

# Tópicos Especiais em Sistemas de Informação

*Aline Chagas Rodrigues Marques*

*Marcelo Takashi Uemura*

## INFORMAÇÕES SOBRE O AUTOR

### **Aline Chagas Rodrigues Marques**

- Mestra em Ciência da Computação pelo Centro de Informática (CIn) (2015).
- Especialista em Análise e Desenvolvimento de Sistemas pela Universidade Federal do Pará (2011).
- Bacharel em Ciência da Computação pela Universidade Federal do Pará (2010).

### **Sobre o Autor**

Sou professora dos cursos a distância de desenvolvimento de sistemas pela Secretaria de Educação do Estado de Pernambuco, vinculada à escola ETEPAC (Escola Técnica Estadual Professor Antônio Carlos Gomes da Costa). Atualmente, sou conteudista das áreas de Engenharia de Software e Gestão de Sistemas de Informação. Sou tutora e orientadora de TCCs EAD pela Unyleya para o curso de Experiência e Interface do Usuário.

## INFORMAÇÕES SOBRE O AUTOR

### **Marcelo Takashi Uemura**

- Mestre em Administração (Inovação) pela Universidade Positivo (2019).
- MBA em Inteligência Financeira pela Universidade Positivo (2018).
- MBA em Business Intelligence pela Universidade Positivo (2016).
- Especialista em Redes Multisserviços e Teleinformática pela Universidade Federal de Pernambuco (2002).
- Especialista em Engenharia de Software pela Universidade Tecnológica Federal do Paraná (2001).
- Bacharel em Engenharia Industrial Elétrica pela Universidade Tecnológica Federal do Paraná (1998).

### **Sobre o Autor**

Sou professor para cursos presenciais e semipresenciais de Análise e Desenvolvimento de Sistemas, Gestão de Tecnologia da Informação, Sistemas para Internet. Atuo como coordenador de cursos de pós-graduação, voltados para desenvolvimento de software, no Centro de Tecnologia da Informação da Universidade Positivo.

## INTRODUÇÃO DO LIVRO

Olá, aluno(a), nesta disciplina, veremos os Tópicos Especiais em Sistemas de Informação.

Na primeira unidade, abordaremos o conceito de Big Data, que faz referência a grandes massas de dados, em sua maioria, de forma não estruturada. Assim, analisaremos dados como os das redes sociais, dos sistemas de comércio eletrônico, das ferramentas de gestão e dos objetos conectados à internet.

Na segunda unidade, veremos conceitos relacionados à internet das coisas (IoT), que é uma das principais fontes de dados para o Big Data. Este, por sua vez, consiste em uma rede de objetos físicos que reúne e transmite dados, como a conexão de veículos, objetos com sensores e conexão com a internet.

Na terceira unidade, veremos a Inteligência Artificial e suas aplicações, bem como a análise de algoritmos, de redes neurais, da relação com a inteligência humana etc.

Por fim, na última unidade, estudaremos a ciência de dados e o aprendizado de máquina. Conheceremos, também, de que forma a máquina será treinada a partir da análise de grande quantidade de dados relacionada aos conceitos de ciência de dados e aplicação de algoritmos.

Assim, desejamos a você um ótimo desenvolvimento com todas essas temáticas!

UNIDADE I

# **BIG DATA**

*Aline Chagas Rodrigues Marques*

## Introdução

Caro(a) aluno(a), na primeira unidade da disciplina de Tópicos Especiais em Sistemas de Informação, vamos estudar sobre Big Data. O termo faz referência a grandes massas de dados, em sua maioria, não estruturados, ou seja, dados que se apresentam de diversas formas devido à variedade de fontes que o produzem, por exemplo: redes sociais, sistemas de comércio eletrônico, ferramentas de gestão e até objetos conectados na internet. Não apenas a quantidade destes dados tem aumentado, mas também a velocidade com que esses dados precisam ser processados. Essa demanda surge com a necessidade de tomadas de decisões, praticamente em tempo real. As tomadas de decisões tornaram-se o principal objetivo das ferramentas de Big Data. Empresas têm investido cada vez mais neste recurso para poder ter sucesso em suas áreas de negócio. Essas tomadas de decisões envolvem problemas complexos e, por este motivo, diversos processos foram modelados para que fosse possível criar soluções de forma otimizada, utilizando os diversos recursos existentes.

Nesta disciplina, vamos entender inicialmente as definições que envolvem a origem do Big Data, como dado e informação. Vamos observar em quais áreas estão sendo aplicadas e suas principais características. Em seguida, verificaremos suas formas de aplicação por meio de diversas ferramentas. E, por fim, vamos analisar as aplicações reais em que o uso dos conceitos estudados auxilia a resolução de algum problema do mundo real.



Fonte: Vasin Leenanuruxsa / 123RF.

## Definição

Nós vivemos na era dos dados. A tendência é que a quantidade de dados gerada por cada indivíduo cresça, mas talvez a quantidade de dados gerados por máquinas, como parte da *internet* das coisas (*Internet of Things* - IoT), cresça mais significativamente. Logs de máquinas, leitores RFID (*Radio Frequency Identification*, ou seja, Identificação por Radiofrequência), rede de sensores de obras, vestígios de GPS (*Global Positioning System*, isto é, Sistema de Posicionamento Global) de veículos e até mesmo transações de varejo contribuem para a crescente montanha de dados.

White (2015) apresenta algumas informações sobre gerações de dados em massa:

- a Bolsa de Nova York gera cerca de 4 a 5 Terabytes de dados por dia;
- o *Facebook* hospeda mais de 240 bilhões de fotos, crescendo a 7 Petabytes por mês;
- o *Ancestry.com* (serviço de genealogia que ajuda pessoas a encontrarem suas raízes familiares) armazena cerca de 10 Petabytes de dados, e
- o *Internet Archive* (biblioteca digital) armazena cerca de 18,5 Petabytes de dados.

## Dado *versus* informação

Dado é o elemento que representa fatos ocorridos antes que tenham sido organizados ou arranjados de maneira que as pessoas possam entender e usar (TAURION, 2013). Pode ser considerado qualquer registro que tenha relação a um evento ou uma ligação a uma entidade (AMARAL, 2016). Dados são representações brutas, como números soltos. Quando os dados são cruzados ou organizados, passam a ser informações. Por exemplo, o número 12 pode representar uma série de coisas: pode ser uma dúzia de ovos, a quantidade de produtos em estoque, ou até um dia se for colocado no formato 12/09/2019. Neste caso, o número 12 passa a ser um dado, mas já com algum tipo de informação. Se associado ao nascimento de uma pessoa, transformamos o 12 na sua data de nascimento. Isso é uma informação sobre ela. Dados só terão significados se cruzados e organizados, transformando-se, portanto, em informações. Reforçando a ideia: uma tabela com nomes e notas de alunos pode ser considerada um conjunto de dados.

Para Davenport (1998), os dados são simples observações sobre o estado do mundo e apresentam as seguintes características:

- são facilmente estruturados;
- obtidos por máquinas;
- frequentemente quantificados;
- facilmente transferidos.

Já a informação pode ser definida como a interpretação de um conjunto de dados. Ela é o resultado da análise realizada a partir de dados apresentados de tal forma que permita a realização dessas interpretações. A informação exige consenso em relação ao significado e requer, necessariamente, a mediação humana (DAVENPORT, 1998).

A informação pode ser vista como a interpretação pessoal para um conjunto de dados, a relação que estabelece com outros dados e o contexto no qual elas estão inseridas. Um mesmo conjunto de dados pode gerar informações diferentes para indivíduos distintos, pois serão analisadas dentro dos próprios contextos. Retomando o exemplo da tabela de notas: as informações geradas a partir das médias dos alunos podem produzir uma série de conclusões acerca da turma avaliada (AMARAL, 2016).

Isso nos direciona a outra definição referente ao uso dos dados: o conhecimento. Segundo Davenport (1998, p. 19): “Conhecimento é a informação valiosa da mente humana, inclui reflexão, síntese e contexto, além disso é de difícil estruturação, transferência e captura em máquinas”. Uma relação dos conceitos pode ser observada na Figura 1.1.



Figura 1.1 - Esquema: dado, informação e conhecimento

Fonte: Oliveira (2016, *on-line*).

Segundo Taurion (2013), os dados podem ser vistos como recursos naturais da sociedade da informação. Porém, para agregar esse valor é necessário tratá-los e analisá-los para que possam ser usados para tomadas de decisões. Como comentamos anteriormente, não é fácil medir o volume total de dados gerados eletronicamente, uma vez que atualmente coletamos informações para armazenamento de diversas fontes e em tempo real.

Os dados podem ser representados em formato analógico, digital ou impresso, como as bibliotecas, que concentram a maior quantidade de dados impressos no mundo (AMARAL, 2016). Porém, no contexto que estamos estudando, vamos nos fixar apenas no formato dos dados digitais, transmitidos por meio de pacotes de bits (zeros e uns), de forma mais eficiente.

Dados em sua forma digital, objeto do nosso estudo, são produzidos por algum tipo de dispositivo eletrônico e passam a integrar um processo que vai da sua preservação até a extração de informações para diversos usos (AMARAL, 2016). Existem várias formas para se extrair e preservar dados, Amaral (2016) apresenta algumas delas:

- o XML (*Extensible Markup Language*) pode representar um texto plano ou registros de um banco de dados relacional;
- o JSON (*JavaScript Object Notation*) é um formato parecido com o XML,

compacto e com padrão independente, utiliza textos para troca de dados rápida entre sistemas;

- as Planilhas Eletrônicas podem ser utilizadas para armazenar dados, sua estrutura básica, composta por linha e colunas, permite a criação de diversas configurações, como uma lista de compra, informações financeiras etc., e
- o processo de ETL (*Extract, Transform and Load*) é mais elaborado e usado para extrair dados para um *data warehouse* (depósito de dados digitais que armazena informações detalhadas). Essa solução requer uma estrutura computacional bem maior e uma de suas utilizações é a análise de dados para auxiliar grandes corporações em tomadas de decisões de diversos níveis.

O processo de produção de dados até a sua utilização final é denominado ciência de dados (*data science*). Como todo processo, também possui um ciclo de vida (Figura 1.2).

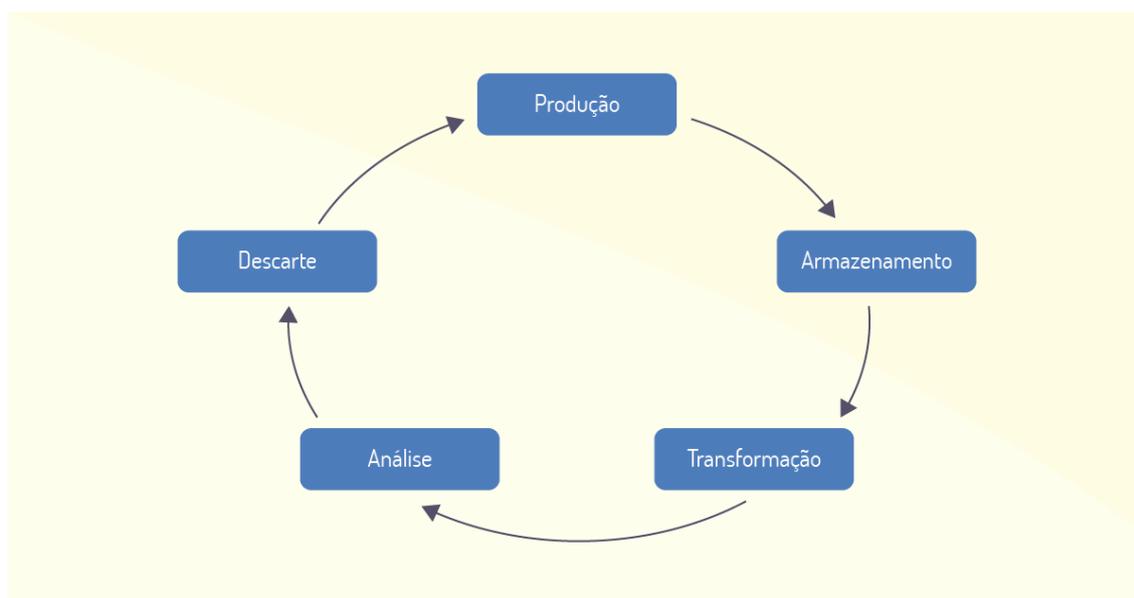


Figura 1.2 - Ciclo de Vida do Dado

Fonte: Amaral (2016, p. 06).

Cada etapa do processo do ciclo de vida demanda diferentes ferramentas que auxiliam as atividades do processo. Esse campo é amplo e tem gerado novas tecnologias que rapidamente são incorporadas para dinamizar os estudos das análises de dados. A Figura

1.3 representa a distribuição de algumas das tecnologias utilizadas em cada etapa do ciclo de vida dos dados.

Para que os dados possam ter o ciclo de vida completo, em algum momento, precisam ser armazenados e analisados, este ponto do processo será o objeto do nosso estudo.

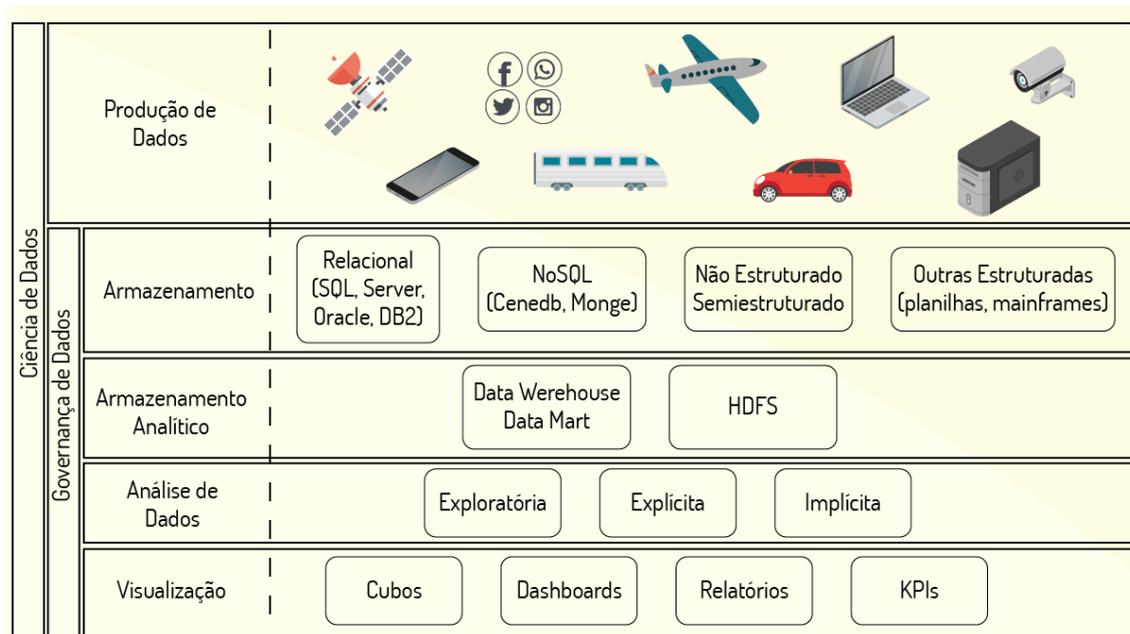


Figura 1.3 - Tecnologias para ciência de dados

Fonte: Amaral (2016, p. 07).

A figura apresenta um esquema para as tecnologias que produzem dados na Web 2.0 por meio das mídias sociais e que são impulsionadas pelo crescimento da IoT. Outras informações apresentadas na imagem são as tecnologias para armazenar, analisar e disponibilizar dados que auxiliam na tomada de decisão.

Como podemos perceber, dados são representações brutas do mundo real. Já as informações são dados processados e transformados em conhecimento, tornando-se úteis para as pessoas. Para que esses dados sejam convertidos em informação, são necessárias técnicas de armazenamento e de processamento. Nos próximos tópicos, vamos conhecer e entender como as ferramentas auxiliam esse processo de transformação dos dados.

## ***Big Data***

Com a necessidade de transformação dos dados surgiu também a demanda por sistemas que armazenam grandes volumes de dados complexos, denominados de *Big Data*. Eles podem ser definidos como um arranjo de dados que o *software* tradicional não consegue gerenciar e, portanto, são utilizadas ferramentas específicas para o seu efetivo tratamento, armazenamento e processamento (TAURION, 2013).

Essas ferramentas têm o intuito de analisar dados não estruturados que não apresentam relação entre si e não possuem uma estrutura organizada. Como exemplo podemos citar os arquivos de áudio e vídeo, comentários em redes sociais, geolocalização utilizada em diversos aplicativos, entre outros.

Ferramentas comuns, como o Excel, são utilizadas para tratar dados estruturados, como os preços dos produtos de uma lista de supermercado. Por outro lado, as ferramentas de *Big Data* analisam dados não estruturados e suportam grandes quantidades, além de realizarem análises em alta velocidade (FAWCETT; PROVOST, 2016).

A criação e a utilização de tecnologias *Big Data* tem crescido consideravelmente, pois estão sendo empregadas no apoio às atividades relacionadas com mineração de dados. A mineração de dados tem por objetivo explorar massas de dados à procura de padrões que sejam consistentes e que possam gerar novos significados a um subconjunto de dados originais (MOURA, 2018). Esse processo foi desencadeado devido à redução de custos de tecnologias para armazenamento de dados, bem como ao crescimento de dispositivos capazes de produzir esses dados de forma massiva. Recentemente, criou-se a "Computação em Nuvem" e a "Internet das Coisas" (AMARAL, 2016).

## **REFLITA**

A série *Black Mirror* (produzida pela Netflix) aborda em seus episódios a relação da sociedade com a tecnologia em uma perspectiva futurista e algumas vezes sombria. No primeiro episódio da segunda temporada, intitulado "*Be Right Back*", depois de perder o namorado em um acidente e descobrir que estava grávida, a personagem Martha resolve contratar um serviço de tecnologia que, por meio dos dados de *e-mail* e de redes sociais de seu companheiro, disponibiliza um serviço que promete diminuir a dor do luto. Depois de entregar os dados da conta de *e-mail* e das redes sociais, o serviço contratado atuava

em etapas: inicialmente, simulava a voz e a personalidade do namorado em conversas telefônicas, posteriormente, na etapa final, Martha recebeu uma versão *android* que se assemelhava fisicamente ao seu namorado. As informações coletadas anteriormente foram projetadas como memórias no robô. Com o passar do tempo, Martha percebeu que faltavam características humanas, uma vez que o robô apenas seguia suas ordens. Ao tentar desfazer seu “erro”, ela pediu para que o robô se jogasse de um penhasco e, para sua surpresa, ele não a obedeceu e implorou por sua “vida”. Por fim, ela atendeu ao pedido, mas o manteve guardado no sótão.

O episódio nos faz refletir sobre quais são as reais aplicações da utilização de dados disponibilizados na rede. Qual sua opinião sobre a atitude de Martha tentar interferir no ciclo natural da vida (como o luto após a morte)? Da forma como foram apresentadas, essas deveriam ser preocupações de quem produz ferramentas para *Big Data*?

Outro fenômeno que auxiliou no crescimento das ferramentas para *Big Data* foi o fato de grandes empresas adotarem as tecnologias web para ampliar suas operações. Na Web 1.0, as empresas começaram a estabelecer presença na *internet* e incorporar suas tecnologias para serviços, como o comércio eletrônico (FAWCETT; PROVOST, 2016). A Figura 1.4 exemplifica um esquema para este novo contexto.

Após marcar presença na *web* e implantar alguns serviços, as empresas passaram a buscar a melhoria do que já ofertavam, aproveitando as mudanças que chegaram com a Web 2.0, começaram a produzir sistemas mais interativos, conectados às mídias sociais (AMARAL, 2016).

O principal objetivo das ferramentas de *Big Data* é processar grandes massas de dados. Esse processamento representa um dos grandes desafios tecnológicos da área, uma vez que se faz necessária a utilização de infraestruturas e tecnologias analíticas que suportem tais atividades. Um elemento que pode facilitar esse processo é a Computação em Nuvem. Essa deve ser a maior mudança nos serviços oferecidos pela *internet*: a consolidação do uso de ferramentas para armazenar e analisar grandes massas de dados (TAURION, 2013). O que alguns autores já definem como Web 3.0.

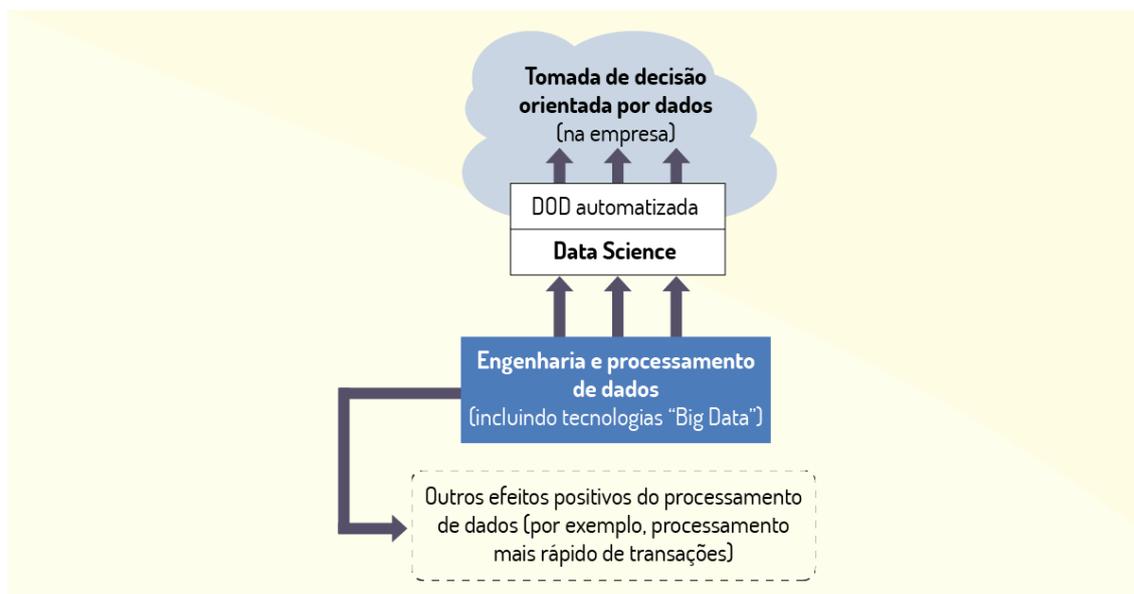


Figura 1.4 - *Data Science* nos contextos das organizações

Fonte: Fawcett e Provost (2016, p. 05).

Entendendo *Big Data* como a produção de ferramentas para produção e armazenamento de dados em grande escala, bem como as tecnologias de extração e análise, precisamos visualizar as mudanças que essas ferramentas trarão para a indústria ao contribuírem com processos produtivos mais eficientes e com menores custos, ampliando o relacionamento das grandes empresas com seus clientes, fornecedores e parceiros comerciais. É a indústria 4.0.

## FIQUE POR DENTRO

A indústria 4.0 (ou a 4a revolução industrial), como definem alguns autores da área, se constitui de uma massiva utilização de ferramentas e de sistemas conectados, trabalhando com análise de grandes massas de dados em tempo real para ampliar as relações entre produtos, empresas, serviços e clientes. Além disso, tem o objetivo de otimizar e reduzir custos de produção. Para saber mais, seguem alguns *links* que abordam esse tema sob diferentes pontos de vista: <https://www.industria40.ind.br/artigo/17534-big-data-na-industria-40-como-transformar-dados-em-lucro> e <https://www.totvs.com/blog/big-data-na-industria-40/>. Acesso em: 14 jan. 2020.

Nesse novo modelo de negócio em que os dados dão suporte às decisões das empresas torna-se mais fácil definir quais abordagens são interessantes para o crescimento do empreendimento. Fawcett e Provost (2016) destacam duas formas nas quais esses modelos de decisão podem ser aplicados: as novas descobertas e as decisões que se repetem com frequência.

Nos próximos tópicos, os demais aspectos relacionados ao *Big Data* serão detalhados para que se construa uma visão mais abrangente dessa ferramenta e da sua atuação dentro dessa nova era da tecnologia.

## ATIVIDADE

1) Ferramentas de *Big Data* podem, entre suas diversas atividades, ampliar o relacionamento do consumidor com um determinado produto ou empresa por meio das análises de dados geradas por informações disponibilizadas por seus usuários. Analise as afirmações a seguir e assinale a alternativa que represente as ações que se encaixam em atividades que têm relação com o uso de *Big Data*.

- I. Sugestão de novos artistas - baseada nas preferências musicais do cliente.
- II. Realização de cadastro do usuário em sistemas de compras *on-line*.
- III. Gerar descontos específicos - baseados nos hábitos de compras do cliente.
- IV. Indicação de produtos relacionados a itens adquiridos em compras anteriores.
- V. Proposta de uma rota de trânsito específica para evitar engarrafamentos.

Estão corretas:

- a) I – II – III – V.
- b) I – II – III – IV.
- c) I – III – IV – V.
- d) II – III – IV – V.
- e) Todas as afirmativas estão corretas.

## Aspectos relevantes

Com a complexidade e a quantidade de dados gerados atualmente, as soluções de *Big Data* precisam ser consistentes. Segundo Taurion (2013), o *Big Data* apresenta alguns aspectos que comumente chamamos de 5 Vs: volume, velocidade, variedade, veracidade e valor. Essas características indicam que ao utilizar ferramentas de *Big Data* a empresa obtém respostas com maior completude, já que possui uma grande quantidade de informações disponibilizadas, e com essas respostas pode-se ter maior confiança nos dados, possibilitando uma tomada de decisão mais precisa. Sobreiro (2018) e Taurion (2013) definiram cada um dos 5 elementos e descreveram sua importância para ferramentas de *Big Data*.

## Volume

Tente imaginar a quantidade de informações geradas através do envio de *e-mails*, mensagens, fotos e vídeos nas redes sociais. Parece incalculável. Agora pense em como armazenar essas mesmas informações utilizando recursos de tecnologia. Essa é a função do *Big Data*: armazenar uma quantidade de dados muito grande utilizando ferramentas tecnológicas, visando melhorar o processamento destes dados.

O aumento na capacidade de processar grandes quantidades de dados é uma das principais relevâncias do *Big Data*. Essa melhoria gerou novos modelos de análise dos dados. Um exemplo é a integração com outras ferramentas de análise de dados, como tecnologias de *Business Intelligence* (BI). BI é o termo aplicado para classificar o uso de ferramentas que empregam os conceitos de análise de dados para auxiliar em atividades de tomada de decisão. Tecnologias para *Big Data* são criadas de forma a ser flexíveis e conseguem exportar dados através de arquivos como Excel, ou elementos que possam ser explorados em outros tipos de sistemas como o JSON (*JavaScript Object Notation*), ou simplesmente conectar-se a tecnologias que façam análise de dados de forma semelhante como o BI.

## FIQUE POR DENTRO

A empresa Heikima desenvolve soluções para *Big Data* e disponibilizou um *e-book* para iniciantes no assunto apresentando os conceitos, ferramentas e alguns casos de sucesso relacionados a *Big Data* e *Data Analytics*. Acesse o *link* e leia o *e-book*: [http://www.mettodo.com.br/ebooks/O\\_guiá\\_definitivo\\_de\\_Big\\_Data\\_para\\_iniciantes.pdf](http://www.mettodo.com.br/ebooks/O_guiá_definitivo_de_Big_Data_para_iniciantes.pdf). Acesso em: 14 jan. 2020.

Em relação ao volume, as ferramentas de *Big Data* também criam um desafio para estruturas de armazenamento de dados como sistemas distribuídos. A tendência da produção de dados é aumentar com o tempo, assim, não seria possível impor um limite de dados para esse tipo de ferramenta. Com o passar do tempo, as atualizações de estruturas de armazenamento de dados precisarão lidar com soluções visando a produção de dados em massa.

### Velocidade

Com o aumento dos dados, algumas características passaram a ser fundamentais em ferramentas de *Big Data* e o retorno dos dados processados é uma delas. As aplicações precisam não apenas armazenar, mas também tratar os dados e deixá-los disponíveis de forma rápida, pois outras aplicações dependem dessa agilidade para o andamento de suas atividades.

A velocidade é a característica das ferramentas de análise de dados e se refere à taxa de tempo para que os dados processados fiquem disponíveis. Pode ser definida com o tempo de retorno de um dado desde a sua entrada até a tomada de decisão relativa ao seu processamento. A etapa pode ser definida como processamento complexo de eventos.

O tempo que o sistema leva para gerar esse processo pode colocá-lo em vantagem no mercado, pois as atividades geralmente estão ligadas às informações que precisam ser enviadas rapidamente, como a liberação de descontos para clientes de um serviço de compras, ou dados de localização em tempo real.

## Variedade

Segundo o relato de diversos autores, a variedade está relacionada ao formato dos dados que, juntamente com o volume e a velocidade, formam os 3 Vs principais do *Big Data*. Os dados podem surgir de diversas formas, podendo servir como base de dados ou simples fonte de informação para outros sistemas.

A variabilidade é a característica do *Big Data* responsável pelo armazenamento e tratamento de tipos de dados em diversos formatos. Essa característica que denota a flexibilidade do *Big Data* torna a ferramenta mais atrativa comercialmente e abre portas para o uso eficaz em diversas áreas de atuação. Sobreiro (2018) lista alguns tipos de dados utilizados em ferramentas de *Big Data* em diferentes tipos de aplicações:

- Excel (Programa de planilhas).
- XML (*Extensible Markup Language*).
- HDF5 (*Hierarchical Data File 5*).
- NetCDF (*Network Common Data Form*).
- DICOM (*Digital Imaging and Communications in Medicine*).

Em ferramentas de análises de dados, deve ser possível a integração desses e outros tipos de dados para diversas finalidades.

## Veracidade

É a característica dos dados relativos à sua precisão e correção. Sistemas de todo tipo recebem dados sem um controle do formato. Com as ferramentas de *Big Data* não seria diferente. Para armazenar dados é preciso ter a certeza de sua qualidade, o *Data Warehouse* (DW), por exemplo, utiliza uma ferramenta chamada ETL (*Extract, Transform and Load*). Essa ferramenta realiza sistematicamente o tratamento e a limpeza dos dados para inserção no DW.

Como o *Big Data* preza pela velocidade e pela variedade dos dados, esses fatores podem influenciar na determinação da veracidade. A velocidade faz com que o tempo disponível não suporte ferramentas com as atribuições do ETL, uma vez que as novas informações recebidas precisam ser processadas e disponibilizadas de forma rápida. A variedade dos

tipos de informações recebidas demanda adaptações para que se possa trabalhar com os diversos formatos de arquivos e as formas personalizadas para lidar com cada tipo de dado.

Um grande desafio posto pelo *Big Data* é determinar a veracidade dos dados disponíveis para a sua empresa de forma que as informações possam servir de guia para um planejamento seguro. Talvez esse seja o aspecto mais relevante, visto que a tomada de decisão baseada em dados incorretos pode ser catastrófica.

### **FIQUE POR DENTRO**

A atividade de transformação dos dados é importante para alimentar o DW, entretanto essa atividade gera uma grande carga de trabalho, já que os dados são carregados de fontes diversas e precisam ser selecionados e padronizados para gerar os repositórios do DW de forma que apenas os dados relevantes possam fazer parte desse repositório. Para entender mais sobre esse processo, acesse o *link* a seguir: <https://www.devmedia.com.br/extract-transformation-and-load-etl-ferramentas-bi/24408#targetText=Veja%20neste%20artigo%20como%20funciona,v%C3%A3o%20alimentar%20o%20Data%20Warehouse&targetText=ETL%20como%20a%20pr%C3%B3pria%20tradu%C3%A7%C3%A3o,Data%20Warehouse%20ou%20Data%20Mart.>

Acesso em: 14 jan. 2020.

### **Valor**

Para que as massas de dados, às quais as ferramentas de *Big Data* têm acesso, possam gerar mais lucros é preciso agregar valor às atividades. O valor é relacionado aos objetivos do uso dos dados. A definição da abordagem que será feita com a massa de dados que está circulando é um dos desafios enfrentados pelas empresas que utilizam *Big Data*. Ou seja, o valor está relacionado ao quão importantes os dados são para as empresas e quais benefícios podem ser extraídos por meio desses dados.

O valor de um dado está relacionado à descoberta de padrões, em outras palavras, seriam as preferências de usuários que podem acarretar ganhos significativos para quem depende de sua interpretação, por exemplo.

A qualidade também pode ser vista como um critério de valor. Algumas características como a exatidão, a integridade, a consistência e a relevância também podem ser concebidos como critério que imputam valor a um dado (GALDINO, 2016).

Nem todo dado é relevante ou útil para quem o coleta. Outro aspecto que pode gerar valor a um dado é a percepção de quem analisa: é necessário realizar uma triagem por intermédio de questionamentos que possam gerar uma análise acerca de sua abordagem para o uso ou não dos dados que foram previamente coletados (TAURION, 2013).

### **Principais ferramentas**

As ferramentas para *Big Data* têm a finalidade de auxiliar no processamento e armazenamento da grande massa de dados gerados de diversas formas, pois com a crescente demanda, surge a necessidade de tecnologias que possam suportar as atividades realizadas (GALDINO, 2016). O *Hadoop* é uma das ferramentas mais utilizada para análise de dados em massa. Sua arquitetura distribuída permite armazenar e processar grandes quantidades de dados de forma rápida. Além do seu poder computacional, outras características que agregam valor na sua utilização são: tolerância a falhas, flexibilidade e escalabilidade. Vamos entender melhor essas características e como essa estrutura funciona no contexto de análises de dados.

### ***Hadoop***

O *Hadoop* é uma das tecnologias que mais se destacam para análise de dados em massa. Foi criado em 2005 em um projeto da empresa Apache e divide as tarefas que serão realizadas em servidores (TAURION, 2013). Entendido como um conjunto de programas que dão suporte às operações voltadas para armazenamento e análise de dados, as atividades relacionadas ao *Hadoop* são combinadas em dois projetos: o *Hadoop MapReduce* (HMR) e o *Hadoop Distributed File System* (HDFS). Sua natureza flexível permite a adaptação dos componentes de seu sistema de acordo com as necessidades específicas da empresa que o utiliza (SOBREIRO, 2018).

O *Hadoop MapReduce* teve origem a partir de uma ferramenta que o *Google* empregava para acelerar o resultado das pesquisas, o *MapReduce*. Esse componente é baseado em

um modelo de programação mais simples que permite o aumento da estrutura, sem necessariamente aumentar os custos (escalabilidade), por meio do uso de servidores de forma paralela (TAURION, 2013). Sua execução consiste em duas atividades: a) o mapeamento dos conjuntos de dados, transformando-os em uma relação chave/valor (ou tupla), chamado de *Map*; e b) o processamento dos dados para extração dos resultados, chamado de *Reduce* (SOBREIRO, 2018).

O *Hadoop Distributed File System* é o componente responsável pelo armazenamento dos dados dentro do *Hadoop*. Esse armazenamento é realizado através de blocos distribuídos em diversos servidores. Essa distribuição permite que a pesquisa pelos dados armazenados seja realizada de forma paralela, acelerando o processo (TAURION, 2013). Essa característica permite o armazenamento de uma grande quantidade de dados com alta performance e recuperação automática a falhas, tornando a ferramenta confiável. A forma de armazenamento de dados em diversos servidores torna a ferramenta escalável, aumentando os recursos de acordo com a necessidade (SOBREIRO, 2018).

O funcionamento do *Hadoop* é baseado em servidores e discos locais, em razão de considerar essa prática menos custosa. Isso implica, como já citado, em um tempo curto de recuperação em casos de falhas, porque cada bloco é copiado em mais dois servidores, assim, em caso de falha de um ou dois servidores, o terceiro servidor é capaz de suprir as demandas de pesquisas realizadas (TAURION, 2013).

O *Hadoop* foi projetado para manipular uma massa de dados grande, assim como um bloco com tamanho grande pode trabalhar com um número razoável de dados processados simultaneamente.

Para realizar o mapeamento dos dados, o *Hadoop* utiliza um servidor chamado *NameNode*. O mapeamento realizado é armazenado na memória e existe apenas um *NameNode*, que deve ter alta disponibilidade e um processo de *backup* eficiente. Por haver apenas um ponto de falha, essa situação deixa o sistema vulnerável.

A distribuição de seus componentes torna o *Hadoop* uma ferramenta que permite ao *Big Data* processar grandes massas de dados de forma rápida. O *Hadoop* é a ferramenta mais utilizada para *Big Data* visto que os dados são tratados em tempo real em diversos servidores, garantido um processamento mais eficiente e gerando economia de dinheiro e tempo (WHITE, 2015).

## FIQUE POR DENTRO

Um exemplo de melhoria pode ser observado no caso da Caesars Entertainment, uma empresa de cassinos americana. Essa empresa implantou um novo sistema *Hadoop* para análise de dados com o intuito de auxiliar nas atividades de criação de campanhas específicas de *marketing*, alcançando os numerosos tipos de clientes. Com a utilização do novo sistema, a empresa estima uma redução no tempo de processamento de 6 horas para 45 minutos, além de visualizar resultados mais precisos nas sugestões de experiências para os usuários e uma maior segurança em suas transações financeiras. Segundo informações da empresa, após a mudança no sistema de análise de dados, o número de registros processados chegou a ultrapassar o valor de 3 milhões por hora.

Leia mais em: <https://cio.com.br/sete-casos-de-uso-do-hadoop-em-aplicacoes-de-big-data/>. Acesso em: 14 jan. 2020.

## ATIVIDADE

2) *Big Data* pode ser definido como um grande volume de dados que trafega pela *internet*. Devido ao avanço da tecnologia, os serviços geram cada vez mais dados. As análises desses dados produzem informações de forma instantânea, ajudando na tomada de decisão em tempo real. O *Big Data* apresenta alguns aspectos que comumente chamamos de 5 Vs: volume, velocidade, variedade, veracidade e valor. Considerando essas características do *Big Data*, assinale a alternativa correta.

- a) A variedade está relacionada com a quantidade extensa de dados armazenada e analisada.
- b) O volume está relacionado com a importância dos dados para a empresa.
- c) A veracidade está relacionada com a alta transmissão de dados recebidos e transmitidos.
- d) O valor está relacionado com a importância dos dados para empresa.
- e) A velocidade está relacionada com a confiabilidade dos dados coletados.

## **Tecnologias envolvidas**

Para realizar o processamento da grande quantidade de dados, bem como para trabalhar com objetivos específicos, algumas tecnologias envolvidas serão apresentadas neste item. Iremos observar o tipo de banco de dados criado para armazenar os diversos tipos de dados gerados na Web 2.0 e também as ferramentas aplicadas nas análises de dados que auxiliam nas tomadas de decisões. Por fim, apresentaremos as novas formas de trabalhar o armazenamento de dados, que podem ser consideradas uma tendência para as ferramentas de *Big Data*.

## **Banco de dados não relacional**

Com o aumento da quantidade de massas de dados geradas no advento da Web 2.0, foi necessário modificar a forma de armazenamento. Os bancos de dados existentes, manipulados de forma relacional, não comportavam mais esses tipos de dados, que passaram a ser modelados de diversas formas (MATOS, 2019). Esses bancos de dados relacionais modelavam os dados de forma que eles chegavam aos usuários no formato de tabelas e estabeleciam uma relação. Geralmente utilizavam o conceito de entidades, registro e colunas, além de representação por chaves (DIANA; GEROSA, 2010). Nesse contexto surgiram os bancos de dados não relacionais.

Banco de dados não relacionais, ou banco de dados NoSQL (do inglês, *Not Only SQL*), foram criados para solucionar a nova forma de armazenamento de dados provenientes da Web 2.0 e também para propor os novos paradigmas de armazenamento de dados em massa como o *Big Data*. Esse tipo de banco é distribuído e não relacional, ou seja, permite receber dados que não estão estruturados e precisam passar por um processo de tratamento para utilização nas ferramentas que tem a finalidade analisá-los (MATOS, 2019). Possuem também uma arquitetura que permite a escalabilidade. Como já vimos, a escalabilidade é uma característica importante nos conceitos relacionados ao *Big Data*, visto que não é possível medir a quantidade de dados que pode ser armazenada (TAURION, 2013).

Diana e Gerosa (2010) descrevem de forma detalhada os itens que levaram ao surgimento das tecnologias para o armazenamento e o processamento desse tipo de dado:

- a natureza de dados gerados pela web atualmente são, em grande parte, semiestruturados, como os arquivos multimídia (imagens, sons e vídeo). Com a natureza particular desses tipos de dados, torna-se necessário a construção de soluções que possam manipulá-los.
- para processar grandes quantidades desses novos tipos de dados, atender as demandas relativas à escalabilidade no armazenamento e impulsionar a baixa no valor dos custos relacionados ao *hardware*, passou-se a utilizar o paralelismo. Trabalhar de forma paralela no processamento de dados permite aumentar o nível de performance e auxilia nas análises de dados para tomada de decisão.
- para atender às novas demandas relacionadas ao contexto apresentado, foi necessário pensar em armazenamento e em análise de dados em massa de forma rápida e eficiente. Pensando-se no da geolocalização, que está distribuída em diversos países, é preciso ter um sistema que permita atender as necessidades relacionadas não apenas a distribuição no armazenamento de dados, mas também que seja tolerante a falhas de *hardware* e infraestrutura de forma geral, mantendo os dados sempre disponíveis.

Desta sorte, podemos afirmar que as tecnologias dos bancos de dados não relacionais cresceram e que algumas ferramentas que se baseiam nesse paradigma foram desenvolvidas para atender as demandas de mercado. Um levantamento realizado em 2018 pelo pesquisador Matos (2019) apresenta as ferramentas mais utilizadas:

- *MongoDB*: é uma ferramenta não relacional que utiliza documentos no formato próximo ao JSON (*JavaScript Object Notation*) para armazenar dados. O documento é semelhante a um registro com campos e valores, suporta esquemas dinâmicos e realiza atividade de ambientes de produção - replicação, indexação, consulta, além de atuar como um sistema de arquivos;
- *Cassandra*: é um mecanismo de banco de dados orientado em colunas. É otimizado para *clusters* (grupo de computadores trabalhando em conjunto) e pode ser utilizado para *datacenters*. Possui um modelo distribuído, otimizado e descentralizado, fato que trouxe popularidade à ferramenta;

- *Redis*: possui uma implementação *key-value store* (atribui valores às chaves para facilitar o acesso e armazenamento, otimizando o processo de busca) e segue o exemplo de *hashmap* e dicionários utilizados em algumas linguagens de programação. Essa característica auxiliou na construção de APIs (*Application Programming Interface*) para linguagens de programação, aumentando a escolha pelos desenvolvedores;
- *Amazon DynamoDB*: é um serviço para banco de dados NoSQL na nuvem fornecido pela AWS (*Amazon Web Service*), disponível para aplicações que precisam de latência (tempo de acesso) abaixo de 10 milissegundos. É compatível com os modelos chave-valor e possui escalabilidade automática de dados transferidos, modelo de dados flexível e desempenho confiável. Ideal para aplicações móveis, web, jogos e tecnologias IoT (*Internet of Things*), e
- *Neo4j*: sistemas de gerenciamento de banco de dados baseado em grafos (possui arestas que funcionam como relacionamento dos vértices ou nós) que cria uma relação com as instâncias de dados. Essa associação auxilia no reconhecimento de padrões em atividades de mineração de dados.

### ***Business Intelligence***

*Business Intelligence* (BI) é um sistema de apoio à decisão integrado a outros sistemas de uma empresa, como um *Data Warehouse*. O seu objetivo é produzir uma análise dos dados e das operações de uma empresa de forma interativa e relacionar essas atividades com outras fontes de informação (FAWCETT; PROVOST, 2016). De forma simplificada, pode ser entendido como uma análise de dados de bases combinadas diferentes (SOBREIRO, 2018).

O BI atua convertendo dados brutos em visão de novos negócios, auxiliando gestores em processos de tomada de decisão. Atua a partir de elementos como *dashboards*, visualizações de informações e relatórios, que são modelados a partir de conjuntos de dados estruturados (MATOS, 2015).

Uma ferramenta de BI combina, basicamente, dados internos coletados de diversas formas (organizações, colaboradores, outros sistemas e *stakeholders* em geral) com ferramentas para transformação e análise dos dados coletados como:

- OLAP (*On-line Analytical Process*): pode ser entendido como uma interface com o usuário, permitindo a visualização de dados em diversas perspectivas. O OLAP recupera informações, após o processamento, como pode ser visto na Figura 1.5, criando uma representação do que foi processado para o usuário. Uma de suas funções básicas é a visualização dos dados em diversas dimensões (como se fosse um cubo), fato que permite a exploração e a rotação, apresentando outras formas de visualização dos dados (LUCAS; VIEIRA, 2019), e
- ETL (*Extract, Transforming and Loading*): é uma ferramenta de integração cujo objetivo é combinar dados de diversas fontes. É utilizado em estruturas de DW e auxilia na extração de dados de fontes externas, transformando em formatos que facilitam as análises e podem ser armazenados em outro sistema. A utilização desse tipo de ferramenta permite o aumento na produtividade de profissionais que trabalham com análise, pois não precisam ter o domínio técnico das atividades relacionadas à análise de dados. O ETL ainda fornece versões consolidadas baseadas em conteúdos específicos para seus usuários (SOBREIRO, 2018).

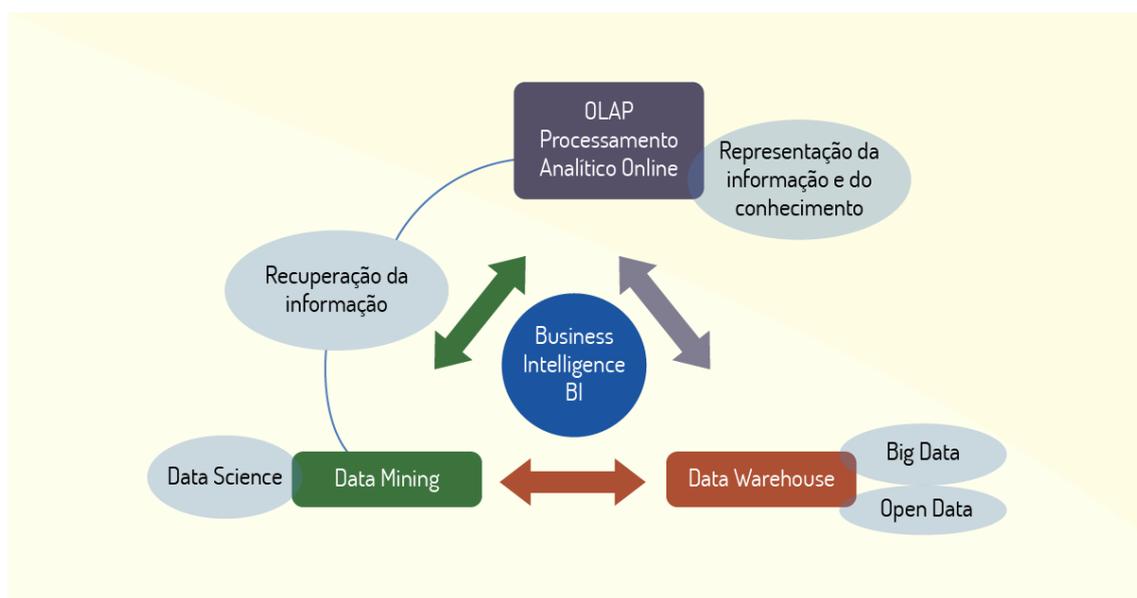


Figura 1.5 - Arquitetura de um BI

Fonte: Lucas e Vieira (2019, p. 01).

Uma ferramenta de BI também auxilia na verificação da qualidade de processos operacionais relacionados à estratégia da empresa por meio de indicadores de desempenho. Mediante recursos de análise e ferramentas de interatividade (como *dashboards*), permite a análise dos dados, em tempo real, admitindo a realização da tarefa de forma eficiente e prática (CALDAS; SILVA, 2016).

Existem diversas ferramentas de análises de dados e armazenamento em massa que podem ser compreendidas como componentes de um BI. Veremos alguns exemplos a seguir.

### ***Data Warehouse***

*Data Warehouse* (DW) é um sistema de análise de dados utilizado por empresas de diversos seguimentos para auxiliar no processo de tomada de decisão. Com a grande massa de dados gerados atualmente, por diversos contextos, como já vimos, se faz necessário uma tecnologia que permita a realização das análises de forma eficiente e que possibilite, por meio do processamento realizado, gerar tomadas de decisões que auxiliem no crescimento dessas empresas (TAURION, 2013).

