficiente de dados, com as ferramentas e práticas de análises corretas, pode ajudar as empresas a enxergar possibilidades e oportunidades que podem melhorar muito o processo interno e a lucratividade. Em resumo, existem 2 grandes áreas quando se trata de armazenamento de dados: armazenamento físico e banco de dados.

O armazenamento físico pode utilizar técnicas de armazenamento magnético ou de armazenamento em estado sólido. O armazenamento magnético é baseado na fragmentação de discos metálicos, onde cada fragmento pode ser magnetizado e desmagnetizado, representando números binários. Esta abordagem é utilizada ainda hoje, possuindo a vantagem de ser de baixo custo, porém de baixa confiabilidade, pois um campo magnético externo ou um impacto no dispositivo pode alterar o estado de magnetização dos fragmentos e corromper os dados. Já o armazenamento de estado sólido, conhecido popularmente como SSD, se baseia na eletrônica digital, fazendo seu armazenamento através de portas lógicas. O SSD é mais confiável, mais rápido e mais compacto que uma tecnologia de armazenamento magnético, porém possui um custo maior. As soluções citadas têm aplicações não só a nível doméstico, mas também industrial. A grande diferença é que, na indústria, os dispositivos de armazenamento são maiores, precisam de um sistema de supervisão e resfriamento e de um sistema de backups.



Dica: Existem muitos tipos de HDs e SSDs no mercado. Cada um tem aplicação específica, com diferentes sistemas de armazenamento, velocidade e padrões. É recomendado que o/a professor/a apresente estes conceitos para os/as alunos/as.

Seguem referências para o assunto:

<https://canaltech.com.br/hardware/saiba-a-diferenca-entre-hd-sata-e-hd-ata-e-os-detalhes-dos-discos-rigidos-68541/>

<https://www.promobit.com.br/blog/conheca-os-tipos-de-ssd-e-saiba-qual-a-melhor-opcao-para-o-seu-tipo-de-uso/>

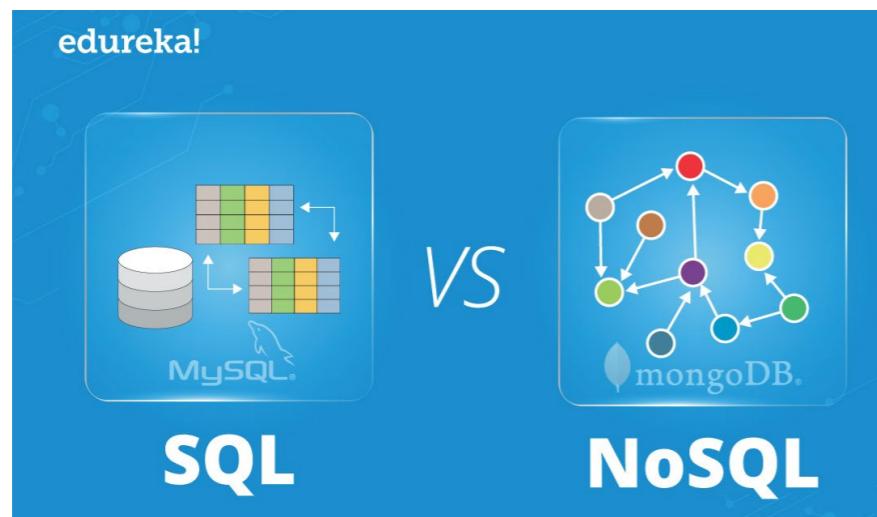
17. <https://www.evaltec.com.br/solucoes/hsm/>



Um banco de dados é um software que permite o registro de dados e informações nos HDs e SSDs. A nível comercial os bancos de dados são divididos em 2 grandes grupos: bancos SQL e NoSQL. O grupo de banco de dados do tipo SQL, são assim classificados pois utilizam a linguagem SQL (*Structured Query Language*) para salvar, manipular e recuperar dados. O grupo de banco de dados do tipo NoSQL (*No Structured Query Language*), são assim classificados pois não utilizam SQL para salvar, manipular e recuperar dados. Apesar disso, as linguagens NoSQL buscam mimetizar a sintaxe SQL, facilitando a interação do profissional com as ferramentas.

A diferença de estrutura entre os bancos do tipo SQL e NoSQL pode ser observada na Figura 17.

Figura 17. Tipos de bancos de dados¹⁸



Como dito anteriormente, os bancos de dados são softwares utilizados para facilitar no armazenamento, manipulação e recuperação de dados. Dessa forma, existem muitas soluções no mercado, sendo elas de código aberto e livres para uso ou de código fechado e adquiridas por assinaturas. A Figura 18 representa as soluções de mercado para banco de dados estruturados, ao passo que a Figura 19 representa os bancos de dados NoSQL.

Figura 18. Bancos de dados SQL¹⁹



18. <https://dev.to/sanchit170054/when-to-use-nosql-over-sql-4gg6>

19. <https://www.edureka.co/blog/sql-vs-nosql-db/>

Figura 19. Bancos de dados NoSQL²⁰



O tipo de banco de dados e a estrutura a ser escolhida para a aplicação em um projeto de gestão de dados depende das características que os dados possuem. Pode-se dizer que quando um banco de dados é alimentado por várias fontes, das quais as informações coletadas não podem ser armazenadas em formato de tabela (seja por estruturas diferentes entre os dados ou pelo volume de informações textuais), é recomendável utilizar bancos NoSQL, caso contrário, recomenda-se utilizar os bancos de dados SQL.



Saiba mais: Professor/a, é interessante dizer aos/as discentes que a escolha da solução a ser aplicada a um projeto é importante e demanda do profissional um conhecimento aprofundado do negócio no qual o banco de dados será utilizado. Escolhas ruins, podem ter impacto tanto no desenvolvimento quanto no resultado do projeto.

<https://www.altexsoft.com/blog/business/comparing-database-management-systems-mysql-postgresql-mssql-server-mongodb-elasticsearch-and-others/>

No ambiente corporativo, é muito comum existirem estruturas de armazenamento de dados mais sofisticadas. Existem dois tipos de estruturas principais que são utilizadas: *data lake* e *data warehouse*.

Data lake é um espaço com grande capacidade de armazenamento de dados. Ele é capaz de abranger todo tipo de informação, seja ela estruturada ou não-estruturada, em qualquer escala. Assim, ele permite que as empresas organizem seus dados a fim de obterem *insights* para tomada de decisão com rapidez e precisão. Neste tipo de arquitetura, diferentes profissionais podem ter acesso simultâneo aos dados brutos, porém os dados precisam ser tratados e organizados de forma que possam ser utilizados em estudos posteriores.

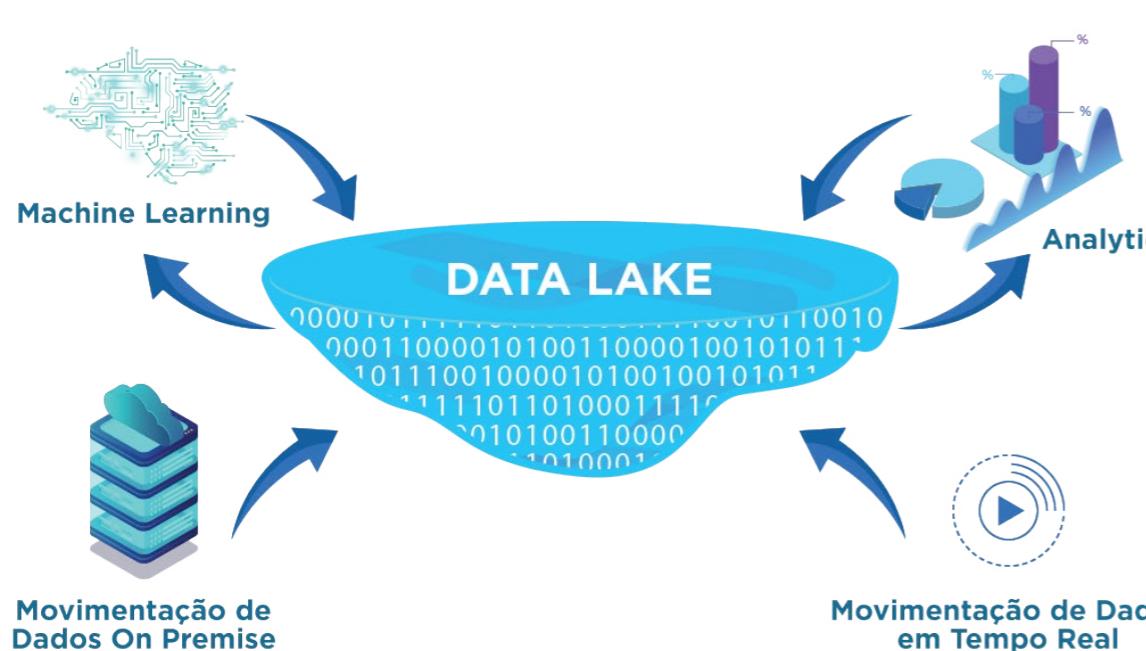
20. <https://www.filecloud.com/blog/2014/08/leading-nosql-databases-to-consider/#.YuLVb3bMK3A>



As principais vantagens de contar com um *data lake* são:

- Permite disponibilidade dos dados a qualquer momento
- Compatibilidade com qualquer formato de dados
- Os dados são disponibilizados de forma bruta, possibilitando que um estudo não interfira em outro

Figura 20. Representação de um data lake²¹



O *data warehouse* (Figura 21) é outra ferramenta para a armazenagem de dados, porém, dedicada a informações já tratadas e padronizadas. Sua proposta é entregar uma visão “limpa” das informações, direcionadas para um objetivo. Normalmente, um *data warehouse* é criado de dados originados do *data lake* que passam por um processo chamado de *Extract, Transform and Load* (ETL). O processo ETL nada mais é que uma cadeia de procedimentos que visa manipular os dados de forma a organizá-los e tratá-los para facilitar o trabalho de análise exploratória dos dados, fazendo com que um *data warehouse* nem sempre possa ser utilizado por todos os profissionais, como um *data lake*.

As principais vantagens de contar com um *data warehouse* são:

- Repositório de dados processados para um estudo de caso
- Contém somente dados estruturados
- Os dados precisam se correlacionar

Figura 21. Representação de um data warehouse²²



Dica: As estruturas de armazenamento de dados apresentadas não são concorrentes e, sim, complementares. Normalmente, a estrutura primária de coleta e armazenamento de dados de uma instituição que lida com um grande volume de dados é um *data lake*, e os dados disponibilizados para diferentes departamentos é um *data warehouse*. É importante explicar aos/as alunos/as os principais usos de cada solução.

<https://www.talend.com/resources/data-lake-vs-data-warehouse/>

2.5. Arquiteturas processamento, otimização e recuperação de banco de dados

Na ciência da computação, existem muitos algoritmos que resolvem o mesmo problema de formas diferentes, a troco de performance e exatidão. Neste sentido, a programação dinâmica é pautada em metodologias de otimização para problemas que requerem decisões interrelacionadas, na qual a escolha imediata afeta o contexto das decisões futuras. O objetivo da programação dinâmica é minimizar o número de estágios necessários e obter uma sequência de decisões ótima para a solução do problema.

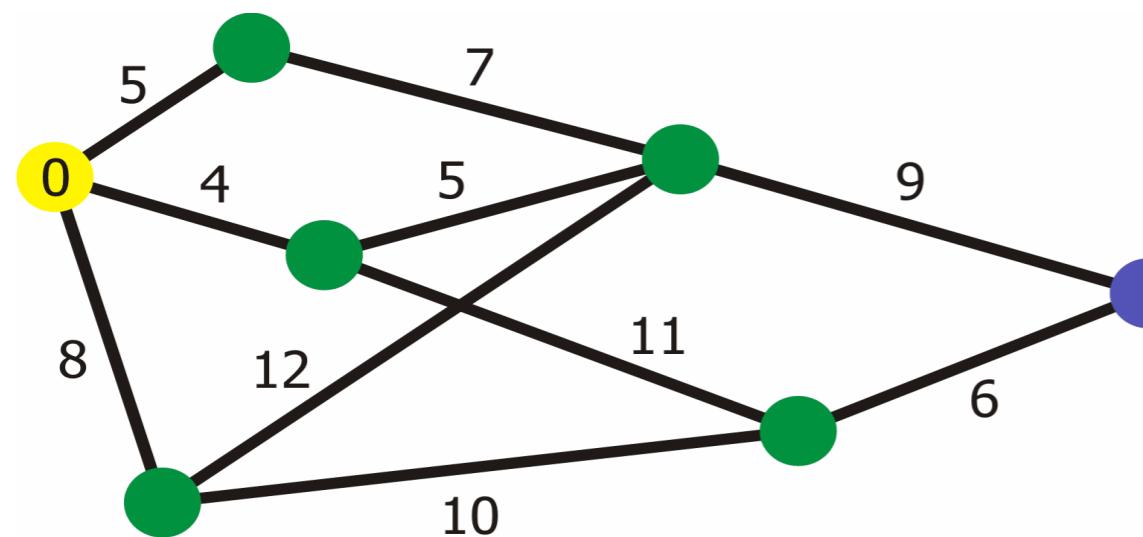
Um problema comum da programação dinâmica é a escolha do caminho de menor esforço, um problema clássico que consiste em encontrar o melhor caminho entre dois pontos. Assim, resolver este problema pode significar determinar o caminho entre dois nós com o custo mínimo, ou com o menor tempo de viagem. Um exemplo deste tipo de problema pode ser observado na Figura 22.

21. <https://solvimm.com/blog/o-que-e-data-lake/>

22. <https://www.salesforce.com/br/blog/2020/10/data-warehouse-e-data-lake.html>



Figura 22. O problema do caminho de menor esforço²³



A programação dinâmica é uma otimização sobre um algoritmo, visando melhorar o seu tempo de processamento e reduzir os recursos utilizados. Ela é aplicada em algoritmos com solução recursiva e se resume em armazenar os dados já calculados durante o processo.

A Figura 23 mostra uma simples solução recursiva utilizada para resolver uma série de Fibonacci e a Figura 24 mostra a solução otimizada utilizando programação dinâmica, a qual armazena os resultados dos cálculos em um vetor f.

Figura 23. Solução recursiva para a série de Fibonacci²⁴

```
FIB (n)
1   se n ≤ 1
2     devolva n e pare
3     devolva FIB (n - 1) + FIB (n - 2)
```

Figura 24. Solução recursiva para a série de Fibonacci utilizando Programação dinâmica²⁵

```
FIB – PD (n)
1   f[0] := 0
2   f[1] := 1
3   para i := 2 até n
4     f[i] := f[i - 1] + f[i - 2]
5   devolva f[n]
```

23. www.bilisim.io

24. Autoria Própria

25. Autoria Própria



Dica: O algoritmo de Fibonacci é de fácil reprodução em qualquer linguagem de programação. Sugere-se ao/a professor/a requisitar aos/as alunos/as que façam um comparativo entre a solução recursiva e a de programação dinâmica em termos de tempo de execução para valores elevados.

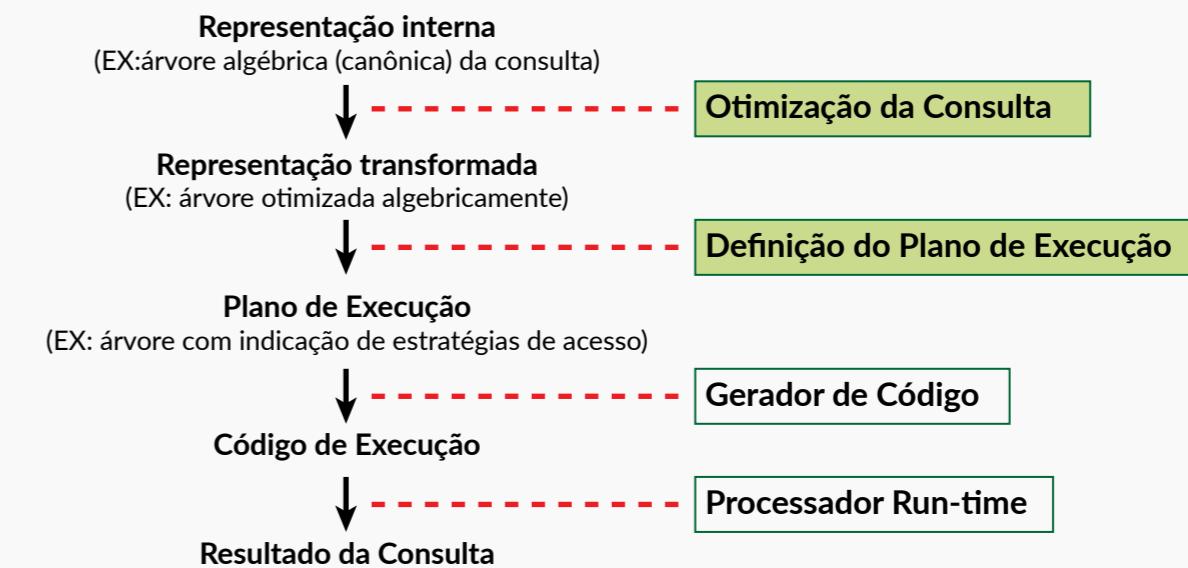
Os bancos de dados possuem o Sistema de Gerenciamento de Banco de Dados, ou SGBD (DBMS em inglês), os quais têm a função de promover a otimização, facilitando a interação e melhorando a velocidade de resposta para os comandos do usuário. Os SGBDs são responsáveis pelas etapas de processamento, desde a tradução do pedido do usuário até a devolução dos itens encontrados. O fluxograma padrão de um SGBD pode ser observado na Figura 25.



Saiba mais: Professor/a, a programação dinâmica e a inteligência artificial são muito utilizadas nos sistemas de busca de bancos de dados. O PostgreSQL faz uso dessas duas estratégias para otimizar as suas consultas. Leve esta referência para acrescentar o conhecimento de seus/suas discentes.

<https://journal.unoeste.br/index.php/ce/article/view/1028>

Figura 25. Etapas realizadas por um SGBD²⁶



Geralmente é preferível que a complexidade e o tempo de execução cresçam linearmente com o tamanho do banco, pois este é o cenário ideal. Portanto, observa-se que várias etapas de otimização de consulta são necessárias para que os softwares sejam eficientes, sendo que o uso da programação dinâmica e de outras técnicas de inteligência artificial voltadas para otimização podem ser utilizadas no mecanismo de busca dos bancos de dados.

26. http://www.decom.ufop.br/guilherme/BCC441/geral/bd2_optimizacao-de-consultas.pdf

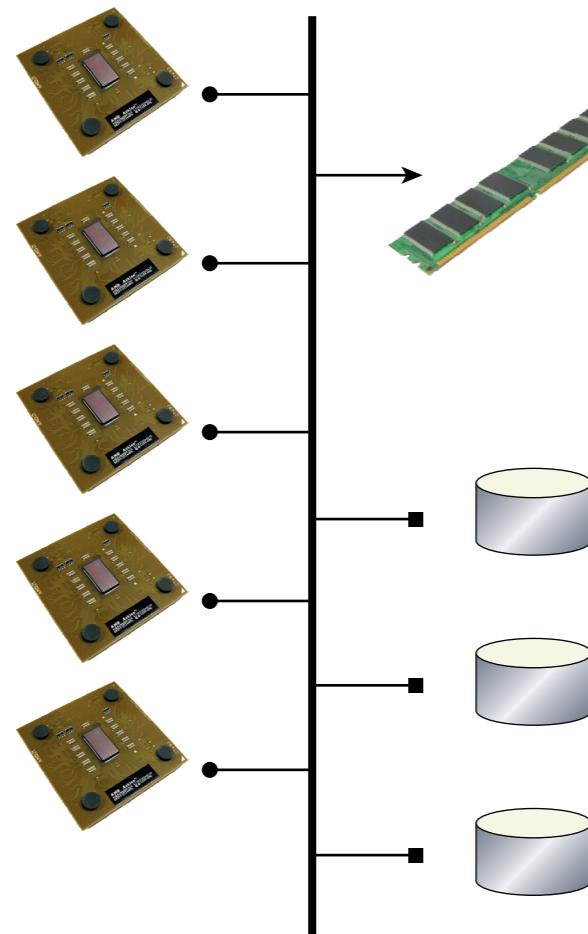


O processamento de grandes bases de dados é um problema complexo e, nesta era de *big data*, este problema ficou ainda mais evidente. Como um banco de dados típico é uma solução de software que é embarcado em um hardware com o intuito de facilitar a gestão de dados, pode-se agregar hardwares para compor um sistema de armazenamento e gestão de dados robusto e confiável, porém, adiciona complexidade de gestão do próprio sistema.

É muito comum utilizar arquiteturas paralelas e distribuídas visando o aumento da velocidade e escalabilidade de um sistema de bancos de dados, sendo que existem 4 principais tipos de taxonomias de *design* para este tipo de arquitetura: Memória compartilhada, discos compartilhados, sem compartilhamento e hierárquico.

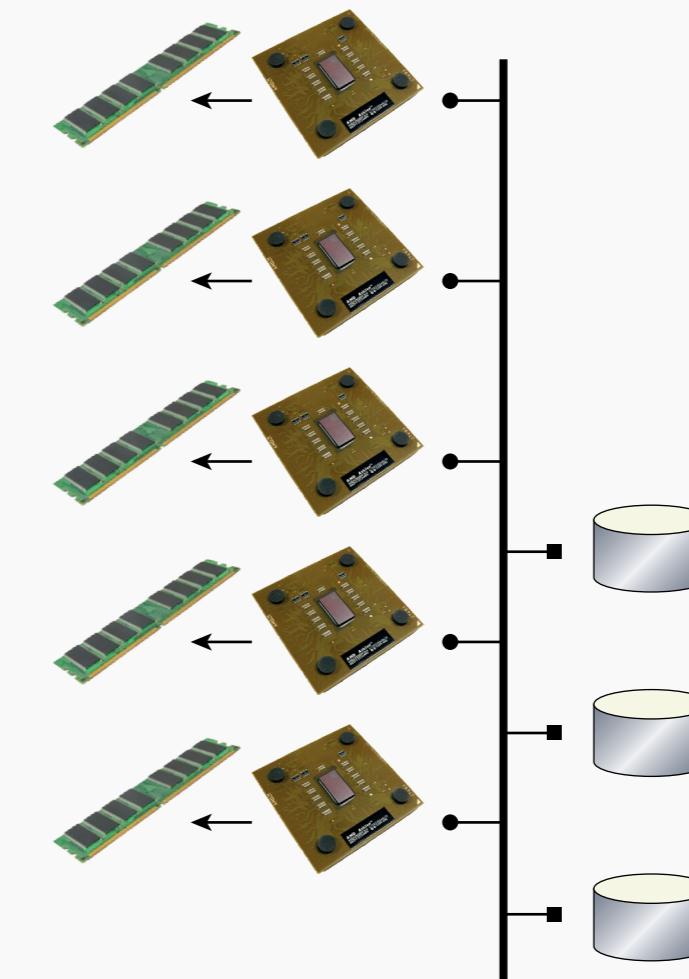
Na arquitetura de memória compartilhada (Figura 26), os processadores e os discos acessam uma memória em comum, normalmente por meio de cabo ou meio de rede de interconexão. A vantagem deste tipo de sistema é a extrema eficiência na comunicação entre os processadores. A desvantagem é o número limite de processadores que podem ser usados, sendo o limite recomendado de 64.