Uma ferramenta de DW possui duas atividades que podem ser descritas como a inserção e a consulta dos dados, através dos elementos que os constituem. O processo é realizado mediante a transformação dos dados em informações (como visto anteriormente, as informações são significados extraídos dos dados). É esse processo que dá suporte às atividades dos gestores, pois cria um ambiente confiável para tomadas de decisões (CALDAS; SILVA, 2016).

O trabalho de uma ferramenta de DW é bem eficiente, visto que ao reunir informações de diversas fontes (bancos de dados), as análises tornam-se mais eficientes. Desta maneira, aproveita-se ao máximo as informações geradas, permitindo que nenhum dado fique inacessível ou não seja utilizado (TAURION, 2013).

Os autores Bispo e Cazarini (1998) descrevem os componentes de um *Data Warehouse*:

- Ferramenta de armazenamento;
- Ferramenta para extração de dados;
- Ferramenta para transformação de dados;
- Ferramenta para a limpeza de dados;

- Repositório de metadados;
- Transferência de dados;
- Gerenciamento e administração;
- Ferramenta para gerenciamento de consultas;
- Ferramenta para gerenciamento de relatórios;
- Ferramenta OLAP (*On-line Analytical Processing*);
- *Data Mining*;
- Ferramenta de simulação.

Os autores ainda descrevem como um DW utiliza essas ferramentas em um processo de análise de dados, da coleta até a construção de relatórios. Após a busca nas bases de dados operacionais, os dados são formatados para o modelo do DW. Os dados podem ser formatados de diversas maneiras: correções, fusões, desmembramentos, atividades que visam auxiliar em pesquisas futuras. Ainda no processo de extração, os metadados (informações sobre a origem e as alterações) são armazenados para dar consistência aos dados. Em seguida, a replicação dos dados é feita. Esse processo dá agilidade às análises, pois acelera os resultados das consultas realizadas.

As ferramentas voltadas para a construção dos relatórios consultam os dados presentes no DW (após todo processo inicial) e elaboram os relatórios em inúmeras formas de visualização e em vários níveis de complexidade. Por exemplo: essas informações podem ser apresentadas de forma sintética (condensando informações), ou analítica (ampliando o campo de visão sobre os dados), ou em outras formas de visualização, como os gráficos. Após a construção dos relatórios, através das ferramentas visuais (OLAP), é possível realizar a análise por meio de comparações (número de vendas de dois produtos), projeções (o que aconteceria com as vendas se juros fossem alterados), informações históricas (dentro de um determinado período qual produto vendeu mais). Essa ferramenta permite a melhor padronização para realização dessas atividades. A Figura 1.6, apresenta um esquema de um DW, representando as etapas e as ferramentas descritas anteriormente.

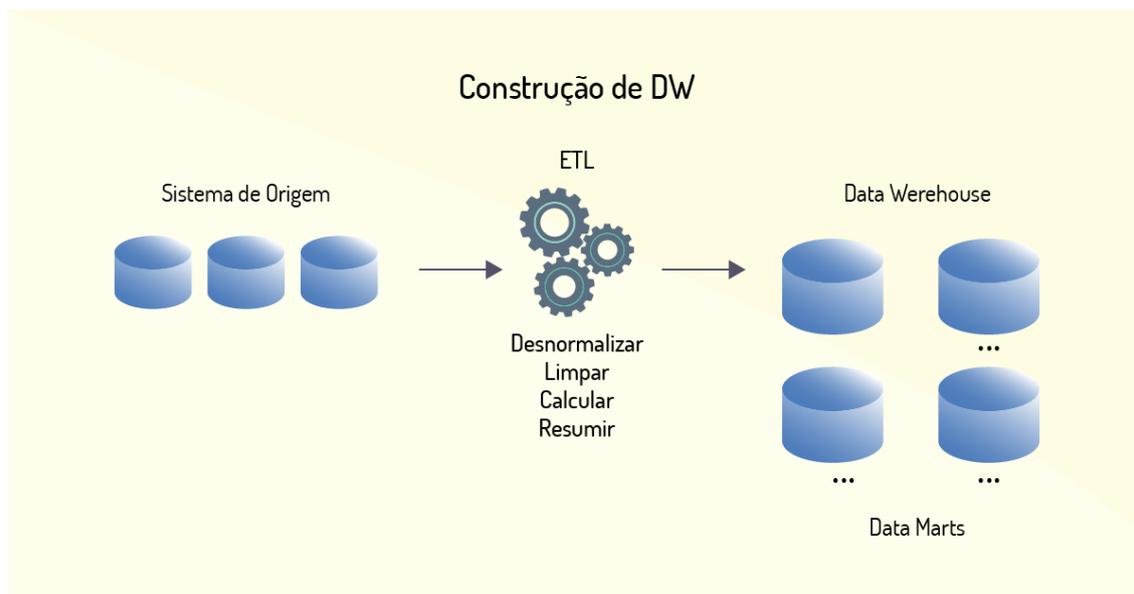


Figura 1.6 - Esquema de um DW

Fonte: Amaral (2016, p. 48).

Os dados gerados partir do processo do DW podem servir de suporte para outras ferramentas de *Big Data*, como veremos a seguir.

### ***Data Mart***

*Data Mart* (DM) pode ser visto como subconjuntos do DW, mas separado por assuntos específicos, por exemplo, os departamentos de uma empresa - neles teríamos os DM do setor comercial, do administrativo, do financeiro. Cada um representando um DM específico e todos como parte de um DW maior (ELIAS, 2014).

Apesar das semelhanças entre o DM e o DW, que podem causar uma certa confusão acerca das definições, devemos ressaltar que existem algumas diferenças. Na parte estrutural, por exemplo, a modelagem de um DM é dimensional (ideal para bancos que se relacionam com um DW). As informações podem se relacionar como um cubo: podendo aprofundar os detalhes de cada dimensão (eixo) ao se visualizar o cubo em fatias, extraindo mais detalhes do que ocorre dentro de cada parte da organização (CALDAS; SILVA, 2016).

Como já vimos, um DW pode receber dados de diversos DM. Depois dessa etapa, o processo inverso passa ser realizado, ou seja, um DW passa a alimentar os DM individualmente. A partir de então, temos duas maneiras de criar um DM, pelas abordagens *Top-down* e *Bottom-up* (CALDAS; SILVA, 2016).

A abordagem *Top-Down* se apresenta a partir da criação de um DW pela subdivisão de um DW em pequenas áreas, ou bancos de dados menores, representando determinados assuntos (por exemplo, os departamentos da empresa) (ELIAS, 2014).

A abordagem *Bottom-up* surge a partir da junção de bancos ou DM menores, essas junções resultam na criação de um DW, visando atender alguma estratégia empresarial, é o modo inverso da abordagem *Top-down* (CALDAS; SILVA, 2016). A Figura 1.7 apresenta um esquema que representa a criação usando as duas abordagens. Sobre as abordagens, Elias (2014) explica que a abordagem *Top-down* possui uma visão mais abrangente, uma vez que parte do nível mais geral (alto) e vai para o mais detalhado (baixo).

O modelo *Top-down* pode ser complexo, pois essa abordagem necessita ter uma visão maior de toda a solução (DW) e será a partir dela que os DM serão construídos. Todo o contexto precisa ser bem definido e compreendido, assim, a implementação não corre riscos de não atingir seus objetivos (CALDAS; SILVA, 2016).

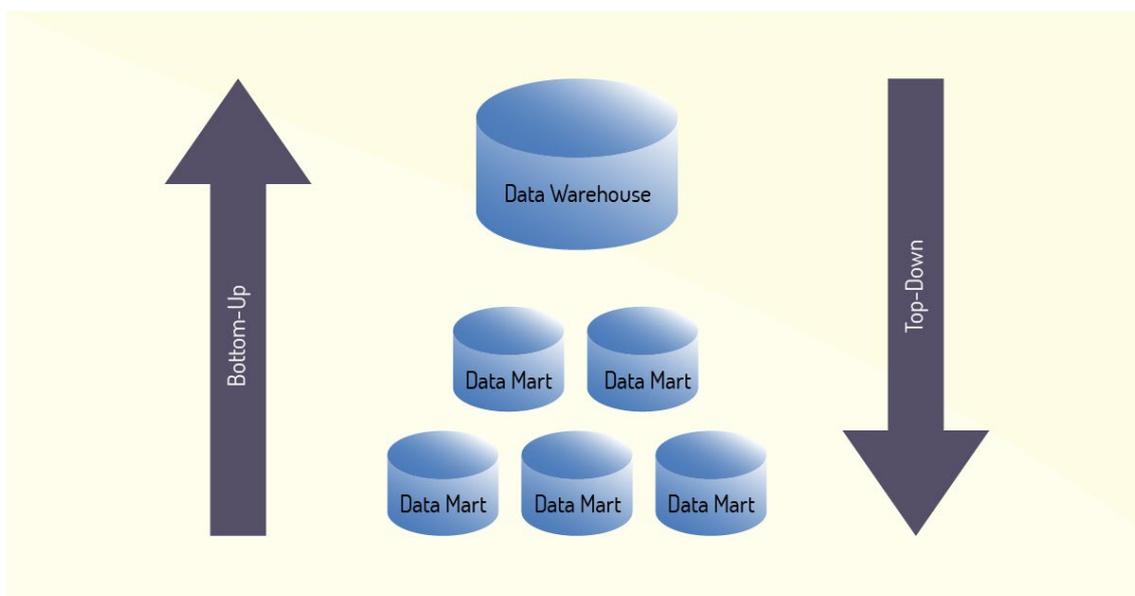


Figura 1.7 - Abordagens para criação de DM

Fonte: Elias (2014, p. 68).

Já a abordagem *Bottom-up* se baseia no conceito de dividir para conquistar, bem difundido na área de Tecnologia da Informação. Nessa concepção o problema maior é dividido, gerando problemas menores. O objetivo é criar uma solução final a partir das soluções das etapas menores. Inicialmente são implementados os DM menores para, posteriormente, construir o DW (maior). Essa abordagem apresenta menos riscos de falhas, já que aponta uma forma de evolução desenvolvida (do menor para o maior) (ELIAS, 2014).

A principal vantagem para o uso de DM é a velocidade na realização das transações, como já vimos, essa é uma das necessidades de quem utiliza esse tipo de ferramenta. Essa característica permite a participação efetiva do usuário final. Imagine que uma grande empresa que comporta alguns departamentos construa uma estrutura para análise de dados que coleta de seus usuários e que faz essa investigação de forma segmentada. A abordagem *Top-Down* pode ser implementada neste caso, uma vez que os dados podem ser todos concentrados em um grande DW e, a partir dele, vários DM podem ser criados visando ao armazenamento dos dados para os respectivos departamentos (comercial, financeiro, *marketing*) e cada DM vai armazenar os dados pertinentes aos seus interesses para análises mais detalhadas.

### ***Data Mining***

*Data Mining* (DN), ou mineração de dados, é um processo de análise de dados em grandes quantidades, normalmente relacionado às áreas de negócio ou às áreas das ciências (TAURION, 2013). A partir da busca de padrões que sejam consistentes ou sistemáticos, realiza um processo de validação para, em seguida, aplicar esses padrões nos conjuntos de dados encontrados (CALDAS; SILVA, 2016).

Com a grande massa de dados existentes (e aumentado cada vez mais) e a demanda de informações relevantes para as diversas áreas de aplicação a partir desses dados, o DN vem sendo tratado como descoberta de conhecimento (mineração) em bases de dados (GALDINO, 2016).

Para transformar essa mineração em conhecimento foi criado um processo denominado de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in*

*Databases – KDD*). Em outras palavras, o KDD é um processo que, a partir da mineração de dados, tem por finalidade a descoberta de conhecimento (MOURA, 2018).

O KDD possui várias etapas que visam criar padrões para a utilização na análise de dados. Esses padrões devem ser relevantes e de fácil compreensão, por isso esse processo pode ser considerado complexo (CALDAS; SILVA, 2016). A Figura 1.8, apresenta as etapas envolvidas no KDD.

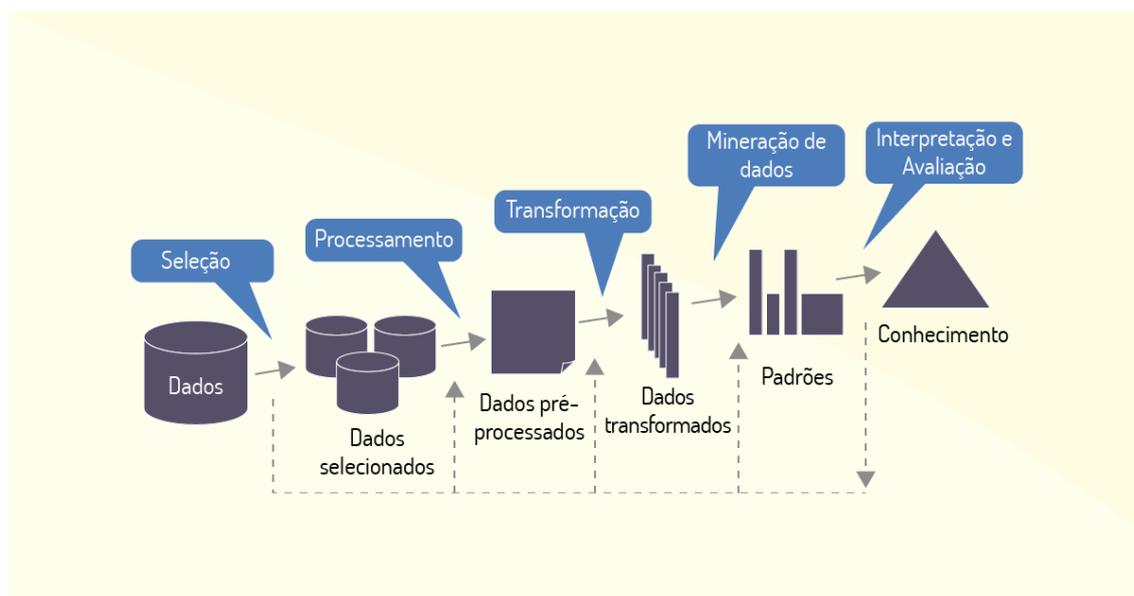


Figura 1.8 - Etapas do KDD

Fonte: Moura (2018, p. 27).

O processo KDD prevê atividades realizadas em sequência e com resultados relacionados entre as etapas, criando uma dependência entre as etapas. Moura (2018) descreve o que é realizado em cada uma dessas etapas do processo.

- **Seleção:** consiste em selecionar um conjunto de dados para análise. Esses dados são selecionados de fontes diferentes com formatos variados. Por exemplo: para mapear as atividades realizadas por usuário de um serviço podem ser selecionadas algumas informações como: sexo, idade e o número de vezes que o usuário acessou o serviço em um período de um ano.
- **Processamento:** consiste em verificar a qualidade dos dados armazenados. Atividades como limpeza, correção e remoção de dados inconsistentes,

verificação de dados ausentes, incompletos ou fora do padrão (*outliers*). Essa atividade pode ser realizada em conjunto com a primeira, pois compreende em desconsiderar valores que estão fora do padrão ou incorretos. Por exemplo: data de nascimento de séculos anteriores, valores de nome com caracteres numéricos, ou idade por extenso.

- **Transformação:** compreende em aplicar as técnicas de normalização, agregação, criação de novos atributos, redução e sintetização dos dados. São agrupados em um mesmo local para aplicação dos modelos de análise. Após a seleção dos dados é preciso colocá-los em um padrão que seja aceito pelas ferramentas de análise. Alguns SGBD já disponibilizam a opção de exportar dados em diversos tipos diferentes.
- **Mineração de Dados:** constitui-se por construir modelos ou aplicar técnicas de mineração de dados. Nessa etapa são realizadas as seguintes atividades: verificação de uma hipótese e descoberta de novos padrões de forma autônoma. As descobertas podem ser preditivas e descritivas. Esses modelos normalmente são aplicados muitas vezes, dependendo dos objetivos do projeto. Um algoritmo é utilizado pela ferramenta para criar os padrões baseando-se na leitura dos dados formatados. Por exemplo: separar os grupos de usuários de maneiras diversificadas, ou seja, baseando-se no sexo, na idade ou na frequência de utilização do serviço. Essas definições de padrões auxiliarão nas atividades de interpretação e análise.
- **Interpretação e Avaliação:** equivale a avaliar o desempenho do modelo, aplicando em outros dados que não foram utilizados na mineração. As validações podem ser feitas utilizando medidas estatísticas ou passando por avaliação de profissionais do ramo.

Após a definição do problema, dos dados utilizados e das ferramentas de análise, a pesquisa é feita pelo DN de forma automática, dentro da massa de dados, buscando por anomalias e relações, deparando-se com problemas não identificados pelos usuários.

Esse tipo de análise pode ser utilizado para extrair informações que a priori não apresentavam uma relação visível. Por exemplo: relações de compras de produtos distintos baseando-se na faixa etária ou sexo do cliente. Os resultados desse tipo de análise

podem ser utilizados para realizar sugestões de compras de novos produtos para um público específico.

Para explorar massas de dados e permitir novas descobertas dentro de um DN é possível utilizar estatística, inteligência artificial, reconhecimento de padrões, algoritmos para aprendizagem e classificação baseada em redes neurais. Nas próximas unidades deste livro, vamos aprofundar os nossos conhecimentos sobre como funciona a Inteligência Artificial (IA) e a *Machine Learning* (ou Aprendizagem de Máquina).

### ***Data Lake***

Um *Data Lake* pode ser descrito como um repositório que armazena uma abundante e diversificada massa de dados, estruturados ou não. É um importante elemento no universo *Big Data* (TAURION, 2013). O paradigma do *Data Lake* baseia-se na ideia de apenas um repositório de dados para todo o ambiente corporativo. Nele são armazenados todos os dados brutos da empresa, ficando disponíveis para quem precisar realizar uma análise sobre esses dados (MATOS, 2015).

Não obstante seja um conceito recente, o *Data Lake* surge como uma evolução nas tecnologias de armazenamento de dados em massa. Visto como fundamental, ele tem como objetivo atuar como ferramenta complementar para os demais repositórios existentes nas empresas, pois utiliza tipos de dados variados e integra tecnologias como o *Hadoop* (a principal tecnologia que oferece suporte ao *Big Data*) (PIRES, 2017).

Segundo Matos (2015), os *Data Lake* são construídos para funcionar de forma que outros repositórios possam consumir seus dados sem a necessidade de uma preparação prévia. Qualquer esquema que precise ser feito pode ser implementado a partir do momento que os dados estiverem passando por análises. Na Figura 1.9, podemos observar como o referido autor apresenta um esquema de comparação entre as formas de consumo de dados de um DW e de um *Data Lake*:

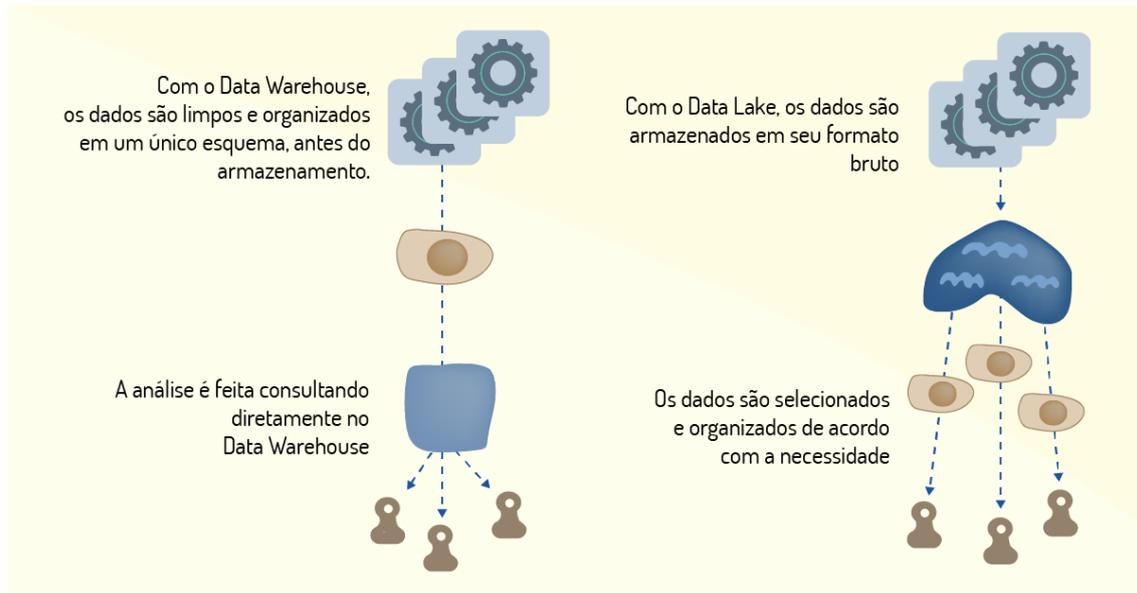


Figura 1.9 - Armazenamento DW e Data Lake

Fonte: Matos (2015, *on-line*).

Pires (2017) lista as principais características de um *Data Lake*:

- baixo custo do armazenamento e processamento de grande massa de dados;
- armazena dados estruturados e não estruturados;
- permite transformar os dados reduzindo o trabalho do ETL no caso de um DW, sem precisar consumir seus dados;
- aceita todos os dados (originais e transformados);
- alimenta diretamente um DW;
- para dados não estruturados, permite definir as suas estruturas no momento em que são utilizados;
- efetua novos tipos de processamentos sobre os dados;
- realiza análises que não seriam possíveis serem feitas em uma base de dados não relacional, e
- processa análises para temas específicos.

A vantagem do uso dos *Data Lake* está no fato deles auxiliarem nas reanálises a partir de novas investigações sobre dados existentes. O *Data Lake* não visa substituir ferramentas de análises existentes, mas complementá-las. O processo seria otimizado, visto que os

dados poderiam ser importados para um repositório com objetivos específicos e, assim, serem analisados de forma isolada (MATOS, 2015).

Atualmente, as empresas utilizam *Data Lake* como ferramenta complementar aos processos existentes para gerir e analisar dados em massa. A partir de novas análises, os dados existentes podem criar um novo significado. Por ser capaz de receber todos os tipos de dados, torna-se flexível e escalável (características importantes para ferramentas de *Big Data*), permitindo bom custo benefício em relação às demais ferramentas existentes (PIRES, 2017).

### ***Cloud Computing***

O termo *Cloud Computing*, ou Computação em Nuvem, faz referência ao armazenamento de dados em uma grande estrutura de servidores que podem ser virtuais ou físicos (TAURION, 2013). É um ambiente computacional que permite o compartilhamento de recursos e uma alta capacidade de armazenamento, admitindo que outros serviços possam ser disponibilizados através de sua estrutura (GALDINO, 2016). Dentre as suas principais características estão a possibilidade de solicitar recursos de forma antecipada (sob demanda), elasticidade no uso de seus recursos e a disponibilidade de recursos de forma expansível, a depender da necessidade do usuário.

A Computação em Nuvem funciona através de *data centers* que disponibilizam recursos de *hardware* para que empresas possam armazenar e processar dados associados aos seus negócios (TAURION, 2013). Essa forma de trabalho é conhecida como *software as service* (*software* como serviço) e que permite que uma empresa ou usuário pague para utilizar recursos computacionais de outra empresa, via internet.

Existem alguns modelos de implantação de serviços em nuvens, são eles:

- nuvem privada: uma empresa ou usuário implanta ou adquire os serviços para seu uso particular, atendendo exclusivamente suas necessidades;
- nuvem comunitária: usuários ou empresas com interesses em comum compartilham um mesmo serviço em estrutura de nuvem;
- nuvem pública: são oferecidas por empresas com grandes estruturas de armazenamentos por meio do modelo de negócio "pague para usar", geralmente com baixos custos para os usuários, e

- nuvem híbrida: estrutura mista de nuvens (pública e privada) usada geralmente para atender necessidades específicas ou baratear os custos com o serviço.

Em ambientes distribuídos, a Computação em Nuvem pode agregar valor ao processo de análise de dados em massa se for utilizada como alternativa para sistemas de distribuição com armazenamento local. Esses ambientes oferecem mais flexibilidade aos sistemas de armazenamento. Além disso, lembramos que a elasticidade é uma característica importante para sistemas que não podem definir a quantidade de dados que irão armazenar. O crescimento e a popularização dos serviços de Nuvem fizeram com que o custo ficasse bem mais acessível, permitindo que empresas de menor porte pudessem investir neste tipo de solução (GALDINO, 2016).

Empresas que já dispõem de serviços que utilizam recurso de ambiente em nuvem podem desenvolver aplicações para utilizar essa infraestrutura existente sem demandar muito custo adicional. Pequenas empresas que ainda não possuem grandes cifras para investimento em infraestrutura podem adquirir pequenos pacotes de serviços, aumentá-los, caso necessário, e, em seguida, reduzir sua utilização, baixando novamente os custos (TAURION, 2013).

Alguns autores descrevem as vantagens de utilizar ambientes em nuvem para realizar análise de dados. Caldas e Silva (2016) relatam que em um sistema distribuído híbrido, composto por serviços privados e usado para implantar sistemas de análise de dados pode ser mais econômico que manter uma estrutura física em um centro de tecnologia de grande ou médio porte. Esse sistema pode ser ampliado com o uso sob demanda de nuvens públicas para armazenamento ou projetos de curta duração.

Outra aplicação relatada pelos autores seria a divisão de fontes (internas e externas). Algumas empresas preferem manter seus dados sob sigilo, por questões de privacidade. Porém, para reduzir custos e manter um sistema desse porte, serviços de nuvens (públicas ou privadas) podem ser contratados para a realização das análises dos dados.

Galdino (2016) alerta que além dos altos custos gerados para implantar e realizar a manutenção em sistemas distribuídos de grande porte (*mainframes*), em alguns casos, a sua capacidade não é totalmente explorada, gerando desperdício de recursos. Os sistemas em nuvem possuem características que podem ser úteis para sistemas de *Big Data*, como configurações por demanda de usuários (número de acessos, por exemplo). Essas

configurações podem ser variar de acordo com situações específicas ou necessidade de um desempenho maior devido a processamento de uma massa de dados em quantidade além do normal. Datas comemorativas ou épocas de vendas elevadas causadas por períodos de grandes descontos, como *Black Friday*, podem gerar situações em que o número de cadastros, visitas a *sites* de lojas e liberação de promoções precisem ser processados em quantidade e velocidade diferentes ao do fluxo normal.

Outra forma de utilização indireta para *Big Data* é contratar empresas que realizam o trabalho de processar dados sob demanda utilizando nuvens, conhecidas como Serviços de Análise (ou *Analysis as a Service*, AaaS). As organizações que dispõem desse tipo de serviço utilizam sistemas híbridos de nuvem pública e privada para manter os riscos sob controle e, quando for demandado, aumentar a capacidade de processamento e armazenamento dos sistemas. Na próxima seção, vamos conhecer alguns sistemas que utilizam essas estruturas tecnológicas para realizar armazenamento e análise de dados de forma massiva.

## ATIVIDADE

3) Com o aumento da quantidade de massas de dados geradas no advento da Web 2.0 foi necessário modificar a forma de armazenamento. Os bancos de dados existentes não comportavam mais esses tipos de dados que passaram a ser modelados de diversas formas (MATOS, 2019). Considerando esse contexto, assinale a alternativa que contém a modelagem de dados utilizada em ferramentas de *Big Data*.

- a) SQL (*Structured Query Language*).
- b) NoSQL (*Not Only SQL*).
- c) ETL (*Extract, Transforming and Loading*).
- d) OLAP (*On-line Analytical Process*).
- e) KDD (*Knowledge Discovery in Databases*).

## Exemplos de usos e casos

As aplicações para as ferramentas de *Big Data* ocorrem em diversas áreas de atuação e visam auxiliar os processos organizacionais que dão suporte às tomadas de decisões. Essas aplicações podem, ainda, criar novas atividades: como realizar previsões sobre tendências no mercado (SOBREIRO, 2018). De acordo com Taurion (2013), algumas aplicações práticas com as soluções de *Big Data* são:

- sugestões de sites de compras para os usuários como Amazon, Americanas, entre outros;
- recomendações de ferramentas de *streaming* como Netflix e Youtube nas páginas que os usuários estão acessando em um determinado momento, e
- com a ajuda das análises dos dados recebidos da ferramenta, uma companhia de viagens percebeu que uma das preocupações de seus usuários era com a bagagem. Para solucionar isso, criou um recurso em seu aplicativo para que os usuários realizassem o rastreamento de suas bagagens.

Vale ressaltar que o *Big Data* obteve maior espaço nesses últimos anos com o advento da IoT. Por meio de vários dispositivos e/ou objetos conectados à *internet* produz uma massa de dados que reúne padrões sobre o uso de determinado produto pelo cliente, bem como o desempenho que esse produto ofereceu ao usuário (TAURION, 2013).

Galdino (2016), por sua vez, apresenta outras áreas de aplicação para ferramentas de *Big Data*:

- a área de saúde faz uso de ferramentas de análise de dados em massa. Ao se observar esses dados é possível realizar a previsão sobre uma possível epidemia, por exemplo, apenas cruzando informações oriundas de redes sociais, ferramentas de pesquisa ou dados de consultas médicas. Galdino (2016) relata os casos do uso da ferramenta *Google Trends* (que cria um *ranking* sobre termos e assuntos mais pesquisados no *Google*) para prever um surto da doença H1N1. E, também, o caso do instituto do coração que utiliza ferramentas de análise de dados para monitorar o avanço e as mutações relacionadas ao vírus HIV.
- empresas de transporte aéreo tem um gasto considerável com combustível: em torno de 30%. Uma empresa americana utilizou os recursos de *Big Data* para

tentar reduzir esses custos. Ela instalou sensores em diferentes partes do avião para coletar vários tipos de dados. Com um volume médio de 640 *Terabytes* em um voo transatlântico é possível conseguir informações que auxiliam nas tomadas de decisão sobre as revisões que serão feitas na aeronave, baseando-se no desgaste e tempo de utilização das peças e sistemas mecânicos da aeronave, bem como saber quando será necessário realizar novos abastecimentos.

- a segurança pública também se utiliza das informações geradas por ferramentas de *Big Data* para preparar estratégias com o intuito de reduzir os crimes. Posterior aos acontecimentos do 11 de setembro, autoridades americanas passaram a usar esses recursos como ferramenta de trabalho. O estado do Tennessee instalou inúmeras câmeras e sensores que passaram a captar informações fora do seu sistema usual de segurança. Desta maneira, foi possível investigar crimes que já haviam acontecido e também reduzir o número de novos delitos. Com essas informações, a polícia conseguiu otimizar as estratégias em caso de perseguição, baseando-se nas rotas de fuga que tinham conhecimento, além de distribuir de forma mais eficiente o seu efetivo em locais onde os crimes geralmente ocorriam. Outro caso relacionado aconteceu em Boston, em 2013. Após analisar imagens enviadas por terceiros e utilizando ferramentas de análise foi possível mapear situações e pessoas suspeitas. Com o cruzamento dos dados a polícia conseguiu solucionar o caso com a prisão do responsável pelo atentado ocorrido durante uma maratona na cidade.

Moura (2018) apresenta alguns casos relacionados à construção de soluções utilizando ferramentas de *Big Data*. As soluções atendem às mais variadas necessidades e empregam múltiplos recursos, como veremos a seguir. Os casos apresentados pela autora são resultados de entrevistas realizadas com profissionais responsáveis por criar soluções que tinham como suporte as tecnologias para *Big Data*.

### **Caso 1: desenvolvimento de algoritmo para análise de dados**

O caso relatado por Moura (2018, p. 42) apresenta o desenvolvimento de uma ferramenta para análise de dados (relatórios) por meio do uso de algoritmos e utilizando como suporte

a linguagem de programação R. A empresa solicitante trabalha com a produção de modelos para análise de dados que atendam às necessidades de seus clientes, dentre as quais se apresentam: controles de estoque, identificação de anomalias, previsão de vendas, identificação do perfil do consumidor.

A Figura 1.10 apresenta as etapas da solução desenvolvida no caso 1. Depois de uma entrevista inicial, em que foi possível identificar as necessidades do cliente, foi feita uma modelagem para acesso aos dados. Para tanto, foi necessário ter acesso à base de dados completa do cliente e, por fim, criar uma modelagem para implementação da solução.

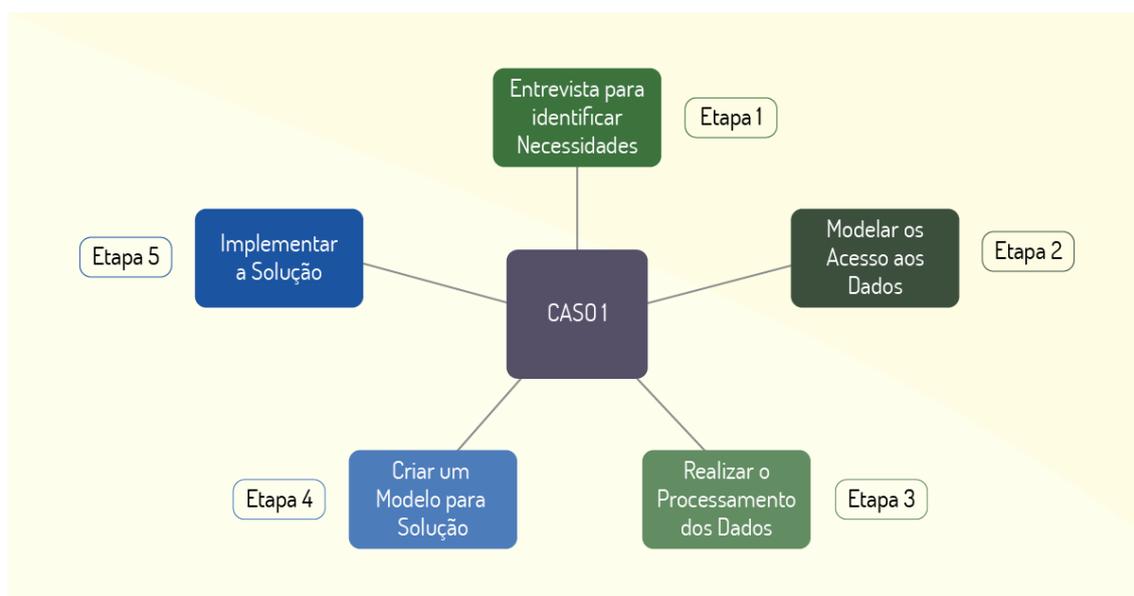


Figura 1.10 - Etapas do Caso 1

Fonte: Elaborada pela autora.

A solução usa modelos matemáticos e estatísticos, sempre verificando qual a melhor estratégia que será empregada (modelos determinísticos ou probabilísticos). Também foram aplicados algoritmos de *machine learning* para realizar descobertas utilizando as bases de dados existentes e já remodeladas. Após a implementação, uma validação é feita para a verificação por parte dos clientes. Nela analisam-se cenários já conhecidos antes da solução e outros que surgiram durante o processo. Segundo a autora, o profissional responsável afirmou em seu relato que a maior dificuldade encontrada foi a modelagem

da solução, uma vez que o modelo desenvolvido deveria atender à grande quantidade de dados e com objetivos distintos.

## **Caso 2: transformação da cultura empresarial para análise de dados**

O Caso 2 debatido por Moura (2018, p. 45) apresenta um projeto cujo objetivo era o de transformar a cultura de uma empresa em uma cultura de *data driven* (cultura orientada pela análise de dados). Ou seja, as decisões e modelagens de novos processos precisariam ser orientadas pela análise dos dados coletados pela empresa. Para realizar tal mudança foi proposta uma solução com a participação ativa dos membros da empresa. O resultado não apresentou apenas a utilização de uma ferramenta, mas sim uma mudança de cultura de trabalho, o que demandou um tempo maior de execução.

A solução se inicia com a realização de uma *hackathon* (eventos de curta duração para construção de soluções de problemas específicos, utilizando programação): os profissionais da empresa e os clientes formaram times para criar projetos de análise de dados. Durante o evento, os times exploraram diversas formas de tomada de decisão. Outros pontos também são analisados: as fontes de onde as informações são coletadas e as formas de armazenamento dessas informações. Essa etapa auxiliou na escolha da modelagem dos dados coletados. A Figura 1.11 apresenta o esquema da solução realizada.

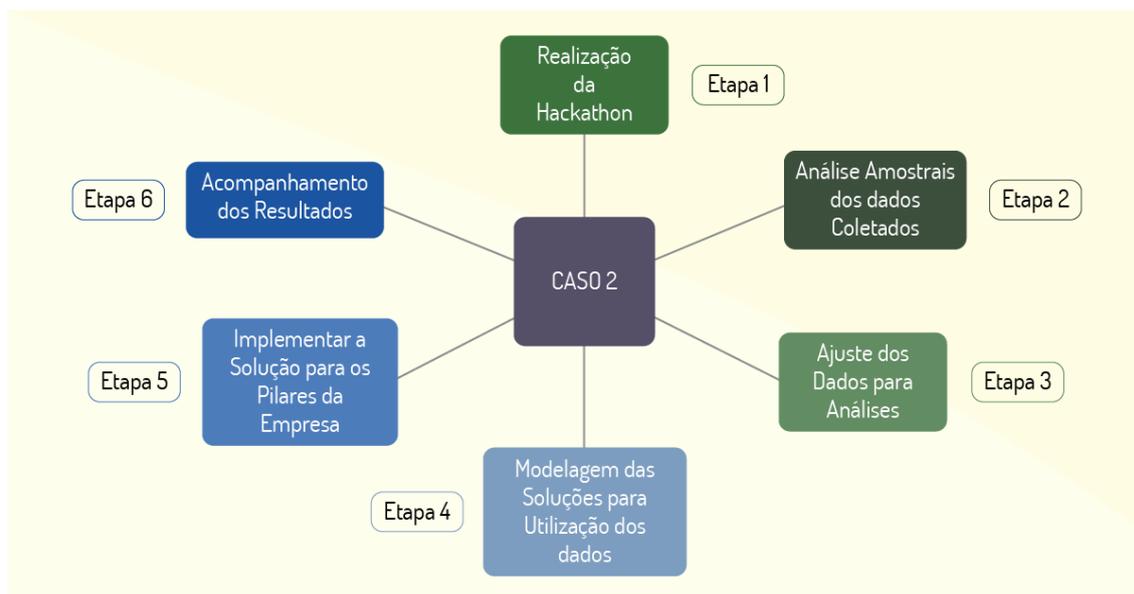


Figura 1.11 - Esquemas da Solução do Caso 2

Fonte: Elaborada pela autora.

Depois da *hackaton*, foi realizada uma análise dos dados coletados em uma modelagem para que os dados pudessem ser apresentados de forma passível de análise pelo cliente. Desta maneira, as análises de dados passam a ser efetuadas de forma constante, como uma atividade cotidiana da empresa. Em seguida, foram criados modelos de soluções que atendessem às necessidades da empresa. As modelagens foram validadas com os gestores de negócio do cliente, que indicaram as mudanças necessárias nas modelagens apresentadas. A implementação tinha como guia atender os pilares do novo projeto de *Data Driven* da empresa: negócio, ciência e tecnologia. Foram disponibilizados também: análises, relatórios, *dashboards* e treinamentos. Na composição da solução foram utilizadas as linguagens de programação R e *Python*, que dão suporte à implementação de ferramentas de *machine learning* (utilizada em soluções de *Big Data*) e a ferramenta Excel (para apresentação dos dados estatísticos).

O profissional entrevistado afirmou que o maior desafio foi o treinamento dos clientes para a cultura de analisar e processar dados, pois este tipo de atividade demandou expertises específicas que levaram um tempo para serem desenvolvidas, como a capacidade analítica dos indivíduos.

### **Caso 3: construção de *software* para tomada de decisão**

O caso 3, apresentado por Moura (2018, p. 46), trata de um projeto para construção de um *software* para análise de dados e tomada de decisão. O cliente utiliza os resultados das análises de dados para áreas de *marketing*; para pesquisar perfil e satisfação; para recursos humanos; para avaliação 360 e pesquisa de clima; para qualidade; para auditorias e análise de *outliers*; para inteligência, ouvidoria e SAQ (*Safety Attitude Questionnaire*, indicador de qualidade).

Inicialmente, após as reuniões de entendimento com o cliente, foram criados protótipos para que o cliente pudesse ter uma aproximação com o projeto que seria desenvolvido. Foram realizadas algumas reuniões para demonstrar as produções desses protótipos, para que tudo que estava sendo desenvolvido fosse validado e para apresentar diagramas de caso de uso (utilizados para auxiliar no entendimento do problema).

Terminada a etapa de definição da solução, foi analisada a forma como os dados do cliente estavam estruturados. Essa análise foi feita para verificar se os dados estavam no formato ideal para o novo produto que estava em desenvolvimento. Depois que se analisaram as estruturas utilizadas para o armazenamento da massa de dados, foi a vez dos dados serem de fato analisados para validar os modelos propostos na solução. Por fim, entregou-se a solução que foi proposta inicialmente. A Figura 1.12 apresenta as etapas do processo da composição da solução encontrada.

Como solução final, observamos um módulo de relatórios que tinha como objetivo auxiliar as análises de dados e as tomadas de decisão. Esse módulo foi adicionado ao *software* existente do cliente. Para a implementação da ferramenta foram empregadas as linguagens de programação PHP, Java, *Python* e R.

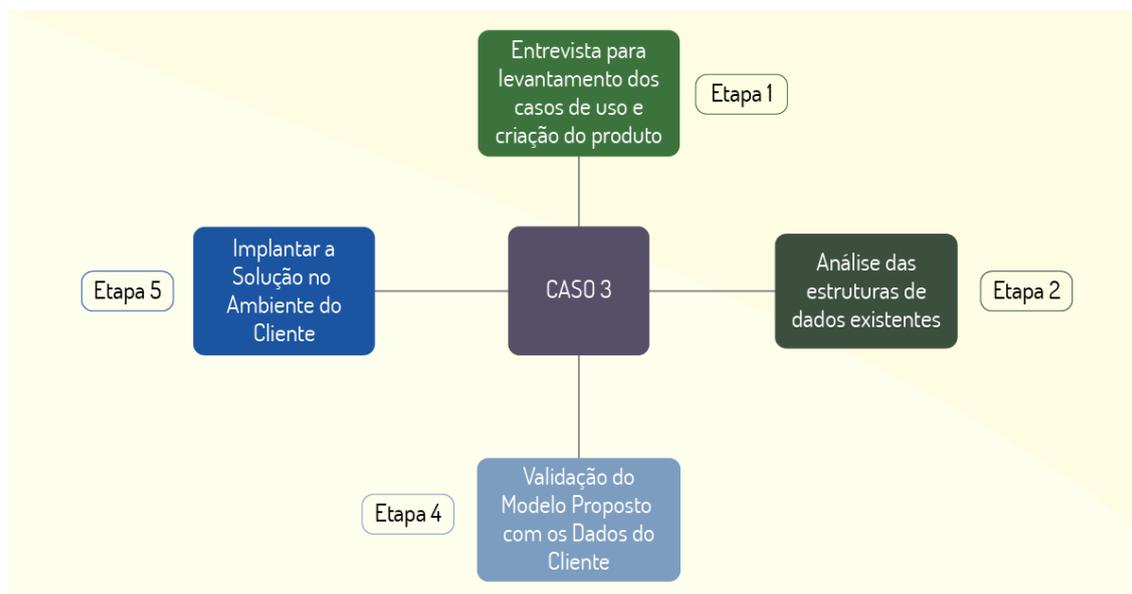


Figura 1.12 - Etapas da Solução do Caso 3

Fonte: Elaborada pela autora.

A maior dificuldade encontrada nos relatos dos especialistas envolvidos no projeto foi desenvolver a parte responsável por mostrar os resultados da análise de forma visual, usando gráficos e tabelas, pois elas precisariam ser rápidas e precisas.

#### **Caso 4: consultoria para análise de novos negócios**

De acordo com os estudos feitos por Moura (2018, p. 49), o caso 4 comporta a entrega de relatórios e *dashboards* baseados em análise de dados, além da realização de uma consultoria cultural que visava auxiliar o cliente nas previsões voltadas para análise de novos negócios, *marketing* e gerenciamento de crises. Inicialmente, em uma conversa com o cliente foi possível identificar os dados necessários para realização das atividades e o nível de maturidade da empresa nas atividades relacionadas a esse tipo de funcionamento.

Após a primeira etapa, foi feita uma análise dos dados utilizados pela empresa. Em seguida, houve uma preparação para que se adequassem as modelagens utilizadas nas soluções entregues. Posteriormente, foi realizado um mapeamento dos tipos de análise

que poderiam compor a solução do problema encontrado. Finalmente, foi implementada uma solução para ser entregue à empresa, como mostra a Figura 1.13.



Figura 1.13 - Etapas da Solução do Caso 4

Fonte: Elaborada pela autora.

A solução entregue foi composta de relatórios, *dashboards* baseados nos dados analisados do cliente e de um relatório de consultoria que visava mudar a cultura da empresa. A empresa passou por uma etapa de acompanhamento para verificar as mudanças em relação às análises de dados praticadas pela empresa. Para a solução foram utilizadas as linguagens de programação R e Python e as ferramentas Qlick Sense (para os *dashboards*), Excel e SAS (*Statistical Analysis System*). As maiores dificuldades encontradas foram realizar a mudança de cultura, explicar os contextos relacionados às análises de dados e tomar decisões baseadas nesses princípios.

### Caso 5: solução para previsão de vendas e novos produtos

O caso 5, explorado por Moura (2018, p. 52), apresenta uma solução conjunta para uma empresa que utiliza análise de dados em massa para otimizar a distribuição de estoque, previsão e sugestão de vendas de novos produtos e análise de sentimento.

A primeira etapa realizada foi um diagnóstico feito durante as visitas ao cliente. Essa etapa visou responder a dois questionamentos iniciais: a) como estava organizada a base de dados; b) quais atividades (análises) eram realizadas com os dados existentes. Após essa etapa seria possível delinear possíveis soluções para os problemas encontrados. Um diagnóstico foi feito, e as validações das soluções foram apresentadas ao cliente.

Posteriormente, realizou-se uma atividade de acesso aos dados, explorando e fazendo uma análise histórica dos dados. Buscou-se, ainda, padronizar os dados para as próximas etapas: a modelagem, o desenvolvimento e a implantação da solução. A Figura 1.14 nos mostra esse processo:



Figura 1.14 - Etapas de Solução do Caso 5

Fonte: Elaborada pela autora.

A solução apresentada, além de contemplar os relatórios, *dashboards*, *softwares* baseados nos algoritmos utilizados para análise dos dados, também ofereceu um treinamento para

auxiliar a empresa a implementar o processo de *data driven* (cultura orientada a análise de dados). Após a implementação da solução, foi oferecido um acompanhamento para diluir as dúvidas encontradas em relação a adaptação dos processos referentes às novas políticas de análise de dados e tomada de decisão. Para o desenvolvimento das soluções foram utilizadas as linguagens de programação R e *Python* e as ferramentas Excel, *Hadoop* e *Spark*.

E com esses exemplos finalizamos o nosso estudo sobre *Big Data*. Nessa primeira unidade vimos as definições, as características, as ferramentas utilizadas e as diversas aplicações para esta área.

#### **ATIVIDADE**

4) Segundo previsões, a internet das coisas (*Internet of Things* - IoT) será uma das maiores responsáveis pela geração dos dados processados em ferramentas de *Big Data*. Dentre as alternativas a seguir, assinale a que se refere a uma aplicação de IoT e que utiliza ferramentas de *Big Data*.

- a) Envio de informações sobre batimentos cardíacos realizado em uma atividade física encaminhado por meio de uma *smart band*.
- b) Recomendações de novas conexões de uma rede social, baseado no ciclo de amizade atual.
- c) Descontos gerados para um cliente imediatamente após atingir um valor total de compras no mês.
- d) Recomendações de novos vídeos de um artista no Youtube, baseado em um vídeo assistido e utilizando ferramentas de *machine learning*.
- e) Sugestão de compra de um produto, baseado em compras conjuntas realizadas por outros clientes.

## INDICAÇÃO DE LEITURA

Título: *Big Data*: O futuro dos dados e aplicações

Autor: Felipe Nery Rodrigues Machado

Editora: Érica

ISBN:9788536527000

Comentário: Esse livro se propõe a apresentar os principais fundamentos de *Big Data*, seu histórico e sua utilização. Apresenta as diferenças entre as ferramentas que utilizam *Big Data*. Mostra, ainda, as tecnologias empregadas e as aplicações reais dessas ferramentas e tecnologias.

UNIDADE II

# Internet das Coisas

*Aline Chagas Rodrigues Marques*

*Marcelo Takashi Uemura*

## Introdução

Caro(a) aluno(a), na segunda unidade da disciplina de Tópicos Especiais em Sistemas de Informação, veremos conceitos relacionados à internet das coisas, IDC – acrônimo equivalente em português (IoT, em inglês).