Figura 26. Estrutura de memória compartilhada²⁷



Na arquitetura de discos compartilhados (Figura 27) os processadores acessam diretamente os discos, normalmente por meio de cabo ou meio de rede de interconexão, mas cada processador possui uma memória privada. A vantagem deste tipo de sistema é que não existem gargalos de memória e é de baixo custo, permitindo que redundâncias sejam implementadas, tornando-o robusto a falhas. A sua principal desvantagem é que como cada processador utiliza de sua própria memória, o sistema tende a crescer muito fisicamente.

Figura 27. Estrutura de discos compartilhados²⁸



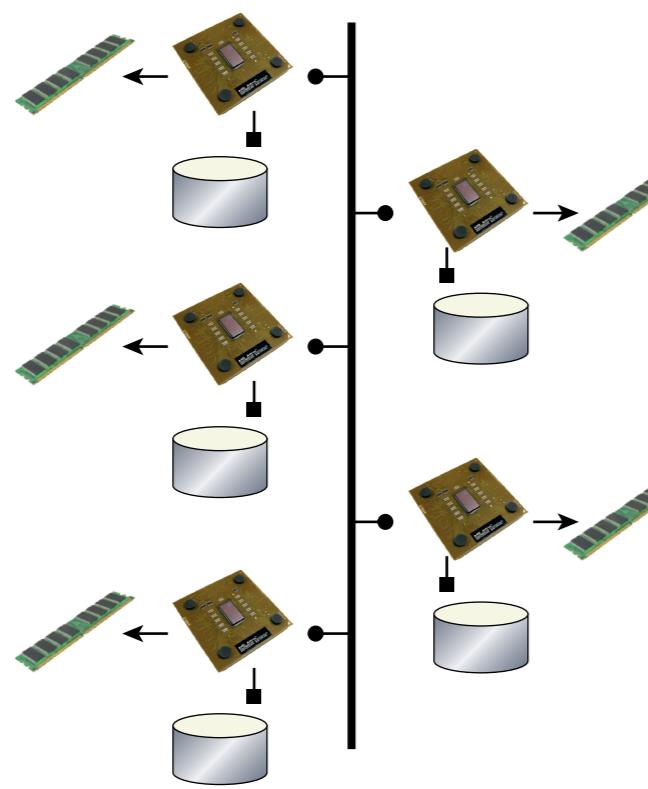
O sistema sem compartilhamento (Figura 28) consiste em um nó onde cada equipamento é um arranjo de processador, memória e disco. A vantagem deste tipo de sistema é que ele dá suporte a muitos processadores, o que reduz drasticamente o tempo necessário para realizar uma busca complexa. A principal desvantagem é ser um sistema custoso, devido ao número de equipamentos necessários para construí-lo.

27. <http://nobios.por.com.br/trabalhos/Bancos%20de%20Dados%20Paralelos.pdf>

28. <http://nobios.por.com.br/trabalhos/Bancos%20de%20Dados%20Paralelos.pdf>

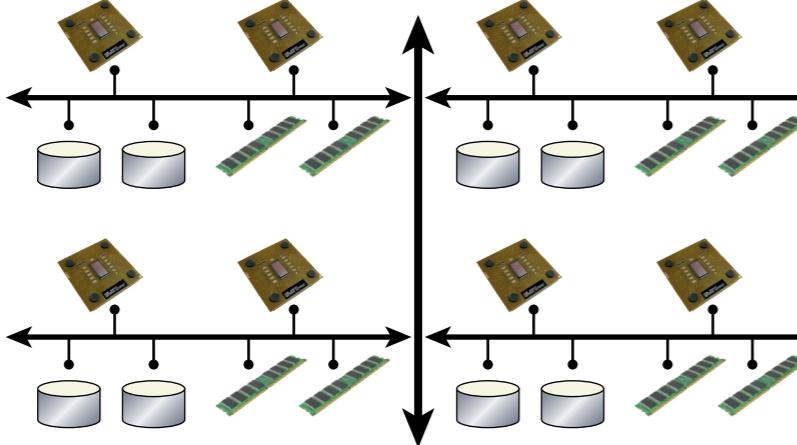


Figura 28. Estrutura sem compartilhamento²⁹



O modelo hierárquico (Figura 29) é o modelo mais flexível dentre os citados. Este modelo combina características de várias arquiteturas anteriores, reduzindo a necessidade e complexidade da comunicação entre processadores. Por isso, utilizar uma estrutura de armazenamento e processamento de dados no formato hierárquico é vantajoso, pois ele é capaz de agregar as vantagens de todos os modelos e as desvantagens de cada um são compensadas pelo formato flexível.

Figura 29. Estrutura hierárquica³⁰



29. <http://nobios.por.com.br/trabalhos/Bancos%20de%20Dados%20Paralelos.pdf>

30. <http://nobios.por.com.br/trabalhos/Bancos%20de%20Dados%20Paralelos.pdf>

Como a tendência atual é de que as empresas adotem cada vez menos arquiteturas de banco de dados instaladas em suas dependências e passem a contratar este tipo de serviço, muitos provedores de computação em nuvem disponibilizam seus serviços através de *Application Programming Interfaces* (APIs). As APIs são muito utilizadas no mundo da programação para permitir que usuários interajam com aplicações sem a necessidade de uma interface gráfica, permitindo assim a orquestração dos sistemas de armazenamento, o que dá ao usuário a liberdade de escolha em construir e operar sua rede de banco de dados.

Como foi dito, a tendência é que os sistemas em nuvem sejam cada vez mais utilizados. Um dos motivos é que a aquisição e a manutenção das arquiteturas de hardware são trabalhosas. Outro motivo bastante relevante é a possibilidade de criar sistemas redundantes com facilidade e com baixo custo para que os dados possuam uma segurança adicional perante falhas.

A recuperação de falhas é um assunto delicado e importante, pois visa garantir a persistência e integralidade de dados durante as transações em um banco de dados. Normalmente, os softwares que lidam com gerenciamento e edição de dados realizam *backups* (salvamento do último estado) de forma periódica, com o objetivo de minimizar os impactos caso uma falha venha a ocorrer. Um banco de dados em um ambiente de produção (disponível para os usuários comuns) faz milhares de transações diárias e está sempre suscetível a falhas. Desta forma, existem basicamente 3 tipos de falhas que podem acontecer:

- Falhas de transação:
 - Erro lógico: Quando a transação não pode ser efetivada por condições externas.
 - Erro do sistema: Quando a transação não pode ser efetivada porque o sistema entrou em um estado inadequado.
- Queda de sistema: Promove a perda de dados em armazenamento volátil (memória RAM), porém a base persistente (não volátil) continua operando corretamente.
- Falha de disco: representa a perda, parcial ou total, do conteúdo não volátil.



Saiba mais: Professor/a, existem muitas técnicas que possibilitam a recuperação de dados em banco de dados. Leve o conhecimento das principais técnicas de backup que são realizadas ao se trabalhar com banco de dados relacionais para os/as alunos/as.

<https://www.devmedia.com.br/artigo-sql-magazine-44-tecnicas-de-backup-e-recuperacao-de-dados-em-mysql/7092>



Saiba mais: Armazenar dados de forma eficiente é muito importante para as empresas e instituições. Isso elevou a demanda por profissionais para atuarem em várias áreas, como desenvolvimento e segurança digital. Além disso, uma pessoa que passa a ter grande importância para o negócio é a que atua em processamento de dados. Neste sentido, é importante que o/a professor/a conscientize os/as discentes sobre a importância de se obter o conhecimento na área de processamento de dados.

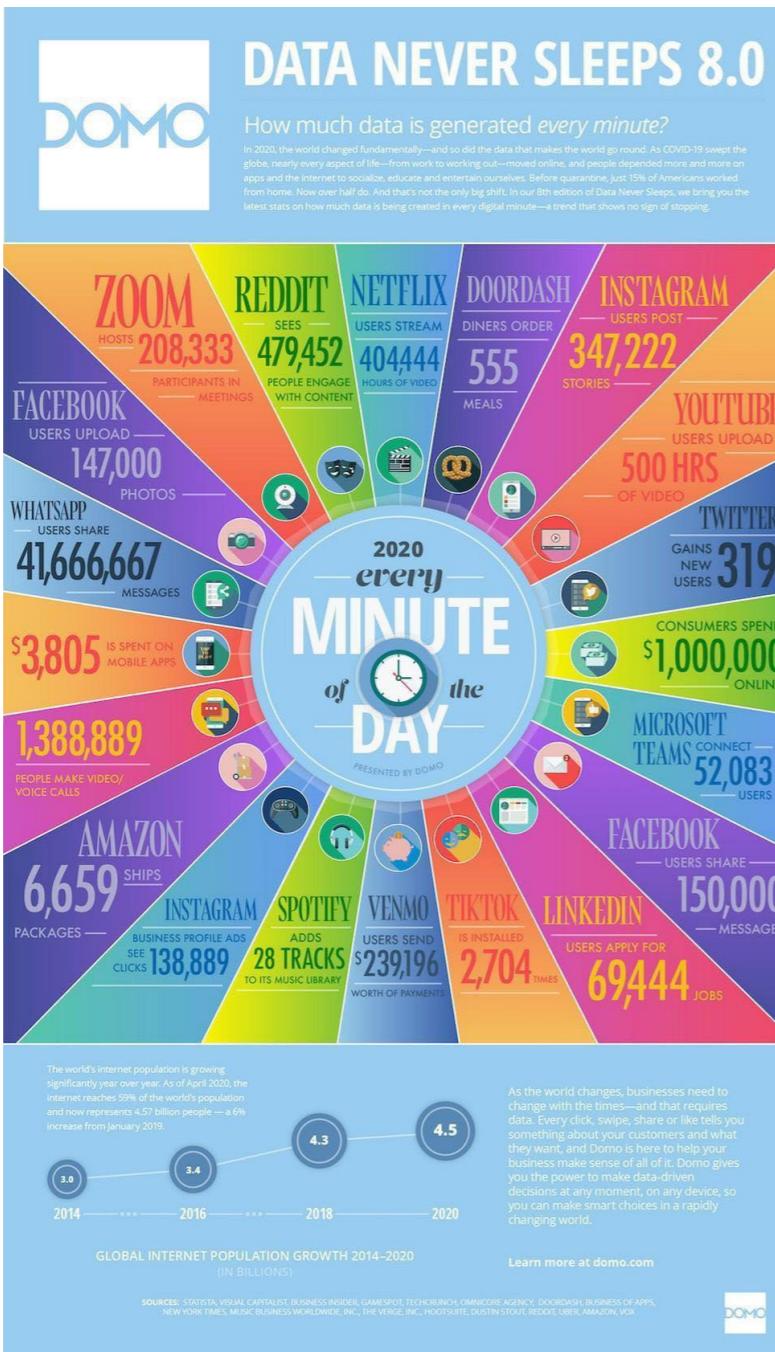
<https://blog.betrybe.com/carreira/processamento-de-dados/>



2.6. Externalização de serviços de armazenamento

Quando se fala em *big data*, é importante visualizar a quantidade de dados gerados diariamente no mundo. Estudos mostram que de 2006 a 2020 o volume saiu da casa de centenas de exabytes para milhares de exabytes. Para se ter uma noção, uma notícia publicada no canal techtudo, mostrou que a quantidade de dados gerados diariamente em 2020, em média, foi de aproximadamente 2,2 milhões de terabytes, sendo que alguns dados curiosos foram apresentados: 41 milhões de mensagens foram trocadas no WhatsApp e cerca de 7 mil encomendas foram enviadas pela Amazon por minuto, como pode ser observado na Figura 30.

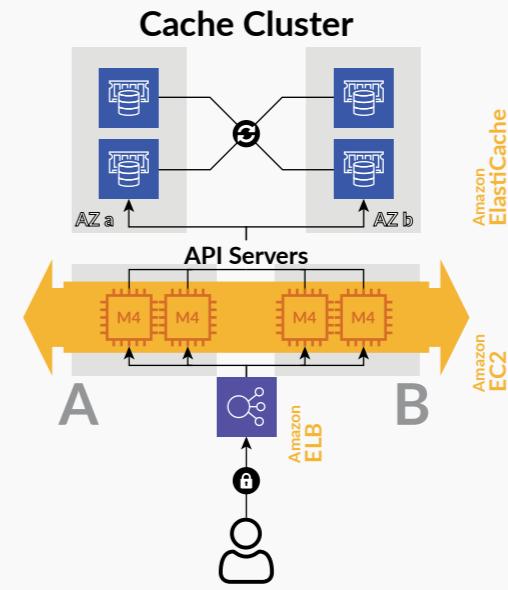
Figura 30. Dados gerados em 1 minuto do dia³¹



31. <https://www.techtudo.com.br/noticias/2020/08/o-que-acontece-a-cada-minuto-na-internet-estudo-traz-dados-surpreendentes.ghtml>

Para conseguir lidar com uma quantidade de dados tão grande, as empresas tiveram a necessidade de adquirir *hardware* (servidores em redundância) com alta capacidade de processamento visando garantir a disponibilidade e a boa experiência para os seus usuários. A grande questão é que para manter estes servidores é necessário um investimento alto, que muitas vezes é mal utilizado, pois os equipamentos devem estar preparados para os piores casos de funcionamento, que não acontecem o tempo inteiro. Neste sentido, observou-se uma nova oportunidade de negócio: serviço de contratação de *hardware* sob demanda, com alta disponibilidade pela internet, o que é atualmente conhecido como *cloud computing*. Basicamente, este tipo de serviço é pautado na disponibilização de uma API para o cliente, que consegue reproduzir uma infraestrutura de *hardware* contendo: redes públicas e privadas, tabelas de roteamento, *hardware* com alta capacidade de armazenamento, *hardware* com alta capacidade de processamento, infraestrutura com redundância e alta disponibilidade. Uma curiosidade é que a redundância e a disponibilidade dos serviços de *cloud computing* são em nível global, como pode ser observado na Figura 31, na qual as zonas 'AZ a' e 'AZ b', podem estar muito distantes, do ponto de vista geográfico.

Figura 31. Solução genérica baseada em cloud computing³²



Os provedores mais conhecidos de *cloud computing* são: AWS, Microsoft Azure e Google Cloud. Eles possibilitam a construção de uma aplicação voltada para desempenho e velocidade, como visto na Figura 31, que se trata de uma construção de *clusters* para armazenamento em cache com redundância, para garantir a disponibilidade.



Dica: Normalmente os serviços de cloud computing oferecem um período gráti para teste das tecnologias. Sugere-se ao/a professor/a experimentar a solução com os/as alunos/as.

Abaixo, algumas referências:

<https://aws.amazon.com/free>

<https://azure.microsoft.com/en-us/free/>

<https://cloud.google.com/free>

32. Autoria própria



Para finalizar o assunto, é importante dizer que existem três tipos principais de modelos que podem ser adotados quando se pensa em disponibilizar uma aplicação: pública, privada e híbrida. Estes modelos podem ser observados na Figura 32:

Figura 32. Modelos de implementação na nuvem³³



Apesar dos modelos baseados em nuvem serem escaláveis e facilitarem muitas outras ações operacionais, muitas empresas ainda optam por manter o sistema localmente, devido à vulnerabilidade e importância dos dados trabalhados. Isto é uma questão estratégica e casual das empresas, porém, com o avanço da cibersegurança, é cada vez mais confiável trafegar dados pela internet e armazená-los totalmente em nuvem. Muitas instituições têm adotado o modelo híbrido, que é uma boa escolha quando já existe uma infraestrutura em suas dependências, mas existe a vontade de usufruir dos benefícios oferecidos pelos serviços *cloud*.



Futuro: Futuro: É importante dizer aos/as alunos/as que a tendência das empresas é migrar a maior parte dos seus sistemas de armazenamento para a nuvem. Neste ebook, já foram citadas muitas vezes as vantagens que podem ser obtidas utilizando os serviços de computação em nuvem. Um estudo realizado em 2019 indicou que em 2022 75% das bases de dados terão migrado para ou sido implementadas na cloud.

<https://portalerp.com/o-futuro-dos-dados-esta-na-nuvem>

2.7. Recapitulando

Ao final deste capítulo, o/a aluno/a deverá ser capaz de:

- Compreender o que é a LGPD e qual a sua importância para a pessoa natural.
- Compreender as atribuições básicas relacionadas à gestão de dados dos agentes do setor elétrico, do ONS e da ANEEL.

33. <https://www.icmpconsultoria.com.br/post/nuvem-publica-privada-ou-hibrida-qual-e-a-melhor-opcao-para-sua-empresa>

- Compreender o conceito de segurança por *hardware* e qual a importância dessa camada adiciona na proteção dos dados.
- Compreender os tipos básicos de bancos de dados existentes e os principais *softwares* do mercado.
- Compreender as técnicas de gestão de dados por meio de *data lakes* e *data warehouses* e as principais características de cada um.
- Compreender as principais arquiteturas para processamento de dados distribuídos e paralelos.
- Compreender os conceitos básicos atrelados à externalização de serviços de armazenamentos e as principais ferramentas disponíveis no mercado

Atividades Sugeridas:

- Apresentações expositivas em slides com os temas: Gestão de dados consciente no setor elétrico (LGPD e GDPR), principais bancos de dados e seus mecanismos (NoSQL, SQL e as soluções do mercado), vantagens e desvantagens de utilizar um módulo de programação distribuída e paralela, principais serviços de *cloud* e estruturas utilizadas.
- Apresentação de vídeos para consolidação do conceito e novidades relacionadas aos temas abordados em aula:
 - https://www.youtube.com/watch?v=8R_26n76kTc
 - https://www.youtube.com/watch?v=0TVV_jO5OU
 - <https://www.youtube.com/watch?v=jBTVe9ck6U>
 - <https://www.youtube.com/watch?v=Aoq-5eyskEk>
 - <https://www.youtube.com/watch?v=QyWigqkgcbU>
 - <https://www.youtube.com/watch?v=JlbIYCM48to>
- Promover atividades em grupo ou individuais, tais como seminários para avaliação do aprendizado. Sugestão de temas:
 - Aplicação dos bancos de dados no setor elétrico: dados históricos e otimização da operação.
 - A importância de se utilizar um serviço *cloud* em uma aplicação *IoT* e como a gestão de dados pode ser mais eficiente.
 - LGPD passo a passo e aplicações no setor elétrico.
- Convidar os/as alunos/as para participarem de webinar e outros eventos disponibilizados pela instituição ou fora dela como forma de intensificar a sua capacitação.

Algumas fontes interessantes:

- [Role of Cloud Computing in smart grid](#)
- [What Is Cloud Computing? A Beginner's Guide](#)
- [5 Reasons Why Cloud Computing Is Important for Your Business](#)
- [How the Cloud Is Transforming the Energy Sector: 7 Practical Use Cases](#)
- [Cloud in Power](#)

Avaliação:

Tendo em vista tudo o que foi visto no capítulo, discorra sobre a importância da utilização de um banco de dados em uma organização. Comente sobre os diferentes tipos de bancos de dados, os estruturados e os não estruturados. Explique também qual a importância de se ter um *data lake* e de se utilizar provedores de *cloud* para manter uma infraestrutura segura e eficiente.



Capítulo 3: Tratamento de dados e machine learning



Nota para o/a professor/a – Este capítulo deve servir como base teórica para que os/as alunos/as compreendam os principais conceitos de tratamento de dados e quais são os principais algoritmos de machine learning. Desta forma, os discentes estarão preparados para dar os passos iniciais em estudos mais aprofundados que envolvam machine learning aplicado ao setor elétrico.

Observação: Neste capítulo é importante manter o foco no conceito de cada algoritmo, pois no próximo capítulo o conhecimento será aplicado.

3.1. Tratamento de dados

Devido ao volume de dados disponíveis atualmente em qualquer processo, é muito comum se deparar com erros de registro inerentes ao armazenamento, à medição ou até mesmo à comunicação e transmissão de dados. Isto gera muitos problemas em estudos e algoritmos que fazem uso destes dados, levando a resultados errôneos de estudos e comportamentos inesperados e inficazes de algoritmos.

Neste sentido, é necessário tratar os dados, descartando ou ressignificando os seus valores, por exemplo. Para tal, pode-se utilizar técnicas estatísticas e de manipulação de dados. É importante ter muita cautela no momento de tratar os dados de uma base, pois o processo de tratamento não deve inserir tendências ou distorcer as informações registradas.

O método mais simples e eficaz de tratamento de dados nasceu da estatística e é chamado de remoção de *outliers*. *Outliers* são registros que diferem estatisticamente da maioria, e geralmente devem ser removidos. Existem casos em que remover um *outlier* pode não ser interessante, pois ele pode não significar uma anomalia e sim um comportamento pouco frequente. Para tratar os dados corretamente, dois requisitos são muito importantes: conhecimento estatístico e conhecimento do problema. A ferramenta estatística fará o uso da matemática para expor os potenciais dados problemáticos, mas é responsabilidade do profissional decidir como tratar o dado.



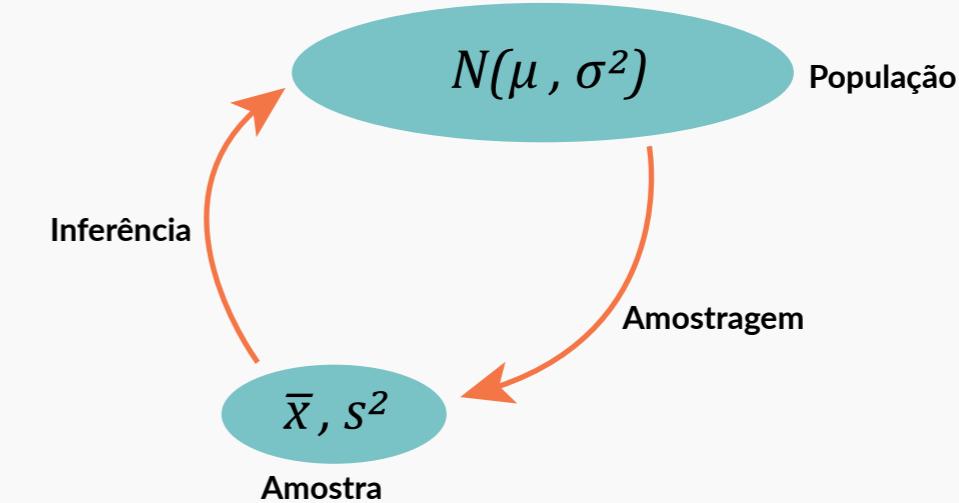
Dica: O conhecimento estatístico necessário para os/as discentes prosseguirem com os estudos é de nível básico. Contudo, caso o/a professor/a sinta a necessidade de um maior formalismo matemático ou precise de um material de apoio, sugere-se o livro **Estatística Básica** do autor Wilton De O. Bussab [23].

Para entendimento mais simples do método de remoção de *outliers* estatístico, fazem-se necessárias algumas definições:

1. **População:** É o conjunto de todas as respostas, medidas ou contagens que sejam de interesse do estudo.
2. **Amostra:** É um subconjunto da população, que deve ser escolhida aleatoriamente, para evitar tendências (vícios).
3. **Parâmetro:** É uma descrição numérica de uma população.
4. **Estatística (Estimadores):** É uma descrição numérica de uma amostra.

A Figura 33 representa o conceito de amostra e população, com suas respectivas descrições numéricas, onde μ e \bar{x} representam as médias e σ^2 e s^2 as variâncias (o desvio padrão é a raiz quadrada da variância) representam as médias e as variâncias (o desvio padrão é a raiz quadrada da variância).

Figura 33. Amostragem e Inferência³⁴



Fórmula para calcular a média:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Fórmula para calcular o desvio padrão da população:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

³⁴ www.lampada.uerj.br



Fórmula para calcular o desvio padrão da amostra:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

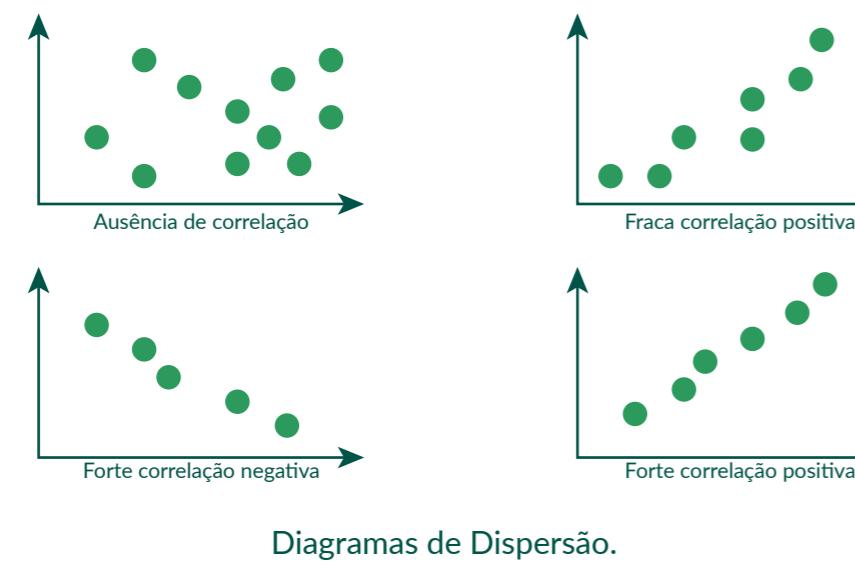
$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$

As variáveis que fazem parte de uma população, ou amostra, são divididas em 4 tipos principais, são eles:

- Qualitativa nominal:** É o tipo de variável que normalmente é selecionada dentro de um grupo de palavras sem hierarquias, que tem sentido para a pesquisa feita.
- Qualitativa Ordinal:** É o tipo de variável que normalmente é selecionada dentro de um grupo de palavras que possuem hierarquias, que tem sentido para a pesquisa feita.
- Quantitativa discreta:** É o tipo de variável numérica para a qual a ordenação é feita através da mesma forma que na ordem matemática e os valores podem variar, desde que em porções fixas.
- Quantitativa Contínua:** É o tipo de variável numérica para a qual a ordenação é feita através da mesma forma que na ordem matemática e os valores podem variar livremente.

Além de todos as definições supracitadas, ainda existe um importante conceito a ser elencado: correlação. É muito comum que a variação de uma variável dependa parcialmente da variação de outra, ou seja, as variáveis são correlatas. A correlação na estatística é responsável por medir o quanto duas variáveis estão correlacionadas. Tipicamente, os valores da correlação variam de 1 à -1, sendo que o extremo positivo indica uma variação linear positiva (diretamente proporcional) e o extremo negativo indica uma variação linear negativa (inversamente proporcional) entre duas variáveis. A Figura 34 representa graficamente o comportamento da correlação entre as variáveis:

Figura 34. Medidas de Correlação³⁵



Na estatística descritiva, o coeficiente de correlação mais utilizado é o de Pearson, que obedece às regras já discutidas. A fórmula para calcular o coeficiente de Pearson entre duas variáveis (x, y) é dada como:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



Saiba mais: Existem muitos tipos de coeficientes de correlação, e cada um deles pode ser aplicado em diferentes casos e apresentam resultados ligeiramente diferentes. É importante que o/a professor/a apresente os outros tipos dos coeficientes em aula.

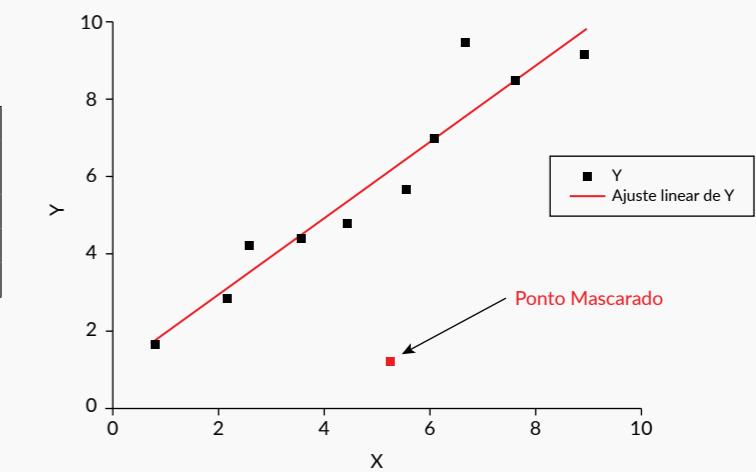
<https://operdata.com.br/blog/coeficientes-de-correlacao/>

Supondo uma base de dados em que a correlação das variáveis é linear, é coerente buscar a melhor reta que se ajuste bem ao redor de todos os pontos em uma dada região, processo que comumente é chamado de “encontrar o modelo do problema”. Com o modelo em mãos, assume-se que a reta é capaz de representar bem todos os registros da base de dados. Em uma base de dados com uma correlação forte, ou seja, com um índice de Pearson próximo à 1 ou -1, os potenciais *outliers* são aqueles valores que se distanciam das medidas estatísticas esperadas.

Neste sentido, os *outliers* são valores discrepantes de todos os outros valores da base. A Figura 35 exemplifica um caso de *outlier* identificado como “Masked Point” (ponto mascarado). É importante observar que para o caso da Figura 35, o modelo linear representa muito bem todos os outros pontos, mas que o índice R^2 utilizado para mensurar a aderência do modelo aos dados decaiu por conta do *outlier*, o que pode levar um analista a entender, erroneamente, que o modelo é ruim para representar o seu problema.

Figura 35. Presença de outlier na modelagem³⁶

Equação	$y = a + b \cdot x$		
Ajuste do índice R^2	0.90163		
		Valor	Erro Padrão
Y	Intercepto	0.98113	0.58789
Y	Angulação	0.99149	0.10851



35. www.segredosdaestatistica.wordpress.com

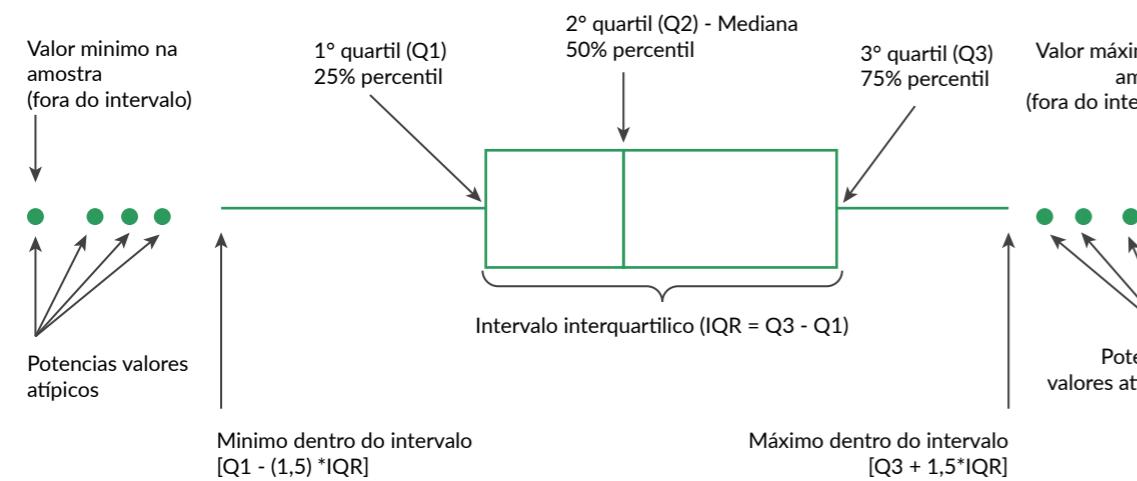
36. <https://www.originlab.com/doc/origin-help/detect-outlier>



A origem dos *outliers* se dá por diferentes causas, podendo ser elas naturais ou artificiais. O erro natural é causado por uma anomalia da própria amostra, ao contrário do erro artificial que é inserido no processo de amostragem ou interpretação.

As técnicas para a identificação de *outliers* são diversas, sendo a mais comum a detecção através da interpretação de um *boxplot*. Em uma amostra ordenada, o primeiro quartil corresponde ao valor que ocupa a posição que divide o tamanho da amostra em 1/4, a mediana corresponde ao valor que ocupa a posição que divide o tamanho da amostra em 1/2 e o terceiro quartil, por sua vez, o valor que divide o tamanho da amostra em 3/4. Figura 36 mostra como é um *boxplot*.

Figura 36. BoxPlot³⁷



Por definição, os limites intervalares mostrados no gráfico correspondem ao dado mais próximo dos limites teóricos máximos encontrados, ou seja, supondo que o limite inferior seja 2 e o limite superior seja 20, mas os valores da amostra que mais se aproximam desses valores são 7 e 18, no gráfico, os limites devem ser representados como 7 e 18. Uma vez que as anormalidades forem detectadas, deve-se considerar algumas opções para o tratamento delas, como remoção, ressignificação ou interpolação de dados.

- **Remoção de dados:** A remoção de dados pode ser utilizada tanto em registros quantitativos quanto em registros qualitativos, desde que os dados a serem removidos não representem um impacto muito grande sobre a base em geral. Tipicamente este método é indicado a ser utilizado em uma base de dados com muitos registros.
- **Ressignificação de dados:** A ressignificação de dados pode ser utilizada tanto em registros quantitativos quanto qualitativos. O processo de ressignificação é adotar um valor típico em sua base (média ou moda) ou forjar um dado que faça sentido dentro da análise que não crie uma tendência errônea no estudo. Este processo pode ser oneroso aos resultados, principalmente quando não é utilizado a média ou a moda, devendo sempre ser aplicado com cautela.
- **Interpolação de dados:** A interpolação de dados pode ser utilizada em registros quantitativos e significa basicamente encontrar valores intermediários entre os dois pontos que cada outlier se posiciona. É um procedimento bem semelhante ao de adotar a média, porém para séries temporais este procedimento pode ser o mais adequado.

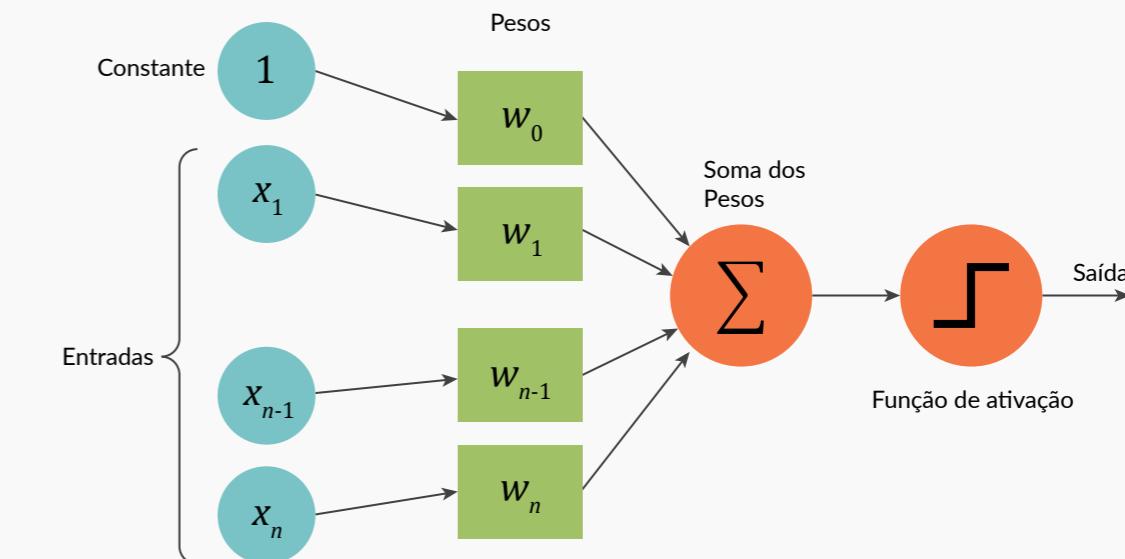
37. <https://medium.com/@claudio.siervi/interpretando-o-diagrama-de-caixa-boxplot-1876b7c099af>

3.2. Machine learning: Modelos de redes neurais artificiais

A aprendizagem de máquina revolucionou o campo de análise de dados. Os modelos são capazes de gerar *insights* sobre dados de um negócio, que nem mesmo os analistas são capazes de identificar, tornando-se grandes aliados na solução de problemas e otimização de produtos e processos. Um dos modelos mais famosos dentro do campo de aprendizagem de máquina são as redes neurais artificiais ou RNA.

Uma RNA é uma modelagem algorítmica que visa simular o funcionamento de um neurônio biológico. A sua implementação no nível mais básico é chamada de Perceptron e pode ser observada na Figura 37.

Figura 37. Perceptron³⁸



O Perceptron busca emular as sinapses cerebrais, através de um modelo linear multivariável que multiplica as entradas pelos pesos W e soma os resultados com o intuito de sensibilizar uma função de ativação, que pode ser um degrau ou uma sigmoide. O processo de aprendizagem se dá através da atualização dos pesos W que são modificados através do feedback de dados rotulados, fazendo com que este Perceptron represente a maior quantidade de casos possível.



Saiba mais: Saiba Mais: A função de ativação é uma das partes mais importantes do Perceptron. Ela é responsável por decidir se o somatório dos pesos é válido ou não. Neste sentido, existem diversas funções de ativação que podem ser utilizadas. É importante que o/a professor/a trabalhar os conceitos básicos de cada uma com os/as discentes.

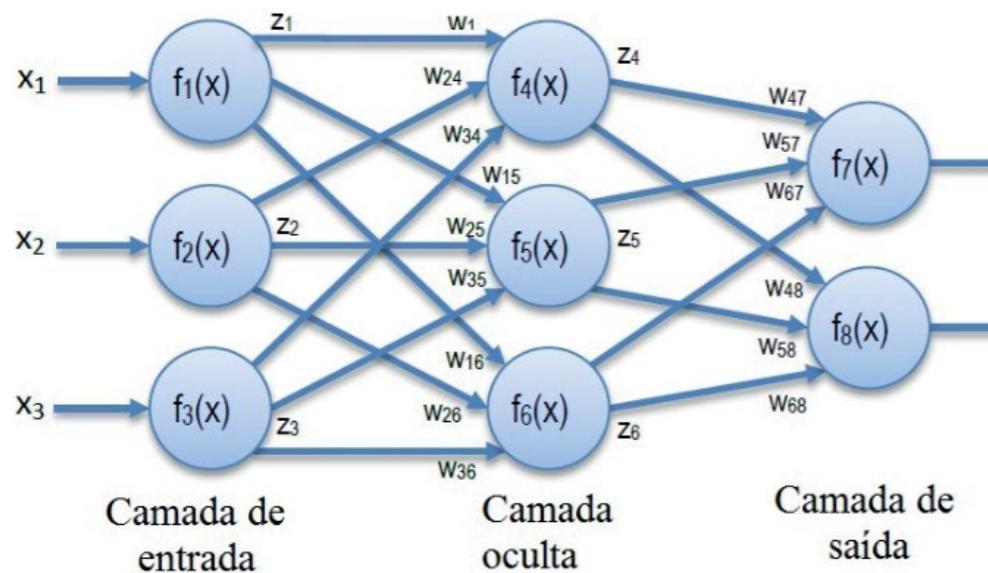
<http://www2.decom.ufop.br/imobilis/redes-neurais-funcoes-de-ativacao/>

38. <https://datascience.eu/pt/aprendizado-de-maquina/perceptron/>



Um Perceptron sozinho não é suficiente para resolver os problemas modernos que usam RNAs como solução. É necessária uma cadeia de Perceptrons, o que forma uma rede neural efetivamente. Na Figura 38 pode-se observar uma rede neural com uma camada oculta. Em resumo, quanto mais camadas possui uma rede neural, maior a sua capacidade de aprendizado. Isto acontece porque as camadas ocultas se organizam, automaticamente, para identificar padrões não triviais nos dados de entrada.

Figura 38. Implementação básica de uma rede neural artificial³⁹



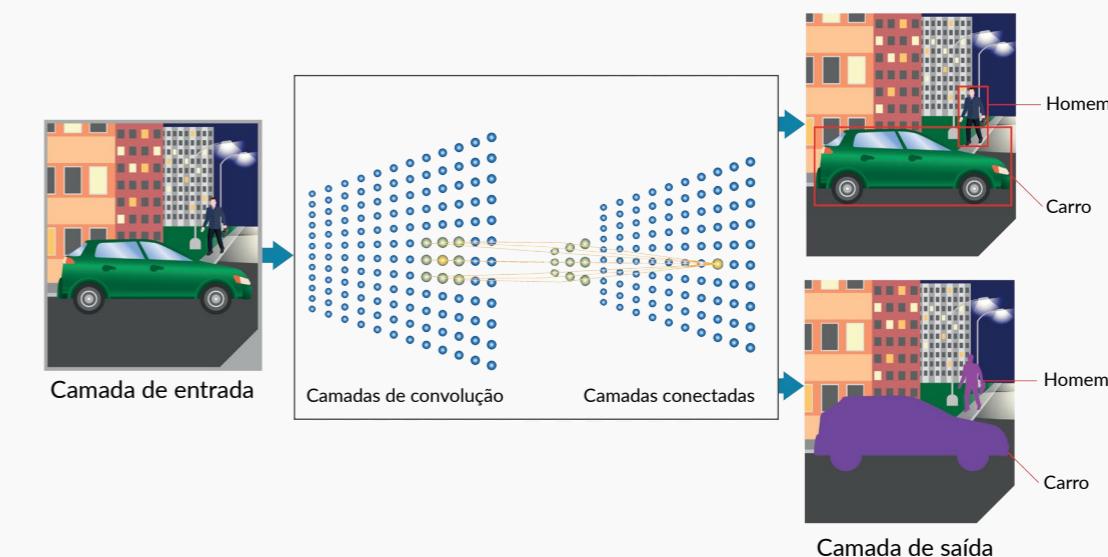
Apesar de serem bastante versáteis, a interpretação do modelo matemático resultante de uma RNA é complexa e muitas vezes humanamente inviável. Por este motivo, é muito comum encontrar trabalhos na literatura que tratam as redes neurais como um bloco de modelo desconhecido. Mesmo sendo complexo conseguir explicar matematicamente o comportamento de uma rede neural, as RNAs podem ser aplicadas em diversas tarefas, já que conseguem encontrar uma relação que representa eficientemente as entradas e saídas.

Neste sentido, muitas novas categorias de redes neurais surgiram com o passar dos anos, com o objetivo de facilitar tarefas específicas. Um exemplo de rede neural muito utilizada hoje no campo da visão computacional é a Rede Neural Convolucional (RNC). Uma RNC é uma rede capaz de aplicar um pré-processamento na imagem, visando reduzir a dimensionalidade e a complexidade na identificação de objetos. Estes processos de minimização de complexidade e dimensionalidade atuam sobre características específicas da imagem como arestas, iluminação e cores, permitindo assim que objetos sejam identificados com maior facilidade.

O nome convolucional vem do processo de convolução aplicado na imagem. Este processo é responsável por aplicar filtros que são responsáveis por transformar um conjunto de pixels em um valor representativo. Desta modo, a RNC aprende não somente os pesos, para ligar as entradas à uma saída, ela pode aprender também os melhores filtros para identificar os objetos, como pode ser observado na Figura 39.