A IoT é considerada uma das principais fontes de dados para o Big Data. O surgimento do conceito de IoT ocorreu em 1999, por Kevin Ashton, com pesquisas no campo da identificação por radiofrequência (RFID, em inglês, *Radio Frequency Identification*) e tecnologias de sensores.

Mas o que é a internet das coisas? Consiste em uma rede de objetos físicos capaz de reunir e transmitir dados, por exemplo, a conexão de carros, veículos e outros objetos dotados de sensores e conexão com a internet. Não é fantástico?

Nesta unidade, iniciaremos com uma introdução sobre a internet das coisas, com a abordagem da história e arquitetura relacionada a IoT, bem como seus aspectos relevantes, como a segurança em seu desenvolvimento e utilização. Em seguida, estudaremos as tecnologias envolvidas utilizando os protocolos *Zigbee*, *6LowPan*, *Bluetooth*, *Lora*, *NB-IOT*, e, por fim, a demonstração de usos e cases utilizados em internet das coisas.

Preparado(a)? Vamos começar!



Fonte: everythingpossible / 123RF.

## **Introdução à Internet das Coisas**

Caro(a) aluno(a), como sabemos, o surgimento do conceito de internet das coisas ocorreu em 1999, mas está em processo de desenvolvimento há décadas. O primeiro aparelho conectado à internet com evidência de IoT surgiu na década de 1980, com a utilização da máquina de Coca-Cola, para determinar a existência ou não de bebida gelada na máquina. Em 1999, a expressão “internet das coisas” foi apresentada pela primeira vez pelo tecnólogo Kevin Ashton, quando a utilizou para descrever como os dados capturados pelos humanos levariam a uma revolução, uma vez que os computadores começaram a gerar e coletar dados sem qualquer intervenção humana. Desde então, a evolução da internet das coisas está convergindo em múltiplas tecnologias, seja por meio de comunicação sem fio ou por diversos sistemas embarcados.

O termo está em expansão desde a década de 1990. Em termos gerais, “internet das coisas” representa dispositivos que detectam diversos aspectos do mundo real, como sensor de temperatura, iluminação, presença ou ausência de objetos e pessoas, além de mostrar os dados do mundo real e o agir sobre a existência desses dados. A utilização dos dados, em vez de serem consumidos pelas pessoas, seriam consumidos pelas máquinas, pela comunicação entre as máquinas, para, assim, melhorar a qualidade de vida dos humanos.

Mas o que exatamente são “coisas”? Entre os diversos contextos de coisas, o termo pode ser representado por qualquer dispositivo que apresenta incorporação com aparelhos eletrônicos, variações de softwares e sensores que realizam comunicação com outros dispositivos.

Existem algumas divergências em relação ao conceito de IoT, não existindo um conceito único sobre o termo. De forma geral, IoT é considerado um ambiente de objetos físicos interconectados com a internet, por meio de diversos sensores pequenos e embutidos, como forma de criar um ecossistema de computação ubíqua para facilitar o cotidiano das pessoas. Essa computação ubíqua representa a presença constante da informática e da tecnologia na vida das pessoas, assim como em suas casas e no ambiente de convívio social.

As seções a seguir mostram a história da internet das coisas e sua arquitetura básica.

## História da IoT

O primeiro dispositivo desenvolvido especificamente para a internet das coisas surgiu em 1990, com a criação da torradeira inteligente apresentada por John Romkey, na conferência INTEROP, em 1989. Essa torradeira surgiu quando o presidente Dan Lynch da conferência prometeu a John Romkey a seguinte situação: caso a torradeira conseguisse ligação com a internet, o aparelho seria representado na conferência.

Diante desse desafio, John Romkey conectou a torradeira inteligente a um computador utilizando a rede TCP (*Transmission Control Protocol*)/IP (*Internet Protocol*) e, dessa forma, a torradeira tornou-se um grande sucesso na conferência. Porém, nesse teste, o pão foi incluído manualmente na torradeira. Em seguida, o requisito foi corrigido e apresentado na conferência INTEROP do ano seguinte já com a utilização do guindaste robótico, que possuía conexão com a internet: assim, pegava a fatia de pão e colocava na torradeira de forma automatizada.

Em 1999, Kevin Ashton, o cofundador da Auto-ID Center, realizou uma palestra para a Procter & Gamble a fim de mostrar uma nova ideia do sistema RFID utilizado na rastreabilidade do produto na cadeia de suprimentos. Como forma de manter a atenção dos executivos da Procter & Gamble, a apresentação possuía no título a expressão *Internet of Things*, e, por isso, Kevin Ashton é considerado o criador dessa expressão.

Em 2005, a discussão sobre a internet das coisas tornou-se pauta da *International Telecommunication Union* (ITU), agência das Nações Unidas para as tecnologias da informação. Nesse relatório, a internet das coisas descreve como dispositivos e objetos do dia a dia, com transmissores, sensores e receptores, possibilitam novas comunicações entre objetos e objetos e entre pessoas e objetos, em qualquer lugar e tempo.

Em 2009, Rob Van Kranenburg publicou o livro “*The Internet of Things*”, com o relato sobre esse novo paradigma pelo qual os objetos conectados produzem informação. Além disso, a Cisco IBSG relatou a existência de mais objetos conectados, tais como *smartphones*, *tablets* e computadores, com relação à população mundial. Esse período é considerado o ano do nascimento da Internet das Coisas.

A partir de 2015, a internet das coisas passou a ser considerada uma realidade, com cerca de 4,9 bilhões de dispositivos conectadas e com utilização real, obtendo um aumento de 30% em relação a 2014 e com tendência para atingir 25 bilhões até 2020. Atualmente, a

internet das coisas possui suporte da comissão europeia, por meio do programa Horizon 2020, com cerca de 80 milhões de euros de financiamento disponíveis ao longo do período de 2014 a 2020 (EUROPEAN COMMISSION, 2019). Em 2020, estima-se que 50 bilhões de equipamentos estarão conectados à internet em todo o mundo, sendo que um terço dos dispositivos serão computadores, *smartphones*, dispositivos móveis e TVs; os tipos de “coisas” como atuadores, sensores e dispositivos inteligentes representaram os outros dois terços. Essas “coisas” são consideradas as invenções que monitoram, controlam, analisam e otimizam a tecnologia mundial.

A Figura 2.1 mostra a estimativa e o crescimento da IoT até 2020.

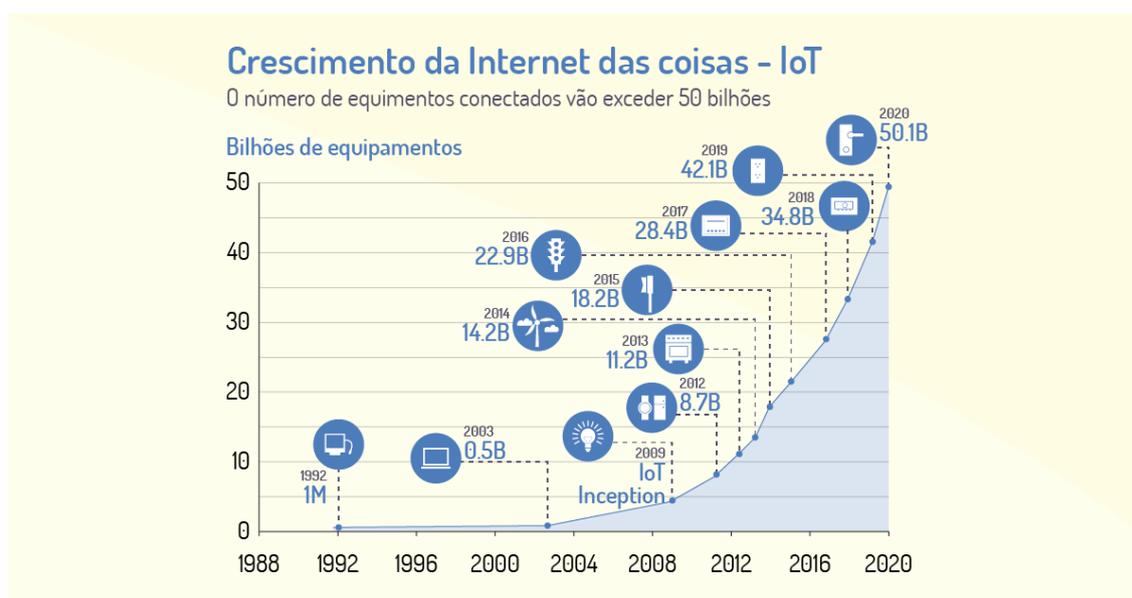


Figura 2.1 - Estimativa e crescimento da IoT até 2020

Fonte: Santos (2018, p. 51).

Segundo Evans (2011), até 2020, 50 bilhões de dispositivos serão conectados à internet. Desse total de dispositivos, a população mundial conectada a eles será de 7,6 bilhão de pessoas. Além disso, possui uma previsão de seis dispositivos conectados por pessoa. Diante dessas observações, existe uma perspectiva da existência de mais dispositivos conectados do que pessoas no mundo.

## Arquitetura da IoT

A arquitetura atual da internet não possui suporte para imersão de grande número de objetos conectados. Com diversos dispositivos da IoT produzindo dados, uma grande diversidade de informação pode gerar um colapso da internet atual, gerando a necessidade de um maior armazenamento de dados para guardar informações desses dispositivos.

Mas o que deve ser adaptado na internet atual para utilizar a arquitetura IoT? Com diversos nós conectados entre si e com a internet, um grande número de endereços de rede deve existir. Com isso, o padrão IPv4 utiliza 32 bits para o endereçamento de máquinas, permitindo a alocação de, aproximadamente, 4 bilhões de dispositivos ( $4 \times 10^9$ ); entretanto, mesmo com esse número, já não é mais suficiente para integrar os diversos dispositivos que estão se conectando à internet, como os sensores IoT, *smartphones*, *notebooks*, dentre outros. Em complemento, para suportar a adição de novos dispositivos à grande rede de computadores, o protocolo IPv6 utiliza 128 bits para a alocação dos endereços IPs, o que equivale a, aproximadamente, 340 undecilhões de endereços ( $1 \times 10^{36}$ ).

Adicionalmente, uma **arquitetura de IoT baseada em componentes** é composta por: dispositivos denominados **objetos (coisas)**, que são entidades físicas e/ou virtuais que podem ser identificadas de acordo com os serviços associados; **gateway**, que é o responsável por intermediar as conexões com a internet para os dispositivos que não conseguem realizar comunicação via protocolo TCP/IP, cujos dados são normalmente armazenados e analisados em nuvem (Cloud), sendo acessados pelas aplicações. Também faz parte da arquitetura IoT o **middleware**, que contempla as ferramentas e os mecanismos que apoiam os funcionamentos desses objetos e da rede IoT como um todo.

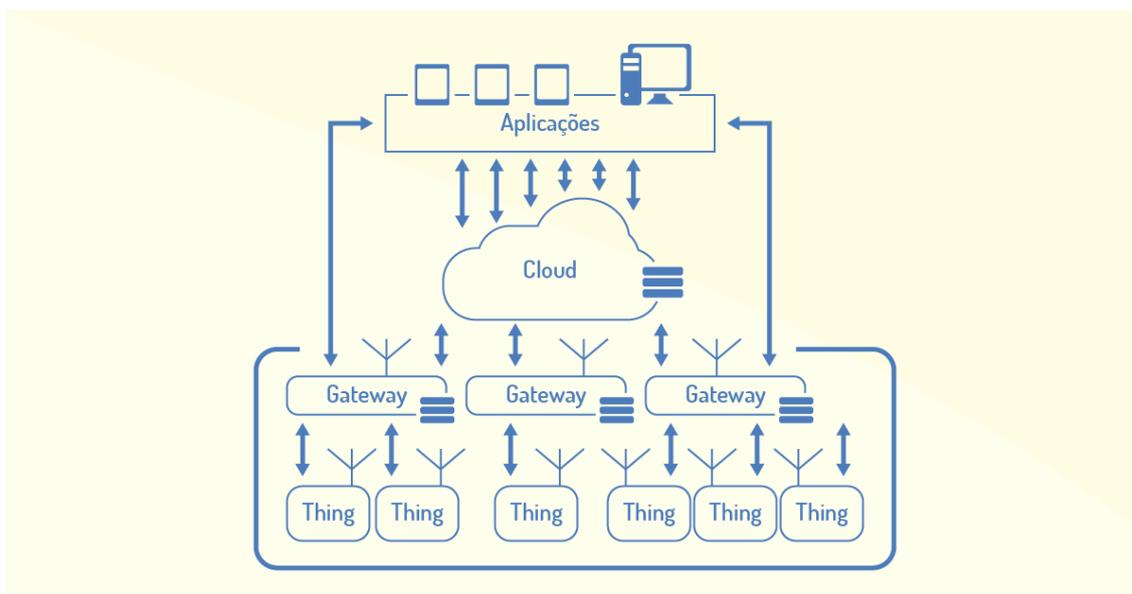


Figura 2.2 - Arquitetura IoT

Fonte: Adaptada de Barros (2016).

Além disso, essas arquiteturas de IoT possuem camadas com as mesmas funções ou funções muito próximas. As camadas servem para prover informações referentes à comunicação na rede, assim como ocorre no modelo TCP/IP utilizado na internet. A arquitetura básica da IoT utilizada como ponto de partida possui três camadas: percepção/atuação, rede e aplicação. A Figura 2.3 mostra a **arquitetura de três camadas da IoT**.

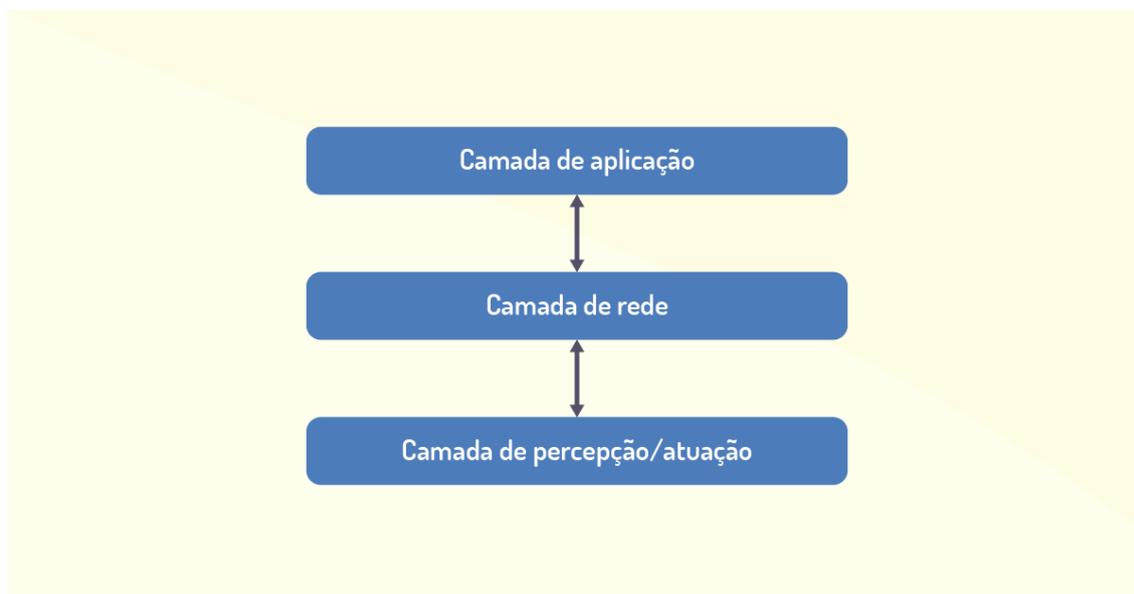


Figura 2.3 - Arquitetura de três camadas da IoT

Fonte: Elaborada pelos autores.

A **camada de percepção** é responsável pela conversão de grandezas físicas do ambiente real em formato do mundo virtual, possibilitando a transmissão dos dados através da rede. Nessa camada de percepção estão as tecnologias de IoT aplicadas em identificação, que coletam informação do mundo real e as convertem para o mundo virtual. Alguns exemplos da camada de rede são as **tags RFID, código de barras, câmeras digitais, GPS, sensores e a rede de sensores**.

Em seguida, a **camada de redes** é responsável pela transmissão de informação da camada de percepção/atuação e enviada para a camada de aplicação. Essa camada de redes possui tecnologias utilizadas para comunicação, assim como o processo de gerenciamento e distribuição de mensagens utilizando roteadores e *gateways*. Por fim, a **camada de aplicação** é responsável por processar as informações recebidas pela camada de rede e fornecer serviços para os clientes ou outras aplicações.

O *Gateway* está posicionado na camada de redes e possui uma grande importância nessa arquitetura de camadas, sendo responsável por conectar diversos dispositivos à internet, realizando a tradução entre as diversas tecnologias utilizadas. Em IoT, esses dispositivos apresentam tecnologias de comunicação heterogêneas, por exemplo, o uso de protocolos como *Bluetooth*, *ZigBee*, *Lora*, dentre outros, apresentando como principal característica

o baixo consumo de energia, tendo em vista a mobilidade e uma maior autonomia de baterias internas.

Um exemplo de uso do *Gateway* são as pulseiras e relógios inteligentes, que coletam dados do usuário pelos sensores (pressão sanguínea, batimentos cardíacos, movimentação etc.) e utilizam *Bluetooth* para enviar os dados para o *smartphone* do usuário (que atua como o *Gateway*), que, por sua vez, encaminha os dados para um servidor (geralmente em nuvem), onde são processados, armazenados e disponibilizados para aplicações (*web* ou celular).

Com a grande quantidade de dispositivos a serem conectados em produtos IoT, uma outra representação da arquitetura para IoT é a **arquitetura orientada a serviços**, conhecida como SOA (*Service Oriented Architecture*), que pode ser utilizada para facilitar o desenvolvimento de aplicações. Além disso, essa arquitetura facilita a comunicação entre os dispositivos, pois relacionam a heterogeneidade de *software*, *hardware*, formato de dados e comunicação.

A arquitetura orientada a serviços, proposta por Atzori, Iera e Morabito (2010), é composta por cinco camadas: objetos, abstração de objetos, gerenciamento de serviços, composição de serviços e aplicações. Dessas cinco camadas, três representam o *middleware* da arquitetura: abstração de objetos, gerenciamento de serviços e composição de serviços. A Figura 2.4 mostra as camadas da arquitetura orientada a serviços.

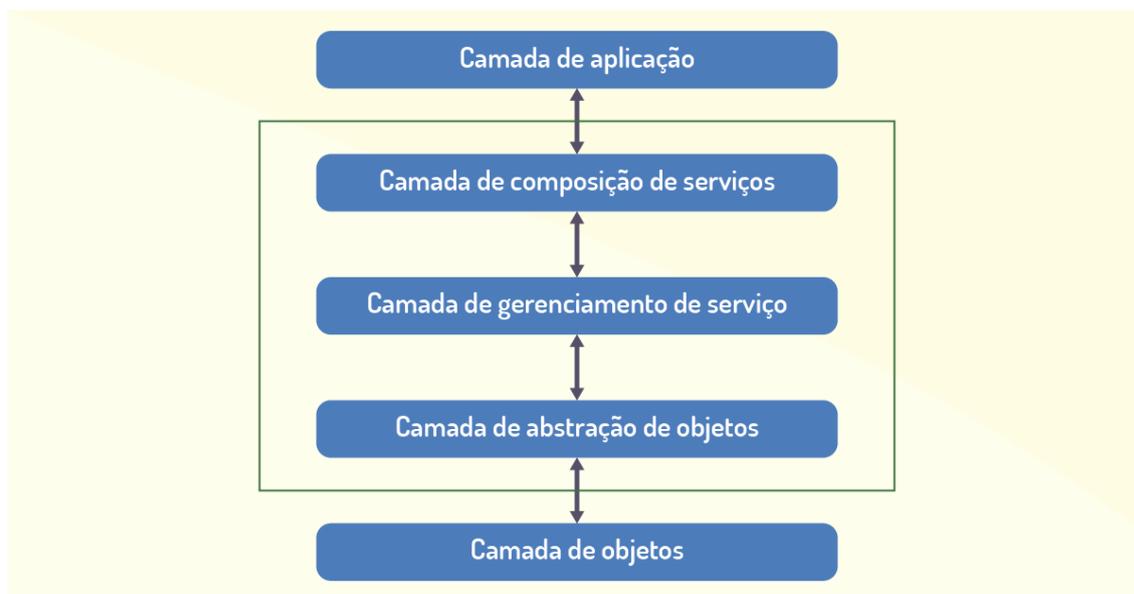


Figura 2.4 - Camadas da arquitetura orientada a serviços

Fonte: Elaborada pelos autores.

A camada objetos representa os dispositivos da solução de IoT, ou seja, as coisas independentemente do local e soluções; sendo assim, os objetos são fonte de dados dessas soluções utilizando a IoT. Os produtos IoT apresentam capacidade de percepção e, assim, os produtos devem transmitir dados para a camada superior dessa arquitetura, a fim de alcançar o objetivo da arquitetura orientado a serviço. Como exemplo podemos citar os objetos inteligentes, que possuem a capacidade de percepção e devem transmitir seus dados para as camadas superiores da arquitetura.

Com relação às três camadas que representam a *middleware* da arquitetura, a primeira camada, abstração de objetos, é responsável por receber os dados da solução, em seguida interpretar e, caso seja necessário, realizar a conversão para a linguagem utilizada pelas camadas superiores. Como exemplo podemos citar os dados de todos os sensores da solução, que são recebidos e convertidos para a linguagem utilizada nas camadas superiores. A segunda camada, gerenciamento de serviço, é responsável por administrar os serviços encontrados em dispositivos, além de disponibilizar recursos dessa solução para a camada superior. Como exemplo, podemos citar o serviço do sistema que não é interrompido para inserção ou exclusão de dispositivos. Por fim, a terceira camada, composição do serviço, realiza a combinação de dados dos produtos IoT, com o objetivo

de criar informações mais complexas. Como exemplo podemos citar a combinação dos serviços de dados disponíveis, que é formulada a partir das necessidades da aplicação, podendo possuir inúmeras combinações possíveis, inclusive fornecendo informações anteriores.

A camada de aplicação, última camada da arquitetura, possui como objetivo utilizar as informações obtidas da camada composição de serviços, como forma de obter resultados para os usuários. As informações são utilizadas como tomada de decisão, enviando ações de acordo com a lógica de negócio da aplicação IoT. Como exemplo podemos citar a comunicação com outros sistemas ou utilização dos serviços utilizados na internet.

## FIQUE POR DENTRO

O Governo Federal do Brasil publicou o plano nacional de internet das coisas. O decreto contendo o plano (n. 8.854 de 2019) possui uma série de objetos para o fomento desse tipo de tecnologia no país relacionado a áreas prioritárias e cria um órgão consultivo por diversos ministérios para avaliar a sua implementação e recomendar ações.

Acesse o link e fique por dentro do plano nacional da internet das coisas: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/decreto/D9854.htm](http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9854.htm). Acesso em: 18 fev. 2020.

## ATIVIDADES

1) O termo IoT (*Internet of Things*) surgiu em 1999, quando o tecnólogo Kevin Ashton o utilizou para mostrar que a utilização dos dados capturados pelos humanos estava relacionada a uma nova revolução da internet, uma vez que o uso de computadores impulsionou a gerar e coletar dados sem a necessidade da intervenção humana.

Assinale a alternativa correta em relação ao conceito de IoT.

- a) Ambiente de objetos virtual com conexão com a internet por meio de diversos *softwares* pequenos e embutidos.
- b) Ambiente de objetos virtuais com conexão com a internet por meio de diversos dispositivos que são considerados embutidos.
- c) Ambiente de objetos nas nuvens sem conexão com a internet por meio de diversos *hardware* pequenos.
- d) Ambiente de objetos físicos com conexão com a internet por meio de diversos tipos de sensores pequenos e embutidos.
- e) Ambiente de objetos nas nuvens sem conexão com a rede local por meio de diversos cabos de redes físicos.

## Os aspectos relevantes da utilização de IoT

Caro(a) aluno(a), nesta seção, vamos estudar os aspectos relevantes encontrados na área de IoT, sendo abordados os fatores-chave para o desenvolvimento de soluções em IoT, relacionados à oportunidade de negócios, e os aspectos voltados à segurança e privacidade e a melhor eficiência da IoT, utilizando os paradigmas *Edge Computing* e *Fog Computing*. A IoT proporciona uma infinidade de aspectos relevantes como oportunidades de negócios para o mercado de trabalho. Essas oportunidades estão classificadas em três categorias estratégicas, cada uma relacionada por um tipo de empresa:

- **Enablers** são empresas que desenvolvem e implementam a sua própria tecnologia voltada para IoT. Por exemplo: HP, IBM e Intel são consideradas empresas orientadas para a tecnologia de internet das coisas.

- **Engagers** são empresas que projetam, criam, integram e fornecem um determinado serviço envolvendo IoT para o cliente.

- **Melhoradores** são empresas que criam os próprios serviços de IoT com valor agregado, além dos serviços fornecidos pelas empresas Engagers.

O uso de produtos IoT pelas empresas é uma forma de aumentar o valor e a competitividade no mercado. Um produto IoT é considerado um sistema comunicante que abrange outros sistemas e pessoas, sendo representado por alguns significados como produto independente, sistema fechado ou ambiente fechado. O produto independente é considerado como algum aparelho independente gerando informação inteligente; como exemplo podemos citar a secadora de roupa conectada. O sistema de IoT é considerado um produto aplicado a algum serviço; como exemplo podemos citar o produto de telemática para logística de transporte. Já o ambiente inteligente é considerado um espaço físico com característica inteligente; como exemplo podemos citar a casa inteligente. Sendo assim, cada termo está relacionado ao produto, serviço ou ambiente com características de IoT.

Dentro do cenário IoT, é perceptível que, ao usar diversos sensores que se comunicam, enviando e recebendo dados de forma frequente, é necessário ter um alto poder de processamento para tratar os dados que estão sendo transmitidos, pois os sensores

somente interpretam o mundo físico e enviam essa informação para alguma outra máquina, podendo ser um roteador, um servidor de dados, dentre outras.

Nesse contexto, surgem dúvidas quanto à segurança dos dados e à forma como é feita essa divisão de processamento. É sabido que enviar muitos dados por uma rede de dados, principalmente para servidores de processamento longe do local de coleta, é relativamente custoso para a organização, pois há mais dados que devem ser trafegados, gerando maiores custos.

Para contornar esse problema, algumas organizações utilizam o conceito de *Edge Computing*, no qual o processamento é realizado em computadores mais próximos do local onde os dados são coletados, evitando o envio destes por muitas redes de dados. Já o *Fog Computing* é um padrão que especifica como o conceito de *Edge Computing* é aplicado na prática.

Nas seções a seguir, você vai aprender um pouco mais sobre as preocupações com segurança ao utilizar IoT, *Edge Computing* e *Fog Computing*.

### **Segurança no desenvolvimento e uso de IoT**

Alguns aspectos relevantes relacionados à segurança e privacidade podem ser abordados para proporcionar benefícios da IoT para indivíduos, sociedades e economias. A seguir, serão abordados os aspectos de segurança e privacidade encontrados na utilização e no desenvolvimento de produtos IoT.

- A garantia de segurança dos produtos e serviços de IoT deve ser considerada uma prioridade para manter a confiança dos usuários dessa tecnologia. Os usuários precisam confiar na segurança proporcionada pelos produtos e serviços de dados oferecidos pela IoT. Diante disso, os desenvolvedores e os usuários de dispositivos de IoT possuem a obrigação de não expor usuários e a própria internet a possíveis danos. Com isso, as ações da indústria, do governo, dos usuários de produtos e serviços de IoT contribuem para o desenvolvimento seguro, a manutenção e a utilização desses dispositivos.
- O processo realizado para coletar, verificar e modificar os dados proporciona valor aos dispositivos e serviços da IoT, porém, esses dados possuem a função de criar perfis com diversas características dos seus usuários. Diante disso, a IoT está relatando as questões de privacidade, por conta da maneira drástica como os dados

são coletados, observados e utilizados. Embora os dados com privacidade sejam falhos, algumas estratégias necessitam ser desenvolvidas para promover a transparência, justiça e escolha do usuário sobre esses dados.

- Alguns mecanismos de segurança da informação devem ser providenciados pelas empresas que provêm as soluções IoT, tendo em vista que os objetos conectados são, normalmente, sistemas embarcados nos quais o usuário não tem condições de instalar dispositivos de segurança, como um antivírus ou *firewall*. Uma das abordagens é o uso de criptografia na comunicação entre os objetos e os *gateways*, que assegura a confidencialidade dos dados por meio da cifragem das mensagens, tornando-as ilegíveis para quem não possui a chave de criptografia. Um exemplo de criptografia que pode ser adotada no IoT é o RC4, também conhecido como ARC4 (Alleged RC4), sendo uma criptografia do tipo simétrica que permite a criptografia de fluxo (*stream*) de mensagens com tamanho variável. Pode-se, ainda, aumentar a segurança com o uso de assinatura digital com algoritmos hash, aumentando a autenticidade dos dados da mensagem enviada.
- Uma outra técnica que pode ser adotada para soluções IoT é o uso de VPN (*Virtual Private Network*) para permitir a criação de um túnel seguro de comunicação entre os dispositivos conectados e os servidores. Algumas soluções IoT baseadas em objetos com simcard para conectividade permitem a conexão direta do simcard através de túneis IPSec. Alguns exemplos em que podem ser empregados as soluções de VPN em IoT são os PDVs (Ponto de Venda ou POS - *Point of Sales*) móveis, em que os sistemas de processamento de pagamentos podem forçar uma conexão criptografada fim a fim com uma VPN IoT e as bicicletas compartilhadas, em que cada bicicleta pode obter um endereço IP privado pela VPN IoT, tornando mais rápida a comunicação para bloqueio e desbloqueio das bicicletas via servidor, comandadas por um aplicativo de usuário.
- Os fabricantes devem ter o cuidado de implementar soluções seguras, tratando as vulnerabilidades conhecidas, bem como permitir correções para erros nos dispositivos de forma eficiente e segura, garantindo uma confiança maior dos clientes nos produtos IoT. Do lado do usuário, é importante que os ambientes de conectividade tenham seus mecanismos de proteção, como é o caso das redes Wi-Fi, em que os roteadores devem sempre ser mantidos atualizados, bem como

configuração de senhas fortes, uso de criptografias para conexão (ex.: WPA2), transformação em rede invisível pela configuração do SSID e que a função *firewall* seja ativada. São ações que aumentam a segurança da rede sem fio que dá cobertura para a conectividade dos objetos.

É importante destacar que a privacidade dos usuários pode ser exposta pelos dados coletados pelos dispositivos inteligentes, presentes em um ambiente IoT. Por isso a ênfase em se observar os aspectos de segurança da informação, recomendando-se testes para mitigação de riscos antes da operacionalização da solução.

### ***Edge Computing***

A *Edge Computing*, também chamada de computação de borda, consiste na inclusão de elementos de processamento parciais instalados próximos aos sensores IoT. Usar uma abordagem *Edge Computing* representa possuir elementos de processamento e armazenamento de dados o mais próximo possível do componente ou aplicação que produz os dados, ou seja, os sensores IoT, pessoas e dispositivos de forma geral.

O principal motivador da abordagem *Edge Computing* é a redução da latência necessária para enviar uma grande quantidade de dados (gerada pelos sensores) dos últimos nós da rede (edge/borda da rede) até um servidor central, geralmente localizado fisicamente distante dos nós de coleta de dados. Esse percurso de enviar dados dos sensores, processar em um servidor central longe da rede e retornar os dados para os sensores ou servidores locais consome muitos recursos de rede, tornando a abordagem *Edge Computing* aplicável quando se tem a produção de muitos dados, como na IoT.

Além disso, a *Edge Computing* não se restringe ao uso de IoT, mas o uso de diversos dispositivos produzindo dados de forma constante na IoT é um representante concreto do uso e vantagens dessa abordagem.

Imagine que em uma casa inteligente temos mais de 100 sensores instalados para coletar diversas informações como: luminosidade, presença de movimento, PH da água, quantidade de água para o consumo, monitoramento dos níveis de oxigênio, presença de tremores na estrutura da casa, dentre diversas outras características de uma casa. Dado que os sensores irão produzir grandes massas de dados, é computacionalmente custoso

enviar esses dados para serem analisados, armazenados e processados em um servidor localizado em outro país ou fisicamente distante de onde é feita a coleta. Claramente, essa abordagem convencional utiliza muitos recursos da rede e também tempo de processamento, sendo que, se usássemos um servidor ou unidades de processamento dos dados mais próximos a esses sensores, reduziria a quantidade e aumentaria qualidade dos dados trafegados pela rede.

Por exemplo, é possível que um conjunto de dispositivos com 5.000 sensores gere até 10GB de dados por segundo! Transferir essa grande quantidade de dados para um servidor distante dos sensores poderá tornar a rede instável e gerar gastos extras para a organização que administra os sensores.

Dado o exposto, a seguir, você vai aprender um pouco mais sobre a *Fog Computing* e como ela se diferencia da *Edge Computing*.

### ***Fog Computing***

*Fog Computing* é um paradigma bem parecido com a *Edge Computing*, já que ambas possuem o intuito de processar os dados o mais próximo possível de onde são coletados, ou seja, as unidades de processamento devem estar próximas da fonte dos dados. No entanto, a *Fog Computing* pode ser considerada uma arquitetura de sistema que distribui a computação, armazenamento, controle e funções de rede próxima aos usuários e sensores.

Antes de nos aprofundar nos conceitos sobre a *Fog Computing*, vamos distingui-la de outro paradigma que pode gerar confusão, a *Cloud Computing*. Um exemplo bastante utilizado para fazer essa distinção é relativo à latência no envio dos dados produzidos pelos sensores para servidores distribuídos, garantindo a qualidade do serviço (QoS). Os nós computacionais na *Fog Computing* são instalados próximos aos sensores IoT, permitindo que a latência de transmissão dos dados seja reduzida quando comparada à abordagem tradicional usada na *Cloud Computing*.

Geralmente, na *Cloud Computing*, temos menos nós de processamento quando comparada à *Fog* ou *Edge Computing*, ou seja, temos uma abordagem mais descentralizada na *Fog* e *Edge Computing*. Os nós de processamento na *Fog Computing* são geograficamente distribuídos.

Outra diferença está na segurança dos dados: na *Fog Computing*, a segurança dos dados é provida nas bordas da rede, ou seja, nos servidores próximos aos sensores, enquanto, na *Cloud Computing*, a segurança é feita nos grandes servidores de processamento e armazenamento dos dados.

A natureza descentralizada da *Fog Computing* permite que os dispositivos presentes nela possam ser usados para pré-processamento e/ou armazenamento dos dados. Em outras palavras, a *Cloud Computing* possui um suporte de *hardware* mais robusto do que a *Fog Computing* e disponibilidade de processamento bem maior do que a *Fog Computing*. A *Fog Computing* fornece mais nós de processamento com menos poder de processamento e menor disponibilidade. São exemplos de nós de processamento na *Fog Computing*: pequenos servidores, roteadores, *switches*, *gateways*, pontos de acesso etc.

Dado que os nós da rede na *Fog Computing* são menores, ou seja, ocupam menos espaço físico, podem ser alocados mais próximos aos usuários quando comparado à *Cloud Computing*. Além disso, não é essencial haver internet para que a rede de *Fog Computing* funcione, pois os sensores podem enviar os dados para os pontos de processamento por uma rede local e, quando disponível a internet, fazer o upload para um servidor na nuvem. Por outro lado, na *Cloud Computing* é sempre necessário haver acesso à internet para enviar os dados para os servidores na nuvem.

Uma diferença crucial entre *Edge* e *Fog Computing* é que a *Edge Computing* possui limitações de processamento nas bordas, enquanto a *Fog Computing* é hierárquica, provendo computação, conectividade, armazenamento, controle e gerencia o envio dos dados locais para a nuvem. Outra característica é que a *Fog Computing* é um pouco mais abrangente, ou seja, atua no âmbito de redes metropolitanas, enquanto a *Edge Computing* se limita às redes locais.

## ATIVIDADES

2) O uso da arquitetura tradicional de *cloud computing* para IoT possui restrição na sua eficiência e desempenho, principalmente quando propõe suporte a serviços da internet das coisas, que necessitam de comunicação frequente e resposta em tempo real entre os dispositivos. Diante disso, surgiu a utilização de outras arquiteturas como *edge computing* e *fog computing* para IoT.

Sendo assim, assinale a alternativa correta em relação às características dessas arquiteturas que melhoraram a relevância da IoT na comunicação de dados.

- a) O aumento do processamento de dados e a diminuição do tempo de comunicação dos dados.
- b) A diminuição do processamento de informação e a diminuição do tempo de transferência dos dados.
- c) O aumento do processamento de dados e o aumento do tempo de comunicação dos dados.
- d) A redução da variação de dados e o aumento do tempo da transferência dos dados.
- e) A redução da permissão de dados e a elevação do tempo de mudança dos dados.

## Os principais protocolos utilizados em IoT

Caro(a) aluno(a), nesta seção, serão abordados os principais protocolos de comunicação utilizados em IoT. Mas o que é um protocolo de comunicação? O protocolo de comunicação é considerado um conjunto de normas responsável por assegurar a efetividade da comunicação de dados entre diferentes máquinas. Sem essas normas, a comunicação seria caótica entre as máquinas.

Os protocolos de comunicação utilizados na internet têm como base o modelo de referência TCP/IP, baseado em camadas, sendo que cada uma apresenta uma função específica. No caso da internet das coisas, não há um modelo de referência específico para a comunicação dos objetos, porém, podemos classificar os protocolos utilizados em três camadas, segundo Sinclair (2018): mídia (MAC/PHY), redes (Network) e aplicativo. Essas três camadas auxiliam na formatação dos dados gerados pelos objetos que serão transmitidos para o *gateway*, para que sejam, posteriormente, disponibilizados e processados por aplicações na rede Internet.

A Figura 2.5 mostra as camadas do protocolo de comunicação para o nível de objetos conectados na internet das coisas.

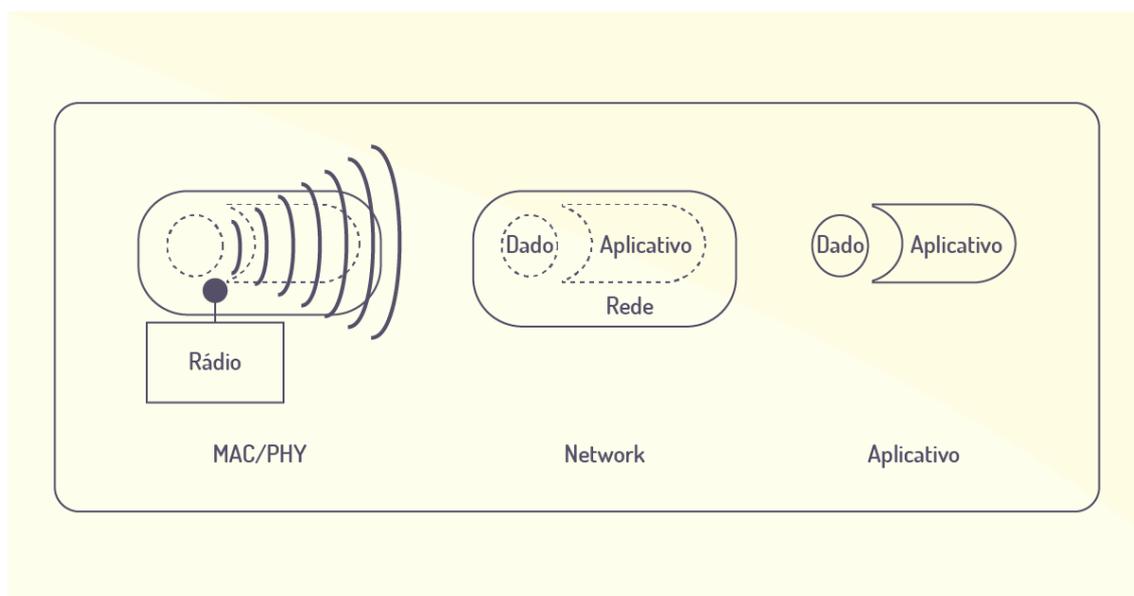


Figura 2.5 - As três camadas do protocolo de comunicação

Fonte: Adaptada de Sinclair (2018).

A camada aplicativo é responsável por prover serviços IoT para os seus clientes por meio da contextualização dos dados utilizados pelos objetos. Por exemplo, uma aplicação solicita medições de temperatura e umidade para os sensores dos objetos conectados de uma solução IoT; os valores dessas medições devem estar contextualizados (como a unidade de medida utilizada, a data e o horário da medição realizada) dentro dos pacotes de dados que serão enviados, para que o devido processamento possa ser dado dentro do contexto estabelecido pelo aplicativo.

Os principais protocolos de aplicativo são o MQTT (*Message Queue Telemetry Transport*, para arquiteturas de *publish-subscribe*, em que clientes “assinam” um tópico e recebem qualquer mensagem encaminhada para este) e o COAP (*Constrained Application Protocol*, utilizado em arquiteturas cliente-servidor, similar ao protocolo HTTP).

A camada de rede, ou *network layer*, é formada pelo empacotamento dos dados provenientes da camada de aplicativo para transporte pela rede conectada de objetos IoT, visando atingir um *gateway* para possibilitar a conexão com a rede internet. Por exemplo, as abstrações das tecnologias de comunicação, serviços de gerenciamento, roteamento e identificação devem ser realizados nessa camada, sendo que, quando o pacote de dados chega no *gateway*, este providencia a readequação dos dados dentro dos protocolos padronizados de internet, para que possam atingir servidores em ambientes de nuvem ou *on premises*.

A camada MAC (*Media Access Control*) / PHY (*PHYSical layer of model*), também conhecida como camada de mídia, está voltada para a estrutura *hardware* utilizada na comunicação dos objetos, por exemplo, as estruturas físicas utilizam sensores para coletar e processar informações. Geralmente, essa camada em IoT é representada pela interface de radiofrequência, que transmite os dados das camadas de aplicativo e rede pelo ar. Os principais rádios empregados em IoT empregam o padrão 802.15.4, usando os protocolos ZigBee, Zwave e 6LowPAN. Outras alternativas são o padrão Wi-Fi (IEEE 802.11), padrão *Bluetooth* (IEEE 802.15.1), tecnologias de redes móveis (4G, CAT M1, NB-IoT e 5G) e rádios de baixo consumo de energia com longo alcance LPWA (Lora, LoraWAN e Sigfox).

Os protocolos mostrados nessas três camadas utilizam as topologias de rede sem fio estrela ou em malha (*mesh*). Na topologia de rede estrela, todos os dispositivos se conectam a um ponto central, conforme apresentado na Figura 2.6, a seguir.

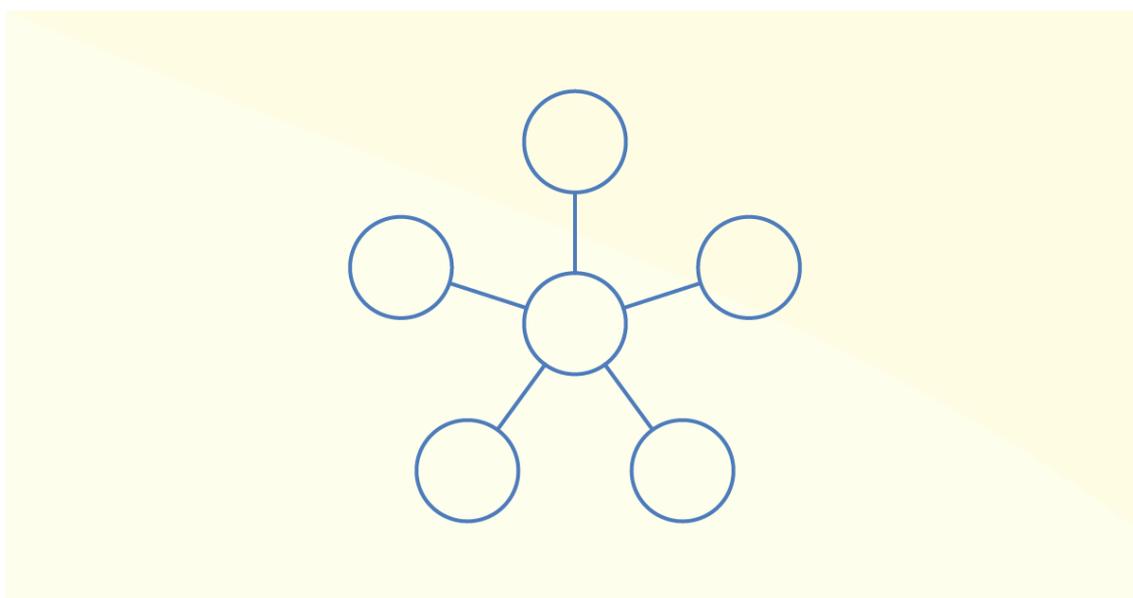


Figura 2.6 - Topologia em estrela

Fonte: Adaptada de Sinclair (2018).

Já na topologia de rede em malha, a conexão pode ser diretamente de um nó a qualquer outro da rede, conforme visualizado na Figura 2.7.

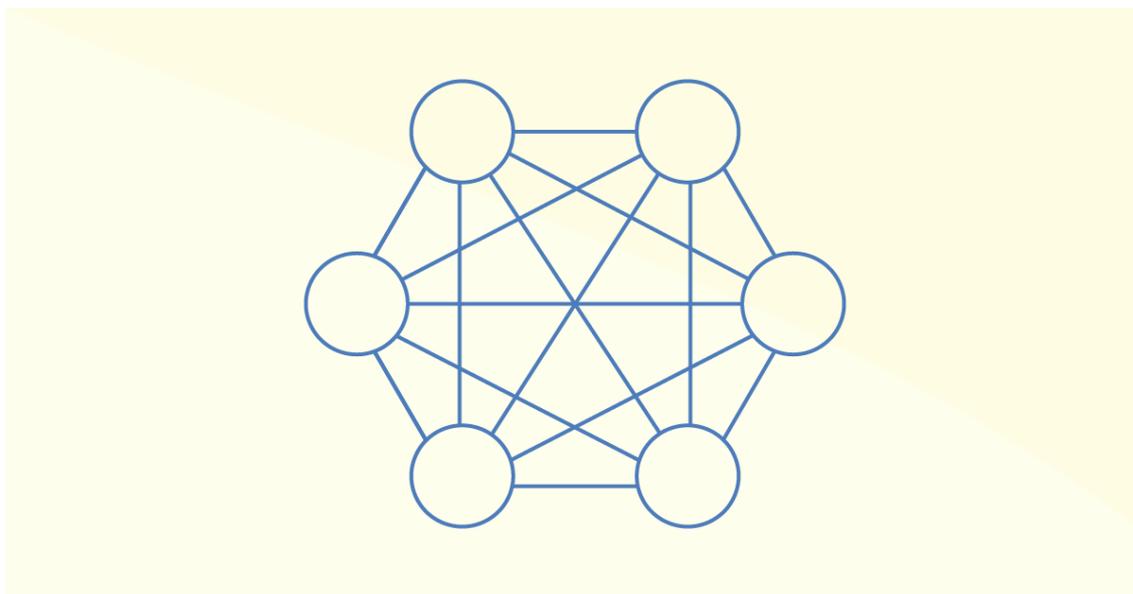


Figura 2.7 - Topologia em malha (*mesh*)

Fonte: Adaptada de Sinclair (2018).

As topologias de estrela são utilizadas em protocolos LoRaWAN e NB-IoT. Já a topologia em malha é utilizada em protocolos Zigbee e Bluetooth.

As aplicações que empregam o uso dos protocolos IOT são implementadas em sistemas embarcados, por exemplo, sensores pessoais, sensores industriais, medidores de energia elétrica inteligentes, computação vestível (*wearables*) e elementos de infraestrutura de redes. Esses protocolos permitem, essencialmente, a comunicação entre os dispositivos conectados com os *gateways* que farão a intermediação com as aplicações presentes em servidores na internet. A maioria desses protocolos apresenta as vantagens de abranger grandes distâncias através de sistemas RF (radiofrequência), para dar uma maior mobilidade para os objetos conectados, além de possibilitarem um menor consumo de energia. Como estes são, normalmente, baseados em baterias internas, possibilitam uma longevidade maior de uso, até que seja necessária a recarga. Nas seções a seguir, serão detalhados os principais protocolos utilizados em IoT.