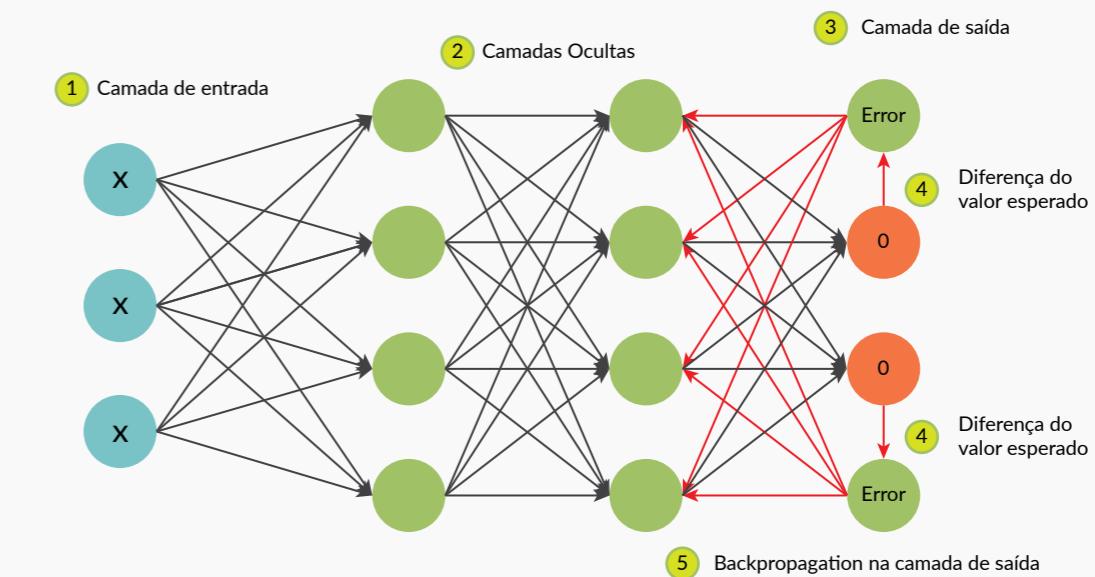
39. <https://www.redalyc.org/jatsRepo/3071/307156938018/html/index.html>

Figura 39. Rede Neural Convolucional na identificação de imagem⁴⁰



Os modelos supracitados consideram, até então, um tipo de modelo conhecido como “feed forward”, ou seja, os resultados das camadas só têm efeito nas camadas posteriores, e os pesos da rede são atualizados todos de uma única vez por iteração na etapa de treinamento. A Figura 40 exemplifica um outro tipo de rede, conhecida como *back propagation*. Este tipo de rede permite a inserção de um certo dinamismo na representatividade, uma vez que os resultados das camadas mais próximas da saída têm impacto nas camadas anteriores. Desta forma, este tipo de rede se torna consciente sobre o contexto em que estão trabalhando e conseguem se adaptar a ele.

Figura 40. Rede Neural Com Back Propagation⁴¹



40. <https://www.algotive.ai/blog/everything-you-need-to-know-about-deep-learning-the-technology-that-mimics-the-human-brain>

41. <https://medium.com/turing-talks/turing-talks-26-modelos-de-pred%C3%A7%C3%A3o-redes-neurais-recorrentes-439198egecf3>



Saiba mais: Existem muitos outros tipos de redes neurais. Como foi dito no texto, elas vão surgindo para resolver problemas específicos, que o uso as RNAs convencionais não são capazes de resolver. Portanto, é muito importante o/a professor/a levar o conteúdo aos/as alunos/as.

<https://iaexpert.academy/2020/06/08/os-tipos-de-redes-neurais/>

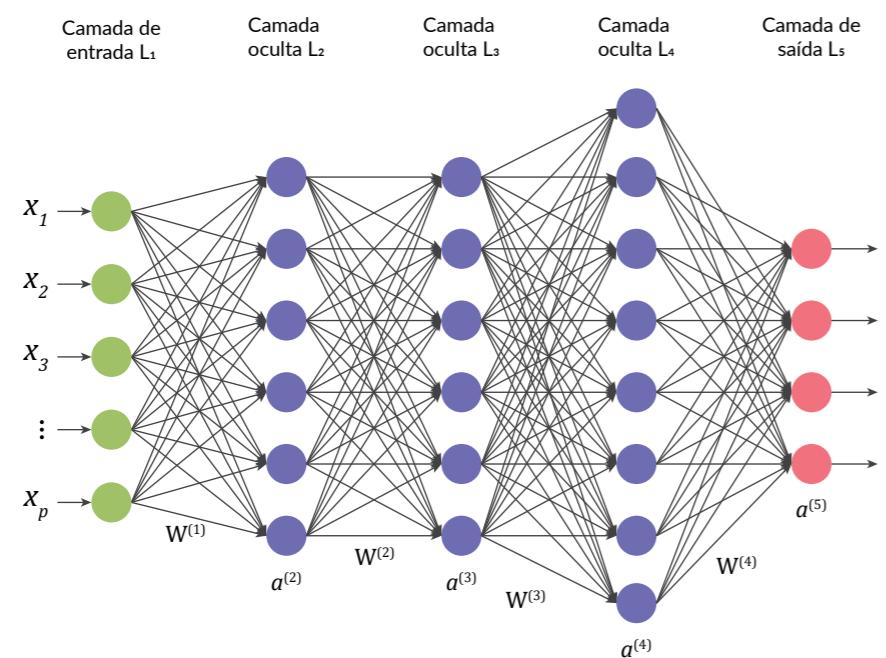
As redes neurais têm um campo de aplicação muito abrangente. Um exemplo é que na engenharia elétrica as RNAs são aplicadas em plantas de geração renovável para auxiliar na tomada de decisão. Neste contexto, a RNA emula o comportamento da planta de acordo com algumas variáveis de entrada e consegue, com uma certa precisão, mensurar quanto a planta vai realmente ser capaz de gerar durante a operação. Esta abordagem ajuda a lidar com o caráter intermitente das fontes renováveis de energia.

3.3. Machine learning: Modelos de redes neurais profundas

Deep learning é uma área da inteligência que busca ensinar as máquinas a fazerem tarefas que são feitas por humanos (ou extrapolando um pouco, mimetizar o cérebro humano). Do *deep learning* nascem os modelos de redes neurais profundas (RNP), que são redes muito grandes e são especializadas em solucionar problemas e identificar padrões complexos. Como exemplo cita-se o DeepL Translator (Tradutor de Linguagens) e a Cortana (assistente personalizado da Microsoft).

Resumidamente, o que difere uma Rede Neural Artificial (RNA) de uma RNP é a quantidade de neurônios e de conexões. Um exemplo de uma RNP pode ser observado na Figura 41:

Figura 41. Rede Neural Profunda⁴²



42. <https://towardsdatascience.com/its-deep-learning-times-a-new-frontier-of-data-a1egefgfega8>



Dica: Professor/a, a figura é meramente ilustrativa. Uma RNP pode possuir milhões de conexões e Perceptrons.

O treinamento deste tipo de rede é extremamente complexo e custoso. Normalmente, é necessário um volume grande de dados na fase de treinamento, sendo que o processo pode durar horas, mesmo em computadores de altíssima performance. Tecnologias mais apuradas permitem a execução do treinamento utilizando programação paralela e distribuída, visando reduzir o tempo necessário para o treinamento.

É importante ressaltar que como a etapa de treinamento de uma rede neural profunda é complexa e como elas são muito utilizadas em aplicações específicas, a comunidade científica já se prontificou em disponibilizar alguns modelos genéricos de RNP pré-treinados em repositórios na internet.

As RNPs são comumente utilizadas na área de processamento e identificação de imagens. A Tesla (conhecida mundialmente por seus investimentos em mobilidade elétrica) faz o uso de redes neurais complexas para adicionar autonomia a seus carros elétricos. A Tesla é a pioneira no “Auto Pilot”, que é um sistema de direção autônomo, o qual permite que o proprietário selecione o local de destino e o carro faz o trajeto completo, sem precisar ser controlado por um motorista.

Um outro exemplo no uso de RNP é no digital twin. Atualmente, existem empresas que utilizam carros com câmeras 360° ou drones para escanear a rede elétrica em alta resolução, levando posteriormente o material para processamento em uma RNP. O resultado é a identificação de anomalias na rede como isoladores rachados, equipamentos mal instalados e até mesmo depreciação de pontos das redes devido à ação de animais.



Futuro: A tendência é o setor de energia implantar cada vez mais soluções baseadas em machine learning. Com estes algoritmos é possível mitigar falhas, elaborar um plano de manutenção preditiva e detectar fraudes.

Professor/a este assunto é latente e existem muitas informações e casos de sucesso. É importante abordar este conteúdo com os/as alunos/as.

Seguem algumas referências para o assunto:

<https://www.unisoma.com.br/inteligencia-artificial-no-setor-de-energia/>

<https://energiahoje.editorabrasilenergia.com.br/novas-tecnologias-e-a-gestao-de-ativos-no-futuro/>

<https://www.byne.com.br/como-a-inteligencia-artificial-vem-promovendo-inovacao-em-diferentes-setores/>



3.4. Machine learning: Classificação de base de dados

O processo de classificação de base de dados é uma das etapas mais importantes na obtenção de informação relevante de dados. É através deste procedimento que é possível reduzir a dimensão e a quantidade de dados de um problema, o que consequentemente melhora o desempenho e reduz o esforço computacional.



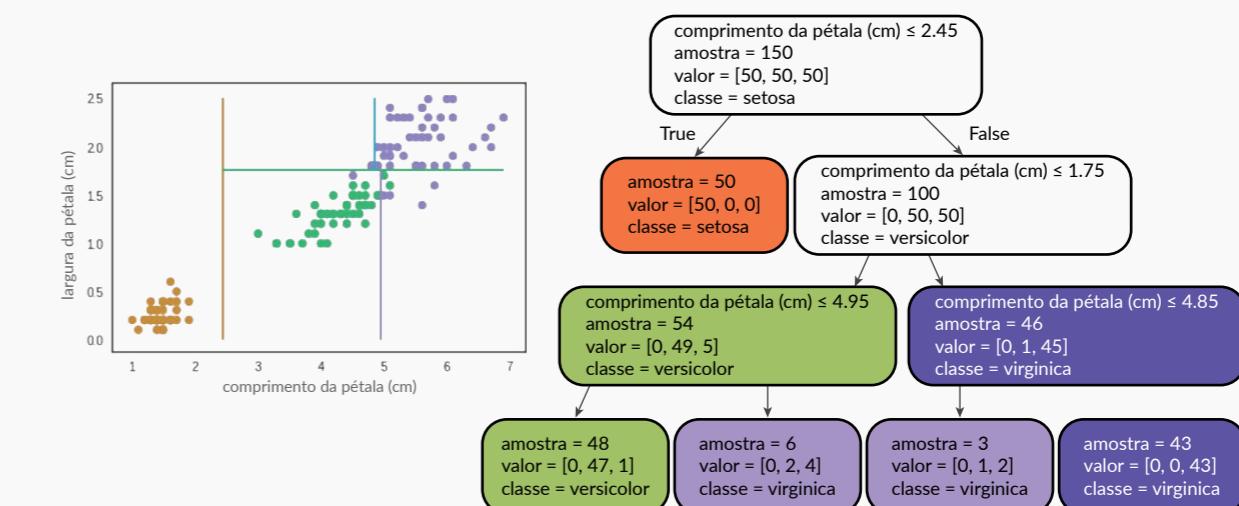
Dica: Dica: Professor/a, é importante orientar os/as alunos/as que o processo de classificação de base de dados e tratamento de dados acontecem de forma ciclica. Primeiramente trata-se os dados e logo é aplicado um algoritmo de classificação. Caso o resultado do algoritmo seja insatisfatório, pode-se escolher entre:

- Coletar mais dados (se possível)
- Aplicar um tratamento mais rigoroso nos dados
- Melhorar o modelo do algoritmo de classificação
- Substituir o modelo do algoritmo de classificação

No geral, existem dois grupos de classificadores de base de dados: classificadores supervisionados e classificadores não supervisionados. Os modelos de aprendizagem supervisionada são aqueles que necessitam de um *feedback* para melhorarem a sua capacidade de representar um conjunto de dados (um bom exemplo é uma RNA). É muito comum escutar termos como: dados “rotulados” ou “tagueados” no processo de treinamento destes modelos, que nada mais é que o dado e a sua identificação, como uma foto de um gato e um rótulo dizendo que a respectiva foto é de um gato. O tagueamento é uma tarefa que deve ser feita por humanos, por isso os modelos que dependem de dados rotulados são conhecidos como modelos de aprendizagem supervisionada.

Dentre os algoritmos de aprendizagem supervisionada encontra-se a árvore de decisões. A árvore de decisões é um classificador que faz o uso de um conjunto mínimo de perguntas para distinguir os dados. Na Figura 42, pode-se observar a árvore de decisões que foi utilizada para classificar os dados do famoso conjunto de flores Íris.

Figura 42. Exemplo de árvore de decisões⁴³



A árvore de decisão é muito utilizada em controladores não lineares de plantas industriais, pois é um modelo bastante simples que é capaz de tomar decisões de forma autônoma depois de treinado. Este tipo de algoritmo tem um problema bem nítido, se as características se diferem muito pouco entre os elementos que compõe a base de dados, ele pode ser ineficaz ou as ramificações podem crescer muito em tamanho.

Um outro tipo de algoritmo de *machine learning* simples com fundamentos em estatística é a classificação probabilística que se baseia no teorema de Bayes (Figura 43) para determinar uma probabilidade de um evento ocorrer ou não. É um modelo adequado para classificação de atributos discretos e, como é baseado em estatística, necessita que a frequência destes atributos na entrada seja conhecida.

Figura 43. Classificador de Naive Bayes⁴⁴

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Probabilidade
Probabilidade original da Classe
Probabilidade posterior
Preditor da probabilidade posterior

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

43. <https://medium.com/machine-learning-beyond-deep-learning/%C3%A1rvores-de-decis%C3%A3o-3f52f6420b69>

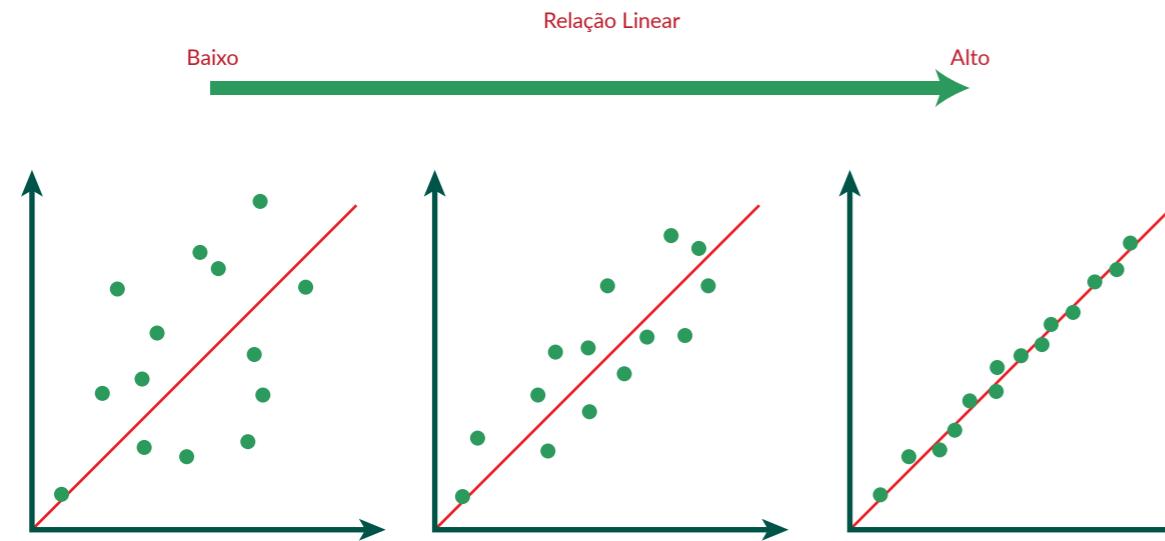
44. <http://inteligenciaartificialbrasil.blogspot.com/2016/09/6-passos-para-aprender-o.html>



Um bom exemplo de classificador de Bayes é a detecção de *spams* na caixa de e-mail. Para tal, o provedor do serviço de e-mail analisa o corpo da mensagem e verifica se no conteúdo existem palavras que são comumente encontradas em *spams*, exemplo: clique, tempo limitado, urgente, imperdível, desconto. O classificador de Bayes é treinado para determinar a frequência admissível de cada palavra, e no final ele é capaz de rotular a mensagem como sendo *spam* ou não.

Os modelos mais complexos de classificadores de base de dados começam a surgir com regressões que visam modelar o comportamento dos dados. O tipo de regressão mais simples para dados quantitativos é a linear. A regressão linear é uma abordagem probabilística que busca descrever como as entradas se relacionam com as saídas, encontrando a melhor reta que representa uma nuvem de pontos. Na Figura 44 é possível ter uma ideia do comportamento deste modelo, sendo a sua fidelidade aos dados reais dependente do quanto distribuídos os pontos estão. A regressão linear é uma ótima ferramenta para descrever tendências.

Figura 44. Exemplo de regressão linear⁴⁵

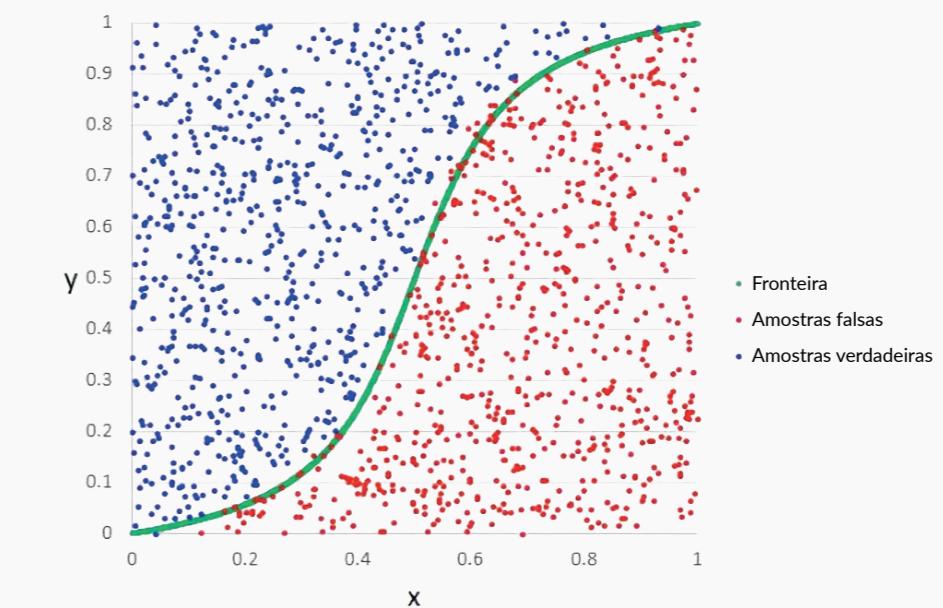


A regressão linear é muito aplicada como primeira abordagem na tentativa de descoberta de padrões em dados. Este modelo é simples e excelente para representar dados pouco dispersos e que seguem uma tendência. Caso o problema a ser modelado apresente mais de uma entrada, pode-se utilizar a regressão linear múltipla, que é uma abordagem matricial para a regressão linear simples. Um exemplo prático é comumente observado no universo dos negócios, quando é necessário prospectar preços, custos ou até mesmo consolidar um modelo que represente os lucros de uma empresa.

A regressão linear é um modelo analítico e funciona muito bem para dados numéricos, porém sabe-se que às vezes é necessário lidar com dados qualitativos, separados em “classes”. Neste caso, o tipo de modelo mais indicado é o de regressão logística. A regressão logística é responsável por medir a relação entre as variáveis de entrada e a resposta final. Essa regressão, normalmente é baseada no modelo exponencial logit. O comportamento da regressão logística pode ser observado na Figura 45.

45. <https://datatab.net/tutorial/linear-regression>

Figura 45. Exemplo de regressão logística⁴⁶



A regressão logística é muito utilizada na gestão de crédito no contexto de risco de crédito. Os modelos são preparados para gerar uma pontuação para cada cliente, a qual é baseada em dados históricos de quantia armazenada e gasta, inadimplência e outras variáveis qualitativas como idade, sexo e árvore genealógica. Desse modo, ao receberem a solicitação de crédito de um cliente que não possui um histórico de dados consolidados na instituição, é possível classificá-lo com base nos perfis conhecidos e atender ou refutar a solicitação do cliente, sem se expor a potenciais prejuízos.



Saiba mais: Existem muitos outros tipos de modelos de regressão. A regressão linear e a logística são os mais básicos. Estes modelos têm ampla aplicação na engenharia, e devem ser conhecidos. É recomendado que o/a professor/a aprofunde o tema com os/as discentes.

<https://iaexpert.academy/2020/06/08/os-tipos-de-redes-neurais/>

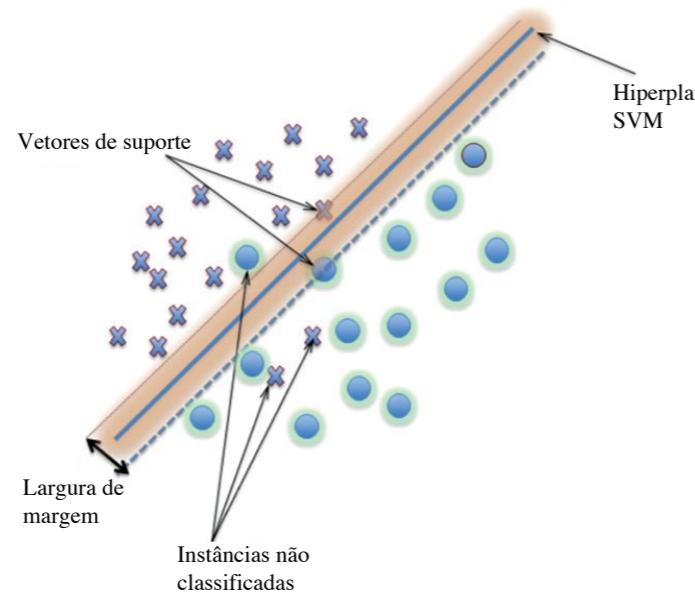
Os modelos de regressão são muito bons em representar um comportamento satisfatório para uma base de dados, porém eles não são capazes de separar dados em grupos, tal como uma árvore de decisão. Dado as limitações apresentadas pela árvore de decisões, existem outros algoritmos que são capazes de agrupar dados por suas semelhanças.

O algoritmo *Support Vector Machine* (SVM) é um algoritmo de classificação que visa separar um conjunto de dados em 2 conjuntos diferentes. Para que este algoritmo funcione corretamente, os seus dados de entrada devem ser linearmente separáveis, então o SVM encontra a reta que melhor separa o conjunto de dados. Esta abordagem é muito utilizada em classificações dicotômicas (separação entre 2 classes distintas e muito bem definidas), como pode ser observado na Figura 46.

46. <https://slidetodoc.com/introduo-ao-tensorflow-programa-de-psgraduaoo-em-cincia/>



Figura 46. Support Vector Machine⁴⁷

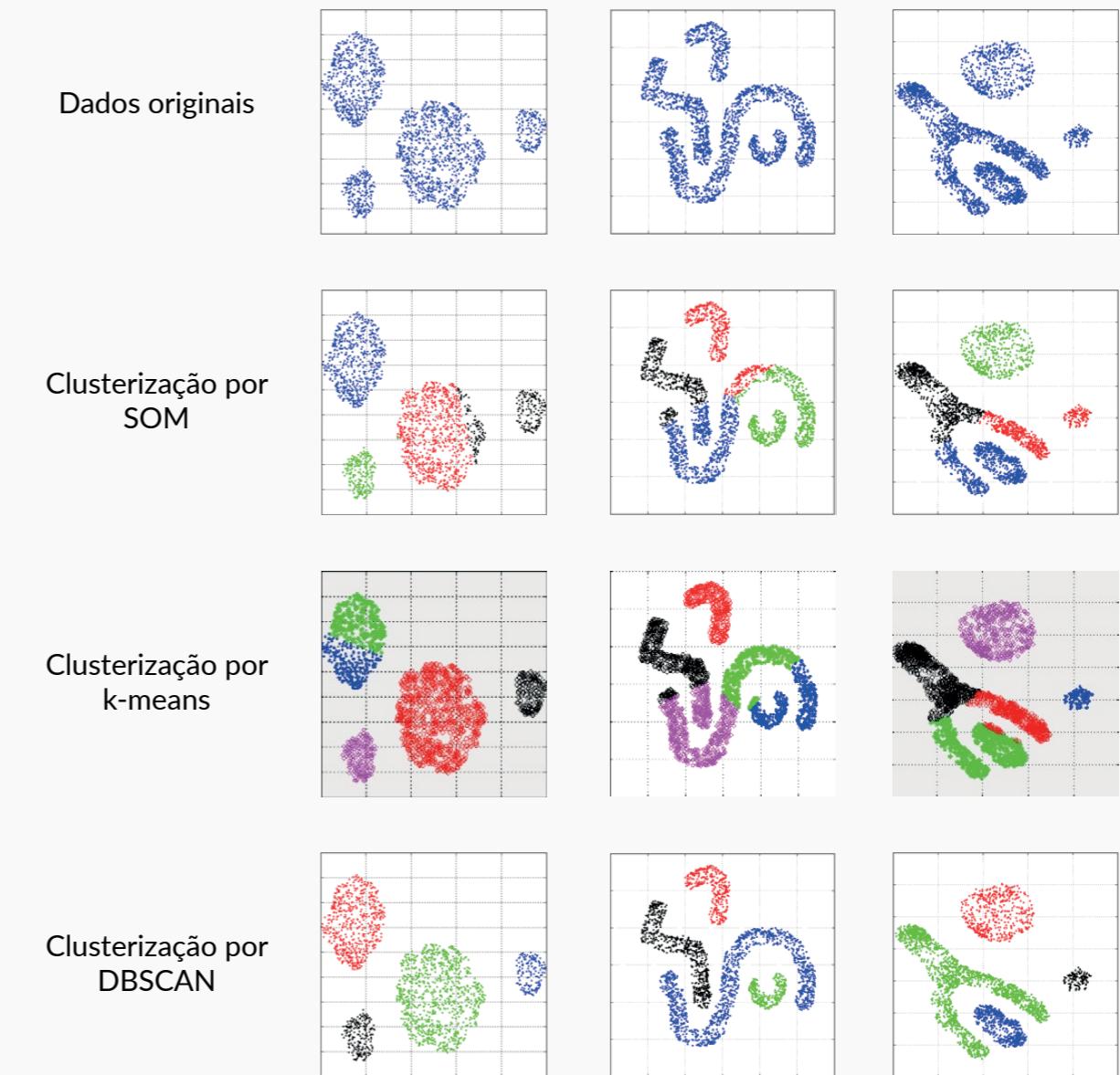


Os algoritmos baseados na técnica de SVM são muito utilizados em problemas de separação binária, ou seja, são muito bons para classificar a ocorrência de um defeito em um equipamento ou indicar uma tomada de decisão que contém somente duas opções (compra ou venda de ações, por exemplo). Alguns estudos têm buscado desenvolver modelos de SVM que possam ser utilizados no mercado livre de energia, indicando cenários de compra e venda em momentos estratégicos, visando maximizar o lucro da comercializadora.

Quando o assunto é separação de dados por características semelhantes, os modelos de aprendizagem não supervisionada aparecem em maioria. Estes modelos são aqueles que não necessitam de *feedback* e, tipicamente, após a primeira parametrização, são capazes de aprender com os dados de entrada e conseguem apresentar resultados satisfatórios.

Um algoritmo famoso, que possui muitas variações e é amplamente utilizado na primeira etapa de extração de conhecimento dos dados, é o algoritmo de clusterização, isso porque ele é capaz de identificar padrões e agrupar os dados em classes que podem, posteriormente, ajudar na modelagem do problema. A técnica de clusterização é responsável por agrupar dados que possuem características semelhantes, sendo que o resultado depende do número máximo de clusters escolhido pelo profissional que elaborou o algoritmo. Existem vários algoritmos de clusterização, como pode ser observado na Figura 47.

Figura 47. Clusterização de dados⁴⁸



Uma aplicação clássica da clusterização na engenharia elétrica é na escolha de curvas de consumo que representarão os consumidores de um dado trecho sem medição de uma distribuidora. Basicamente, em estudos de rede de distribuição, o consumo de cliente que não pode ser medido é aproximado como sendo o consumo dos outros clientes aos quais ele se assemelha.

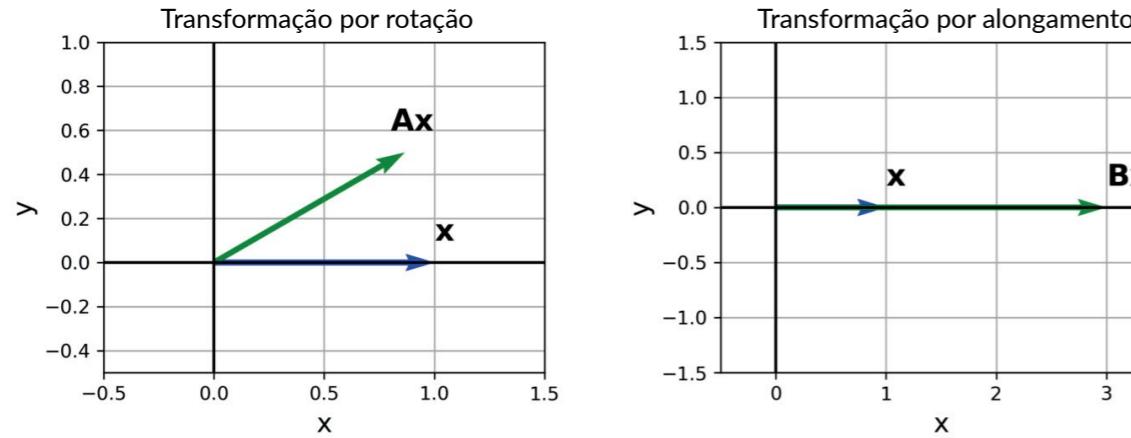
Uma outra técnica um pouco mais avançada que permite a classificação de valores em uma base de dados é a decomposição de vetores singulares (*Singular Value Decomposition - SVD*). O SVD é um método utilizado para promover a redução de dimensionalidade em conjunto de dados através de uma matriz de fatoração. É amplamente utilizado em processamento e *machine learning* aplicado a imagens. Um exemplo ilustrativo pode ser observado na Figura 48.

47. <https://www.sciencedirect.com/science/article/abs/pii/S0924271610001140>

48. https://www.researchgate.net/figure/Figura-2-Desempenho-de-Diferentes-Metodos-de-Clusterizacao-para-Dados-Espaciais-Fonte_fig2_269203176



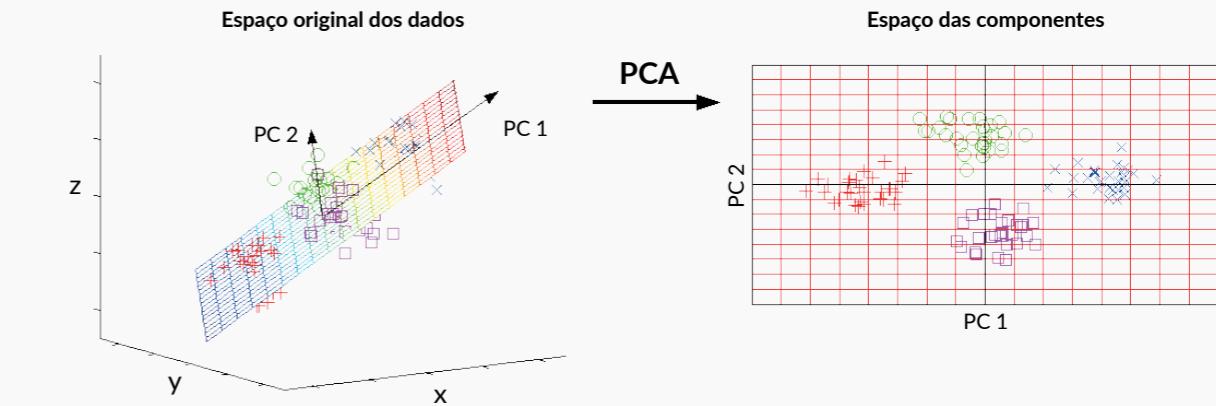
Figura 48. SVD⁴⁹



Os algoritmos baseados na técnica SVD são comumente utilizados na etapa de “tunning” em sistemas de controle. Esta etapa diz respeito à escolha das melhores variáveis para medição, manipulação e controle de uma planta industrial. Dessa forma, é possível otimizar as malhas de controle e obter respostas cada vez mais precisas e rápidas, sem precisar de uma análise custosa e demorada no início do projeto.

O SVD pode ser muitas vezes complexo de se implementar, e por este motivo surgiu o algoritmo de análises por componentes principais (*Principal Component Analysis – PCA*). O PCA é uma simplificação do SVD, pois aplica somente uma transformação ortogonal nos dados, buscando observações que apresentem variáveis linearmente independentes. Desta forma, é possível classificar um conjunto de dados, como pode ser observado na Figura 49.

Figura 49. PCA⁵⁰



O algoritmo PCA é muitas vezes utilizado na redução de dimensionalidades de problemas matemáticos, sem a perda de informação útil para a solução. Um exemplo de aplicação deste tipo de algoritmo é no planejamento de operação energética de médio prazo, o qual usa o PCA para reduzir a dimensionalidade de um modelo muito grande de programação dinâmica. Dessa forma, a solução do planejamento da operação é simplificada.



Futuro: *Machine learning* pode oferecer vantagem competitiva às empresas ao solucionar problemas e revelar percepções mais rapidamente e com mais facilidade do que a análise convencional. Saber como tratar os dados e interpretá-los é o primeiro passo para propor soluções arrojadas e adaptativas ao longo do tempo. Por este motivo, é muito importante que o/a professor/a enfatize a importância destes algoritmos não só para o setor elétrico, como também para etapas operacionais e financeiras.

Seguem algumas referências para o assunto:

<https://www.intel.com.br/content/www/br/pt/analytics/machine-learning/machine-learning-data-and-predictive-analytics.html>

<https://www.aquare.la/ia-na-negociacao-de-compra-e-venda-de-energia/>

49. <https://towardsdatascience.com/understanding-singular-value-decomposition-and-its-application-in-data-science-388a54beg95d>

50. <https://github.com/Joaocarabetta/PCA-Explorer>



3.5. Recapitulando

Ao final deste capítulo, o/a aluno/a deverá ser capaz de:

- Compreender a estatística básica e as principais etapas para o tratamento dos dados.
- Compreender o que são redes neurais artificiais, suas variações e aplicações.
- Compreender o conceito de rede neural de aprendizagem profunda e suas aplicações.
- Compreender os principais algoritmos de classificação de base de dados.
- Diferenciar algoritmos de aprendizagem supervisionada e não supervisionada.
- Compreender que os algoritmos de aprendizagem não supervisionada são excelentes para extrair informações iniciais dos dados.
- Conhecer os principais algoritmos de *machine learning*, sendo capaz de distinguir os eficazes em gerar modelos comportamentais dos mais eficazes em separar uma base de dados em grupos semelhantes.
- Compreender que os algoritmos de *machine learning* são muito empregados em etapas de tomada de decisão.

Atividades Sugeridas:

- Apresentações expositivas em slides com os temas: Principais algoritmos de *machine learning*, estatística básica voltada para o tratamento de dados, modelagem de sistemas e comportamentos utilizando aprendizagem supervisionada e RNAs, modelagem comportamental utilizando regressões, classificação de dados utilizando algoritmos de agrupamentos.
- Apresentação de vídeos para consolidação do conceito e novidades relacionadas aos temas abordados em aula:
 - https://www.youtube.com/watch?v=aircAruvnKk&list=PL_h2yd2CGtBHEKwEH5iqT-ZH85wLS-eUzv&index=3
 - https://www.youtube.com/watch?v=IHZwWFHWa-w&list=PL_h2yd2CGtBHEKwEH5iqT-ZH85wLS-eUzv&index=5
 - <https://www.youtube.com/watch?v=GBSTT5NBE4g>
 - https://www.youtube.com/watch?v=Ilg3gGewQ5U&list=PL_h2yd2CGtBHEKwEH5iqT-ZH85wLS-eUzv&index=5
- Promover atividades em grupo ou individuais, tais como seminários para avaliação do aprendizado. Sugestão de temas:
 - Aplicação dos algoritmos de *machine learning* no setor elétrico: manutenção preditiva e previsão de geração.
 - A importância de se tratar os dados antes de realizar estudos: casos em que a falta do tratamento de dados pode alterar os resultados.
 - Aplicação dos algoritmos de classificação de dados no setor elétrico: como é possível dar autonomia para os algoritmos tomarem decisões que antes eram feitas por especialistas do setor.
- Convidar os/as alunos/as para participarem de *webinar* e outros eventos disponibilizados pela instituição ou fora dela como forma de intensificar a sua capacitação.
- Convidar os/as alunos/as à se capacitarem na linguagem de programação Python e reproduzirem alguns dos algoritmos apresentados neste capítulo.

Algumas fontes interessantes:

- [Comparison of Deep Learning and the Classical machine learning Algorithm](#)
- [digital twin and big data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison](#)
- [Basics Statistics for Data Analysis](#)
- [A Tour of machine learning Algorithms](#)

Avaliação:

Comente brevemente sobre as características de um *Perceptron* e de uma rede neural artificial padrão. Cite os principais casos de uso e explique como é possível uma arquitetura tão simples aprender e gerar bons modelos. Discorra também brevemente sobre os modelos de clusterização e explique o porquê de sua importância.



Capítulo 4: Machine learning e big data aplicados ao setor elétrico



Nota para o/a professor/a – Este capítulo deve servir como base para o entendimento básico de como o machine learning e big data podem ser aplicados no setor elétrico. É importante nivelar os/as alunos/as na linguagem de programação Python (um entendimento de como manipular listas, dicionários e ser capaz de realizar operações matemáticas é o suficiente). Encoraje os/as discentes a se atualizar e aprofundar nos assuntos aqui apresentados, pois são muito pertinentes atualmente.

4.1. Big data no setor elétrico: Manipulação de dados para tomadas de decisão

As técnicas de big data aplicadas ao setor elétrico podem ser decisivas para o sucesso de um projeto de *Business Intelligence* (BI). Na realidade, um projeto de BI deve ter seu maior esforço voltado ao tratamento dos dados, pois este terá impacto direto nos modelos, que poderão ser utilizados posteriormente em setores de gestão de ativos, gestão comercial e de manutenção.

É muito importante vincular todos os procedimentos teóricos voltados para construção e análise de base de dados com problemas que o setor de energia enfrenta, pois dessa forma estudos voltados para BI podem ser realizados com maior frequência e exatidão. Uma solução tecnológica para o setor elétrico deve ser bem pensada para ser aderente às demandas que o setor possui. Existem diversos pontos de melhorias no setor e muito espaço para novas tecnologias.



Saiba mais: Os Sistemas Elétricos de Potência são responsáveis pelo suprimento de energia elétrica aos consumidores, tendo aumentado consideravelmente a sua complexidade ao longo dos anos.

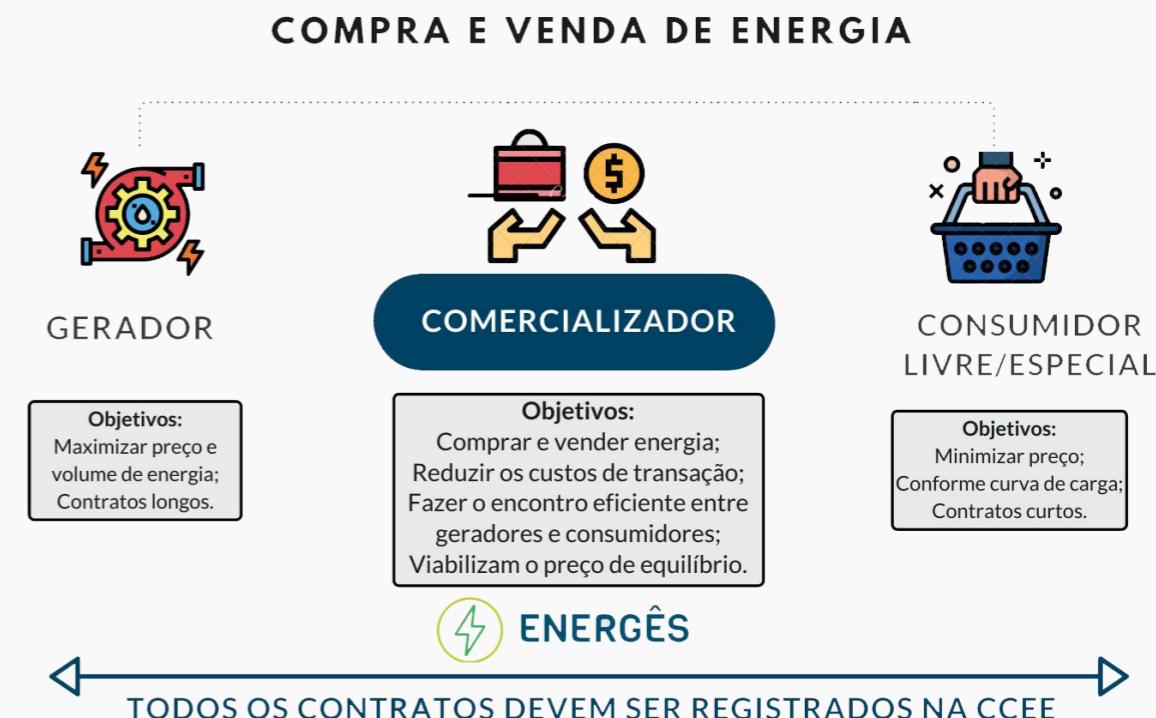
Devido às características de volume, velocidade e variedade associadas aos dados gerados pelas smart grids, diversos autores apontam a necessidade de aplicação de big data analytics para que os dados possam ser analisados e utilizados de forma efetiva. Esse processo é definido como a aplicação de técnicas avançadas de análise a conjuntos de dados associados ao conceito de big data. A partir desta aplicação, podem ser realizadas análises descritivas, diagnósticas, preditivas e prescritivas sobre os parâmetros de operação dos sistemas elétricos.

É recomendado que o/a professor/a incentive os/as alunos/as sobre a leitura de artigos científicos e teses sobre este assunto.

<https://bibliodigital.unijui.edu.br:8443/xmlui/bitstream/handle/123456789/6769/Ivan%20Eduardo%20Metz%20K%C3%B3Chne.pdf?sequence=1&isAllowed=y>

O primeiro passo a ser seguido após se identificar uma possibilidade de melhoria no setor e encontrar a potencial solução, é definir a estrutura de transferência e armazenamento de dados, se é uma solução pautada em geração, gestão e consumo de dados. A título de exemplo, a Figura 50 apresenta um relacionamento básico entre os geradores, comercializadores e consumidores do setor elétrico, com os seus respectivos objetivos.

Figura 50. Relações na compra e venda de energia⁵¹



De forma resumida, abaixo é apresentado alguns tipos de dados que cada parte tem que gerenciar e interpretar.

Os dados que o Gerador tem que lidar são tipicamente:

- Dados de medição do seu processo
- Gestão de manutenção
- Dados de previsões climáticas
- Dados do mercado de energia
- Gestão de contratos
- Previsão de geração

Os dados que o Comercializador tem que lidar são tipicamente:

- Dados de previsões climáticas
- Dados de consumo
- Gestão de pacotes de energia
- Dados do mercado de energia
- Dados dos clientes
- Previsão de preços

51. <https://energes.com.br/cartilha-do-setor-elettrico/>

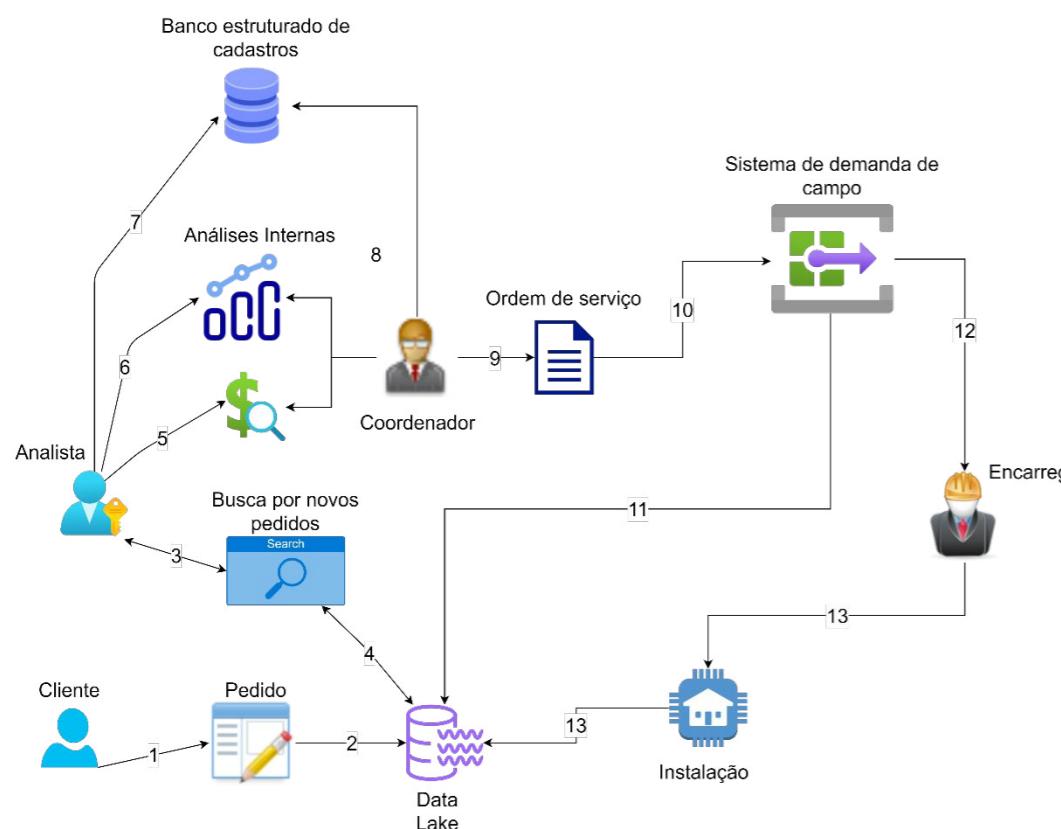


Os dados que o Consumidor Livre tem que lidar são tipicamente:

- Gestão de consumo do seu processo
 - Dados do mercado de energia
 - Dados dos clientes
 - Gestão de manutenção

A título de exemplo, considere uma visão micro sobre como uma concessionária de energia pode gerenciar um pedido de uma nova ligação de uma unidade consumidora (UC) dentro de sua área de concessão. Tipicamente, o cliente deve ir às instalações da concessionária ou pode pela internet realizar o seu pedido. Este pedido deve ser processado e, de alguma forma, a equipe de campo deve se organizar para que o religamento da UC seja realizado. Um trabalho imaginativo interessante de se fazer é o que acontece internamente dentro da concessionária, uma vez que o cliente faz uma solicitação de religamento. A Figura 51 ilustra uma possibilidade de gestão da informação.