## Protocolo Zigbee

O protocolo Zigbee foi fundado em 2002, pela aliança Zigbee, sendo considerado um protocolo de comunicação utilizado em ambientes industriais e definido pelo IEEE 802.15.4, com as especificações de comunicação da camada física e da camada de acesso ao meio das redes de comunicação sem fio que operam com baixa taxa de transmissão de dados.

O protocolo Zigbee possui uma pilha de protocolos de implementação simplificada, sendo estruturado por três características: baixo custo, baixo consumo de energia e reduzida taxa de transferência de dados de 250 kbps, com o alcance máximo de 100 metros. Por causa disso, o protocolo possui aplicação em monitoramento e controle industrial, automação residencial e de sistemas de energia. Essa pilha da plataforma ZigBee é definida em camadas, sendo que o padrão IEEE 802.15.4 define as camadas física (PHY) e controle de acesso ao meio (MAC). As camadas de rede (NWK) e suporte à aplicação (APS) são definidas pela *ZigBee Alliance*, deixando para o usuário a definição da camada de aplicação. A Figura 2.8 mostra a descrição da arquitetura do protocolo Zigbee.

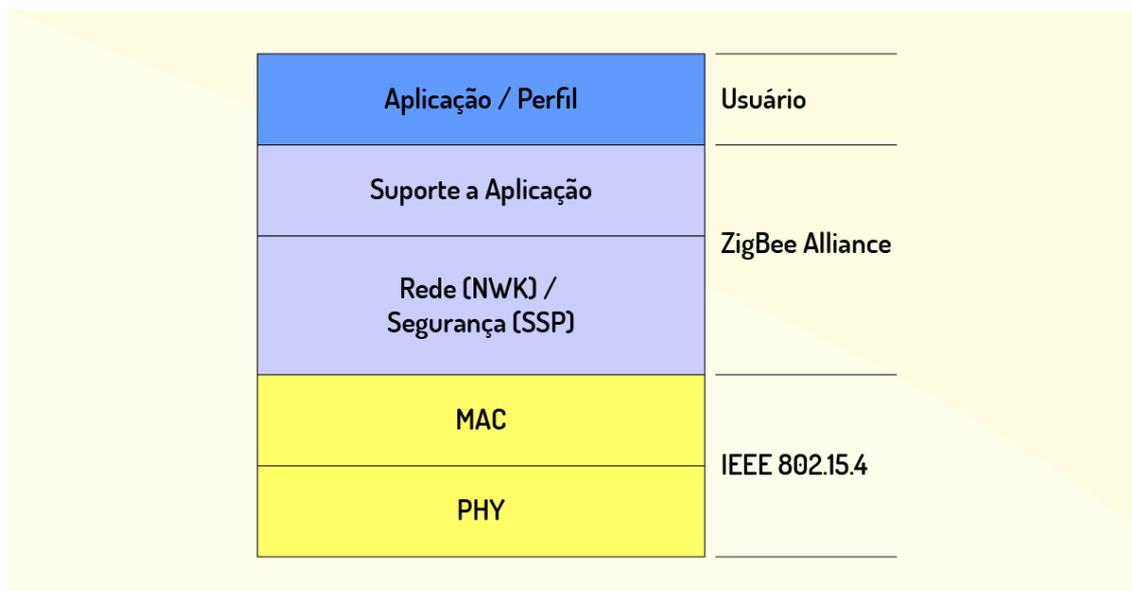


Figura 2.8 - Arquitetura do protocolo Zigbee

Fonte: Kalebe / Wikimedia.

As aplicações que utilizam o protocolo Zigbee são voltadas para a indústria, mas não limitadas a esse cenário. Por exemplo, se as entidades comunicantes transmitem poucas informações, mas precisam que estas sejam transmitidas por uma distância maior e com segurança, é aconselhável o uso do Zigbee. Um exemplo prático é o acompanhamento de características de máquinas em uma linha de produção (imagine uma montadora de carros), observando os sensores de temperatura, pressão, soldagem etc.

Em complemento, na automação industrial, o protocolo Zigbee realiza a redução da latência e o consumo de energia, além de reduzir significativamente o custo de comunicação e o processo geral de controle. Por fim, por meio da aplicação de monitoramento sem fio com o protocolo Zigbee, é possível detectar falhas com precisão. Além disso, é possível utilizar uma malha *mesh* para aumentar o seu alcance em extensão doméstica ou industrial, sem necessidade de utilizar ligações entre elas.

### **Protocolo 6LowPan**

O protocolo 6LowPan é um acrônimo para *IPv6 over Low power Wireless Personal Area Network*, sendo considerado uma adaptação do protocolo IPv6 com o protocolo 802.15.4. O 6LoWPAN é considerado um protocolo de rede que possui endereçamento e mecanismos de compressão dos pacotes IPv6 em protocolo 802.15.4, permitindo que sejam enviados e recebidos nessas redes. O endereçamento do protocolo 802.15.4 utiliza 16 bits de endereços para cada dispositivo na rede, obtendo uma identificação IPv6 única. A compressão do 6LoWPAN utiliza informações de protocolos presentes em outras camadas. Por exemplo, o 6LoWPAN pode utilizar parte do endereço MAC do dispositivo para atribuir um respectivo endereço IPv6. A Figura 2.9 mostra a descrição da arquitetura do protocolo 6LowPan.

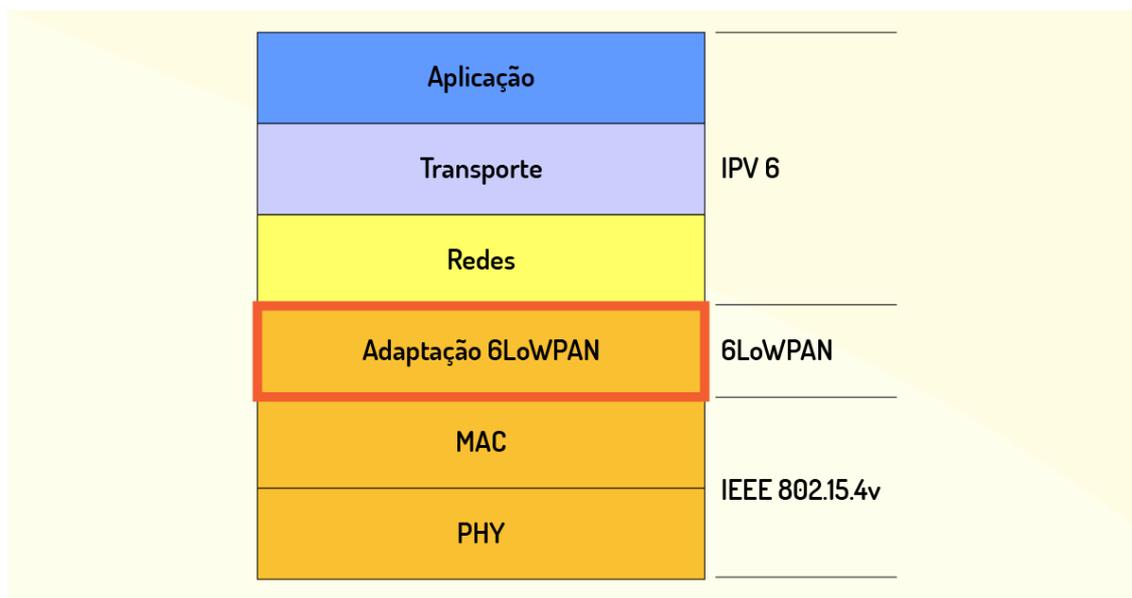


Figura 2.9 - Camada do protocolo 6LoWPAN

Fonte: Elaborada pelos autores.

O protocolo 6LowPAN possui uma camada de adaptação entre o IEEE 802.15.4 e o IPv6, com a presença de cabeçalhos específicos que podem ser removidos ou adicionados dependendo da necessidade, assim como permite o envio da informação útil. Essa camada possui adaptação do protocolo 6LowPan utilizando dois tipos de endereços: o endereço de camada de ligação lógica, com sua utilização em todos os dispositivos e com endereços de 16 bits e de 64 bits, e o endereço da camada de rede, com a ligação entre a rede IPv6 e 6LoWPAN.

### **Protocolo *Bluetooth***

O protocolo *Bluetooth* surgiu em 1998, quando a Ericsson formou, junto com as empresas IBM, Intel, Nokia e Toshiba, o Bluetooth SIG (*Special Interest Group*), com o objetivo de criar um padrão universal e aberto para a comunicação sem fio entre dispositivos.

Mas o que é *Bluetooth*? *Bluetooth* é considerado uma tecnologia de comunicação sem fio que permite que computadores e *smartphones* realizem trocas de dados entre si, com conexão com diversos dispositivos, como: mouses, teclados, fones de ouvido, impressoras, caixas de som e outros acessórios por meio das ondas de rádio.

Diante do conceito de *Bluetooth*, como podemos conectar diferentes tipos de aparelhos, incluindo sensores e vestíveis ao conceito de internet das coisas? O *Bluetooth Smart* (ou *Bluetooth Low Energy – BLE*) é o conceito utilizado em aplicações com agentes de transferência de dados. O BLE é concedido pela *Bluetooth Special Interest Group*, sendo considerado uma rede sem fio. Esse *Bluetooth Smart* é responsável por fornecer o consumo de energia e custos reduzidos em diversos conceitos de IoT. A Figura 2.10 mostra a comunicação de diversos dispositivos por *Bluetooth*.



Figura 2.10 - Comunicação de diversos dispositivos por *Bluetooth*

Fonte: Souza (2016, p. 12).

As primeiras versões do *Bluetooth*, conhecidas como *Bluetooth Classic*, não possuíam integração aos produtos IoT, por possuir uma distância de alcance baixo e curta limitações de gasto de energia, tornando-se impraticável na utilização de IoT. Como forma de atender à crescente demanda por conectividade sem fio em dispositivos com IoT, surgiu a versão 4.0 *Bluetooth* com a categoria *Bluetooth Low Energy* (BLE). Essa versão BLE é utilizada em aplicações de baixo consumo de energia, com uma taxa de transferência de dados menor, com 1 Mbps. Por fim, como forma de expandir em novas aplicações em IoT, surgiu a versão 5 do *Bluetooth*, com o dobro de velocidade de transmissão de dados e com alcance de comunicação quatro vezes maior.

## Protocolo LoRa

A tecnologia LoRa (*Long Range*), desenvolvida pela Semtech Corporation, possui operação nas faixas de frequências de tecnologia sem fio chamada sub-GHz (Sub-Gigahertz, frequências de 2,4 GHz, 868 MHz a 928 MHz, 433 MHz e 315 MHz), com transmissão de dados para servidores distantes, permitindo a conectividade com a internet por meio de *gateways*. Essa tecnologia permite um alcance de transmissão de vários quilômetros com baixo consumo de energia. O padrão LoRaWAN (*Long Range Wide Area Network*) é considerado um protocolo que define a arquitetura do sistema, assim como os parâmetros de comunicação, usando a tecnologia LoRa.

O protocolo LoRaWAN possui uma arquitetura interconectadas com acesso em vários dispositivos que possam se comunicar com um *gateway* que usa a modulação LoRa. Essa arquitetura possui topologias do tipo estrela, com a conexão de três tipos dispositivos, os nós conectados por um link de único salto a um ou mais *gateways*, quem, por sua vez, possuem conexão por meio da rede IP. A Figura 2.11 mostra o diagrama de interconexão dos elementos da tecnologia LoRa.

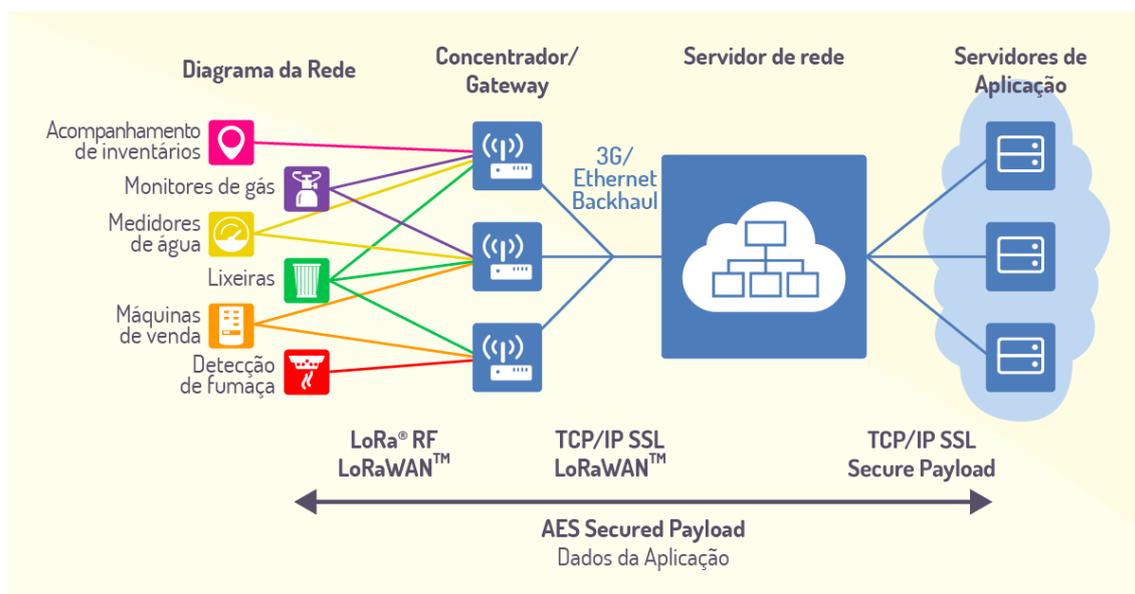


Fig. 2.11 - Diagrama de interconexão dos elementos da tecnologia LoRa

Fonte: Adaptado de Ram (2018).

O diagrama de interconexão dos elementos da tecnologia LoRa possui quatro módulos. O primeiro módulo do diagrama de rede considera os dispositivos conectados na rede através dos módulos LoRa, sendo que alguns exemplos são as máquinas de venda, lixeiras e equipamentos de detecção de fumaça. O segundo módulo concentrador/*gateway* é representado pelos concentradores responsáveis por receber os sinais enviados pelos módulos LoRa dos dispositivos conectados na rede e enviar para a internet. O terceiro módulo, o servidor de rede, é responsável pela disponibilização dos dados provenientes da rede LoRa através dos concentradores/*gateways* para que possam ser utilizados por aplicações na internet. Por fim, o quarto módulo, representado pelo servidor de aplicações, é responsável por prover acesso aos dados para o cliente final, na forma relacionada a alguma aplicação útil.

### **Protocolo NB-IOT**

O padrão NB-IoT (*Narrowband Internet of Things*) é considerado uma tecnologia de rádio LPWAN (*Low Wide Area Network*), desenvolvido pela 3GPP (*3rd Generation Partnership Project*), que serve para conectar diversas variedades de serviços com dispositivos celulares. Esse protocolo concentra-se, especificamente, na cobertura interna, baixo custo, longa duração da bateria e alta densidade de conexão. Em termos de radiofrequência na camada de mídia, utiliza um espectro de frequências licenciado, com a modulação OFDMA (*Orthogonal Frequency Division Multiple Access*), e uma taxa de dados de 20 kbps, sendo que a antena apresenta uma cobertura de, aproximadamente, 20 km. Diante disso, o padrão NB-IoT torna-se importante para funções que realizam monitoramento do sensor de rastreamento com difícil alcance, possibilitando a conexão de diversos serviços em IoT.

A utilização do protocolo NB-IoT surgiu da necessidade encontrada na internet das coisas com a conexão de grande quantidade de equipamentos, seja por questão de duração de bateria, custo e eficiência do produto. Por isso, o protocolo permite conectar bicicletas, patinetes e equipamentos industriais com custo reduzido de consumo e possibilita a utilização dos dispositivos sem a necessidade de troca de bateria por mais de 10 anos.

Além disso, esse equipamento, utilizando o protocolo NB-IoT, mesmo operando com banda estreita, permite a comunicação com longas distâncias.

A aplicação NB-IoT utiliza mediador inteligente, um dispositivo eletrônico responsável por registrar o consumo de energia elétrica e comunica as informações ao fornecedor de eletricidade por meio de monitoramento e faturamento. Esse mediador inteligente possui vários módulos conectados da residência usando eletricidade, gás e água, com conexão com um módulo NB-IoT, para, assim, esse módulo conectar um respectivo data center. A Figura 2.12 mostra uma aplicação com NB-IoT.

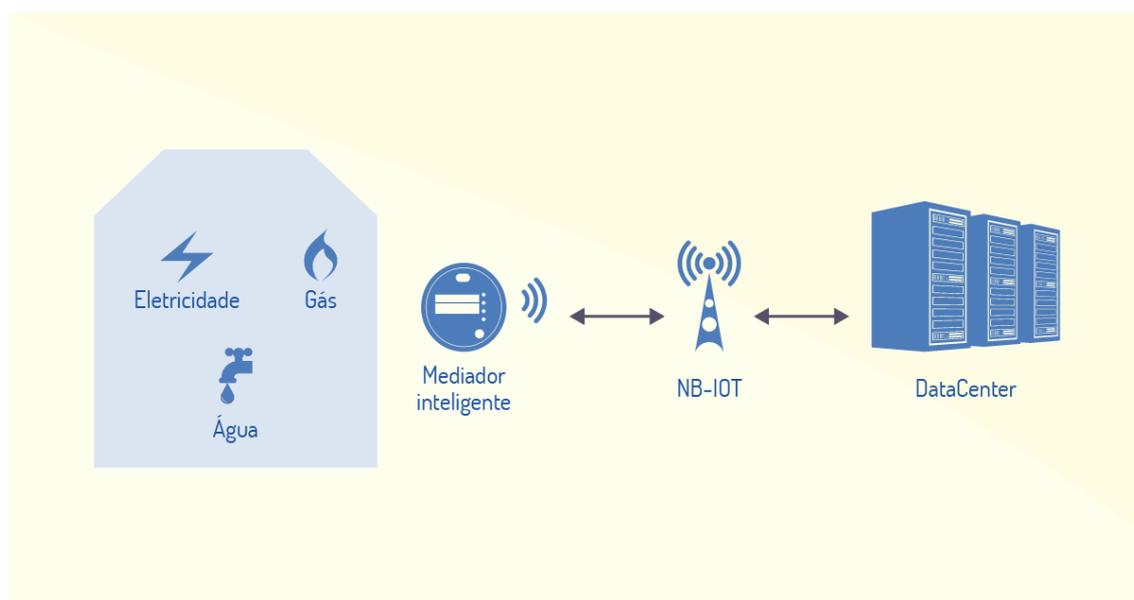


Figura 2.12 - Aplicação com NB-IoT

Fonte: Elaborada pelos autores.

O protocolo NB-IoT é considerado uma IoT Massiva. Mas o que é isso?

Bem, a IoT massiva possui como principal alvo grandes volumes de dispositivos de baixa complexidade, que raramente enviam ou recebem mensagens. O tráfego da IoT massiva é frequentemente utilizado em aplicações do protocolo NB-IoT utilizando medidores inteligentes de residência e tecnologias vestíveis, como aplicações em pulseiras e relógios. Além disso, pode ser utilizado em fábrica inteligente como forma de rastrear os seus respectivos estoques.

## **FIQUE POR DENTRO**

O desenvolvimento do projeto de internet das coisas necessita de três elementos: o objetivo específico, os dispositivos inteligentes e o protocolo de comunicação.

A definição do objetivo do projeto é o primeiro passo para definir o protocolo utilizado e os dispositivos mais adequados em relação à distância de comunicação, à taxa de dados e ao consumo de energia. Esses protocolos de comunicação permitem a conexão entre diversos dispositivos, assim como o acesso em praticamente qualquer lugar do mundo.

Acesse o *link* e fique por dentro dos protocolos mais utilizados para IoT. Disponível em: <https://www.datamon.com.br/Noticia/quais-os-protocolos-mais-utilizados-para-iot>.

Acesso em: 18 fev. 2020.

## ATIVIDADE

3) Os protocolos de comunicação são considerados a base de uma rede de comunicação e padronização na internet das coisas. Esses protocolos na internet das coisas são organizados por três camadas, utilizadas pelo modelo TCP/IP.

Assinale a alternativa que apresenta corretamente as três camadas do protocolo de comunicação utilizado em IoT.

- a) Camada de mídia, camada da rede e camada de aplicativo.
- b) Camada de mídia, camada da rede e camada de enlace.
- c) Camada de mídia, camada de transporte e camada de aplicativo.
- d) Camada de mídia, camada física e camada de aplicativo.
- e) Camada de sessão, camada da rede e camada de aplicativo.

## Os Principais usos encontrados em IoT

Caro(a) aluno(a), nesta seção, serão abordados alguns exemplos práticos utilizados por empresas que estão se apropriando dos conhecimentos da área de IoT para realizar melhorias e inovações em seus processos.

As aplicações IoT devem possuir dispositivos inteligentes e conectados como maneira de utilizar cada informação que compartilha na sua vida cotidiana e, conseqüentemente, as aplicações que utilizam dados para interagir com você diariamente. Vamos descrever a seguir, algumas utilizações de IoT.

- A IoT aplicada a casas inteligentes (*smarthomes*): Imagine que, antes de chegar à casa, você pudesse ligar o ar-condicionado ou as luzes da casa. Há diversos estudos e protótipos de casas inteligentes, em que sensores monitoram diversas características da casa, como: luminosidade, presença de pessoas, pressão da água, qualidade de água etc. Imagine que a compra de comida em sua residência foi automatizada, ou seja, se a sua geladeira possuísse sensores para identificar quais alimentos você possui dentro dela; a partir desses sensores, seria possível que a sua geladeira fizesse o pedido dos itens que estão faltando em um supermercado que, de forma automática, enviaria os produtos para sua casa. É possível identificar os alimentos presentes a partir da cor, formato, posição dentro da geladeira, dentre outras características.
- A IoT aplicada à iluminação pública: estudos apontam que os gastos de energia com iluminação pública em uma cidade estão entre 30% e 50% do gasto total. A IoT pode auxiliar na redução desses gastos, por exemplo: é possível detectar se há movimentação e ativar ou desativar as luzes; perceber se há alguma lâmpada queimada e avisar à equipe de manutenção; monitoramento do consumo por poste ou quarteirão, dentre outras especificidades. Esse controle aprimorado permite a inserção de técnicas que reduzem o consumo de energia, permitindo que as cidades economizem recursos financeiros e energéticos.
- A IoT aplicada a carros conectados: imagine você possuindo um carro conectado para otimizar as execuções internas? Com IoT isso é possível, por meio de carros conectados utilizando sensores de bordo e conectividade com a internet. As evoluções de carros conectados estão sendo trabalhadas por grandes empresas

como Tesla, BMW, Apple e Google.

- A IoT aplicada a cidade inteligente: imagine você andando por uma cidade com vigilância inteligente, transporte automatizado, sistemas de estacionamento inteligente e monitoramento ambiental? Com IoT isso é possível, por meio da instalação de sensores nas coisas, possibilitando conexão em tempo real com diversos objetos.
- A IoT aplicada à agricultura: imagine você monitorando o clima da região e o solo da plantação em tempo real? Com IoT é possível, com a utilização de sensores para detectar umidade e nutriente do solo, controlar o uso da água para o crescimento da planta e determinar fertilizantes personalizados.
- A IoT aplicada em sistema de energia: imagine você controlando o consumo da sua energia diariamente? Com IoT é possível, pois, com a utilização de Smart Grids, pode-se coletar dados inteligentes para automatizar e analisar o comportamento dos consumidores, a fim de melhorar a eficiência, assim como o uso da eletricidade.
- A IoT aplicada na saúde: imagine você analisar a sua saúde em tempo real e receber a estratégia para combater a doença? Com IoT é possível, com a coleta personalizada de dados dos usuários e o fornecimento de estratégias personalizadas para combater as doenças. A IoT pode ser aplicada na saúde para detectar e prever anomalias no corpo humano – por exemplo, há sensores que ficam acoplados na pele para identificar a quantidade de açúcar no sangue e alertar caso o nível de glicemia fique muito alto ou muito baixo. Agora, imagine diversos sensores acoplados no seu corpo de forma a monitorar diversas características e se comunicando com alguns atuadores que estão presentes em sua casa – por exemplo, se o nível de estresse estiver alto, é possível colocar música ambiente e ligar a TV em um programa humorístico.

As empresas possuem diversos cases de IoT com equipamento do tipo *Gadgets*. Mas o que são *Gadgets*? São dispositivos eletrônicos portáteis, como celulares, cafeteiras e leitores de MP3, criados para melhorar funções específicas da vida cotidiano, com a produção de produtos inteligentes e avançados. A seguir, vamos mostrar alguns cases de *gadgets* com recursos de IoT e um caso de solução para serviços de transporte.

### Case 1: Cafeteira Behmor (SANTOS, 2018,)

O dispositivo da Amazon, Dash Wand, possui funcionalidade de IoT com controle de voz da Alexa, por meio da solicitação entre resposta e pergunta do usuário, assim como realiza a conexão com controle de dispositivos domésticos e realiza solicitação de compra com os usuários.

A cafeteira Behmor, dispositivo da Amazon mostrado na Figura 2.13, realiza conexão da cafeteira via Wi-Fi, utilizando aplicativo IOS ou Android. As principais funcionalidades de IoT dessa cafeteira são: controle preciso da seleção do grão do café e o fluxo de aquecimento da água; por fim, a cafeteira informa ao usuário a finalização desse café.



Figura 2.13 - Dispositivo com cafeteira Behmor

Fonte: Santos (2018, p. 166).

### Case 2: Travesseiro Zeek Smart Pillow (SANTOS, 2018)

O Zeek Smart Pillow mostrado na Figura 2.14 é um travesseiro com memória que detecta automaticamente o ronco do usuário e, em seguida, vibra suavemente para mudar sua posição de dormir, sem que a pessoa acorde.



Figura 2.14 - Dispositivo Zeek Smart Pillow

Fonte: Santos (2018, p. 163).

Além disso, esse travesseiro pode rastrear o movimento do sono e o ronco em decibéis para fornecer um relatório diário sobre a sua qualidade de sono.

### Case 3: Trens inteligentes com IoT

O uso de trens para o transporte de passageiros é comum em países dentro da Europa e da Ásia. Alguns trechos chegam a durar até 24 horas, tornando a viagem bastante cansativa para os usuários. Para contornar esse problema, há projetos que utilizam IoT para melhorar a experiência dos usuários dentro dos vagões – por exemplo, é possível monitorar os corredores para verificar se algum passageiro está fora do seu local de dormir, podendo inferir perigo em casos de mal funcionamento do trem; é possível instalar câmeras atrás dos espelhos do banheiro e, a partir das imagens, avaliar o humor do usuário e propor melhorias como: uma refeição ou bebida, alguma medicação não controlada, dentre outras; também é possível usar sensores e câmeras instaladas na frente do trem para monitorar o percurso e aplicar medidas de segurança, caso haja obstáculos à frente. Nesse último caso, as imagens e dados dos sensores são enviados para um *Gateway* localizado dentro do vagão e, então, o processamento dos dados é realizado em

tempo real, podendo acionar atuadores de segurança no trem como: freio, avisos de perigo aos usuários, dentre outras ações.

Os conceitos descritos anteriormente também podem ser expandidos para os ônibus municipais e interestaduais do Brasil, permitindo que eles se comuniquem com servidores durante o trajeto e também com outros veículos, evitando acidentes e trocando dados que podem ser usados para a melhoria contínua do sistema de transporte.

A Nokia e a Altran lançaram uma solução em conjunto para agilizar e otimizar a manutenção de trens por meio de IoT. O projeto tem como base a incorporação de sensores inteligentes para monitorar e coletar dados sobre a condição e o desempenho do material circulante, a rede sem fio e as conexões de IoT para agregar e transmitir dados. É uma plataforma de IoT escalável e adaptável para gerenciar a grande variedade de sensores e outros dispositivos necessários, além de um mecanismo avançado de análise para interpretar dados e recomendar medidas de manutenção preventiva. Esses recursos podem fornecer uma melhoria de até 30% na confiabilidade do trem, uma redução de 20% nos atrasos e cancelamentos, e uma redução de 10% nas horas-homem de manutenção (IOT..., 2018).

A IBM também propõe soluções de IoT para trens inteligentes, a fim de trazer maior confiabilidade e segurança no transporte, bem como uma experiência moderna e otimizada para os passageiros. Foi implementado na **Taiwan High Speed Corporation**, uma empresa de sistema ferroviário totalmente elétrico, um sistema totalmente avançado, baseado em IoT, capaz de acionar automaticamente atividades de manutenção, detectando possíveis problemas na rede por meio da automatização de alarmes e resolvendo-os antes que se tornem um risco para a segurança dos passageiros. Os sistemas de monitoramento e telemetria rastreiam e relatam a condição de mais de 320.000 ativos, como trens e trilhos. Como resultado, 99,15% das chegadas e partidas de trens ficam dentro de seis segundos após o horário previsto, obtendo um crescimento rápido no número de passageiros (BELLIAS, 2016).

#### Case 4: IoT na manufatura avançada de uma usina de cana

Em uma usina de cana, presente entre os maiores grupos sucroenergéticos do Brasil e com capacidade aproximada de moagem de mais de 20 milhões de toneladas de cana, foi

implementada uma melhoria de controle e gestão da sua produção por meio de informatização, que até então eram feitos de forma manual (VALENTE; COLENCI NETO, 2017). Como solução, foi utilizada uma arquitetura IoT que atenderia às regras de negócio por meio de uma aplicação, tendo a tecnologia de *tags* UHF RFID para o rastreamento da movimentação das mudas de cana e sensores para medição de temperatura, umidade e radiação nos cinco viveiros existentes. A visualização da rastreabilidade de cada muda desde o seu nascimento até sua expedição para o plantio, bem como as temperaturas submetidas foram apoiadas pela captura dos dados de movimentação de cada muda produzida, juntamente com os dados dos sensores capturados em tempo real. Entre os benefícios obtidos destacam-se: a redução de tempo na coleta dos dados, que passou a ser automática e em tempo real; diminuição no tempo de localização e geração de informação; controle visual das posições dos viveiros com taxa de ocupação em tempo real; rapidez e acurácia na análise dos dados gerados pelos sensores; relatório com dados dos lotes mais antigos para expedição (FIFO); institucionalização de um processo de previsão de lotes a serem expedidos por período.

## ATIVIDADE

4) As aplicações IoT devem possuir dispositivos inteligentes e conectados como maneira de utilizar cada informação compartilhada na sua vida cotidiana. Existem diversas aplicações de IoT no mundo real, como em casas inteligentes, carros conectados e cidades inteligentes.

Sendo assim, assinale a alternativa correta em relação à aplicação IoT que torna os espaços da residência tão comuns, como a utilização dos *smartphones*.

- a) *Smart City*.
- b) *Smart TV*.
- c) *Smart Switch*.
- d) *Smart Bracelet*.
- e) *Smart Home*.

## **INDICAÇÕES DE LEITURA**

Nome do livro: IoT: como usar a "internet das coisas" para alavancar seus negócios

Editora: Autêntica Business

Autor: Bruce Sinclair

ISBN: 9788551303566

Comentário: Este livro mostra uma visão detalhada da IoT em relação à maneira como consumidores e empresas adquirem e usam os produtos. Muito além da tecnologia, a IoT é considerada uma questão estratégica e, conseqüentemente, assunto a ser pautado na agenda de qualquer dirigente ou gestor que não queira ser surpreendido pela concorrência ou por novos modelos de negócios disruptivos. Além do conhecimento sobre IoT, esse livro proporciona a segurança necessária para aprofundar a discussão sobre IoT na sua organização e como usá-la para alavancar seus negócios.

## **INDICAÇÕES DE FILME**

Nome do filme: Minority Report

Gênero: Ficção Científica

Ano: 2002

Elenco principal: Tom Cruise, Colin Farrell, Samantha Morton e Max von Sydow

Comentário: O filme foi ambientado em 2054, em que a polícia PreCrime de Washington segura os assassinos antes de agir, reduzindo, assim, a taxa de assassinatos a zero. Com isso, os assassinatos são identificados usando três seres mutantes, chamados "Precogs", que realizaram a pré-visualização de crimes recebendo visão do futuro. No filme, surgem cenas de internet das coisas quando o personagem principal, John Anderton, caminha na estação do metrô em Washington e diversos outdoors interativos com sensores detectam sua presença, além de chamarem por seu nome e sugerirem bebidas e viagens com base no seu perfil de consumo. Além disso, os carros do filme são guiados sem a intervenção humana.

UNIDADE III

# Inteligência Artificial (IA)

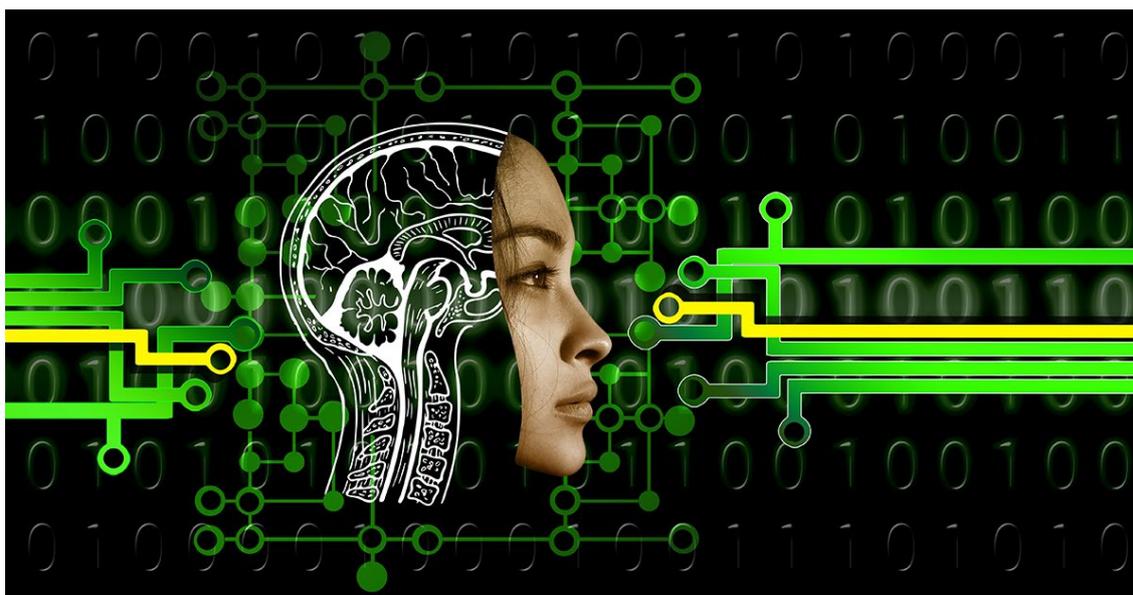
*Aline Chagas Rodrigues Marques*

*Marcelo Takashi Uemura*

## Introdução

Caro(a) aluno(a), imagine a seguinte situação: você acorda pela manhã com a sua assistente pessoal lhe despertando e desejando “bom dia” ao som da música que você quer escutar. Ela também já analisa as melhores notícias para que você esteja atualizado sobre o que está ocorrendo em sua região e pelo mundo. Quando você está escovando os dentes, ela analisa sua caixa de e-mails e já o notifica sobre quais e-mails importantes você precisa ler, quais reuniões precisa participar e lembra, também, do aniversário de sua mãe. Além disso, a assistente já envia como está o fluxo de trânsito e em qual horário você chegará ao trabalho. Ela também já preparou seu café da manhã e já deixou o carro na melhor direção para o momento de sua partida. Por fim, se despede e lhe deseja um bom retorno para casa.

É esse o contexto do nosso assunto: a Inteligência Artificial (IA). Com essa abordagem, percebemos que o assunto pode ser aplicado no nosso dia a dia de forma bem simplista. É possível perceber, também, que a análise humana ficou a cargo totalmente da assistente pessoal, e é isso que um dos ramos da inteligência artificial pretende: tarefas que antes eram atribuições somente humanas passam a ser realizadas por máquinas. Não é fantástico? Mas, para que tudo isso ocorra, é necessária uma série de recursos que possibilitam o crescimento e a aplicação da inteligência artificial, como: algoritmos, redes neurais, conceitos atrelados à inteligência humana, entre outros. Veremos alguns desses recursos em nossos próximos tópicos. Preparado(a)? Vamos começar!



Fonte: geralt / Pixabay.

## Definição

É comum que muitos se assustem com a ideia de termos uma máquina com características do ser humano, como o ato de pensar e se movimentar. Porém, essa é uma realidade cada vez mais presente no cotidiano das pessoas, sobre a qual muitos estudos estão sendo realizados em todas as partes do mundo, envolvendo diferentes disciplinas do conhecimento humano e a computação, no campo denominado Inteligência Artificial.

A evolução da IA, atrelada ao poder computacional, começou nos anos 1930, quando Alan Turing (Figura 3.1), que trabalhou no escritório de decodificação de mensagens da Inteligência Britânica, forneceu uma visão matemática à ciência da computação e, a partir disso, a computação pôde se expandir, não se restringindo à visão limitada de uso para operações de cálculos matemáticos (MEDEIROS, 2018).

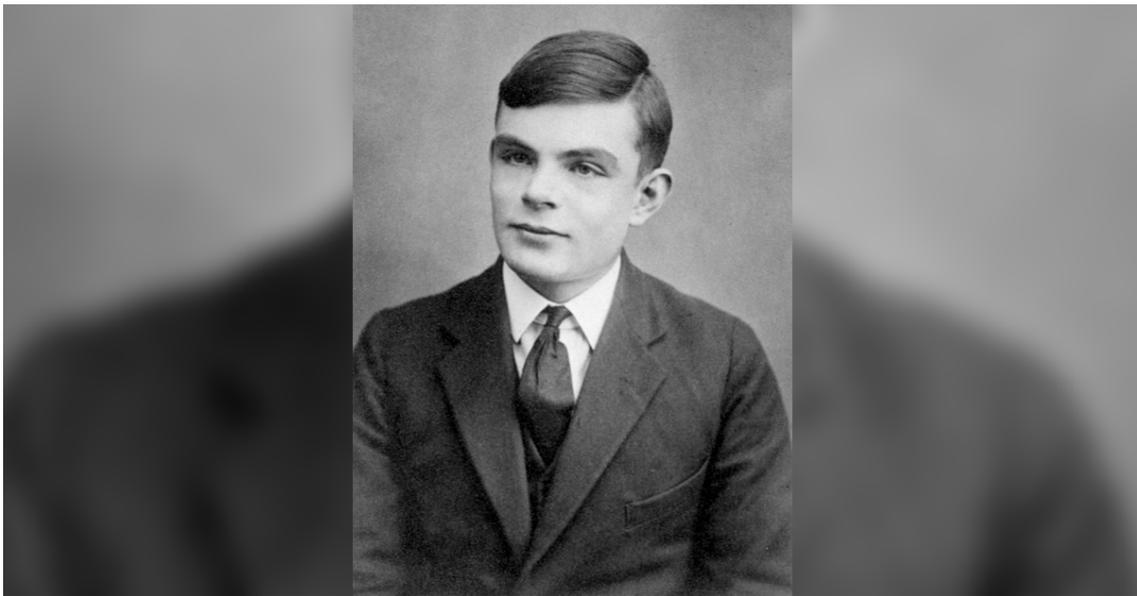


Figura 3.1 - Alan Turing, considerado o pai da inteligência artificial

Fonte: WikiPedant / Wikimedia.

Tudo começou com o experimento de Turing, conhecido como a máquina de Turing: era a Segunda Guerra Mundial, e os alemães, que estavam em vantagem computacional, possuíam uma máquina chamada Enigma, a qual escrevia mensagens alemãs com rotores,

que mudavam seus códigos diariamente. Isso servia como forma de embaralhar as mensagens trocadas pelas tropas alemãs de Hitler. Caso alguém desvendasse esse segredo, era capaz de ler essas mensagens. A única saída era utilizar a mesma chave de encriptação original, a qual era trocada mensalmente pelos rotores e formulavam milhares de combinações (ALAN..., *on-line*).

## FIQUE POR DENTRO

De forma similar à máquina Enigma, as mensagens trocadas via aplicativo WhatsApp baseiam-se nesse tipo de criptografia, em que uma pessoa precisa ter a chave para poder ler a mensagem trocada no aplicativo.

Você pode entender melhor sobre isso em: [https://faq.whatsapp.com/pt\\_br/android/28030015/](https://faq.whatsapp.com/pt_br/android/28030015/). Acesso em: 20 fev. 2020.

Nesse contexto, Turing e sua equipe conseguiram descobrir como decifrar a máquina Enigma utilizando uma técnica eletromecânica nomeada de “bombas”, com mecanismos que permitiam a análise simultânea de muitos textos em uma velocidade superior aos humanos. Por exemplo, os ingleses passaram a ler mais de 50 mil mensagens por mês, em que cada mensagem era lida em 1 minuto (ALAN..., *on-line*). Dessa forma, os ingleses conseguiram ler todas as mensagens trocadas entre as tropas alemãs e, assim, surgiu a máquina de Turing (Figura 3.2):

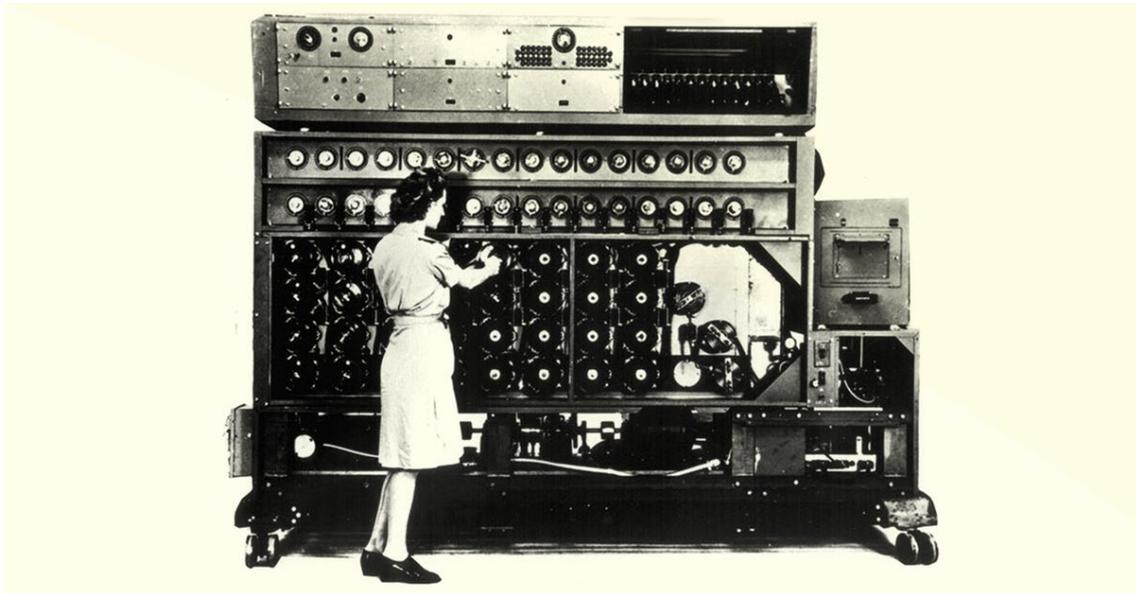


Figura 3.2 - Máquina de Turing

Fonte: War... (*on-line*).

Uma grande contribuição da máquina é o Teste de Turing, criado em 1950, que serve para identificar uma IA, em que, durante uma conversa textual, por um longo período de tempo, se não for possível classificar o sujeito com quem se comunica, entre humano ou IA, o sistema é considerado um autômato inteligente (ALAN..., *on-line*).

Com base nesse contexto, a IA se apoia nisso, quando utiliza o computador para processar símbolos de maneira eficiente e eficaz. A modelagem dos processos de IA pode ser feita primeiramente no papel, mas, depois, é necessária a automatização disso para se fazer analogias completas e complexas com o cérebro, utilizando, para isso, uma linguagem de símbolos.

Em uma visão generalizada, a computação engloba tudo o que é um processo efetivo, com o objetivo de obter um resultado, seja por meio de padrões, símbolos e desde que estes estejam bem-definidos no processamento computacional. E, assim, o computador é utilizado como recurso para executar todo esse processo.

Mas o que vem a ser IA? Vejamos algumas definições:

- É uma disciplina que faz uso das capacidades de processamento de símbolos da computação com o objetivo de encontrar métodos de forma genérica para automatizar atividades de percepção e cognição e manipulação por intermédio do computador (TEIXEIRA, 2009).
- É uma área da ciência da computação que investiga o descobrimento de métodos ou dispositivos computacionais, com o intuito de simular a capacidade humana na solução de problemas (MEDEIROS, 2018).
- É considerada um avanço tecnológico que faz com que sistemas promovam a simulação de uma inteligência semelhante à humana. Além disso, está além de somente códigos de programação para tornar decisões autônomas; está envolvida com as capacidades humanas, como: aprender, discernir sobre caminhos a seguir e pensar de forma racional diante de situações problemas (SELVAMANIKKAM, 2018).

Todas essas definições visam tentar simplificar o conceito de IA, o que é bem difícil. Pode-se basear em um conceito genérico que envolve o estudo de capacidades humanas apoiadas pelo uso do processamento computacional de símbolos.

Adicionalmente, ao longo da evolução de IA, ela foi dividida em quatro grandes linhas de pensamento baseadas na divisão de sistemas, que são (RUSSEL; NORVIG, 2013):

- I. **Sistemas que pensam como seres humanos:** nesse tipo de sistema, tem-se todo o conhecimento acerca do funcionamento da mente humana, como tomadas de decisão, resolução de problemas e aprendizagem, e essas teorias são executadas em programas de computador. A ciência cognitiva apresenta modelos computacionais da IA e técnicas experimentais da psicologia para a construção de teorias dos processos de funcionamento da mente humana. Um exemplo de sistemas que pensam como seres humanos são as redes neurais artificiais.
- II. **Sistemas que agem como seres humanos:** essa linha está relacionada a sistemas que requerem a atuação como humanos no sentido comportamental. A

inteligência, aqui, é criada para executar funções que também podem ser executadas por humanos. Um exemplo é o Teste de Turing, em que o computador deve ter as capacidades de processamento de linguagem natural, representação de conhecimento, raciocínio automatizado e aprendizado de máquina, a fim de responder a questionamentos como um ser humano.

- III. **Sistemas que pensam de forma racional:** o sistema é considerado racional se faz uso da notação lógica, simulando o pensamento lógico racional do ser humano, dentro das limitações dos recursos computacionais. Um exemplo de sistema que pensa de forma racional são os sistemas inteligentes.
- IV. **Sistemas que agem de forma racional:** esse tipo de sistema requer um agir racional, que consiste na utilização de um agente, o qual é uma entidade com sensores que percebe o ambiente e possui algumas formas de pensar racional programadas, para que, assim, possa tomar a ação correta no intuito de atingir um objetivo proposto. Um exemplo que podemos citar é a criação de agentes inteligentes para completar formulários de busca de passagens aéreas, ao invés do cliente com o objetivo de monitorar os preços de um determinado trecho.

Vale ressaltar que os sistemas I e III estão ligados ao pensamento e raciocínio, enquanto as demais estão atreladas ao comportamento humano. Além disso, os sistemas I e II envolvem estudos empíricos, formulação de hipóteses, confirmação do experimento e realizam a medição do sucesso da solução encontrada com base no quão fidedigna suas ações são das atitudes humanas. Já os sistemas III e IV comparam o sucesso com um conceito ideal de inteligência – que é a racionalidade – e utilizam a matemática e a engenharia na busca de soluções.

Por conseguinte, desde a sua evolução histórica, a IA tem alcançado um grande desenvolvimento através dos estudos desses quatro tipos de sistemas. Tais sistemas contribuem para realizar abordagens tanto centradas em seres humanos, quanto em torno da racionalidade. Agora, vamos entender a evolução da história da IA com base nos períodos a seguir (SELVAMANIKKAM, 2018):

**1943 – 1950:** As pesquisas estavam concentradas no modelo de neurônios artificiais para que pudessem ser desenvolvidas máquinas que fossem capazes de aprender.

**1951 – 1969:** O computador não é somente visto como uma máquina de calcular e, a partir disso, surgiram programas que jogavam xadrez, imitavam a forma de raciocínio humano, comunicavam-se em linguagem natural e analisavam as estruturas moleculares.

**1970 – 1980:** Aqui, tem-se o problema relacionado ao armazenamento dos dados e ao tempo de processamento. Nessa época surgiu a Teoria da Complexidade Computacional, que corroborou para que a solução dos problemas não fosse baseada somente em quantidade de memória adicional e tempo de processamento mais rápido.

**1980 – atual:** Houve o surgimento da linguagem de programação PROLOG, criada pelos japoneses, capaz de realizar milhões de inferências por segundo. A partir disso, grandes investimentos foram realizados na Europa e nos Estados Unidos, fazendo com que a área de IA se tornasse muito ativa, estando, atualmente, voltada para aplicações muito práticas como: manufatura, robótica, visão computacional, entre outras.

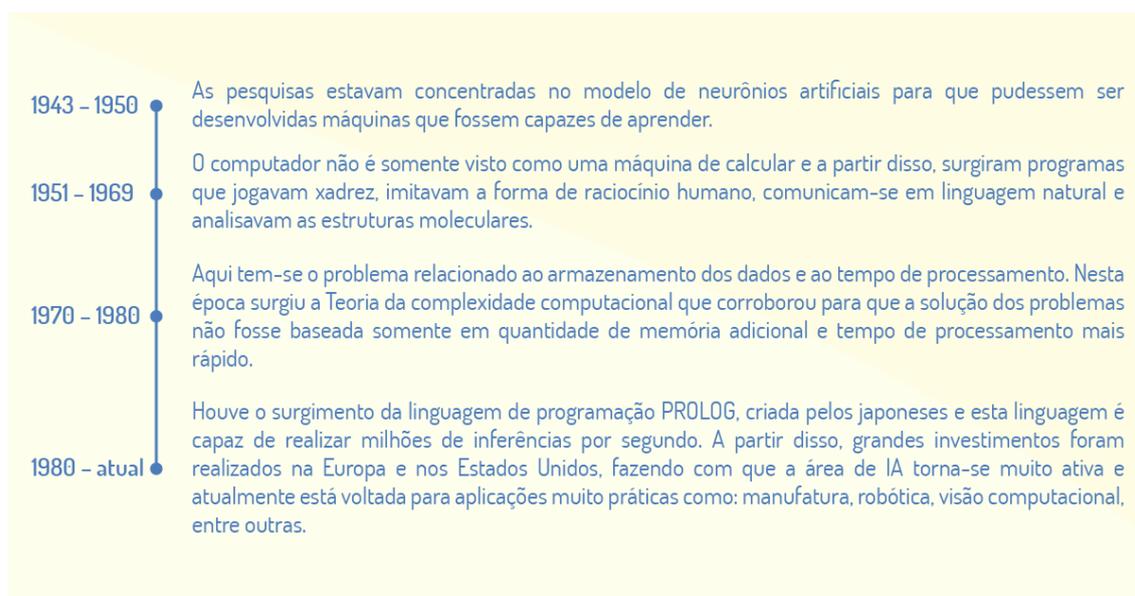


Figura 3.3 - A evolução histórica da IA

Fonte: Elaborada pelos autores.

Além dessa evolução história de IA, tem-se as três principais abordagens que podem ser utilizadas nessa área: **conexionista**, que está relacionada à hipótese de causa e efeito em que há a modelagem do cérebro humano para reproduzir as ações humanas – um exemplo são as redes neurais; **simbólica**, baseia-se na hipótese de sistemas de símbolos físicos, que dado um conjunto de estruturas simbólicas e regras de manipulação, são recursos suficientes para criar a inteligência – um exemplo são os sistemas; **evolucionária**, está relacionada à teoria evolutiva de Darwin, em que sistemas inteligentes são modelados a partir da evolução de indivíduos aleatórios e utilizam operações genéticas de recombinação e mutação – um exemplo são os algoritmos genéticos. Ainda se tem uma quarta abordagem em desenvolvimento, denominada **IA Híbrida**, que utiliza diversas ferramentas e mescla as abordagens acima para encontrar a solução de um problema (SELVAMANIKKAM, 2018).

Por fim, vimos como a IA é uma grande área de pesquisa e, atualmente, considerada uma grande evolução da tecnologia, dividindo os sistemas nas diversas formas de pensar e raciocinar e também com suas diferentes abordagens. Tudo isso corrobora para que a IA se sedimente como uma área de inovação na criação de máquinas que podem ser comparadas à forma de pensar e agir de forma humana!

## ATIVIDADE

1) Agentes inteligentes são capazes “de fazer uma leitura do ambiente em que está inserido e, a partir daí, propor soluções que maximizem as chances de sucesso. A Inteligência Artificial, que antes estava restrita às grandes empresas, hoje influencia a compra de produtos, muda o ambiente, melhora o dia a dia corporativo e facilita a vida das pessoas, de maneira racional” (ALVES, 2019, *on-line*).

Conforme a classificação dos sistemas ao longo da evolução da IA (Inteligência Artificial), os agentes inteligentes citados no texto, podem ser classificados corretamente como:

- a) sistemas que pensam como seres humanos.
- b) sistemas que agem como seres humanos.
- c) sistemas que pensam de forma racional.
- d) sistemas que agem de forma racional.
- e) sistemas especialistas.

## **Tipos de problemas onde a IA é indicada**

A inteligência artificial foi elaborada com dois grandes objetivos: o primeiro é ter a habilidade de entender que a eficácia foi atingida em determinada situação e, assim, aplicar esse aprendizado para outros contextos de problemas; o segundo se refere à possibilidade de resolver situações e problemas novos com êxito, com base no conhecimento que a IA adquiriu durante a solução do problema (LUGER, 2013).

No que se refere aos tipos de problemas que a IA pode auxiliar, temos uma diversidade de exemplos que corroboram que utilizar IA não é novidade e que já faz parte do nosso dia a dia. Vamos categorizar os problemas e exemplos a partir das seções seguintes.

### **Indústria**

O problema de automatizar os processos industriais vem ocorrendo há muito tempo. Desde que mudaram os modelos de fabricação, por exemplo, já foi uma tentativa de fazer com que as tarefas manuais fossem automatizadas, principalmente com a evolução das máquinas industriais e até mesmo com os computadores (LUGER, 2013).

A partir disso, a ideia era produzir mais em menor tempo, para obter o maior lucro possível. Esse contexto está atrelado aos processos produtivos contínuos, que possuem atividades repetitivas e frequentes, fazendo com que a IA aprenda esse tipo de atividade, ou seja, as máquinas copiam esses padrões humanos de atividade.

Quando esse aprendizado ocorre, muitos dados podem ser combinados de forma rápida, os quais são processados de forma eficiente, fazendo com que haja uma interpretação dos dados gerado, para, assim, serem adquiridos como conhecimento pelas máquinas.

Um dos grandes benefícios atrelados ao uso de IAs em processos contínuos das empresas é a eliminação de erros em virtude da falta de atenção humana naquele processo ou até mesmo a fadiga, pois as máquinas trabalham ininterruptamente e, assim, produz-se um nível de produtividade muito maior (LUGER, 2013).

Outro benefício que podemos citar é a redução de custos com pessoal. Em algumas lojas de comércio eletrônico, são utilizados *chatbots*, programas de IA que simulam uma conversa com seres humanos, enquanto o atendente humano só é solicitado quando envolve um problema de complexidade maior. Sendo assim, os custos com os funcionários são menores, ou até podem alocar alguns funcionários para áreas mais estratégicas da indústria para maximizar o lucro da empresa e focar nos negócios da organização (LUGER, 2013).

## REFLITA

Mas, na prática, como podemos perceber que as IAs resolvem os problemas da indústria? Vejamos, a seguir, alguns exemplos (LUGER, 2013).

1. Análise de dados: podem ser implantados sensores no processo produtivo, por meio de soluções de IoT (Internet das Coisas), nas tarefas repetitivas e frequentes. Com base nos dados recebidos dos sensores, as IAs podem chegar a padrões que serão utilizados para otimização das atividades de rotina.
2. Manufatura avançada: atualmente considerada manufatura inteligente, dentro do contexto da Indústria 4.0, está adicionando robôs industriais em seus processos operacionais. Além disso, auxilia na demanda de atividades da logística.
3. Gestão de pessoas: as IAs podem ajudar a prever um padrão de candidato que poderia se encaixar para a vaga ofertada pelos recursos humanos da empresa. Pode também ser utilizada para automatizar as rotinas de trabalho do setor.
4. Manutenção preditiva: esse tipo de manutenção requer o acompanhamento frequente do funcionamento das máquinas, podendo ser adotadas soluções de IoT (Internet das Coisas), com uso de sensores para detectar falhas antes que ocorram paradas nas linhas de produção. Para isso, são utilizadas as IAs para realizarem esse levantamento e diagnósticos precisos para a análise das potenciais falhas que poderão ocorrer.

Além dessas áreas, podemos citar algumas situações em que as IAs já substituíram os humanos nos processos produtivos, como: algumas máquinas inteligentes já realizam a conferência dos produtos, sem passar por operadores humanos; já existem máquinas atreladas ao processo criativo da empresa, que, por conta própria, executam novos projetos de tarefas criativas, similares aos humanos.

## Saúde

Nessa área, as IAs estão sendo aplicadas para a prevenção de doenças, aprimoramento de diagnósticos e a indicação de tratamentos em diversos casos de saúde. Em um estudo publicado pela Nature (YU; BEAM; KOHANE, 2018), alguns robôs já são utilizados para fornecer orientações de conduta dentro de um ambiente hospitalar, baseado nos padrões adquiridos dentro desse ambiente.

Além disso, as IAs possuem uma quantidade significativa de dados e algoritmos (sequência lógica de dados) com o objetivo de fornecer informações precisas nas tomadas de decisões para auxiliar os profissionais de saúde. Para isso, todos os dados (prontuários, pesquisas acadêmicas relevantes, entre outros) são armazenados em nuvem, e, após realizar o processamento e o cruzamento de informações, as IAs oferecem aos médicos uma diretriz para fazer um diagnóstico mais preciso ou até mesmo prescrever tratamentos terapêuticos (NOSÉ, 2019).

Dentre as diversas áreas de saúde podemos citar a oftalmologia, que já se prevalece do uso de IAs em cirurgias de cataratas. A IA calcula em tempo real todas as informações sobre o globo ocular no momento da cirurgia e sugere a lente intraocular mais recomendada após a remoção do cristalino (lente natural do olho); dessa forma, evita limitações visuais do paciente após a cirurgia (NOSÉ, 2019).

Um exemplo que podemos citar em precisão de diagnósticos com câncer é o *software IBM Watson for Oncology*, que, baseado em uma gama de informações disponibilizadas a ele sobre o paciente em questão, consegue chegar a um diagnóstico mais preciso e assertivo (IBM WATSON, 2019).

Outro problema que as IAs podem solucionar é a triagem automática de pacientes para priorizar os casos emergenciais. Dessa forma, aqueles casos mais críticos aparecem em primeiro para atendimento médico (OS BENEFÍCIOS..., 2020).

Outra questão é a utilização de rede neural como forma de fornecer uma análise mais aprofundada nos prontuários para detectar descobertas de doenças em nível sobre-humano (OS BENEFÍCIOS..., 2020).

Adicionalmente, alguns aparelhos já vêm com IAs programáveis, que notificam o médico instantaneamente sobre mudanças no estado clínico do paciente. Dessa forma, o médico já registra isso no prontuário e, assim, começa o atendimento de forma eficaz.

Outra utilização é na precisão de um diagnóstico muito complexo, pois a demora em acertar o diagnóstico do paciente pode prejudicá-lo, pela demora no início do tratamento. Sendo assim, a IA faz a associação dos sintomas com todo o histórico médico do paciente e sugere as possíveis doenças que este pode desenvolver, para, então, o médico poder iniciar as profilaxias adequadas (OS BENEFÍCIOS..., 2020).

Uma perspectiva futura nessa área é na predisposição de determinadas doenças, mas com base no histórico familiar. Para isso, todos os prontuários precisarão estar armazenados na nuvem, sendo necessárias mais regras de segurança e, até mesmo, permissões do conselho de ética da área de saúde para aprovar o registro desses prontuários na internet (OS BENEFÍCIOS..., 2020).

Por fim, o setor de saúde está sendo muito otimizado com a aplicação das IAs para buscar recursos na melhoria das experiências dos pacientes, agilizando seus atendimentos, e dos melhores procedimentos para a redução de custos dentro do hospital.

## **Educação**

Um das questões em educação e que as IAs estão auxiliando é na criação de plataformas adaptativas, que consistem na sugestão de trilhas de aprendizado individualizadas com base nos dados que se tem de cada aluno (STEFANINI, 2019).

Para fazer esse trabalho, a IA mapeia todo o conteúdo de uma determinada área de conhecimento e associa com as dificuldades que o aluno tem na disciplina. Assim, para ajudar ao aluno, a IA propõe um plano de estudo adequado para ele, como forma de minimizar as dificuldades encontradas naquele conteúdo.

Outra solução que a IA pode resolver são para os cursos de educação a distância no que se refere à correção automática de provas, como forma de auxiliar a rotina do professor, sendo mais rápido que a correção manual, melhorando a sua eficiência nos planos de aula (MARTINS, 2010).

Uma nova aplicação é o Professor 24 Horas, que consiste em uma IA que se propõe a tirar as dúvidas dos alunos 24 horas por dia e pode ajudá-los a aprofundar os conhecimentos nas disciplinas do seu curso à distância. A plataforma ainda oferece um conteúdo extenso e respostas bem precisas para as dúvidas dos alunos, que podem acompanhar sua evolução a partir de relatórios e, ainda, compartilhar tudo isso em uma rede social própria (STEFANINI, 2019).

Por conseguinte, a gamificação pode ser melhorada para promover o aprendizado dos alunos por meio de jogos. Fazendo o equilíbrio entre o jogo ser desafiador e seu nível de dificuldade adequado, a IA analisa o desempenho de cada estudante e faz adaptações no jogo para cada nível de atuação do aluno. Se o aluno sempre ganha, o jogo evolui para se tornar mais difícil, mas, se o aluno está sempre perdendo, o jogo também é adaptado para as suas capacidades. É assim que a gamificação está se tornando inteligente (MARTINS, 2010).

Por fim, as IAs estão sendo utilizadas para conter alunos propensos à evasão. A partir disso, a escola pode trazer iniciativas eficazes (seja acompanhamento ou até comunicado para a família) para promover a aprendizagem mais eficiente e o favorecimento da permanência do aluno na instituição.

E como podemos perceber a utilização da IA em nosso dia a dia? Vamos entender, agora, como isso funciona!

De acordo com Martins (2010), a IA aparece em nosso cotidiano das seguintes formas:

1. assistentes pessoais: esse tipo de IA está atrelado ao reconhecimento de voz e, a partir disso, pode fazer diversas tarefas, como é o caso nos *smartphones*: pode fazer uma ligação, abrir um aplicativo, informar o tempo. Cada vez mais elas estão aprendendo os diferentes idiomas e pronúncias e estão se tornando mais naturais.
2. sugestão de filmes: se você tem acesso à Netflix ou ao YouTube já deve ter recebido uma recomendação de filme. Tudo isso é realizado por IAs, que geram um padrão do seu gosto por filmes e, a partir disso, fazem as recomendações. Essas IAs também são utilizadas em aplicativos de música, como no Spotify.
3. *e-mail*: com base em certas características do e-mail recebido, as IAs já sabem que isso se trata de lixo eletrônico e envia diretamente para a pasta de spam. Esse aprendizado ocorreu quando os usuários realizavam essa ação de forma frequente. Hoje em dia, a IA já faz isso de forma automática.
4. segurança dos aparelhos móveis: as IAs já fazem uso de reconhecimento facial e por voz para realizar o desbloqueio dos telefones dos usuários. Até mesmo rostos disfarçados já foram descobertos com o uso das IAs, que têm uma taxa de erro bem pequena – praticamente se igualando a um humano nesse tipo de reconhecimento – e realizam a segurança dos dados do aparelho de forma eficiente.
5. *chatbots*: são IAs que conversam com o usuário sobre dúvidas frequentes em relação aos produtos da empresa. Eles estão disponíveis 24 horas por dia para ajudar ao consumidor, por meio de troca de mensagens.

## REFLITA

De repente, você está em casa sozinho à noite e a sua assistente pessoal Alexa começa a rir sozinha. Desesperador não? Uma assistente pessoal totalmente programável, criando atitudes de forma autônoma. Foi isso que ocorreu nos Estados Unidos com alguns usuários que registraram com vídeos e enviaram à Amazon. A empresa confirmou o problema e estava corrigindo-o.

Será que, ao inserirmos tanta Inteligência Artificial em nosso cotidiano, fazendo com que elas assumam nossas tarefas, chegará um dia em que elas começarão a decidir da forma que quiserem o que pretendem ou não fazer com a inteligência que lhes foi atribuída? (YUGE, 2018).

Por fim, vimos algumas das áreas em que as IAs podem ser utilizadas, sendo que algumas já estão totalmente imersas em nosso dia a dia e até mesmo algumas perspectivas futuras em cada área. É importante que as IAs possam contar com o acesso a um grande volume de dados para prover processamentos cada vez mais assertivos, havendo, nesse sentido, uma relação muito forte com o uso de tecnologias como Big Data.

Muitas pesquisas ainda estão sendo realizadas dentro da área de inteligência artificial para fazer com que ela se solidifique e se intensifique, seja em nossa vida pessoal ou profissional. A IA passa de uma perspectiva meramente de ficção científica para se adequar à otimização de tarefas feitas por humanos. No entanto, é necessário, também, compreender os riscos inerentes à utilização das IAs e seus impactos na vida cotidiana das pessoas.

## ATIVIDADES

2) Leia o trecho a seguir.

“São *softwares* movidos por Inteligência Artificial, capazes de executar pequenas tarefas e fornecer informações solicitadas pelo usuário. Eles partem, em resumo, da conta do usuário, reconhecimento de voz e de sua geolocalização para acessar várias fontes sobre clima, trânsito, horários, entre outros temas. São exemplos desses *softwares*: Siri, Alexa, Cortana entre outros” (RAHDE, 2019, *on-line*).

Com base no texto acima, podemos afirmar que esses *softwares* podem ser definidos como qual tipo de utilização de inteligência artificial (IA) em nosso dia a dia?

- a) Assistentes pessoais.
- b) *Chatbots*.
- c) Leitor de *e-mail*.
- d) Análise de dados.
- e) Gestores de pessoas.

## Lógica *fuzzy*

Vimos, até agora, os conceitos inerentes à Inteligência Artificial e em quais contextos ela pode ser aplicada, certo? Mas, agora, precisamos ir um pouco mais além: se precisássemos de sistemas inteligentes que aceitassem parametrização de dados subjetivos ou até mesmo incertos (ex.: pouco alto, pouco caro, entre outros), de que forma poderíamos realizar isso? Iríamos precisar utilizar a lógica *fuzzy* para realizar esse tipo de tratamento com esses dados. Vamos começar!

O termo “*fuzzy*”, quando traduzido do inglês para o português, significa “difuso, confuso, vago”, e a palavra “lógica” denota um estudo para saber se os argumentos são válidos ou não. Sendo assim, a lógica *fuzzy* pode ser entendida como a forma com que os argumentos considerados vagos podem ser válidos ou não em determinadas situações (SIMÕES, 2007).

Em outra definição, pode-se dizer que a lógica *fuzzy* consiste em uma forma da lógica multivalorada, em que são considerados os valores no intervalo entre 0 e 1 para representar situações de imprecisão ou incertezas do mundo real (ZADEH, 1976).

Além disso, a lógica *fuzzy* é um ramo da inteligência artificial, de sistemas computacionais e para o controle de processos. Ela permite que as informações descritas em linguagem natural possam ser convertidas em formatos binários. Essas informações formam o que são denominados de termos *Fuzzy* (MARRO *et al.*, 2010).

Mas o que são termos *Fuzzy*? Vejamos situações do nosso dia a dia que refletem isso (MARRO *et al.*, 2010):

- Em placas de trânsito, quando temos a informação: siga por “alguns km”.
- Em previsões do tempo: O dia está “parcialmente” nublado.
- Classificação de objetos em: “largo”, “sujo”.
- Descrição de características físicas como: “alto”, “baixo”, “gordo”.
- Descrição das pessoas pela faixa etária: “idoso”, “criança”.

Os termos entre aspas refletem medidas incertas ou imprecisas e são considerados conceitos muito vagos. No exemplo da altura, não existe um limite preciso que determine

essa medida; mesmo se considerarmos que homens altos são aqueles que medem a partir de 1.90 m, então, os homens que estão com 1.89 m são considerados “quase altos”, e não altos.

Perceba que, com a lógica *fuzzy*, não há como responder somente “sim” ou “não” para as situações específicas com todas as informações conhecidas. É necessário informar valores mais relativos, como “talvez”, para que seja mais apropriado para o contexto informado.

## Histórico

Os conceitos de lógica *fuzzy* começaram por volta do ano 1920, por Jan Lukasiewicz, que iniciou o conceito de graus de pertinência que, combinados com a lógica clássica desenvolvida por Aristóteles, formaram a base da lógica *fuzzy*. Os graus de pertinência iniciaram com os valores 0,  $\frac{1}{2}$  e 1 e, depois, expandiram para um conjunto infinito de valores entre 0 e 1 (SIMÕES, 2007).

Vale ressaltar que um grau de pertinência se refere a diferentes graus que são considerados verdadeiros. Assim, os graus de verdade estavam dispostos entre 0 e 1, e as informações que se quer medir admitem graus. Como exemplos dessas informações pode-se citar: altura, velocidade, temperatura, entre outros (SIMÕES, 2007).

A partir desse estudo, em 1960, um professor de ciência da computação da universidade de Califórnia, Lotfi Asker Zadeh, começou a se basear nesse estudo e foi o autor que publicou o primeiro artigo denominado “*Fuzzy sets*”.

Zadeh (1976) começou a observar que muitas das regras que as pessoas costumeiramente usavam não podiam ser explicadas com base na lógica tradicional. Por exemplo, pode-se olhar para uma pessoa e supor que ela tem 40 anos, mas não se sabe explicar esse fato.

Zadeh procurou explicar ideias assim por meio da lógica *fuzzy*. Porém, os estudos de Zadeh (1976) não foram aceitos de forma branda e foram alvo de muitas críticas que não consideravam os conceitos divergentes da lógica clássica, já que acreditavam que a lógica tradicional ainda poderia encontrar uma forma de explicar os valores difusos.

Contudo, conforme os estudos de Zadeh (1976) foram avançando, alguns precursores foram utilizando-os na prática, como os japoneses, que construíram os primeiros sistemas de controle *fuzzy*, em 1987, para ser utilizado na estrada de ferro de Sendai, localizada no Japão. Esse sistema visava controlar a aceleração, frenagem e a parada nas estações metroviárias. Além desse uso, houve, também no Japão, a aplicação dessa lógica em um sistema de água. Somente nos anos 1990 é que a lógica *fuzzy* ganhou a importância em empresas nos Estados Unidos (ZADEH, 1976).

Dessa forma, a lógica *fuzzy* foi aplicada em diversos setores, principalmente os industriais, que serviram para solidificar os conceitos desse tipo de lógica. Por isso é considerada uma técnica padrão e tem uma boa aceitação tanto em processos industriais quanto comerciais.

### **Conjuntos *fuzzy***

Antes de começarmos a conceituar os conjuntos *fuzzy*, os quais são de extrema importância para o entendimento da lógica *fuzzy*, vejamos, a seguir, alguns princípios que norteiam a lógica (MARRO *et al.*, 2010).

1. Está baseada em palavras, de forma semântica, ou seja, os valores expressos são realizados em forma de palavras. Ex.: médio, quente, alto, frio, entre outros, que podem ser utilizados para a definição de estados de uma variável.
2. Podem ser utilizados alguns advérbios ou modificadores na variável utilizada, como: mais ou menos, muito, bastante, pouco, entre outros.
3. Também podem ser utilizados alguns quantificadores do estado da variável como: vários, muitos, usualmente, entre outros.

Além desse princípio, faz-se necessário explicar também como a lógica *fuzzy* diverge da lógica clássica (GOMES; RODRIGUES, *on-line*).

- As variáveis utilizadas na lógica clássica exigem uma definição exata, como: “é mortal”, “é mulher”. Já na lógica *fuzzy*, não se tem a exatidão dos predicados, ou seja, da declaração exata dos valores das variáveis, por exemplo: “é alto”, “é

velho”, “é jovem”, entre outros.

- A lógica clássica só admite respostas consideradas como verdadeiras ou falsas, enquanto, na lógica *fuzzy*, os valores possuem um grau de veracidade que pode variar entre totalmente falso (0) e totalmente verdadeiro (1).
- Na lógica clássica, uma proposição lógica só pode possuir dois extremos: completamente falsa ou verdadeiro, ou seja, só há a possibilidade de um resultado. Já na lógica *fuzzy*, os valores podem ser parcialmente verdadeiros ou parcialmente falsos.
- A função de pertinência, na matemática clássica, é a função que define se um elemento pode ou não pertencer a um determinado conjunto e somente poderá assumir os valores 0 (zero) ou 1 (um), conforme demonstra a função de pertinência na matemática clássica:

$$f_A(x) = \begin{cases} 1, & \text{se e somente se } x \in A \\ 0, & \text{se e somente se } x \notin A \end{cases}$$

Na lógica *fuzzy*, o resultado de uma proposição pode ter variância em graus de verdade, podendo ser parcialmente verdadeiro ou parcialmente falso. Sendo assim, a função de pertinência  $\mu(x)$  pode assumir valores no intervalo entre 0 e 1, podendo ser totalmente verdadeiro ou totalmente falso, conforme explicitado na função de pertinência de um conjunto *fuzzy* (GOMES; RODRIGUES, *on-line*):

$$\mu_A(x): X \rightarrow [0,1]$$

Agora, podemos conceituar os conjuntos *fuzzy* como funções que fazem o mapeamento de um valor escalar no intervalo entre 0 e 1, o qual indica seu grau de pertinência a um determinado conjunto (GOMES; RODRIGUES, *on-line*).

Podemos dizer, ainda, que o conjunto *fuzzy* é representado pela letra A e que o grau de pertinência determina os valores do conjunto  $\mu$  em relação ao conjunto *fuzzy* A, ou seja, quanto é possível para um elemento x pertencer ao conjunto A.

Vamos entender melhor a aplicação da fórmula, por meio de alguns exemplos (MELO NETO, *on-line*)?

1. Vamos representar o conceito de céu ensolarado através do conjunto:

$$A: \{(0, 1.0), (20, 0.8), (30, 0.4), (75, 0.0)\}$$

Os valores de pertinência são: 1.0, 0.8, 0.4 e 0.0. Podemos, então, interpretar que: a pertinência de 1.0 tem uma cobertura de nuvens de 0%, a de 0.8 tem uma cobertura de nuvens de 20%, a de 0.4 tem cerca de 30% de nuvens, e a de 0.0 tem cerca de 75% de nuvens ou mais. Logo, o céu ensolarado está presente no primeiro par ordenado (0, 1.0).

2. Sendo A o conjunto de pessoas altas, tem-se os seguintes gráficos para representar se uma pessoa com 1.70 é alta ou não, diferenciando-se pela definição de conjuntos clássico e *fuzzy*:

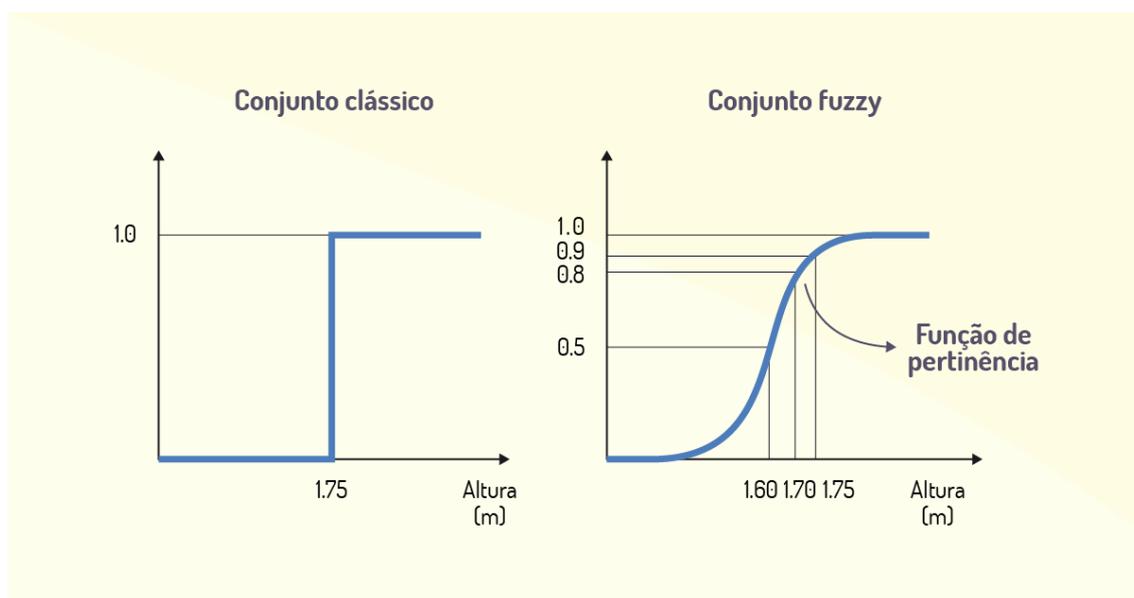


Figura 3.4 - Conjunto Clássico x Conjunto *Fuzzy*

Fonte: Adaptada de Melo Neto (*on-line*).

Pode-se visualizar que, no conjunto clássico, uma pessoa com 1.70 de altura não pertence ao conjunto A de pessoas altas e tem grau de pertinência 0. Já no conjunto *fuzzy*, essa pessoa em questão pertence ao conjunto A de pessoas altas e tem grau de pertinência 0.8.

Por fim, entendemos que a lógica *fuzzy* é uma técnica que permite capturar as informações precisas ou incertas descritas em linguagem natural para, depois, serem convertidas em formato numérico, formando os conjuntos *fuzzy*.

### Estudo de caso

Com o conceito dos conjuntos *fuzzy* para representar as incertezas de algumas informações, tem-se a necessidade de entender como se dará a modelagem de um sistema que se baseia na lógica *fuzzy*. Esse tipo de sistema é nomeado de Sistemas de inferência *fuzzy*, ou *Fuzzy Inference Systems* (FIS), que visam representar a modelagem do raciocínio humano em um formato de regras (GOMES; RODRIGUES, *on-line*).

É comum expressarmos o raciocínio de forma lógica, com alguns conectivos que antecedem uma ação e, depois, um conectivo para representar a reação disso. São utilizados, por exemplo, SE (antecedente), ENTAO (consequência), e ainda podem ser utilizados diversos operadores lógicos como “e” e “ou”.

Nesse contexto, uma arquitetura para sistemas de inferência *fuzzy* é proposta (Figura 3.5) (GOMES; RODRIGUES, *on-line*):

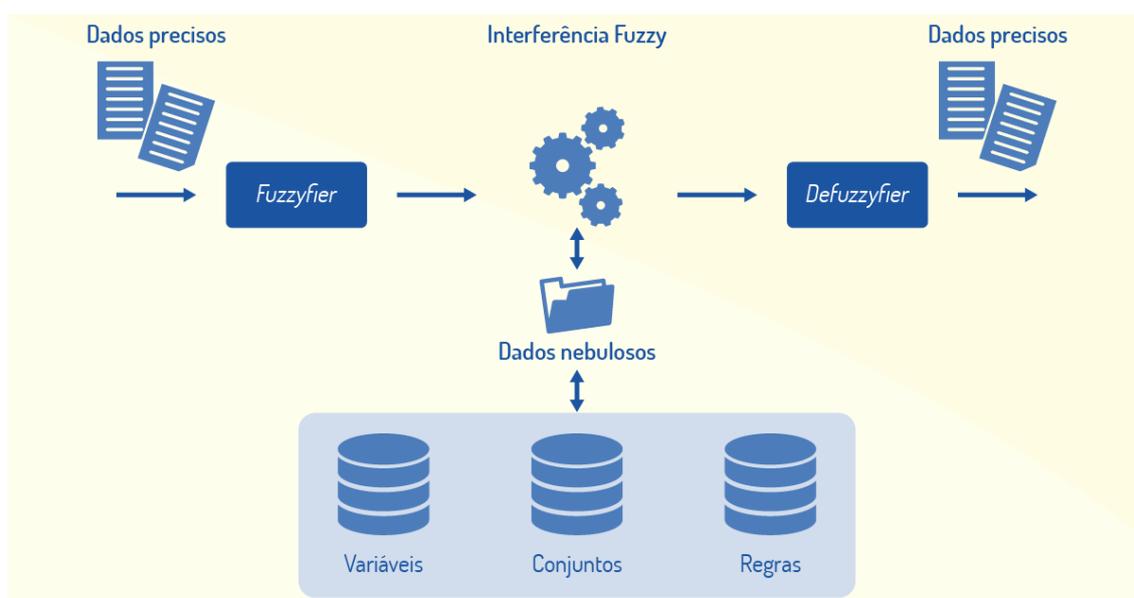


Figura 3.5 - Arquitetura de um sistema de inferência *fuzzy*

Fonte: Adaptada de Gomes e Rodrigues (*on-line*).

A arquitetura possui quatro componentes: processador de entrada (*fuzzyfier*), que transforma os dados de entrada precisos em representações nebulosas; uma base de conhecimento, representada pelas variáveis nebulosas, suas funções de pertinência e um conjunto de regras descritas por um especialista; uma máquina de inferência *fuzzy*, responsável pela avaliação das regras e inferência da saída; e um processador de saída (*defuzzifier*), que fornece um número real como saída da inferência.

Dessa forma, os dados precisos fornecidos como entradas para o sistema de inferências são transformados em dados nebulosos (*fuzzification*). Estes são processados pelo núcleo de cálculo, em função do mapeamento de regras definidas e, então, transformados novamente em dados precisos (*defuzzification*), que são retornados como saída pelo sistema de inferências. É importante destacar que tanto as entradas como a saída de um sistema de inferências *fuzzy* são valores numéricos, que são internamente processados utilizando-se definições *fuzzy*.

Vejamos como um sistema desses pode auxiliar na tomada de decisão em um sistema de RH (Recursos Humanos), em que se tem a implantação de uma nova funcionalidade: determinar os valores de gratificação dada aos funcionários, evitando a parcialidade e promovendo a isenção reivindicada pelo quadro de funcionários insatisfeitos (GOMES; RODRIGUES, *on-line*). O projetista informou que os critérios de experiência e de capacitação serão usados como variáveis de entrada do sistema, sendo a experiência expressa entre 0 e 30 anos, enquanto o tempo de capacitação é expresso entre 0 e 15 anos. A gratificação é a variável de saída, com valores entre R\$ 0,00 e R\$ 1.000,00. Esses intervalos são conhecidos como o universo de discurso das variáveis. O especialista definiu que a variável de entrada **experiência** se divide em três segmentos: pouca, média e muita. A variável de entrada **capacitação** seguiu uma divisão similar em três partições: fraca, média e forte. E a saída, representada pela variável **gratificação**, foi dividida em cinco termos: muito baixa, baixa, média, alta e muito alta.

Na sequência, o especialista determina as regras que direcionarão a decisão feita pelo sistema, constituídas pelo encadeamento de condições e uma conclusão, na forma: SE

<condição1> E <condição2> ENTÃO <conclusão>. A conclusão indica a qual conjunto a variável de saída pertence. As regras para o problema de valoração de gratificação definidas pelo especialista são as seguintes:

*SE capacitação é “Fraca” E experiência é “Pouca”, ENTÃO, gratificação é “Muito-Baixa”.*

*SE capacitação é “Fraca” E experiência é “Média”, ENTÃO, gratificação é “Baixa”.*

*SE capacitação é “Fraca” E experiência é “Muita”, ENTÃO, gratificação é “Média”.*

*SE capacitação é “Média” E experiência é “Pouca”, ENTÃO, gratificação é “Baixa”.*

*SE capacitação é “Média” E experiência é “Média”, ENTÃO, gratificação é “Média”.*

*SE capacitação é “Média” E experiência é “Muita”, ENTÃO, gratificação é “Alta”.*

*SE capacitação é “Forte” E experiência é “Pouca”, ENTÃO, gratificação é “Média”.*

*SE capacitação é “Forte” E experiência é “Média”, ENTÃO, gratificação é “Alta”.*

*SE capacitação é “Forte” E experiência é “Muita”, ENTÃO, gratificação é “Muito-Alta”.*

Uma visão segmentada dos processos de fuzzificação e defuzzificação pode ser observada no diagrama da Figura 3.6, em que é avaliado o caso de um profissional com 10 anos de capacitação e 20 anos de experiência. O processo usa a capacitação e a experiência informada para rastrear as regras que foram ativadas e conclui que o profissional é merecedor de uma gratificação de R\$ 600,00.

Observando as regras já definidas e o diagrama da Figura 3.6, nota-se que as regras 5, 6, 8 e 9 foram ativadas por atenderem às condições impostas. Com os graus de inclusão dos conjuntos *fuzzy* de gratificação já ponderados, a defuzzificação é efetuada, gerando a saída única, que é o valor de gratificação esperado como resultado final do processo de decisão nebuloso.

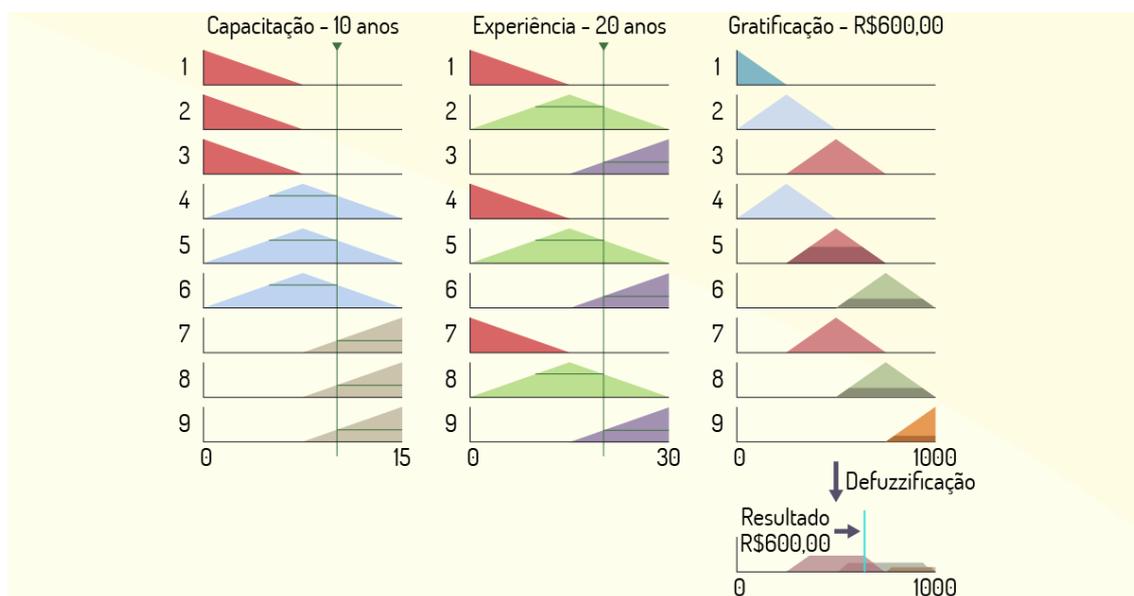


Figura 3.6 - Processo de fuzziificação e defuzziificação

Fonte: Adaptada de Gomes e Rodrigues (*on-line*).

Para se chegar ao valor de R\$ 600,00, foram consideradas as áreas em negrito da terceira coluna na direita do diagrama, realizadas por uma API JFuzzyLogic, que recebeu os valores para realizar a defuzziificação e gerou a saída por meio do cálculo do centro de gravidade das áreas em negrito da coluna gratificação do diagrama. Esse cálculo já está implementado como método dentro da API e é transparente para o projetista da lógica *fuzzy*.

Por fim, esse exemplo da determinação do valor de gratificação, apoiado pela lógica *fuzzy*, esclarece o processo de modelagem de uma solução orientada a regras de negócio encontrada no dia a dia de muitos sistemas corporativos.

## ATIVIDADE

3) Termos *fuzzy* são utilizados diariamente em linguagem natural para representar dados incertos ou, como são denominados em computação, nebulosos. Tais dados são considerados muito vagos e/ou subjetivos e não possuem um limite muito bem definido.

Com base na definição dos termos *fuzzy*, qual das alternativas abaixo aplica corretamente esse conceito ao termo sublinhado? Assinale a alternativa correta.

- a) O dólar está com cotação de R\$ 3,50 (reais) na bolsa de valores.
- b) O trabalho está realizado, de forma concluída, em 100%.
- c) A água está muito quente.
- d) Siga a estrada por 10 km.
- e) O idoso tem uma escala de idade a partir de 50 anos.

## Algoritmos genéticos

Os algoritmos genéticos (AG) podem ser definidos como uma classe particular de algoritmos evolutivos que utilizam as técnicas da biologia evolutiva, seleção natural e *crossing over* (recombinação) e se baseiam na teoria da evolução de Darwin (FERNANDES, 2003).

Os AGs são considerados algoritmos de otimização global que se baseiam nos mecanismos de seleção natural e de genética. Eles utilizam uma estratégia de busca paralela, de forma estruturada, aleatória e direcionada para buscar informações históricas de novos pontos de busca para obter o melhor desempenho (GILLIARD; PACHECO, *on-line*).

Adicionalmente, técnicas de busca e otimização tradicionais operam com um único candidato que, de forma iterativa, utiliza algumas estatísticas a serem associadas ao problema a ser solucionado. Tais processos estatísticos não são algorítmicos, e a simulação nos computadores pode se tornar muito complexa (GILLIARD; PACHECO, *on-line*).

Já as técnicas de computação evolucionária trabalham com uma população de candidatos em paralelo e realizam a busca em diferentes áreas para a solução do problema, tendo um número de membros específico para cada região a ser realizada a busca da solução (FERNANDES, 2003).

Por conseguinte, os AGs diferem-se dos métodos tradicionais de busca e otimização nos seguintes aspectos (FERNANDES, 2003):

1. operam por meio de uma codificação de um conjunto de parâmetros, e não com parâmetros isolados e individualizados.
2. iniciam com um certo número populacional, e não com um único indivíduo.
3. utilizam regras de transição baseadas na probabilidade, e não determinísticas.
4. fazem uso de informações de custo ou recompensa nos algoritmos, e não de conhecimento auxiliar, como ocorre com os métodos tradicionais.

Por fim, os algoritmos genéticos são utilizados para a busca de soluções ótimas de forma eficiente, ou até mesmo para solucionar uma grande variedade de problemas, já que não há tantas limitações como ocorre com os métodos tradicionais. Além disso, os AGs são considerados uma classe de procedimentos com várias etapas separadas, em que cada uma dessas etapas possui uma grande variação de soluções possíveis.

### Características gerais

Antes de começarmos a descrever os AGs, vamos conceituar alguns termos relevantes para o entendimento dessa área (RUSSEL; NORVIG, 2013). Os AGs atuam sobre uma **população**, que representa um conjunto de **indivíduos** dentro de um espaço de busca. Realizando uma ligação com a genética, um indivíduo corresponde a um **cromossomo**, que contém **genes** que o caracterizam. Do ponto de vista computacional, cada indivíduo (ou cromossomo) é codificado por uma cadeia de valores representativos, em uma função objetivo utilizada para buscar a solução de um determinado problema de otimização, sendo que os genes são os seus parâmetros utilizados no formato binário, inteiro ou real.

Assim, a proposta de um AG é selecionar, a partir de uma população com indivíduos gerados aleatoriamente, indivíduos melhor adaptados (aptidão) a uma função objetivo para resolver um problema por meio de uma análise conhecida como **adaptação (*fitness*)**. Essa seleção é feita para que possam ser criados descendentes que irão estar mais aprimorados para a solução, sendo esta conhecida como **reprodução**. A descendência de indivíduos é criada por meio de processos de **cruzamento (*crossover*)** (combinação de dois indivíduos/cromossomos intercambiando sua informação genética) ou **mutação** (introdução de novas combinações genéticas nos indivíduos), aumentando, assim, a diversidade e variabilidade da população. Dessa forma, novos indivíduos com novos códigos genéticos serão gerados, como pode ser visto na Figura 3.7, a seguir.

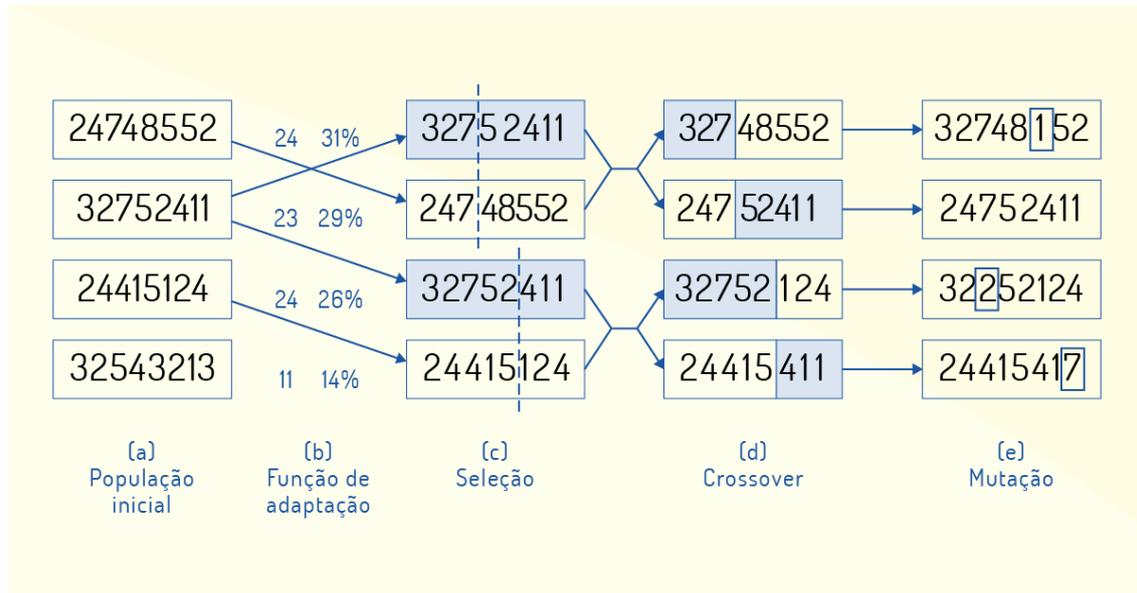


Figura 3.7 - Sequência de passos de um algoritmo genético

Fonte: Russel e Norvig (2013, p. 166).

Assim como ocorre na evolução dos seres humanos – que não aconteceu apenas uma única vez, mas continua ocorrendo através de vários ciclos ao longo da história –, os algoritmos realizam, de forma interativa e sequencial, cada processo de seleção e reprodução, gerando uma nova população, denominada **geração**. Cada indivíduo de uma geração apresenta uma **aptidão bruta**, um valor gerado pela função objetivo, que pode ser normalizada (**aptidão normalizada**) para o algoritmo de seleção. Assim, a média das aptidões de uma população é representada pela sua **aptidão média**, sendo que o indivíduo que apresenta melhor adaptação é encontrado pela **aptidão máxima** desta.

Nos subtópicos a seguir, veremos com mais detalhes como esses passos retratados pelos conceitos de um AG são tratados.

## Operadores e parâmetros genéticos

Operadores genéticos são formas de diversificar uma população que mantenha as características de adaptação adquiridas em gerações anteriores. O objetivo desses operadores é que, após sucessivas gerações, a população se transforme até chegar em um resultado satisfatório (CARVALHO, *on-line*).

Um exemplo é o operador de mutação que visa introduzir e manter a diversidade genética de uma população, alterando, de forma arbitrária, uma estrutura da população para gerar novos elementos, conforme podemos visualizar na Figura 3.8:

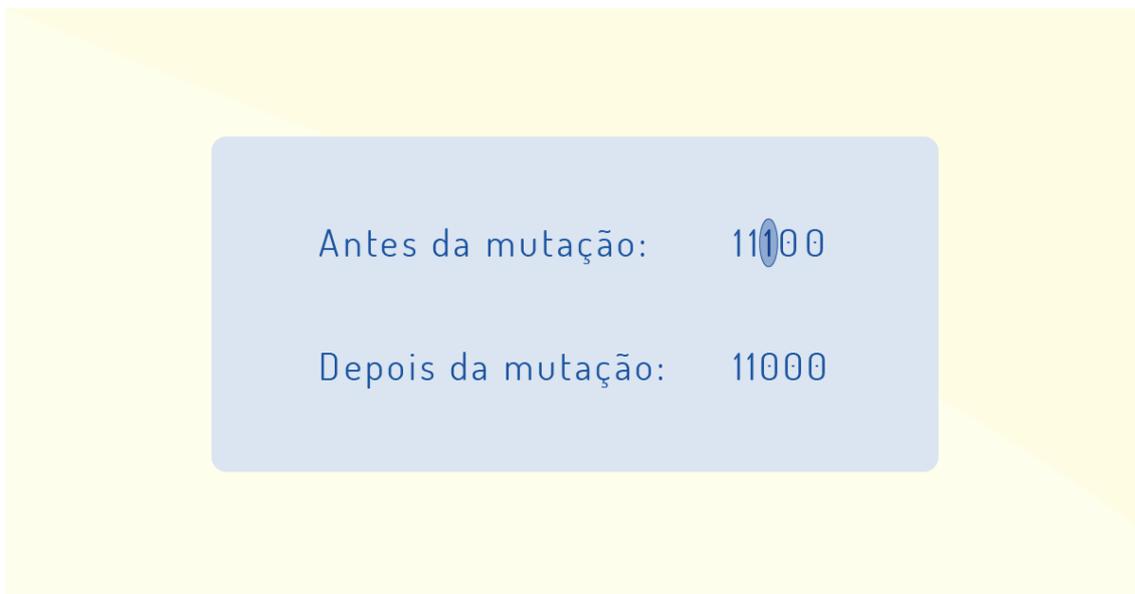


Figura 3.8 - Processo de mutação

Fonte: Adaptada de Carvalho (*on-line*).