Figura 51. Pedido de religamento da UC⁵⁴



Cada etapa da Figura 51 é descrita abaixo:

1. O pedido do/a cliente é enviado ao *data lake* da distribuidora em um formato de arquivo.
 2. O/A analista responsável periodicamente checa a existência de novos pedidos.
 3. A ferramenta de busca procura por novos pedidos no *data lake* e, ao encontrar um novo pedido, disponibiliza os dados para o/a analista.
 4. O/A analista então faz uma busca sobre a situação financeira do/a cliente que fez o pedido, religamento e gera um relatório.

5. O/A analista então faz um estudo para verificar a possibilidade e impactos que a nova ligação pode ter para o sistema de distribuição da empresa.
 6. O/A analista registra todos os dados que usou e gerou em um *data warehouse* mantido pelo seu departamento.
 7. O/A coordenador/a de área tem acesso a todo o trabalho do/a analista e pode utilizar o que foi gerado para a tomada de decisão.
 8. Ao verificar que está tudo dentro do esperado, o/a coordenador/a gera uma nova ordem de serviço.
 - 8.1. A ordem de serviço é postada no sistema que registra as demandas de campo
 9. O sistema então popula o *data lake* da distribuidora, com a nova ordem de serviço, uma vez que pode ser um caso de consulta posterior.
 10. O/A encarregado/a checa o sistema e verifica que existe um novo pedido de ligação
 11. O/A encarregado/a instala um medidor inteligente na unidade consumidora do/a cliente.
 12. O medidor inteligente começa a popular o *data lake* da concessionária com dados de medição.

É importante perceber que em uma estrutura organizada como a supracitada, o volume de dados que deve ser gerenciado pela concessionária só tende a aumentar. A cada novo pedido, novos estudos são feitos, uma nova ordem de serviço é gerada e mais dados de medição estão à disposição para gerar novos estudos e novas informações. Como existem milhares de UCs em uma área de concessão de uma distribuidora, milhões de dados podem ser gerados. Este é um exemplo de como o *big data* pode se expressar no setor elétrico.



 **Saiba mais:** Sabe-se que o Brasil enfrenta algumas dificuldades quando o assunto é viabilizar a implantação de uma tecnologia em larga escala no setor elétrico. Apesar disso, algumas distribuidoras têm se esforçado para conseguir digitalizar os seus sistemas. A Enel Distribuição São Paulo anunciou em 2021 que instalaria mais de 300 mil medidores inteligentes através de um programa de pesquisa e desenvolvimento.

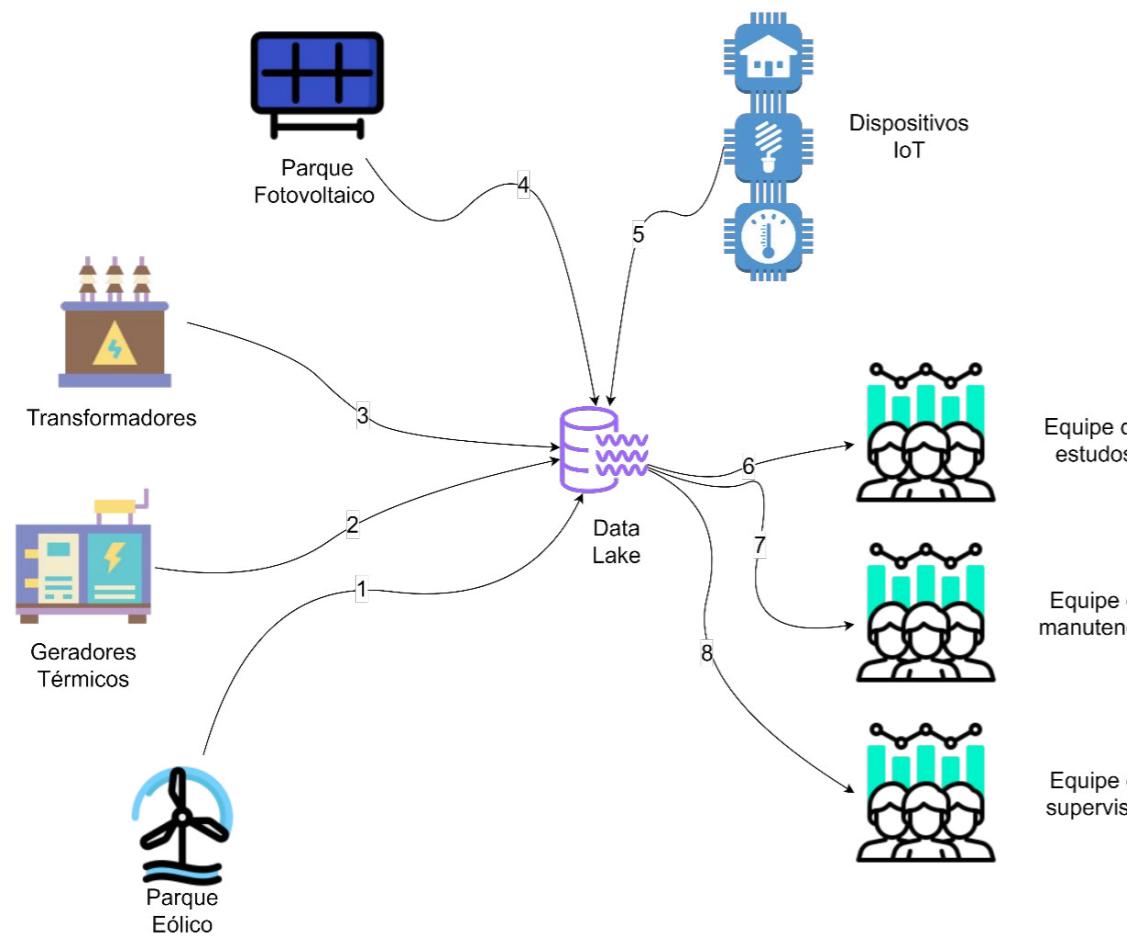
É recomendado que o/a professor/a incentive os/as alunos/as a buscarem notícias sobre os avanços tecnológicos no setor elétrico e como que os dados que as tecnologias disponibilizam podem melhorar os resultados de uma empresa.

<https://www.canalenergia.com.br/noticias/53162022/enel-sp-vai-instalar-300-mil-medidores-inteligentes-esse-ano>

Pode-se analisar também como é feita a gestão de dados de campo, que são gerados por sensores e dispositivos IoT. A Figura 52 uma possibilidade de tráfego de dados.



Figura 52. Dados de medição em data lake⁵³



Em 1, 2, 3, 4 e 5 os dados coletados podem ser de:

- Corrente nas fases
- Tensões nas fases
- Energia registrada
- Temperatura de operação

Em 6, 7 e 8 as equipes utilizam os dados armazenados para realizar estudos e gerar planos estratégicos para a supervisão, manutenção e controle dos ativos.

Na estrutura apresentada, observa-se que cada equipe tem a possibilidade de extrair dados brutos do *data lake*, construir seus *data warehouses* e realizar estudos independentes utilizando as mesmas fontes de dados. Esta estrutura de gestão de dados está cada vez mais sendo utilizada dentro das distribuidoras, pois adiciona flexibilidade e independência para cada equipe, mantendo o histórico dos dados intacto.

53. Autoria própria

4.2. Machine learning no setor elétrico: Manutenção preditiva e sistemas de previsão

A manutenção elétrica é fundamental para garantir o bom funcionamento de painéis, máquinas e equipamentos industriais. Por meio dela, é possível prevenir danos e corrigir diversos problemas, evitando acidentes que podem comprometer a segurança dos colaboradores ou paralisar a produção. A manutenção preditiva está associada às práticas de manutenção com caráter preventivo e antecipado de todo o aparato físico utilizado nas operações. Na prática, a manutenção preditiva é uma metodologia para a prevenção de danos e previsibilidade de falhas.

Para que a manutenção preditiva seja eficaz, é necessário traçar um plano que deve ser bem executado desde a aquisição dos dados de medição até a análise de dados que gerará o plano de ação final. Os ativos devem estar devidamente sensoriados e os dados gerados devem popular um *data lake* com todos os dados de medição. Posteriormente os dados devem ser processados e estruturados para permitir que os engenheiros utilizem ferramentas computacionais para identificar padrões e apontem possíveis condições de falha.

Pode-se dizer que as vantagens de se utilizar a manutenção preditiva no setor elétrico são:

- aumento da vida útil do equipamento;
- controle dos materiais (peças, componentes, partes, etc) e melhor gerenciamento;
- diminuição dos custos nos reparos;
- melhoria da produtividade da empresa;
- diminuição dos estoques de produção;
- limitação da quantidade de peças de reposição;
- melhoria da segurança;
- credibilidade do serviço oferecido;
- motivação do pessoal de manutenção;
- boa imagem do serviço após a venda, assegurando o renome do fornecedor.

Sendo sua principal limitação a dependência de dados confiáveis que representem exatamente as condições operacionais dos equipamentos.

Pode-se citar que as principais técnicas e dados de medição utilizadas para alimentar estudos e algoritmos de manutenção preditiva são:

- **Imagem termográfica do equipamento em operação:** utilizando um termovisor é possível converter a radiação infravermelha normalmente invisível emitida por um objeto em uma imagem monocromática ou multicolor que é representativa dos padrões térmicos em toda a superfície do objeto.
- **Dados de inspeção ultrassônica:** é uma técnica que detecta e isola sons de alta frequência que de outra forma seriam inaudíveis para o ouvido humano. É usado para diferenciar os ruídos de descarga elétrica do padrão normal de som.
- **Assinatura de corrente:** Analisa registro e dados atuais de medição de corrente, buscando encontrar padrões em que defeitos possam ocorrer.
- **Monitoramento de temperatura à distância:** Mede a temperatura com dispositivos IoT e através de modelos matemáticos encontra a distribuição de temperatura ao redor do objeto, buscando determinar a performance do equipamento.
- **Análise do óleo:** mede as porcentagens existentes de gases combustíveis específicos, tais como hidrogênio, oxigênio, metano, etano, etileno e acetileno, e indica combinações e proporções que indicam problemas potenciais.



Dica: A manutenção preditiva é uma área muito importante no setor elétrico. Existem muitos projetos e iniciativas que visam implantar a manutenção preditiva nos ativos de uma empresa.

Professor/a é muito importante nivelar o conhecimento dos/as alunos/as em relação a este assunto. Segue uma referência com os conceitos básicos de manutenção preditiva e detectiva.

https://www.osetoreletrico.com.br/wp-content/uploads/2010/03/Ed50_fasc_manutencao_industrial_cap3.pdf

Uma das vertentes da manutenção preditiva é utilizar algoritmos de *machine learning* para reconhecer padrões nas variáveis observadas que podem indicar a tendência de um defeito ocorrer. Os algoritmos de *machine learning* tem uma vasta aplicação no setor elétrico e podem ser aplicados também em atividades do mercado livre de energia, previsão de geração de energia renovável, previsão de consumo e até mesmo em otimização de rotas para atendimento de clientes.



Dica: Professor/a, a ANEEL disponibiliza em seu portal transparência todas as iniciativas de pesquisa e desenvolvimento que foram concluídas nos últimos anos desde 2008. Em uma rápida consulta neste portal, é possível encontrar muitos exemplos da utilização do machine learning no setor elétrico.

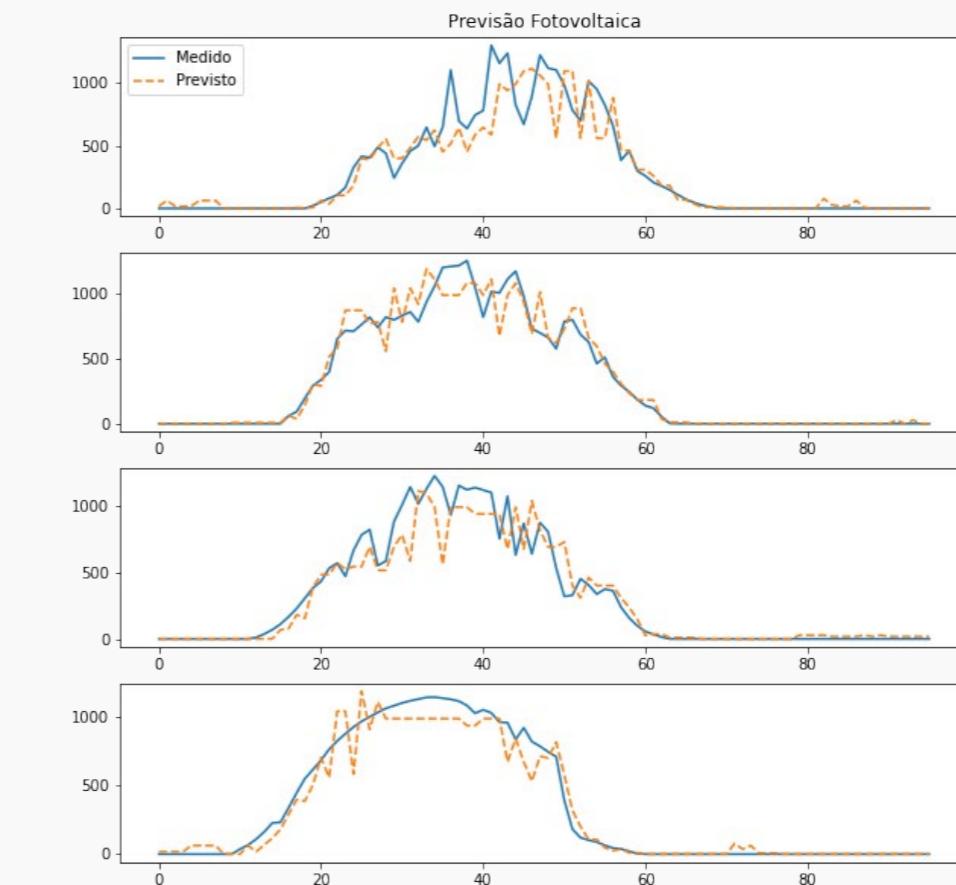
É importante trazer alguns dos assuntos para discussão em sala de aula com os/as discentes.

<https://dados.gov.br/dataset/projetos-de-p-d-em-energia-eletrica#:~:text=O%20Programa%20de%20P%26D%20tem,e%20usos%20finais%20de%20energia.>

Um exemplo de aplicação para os algoritmos de *machine learning* é na tentativa de detecção de consumidores irregulares na rede de uma concessionária de energia. Nos últimos anos, técnicas modernas de mineração de dados têm sido utilizadas para extrair conhecimento das bases de dados das distribuidoras, identificando possíveis padrões de consumo que possam estar relacionados a furtos ou fraudes de energia. Algoritmos são treinados com os dados históricos de consumo (tanto dados de consumidores regulares como de consumidores irregulares), gerando um modelo que pode ser posteriormente aplicado para classificar as unidades consumidoras a partir de seus dados mais recentes de consumo. Ou seja, após ser treinado, o modelo é capaz de apontar qual a probabilidade de uma unidade consumidora estar cometendo uma irregularidade.

Uma outra aplicação famosa destes algoritmos é na previsão de energia renovável, principalmente fontes eólicas e fotovoltaicas. Através de dados históricos de medição e dados de operação, é possível prever com uma boa assertividade o quanto uma planta de geração será capaz de produzir em energia elétrica. Esta é uma área que ainda existem muitos estudos sendo realizados. A Figura 53 mostra um exemplo de previsão de geração fotovoltaica utilizando um algoritmo de *machine learning*.

Figura 53: Previsão de geração fotovoltaica⁵⁴



As aplicações dos algoritmos de *machine learning* no setor elétrico são infinitas. A grande dificuldade é conseguir modelar o processo em termos computacionais para que eles possam representar com fidelidade os fenômenos que acontecem na realidade. Um outro impasse que o setor elétrico apresenta é a disponibilidade de dados estruturados e de qualidade. É comum algumas distribuidoras nem possuírem dados do processo que pretendem otimizar, precisando investir em uma etapa anterior, que é a de instrumentação e aquisição de dados. Esta etapa é custosa e muitas vezes desencoraja os investimentos no setor elétrico.



Futuro: A manutenção preditiva está presente em muitas organizações que precisam cuidar dos seus ativos. Isto porque ela permite um planejamento muito mais preciso sobre a manutenção preventiva dos equipamentos e consequentemente evita gastos desnecessários. Com a implantação do 5G e com o surgimento de tecnologias cada vez mais robustas e baratas, a tendência é que ocorra uma disseminação ainda maior da manutenção preditiva no setor elétrico, visando assim melhorias contínuas na disponibilidade e confiabilidade do serviço.

A aprendizagem profunda e outras formas de inteligência artificial produziram ferramentas e estratégias poderosas para a engenharia elétrica. Pesquisadores/as estão investigando como as redes neurais poderiam oferecer mais vantagens para controlar e melhorar o desempenho em sistemas dinâmicos. As redes neurais também podem ajudar a produzir com maior confiabilidade energia renovável. Uma das maiores desvantagens de confiar em fontes como a energia solar e eólica é que o sol nem sempre brilha e o vento nem sempre sopra. Isso torna a programação do sistema de energia vital para manter um abastecimento consistente. Uma rede neural pode informar o planejamento através do processamento rápido de grandes quantidades de dados para desenvolver previsões mais precisas da velocidade do vento ou da irradiação solar.

Professor/a, é importante enfatizar para os/as discentes que na engenharia, os avanços nas redes neurais proporcionam infinitas oportunidades de experimentação e de resolução criativa de problemas. Construir conhecimentos especializados em engenharia elétrica e informática pode ajudar os/as profissionais a descobrir novas formas de alcançar resultados com inteligência artificial e *big data*.

<https://www.plantservices.com/articles/2016/light-touch-predictive-maintenance-for-electrical-systems-part-1/>

<https://online.egr.msu.edu/articles/neural-networks-future-electrical-computer-engineering/>

4.3. Machine learning aplicado à modelagem da geração de energia



Dica: Professor/a, é importante orientar os/as alunos/as que se capacitem em alguma linguagem de programação especializada para análise de dados. Neste capítulo será utilizado Python, mas é importante ressaltar que a linguagem R também é muito utilizada.

Seguem algumas referências para o assunto:

<https://www.python.org/>

<https://www.devmedia.com.br/guia/python/37024>

<https://gabriel-schade-cardoso.gitbook.io/python-aprendendo-a-programar/>

Como visto anteriormente, existem infinitas possibilidades para a aplicação dos algoritmos de machine learning no setor elétrico. O objetivo deste capítulo é exemplificar como modelos de machine learning podem ser construídos com dados disponíveis na internet e que são acessíveis para

todas as pessoas. Professor/a, ressalta-se que o objetivo do exemplo é ilustrar para o/a aluno/a a aplicação dos algoritmos, desmistificar a implementação e fazer um comparativo inicial sobre o desempenho deles. Em nenhum momento o objetivo é obter algoritmos otimizados e comerciais.

O exemplo a seguir trata-se da aplicação de algoritmos de regressão linear, polinomial e de árvore de decisões, visando obter um modelo que possa ser utilizado como previsão de geração. Os dados utilizados foram obtidos da base de dados disponível no site da ANEEL.

Para a reprodução do exemplo você precisará:

- Ter acesso à internet
- Ter conhecimento básico em Python
- Ter conhecimento básico em *machine learning*
- Ter o Python instalado: <https://www.python.org/downloads/>
- Instalar as bibliotecas Python:
 - Virtualenv: pip install virtualenv
 - Pandas: pip install pandas
 - Numpy: pip install numpy
 - Matplotlib: pip install matplotlib
 - SciKit-Learn: pip install scikitlearn
 - Seaborn: pip install seaborn
- Baixar a base de dados de geração por fonte no link: https://drive.google.com/drive/folders/1RMxj4KY_q_EDfJIV5VIrxYHLiN3FaPQF
- Ter uma plataforma para reproduzir o código (recomenda-se o Visual Studio Code com o Jupyter Notebook instalado ou a ferramenta online Google Colab em <https://research.google.com/colaboratory/>).