Na Figura 3.8, um indivíduo, representado pelo número 1 (um), sofreu mutação e gerou um novo indivíduo, representado pelo número 0 (zero). Dessa forma, a mutação assegura que a probabilidade de se chegar a qualquer ponto do espaço de busca nunca será zero,

além de contornar o problema de mínimos locais, pois, com esse mecanismo, altera-se levemente a direção da busca.

Outro operador é o cruzamento (*crossover*), que faz a recombinação de características dos pais que, durante a reprodução, faz com que as gerações seguintes herdem essas características. Conforme aponta Carvalho (*on-line*), ele é um operador genético dominante e pode ser utilizado de várias maneiras, como:

1. um ponto: um ponto de cruzamento é selecionado entre as gerações, que trocarão suas características genéticas para gerar uma próxima geração.
2. multipontos: muitos pontos de cruzamento podem ser usados e há uma generalização de troca de material genético.
3. uniforme: por meio de um parâmetro global, realiza a probabilidade de cada variável poder ser trocada entre os pais.

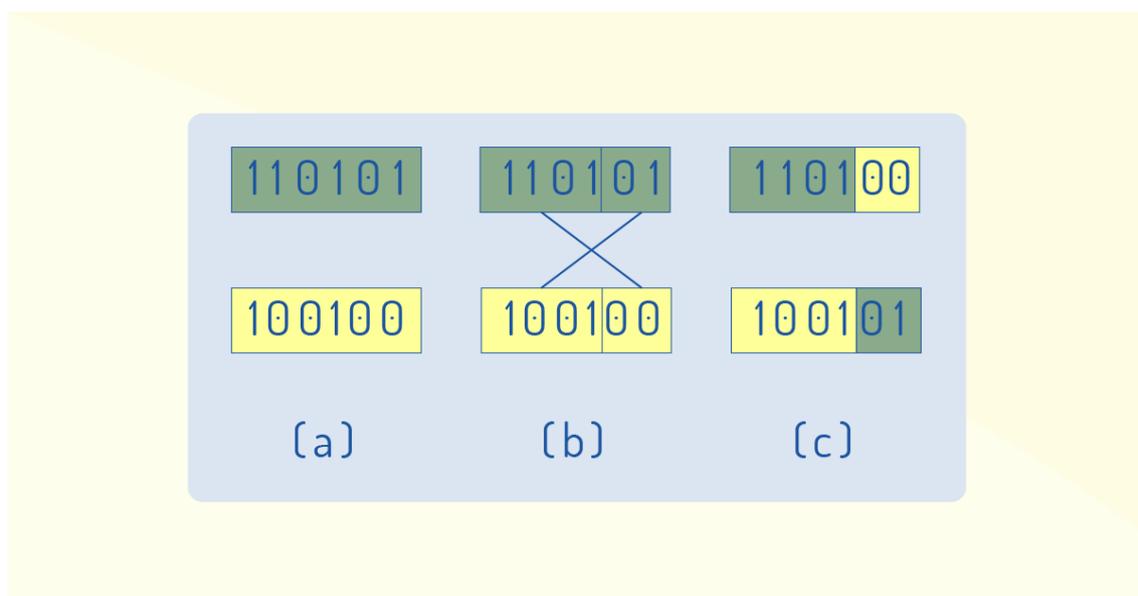


Figura 3.9 - Processo de *crossover*

Fonte: Adaptada de Carvalho (*on-line*).

A Figura 3.9 ilustra o *crossover* de um ponto em que dois indivíduos (a) foram selecionados de forma aleatória, um ponto (4 genes) de cruzamento foi selecionado (b) e foram recombinadas as características, gerando dois novos indivíduos (c).

Além dos operadores, os parâmetros também influenciam o comportamento dos AGs, que são configurados diante da necessidade do problema em questão e dos recursos que estão disponíveis para a solução do problema. Alguns desses parâmetros são (GILLIARD; PACHECO, *on-line*):

1 – tamanho da população: afeta de forma direta o desempenho global e a eficiência dos AGS. Quanto menor uma população, menor o desempenho, pois a cobertura de espaço de busca do problema é menor. Para populações maiores, isso funciona de forma inversa e também requer maiores recursos computacionais e de tempo de execução do algoritmo.

2 – taxa de cruzamento: se essa taxa for alta, novas características serão introduzidas na população, porém, é necessário cautela, pois, com a velocidade muito alta, também pode-se perder características de aptidão importantes para a população.

3 – taxa de mutação: uma baixa taxa desse tipo funciona de forma preventiva para que uma dada posição da busca não fique parada. Entretanto, se for muito alta, o espaço de busca torna-se aleatório.

4 – intervalo de geração: é importante para controlar qual parte da população será substituída.

Dessa forma, tanto os parâmetros quanto os operadores são importantes para alcançar uma população com características diversas e que mantenham aqueles mais aptos ao ambiente configurado, para se chegar à solução do problema.

### **Estrutura básica de um AG simples**

Um AG simples tem seu funcionamento de forma que, na inicialização, seja gerada uma população aleatória de  $n$  indivíduos, em que cada indivíduo representa um ponto no

espaço de possíveis soluções para o problema. Após cada iteração do algoritmo, são realizadas quatro operações básicas (GILLIARD; PACHECO, *on-line*):

1 – cálculo de aptidão: determinado pelo cálculo da função de avaliação (*fitness*) especificada no projeto. Cada indivíduo da população da inicialização é uma entrada para uma ferramenta de análise de desempenho. Após isso, são geradas saídas que fornecem a aptidão de cada indivíduo e eles são ordenados com base nos valores dessa saída.

2 – seleção: os indivíduos mais aptos serão selecionados para gerar uma nova população, por meio de *crossover*. Uma ressalva deve ser feita, pois podem ser selecionados indivíduos iguais, enquanto outros podem até desaparecer.

3 – cruzamento (*Crossover*): em seguida, de posse da primeira lista de indivíduos da fase de seleção, eles são embaralhados para formar uma segunda lista de parceiros. Cada indivíduo da primeira lista é cruzado com um da segunda lista, por meio de um ponto de corte feito de forma aleatória (conforme já foi mencionado na Figura 3.8). Um novo cromossomo (cadeia de bits) é gerado, permutando-se a metade inicial de um cromossomo com a metade final do outro, sendo essa a metodologia mais adotada.

4 – mutação: é realizada para que o algoritmo realize uma varredura maior no espaço de possíveis soluções do problema. É sorteado um indivíduo aleatoriamente e que terá seu gene modificado para compor a nova população.

Todas essas operações podem ser visualizadas no fluxo de atividades da Figura 3.10:

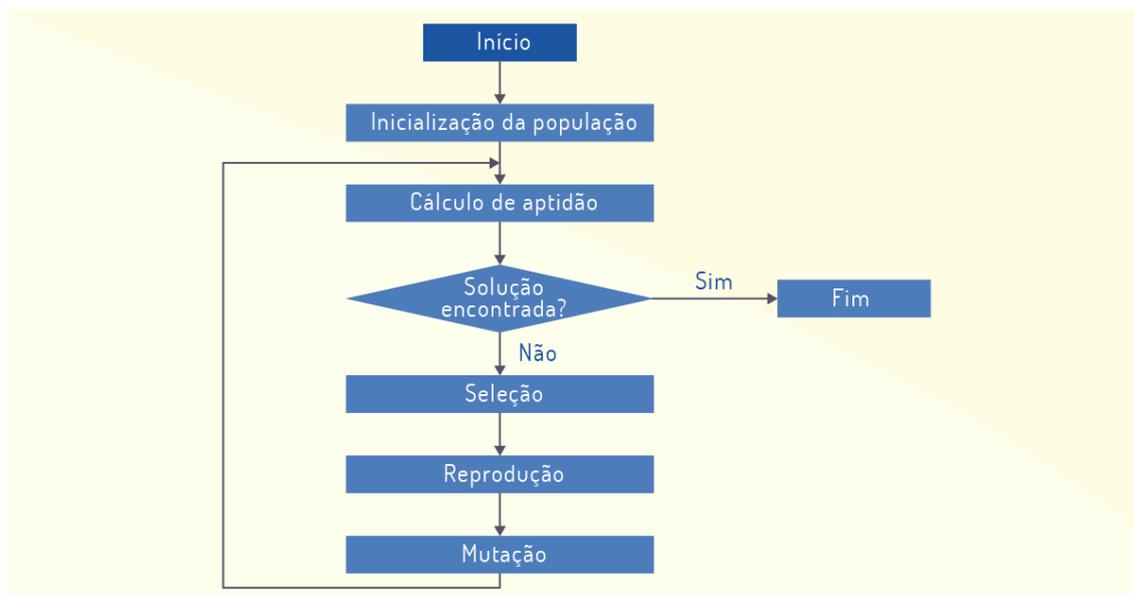


Figura 3.10 - Fluxo de operações de um AG

Fonte: Elaborada pelos autores.

Caso o algoritmo não encontre a solução para o problema, uma nova iteração é realizada a partir do cálculo de aptidão para buscar novos indivíduos considerados aptos, que, se não forem a solução para o problema, passarão pelas quatro operações básicas de um AG simples. Quando encontrarem uma estagnação na varredura da solução dos problemas, as iterações param e fornecem a solução ótima.

### Aplicações de um AG

Os AGs são utilizados em larga escala para a solução de problemas de otimização em muitas áreas. São considerados uma classe específica de algoritmos que se baseiam na evolução de Darwin e na biologia evolutiva.

Vejamos, a seguir, algumas das áreas em que os AGs podem ser implementados (MIRANDA, *on-line*).

- Circuitos analógicos: tendo como uma entrada e uma saída desejada, pode ser especificada como Tensão. O AG, então, gera uma topologia, o tipo e os valores do circuito para gerar a tensão desejada.
- Síntese de protocolos: determinação de quais funções do protocolo devem ser implementadas em *hardware* e quais devem ser implementadas em *software* para que um certo desempenho seja alcançado.
- Programação genética: gera a listagem de um programa, em uma determinada linguagem especificada, para que um determinado conjunto de dados de entrada forneça uma saída desejada.
- Ciências biológicas: modela processos biológicos para o entendimento do comportamento de estruturas genéticas.
- Problemas de otimização complexos: problemas com muitas variáveis e espaços de soluções de dimensões elevadas. Ex.: problema do caixeiro viajante, gerenciamento de carteiras de fundos de investimento.

Além desse tipo de AG, podemos descrever outros como o Genitor, que é um algoritmo que conserva os melhores pontos encontrados na população, realizando elitismo e o CHC (*Cross generational elitist selection, Heterogeneous recombination and Cataclysmic mutation*), mais utilizado atualmente, pois os N melhores indivíduos são selecionados a partir da população atual e após o cruzamento, removendo os indivíduos duplicados.

Portanto, os AGs têm inúmeras aplicações, podem ser adaptáveis para uma série de problemas e possuem elevado custo computacional. Contudo, para problemas específicos, aconselha-se o uso de algoritmos híbridos que mesclam técnicas de AG com métodos de otimização tradicionais.

## ATIVIDADE

4) Considere o trecho a seguir:

“Este estudo usa ‘algoritmos evolutivos genéticos’ para prever o crescimento urbano, mirando especificamente o distrito de Minato em Tóquio. O arquiteto Ivan Pazos, principal autor do novo estudo, explicou a ciência por trás do algoritmo: ‘Operamos dentro da computação evolutiva, um ramo da inteligência artificial e aprendizagem automática que usa as regras básicas da genética e a lógica da seleção natural de Darwin para fazer previsões’” (ABDALLAH, *on-line*).

As regras básicas de um algoritmo simples como o da matéria acima fazem uso de operações básicas da genética. Diante disso, analise as assertivas a seguir.

I. A primeira operação é a inicialização da população, que é uma amostra de indivíduos coletada de forma padronizada para descartar indivíduos que não se encaixam no problema.

II. A segunda operação é o cálculo de aptidão, que consiste em analisar os indivíduos da população inicial e fornece um ranking selecionando os melhores indivíduos.

III. A terceira operação é, após receber os indivíduos ranqueados, a seleção dos indivíduos para gerar uma nova população.

IV. A quarta operação é o cruzamento, em que é realizado um corte de forma configurada na primeira lista de indivíduos da seleção, para, assim, gerar duas listas de indivíduos que serão cruzados.

V. A quinta operação é a mutação que seleciona um indivíduo qualquer da nova geração e modifica seu gene, para que seja gerada uma nova população de indivíduos.

Estão corretas as afirmativas:

a) I, II e III, apenas.

b) I, III e IV, apenas.

c) II e III, apenas.

d) III e IV, apenas.

e) III e V, apenas.

## Redes neurais

Os computadores foram criados como formas de acelerar atividades e raciocínios humanos. As máquinas realizam tarefas de forma muito mais eficaz que um ser humano, desde que configuradas corretamente sobre a forma que devem solucionar um determinado problema (DEEP LEARNING, *on-line*).

Baseando-se nisso, muitos pesquisadores procuraram entender como funciona a inteligência, que pode ser entendida como uma série de pensamentos lógicos, ou que a inteligência é estruturada sob a forma como o cérebro é organizado. Assim, foram criados sistemas lógicos para se adequar a esse tipo de inteligência (DEEP LEARNING, *on-line*).

Foi com base nesse contexto que se procurou entender como os neurônios funcionam, já que estão relacionadas ao processo de aprendizagem de novos conhecimentos e fazem parte do principal órgão que estava sendo estudado: o cérebro humano. A partir desse estudo, foi criado o conceito das **redes neurais**. Mas o que vem a ser uma rede neural?

Existem várias definições para redes neurais (RN). Vejamos, a seguir, algumas delas (REDES..., 2019).

- São considerados sistemas computacionais que possuem nós interconectados e a capacidade de funcionar, de forma similar, como os neurônios do cérebro humano.
- São estruturas lógicas que imitam o cérebro humano, de forma metafórica.
- É uma nova forma de abordar a solução de problemas de previsão de séries temporais, sem utilizar métodos estatísticos tradicionais.
- Em computação, são modelos que se baseiam no sistema nervoso central dos humanos e realizam o aprendizado de máquina, até mesmo, como o reconhecimento de padrões.

Haykin (2001) enfatiza a visão de uma rede neural como uma máquina adaptativa, sendo um processador maciça e paralelamente distribuído, que tem a propensão natural para armazenar conhecimento experimental, por meio das forças de conexão entre células computacionais simples conhecidas como **neurônios** ou unidades de processamento.

Como pudemos perceber nas definições apresentadas, uma rede neural procura funcionar da mesma forma que o cérebro humano para resoluções de problemas, com base no conhecimento aprendido. Baseiam-se na estrutura de redes neurais biológicas de forma associada ao processamento paralelo de informações do cérebro humano.

Por fim, o neurônio artificial foi criado para se basear na modelagem de sistemas inteligentes que podem realizar várias tarefas, como classificação, reconhecimento e processamento de imagens, dentre outras. *A posteriori*, foi criado um sistema com diversos neurônios artificiais, o que foi denominado rede neural artificial (RNA).

### **Histórico das redes neurais**

Os estudos sobre redes neurais iniciaram na década de 1940, mais precisamente em 1943, quando o neurofisiologista Warren McCulloch e o matemático Walter Pitts publicaram um estudo sobre o funcionamento dos neurônios e, então, puderam modelar uma rede neural com base nos conhecimentos em circuitos elétricos. Além disso, em 1949, um outro estudo importante, de Donald Hebb, descreveu que os neurônios são fortalecidos em suas sinapses, à medida que são utilizados diversas vezes, o que é uma base para entender a maneira como os humanos aprendem (DEEP LEARNING, *on-line*).

Em 1956, um neurobiologista chamado Frank Rosenblatt começou a trabalhar no desenvolvimento de um projeto nomeado por Perceptron, cujo objetivo é calcular uma soma ponderada das entradas informadas, fazer a subtração de um limite e, a partir disso, informar um dos valores como possíveis resultados. Essa pesquisa é resultante da análise comportamental do olho de uma mosca, que, a partir das entradas informadas em seu campo de visão, subtrai algumas entradas e, então, toma a decisão para solucionar seu problema de fuga do ambiente (DEEP LEARNING, *on-line*).

Nos anos 1960, um projeto importante chamado MADALINE impulsionou os estudos nessa área. Esse foi o primeiro projeto aplicado a um problema real, que visava ao reconhecimento de padrões binários em uma chamada telefônica, em que, ao ler um bit, seria possível prever o próximo. Além disso, foi utilizado um filtro que tinha o intuito de

eliminar ecos nas linhas telefônicas. Apesar de ser antigo, esse projeto ainda é utilizado de forma comercial (DEEP LEARNING, *on-line*).

Após isso, houve um grande hiato nas pesquisas em redes neurais, por vários estudos que provocam a inviabilidade das pesquisas e os efeitos negativos associados à área. Somente nos anos 1980 é que houve um crescimento das pesquisas: em 1982, foram criados dispositivos úteis com base na modelagem do cérebro; em 1985, foi criada a reunião anual de RN; em 1987, foi realizada a primeira conferência sobre redes neurais do *Institute of Electrical and Electronic Engineers* (IEEE); em 1986, foi criado o conceito de redes de *Backpropagation*, que aprendem de forma mais lenta, porém, têm um resultado mais preciso (DEEP LEARNING, *on-line*).

Por fim, as redes neurais, a partir desse processo histórico, mais precisamente a partir dos anos 1980, ganhou muitos adeptos e, hoje em dia, são utilizadas em diversas aplicações. A essência por trás do desenvolvimento das redes neurais é que, se uma rede funcionar na natureza, ela tem uma alta probabilidade de funcionar nos computadores. Contudo, o desenvolvimento dessa área é dependente do avanço das pesquisas em *hardware* específico para uso, para, assim, desenvolverem redes neurais mais rápidas e eficientes.

### **Neurônios e funcionamento das redes neurais**

Para que possamos entender melhor as redes neurais, é necessário entender as partes que compõem um neurônio: os dendritos são as zonas receptoras que têm como função captar as informações do ambiente ou de outras células que serão processadas pelo corpo celular e, posteriormente, distribuídas para outros neurônios ou células do corpo humano pelo axônio (vide Figura 3.11) (ASSIS, 2009):

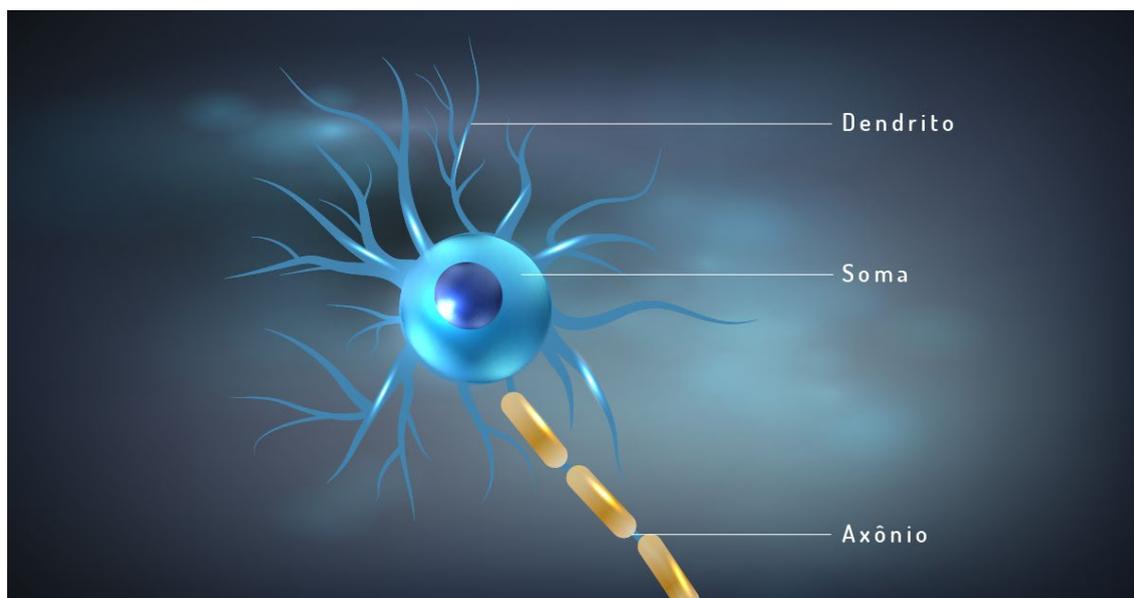


Figura 3.11 - Estrutura de um neurônio

Fonte: Adaptada de Freepik.

Cabe ressaltar que o cérebro humano tem milhões de neurônios, que, através das sinapses (ligações entre eles), conseguem resolver problemas complexos. Portanto, não são os neurônios isoladamente que possibilitam que os seres humanos realizem as mais variadas tarefas, mas, sim, as ligações estabelecidas por eles (DEEP LEARNING, *on-line*).

Assim, podemos ver que os neurônios podem ser vistos como pequenas unidades computacionais que realizam o processamento de informações provenientes do seu ambiente e se interligam com outros neurônios para a aprendizagem de um novo conhecimento. Com base nisso, foram produzidos os neurônios artificiais.

Cada neurônio possui dois ou mais receptores, que servem para a entrada de dados e percepção do tipo de sinal recebido; o corpo é formado por processadores que têm um sistema de *feedback*, que altera sua programação lógica dependendo dos dados de entrada, e a saída binária, que informa uma resposta formatada como “sim” ou “não”, dependendo do resultado que ocorreu no processamento (DEEP LEARNING, *on-line*).

Haykin (2001) identifica três elementos básicos para um modelo neuronal não linear, conforme apresentado na Figura 3.12 a seguir. Os sinais de entrada correspondem a estímulos provenientes do ambiente ou de sinapses (ligações) com outros neurônios. Cada entrada apresenta um **peso sináptico**, que será multiplicado pelo sinal entrante e que serão todos somados por meio de uma função de **junção aditiva**. A **função de ativação** é responsável pela ativação da saída de um neurônio, restringindo o intervalo permissível da amplitude do sinal de saída, sendo, geralmente, não linear e no formato sigmoideal. Esse modelo também apresenta a inclusão de um *bias*, que tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo.

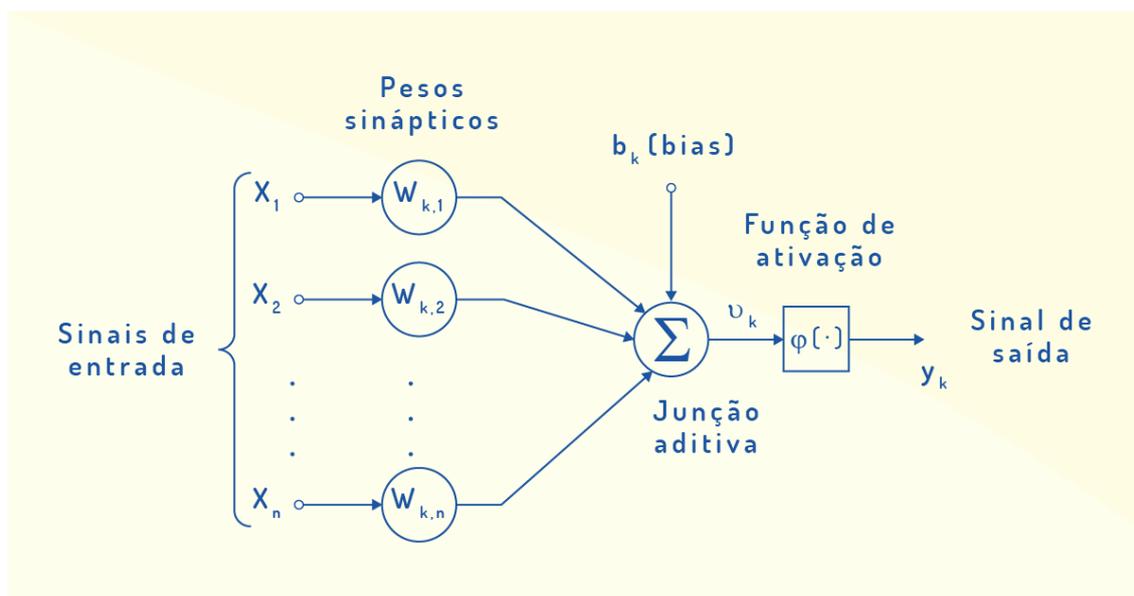


Figura 3.12 - Modelo não linear de um neurônio

Fonte: Adaptada de Haykin (2001, p. 36).

Como um neurônio pode receber várias entradas e tipos de sinais e realizar um único processamento, faz-se necessário conectar vários neurônios que são similares em rede para melhorar o desempenho do processamento de informações e, assim, produzir mais

resultados. Tendo em vista essa conexão entre neurônios, arquiteturas de redes neurais são propostas em relação à organização na forma de camadas, como a de camada única e a multicamadas.

Na arquitetura de rede de camada única, temos uma camada de entrada que se projeta sobre uma camada de saída de neurônios (nós computacionais), conforme Figura 3.13, a seguir:

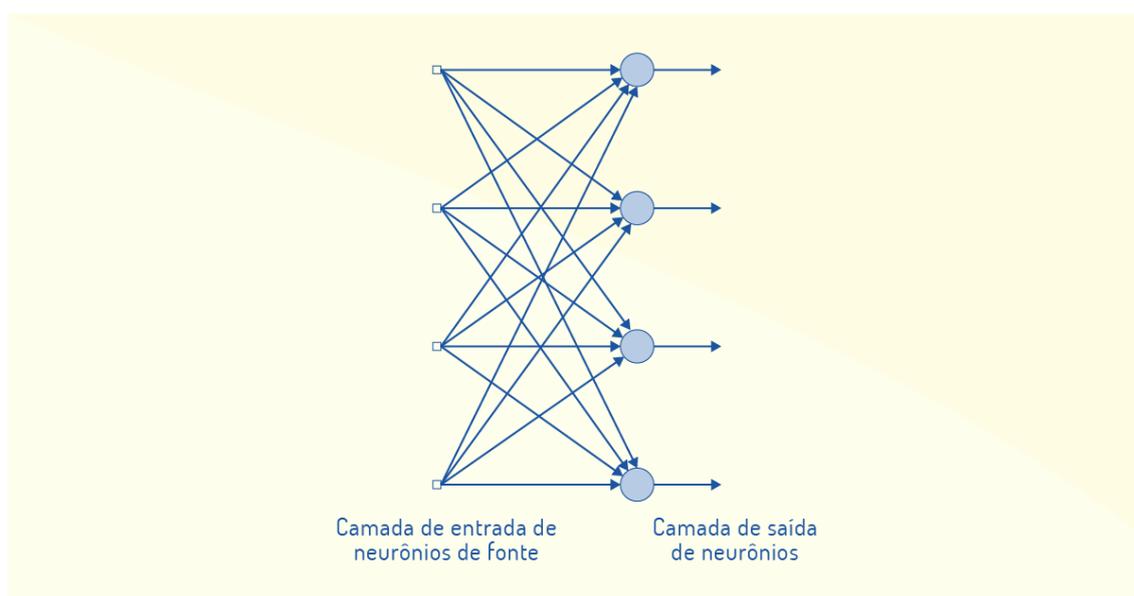


Figura 3.13 - Rede alimentada adiante com uma única camada

Fonte: Adaptada de Haykin (2001, p. 47).

Essa rede se comporta como uma rede alimentada adiante (*feed forward*) ou acíclica, não permitindo *loops* na rede, como no caso das redes recorrentes, que apresentam laços de realimentação. As redes alimentadas adiante também são aplicadas na arquitetura de redes multicamadas, em que há camadas intermediárias de neurônios para processamento entre as camadas de entrada e saída. Essas camadas intermediárias são conhecidas de camadas ocultas e ajudam na extração de estatísticas de ordem elevada. Na Figura 3.14, temos uma ilustração de rede multicamada:

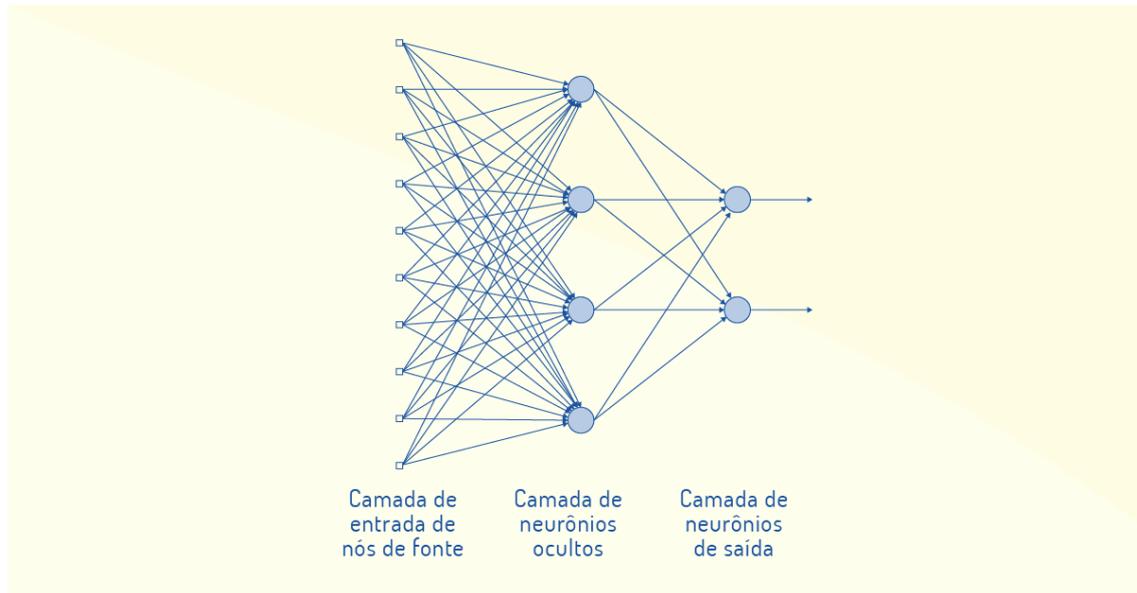


Figura 3.14 - Rede alimentada adiante, totalmente conectada com uma camada oculta e uma camada de saída

Fonte: Haykin (2001, p. 48).

Os elos de comunicação entre os nós (neurônios) de diferentes camadas são chamados de conexões sinápticas e podem ser refletidos em uma rede neural totalmente conectada (todos os nós de uma camada ligados a todos os nós da camada adjacente) ou parcialmente conectada (algumas conexões faltando).

Vamos a um exemplo para entender melhor o funcionamento dos neurônios em uma rede neural: pode ser criado um sistema que visa identificar e diferenciar bananas de maçãs. Sendo assim, os neurônios serão sensíveis à cor e à forma, totalizando quatro: dois farão a identificação das cores amarelo e vermelho e, os outros, farão a percepção da forma (redondo ou comprido). Assim, cada neurônio poderá receber quatro entradas, mas, de acordo com sua configuração, ele só receberá a entrada para a qual foi configurado (ASSIS, 2009).

Por conseguinte, para obter um melhor desempenho desse sistema, é criada uma rede em camadas. Na primeira camada serão adicionados quatro neurônios, que recebem os estímulos do ambiente (cor amarela, vermelha, forma redonda ou comprida). Uma

segunda camada (oculta) com três neurônios vai realizar o processamento dessas informações de entrada, enquanto a terceira camada vai informar se é maçã ou banana, por meio de dois neurônios (ASSIS, 2009). Veja a Figura 3.15:

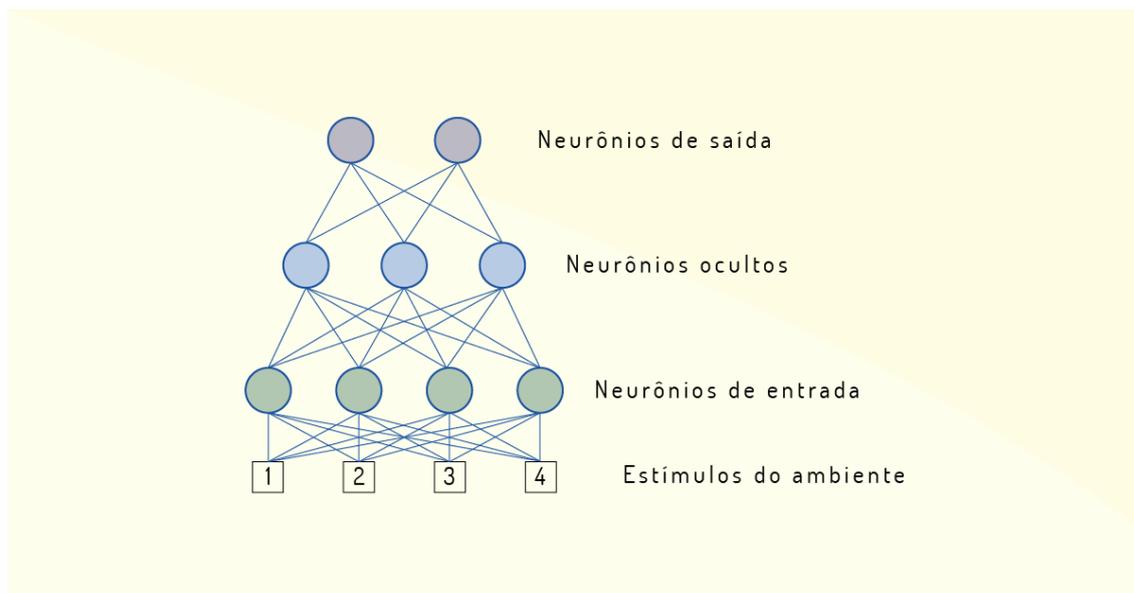


Figura 3.15 - Exemplo de rede em camadas

Fonte: Elaborada pelos autores.

O modelo mais simples de uso de uma rede de camada única ou multicamadas é o *perceptron*, que consiste em um único neurônio na camada de saída, utilizado para a classificação de padrões ditos linearmente separáveis (HAYKIN, 2001). Na Figura 3.16, temos um exemplo de rede *perceptron* de múltiplas camadas (também conhecida como MLP - *Multi Layer Perceptron*).

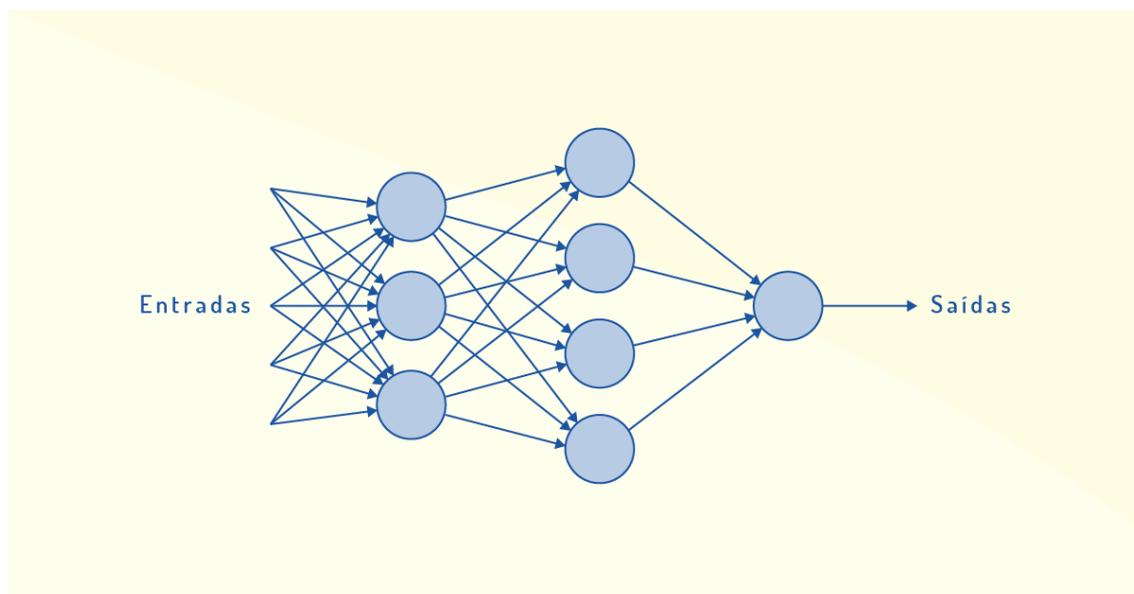


Figura 3.16 - Exemplo de rede *perceptron*

Fonte: Adaptada de Deep Learning (*on-line*).

Nesse modelo, a primeira camada toma decisões simples, ponderando as evidências de entradas. Na segunda camada, novas decisões são tomadas, ponderando os resultados da primeira camada, em um nível mais complexo. E a complexidade da decisão aumenta à medida que outras camadas são utilizadas, até que seja atingida a camada de saída. Para tanto, os neurônios fazem uso de filtros adaptativos e algoritmo do mínimo quadrado médio (LMS).

O importante da rede neural é a forma como elas processam as informações; elas não executam programas, mas aprendem. Mas como isso é realizado? A rede neural de entrada testa várias vezes se o valor recebido de entrada é igual ao valor para o qual ela está configurada. Quando ela acerta, ela recebe uma pontuação, e a rede estabelecida entre ela é reforçada como o melhor caminho. Com os erros, os neurônios perdem pontos. Assim, o sistema sempre segue o caminho com maior pontuação e, quanto mais tentativas forem realizadas, mais aprimorado o sistema vai se tornando, passando a executar as tarefas quase sem erro (DEEP LEARNING, *on-line*).

## FIQUE POR DENTRO

O processo de aprendizagem é uma etapa fundamental para as redes neurais. Para tanto, podem ser adotadas aprendizagens supervisionadas, não supervisionadas, com ciclos iterativos de ajustes aplicados aos pesos sinápticos e boas para a melhoria do desempenho da rede.

Você pode entender melhor sobre isso no artigo *Redes neurais artificiais*, de Thomas Walter Rauber, disponível em: [https://www.researchgate.net/profile/Thomas\\_Rauber/publication/228686464\\_Redес\\_nеurais\\_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf](https://www.researchgate.net/profile/Thomas_Rauber/publication/228686464_Redес_nеurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf).

Acesso em: 19 fev. 2020.

Portanto, uma rede neural funciona de forma diferente da lógica tradicional: ela aprende o que precisa ser realizado e executa essa função. Assim, uma rede capacitada possui os neurônios necessários para realizar uma tarefa de forma muito mais eficiente.

## Aplicações

As redes neurais são utilizadas em diversas aplicações e podemos, então, dividir as redes em alguns temas macro, como: processamento de sinais, imagem e visão; categorização e controle de sistemas; identificação e classificação de padrões. Vale ressaltar que tais temas podem se comportar de diferentes formas, dependendo de como a rede neural está aprendendo.

Podemos citar algumas aplicações que utilizam redes neurais, como: reconhecimento de voz, identificação de e-mails como spam, computação em nuvem (*cloud computing*), mercado financeiro, agricultura, medicina, entre outros. São diversas áreas que realizam o uso de RN (DEEP LEARNING, *on-line*).

No reconhecimento de voz, por exemplo, a rede neural é utilizada para aprender a conhecer a voz de pessoas específicas; na identificação de e-mails como spams, a rede neural consegue identificar o que é um spam e os apaga, tendo uma boa margem de acertos. Um outro exemplo é a utilização de *softwares* OCR (*Optical Character Recognition*), utilizados em scanners que precisam aprender a reconhecer caracteres de imagem e utilizam a rede neural para isso. Algumas redes neurais são utilizadas em robôs para realizar o desarmamento de bombas (REDES..., 2019).

Um termo relacionado às redes neurais ganhou popularidade a partir dos anos 2000: aprendizagem profunda (*deep learning*), que é um tipo de aprendizado de máquina que é capaz de compreender, reproduzir e até melhorar comportamentos e padrões configurados previamente para aplicar de forma automática a mais dados, criando modelos que funcionam em tempo real e são executados em computadores, formando gigantescas redes neurais artificiais que se assemelham ao cérebro humano (REDES..., 2019).

Algumas empresas vêm utilizando esse algoritmo de aprendizagem profunda, como: a Google, que é capaz de identificar gatos nas fotos armazenadas por seus usuários; o Facebook, que criou o *software* DeepFace, que objetiva identificar e marcar os usuários de forma automática nas fotos armazenadas na rede social. A aprendizagem profunda (*deep learning*) é muito utilizada para reconhecimento de imagens, como podemos perceber nos exemplos citados (REDES..., 2019).

Por fim, o crescente processo evolutivo da inteligência artificial corrobora para o crescimento das redes neurais, em especial das classificadas como aprendizagem profunda. Dessa forma, é possível criar redes inteligentes autônomas que podem impactar diversos ramos da sociedade, a partir do uso de modelos matemáticos agregando padrões biológicos em IAs.

## ATIVIDADE

5) A Amazon Alexa, uma assistente virtual inteligente que conversa com o usuário para realizar suas tarefas diárias, e o Amazon Polly, um serviço que realiza a transformação de texto em falas reais, são produzidos pela empresa Amazon, a qual informou que tais serviços já estão realizando a conversão do texto em falas utilizando redes neurais em camadas.

Sobre as redes neurais em camadas, pode-se afirmar que:

- a) a rede em camadas possui somente duas: a de entrada e a de saída. O processamento é realizado por uma IA externa.
- b) na rede em camadas, a primeira etapa é receber os estímulos do ambiente, que só podem ser aqueles que são pré-configurados na rede.
- c) em uma rede em camadas, a camada de neurônios de saída emite informações vagas e imprecisas.
- d) as camadas de uma rede neural são: neurônios de entrada, processamento e saída, que visam responder o problema de forma binária.
- e) em uma rede em camadas, a camada de entrada recebe os dados advindos do estímulo de forma padronizada.

## Sistemas especialistas

Antes de iniciarmos o entendimento sobre esse tipo de sistema, é necessário definir o que é um especialista: é uma pessoa ou organização que detém o conhecimento, experiência e a habilidade de aplicar esse conhecimento para fornecer conselhos e solucionar problemas (LAUDON; LAUDON, 2015).

Partindo disso, podemos definir que um sistema especialista utiliza o conhecimento humano para solucionar problemas que necessitam da presença de um especialista. Esse tipo de sistema é uma categoria dos denominados sistemas baseados em conhecimento, já que a ferramenta principal para esse tipo de sistema é o conhecimento, e eles se propõem a solucionar problemas (O'BRIEN; MARAKAS, 2008).

Outra definição para esse tipo de sistema pode ser entendida como um programa de computador que usa tecnologias de inteligência artificial para simular o julgamento e o comportamento de um ser humano ou de uma organização que possui conhecimento especializado e experiência em um campo específico (LAUDON; LAUDON, 2015).

Normalmente, um sistema especialista incorpora uma base de conhecimento contendo experiência acumulada e um mecanismo de inferência ou regras – um conjunto de regras para aplicar a base de conhecimento a cada situação específica descrita no programa. Os recursos do sistema podem ser aprimorados com acréscimos à base de conhecimento ou ao conjunto de regras. Os sistemas atuais podem incluir recursos de aprendizado de máquina que lhes permitem melhorar seu desempenho com base na experiência, assim como os humanos (O'BRIEN; MARAKAS, 2008).

Em suma, uma base de conhecimento é uma coleção organizada de fatos sobre o domínio do sistema. Um mecanismo de inferência interpreta e avalia os fatos na base de conhecimento para fornecer uma resposta. Tarefas típicas para sistemas especialistas envolvem classificação, diagnóstico, monitoramento, projeto, programação e planejamento para empreendimentos especializados.

Fatos para uma base de conhecimento devem ser adquiridos de especialistas humanos por meio de entrevistas e observações. Esse conhecimento é geralmente representado na

forma de regras “se-então” (regras de produção): “Se alguma condição for verdadeira, então a seguinte inferência pode ser feita (ou alguma ação tomada)”.

A base de conhecimento de um grande sistema especialista inclui milhares de regras. Um fator de probabilidade é frequentemente associado à conclusão de cada regra de produção e à recomendação final, porque a conclusão não é uma certeza. Por exemplo, um sistema para o diagnóstico de doenças oculares pode indicar, com base em informações fornecidas a ele, uma probabilidade de 90% de que uma pessoa tenha glaucoma e pode listar conclusões com probabilidades menores. Um sistema especialista pode exibir a sequência de regras por meio das quais chegou à sua conclusão. O rastreamento desse fluxo ajuda o usuário a avaliar a credibilidade de sua recomendação e é útil como uma ferramenta de aprendizado para os alunos (O'BRIEN; MARAKAS, 2008).

Além disso, sistemas especialistas apoiados por humanos frequentemente empregam regras heurísticas, ou “regras práticas”, além de simples regras de produção, como aquelas obtidas de manuais de engenharia. Assim, um gerente de crédito pode saber que um candidato com um histórico de crédito ruim, mas um registro limpo desde a aquisição de um novo emprego, pode, na verdade, ser um bom risco de crédito. Os sistemas especialistas incorporaram essas regras heurísticas e cada vez mais têm a capacidade de aprender com a experiência. Os sistemas especialistas continuam sendo ajudantes – em vez de substitutos – de especialistas humanos (LAUDON; LAUDON, 2015).

Vale ressaltar que o conceito de sistemas especialistas foi desenvolvido pela primeira vez na década de 1970, por Edward Feigenbaum, professor e fundador do Laboratório de Sistemas de Conhecimento da Universidade de Stanford. Feigenbaum explicou que o mundo estava passando do processamento de dados para o "processamento de conhecimento", uma transição que estava sendo possibilitada pela nova tecnologia de processador e pelas arquiteturas de computadores (LAUDON; LAUDON, 2015).

Sistemas especialistas têm desempenhado um papel importante em muitos setores, incluindo serviços financeiros, telecomunicações, saúde, atendimento ao cliente, transporte, videogames, fabricação, aviação e comunicação escrita. Dois sistemas especializados iniciais abriram espaço no setor da saúde para diagnósticos médicos:

Dendral, que ajudou químicos a identificar moléculas orgânicas, e MYCIN, que ajudou a identificar bactérias como bacteremia e meningite e a recomendar antibióticos e dosagens.

Um sistema especialista desenvolvido mais recentemente, o ROSS, é um advogado artificialmente inteligente baseado no sistema de computação cognitiva Watson da IBM. O ROSS conta com sistemas de autoaprendizagem que usam mineração de dados, reconhecimento de padrões, aprendizado profundo e processamento de linguagem natural para imitar a maneira como o cérebro humano funciona (ROSS INTELLIGENCE, 2019).

Agora, vejamos a utilidade dos sistemas especialistas (LAUDON; LAUDON, 2015).

- É um repositório de conhecimento que será repassado para outros profissionais da área em que o sistema está tratando, com o intuito de capacitar as pessoas que não são especialistas nessa área.
- Servem como apoio aos especialistas no tratamento de alguma informação específica da área.
- Auxilia a tomar decisões mais consistentes, uma vez que sejam inseridas no sistema especialista as mesmas condições para fornecer um conjunto de decisões coeso.
- Auxiliam soluções que têm o processo demorado para conduzir a soluções de forma rápida.

Podemos ressaltar, ainda, que os sistemas especialistas atuam em várias aplicações para prover maior agilidade e fornecer soluções produtivas para as empresas. Podemos citar que eles podem atuar em algumas categorias de aplicações das organizações, como no gerenciamento de decisões, monitoração e/ou controle de processos da empresa, diagnóstico e/ou problemas operacionais relacionados a uma tarefa em específico, entre outras.

Além disso, podemos dizer que os sistemas especialistas podem atuar nos sistemas relacionados à contabilidade, utilizados em fluxos de caixa, contas a receber e a pagar, entre outras áreas relacionadas à contabilidade; no planejamento de recursos de capital quando auxilia nas decisões de investimentos da organização; na produção em que as

tarefas repetitivas podem ser realizadas por robôs; na gestão financeira em que podem ser utilizados na análise de investimentos, entre outros (O'BRIEN; MARAKAS, 2008).

No entanto, segundo O'Brien e Marakas (2008), algumas limitações dos sistemas especialistas podem ser elencadas, como:

1. alto custo no desenvolvimento e manutenção do sistema.
2. alterações no *know-how* dos especialistas para se atualizarem sobre os avanços em suas especialidades.
3. algumas limitações em relação ao acesso do conhecimento do sistema especialista, já que são limitados a regras definidas.
4. a medição do desempenho de Sistemas Especialistas é muito difícil, porque não se sabe quantificar o uso de conhecimento.
5. constituem-se ineficiente nas tomadas de decisões gerenciais de caráter subjetivo.
6. têm pouca utilização e poucos casos de teste.
7. são considerados difíceis para utilização pelo usuário. Em alguns casos, torna-se necessária a assistência de um computador pessoal ou pessoas treinadas para orientar o usuário.
8. são passíveis de erro, pois lidam com conhecimento de base humana. Se as informações não forem corretas ou estiverem incompletas, o sistema será afetado negativamente.

Portanto, podemos dizer que os sistemas especialistas são muito eficazes nas empresas e a sua utilização promove resultados positivos nas organizações. Vale ressaltar que é um tipo de tecnologia que fornece para as empresas maior agilidade, retorno eficaz na gestão, qualidade do serviço, soluções rápidas para os problemas organizacionais, aumento da produtividade, bem como mantém as empresas com competitividade no mercado.

## **ATIVIDADE**

6) Sistemas especialistas são programas de computadores que utilizam conceitos de Inteligência Artificial utilizado para a solução de problemas complexos da mesma forma que um especialista. Sobre o especialista, é correto afirmar que:

a) é uma pessoa ou organização que terá a formação de especialista para poder utilizar o sistema especialista.

b) é uma pessoa ou organização que detém o conhecimento, experiência e a habilidade de aplicar esse conhecimento para fornecer conselhos e solucionar problemas.

c) é um sistema de informação que utiliza dados e informações processadas para gerar conhecimento.

d) é um sistema ou organização que, apoiado na abordagem da inteligência artificial, gera o conhecimento para que uma pessoa o analise e, assim, gere informação relevante.

e) é uma pessoa que detém poucos conhecimentos na área e será especialista após utilizar esse tipo de sistema.

## **INDICAÇÃO DE LEITURA**

Título: Inteligência Artificial

Autores: Peter Norvig e Stuart Russel

Editora: Elsevier

ISBN: 9788535237016

Comentário: Esse é um livro de referência e foi o primeiro a abordar, de forma didática, os principais conceitos inerentes à Inteligência Artificial. Com ele, você aprenderá sobre o histórico de IA, suas principais aplicações e sua evolução em diferentes áreas de pesquisa.

UNIDADE IV

# Ciência de Dados e Aprendizado de Máquina

*Aline Chagas Rodrigues Marques*

*Marcelo Takashi Uemura*

## Introdução

Olá, aluno(a). Em 2017, uma das notícias que mais chocou os jogadores de Go, um jogo de tabuleiro, foi que uma máquina venceu por 2 x 0 o melhor jogador de Go, Ke Jie. Foi algo impressionante naquela época. E quem seria a máquina? A máquina foi denominada como AlphaGo, uma inteligência artificial (IA) desenvolvida pela Empresa Google no departamento chamado Google DeepMind. Demis Hassabis começou a criação dessa IA em 2014 e disse que o desfecho do jogo era algo muito imprevisível, já que a máquina operaria em sua capacidade máxima.

Ke tinha apenas 19 anos e se surpreendeu com a estratégia adotada pela AlphaGo, a qual informou que Ke jogou perfeitamente em seus 50 primeiros movimentos no jogo de tabuleiro. A disputa foi realizada no leste da China, e Ke foi o jogador mais próximo em comparação com as habilidades da IA. Para entender como a AlphaGo ganhou esse jogo, foi utilizando os conceitos de inteligência artificial, *machine learning* e *deep learning*, sendo estes dois últimos os objetos de estudo desta unidade! Vale ressaltar que as três áreas estão correlacionadas, mas não podem ser definidas como iguais.

*Machine learning* refere-se a como a máquina será treinada a partir da análise de grande quantidade de dados, utilizando os conceitos de ciência de dados e com a aplicação de algoritmos que permitem que a máquina tenha a habilidade de aprender a tarefa pela qual ela foi programada. Já *deep learning* permite que suas redes neurais artificiais calibrem mais suas respostas e precisão em seus acertos. Dessa forma, o AlphaGo da Google, após jogar contra si por várias vezes, aprendeu como funcionavam as regras do jogo e, assim, treinou sua partida, calibrando sua rede neural.



Fonte: ahmedgad / Pixabay.

## Ciência de dados - Definição

Ciência de dados, também conhecida como *Data Science*, é uma área emergente na computação de forma multidisciplinar, pois abrange várias áreas e possui um extenso *modus operandi* para realizar análise futura ou passada sobre os dados, que são disponibilizados na internet em uma grandeza de forma exponencial (PROVOST, 2016).

Em outra definição, podemos afirmar que é a habilidade em realizar a obtenção, compreensão, exploração, processamento, extração de valor e visualização dos dados, de diferentes segmentos do mercado (AMARAL, 2016).

Adicionalmente, a ciência de dados está envolvida na combinação de diversas técnicas de visualização e estrutura de armazenamento de dados para se chegar às conclusões dessas extensas quantidades de dados e, assim, poder aplicar esse conhecimento em situações reais (COELHO, 2019). Vejamos as principais técnicas, na Figura 4.1:



Figura 4.1 - Principais áreas englobadas pela ciência de dados

Fonte: Elaborada pelos autores.

- **Visualização:** está relacionada à representação gráfica de informações e dados com o objetivo de analisar grandes quantidades de dados e facilitar a tomada de decisão com base nos dados analisados. Para melhor compreensão, um exemplo seria quando um cientista de dados projeta visualizações de dados para um público menos técnico, talvez a fim de ajudar os membros desse público a tomar melhores decisões de negócio ou buscar ações mais assertivas. Na Figura 4.2, temos um exemplo de visualização de dados relacionados à renda *per capita* e expectativa de vida em 2007.

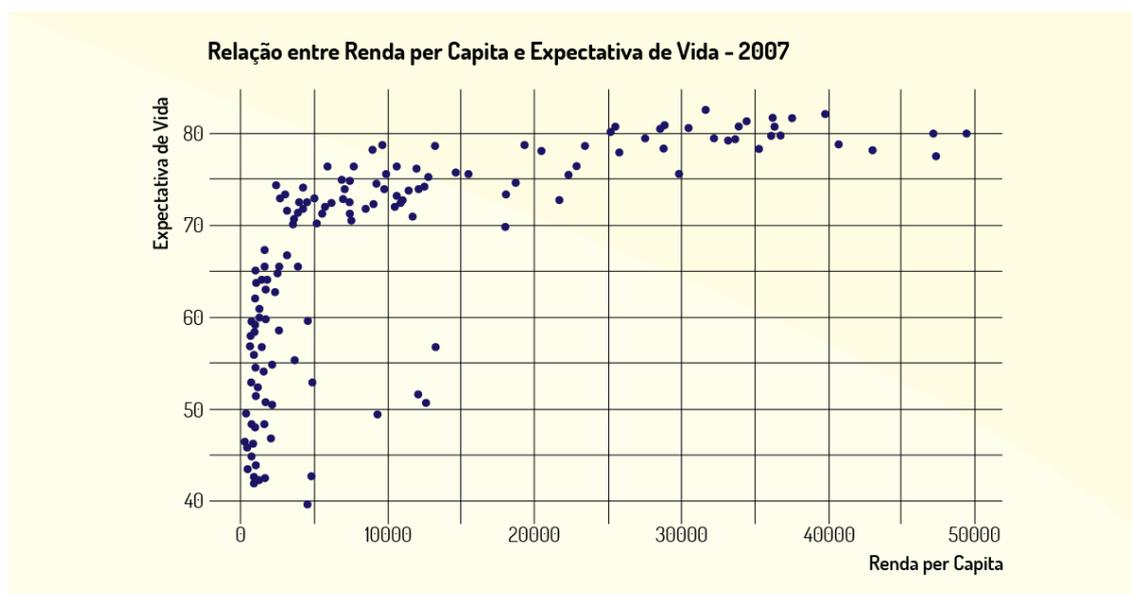


Figura 4.2 - Exemplo de visualização de dados

Fonte: Oliveira, Guerra e McDonnell (2018, p. 108).

- **Mineração de dados:** é o processo para buscar padrões e correlações entre os dados, com base em um grande conjunto de dados para realizar previsões dos resultados.
- **Processamento de dados:** é a sequência de operações executadas para converter dados brutos em forma utilizável (automática ou manualmente).
- **Neurocomputação (*Neurocomputing*):** é um processo de computação de dados

usando redes neurais artificiais, com a simulação do cérebro humano (Inteligência Artificial), muito utilizado em *deep learning*.

- **Aprendizado de máquina (*Machine Learning*):** é definido como um método de análise de dados que faz com que a máquina aprenda continuamente com os dados e faça previsões.
- **Reconhecimento de padrões:** é o reconhecimento automatizado de padrões, que são unidades de informação que se repetem, e regularidades nos dados. Não se deve confundir com mineração de dados, pois esta é um processo que utiliza algoritmos de reconhecimento de padrões para análises de grandes volumes de dados.
- **Estatística:** esta área realiza a coleta e a classificação de dados, com o objetivo de realizar deduções de proporções sobre o total do conjunto de dados com base em uma pequena amostra, trabalhando em conjunto com as outras técnicas de ciência de dados para obtenção de melhores resultados.

Sendo assim, cada uma dessas áreas visa oferecer diversas ferramentas para realizar toda uma análise dos dados para efetuar *insights* (visões) sobre o futuro dos dados processados e poder resolver problemas do mundo real e auxiliar na tomada de decisão das empresas, a depender da situação a ser analisada, podendo ser utilizadas em conjunto para se chegar à solução.

Por conseguinte, podemos ressaltar como exemplo prático o uso da área de ciência de dados no Design, no que se refere ao uso da análise dos dados coletados no mercado para apoiar à concepção de um novo produto para o cliente, desde a especificação com os aspectos do produto, definição da arquitetura, distribuição do produto e finalizando com a experiência final do usuário, maximizando as chances de sucesso (COELHO, 2019).

Podemos dizer que, apesar de a ciência de dados ter seu reconhecimento na área da computação, ela é considerada multidisciplinar, perpassando por várias áreas de conhecimento, como podemos visualizar na Figura 4.3:

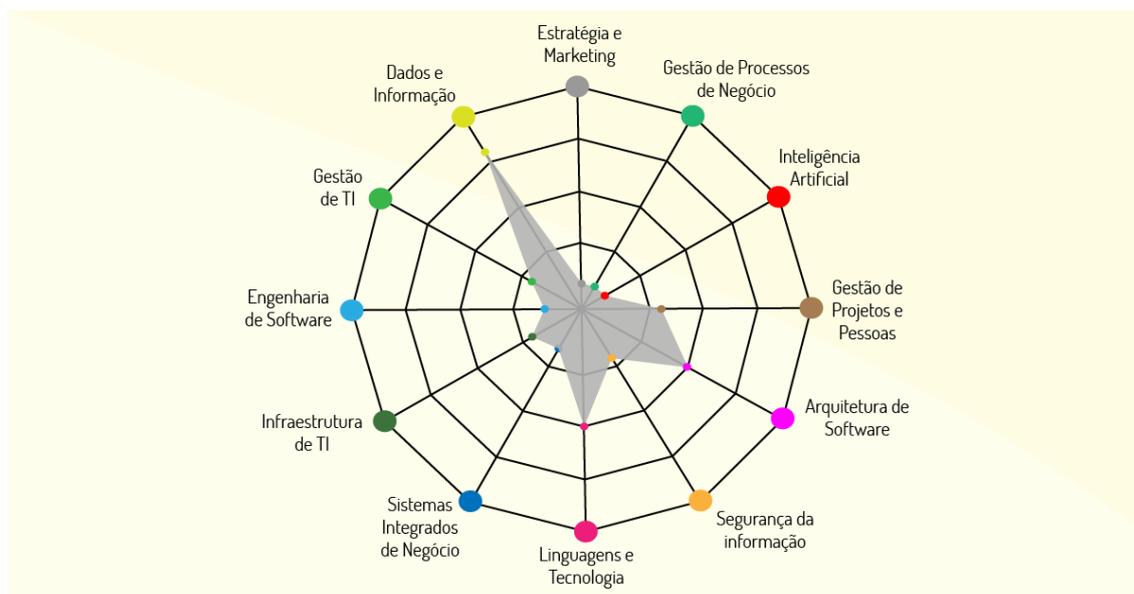


Figura 4.3 - Multidisciplinaridade de ciência de dados

Fonte: Adaptada de Portal GSTI (*on-line*).

Todas essas áreas contribuem para que a ciência de dados possa ter sua ampla divulgação dentro do mundo de análise de dados. Cada qual com sua técnica, pode auxiliar as empresas na melhor classificação e organização de seus dados e, assim, se tornar mais competitiva dentro do mercado.

Outras definições sobre a ciência de dados podem ser (PORTAL GSTI, *on-line*):

- 1 – É um processo para facilitar a tomada de decisão das empresas.
- 2 – É uma abordagem para unificar áreas como estatística e análise de dados.
- 3 – É o processamento e utilização de ferramentas para poder extrair informações a partir dos dados obtidos junto à organização.
- 4 – É uma ciência que trabalha com dados estruturados, que são dados coletados a partir de bancos de dados como aqueles utilizados por *softwares* de ERPs (*Enterprise Resource Planning*) ou dados não estruturados; são todos aqueles que podem ter diversas origens para a coleta de dados, como áudio, imagem, vídeo, entre outros.

Por fim, a ciência de dados visa desenvolver novas ferramentas e técnicas para análise de dados em função do extenso volume de dados que se tem na internet, posto que os métodos tradicionais acabaram ficando ultrapassados para analisar essas grandes quantidades de dados.

E qual o nome do papel de quem atua em Ciência de Dados? Dentro dessa área, tem-se o cientista de dados, que é o responsável por tratar dados (estruturados ou não), buscando transformá-los em oportunidades de negócio. Esse profissional precisa ter sólidos conhecimentos em Ciência da Computação, Modelagem Preditiva, Estatística, Matemática e Programação (AZAM, 2014).

Dentro da área de ciência de dados, é usual encontrarmos matemáticos, economistas, analistas de sistemas, estatísticos, engenheiros e profissionais de marketing, que estão sempre em busca de um produto perfeito; esses profissionais se utilizam da ciência para realizar a obtenção de dados valiosos.

Há uma série de habilidades necessárias para o cientista de dados, como: ser ágil na resolução de problemas, ter uma certa curiosidade para explorar os dados minuciosamente e padrões que não foram pensados para a solução de determinado problema; ter um bom raciocínio lógico e possuir uma boa visão de negócios, empreendedorismo e marketing. Tudo isso corrobora para que o cientista de dados seja versátil em um conjunto de características necessárias para a área (AZAM, 2014).

As funções relacionadas a esse profissional são: realizar análises quantitativas apoiando-se na modelagem matemática e análise estatística, transformar dados brutos para que sejam visualizados por todo o negócio, e até mesmo criar tendências aplicáveis na solução do problema e possuir uma visão profunda sobre o negócio da empresa e entender sobre gestão e tecnologia (BAUDISCH, 2016).

Conforme vimos, a ciência de dados, apesar de ser uma profissão recente, está envolvida com o conhecimento de diversas áreas e tem a responsabilidade de colocar em prática todo o processo de ciência de dados para a resolução de problemas. Alguns consideram o cientista de dados como um “pescador de dados”, já que busca os dados certos para a solução dos problemas.

Portanto, a área da ciência de dados está buscando descobrir conhecimento que tenha valor para o negócio, utilizando, para isso, a aplicabilidade de várias áreas, desde tecnologia à análise estatística. Para que ela possa ter uma maior efetividade, é necessária uma grande quantidade de dados, o que implica a busca de tecnologias que possam trazer novas fontes de dados, como a Internet das Coisas, e aumentar o desempenho de análise com *Big Data*. A ciência de dados objetiva não somente interpretar números, mas fazer previsões sobre eles, para que possam auxiliar na tomada de decisões. Para isso, tem-se o papel do cientista de dados, que faz a análise dos dados em massa, quase que em tempo real, para apoiar as empresas na geração de valor para o negócio.

## **REFLITA**

O cientista de dados precisa ter uma série de conhecimentos em várias áreas, da estatística até a ciência da computação. Porém o mercado está com dificuldades de encontrar profissionais habilitados em ciência de dados. Você acha que a área precisaria segregar melhor seus campos de conhecimento? De que forma se pode melhorar a oferta de profissionais qualificados em ciência de dados?

## ATIVIDADE

1) Um dos principais papéis em ciência de dados é o do cientista de dados, o qual possui um conjunto de características importantes para realizar a análise de extensas quantidades de dados nessa área. Além disso, é correto afirmar que a função principal de um cientista de dados é:

- a) ser ágil na resolução dos algoritmos de ciência de dados.
- b) encontrar dados que podem ser relevantes ou não para os negócios.
- c) é considerado um “pescador de dados”, que busca os dados certos para a solução do problema proposto.
- d) propor-se a criar redes neurais, como recurso trivial para buscar o conhecimento para a solução do problema proposto.
- e) implantar soluções para aprendizado de máquina.

## Ciência de dados - Aspectos relevantes

Vimos que a ciência de dados é composta por diversas áreas de conhecimento, produzindo um produto que objetiva direcionar os usuários e/ou organizações em suas tomadas de decisões, por meio de uma série de etapas da análise de dados.

Para estruturar melhor as etapas para se realizar a análise de dados, foi criado um fluxo de atividades, conforme demonstrado na Figura 4.4:

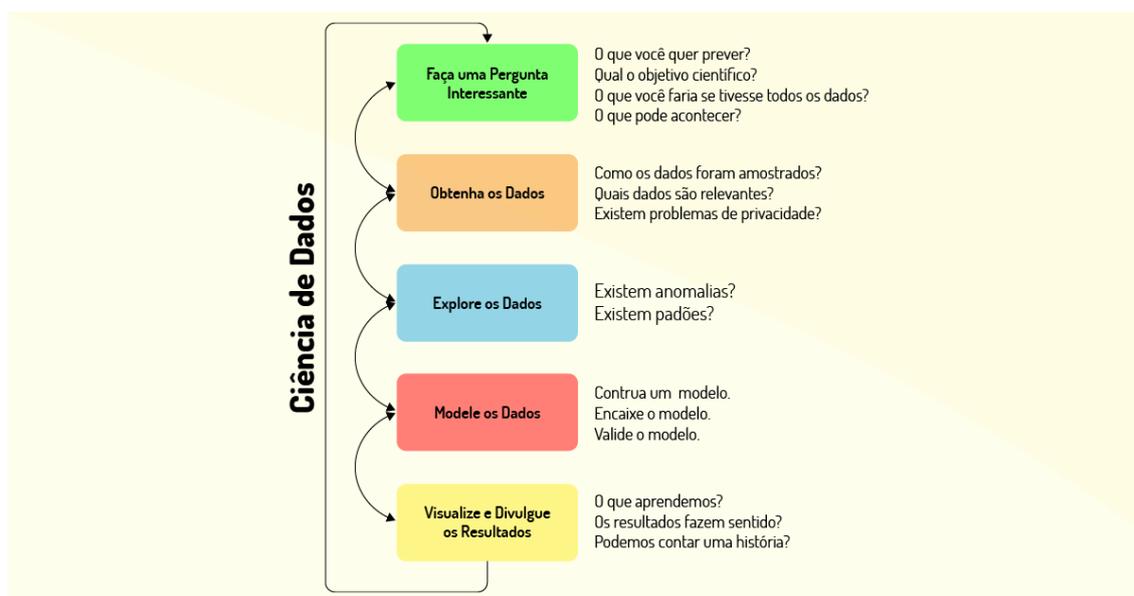


Figura 4.4 - Fluxo de atividades de ciência de dados

Fonte: Adaptado de Wickham e Grolemond (2017, p. 135).

Agora, vamos explicar um pouco sobre o fluxo de atividades:

- Na etapa 1 (*Faça uma Pergunta Interessante*), é recomendado realizar uma pergunta interessante sobre o tema dos dados pesquisados. Algumas perguntas podem ser feitas como: O que você deseja prever? Qual o objetivo científico? O que pode ocorrer?

- Na etapa 2 (*Obtenha os Dados*), é realizada a obtenção dos dados que devem ser descritos, informando de que forma foi feita a amostra, quais dados são considerados relevantes e se existem problemas de privacidade atrelados a esses dados.
- Na etapa 3 (*Explore os Dados*), os dados são explorados massivamente até que se encontre alguma anomalia ou padrão que possa representar uma perspectiva de solução do problema.
- Na etapa 4 (*Modele os Dados*), é realizada a modelagem dos dados, por meio da construção e validação de um modelo que represente a estrutura dos dados analisados até chegar na resposta ao problema.
- Na etapa 5 (*Visualize e Divulgue os Resultados*), os dados podem ser visualizados e divulgados aos interessados. É interessante que algumas perguntas sejam feitas para verificar o aprendizado com o fluxo proposto, por exemplo: “o que foi aprendido com o fluxo?”; “os resultados têm noção lógica para a resposta do problema?”.

Essas etapas visam facilitar o processo para se realizar a análise de dados em ciência de dados. Essa abordagem permite que os dados sejam obtidos, preparados, analisados, visualizados e gerenciados, preservando as informações que podem ser obtidas com a ciência de dados.

Com base no livro *R for Data Science*, é definido um fluxograma do processo para ciência de dados com base em seis etapas, iniciando pela coleta de dados e finalizando com a comunicação dos resultados para os envolvidos de maneira automatizada e eficiente. As etapas são (vide Figura 4.5):

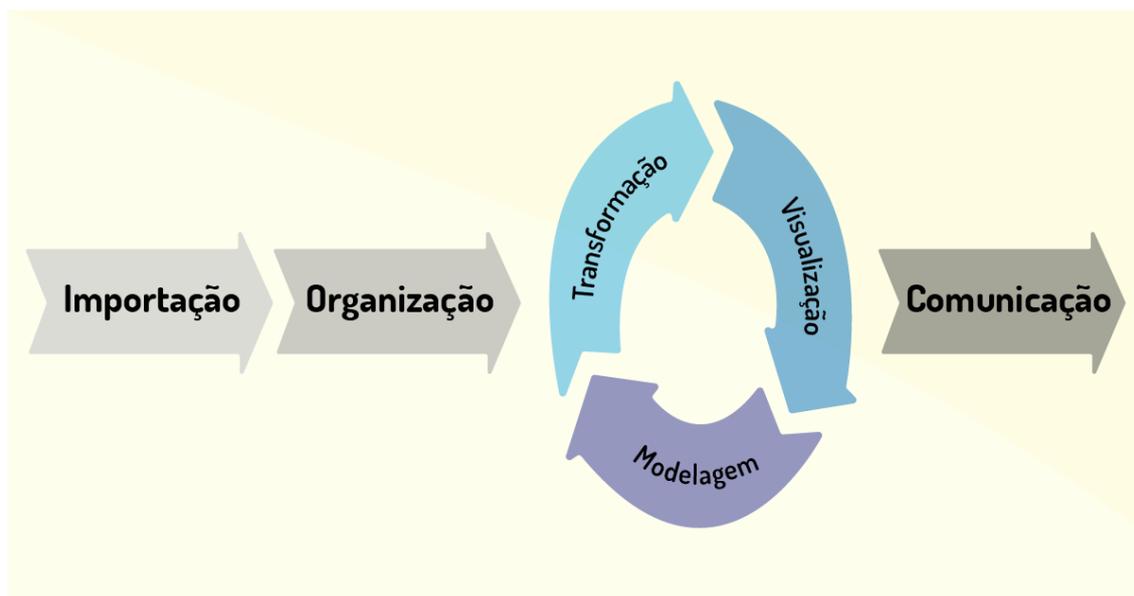


Figura 4.5 - Processo para Ciência de Dados

Fonte: Adaptada de Wickham e Grolemund (2017, p. 145).

**1. Importação:** os dados são importados de diversos programas e/ou fontes de dados em que estão armazenados. Ex.: formato em arquivo do Word, em banco de dados como o MongoDB, entre outros.

**2. Organização:** aqui os dados serão organizados para poder armazenar a base de dados de forma consistente. Podem ser dispostos no formato de coluna, contendo uma variável, ou linha, contendo uma observação.

**3. Transformação:** nesta etapa podem ser criadas variáveis a partir das variáveis já armazenadas ou por meio de agrupamentos de variáveis, ou até mesmo utilizando cálculos estatísticos.

**4. Visualização:** consiste na primeira etapa em que o conhecimento é gerado, em que os dados são apresentados conforme um questionamento levantado. Na Figura 4.6, temos um exemplo de visualização dos dados de salário em função da idade e da educação de trabalhadores de uma empresa. A visualização de dados permite realizar novos

questionamentos sobre eles, para que outras análises possam ser feitas, por exemplo: para os maiores salários com formação superior, qual a representatividade de cargos de chefia?

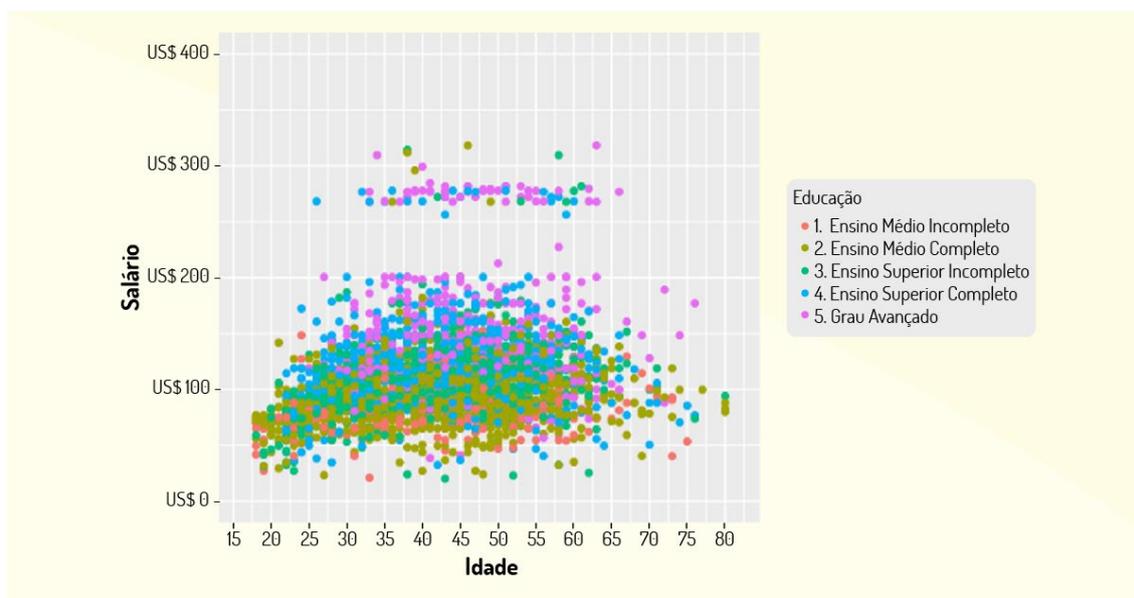


Figura 4.6 - Exemplo de visualização de dados

Fonte: Adaptada de Oliveira, Guerra e McDonnell (2018, p. 82).

**5. Modelagem:** os modelos são ferramentas que complementam a etapa de Visualização, vista no ponto anterior. Podem ser utilizados modelos computacionais e até mesmo estatísticos para realizar o processamento dos dados que irão responder a questionamentos propostos para análise de dados.

**6. Comunicação:** são reportados a todos os interessados do processo os conhecimentos que foram obtidos com a extração dos dados.

Perceba que a diferença desse processo é que são utilizadas diversas ferramentas para que a análise de dados possa ser feita de forma automatizada e menos manual, como ocorre no processo descrito anteriormente.

Vale ressaltar que a essência em ciência de dados é iniciar com uma base de dados extensa e analisar esses dados para que sejam geradas informações importantes. Após isso, por

meio de combinações de informações, pode-se gerar o conhecimento que possibilita solucionar o problema proposto inicialmente. Por conseguinte, alguns *insights* (previsões), de forma isolada, podem ser adquiridos e, após analisá-los de forma extensa, pode ser obtida a sabedoria que se precisa para solucionar o problema corrente e, se eventualmente surgirem outros problemas similares, ser a base para a solução (WICKHAM; GROLEMUND, 2017). Podemos visualizar melhor na Figura 4.7:

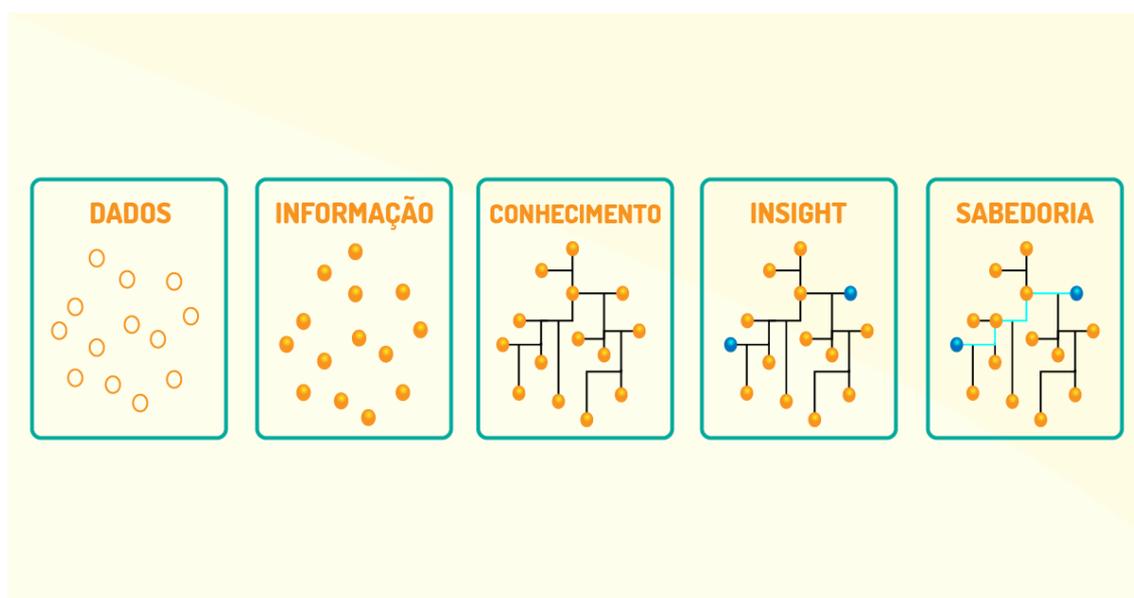


Figura 4.7 - Dos dados à sabedoria

Fonte: Adaptada de Wickham e Grolemond (2017, p. 150).

Por fim, a ciência de dados pode ajudar a solucionar problemas de diversas áreas como: indústrias, setor financeiro, varejo, entre outras. Para isso, devem ser empregados algum processo para auxiliar as organizações em suas tomadas de decisões, gerenciamento de riscos, análises preditivas sobre o negócio, possibilidade de aumentar a oportunidade de negócios. Enfim, a ciência de dados é considerada uma técnica avançada para analisar uma extensa quantidade de dados em busca de padrões que possam auxiliar na resolução de problemas ou, ainda, na busca da melhoria de processos e/ou negócios.

## ATIVIDADE

2) Um fluxo de atividades determina quais passos devem ser seguidos para se chegar a um resultado. Os passos se propõem a ser esquemas de tarefas que, se seguidos sequencialmente em uma ordem predefinida, garantem a solução do problema.

Em ciência de dados, tem-se um fluxo de atividades com cinco tarefas, que são:

- a) faça uma pergunta interessante, obtenha os dados, explore os dados, modele os dados e visualize e divulgue os resultados.
- b) faça uma análise prévia dos dados, obtenha os dados, explore os dados, modele os dados e visualize e divulgue os resultados.
- c) minere os dados, obtenha os dados, explore os dados, modele os dados e visualize e divulgue os resultados.
- d) faça uma pergunta interessante, obtenha os dados, explore os dados, modele os dados e visualize e categorize os resultados.
- e) faça uma pergunta interessante, obtenha os dados, explore os dados, minere os dados e visualize e divulgue os resultados.

## **Ciência de dados - Tecnologias/conhecimentos envolvidos**

A ciência de dados é uma área multidisciplinar, conforme já vimos em nossos estudos. Requer, basicamente, os conhecimentos de estatística, ciência da computação e matemática para que análise de dados seja efetuada de forma metodológica. Para isso, são utilizadas diversas tecnologias que podem ser agrupadas em três categorias: linguagens de programação, armazenamento e processamento de dados e estruturas de banco de dados SQL e NOSQL.

## **Linguagens de programação**

Várias linguagens de programação podem ser utilizadas para ciência de dados, como Java, R, Python, Julia, dentre outras opções. A linguagem de programação Python tem sido a mais utilizada, porém, a linguagem Julia tem despontado, sendo uma linguagem de programação de alto nível, dinâmica e de alto desempenho. Neste tópico, veremos as linguagens de programação R e Python, bem como suas especificidades. Vamos lá?

## **Linguagem R**

É definida como uma linguagem de programação com o objetivo de solucionar problemas estatísticos, sendo bastante utilizada para a visualização gráfica de dados. Foi criada pelos pesquisadores Ross Ihaka e Robert Gentleman, que, com a letra inicial dos seus nomes, nomearam a linguagem como R (MSPERLIN, *on-line*).

O R apresenta algumas vantagens, dentre as quais destacam-se: definida como uma plataforma estável e utilizada amplamente na indústria; possui o suporte necessário para correção de erros e atualizações de versões; tem pacotes de códigos para as mais diversas funcionalidades; altamente compatível com outras linguagens de programação e sistemas operacionais, e é uma linguagem gratuita, o que influencia a sua seleção em um ambiente empresarial, sem custos com licença e distribuição (WICKHAM; GROLEMUND, 2017).

Para entender melhor como o R funciona, vamos observar sua arquitetura, conforme Figura 4.8:

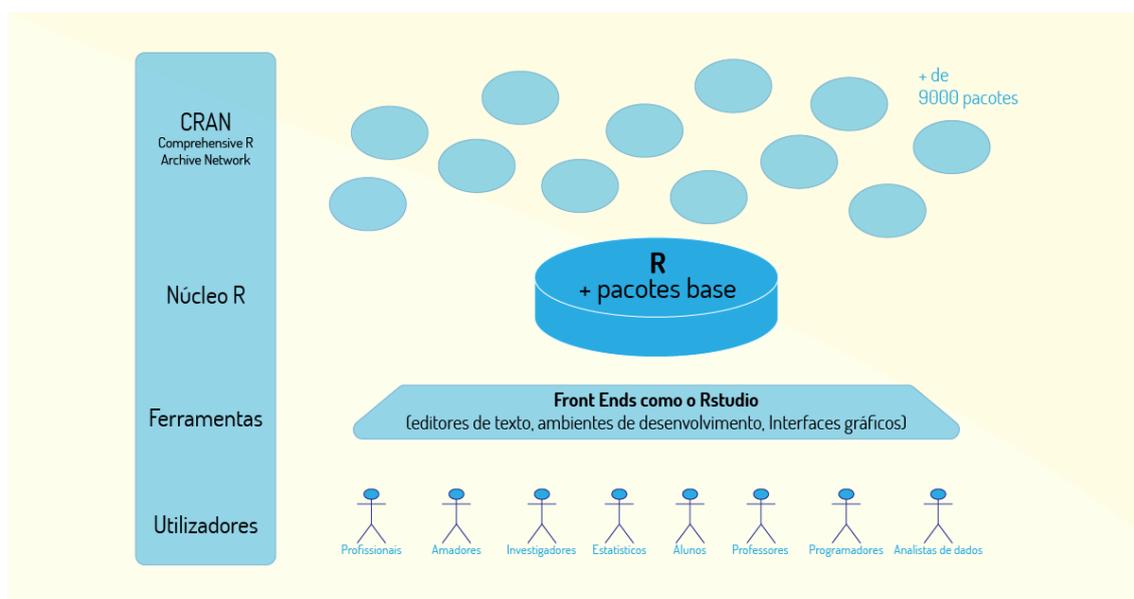


Figura 4.8 - Arquitetura do R

Fonte: Adaptada de Gouveia (2017).

Basicamente, os componentes da arquitetura R podem ser descritos como: utilizadores, aqueles que podem utilizar o R para diversos fins; ferramentas, são disponibilizadas várias ferramentas, sendo a mais utilizada o RStudio; o núcleo R é composto por pacotes de código base do R e CRAN (*Comprehensive R Archive Network*), que é uma rede de servidores web, e FTP (*File Transfer Protocol*), que armazena códigos e documentação atualizados do R (GOUVEIA, 2017).

Adicionalmente, R é utilizado em ciência de dados quando as tarefas de análise de dados precisam ser efetuadas em um único computador (*standalone*), ou quando a análise precisa ser feita em servidores individuais. Além disso, é muito eficiente para a análise

de dados, pois possui diversos pacotes com fórmulas e testes estatísticos (GOUVEIA, 2017).

## Python

É uma linguagem de programação com um diverso conjunto de funcionalidades que vão desde a coleta de dados até a construção de aplicativos. Por ser uma linguagem orientada a objetos, tem ampla aceitação no mercado, já que esse é um padrão para as novas linguagens de programação e, com ela, é mais acessível escrever um código em larga escala e robusto (PYTHON SOFTWARE FOUNDATION, 2010).

Quando Python é empregado com outras ferramentas (bibliotecas) conhecidas por cientista de dados – como Pandas (manipulação de dados), Numpy e Scipy (computação científica), Scikit-learn (Aprendizado de máquina) e Seaborn (visualização de dados) –, torna-se a linguagem preferida por eles e facilita o trabalho estatístico (MATOS, 2018).

Para que os usuários possam contribuir com a linguagem, podem adicionar novos pacotes no repositório Pypi, que também possui diversas bibliotecas disponíveis para os usuários, disponibilizando várias funcionalidades.

Python pode ser utilizado para análise de dados de forma integrada com aplicativos web ou por meio de códigos estatísticos integrados em um servidor em ambiente de produção. É considerada, também, uma ótima opção para scripts de automatização de mineração de dados, além de possibilitar o seu uso em partes do desenvolvimento de *softwares frontend* e *backend*.

Uma de suas ferramentas IDE é o Jupyter Notebook (antigo IPython Notebook), uma ferramenta para análise exploratória e apresentações de dados, sendo um ambiente computacional interativo e incremental que combina execução de código e cálculos matemáticos (MATOS, 2018; PYTHON SOFTWARE FOUNDATION, 2010).

Vale ressaltar que alguns atrativos podem ser mencionados para que Python seja utilizada como linguagem para análise de dados, como: código aberto, em que o usuário é livre

para instalar onde quiser; comunidade *on-line* disposta a ajudar na criação dos pacotes para o repositório; maior facilidade de aprendizado sobre a linguagem, e pode se tornar a linguagem mais acessível para análise de dados (PYTHON SOFTWARE FOUNDATION, 2010).

Por fim, Python é uma linguagem de amplo uso, que possui características inerentes às linguagens de programação e pode ser utilizada por usuários com diferentes tipos de conhecimentos. Possui uma gama de pacotes estatísticos que auxiliam na análise de dados e pode executar trechos de código em R. Está se tornando uma das ferramentas mais utilizadas atualmente em ciência de dados.

### **Armazenamento e processamento de dados**

Vamos conhecer, a seguir, o armazenamento e processamento de dados, por meio dos sistemas Pig e Hive, que têm como base o Apache Hadoop, plataforma que adota o conceito de *MapReduce*, a fim de prover a análise de dados em um ambiente de *Big Data*.

#### **Pig**

O Pig é um sistema de fluxo de dados de alto nível e de código aberto, que fornece uma linguagem simples chamada Pig Latin, para consultas e manipulação de dados. O Pig está sendo utilizado por empresas como *Yahoo*, *Google* e *Microsoft* para coletar grandes quantidades de conjuntos de dados na forma de fluxos de cliques, registros de pesquisa e rastreamentos da Web. O Pig também é usado em alguma forma de processamento e análise ad-hoc de todas as informações (EDUREKA, 2019).

Foi desenvolvido anteriormente pelo Yahoo, em 2006, para que os pesquisadores tenham uma maneira ad-hoc de criar e executar tarefas do Map Reduce em conjuntos de dados muito grandes. Foi criado para reduzir o tempo de desenvolvimento por meio da abordagem de várias consultas (EDUREKA, 2019).

Por fim, o Pig fornece operações de dados como filtros, junções, ordenações, suporte a tipos de dados aninhados, como tuplas e mapas, que estão faltando no *Map Reduce*, além de ser fácil de escrever e de ler, principalmente se houver conhecimento sobre SQL (*Structured Query Language*).

## **Hive**

O *Hive* é um pacote de *data warehouse* criado sobre o Hadoop para executar a análise de dados. É direcionado para usuários que se sentem confortáveis com o SQL, pois possui uma linguagem de programação chamada 'HiveQL', que é semelhante ao SQL. O Hive é usado para gerenciar e consultar dados estruturados (EDUREKA, 2019).

Ele abstrai a complexidade do Hadoop, ou seja, você não precisa escrever um programa de Map Reduce. Com o Hive, também não é necessário que o usuário aprenda as APIs Java e Hadoop. Com os incríveis recursos do Hive, o Facebook agora pode analisar vários Terabytes de dados todos os dias, já que a empresa queria maneiras diferentes de armazenar, extrair e analisar dados. Hoje em dia, os dados do Facebook são coletados por *cronjobs* (tarefas agendadas) e armazenados no OracleDB (EDUREKA, 2019).

## **Estruturas de banco de dados**

Como todos os dados disponibilizados podem ser armazenados de forma temporária ou não, é necessário ter conhecimento sobre tipos de banco de dados, relacionais ou não, para que seja gerado o conhecimento necessário para alcançar o objetivo da solução de um problema determinado.

Uma das formas de manipulação dos dados é utilizando a linguagem SQL (*Structured Query Language*), que é uma linguagem padrão universal para realizar essa manipulação por meio de SGBDs (Sistema de Gerenciamento de Banco de Dados), que podem ser definidos como um conjunto de *softwares* responsáveis por realizar o gerenciamento de uma base de dados (SQL, *on-line*).

O SGBD permite realizar várias tarefas relacionados aos dados, como salvá-los no HD, fornecer uma interface para que programas e usuários externos possam acessá-los, realizar a ligação de dados e metadados, permitir o controle de acesso às informações do banco de dados, o que pode ser realizado utilizando a linguagem SQL (SQL, *on-line*).

Em ciência de dados, os dados podem ser obtidos por meio de uma grande base de dados e, então, é interessante conhecer a linguagem SQL, que serve como linguagem universal para realizar qualquer operação em um SGBD, independentemente do tipo de programa, código e/ou tarefas utilizados nesse SGBD (VICTÓRIA, 2019).

Mas como a quantidade de dados a ser armazenada é muito grande, foi nesse contexto que surgiram os bancos de dados não relacionais (NOSQL), que permitem armazenar dados estruturados ou não, de forma mais veloz e flexível e têm escalabilidade, sendo ideais para soluções de *Big Data* (VICTÓRIA, 2019).

O NOSQL tem uma série de tipos de bancos de dados, como: colunas, em que os conjuntos de dados são armazenados em linhas particulares de uma tabela; grafos, em que os dados são armazenados em formato de arcos ou vértices conectados por arestas (conjunto de linhas); chave-valor, em que o banco armazena um conjunto de chaves com seus respectivos valores, simulando tabelas hash, e documento, em que todo dado não estruturado é um documento que pode possuir diversos formatos. Vale ressaltar, também, que o NOSQL é gratuito e possui um alto desempenho em consultas ao banco de dados (VICTÓRIA, 2019).

Por fim, essas estruturas de banco de dados, relacionais ou não, são adaptáveis para cada contexto em ciência de dados. Elas não são competitivas, pois, se tivermos um grande volume de dados, é recomendado o NOSQL, e, caso a análise e tratamento de dados não precise ser tão profunda, podem ser utilizados os bancos de dados relacionais.

## ATIVIDADE

3) Para que a ciência de dados possa ser implementada são necessários alguns recursos tecnológicos como linguagens de programação, estruturas de armazenamento e processamento de dados e banco de dados. Qual das alternativas a seguir refere-se a exemplos de banco de dados?

- a) Java e Mysql.
- b) Mysql e Nosql.
- c) Nosql e Java.
- d) PHP e Mysql.
- e) Ruby e Nosql.

## Ciência de dados - Exemplos de usos e cases

Agora que entendemos os conceitos acerca de ciência de dados, vamos compreender melhor como as organizações estão aplicando essa área em alguns casos de sucessos que corroboram para o crescimento dessa nova forma de analisar os dados.

### Case 1: Airbnb

O serviço de hospedagem *on-line* Airbnb realiza sugestões de hospedagem para seus usuários baseadas nos parâmetros informados por eles. A busca retorna uma sugestão de local para que o usuário se hospede (AIRBNB, *on-line*).

Assim, o Airbnb sempre precisou dos dados para que o negócio pudesse ter sucesso, procurando entender seus dados com base na demografia dos locatários, preços, período e número de hóspedes. Baseando-se na enorme quantidade de clientes cadastrados na plataforma, demanda por aluguéis e locais, foi necessário analisar mais profundamente esses dados e, assim, se apoiaram na ciência de dados (AIRBNB, 2019).

A empresa criou um sistema de preço dinâmico denominado Aerosolve, que realiza uma predição de preço com base na localização do usuário que tem o local para alugar, período do ano e uma série de outros parâmetros para se chegar a uma faixa de preços ideal e compatível com o mercado de aluguel, e, assim, tanto a empresa quanto o locatário podem maximizar seus lucros (AEROSOLVE, *on-line*).

Além disso, a empresa divide suas equipes em duas grandes categorias: **negócios**, na qual os funcionários estão preocupados em como o locatário pode oferecer um serviço de qualidade e excelência, em que os locadores podem acompanhar suas despesas na locação do imóvel, e **plataforma**, que realiza o desenvolvimento de APIs (*Application Programming Interface*) para processar os pagamentos dos usuários, manter a segurança da comunidade Airbnb e fornecer a infraestrutura de dados para as ferramentas de análise (ILUMEO, *on-line*).

Nessas duas equipes há cientistas de dados que se propõem a descobrir as necessidades e desafios de cada equipe e trabalham de forma otimizada com vários setores, como: engenharia, design, arquitetura, entre outros, para que os dados sejam considerados itens avaliativos na maioria das decisões dos grupos (ILUMEO, *on-line*).

Por fim, Airbnb possui no seu quadro de funcionários de cientista de dados e especialistas de diversas áreas, como: ciência da computação, para a construção de algoritmos complexos de aprendizagem; economistas que avaliam como maximizar o lucro da empresa, e estatísticos que realizam análises sobre o impacto de cada mudança (ILUMEO, *on-line*).

## **Case 2: Netflix**

Um dos maiores serviços de *streaming*, a Netflix obteve grande sucesso pela utilização da ciência de dados. Em todo o tempo que o usuário está utilizando o serviço, ele fornece informações valiosas para os cientistas de dados que fazem análise minuciosas sobre esses dados para aumentar o poder competitivo da empresa (COMO..., 2018).

Em 2015, a Netflix disponibilizou um infográfico em que apontou os episódios de séries mais populares, a partir do qual 70% dos espectadores assistiram até finalizar a primeira temporada. Dessa forma, a Netflix proporcionou o *binge watching* (“maratona de episódios”), em que os usuários assistem vários episódios de uma só vez, enquanto as outras emissoras só liberam um episódio por vez. Foi com base nessa pesquisa que a Netflix passou a disponibilizar todos os episódios de uma só vez e, assim, conseguiu fidelizar o seu usuário na plataforma (COMO..., 2018).

Além desse caso, mais precisamente em 2011, a Netflix analisou todos os dados de consumo do usuário e percebeu que os usuários gostavam de filmes estrelados por Kevin Spacey, os longas metragens produzidos por David Fincher e uma minissérie britânica (1990) sobre política. A partir disso, ela passou a produzir, em 2013, a série *House of Cards*, que é uma das séries de maior sucesso que consiste em um drama político produzido por Fincher e tem o ator Spacey como papel principal (COMO..., 2018).