O exemplo a seguir foi desenvolvido no Google Colab. Considere que cada quadrado escuro é uma célula. O código apresentado possui comentários explicativos que se caracterizam por um hashtag (#) seguido da explicação. Professor/a, destaca-se a importância de mostrar o caminho da modelagem, para que os/as alunos/as compreendam que a utilização de um algoritmo de machine learning precisa da etapa de obtenção de dados, tratamento de dados, treinamento do modelo, e análise de resultados.

```
# Import das bibliotecas usadas
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.ensemble import AdaBoostRegressor

df = pd.read_csv('./GeracaoFonte.csv') #carregando o arquivo csv que deve estar
#na mesma pasta que o arquivo .py
df.head() # Exibe as 5 primeiras linhas dos dados
```



ideGeracaoFonte	nomFonteGeracao	mdaEnergiaDespachadaGWh	mesReferencia	anoReferencia	dthProcessamento
0	1 Hidreleticas exclusive Itaipu	21074.749841	1	2000	NaN
1	2 Itaipu	6523.780292	1	2000	NaN
2	3 Oleo Diesel / Combustivel	112.026593	1	2000	NaN
3	4 Gas Natural	512.536500	1	2000	NaN
4	5 Carvao	772.954908	1	2000	NaN

```
df['nomFonteGeracao'].unique() # Mostra os nomes unicos na coluna que contém os nomes das fontes
```

```
array(['Hidreleticas exclusive Itaipu', 'Itaipu', 'Oleo Diesel / Combustivel', 'Gas Natural', 'Carvao', 'Eolicas', 'Biomassas', 'Nuclear', 'Residuos Processos Industriais', 'Energia produzida fora do SIN', 'Solar Fotovoltaica'], dtype=object)
```

```
# Remove dados desnecessários  
df = df.drop(["ideGeracaoFonte", "dthProcessamento"], axis=1)  
df.head()
```

nomFonteGeracao	mdaEnergiaDespachadaGWh	mesReferencia	anoReferencia
0 Hidreleticas exclusive Itaipu	21074.749841	1	2000
1 Itaipu	6523.780292	1	2000
2 Oleo Diesel / Combustivel	112.026593	1	2000
3 Gas Natural	512.536500	1	2000
4 Carvao	772.954908	1	2000

```
# Cria função que filtra por nome da fonte  
def filter_source_name(df, source_name):  
    return df[df['nomFonteGeracao'] == source_name]  
  
df = filter_source_name(df, "Hidreleticas exclusive Itaipu") # Atualiza a tabela com os dados filtrados para a fonte do argumento  
df.head()
```

nomFonteGeracao	mdaEnergiaDespachadaGWh	mesReferencia	anoReferencia
0 Hidreleticas exclusive Itaipu	21074.749841	1	2000
10 Hidreleticas exclusive Itaipu	20519.823366	2	2000
20 Hidreleticas exclusive Itaipu	21184.922627	3	2000
30 Hidreleticas exclusive Itaipu	7128.300200	4	2000
40 Hidreleticas exclusive Itaipu	21986.323500	5	2000

```
# Separa os dados anualmente, já que o conjunto de dados varia de 2000 a 2020 e transforma em um dicionário  
def split_years(df):  
    return {y:df[df['anoReferencia'] == y].drop(["anoReferencia"], axis=1) for y in df['anoReferencia'].unique()}
```

```
df_dict= split_years(df)
```

```
# A ideia é sempre usar o ano anterior para tentar prever o ano futuro  
# Aqui temos o modelo de REGRESSÃO LINEAR  
key_list= list(df_dict.keys())  
for k, v in df_dict.items(): # pega a chave k (ano) e o valor v (dados de geração)  
    # Se for o ano de 2019 ele não calcula mais, já que os dados de 2020 estão faltantes  
    if k==key_list[-2]:  
        break  
    # Define os valores de treinamento e reais  
    x_train= df_dict[k]["mesReferencia"].to_numpy().reshape(-1, 1)  
    y_train= df_dict[k]["mdaEnergiaDespachadaGWh"].to_numpy().reshape(-1, 1)  
    x_real= df_dict[k]["mesReferencia"].to_numpy().reshape(-1, 1)  
    y_real= df_dict[k+1]["mdaEnergiaDespachadaGWh"].to_numpy().reshape(-1, 1)  
    model= LinearRegression().fit(x_train, y_train) # Treina o modelo
```

```
# Plota o resultado  
plt.scatter(x_real, y_real, color= 'red', label= 'Dados Reais')  
plt.plot(x_train, model.predict(x_train), color= 'green', label= 'Modelo de Previsão', linewidth=1.0)  
plt.title('Regressão Linear')  
plt.xlabel('mês')  
plt.ylabel('Potência Gerada')  
plt.legend()  
plt.grid(True)
```



Figura 54. Regressão linear aplicado aos dados de geração da ANEEL⁵⁵

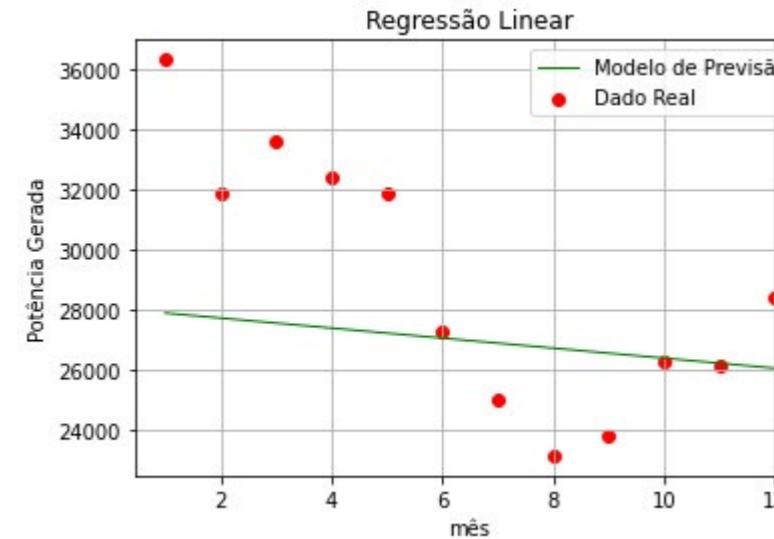
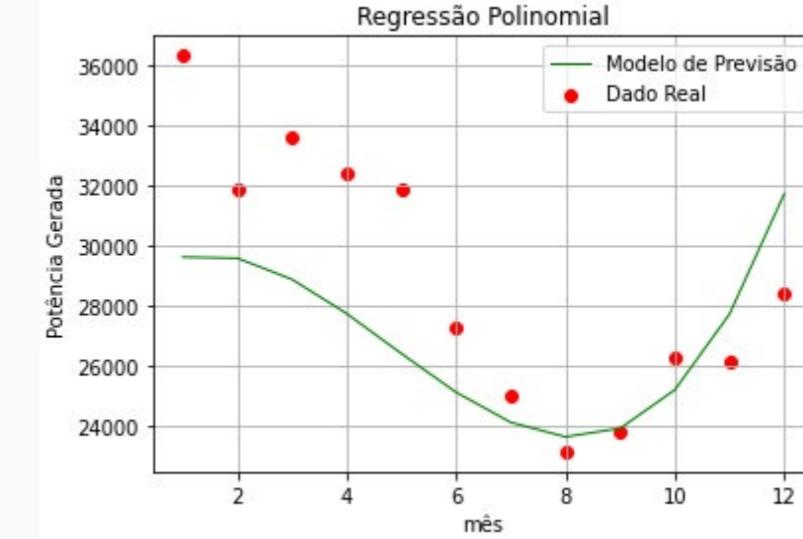


Figura 55. Regressão polinomial aplicada aos dados de geração da ANEEL⁵⁶



```
# A ideia é sempre usar o ano anterior para tentar prever o ano futuro
# Aqui temos o modelo de REGRESSÃO POLINOMIAL
key_list= list(df_dict.keys())
for k, v in df_dict.items(): # pega a chave k (ano) e o valor v (dados de geração)
    # Se for o ano de 2019 ele não calcula mais, já que os dados de 2020 estão
    # faltantes
    if k==key_list[-2]:
        break

    # Define os valores de treinamento e reais
    y_real_poly= df_dict[k+1][“mdaEnergiaDespachadaGWh”].to_numpy().reshape(-1, 1)
    x_real_poly= df_dict[k+1][“mesReferencia”].to_numpy().reshape(-1, 1)
    x_train_poly= df_dict[k][“mesReferencia”].to_numpy().reshape(-1, 1)
    y_train_poly= df_dict[k][“mdaEnergiaDespachadaGWh”].to_numpy().reshape(-1, 1)
    training_data_poly = PolynomialFeatures(3).fit_transform(x_train_poly)
    model_poly= LinearRegression().fit(training_data_poly, y_train_poly) # Treina
    o modelo

    # Plota o resultado
    plt.scatter(x_real_poly, y_real_poly, color= ‘red’, label= ‘Dado Real’)
    plt.plot(x_train_poly, model_poly.predict(training_data_poly), color= ‘green’,
    label= ‘Modelo de Previsão’, linewidth=1.0)
    plt.title(‘Regressão Linear’)
    plt.xlabel(‘mês’)
    plt.ylabel(‘Potência Gerada’)
    plt.legend()
    plt.grid(True)
```

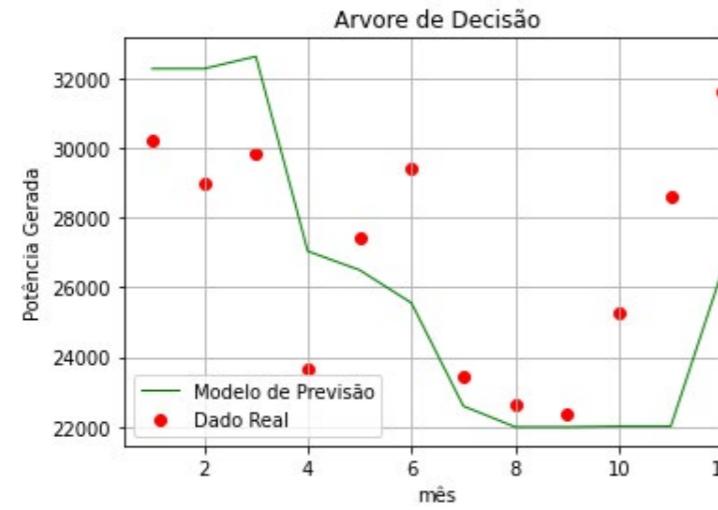
```
# A ideia é sempre usar o ano anterior para tentar prever o ano futuro
# Aqui temos o modelo de REGRESSÃO POR ARVORE DE DECISÃO
key_list= list(df_dict.keys())
for k, v in df_dict.items(): # pega a chave k (ano) e o valor v (dados de
    # geração)
    # se for o ano de 2019 ele não calcula mais, já que os dados de 2020 estão
    # faltantes
    if k==key_list[-2]:
        break

    # Define os valores de treinamento e reais
    y_real_tree= df_dict[k+1][“mdaEnergiaDespachadaGWh”].to_numpy().reshape(-1, 1)
    x_real_tree= df_dict[k+1][“mesReferencia”].to_numpy().reshape(-1, 1)
    x_train_tree= df_dict[k][“mesReferencia”].to_numpy().reshape(-1, 1)
    y_train_tree= df_dict[k][“mdaEnergiaDespachadaGWh”].to_numpy().reshape(-1, 1)
    model_tree = AdaBoostRegressor(n_estimators=10, learning_rate=0.05).fit(x_
    train_tree, y_train_tree) # Treina o modelo

    # Plota o resultado
    plt.scatter(x_real_tree, y_real_tree, color= ‘red’, label= ‘Dado Real’)
    plt.plot(x_train_tree, model_tree.predict(x_real_tree), color= ‘green’, label=
    ‘Modelo de Previsão’, linewidth=1.0)
    plt.title(‘Regressão Linear’)
    plt.xlabel(‘mês’)
    plt.ylabel(‘Potência Gerada’)
    plt.legend()
    plt.grid(True)
```



Figura 56. Árvore de decisão gerada a partir dos dados de geração da ANEEL⁵⁷



Os dados exibidos nas Figura 54, Figura 55 e Figura 56, são apenas previsões para o ano de 2019 usando como base o ano de 2018. Professor/a, você pode sugerir os/as discentes a reduzir o índice em `if k==key_list[-2] para if k==key_list[-3]` para visualizar os modelos para os outros anos.

Como pode ser observado, o modelo linear não teve nenhuma aderência com os dados. Os outros modelos tiveram uma aderência melhor, mas ainda não são ideais. Isso pode acontecer por três principais motivos:

- O conjunto de dados é muito pequeno em relação à necessidade de previsão, pois estamos usando os dados puros dos 12 anteriores meses para prever 12 meses posteriores.
- Considera-se que é um problema de entrada única, só estamos usando potência para prever a própria potência, e, isso seria muito assertivo, se e somente se existisse um padrão de comportamental e de amplitude. Pode até se perceber um padrão comportamental (no Início e no Fim do ano os despachos são maiores) para a usina de Itaipu (que é o caso do exemplo), mas a amplitude não se repete durante o ano.
- O primeiro ajuste dos parâmetros dos modelos pode não ser o ideal, é necessário fazer uma equalização dos parâmetros para encontrar a resposta que tem a melhor aderência ao problema.

4.4. Recapitulando

Ao final deste capítulo, o/a aluno/a deverá ser capaz de:

- Compreender a abordagem das técnicas de *big data* no setor elétrico.
- Compreender a importância da manutenção preditiva e como ela é aplicada ao setor elétrico.
- Compreender a aplicação de algoritmos de *machine learning* no setor elétrico e como estes podem ser utilizados em modelos de previsão.
- Citar casos de sucesso da aplicação de técnicas de *machine learning* e *big data* no setor elétrico.
- Conhecer as principais entidades, que promovem estudos e disponibilizam dados, sobre o tema de *machine learning* e *big data* no setor elétrico.

57. Autoria própria

Atividades Sugeridas:

- Apresentações expositivas em slides com os temas: Principais algoritmos de *machine learning* aplicados no setor elétrico; abordagem do *big data* no setor elétrico e o fluxo de informações; como o *machine learning* pode ser utilizado para prever a geração de fonte renovável.
- Apresentação de vídeos para consolidação do conceito e novidades relacionadas aos temas abordados em aula:
 - <https://www.youtube.com/watch?v=C1sKqenTO98>
 - <https://www.youtube.com/watch?v=KndHOYnJ5RE>
 - https://www.youtube.com/watch?v=My1_ttlLsgf
 - <https://www.youtube.com/watch?v=zA0OkLGZIJ8>
- Promover atividades em grupo ou individuais, tais como seminários para avaliação do aprendizado. Sugestão de temas:
 - Aplicação dos algoritmos de *machine learning* no setor elétrico: manutenção preditiva e previsão de geração.
 - A importância de se tratar os dados antes de realizar estudos: casos em que a falta do tratamento de dados pode alterar os resultados.
 - Aplicação dos algoritmos de classificação de dados no setor elétrico: como é possível dar autonomia para os algoritmos tomarem decisões que antes eram feitas por especialistas do setor.
- Convidar os/as alunos/as para participarem de webinar e outros eventos disponibilizados pela instituição ou fora dela como forma de intensificar a sua capacitação.
- Convidar os/as alunos/as a se capacitarem na linguagem de programação Python e reproduzirem alguns dos algoritmos apresentados neste capítulo.

Algumas fontes interessantes:

- [Maintenance Tools applied to Electric Generators to Improve Energy Efficiency and Power Quality of Thermoelectric Power Plants](#)
- [Data Mining Techniques for Transformer Failure Prediction Model: A Systematic Literature Review](#)
- [Past, present and future of Industry 4.0 - a systematic literature review and research agenda proposal](#)
- [Data Driven Predictive Maintenance of Distribution Transformers](#)
- [Industrial Internet of Things monitoring solution for advanced predictive maintenance applications](#)
- [A Scalable Predictive Maintenance Model for Detecting Wind Turbine Component Failures Based on SCADA Data](#)
- [machine learning Algorithms in Forecasting of Photovoltaic Power Generation](#)
- [Short-term load forecasting based on support vector machines regression](#)
- [Study of electricity load forecasting based on multiple kernels learning and weighted support vector regression machine](#)

Avaliação:

De acordo com o conteúdo visto neste capítulo e levando em consideração as leituras direcionadas, cite as principais aplicações da manutenção preditiva para prever defeitos em um transformador de uma subestação e de *machine learning* para a previsão da geração de uma planta fotovoltaica.



Casos de sucesso

Caso 1: Urban Futurability – ENEL SP.

Local: Brasil

Descrição: Urban Futurability é um projeto realizado pela Enel SP em parceria com a Fundação CERTI. O projeto abrange diversas frentes de atuação, dentre as quais temas relacionados a digitalização e *machine learning*, e conta com investimentos que somam R\$ 125 milhões.

Um resultado importante do projeto é a implantação de um Laboratório Vivo no bairro Vila Olímpia na cidade de São Paulo, planejado para ser transformado num bairro digital, inteligente e sustentável, beneficiando mais de 300 mil clientes na região.

Entre os destaques está a digitalização de 150 km da rede elétrica da Vila Olímpia. Com uma réplica digital tridimensional da rede, denominada *digital twin*, e instalação de sensores, a Enel está coletando informações sobre as condições do sistema para mapear possíveis falhas antes que ocorram, como a proximidade da vegetação na rede elétrica, qualidade dos equipamentos e da própria fiação.

Continue lendo em:

<https://www.abcdabc.com.br/abc/noticia/enel-sp-transforma-vila-olimpia-primeiro-bairro-digital-sp-138173>

Caso 2: Projeto de Internet das Coisas no segmento de distribuição – CPFL.

Local: Brasil

Descrição: A CPFL Energia, maior grupo privado do setor elétrico brasileiro, com 14,3% do mercado nacional de distribuição, escolheu a NEORIS, empresa de consultoria e TI, como parceira na implementação de um grande projeto de Internet das Coisas.

Com o intuito de reduzir custos, melhorar a qualidade das telemedições e ajudar no controle da distribuição de energia elétrica, a CPFL introduziu o projeto de *smart grid* Tauron, que começou a ser estruturado em 2012 e terminou de ser implementado em 2015.

O projeto Tauron foi dividido em três áreas: campo, automatização de medição da operação e telemedição.

Continue lendo em:

<https://www.convergenciadigital.com.br/Cloud-Computing/CPFL-Energia-adota-smart-grid-e-adota-cloud-privada-44915.html?=&UserActiveTemplate=site%2Csite%252Csite%25252Csite>

Caso 3: Inteligência Artificial Aplicada ao Setor Elétrico – EDP

Local: Brasil

Descrição: A EDP Brasil deu início, esta semana, a um projeto pioneiro de Pesquisa & Desenvolvimento para sistemas de inteligência artificial aplicados à área de distribuição de energia. Com investimento de R\$ 8,3 milhões ao longo de 18 meses, o projeto irá estudar os impactos da automação de processos com robôs de última geração.

Com o apoio da EY e de Doutores da Universidade de São Paulo (USP), a iniciativa visa a estruturação de um observatório de tecnologias disruptivas que analisará, entre outros temas relevantes, os impactos na formação e gestão da força de trabalho. A partir desse estudo, será possível identificar, medir e mitigar os aspectos socioeconômicos e culturais, ajudando a definir um novo perfil para os colaboradores e corpos diretivos das empresas diante das mudanças trazidas pela aplicação de novas tecnologias avançadas.

O estudo inédito no setor de energia permitirá o uso de robôs de primeira geração, que são capazes de automatizar processos manuais e repetitivos, liberando as pessoas para atividades mais analíticas e criativas. Na sequência, será realizado o mesmo estudo usando robôs de terceira geração e superior, com o uso de inteligência artificial, capazes de aprender a identificar “padrões” e tomar decisões.

Continue lendo em:

<https://www.edp.com.br/noticias/edp-e-pioneira-na-aplicacao-de-inteligencia-artificial-no-setor-eletrico>

Caso 4: Plataforma Digital de Gestão de Energia – AES Brasil

Local: Brasil

Descrição: O caso trata de um projeto que consiste no desenvolvimento de uma plataforma digital customizada para atender as demandas dos clientes do nosso Comercializador Varejista, com apresentação de dados em dashboards, criação de alertas e recomendações automáticas.

Como objetivo principal do projeto pode-se citar: Digitalização e automatização de processos internos para gestão energética de nossos clientes varejistas, com o intuito de elaborar um modelo de otimização utilizando aprendizado de máquina (*machine learning*) para oferecer ao mercado uma ferramenta de auxílio à tomada de decisão.

Continue lendo em:

<https://www.aesbrasil.com.br/pt-br/aes-brasil-lanca-plataforma-digital-energia>
<https://www.aesbrasil.com.br/pt-br/Inovacao>



Caso 5: Machine learning na Predição de Geração Eólica - DeepMind

Local: Reino Unido

Descrição: Em 2018, a empresa DeepMind, posteriormente adquirida pela Google, começou a aplicar algoritmos de aprendizado de máquina a 700 MW da capacidade de energia eólica do Google no centro dos EUA (700 MW é eletricidade suficiente para abastecer uma cidade de médio porte).

Usando uma rede neural que trabalha nas previsões meteorológicas disponíveis e nos dados históricos das turbinas, a solução implementada tornou possível prever razoavelmente a produção de energia eólica com 36 horas de antecedência.

Em apenas um ano, esses algoritmos de aprendizado de máquina aumentaram a produtividade da geração de energia eólica em cerca de 20%.

Continue lendo em:

<https://deepmind.com/blog/article/machine-learning-can-boost-value-wind-energy>

Caso 6: Machine learning aplicado para Confiabilidade da Rede – PG&E

Local: Estados Unidos

Descrição: Devido à experiências passadas com problemas com blecautes no sistema elétrico da PG&E (grande empresa de energia dos Estados Unidos), a empresa vem investindo cada vez mais na aplicação de técnicas de inteligência artificial em prol de melhoria da segurança energética.

A PG&E atualmente tem duas ferramentas de aprendizado de máquina para lidar com a confiabilidade da rede. A primeira ferramenta, “Localização de Falhas, Isolamento e Restauração de Serviços” (FLISR), está relacionada ao tratamento do tipo de problema que resultou num apagão que ocorreu em 2003. Esta ferramenta usa algoritmos para permitir que a rede seja “auto curável”, restaurando o serviço automaticamente após uma falha elétrica por meio de redirecionamento de fluxos de potência.

Essa ferramenta está em uso há vários anos e teve um efeito imediato. No futuro, a PG&E buscará tornar isso mais robusto por meio da instalação de sensores inteligentes nas linhas de transmissão para permitir reparos físicos mais rápidos e ações preventivas aprimoradas (por exemplo, corte de galhos de árvores de alto risco).

A segunda ferramenta, “Ferramenta de Sistema para Risco de Ativos” (STAR), é um modelo de pontuação de risco dinâmico para todos os ativos da PG&E, permitindo a priorização da substituição e manutenção de ativos. Essa capacidade de manutenção preditiva possui vasta aplicabilidade, desde grandes subestações até postes de energia individuais, em última análise, reduzindo custos e reduzindo ainda mais a frequência de interrupções. A manutenção preditiva é uma aplicação de aprendizado de máquina com popularidade crescente.

Continue lendo em:

<https://digital.hbs.edu>

Caso 7: Visão artificial e machine learning na rede de distribuição – Minsait

Local: Brasil

Descrição: Visando reduzir deslocamentos improcedentes e desnecessários, a empresa Minsait desenvolveu, em conjunto com a SP Energy Networks (SPEN), o iIdentify, solução que vai ajudar distribuidoras de energia a terem mais controle sobre o estado dos ativos de rede nas proximidades de seus clientes. Baseada em visão artificial, aplicando algoritmos de *machine learning* sobre a câmera dos clientes, a ferramenta pode impactar positivamente as operadoras brasileiras, gerando uma diminuição de cerca de 15% nos custos com deslocamentos e otimizar o tempo de atendimento aos clientes, segundo levantamentos da Minsait.

Por meio de um aplicativo a solução facilita a identificação de elementos na rede, capturando imagens por meio das câmeras de celular e processando-as por visão artificial.

Além disso, a análise do estado de equipamentos como fusíveis, painéis elétricos, disjuntores, postos de recarga de veículos elétricos, entre outros, auxiliará na detecção de áreas de alta demanda de energia e saber com antecedência o impacto no desempenho da rede para definir as ações necessárias que lhe permitam manter a estabilidade e evitar sobrecargas.

Já em utilização na SP Energy Works, o iIdentify espera reduzir até 10 mil dos 80 mil deslocamentos anuais que a companhia realiza.

Continue lendo em:

<https://tiinside.com.br/04/03/2021/minsait-utiliza-visao-artificial-e-machine-learning-para-reduzir-deslocamentos-improcedentes-na-distribuicao-de-energia/>



Entidades a consultar sobre o tema *machine learning* e *big data* no setor elétrico

Neste tópico são apresentadas algumas entidades relevantes para consultas sobre o tema de *machine learning* e *big data* aplicados no setor elétrico, onde pode-se encontrar, por exemplo: informações técnicas e científicas, estudos de casos e aplicações, tendências, rede de parceiros, entre outros itens.

IEEE CIS – IEEE Computational Intelligence Society (Mundo)

Acesso: <https://cis.ieee.org/>

Atuação:

- O IEEE CIS é uma organização vinculada ao IEEE (Institute of Electrical and Electronics Engineers) que se posiciona como líder internacional em pesquisas interdisciplinares e inovadoras, transferência de tecnologia, aplicações do mundo real e educação em Inteligência Computacional.
- Abrange diversos temas, entre eles: *big data*, *Machine* e *Deep Learning*, Técnicas de Inteligência artificial, Redes Neurais, etc.
- Organiza e dissemina publicações técnicas e científicas, congressos, eventos, capacitações e networking de especialistas e interessados no em inteligência computacional.

Por que consultar: Por meio da IEEE CES é possível acessar o estado-da-arte em termos de conhecimento, avanços e estudos de casos de aplicações de *big data* e *machine learning*, com muitas referências voltadas para o setor elétrico.

MIT.EI – Massachusetts Institute of Technology Energy Initiative (EUA)

Acesso: <https://energy.mit.edu/>

Atuação:

- MIT.EI é um centro do MIT para pesquisa, educação e divulgação de temas relacionados à geração, distribuição e consumo de energia.
- A missão no MIT.EI é desenvolver soluções com baixo ou zero carbono, que atendam com eficiência as necessidades globais de energia, minimizando os impactos ambientais e mitigando as mudanças climáticas.
- Dentre os diversos conteúdos e áreas englobadas na atuação do MIT.EI, os assuntos relacionados a *machine learning* e *big data* possuem grande destaque, com número crescente de pesquisas e desenvolvimento de soluções sendo realizadas.

Por que consultar: O MIT.EI reúne e disponibiliza uma ampla base de dados, publicações de elevado nível e pesquisadores e especialistas consultores, que podem ser acessados para trabalhos nas áreas de *machine learning* e *big data* para o setor elétrico, dentre outros temas correlatos.

Fechamento

A digitalização do setor elétrico é um tema muito importante e ainda muito incipiente, sendo que existe muito espaço para novas tecnologias e inovação, desde a camada de gestão inteligente de dados até sofisticados dispositivos IoT plugados em equipamentos da rede e algoritmos de aprendizado de máquina, visando automatizar tarefas, prever condições operativas e falhas em equipamentos.

Do ponto de vista de gestão eficiente de dados, é muito importante entender o básico sobre processamento de dados, devido ao grande volume de informação gerada. É tão importante quanto saber as tecnologias empregadas na era do *big data* e do *cloud computing*, pois estas soluções visam facilitar o processo árduo e custoso de lidar com grande quantidade de dados, como a existente no setor elétrico.

Os algoritmos de *machine learning* desempenham um importante papel, principalmente quando aplicados à manutenção preditiva e previsão de condições operativas para a geração e consumo. Com estes modelos é possível especular, com um pouco menos de risco, o que acontecerá com o mercado de energia nos próximos dias ou se um importante transformador de uma subestação pode estar prestes a falhar, por exemplo.

Além disso, profissionais desse setor precisam se manter atualizados sobre os assuntos abordados neste ebook, pois é um campo de possibilidades extremamente vasto e que está em constante modernização. Na corrida do desenvolvimento de tecnologias, qualquer algoritmo que consegue ter um resultado 1% melhor tem o potencial de ser extremamente lucrativo.

É nesse sentido que esse ebook se mostra importante ao compilar os conteúdos mais relevantes a serem passados por docentes/as de disciplinas relacionadas a essa temática. Espera-se que, com disciplinas atualizadas, estudantes de cursos que têm interação com o setor elétrico finalizem seus cursos com conhecimentos mais aderentes aos exigidos pelo atual mercado de trabalho.



Glossário

AWS	<i>Amazon Web Services</i>	NoSQL	<i>No Structured Query Language</i>
5G	Quinta geração de transmissão para Telefones Celulares	ONS	Operador Nacional do Sistema Elétrico
ANEEL	Agência Nacional de Energia Elétrica	P&D	Pesquisa e Desenvolvimento
API	<i>Application Programming Interfaces</i>	Payback	Tempo de retorno de Investimento
Asset Twins	Gêmeos ativos	PCA	<i>Principal Component Analysis</i>
BDO	Boletim Diário da Operação	PDA	Plano de Dados Abertos
BI	<i>Business Intelligence</i>	PDE	Plano Decenal de Expansão
Big Data	Grande Quantidade de Dados	PIB	Produto Interno Bruto
BSO	Boletim Semanal da Operação	PME	Pequenas e Médias Empresas
CCEE	Câmara de Comercialização de Energia Elétrica	Process Twins	Gêmeos de processo
Cloud Computing	Computação da Nuvem	PRODIST	Procedimentos de Distribuição de Energia Elétrica
Data Lake	Repositório de Dados	PwC	PricewaterhouseCoopers
Data Warehouse	Repositório de Dados Organizado	RAM	<i>Random Access Memory</i> – Memória de Acesso Aleatório
Digital Twin	Gêmeo Digital	RED	Recursos Energéticos Distribuídos
DPO	<i>Data Protection Officer</i>	RMS	<i>Root Mean Square</i>
EPE	Empresa de Pesquisa Energética	RNA	Rede Neural Artificial
ETL	<i>Extract, Transform and Load</i>	RNC	Rede Neural Convolucional
EUGDPR	<i>Europe General Data Protection Regulation</i>	RNP	Rede Neural Profunda
GDPR	<i>General Data Protection Regulation</i>	SCDE	Sistema de Coleta de Dados de Energia
HD	<i>Hard Disc</i> - Disco Rígido	SG	<i>Smart Grid</i> – Rede Elétrica Inteligente
IoT	<i>Internet of Things</i> - Internet das Coisas	SGBD	Sistema de Gerenciamento de Banco de Dados
IPDO	Informativo Preliminar Diário da Operação	SIN	Sistema Interligado Nacional
JSON	<i>JavaScript Object Notation</i> – Notação para objetos em Javascript	SMF	Sistema de Medição para Faturamento
KPI	<i>Key Performance Indicators</i>	SQL	<i>Structured Query Language</i>
kWh	Quilowatt-hora	SSD	<i>Solid State Drives</i> – Disco de Estado Sólido
LGPD	Lei Geral de Proteção de Dados	SVD	<i>Singular Value Decomposition</i>
M2M	<i>Machine to Machine</i> – Máquina para Máquina	UC	Unidade Consumidora
Machine Learning	Aprendizado de Máquina	Unit Twins	Gêmeos do sistema
MUST	Montante de Uso do Sistema de Transmissão	WEF	<i>World Economic Forum</i> – Fórum Econômico Mundial
NIST	National Institute of Standards and Technology		



Referências

- [1] MORETTIN, Pedro Alberto. Estatística básica. 9. ed. São Paulo, SP: Saraiva, 2017. E-Book (554 p.)
- [2] TRIOLA, Mario F. Introdução à estatística. 12. ed. Rio de Janeiro: LTC - Livros Técnicos e Científicos, 2017. E-book (812 p.)
- [3] Montgomery, Douglas. C.; Runger, George C. Estatística Aplicada e Probabilidade para Engenheiros. 4^a Ed. Rio de Janeiro. LTC, 2009. 652p.
- [4] Hair Jr., Joseph F.; et al. Análise Multivariada de Dados. Porto Alegre. Bookman, 2009. 688p.
- [5] FACELI, Katti et al. Inteligência artificial: uma abordagem de aprendizado de máquina. Rio de Janeiro, RJ: LTC, 2011. XVI, 378 p.
- [6] Dorian Pyle. Data Preparation for Data Mining. The Morgan Kaufmann Series in Data Management Systems. 1999. ISBN-10: 1558605290
- [7] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems, 2011. ISBN-10: 9780123814791
- [8] SMART GRIDS: O QUE É E COMO FUNCIONA ESSA INOVAÇÃO EM DISTRIBUIÇÃO DE ENERGIA ELÉTRICA, <https://www.neoenergia.com/pt-br/te-interessa/inovacao/Paginas/smart-grid-o-que-e-como-funciona-inovacao-em-distribuicao-de-energia-eletrica.aspx>, visualizado em 28 de agosto de 2021
- [9] Digitalização da rede elétrica, <https://certi.org.br/blog/digitalizacao-da-rede-eletrica/>, visualizado em 28 de agosto de 2021
- [10] Um debate sobre a digitalização no setor elétrico, <https://www.ecomergencia.com.br/blog/no-giro-energia-um-debate-sobre-a-digitalizacao-no-setor-eletroico/>, visualizado em 28 de agosto de 2021
- [11] Digitalização revoluciona o sistema energético brasileiro, <https://new.siemens.com/br/pt/empresa/stories/energia/digitalizacao-revoluciona-sistema-energetico-brasileiro.html>, visualizado em 28 de agosto de 2021
- [12] A era digital já é realidade no setor elétrico, <https://www.energiaquefalacomvoce.com.br/2019/04/22/artigo-a-era-digital-ja-e-realidade-no-setor-eletroico/>, visualizado em 29 de agosto de 2021
- [13] Novas tecnologias colocam mais pressão para mudanças na regulação de energia, <https://www.blocknews.com.br/corporativo/novas-tecnologias-colocam-mais-pressao-para-mudancas-na-regulacao-de-energia/>, visualizado em 29 de agosto de 2021
- [14] Innovation Process in the Brazilian Electric Sector, http://www.gesel.ie.ufrj.br/app/webroot/files/publications/07_castro.pdf, visualizado em 29 de agosto de 2021
- [15] The Future of Electricity New Technologies Transforming the Grid Edge, http://www3.weforum.org/docs/WEF_Future_of_Electricity_2017.pdf, visualizado em 30 de agosto de 2021
- [16] Procedimentos de rede Submódulo 25.2 Apuração dos dados, http://www.ons.org.br/-2FProcedimentosDeRede%2FM%C3%B3dulo%2025%2FSubm%C3%B3dulo%2025.2%2FSubm%C3%B3dulo%2025.2_Rev_1.1.pdf, visualizado em 30 de agosto de 2021
- [17] Procedimentos de rede Submódulo 6.17 Coleta de dados de medição para faturamento, http://apps08.ons.org.br/ONS.Sintegre.Proxy/ecmprsite/ecmfragmentsdocuments/Subm%C3%B3dulo%206.17-RS_2020.12.pdf, visualizado em 30 de agosto de 2021
- [18] Digital Twin: entenda o impacto na indústria e na preservação dos seus ativos, <https://kotengenharia.com.br/digital-twin-impacto-na-industria-e-preservacao-de-ativos/>, acesso em 13 de setembro de 2021
- [19] Principais Algoritmos de Machine Learning para você conhecer em 2022, <https://insightlab.ufc.br/principais-algoritmos-de-machine-learning-para-voce-conhecer-em-2022>, acesso em 28 de julho de 2022
- [20] Capítulo 10 – As 10 Principais Arquiteturas de Redes Neurais, <https://www.deeplearningbook.com.br/as-10-principais-arquiteturas-de-redes-neurais/>, acesso em 14 de setembro de 2021
- [21] Segredos da estatística, <https://segredosdaestatistica.wordpress.com/>, acesso em 16 de setembro de 2021
- [22] Outlier: o ponto fora da curva. <https://medium.com/ensina-ai/outlier-o-ponto-fora-da-curva-1f28f3d9c23>, acesso em 16 de setembro de 2021
- [23] Bussab, WO; Morettin, PA. Estatística Básica. São Paulo: Editora Saraiva, 2017 (9^a Edição).
- [24] Plano de Dados Abertos 2020-2022, <https://dadosabertos.aneel.gov.br/>, visualizado em 28 de julho de 2022
- [25] Como machine learning e big data se relacionam, <https://gatefy.com/pt-br/blog/como-machine-learning-e-big-data-se-relacionam/>, acesso em 18 de setembro de 2021
- [26] WEG utiliza inteligência artificial para planejar manutenção preditiva em aerogeradores, <https://www.weg.net/institutional/BR/pt/news/produtos-e-solucoes/weg-utiliza-inteligencia-artificial-para-planejar-manutencao-preditiva-em-aerogeradores>, acesso em 18 de setembro de 2021
- [27] Inteligência artificial inicia a revolução do setor elétrico, <https://www.dci.com.br/servicos/energia/inteligencia-artificial-inicia-a-revolucao-do-setor-eletroico/14308/>, acesso em 18 de setembro de 2021



Respostas das avaliações

Avaliação do capítulo 1

Como foi visto, uma *smart grid* é definida pelo grau de automação, tecnologias e recursos energéticos distribuídos na rede, permitindo que ela tenha certa autonomia e que às vezes opere desacoplada do sistema principal. Para elevar o nível de automação e medição, os dispositivos IoT entram em peso, instrumentando equipamentos e dando controle para os operadores reconfigurarem o ponto operacional da rede remotamente. Uma vez que os dispositivos IoT são responsáveis também por coletar dados. Estes dados devem ser armazenados em algum local para análise posterior e visualização, sendo comumente utilizados serviços de computação em nuvem para fazer este papel. Além disso, é muito comum fornecedores de dispositivos IoT terem a sua própria nuvem para prestar serviços.

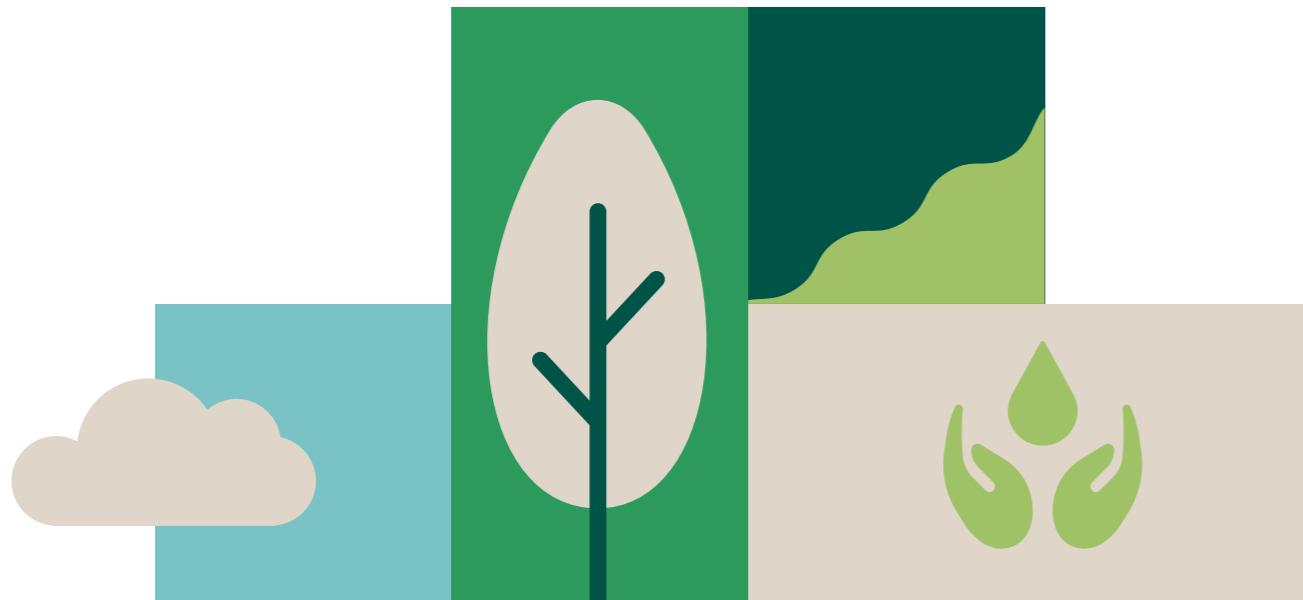
Com tantos dados disponíveis, é possível gerar *digital twins* da rede em operação, utilizando computação gráfica e processamento eficiente de dados. Desta forma, os especialistas podem inspecionar as condições operativas de equipamentos com mais precisão. Os dados também abrem portas para a manutenção preditiva, que vem ganhando força no setor elétrico, pois ajuda a manter a confiabilidade e a disponibilidade dos sistemas de geração, transmissão e distribuição de energia. Vale ressaltar que os modelos de *machine learning* empregam um papel importantíssimo na manutenção preditiva e na análise posterior dos dados de campo ou de processos.

Avaliação do capítulo 2

Basicamente, os bancos de dados nasceram para lidar com grandes volumes de dados. Eles apresentam sistemas eficientes de busca nos quais o tempo de execução não cresce exponencialmente com a quantidade de dados disponível. Uma vez entendido isso, existem bancos de dados estruturados (tabelados) e não estruturados, nos quais as aplicações são especificadas de acordo com o tipo de dados disponíveis: quando existem dados que possuem características imutáveis e bem definidas utiliza-se um banco de dados SQL, ao passo que o NoSQL permite dados com característica mutáveis e sem estrutura.

Ainda falando sobre armazenamento de dados, é importante ressaltar as estruturas de armazenamento genéricas que lidam com um volume gigantesco de dados: Os *data lakes* e os *data warehouse*. Estas estruturas nada mais são que repositórios de dados de uma organização, que armazenam dados de todas as fontes. Os *data lakes* aceitam qualquer tipo de dado, formato ou estrutura, ao passo que os *data warehouses* requisitam dados estruturados e pré-tratados.

Uma vez que o volume de dados armazenados e que trafega tende a crescer cada vez mais, *hardware*s específicos, robustos e de alta performance são requisitados. Para lidar com essa problemática, pode-se utilizar a *cloud computing* em modo privado ou híbrido, permitindo assim um processamento de dados ágil e o armazenamento seguro dos dados. Estes dois modelos permitem uma camada de segurança maior e garantem que somente os setores da empresa possam ter acesso aos dados no banco.





Avaliação do capítulo 3

O *Perceptron* é a estrutura mais básica de uma rede neural. Basicamente ele é composto por uma interface de entrada, na qual os dados (uma constante e um peso) são recebidos, representando o quão aquele dado é importante e uma função de ativação para a saída, que representa se o dado é relevante ou não para o caso que está sendo analisado.

Combinando vários *Perceptrons*, é possível construir modelos que consigam gerar resultados com maior precisão. A combinação dos *Perceptrons* é conhecida como uma rede neural artificial, a qual precisa passar por um treinamento que irá equalizar os pesos, utilizando dados rotulados. Normalmente, é importante ter um conjunto de dados amplo e confiável para treinar esses modelos de redes neurais, pois eles são extremamente dependentes da qualidade dos dados e quanto os dados realmente são capazes de representar todos os casos reais.

Existem também, outros modelos de aprendizagem de máquina que não são baseados no *Perceptron* e não precisam passar por treinamentos com dados rotulados. Estes modelos são conhecidos como de aprendizagem não supervisionada e, essencialmente, são capazes de encontrar padrões dentro de um conjunto de dados e apontar as características mais marcantes dos dados em cada conjunto. Este tipo de abordagem é excelente para etapas de pré-processamento e rotulação de dados, pois é onde a qualidade dos dados é analisada e potenciais características que podem ser usadas como rótulos aparecem.



Avaliação do capítulo 4

Os transformadores são uns dos equipamentos mais importantes de uma subestação e, por isso, possuem um alto nível de instrumentação e sensoriamento, com a finalidade de coletar dados e observar minuciosamente todas as suas condições operativas.

Com muitos dados coletados, é possível treinar alguns modelos de aprendizagem de máquina para interpretar e prever o funcionamento de um dado transformador. Por exemplo: com dados de temperatura interna e externa, pressão, tensão, corrente e informação harmônica é possível treinar um modelo de *machine learning* com um viés de manutenção preditiva, que observa essas variáveis em tempo (real ou não) e aponta em quanto tempo será necessária a próxima manutenção preventiva do equipamento, ou se existe alguma urgência que deve ser resolvida imediatamente.

Um modelo de *machine learning* pode ser utilizado também para prever o comportamento de geração de uma planta fotovoltaica. Através dos dados de clima, tempo e de gerações históricas, é possível prever, com uma precisão razoável, quanto será gerado no curto, médio e longo prazo. Para tal, pode-se utilizar modelos de aprendizado de máquina baseados em clusterização, agrupando pontos operativos semelhantes considerando que a previsão de geração é a média da geração dos clusters, ou até mesmo, confeccionar uma rede neural que representa o modelo da usina e que sua saída é a potência a ser gerada, com base nos dados de entrada.



Este ebook faz parte de uma coleção de quatro apostilas desenvolvidas pelo projeto Profissionais do Futuro: Competências para a Economia Verde.

Confira abaixo os demais exemplares:



Sistemas de armazenamento de energia



Cibersegurança e Ética de Dados no Setor Elétrico



Recursos Energéticos Distribuídos e Redes Elétricas Inteligentes