Adicionalmente, ainda utilizando a ciência dos dados, a empresa produziu corretamente *teasers* para promover a série, e cada *teaser* era especial para cada perfil de usuário: se um usuário gostava mais de filmes com Spacey, o *teaser* teria mais cenas do ator; se outro usuário apreciava mais cenas do sexo feminino, então, o foco do *teaser* era no elenco feminino da série. Tudo isso atraía o usuário para o que ele realmente apreciava em uma série.

Um outro caso de sucesso da Netflix foi o lançamento, em 2016, da série *Stranger Things*, uma das séries de maior sucesso da empresa. A série foi rejeitada por, pelo menos, 15 redes de TV até ser recebida pela Netflix para o lançamento. Essas redes de TV consideraram que não havia público-alvo para uma série estrelada por crianças que não fosse especificamente infantil. No entanto, a Netflix, utilizando os dados de seus usuários, percebeu que havia, sim, um público-alvo e conseguiu um grande sucesso lançando a série (GALLI, 2017).

Por fim, a empresa relata que utiliza da linguagem Python para análise de dados em seu serviço, realizando tarefas de monitoramento de dados, movimentação e sincronização dos dados, ativação de aplicativos dentro da plataforma para obter informações, entre outros. Dessa forma, a linguagem, juntamente com uma equipe de excelentes profissionais, auxilia no sucesso que a empresa é hoje!

### **Case 3: Bayes Impact**

A empresa Bayes Impact, fundada em 2014 e sediada em São Francisco (Norte da Califórnia), composta por vários cientistas de dados, auxilia ONGs (Organizações sem fins lucrativos) a resolver vários problemas, sejam de ordem social e até mesmo de saúde (IMPACT..., *on-line*).

Um dos casos que a empresa ajudou a solucionar foi junto ao departamento de saúde dos EUA para realizar melhores combinações entre as pessoas que se dispõem a doar órgãos e as que precisam de transplantes. A partir dos dados informados por esses tipos de pessoas, os cientistas de dados conseguiram realizar uma análise melhor desses dados

e, assim, propor uma combinação mais eficiente e que teve a maioria de acertos (IMPACT..., *on-line*).

Outros casos em que a empresa teve grande importância com a utilização dos dados foi para a Michael J. Fox Foundation, em que desenvolveu melhores métodos de ciência de dados para corroborar com as pesquisas de Parkinson.

Além disso, a Bayes criou rotinas específicas para combater fraudes em microfinanças. As empresas de microcrédito estavam com altas taxas de fraude e esses obstáculos impediam sua capacidade de fornecer crédito rapidamente aos interessados, mas também no quanto eles poderiam emprestar (IMPACT..., *on-line*).

A Bayes Impact trabalhou com a organização de microfinanças Zidisha, que conecta diretamente credores e mutuários sem passar por uma estrutura intermediária. Zidisha trabalha, principalmente, no Quênia e Gana. O Bayes Impact desenvolveu algoritmos para o Zidisha para detectar rapidamente possíveis tentativas de fraude. Após essa intervenção, a fraude caiu 35% na plataforma e 10% dos novos pedidos de empréstimo foram aceitos graças à detecção de fraudes da Bayes (IMPACT..., *on-line*).

Vale ressaltar que a empresa está construindo um modelo com base na ciência de dados para auxiliar a cidade de São Francisco na otimização de serviços básicos e essenciais da cidade, como otimizar as taxas de respostas em emergências.

Percebe-se que a Bayes maximiza o uso da ciência de dados para tentar solucionar problemas sociais. Eis aqui uma outra finalidade desse tipo de ciência: enquanto algumas empresas preferem utilizar os dados para auxiliar na geração de lucro e tomada de decisão, essa empresa está voltada a promover iniciativas que solucionem problemas de ordem social e que possam causar um grande impacto em um mundo que é orientado por dados.

#### **Case 4: Merck**

A Merck KGaA, em Darmstadt, Alemanha, é uma empresa líder em ciência e tecnologia nas áreas de cuidados com a saúde, avaliada em 40 bilhões e presente em 140 mercados em todo o mundo (MERCKGROUP, *on-line*).

Um dos grandes problemas da empresa era que seus engenheiros dispndiam entre 60% e 80% para analisar, acessar e visualizar os dados valiosos para cada projeto e, assim, o esforço era muito alto e lento, o que fazia com que os negócios da empresa fossem comprometidos pela baixa velocidade na aquisição de informações (MERCKGROUP, *on-line*).

Dessa forma, a empresa implantou o MANTIS (*Manufacturing and Analytics Intelligence*) para obter dados do ERP e sistemas centrais de fabricação e controle de estoque para obter mais *insights* para os negócios (MERCKGROUP, *on-line*).

Esse sistema, classificado como armazenamento de dados, é composto por vários bancos de dados e ferramentas de código aberto que processam dados estruturados ou não e que estejam armazenados nos sistemas da empresa. Os dados processados pelo MANTIS poderiam ser textos, áudios, vídeos, imagens, entre outros (MERCKGROUP, *on-line*).

A partir disso, analistas de negócios, que não têm a expertise técnica, começaram a visualizar os dados a partir de uma interface de visualização dos dados desse sistema e, assim, puderam sugerir melhores *insights* para a empresa com essa melhoria.

Outro ponto foi que os cientistas de dados da empresa melhoraram suas análises a partir de ferramentas de alta tecnologia desse sistema, como as de simulação e modelagem dos dados, auxiliando o trabalho dos engenheiros da empresa.

Assim, o MANTIS ajudou nos processos da empresa e reduziu em 45% o tempo e o custo do portfólio geral de projetos de análise de TI (Tecnologia da Informação) da empresa. Além disso, trouxe resultados comerciais tangíveis, com uma redução de 30% no tempo médio de entrega dos projetos e uma redução de 50% nos custos médios de estoque com os medicamentos. Vale ressaltar que a empresa realizou um teste piloto desse sistema em uma fábrica da Ásia e Pacífico e percebeu que haveria um grande retorno para a empresa; dessa forma, seguiu para implantar por toda as filiais da empresa (MERCKGROUP, *on-line*).

## ATIVIDADE

4) Considere o trecho a seguir:

“A equipe do Booz Allen conseguiu desenvolver um aplicativo para os treinadores da MLB para prever qualquer lançamento de arremessador com até 75% de precisão, mudando a maneira como as equipes se preparam para um jogo, formação da equipe adversário e o melhor arremesso com base em alguns parâmetros” (3 EXEMPLOS..., 2019, *on-line*).

Pode-se afirmar que o aplicativo mencionado no contexto anterior está utilizando técnicas de ciência de dados, pois:

- a) com as predições do aplicativo, o treinador procura formas específicas para se preparar para o jogo.
- b) por meio das previsões do aplicativo, a equipe fica limitada em suas jogadas.
- c) as equipes treinam movimentos específicos, mas não sabem qual é o melhor momento para o arremesso.
- d) as equipes antecipam as suas jogadas, mas não conseguem analisar como será a formação da equipe adversária.
- e) mediante as predições do aplicativo, o jogo não mudará a forma como acontece.

## Aprendizado de máquina - Definição

Existem várias definições para esse termo, mas vamos, aqui, tentar resumir para você. Do inglês *machine learning*, o aprendizado de máquina pode ser definido, basicamente, como uma subárea de inteligência artificial que permite a construção de sistemas automatizados, possuindo a capacidade de aprender com a grande quantidade de dados recebidas e auxiliar na tomada de decisões sem a interferência humana (SAS, *on-line*).

Outros conceitos estão atrelados ao aprendizado de máquina, como:

- é uma técnica de análise de dados que pode ser entendida como a capacidade de transformar dados brutos em informações relevantes. Essa técnica automatiza a concepção de modelos analíticos (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012).
- é um método que permite que as máquinas aprendam, de forma isolada, determinadas tarefas humanas a partir da análise de um grande volume de dados (MAGNUS, 2018).
- é um ramo da ciência da computação que fornece respostas automatizadas ao usuário, utilizando conceitos de *Big Data* e Inteligência Artificial (SILVA *et al.*, 2007).

Dessa forma, percebe-se que o aprendizado de máquina tem um conceito bem amplo e está intrinsecamente ligado à *Big Data* e à inteligência artificial (conceitos vistos anteriormente) e, assim, conseguem analisar essa quantidade de dados por meio do uso de uma infinidade de algoritmos para estabelecer padrões nos dados analisados. Mas o que são esses algoritmos? Bom, vamos prosseguir para explicar melhor.

Os algoritmos são a base fundamental para que o aprendizado de máquina possa funcionar. Tais algoritmos fazem uso de análises estatísticas para realizar previsões das respostas de forma mais precisa e, assim, entregar o melhor resultado da predição com uma chance de erro mínima.

Esses algoritmos podem ser classificados em (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012):

**1. supervisionados:** o aprendizado do algoritmo consiste em receber os dados que possuem a resposta certa. Dessa forma, há uma intrínseca relação entre a entrada e a saída de dados, servindo para treinar o algoritmo. Nesse tipo de algoritmo, tem-se a intervenção humana para controlar a entrada, a saída e o resultado das previsões do algoritmo. Assim, o algoritmo aprende a melhor resposta e a utiliza para a sua próxima interação com os dados. Ex.: ao contratar um empréstimo em um banco, toda a informação a ser analisada é com base no histórico de crédito do cliente (entradas) e a concessão de crédito; com base nesses dados, pode retornar de forma positiva ou negativa (saída). Em resumo, os dados de entrada já têm rótulos que servirão para a saída de dados e precisam ser treinados por humanos no algoritmo para concessão de crédito.

**2. não supervisionados:** é feito o processamento de tarefas muito complexas, sem o treinamento realizado por um humano, e os dados que serão a entrada não possuem rótulos específicos, ou seja, os efeitos das variáveis de entrada e saída não conseguem ser previstos previamente. Além disso, os resultados não têm relação com o padrão encontrado na análise de dados. Ex.: os artigos científicos são categorizados conforme alguns parâmetros, como: número de páginas, palavras-chaves, temas, entre outros. Ao buscar artigos científicos em um buscador, ele encontra um padrão conforme os parâmetros especificados para os artigos selecionados, e o algoritmo procura encontrar novos padrões para, assim, melhorar o resultado da busca.

Com relação à aprendizagem, eles podem fazer uso do algoritmo supervisionado, não supervisionado, ou de nenhum deles, podendo ser classificados em quatro aprendizados de máquina existentes, que são (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012):

- **aprendizado supervisionado:** com a base do algoritmo supervisionado, o sistema tem como entrada um conjunto de dados com a resposta correta (dados rotulados), e o que a máquina precisa é retornar a resposta correta a partir do aprendizado dessas variáveis de entrada e saída. Ex.: busca de imagens no Google, em que o algoritmo busca a origem da imagem, já classificada como a resposta correta para aquela determinada busca, com base nos dados fornecidos. A supervisão humana ocorre, pois os que buscam a imagem inserem novos dados de busca, melhorando o algoritmo de forma que os novos dados aperfeiçoem, de forma semântica, as chances de acerto na busca pela solução correta.

- **aprendizado não supervisionado:** não se tem um resultado prévio (dados não rotulados), conforme já vimos nos algoritmos não supervisionados. Por ser mais complexo, a descoberta de padrões é sempre uma nova descoberta e depende das variáveis de entrada no sistema. Ex.: para descobrir hábitos alimentares, é necessário realizar um conglomerado de informações como frequência da compra de certos alimentos, registros de compras e perfil do cliente. Tudo isso deve ser associado para se chegar a um padrão. O algoritmo realiza essa associação de forma autônoma, sem a intervenção humana.

- **aprendizado semissupervisionado:** utiliza o algoritmo supervisionado, mas pode receber dados rotulados e não rotulados. Esse tipo de aprendizado é utilizado quando o custo com dados rotulados (aqueles em que o computador já conhece a resposta advinda desses dados) é muito caro e precisa de um treinamento rotulado. As respostas para as incertezas do algoritmo são pequenas e, assim, auxiliam no direcionamento de descobertas de padrões da máquina. Ex.: identificação do rosto de uma pessoa por meio do recurso de webcam. O rosto terá partes como dados rotulados e partes como não rotulados para o aprendizado (treinamento) que irão auxiliar a máquina na redução de erros para a sua identificação.

- **aprendizado por reforço:** não utiliza nenhum tipo de algoritmo (supervisionado ou não supervisionado). Ele é baseado em tentativa e erro e, a partir disso, analisa quais das suas ações tem os melhores resultados e as guarda para futuras iterações no algoritmo. Esse aprendizado é bastante empregado em áreas como robótica e jogos, em que o robô é lançado em um ambiente e faz diversos testes para coletar informações e se adaptar. Um outro exemplo é quando a máquina estrutura um portfólio de ações para investimentos; ela analisa o retorno financeiro e como o mercado está evoluindo naquele momento a fim de prever a melhor solução de investimento, sem possuir nenhum treinamento prévio disso.

Com isso, temos diferentes tipos de aprendizados de máquinas, sendo que não há um melhor entre eles. Lembre-se de que as máquinas são programadas por humanos para executarem as tarefas que solucionem o problema. Diante disso, é o humano quem treina a máquina para escolher o melhor algoritmo para um determinado contexto.

## Origem do aprendizado de máquina

Quando, em 1946, foi criado o primeiro sistema de computador, o ENIAC (*Electronic Numerical Integrator and Computer*), quase tudo era realizado pelos humanos, até mesmo a busca de conexões entre as diferentes partes do sistema para se obter um resultado. Nesse contexto, a máquina se equiparava apenas com as habilidades humanas em realizar cálculos matemáticos, porém, com maior desempenho (MAGNUS, 2018).

Já em 1950, o pesquisador e pai da computação, Alan Turing, começou a questionar se as máquinas poderiam pensar e começou a desenvolver o Teste de Turing, que realizava análises de interações entre a máquina e os seres humanos (MAGNUS, 2018).

A partir disso, um grande estudioso de Inteligência Artificial, Arthur Samuel, criou o primeiro sistema capaz de aprender a jogar damas. Samuel foi vencido pela máquina nesse jogo, em que a máquina ia melhorando suas jogadas e seu desempenho a cada nova partida, realizando estudos dos melhores movimentos no jogo e até mesmo propondo novas estratégias. Foi a partir disso que Samuel começou a empregar o termo “Aprendizado de Máquina”, em 1959 (SAS, *on-line*).

Outro grande *software* que ajudou a propagar o termo “aprendizado de máquina” nos anos 1960 foi a construção do sistema ELIZA, desenvolvido pelo professor Joseph Weizenbaum, do laboratório de Inteligência Artificial do Instituto de Tecnologia de Massachusetts (Massachusetts Institute of Technology – MIT). O objetivo desse sistema é a simulação de diálogos, como uma psicóloga, construindo novas perguntas com base no que o paciente informa. Muitos usuários acreditavam que a ELIZA era uma psicoterapeuta humana. Na Figura 4.9, tem-se o professor interagindo com ELIZA:



Figura 4.9 - Weizenbaum interagindo com o sistema ELIZA

Fonte: Eliza (*on-line*).

A partir desse contexto de origem, podemos perceber que a resposta à pergunta que Alan Turing fez (“As máquinas podem pensar?”) é: sim, as máquinas são capazes de reproduzir o raciocínio humano e estão cada vez mais inteligentes!

### **FIQUE POR DENTRO**

Vimos a Eliza em aprendizado de máquina, certo? Ela foi considerada o primeiro chatbot que simula diálogos, comportando-se como uma psicóloga, fazendo novas perguntas com base no que o paciente está falando naquele momento. Interessante não é mesmo? Se você tem facilidade com o inglês, acesse o link indicado e converse diretamente com a Eliza, em uma implementação desenvolvida em JavaScript, disponível em: <http://www.masswerk.at/elizabot/>. Acesso em: 26 fev. 2020.

Vale ressaltar, aqui, a diferença entre inteligência artificial (IA) e aprendizado de máquina. Em IA, temos que a máquina será programada para simular algumas

características humanas, como reconhecimento de voz, percepção visual, traduzir idiomas e tomada de decisão com base em algum contexto. Já em aprendizado, os computadores são programados para aprender com os dados analisados e, assim, fornecer uma resposta mais exata. Apesar do aprendizado ser uma subárea de IA, em que todo aprendizado utiliza recursos de IA, nem toda IA tem o aprendizado automático da máquina (SAS, *online*).

Adicionalmente, o aprendizado de máquina tem a capacidade de adaptar-se, conforme é exposto, a outras quantidades de dados; isso faz com que ele seja um sistema dinâmico e não precise de especialistas humanos para realizar certas modificações no algoritmo. Por fim, a adoção de aprendizado de máquina só tende a aumentar e se tornar mais presente em aplicativos, assistentes digitais e na inteligência artificial de forma geral, haja vista que a demanda por dados e algoritmos está cada vez mais em expansão.

## ATIVIDADE

Analise as definições a seguir em relação aos tipos de aprendizado de máquina.

- 1) Supervisionado.
- 2) Semissupervisionado.
- 3) Não supervisionado.
- 4) Por reforço.

I - ( ) É um tipo de aprendizagem de máquina que investiga como agentes de *software* devem agir em determinados ambientes, utilizando técnicas de tentativa e erro.

II - ( ) É considerado um tipo de aprendizado de máquina que realiza o treinamento de dados rotulados para solução de uma tarefa, tendo a intervenção humana.

III - ( ) É um tipo de aprendizado de máquina que trabalha tanto com dados rotulados quanto com dados não rotulados, melhorando significativamente a acurácia.

IV - ( ) É classificado como um tipo de aprendizado de máquina em que a aprendizagem ocorre com dados não rotulados, ou seja, não dizemos ao computador o que é aquela entrada.

V - ( ) É uma forma de ensinar ao computador qual ação priorizar dada uma determinada situação, por meio de várias tentativas. Por reforço.

Em seguida, assinale a alternativa que relaciona cada item das descrições acima com o tipo de aprendizado de máquina de forma correta:

a) I - 1, II - 1, III - 2, IV - 3 e V - 4.

b) I - 4, II - 3, III - 2, IV - 3 e V - 4.

c) I - 4, II - 1, III - 2, IV - 3 e V - 4.

d) I - 4, II - 1, III - 2, IV - 4 e V - 4.

e) I - 1, II - 1, III - 2, IV - 3 e V - 4.

## **Aprendizado de máquina - Aspectos relevantes**

Agora que entendemos melhor o que é *machine learning* e seus diferentes tipos de aprendizados de máquina, que constituem a base do funcionamento dessa área, veremos quais os principais métodos estatísticos que auxiliam no alcance dos objetivos esperados ao utilizar esse tipo de aprendizado.

### **Tipos de métodos estatísticos**

No aprendizado de máquina, são utilizados diversos métodos estatísticos, que também são empregados pelos cientistas de dados, a depender do problema a ser solucionado e, assim, alcançar o desempenho que se espera na utilização do método selecionado. Vejamos, a partir de agora, os três principais métodos (SILVA *et al.*, 2007).

### **Regressão**

Esse tipo de método é aplicável em algoritmos supervisionados com o objetivo de prever os resultados em uma saída de forma contínua, com base no mapeamento das variáveis de entrada (dados rotulados), ou seja, realiza-se uma análise das variáveis de entrada para descrever as suas características e, assim, prever uma saída contínua, aquela que possui somente uma categoria e já é conhecida com base nos dados rotulados.

Na Figura 4.10, temos um exemplo de gráfico gerado em uma regressão linear, em que, no eixo x, temos as variáveis independentes (dados de entrada) e, no eixo y, as variáveis dependentes (dados de saída). Os pontos dos dados plotados permitem a construção da linha de regressão, que, por sua vez, permite a predição de valores de saída para um certo valor de entrada.

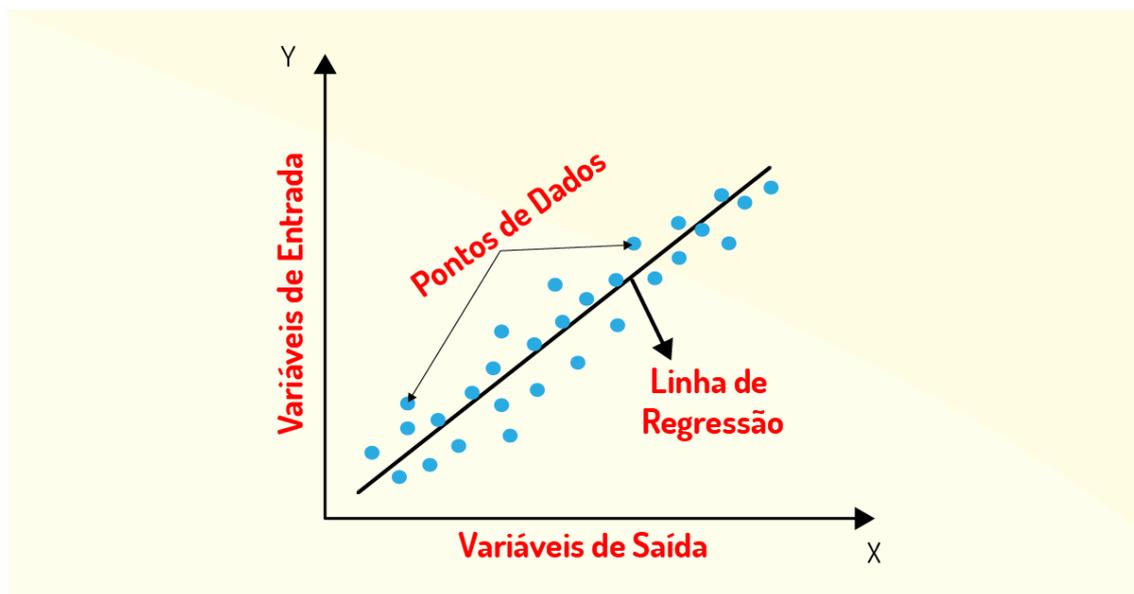


Figura 4.10 - Gráfico de Regressão Linear

Fonte: Elaborada pelos autores.

Vejam, a seguir, dois exemplos para clarificar melhor o uso da regressão.

Em um contexto imobiliário, recebe-se como dado de entrada o tamanho das casas e é solicitada a previsão de preço. Logo, sabe-se que existe uma relação entre tamanho da casa e preço, e o algoritmo só vai ratificar essa relação. A saída será contínua, pois será determinado o preço da casa conforme a média dos tamanhos. Assim, é um problema de regressão que vai buscar mapear o tamanho das casas para se chegar a um preço ideal.

Um outro exemplo que podemos citar é que podemos ter como variáveis de entrada a imagem de um homem ou mulher, sendo necessário informar uma saída, já conhecida, que é a idade. O método de regressão, então, analisa os dados da imagem e fornece uma saída contínua, que é a idade.

É importante frisar que o método da regressão faz o mapeamento procurando entender como as variáveis de entrada podem evoluir para gerar uma saída contínua. A ideia é encontrar uma saída que seja a média das oscilações que ocorrem ao analisar as entradas fornecidas.

Por fim, o método da regressão procura analisar de que forma ocorre a relação entre duas ou mais variáveis classificadas como quantitativas e, assim, tentar prever uma saída com base nessa relação.

## Classificação

É um tipo de método aplicável a algoritmos supervisionados em que o intuito é prever duas ou mais saídas com base nos dados rotulados de entrada, ou seja, será realizado um mapeamento das variáveis de entrada, agrupando-as em diferentes categorias para se chegar às saídas correspondentes.

Na Figura 4.11, temos um exemplo de gráfico utilizado pelo método de classificação, em que os dados de entrada são classificados nas categorias classe A e classe B.

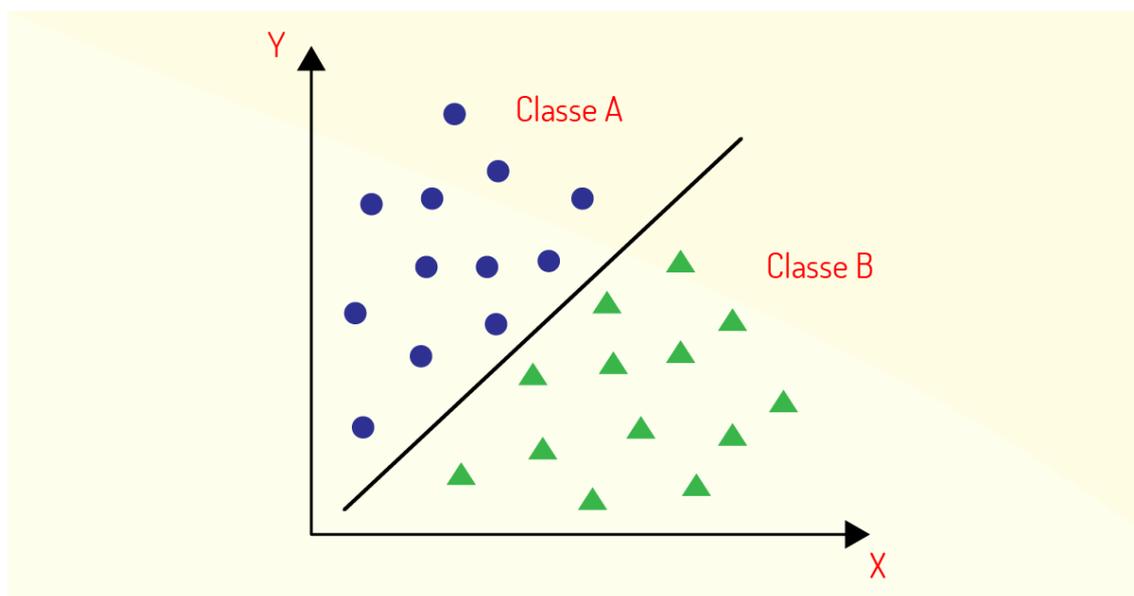


Figura 4.11 - Gráfico de classificação

Fonte: Elaborada pelos autores.

Vejam, a seguir, dois exemplos para ilustrar melhor o método da classificação.

Ainda no contexto imobiliário descrito acima, ao utilizar o método de classificação, com base nos dados fornecidos como tamanho da casa, o método iria agrupar as casas em que possuem um determinado tamanho e seriam classificadas com um preço abaixo do que foi estipulado como saída e as casas com tamanhos maiores estariam classificadas com um preço acima dessa saída.

Um outro exemplo da classificação, pode ser utilizado na oncologia, em que se conhece um tumor cancerígeno por algumas variáveis de entrada como tamanho e forma, e, com base nisso, tenta-se prever se o tumor é maligno ou benigno. A classificação vai categorizar os dados de entrada (tamanho e forma) para se chegar à conclusão do tipo de tumor.

Vale ressaltar que esse tipo de método busca analisar, entender e realizar previsões de variáveis em categorias e não quantitativas, como ocorre nos métodos de regressão. Essas categorias podem ser do tipo:

- binárias: que possuem somente duas categorias: ou é 1 (um), que significa que um evento ocorreu, ou 0 (zero), em que não existe ocorrência de evento.
- faixas: em que os dados são classificados por faixas que podem ser de idade, de renda, de escala, dentre outras.
- categóricas não ordenadas: em que cada categoria possui uma variável única em comparação com outras categorias, como exemplos têm-se gênero, cor, raça, entre outras.

Adicionalmente, o método de classificação recebe um dado de entrada e tenta atribuir algum rótulo a essa entrada, para, assim, classificar e realizar as previsões de saída. Geralmente, as respostas são simples e correspondem a “sim” ou “não”.

Por fim, o método da classificação é muito utilizado quando se tem um objetivo bem descrito e as saídas conhecidas, para que assim seja utilizado o algoritmo de aprendizagem supervisionada e poder informar resultados com base em categorias e não numéricas.

## Clustering

Esse tipo de método é aplicável a algoritmos não supervisionados, em que não se tem ideia do que os resultados podem apresentar, com base em um banco de dados. Além disso, esse método busca encontrar padrões e os dados só serão agrupados conforme se encontram as relações entre as variáveis.

Na Figura 4.12, temos um exemplo de gráfico que representa o processo do algoritmo de *clustering*, em que os dados que apresentam algum padrão são agrupados em *clusters* (no caso, *cluster 1*, *cluster 2* e *cluster 3*).

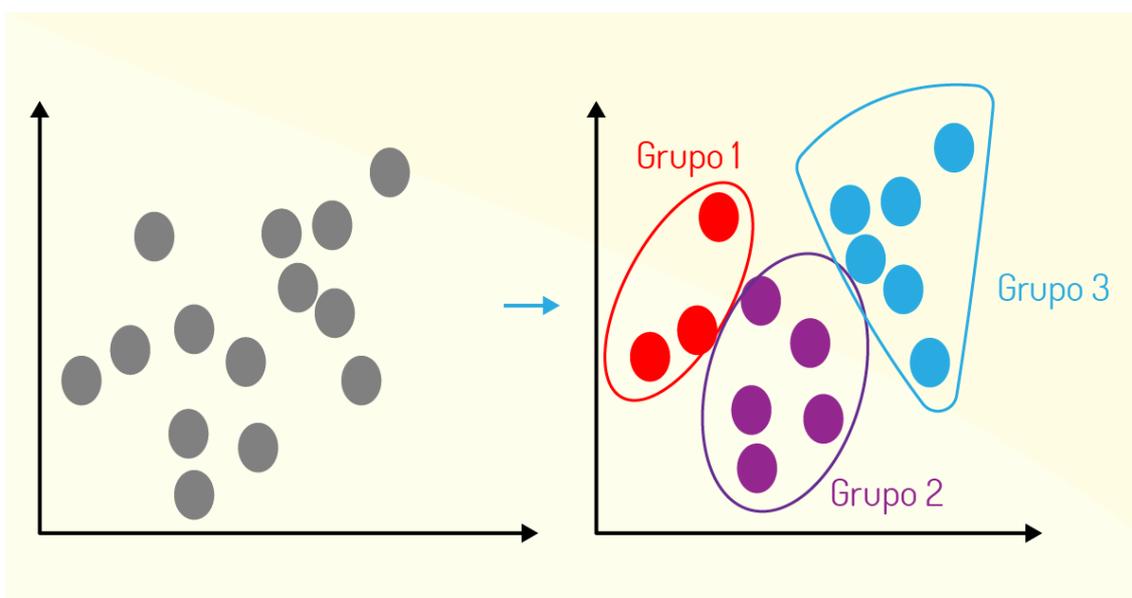


Figura 4.12 - *Clustering*

Fonte: Elaborada pelos autores.

Além disso, esse método pode ser bastante útil quando se deseja reduzir o espaço amostral, dado um conjunto de dados para se ater somente nos atributos que têm relevância para o problema, ou até mesmo realizar detecção de tendências.

Vejam, a seguir, um exemplo para exemplificar melhor o uso do método.

Caso você tenha uma coleção de 500 pesquisas realizadas por uma universidade e deseje encontrar uma maneira de agrupá-las, o método de *Clustering* pode auxiliar fazendo o agrupamento a partir de grupos que podem ser semelhantes, ou não, e que estariam relacionados por algumas variáveis, como número de páginas, frequência de palavras, frases, entre outras.

Um outro exemplo é a aplicação do algoritmo *Cocktail Party*, que consiste encontrar um padrão com base em uma estrutura de dados desorganizada. Esse algoritmo também é conhecido como separação de sinais de áudio. Caso você tenha duas pessoas conversando em uma sala, ambas utilizando um microfone com captação de áudio, o algoritmo tentará agrupar (*Clustering*) as vozes das pessoas e, assim, dizer de quem é cada voz (SILVA *et al.*, 2007).

Por fim, o método do *Clustering* pode gerar saídas que podem nem ser explicáveis, posto que as relações entre os dados de entrada ainda serão criadas e analisadas. Dessa forma, vários *clusterings* (agrupamentos) podem ser criados e, assim, serão utilizados somente aqueles que respondem ao problema descrito.

Agora que você entendeu os principais métodos em *machine learning*, pode selecionar o método mais adequado para o contexto do problema, tendo como ponto principal a pergunta de negócio que será respondida e, a partir das análises encontradas com o método, poder gerar valor com essa informação. O método é apenas o meio para se encontrar a resposta da pergunta do problema informado.

Portanto, com base no que você já sabe dos algoritmos, tipos de aprendizado de máquina e classificação dos métodos utilizados em *machine learning*, vamos entender como funciona o fluxo do processo de aprendizado de máquina. Vamos lá!

### **Fluxo do processo de aprendizagem de máquina**

O processo de aprendizado de máquina possui um fluxo que consiste em fornecer etapas de forma sequencial, que provê um aprendizado de forma automatizada para que, assim,

consiga chegar às respostas de forma mais eficiente (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012, p. 89). Vide Figura 4.13:



Figura 4.13 - Processo do fluxo de aprendizagem de máquina

Fonte: Elaborada pelos autores.

O primeiro passo é ter um banco de dados que possua uma certa qualidade dos dados, pois isso influencia nas respostas. Para tanto, pode ser necessário um pré-processamento dos dados, para que possa ser realizado um tratamento, em especial para dados de diferentes fontes, com diferentes formatos, desconhecidos ou incompletos. Se o banco de dados for automatizado, erros básicos serão evitados, como erros de digitação, valores errados, entre outros. Há uma relação direta entre a quantidade de dados de qualidade e a precisão da aprendizagem, pois, quanto maior for a quantidade desses dados, maior será a aprendizagem do sistema e com muito mais precisão.

Em seguida, os dados são separados em duas amostras: uma base para treinamento e outra para verificação do desempenho do modelo treinado. Na base de treinamento, os dados serão analisados, e o sistema tenta encontrar padrões nas relações (existentes ou não) entre

as variáveis e passa a aprender com esses dados. Aqui, a máquina está em processo inicial de treinamento e tem os seguintes objetivos: o que e como procurar essas relações, o que pode ser encontrado e como irá encontrar o resultado.

Por conseguinte, selecionando o modelo (algoritmo e o método adequado), a máquina é capaz de realizar previsões. Nesse momento, ela realiza apenas uma previsão, pois está aprendendo. O treinamento ocorre por meio de ciclos iterativos, aprendendo com os erros, sendo que cada resultado gerado pode ser ajustado e melhorado para que a máquina se aperfeiçoe em suas previsões e seja mais preciso em suas avaliações. A base de verificação, que contém dados não conhecidos e diferentes da base de treinamento, deve ser utilizada para avaliar a eficácia do modelo, que tem como objetivo atingir um nível de acerto almejado.

Por fim, quando a máquina passa a gerar resultados mais precisos e confiáveis com base nesse fluxo e sem a interferência de melhorias em seus resultados, tem-se que a máquina de aprendizagem está utilizando o processo de forma automatizada.

## ATIVIDADE

6) Analise a situação a seguir:

“A partir de dados de animais em um zoológico, deve-se aproximar animais por suas características, ou seja, a partir dos dados como ‘quantidade de pernas’, ‘quantidade de dentes’, ‘põe ovo’, ‘tem pelos’ e vários outros, procuramos animais que estão mais próximos. Poderíamos, assim, separar animais em mamíferos, aves ou répteis, mas sem ‘contar’ ao algoritmo sobre estas classificações. Apenas comparando a distância entre dados o algoritmo mostraria que um tigre está “mais próximo” de um leão do que de uma garça” (HONDA, 2017, *on-line*)

Nesse contexto, pode-se utilizar que tipo de método estatístico de aprendizado de máquina?

- a) Regressão.
- b) Classificação.
- c) Supervisionado.
- d) Aprendizagem por reforço.
- e) Clusterização.

## **Aprendizado de máquina - Tecnologias/conhecimentos envolvidos**

Uma das tecnologias envolvidas na área de *Machine Learning* é o que está sendo denominada como *Deep Learning*. Na seção seguinte, serão apresentados a você a definição, o funcionamento e alguns aspectos relevantes dentro dessa área. Vamos começar!

### ***Deep learning***

Em português, o termo é traduzido como aprendizagem profunda, a qual é uma subcategoria do aprendizado de máquina e tem como objetivo realizar aprendizagens profundas e extensas com o uso de redes neurais (DEEP LEARNING, *on-line*).

Outra definição pode ser descrita como o treinamento de um modelo da computação que utiliza algoritmos para realizar o processamento dos dados e imitar esse processamento como se fosse um cérebro humano (DEEP LEARNING, *on-line*).

Os algoritmos utilizados pela *deep learning* são bem mais complexos dos que os que são empregados em *machine learning* e estão baseados nos princípios das redes neurais (como já vimos nesta unidade) para que possa simular o processamento de um cérebro humano com muito mais fidelidade e precisão, no que se refere à forma como as informações são compreendidas e quais resultados são gerados a partir dessas informações (DEEP LEARNING, *on-line*).

Outras tarefas inerentes aos seres humanos podem ser realizadas com a *deep learning*, como: reconhecimento de fala, percepção, detecção de imagens, realizar previsões, entre outros. Para isso, a *deep learning* realiza configuração de alguns parâmetros sobre os dados de entrada e faz com que a máquina seja treinada para que aprenda a reconhecer, sozinha, todos os padrões estabelecidos em várias camadas de processamento.

Vale ressaltar que tanto a *deep learning* quanto o *machine learning* são os pilares da Inteligência Artificial (IA) e que fazem com que a área de IA esteja evoluindo e se

propagando de forma rápida e que viabilizam a aplicação de IA no nosso dia a dia. A Figura 4.14 expõe essa relação entre as áreas (DEEP LEARNING, *on-line*):

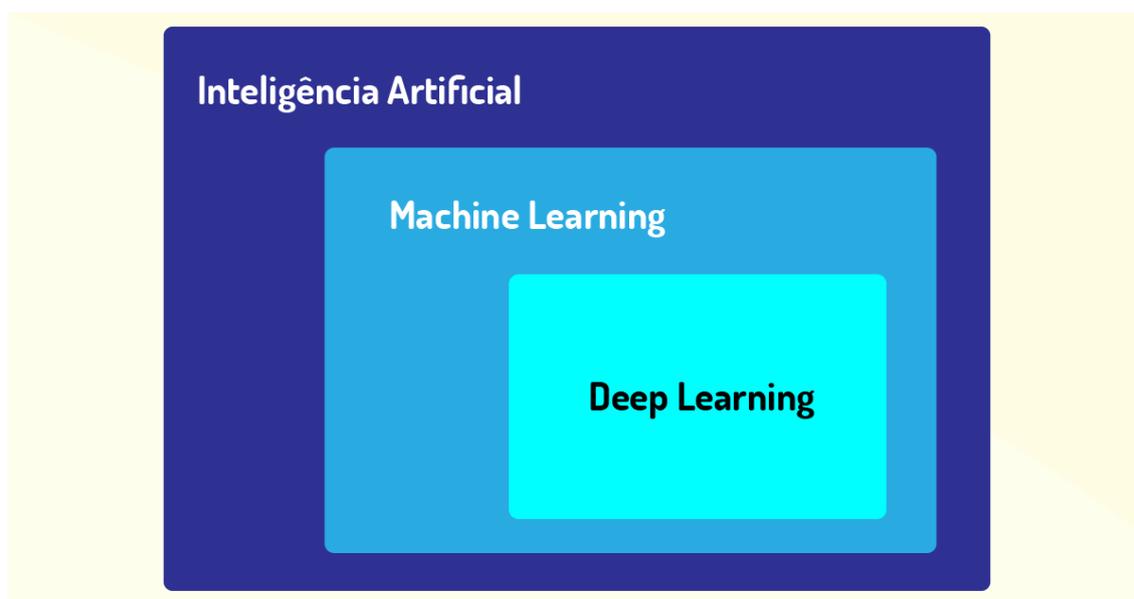


Figura 4.14 - Relacionamento entre as áreas

Fonte: Elaborada pelos autores.

Conforme se percebe na imagem, a área de IA engloba as duas outras áreas e que *deep learning* é uma subárea dentro de *machine learning*, já que, como o próprio nome diz, é uma aprendizagem profunda do aprendizado de máquina.

E como funciona a *deep learning*? Após receber uma grande quantidade de dados, a *deep learning* utiliza um modelo computacional para relacionar os termos e as palavras encontrados na análise dos dados para inferir significados e, assim, decifrar a linguagem natural (DEEP LEARNING, *on-line*).

Esse modelo computacional possui uma série de camadas de neurônios que são utilizados para diversos fins, como processar os dados e reconhecer a fala humana e objetos de forma visual. Assim, a primeira camada de neurônio (de entrada) recebe a informação e passa para as camadas seguintes, denominadas de camadas ocultas que recebem a

informação da camada anterior, realizam o processamento e repassam para a camada seguinte, até gerarem um resultado na camada de saída (DEEP LEARNING, *on-line*).

Cada camada possui um algoritmo simples para realizar o processamento computacional e possui uma determinada função, que lhe foi atribuída para que possam ser treinadas e, assim, gerarem um resultado esperado. Vale ressaltar que uma rede neural simples possui apenas uma camada oculta, enquanto a *deep learning* possui N camadas ocultas, conforme podemos visualizar na Figura 4.15 (DEEP LEARNING, *on-line*):

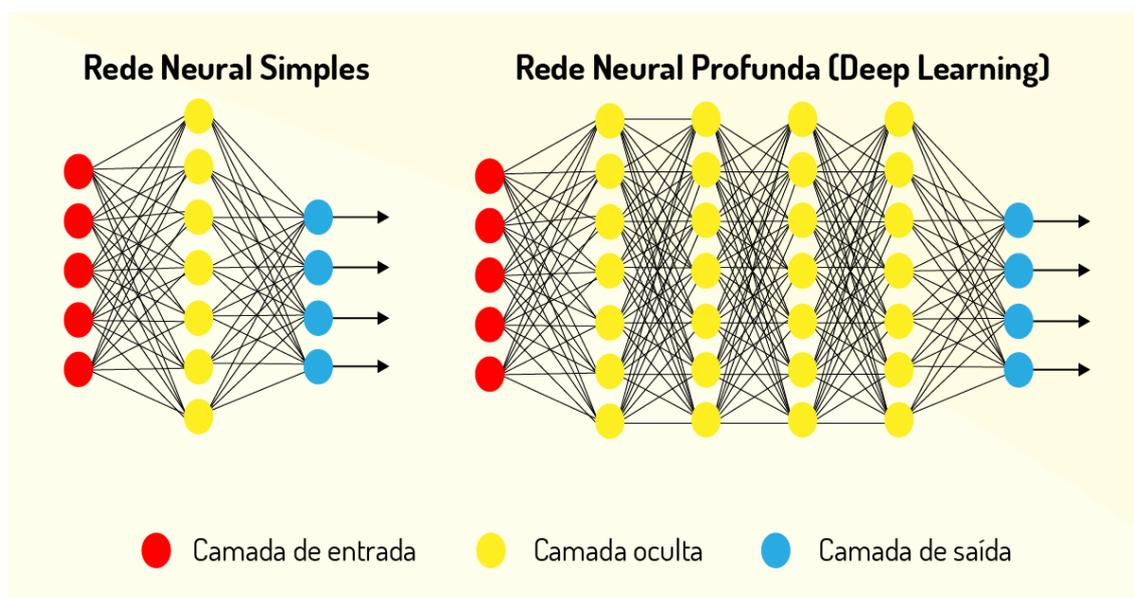


Figura 4.15 - Diferenças entre redes neurais

Fonte: Adaptada de Deep Learning (*on-line*).

Por meio da Figura 4.15, pudemos descrever melhor que os dados são submetidos a diversas camadas não lineares de processamento. Esses neurônios artificiais das camadas de *deep learning* são capazes de realizar diversas atividades humanas e complexas, sem a necessidade da intervenção de um humano neste processamento.

Além de se basear em redes neurais simples, outra característica importante da *deep learning* é a extração de recursos dos dados brutos, a qual utiliza um algoritmo para construir, de forma automática, os recursos essenciais dos dados para que seja possível treinar e prover aprendizado e compreensão da rede neural. Esses recursos são aprendidos de forma hierárquica em camadas, sendo que na camada inicial são aprendidos os recursos de baixo nível, obtendo um nível de abstração maior à medida que sobe na hierarquia. Na Figura 4.16, temos um exemplo de extração de recursos, como valores médios de pixels, forma, textura, cor, posição e orientação e os respectivos níveis de abstração. Essa extração de recursos feita pelo algoritmo é supervisionada, seja pelo cientista de dados ou pelo engenheiro de IA.

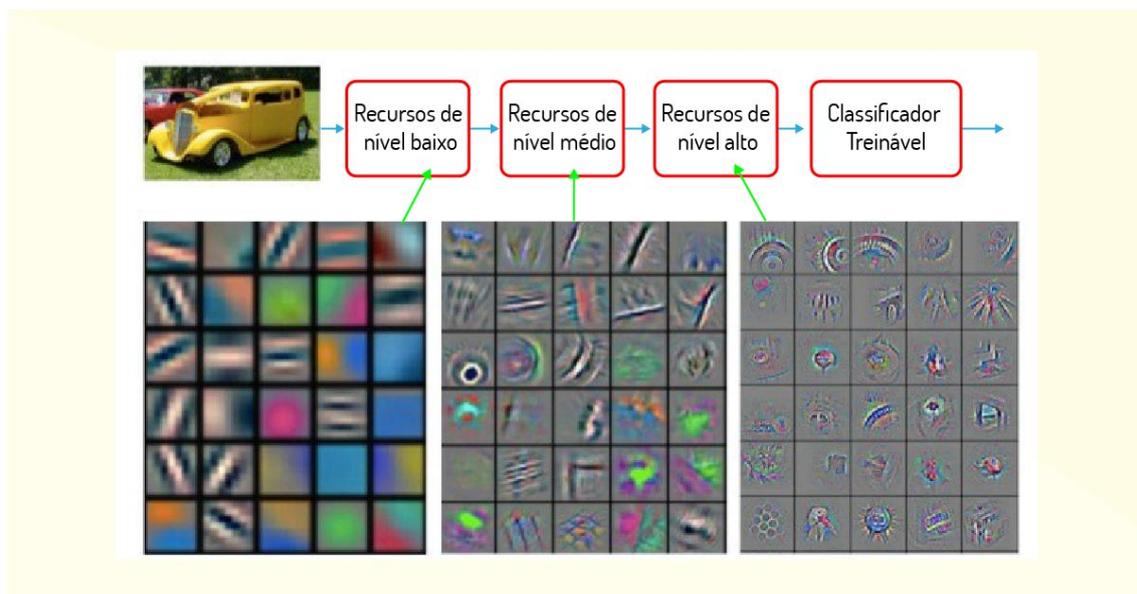


Figura 4.16 - Recursos de níveis baixo, médio e alto de abstração

Fonte: Adaptada de Prasad (2018).

Por conseguinte, um exemplo prático da *deep learning* foi realizado pela empresa Google em seu sistema de tradução. Com apenas aprendizado de máquina, a tradução era feita considerando apenas as palavras do texto, o que muitas das vezes não fazia sentido para

o contexto da frase. Então, para solucionar isso, a Google implementou o *Google Neural Machine Translation* (GNMT), com base na *deep learning*, que traduz frases completas e as adapta conforme o contexto do texto. Assim, o serviço de tradução tornou-se mais preciso e inteligente (DEEP LEARNING, *on-line*).

Por fim, podemos perceber que as duas áreas *Deep Learning* e *Machine Learning* tornaram-se a essência de processamento para que as máquinas se tornem cada vez mais inteligentes e possam evoluir de forma muito rápida e sem intervenção humana.

## **Ferramentas de aprendizado de máquina**

Vejamos, agora, algumas ferramentas que utilizam o aprendizado de máquina para prover funcionalidades, seja para aplicativos individuais, seja para soluções complexas.

**1 - Scikit-learn:** é uma biblioteca de aprendizagem baseada na linguagem de programação em Python, que permite construir aplicações sobre os pacotes disponíveis dessa linguagem, utilizando conceitos de matemática e estatística. A ferramenta está disponível com a licença BSD, com código aberto e reutilizável (SCIKIT-LEARN, *on-line*).

**2 - Shogun:** é uma das bibliotecas de aprendizado de máquina mais antigas que existem. A princípio, essa ferramenta poderia ser integrada com C++, mas, por meio de uma biblioteca denominada SWIG, ela pode ser utilizada junto com outras linguagens de programação como Java, Ruby, R, entre outras e ambientes como Matlab. Além disso, possui a biblioteca MIPack, que utiliza aprendizagem de máquina que faz com que seja mais eficaz e fácil de trabalhar, por possuir um conjunto mais completo de APIs (SHOGUN, *on-line*).

**3 - Accord Framework/AForge.net:** é um *framework* baseado na linguagem .net que utiliza diversos algoritmos de aprendizado de máquina e processamento de sinais a serem aplicados no processamento de áudios e imagens, os quais podem ser úteis para aplicações de reconhecimento facial (ACCORD.NET, *on-line*).

**4 - GoLearn:** é uma biblioteca da Google que se propõe a reunir todos os conceitos relativos à aprendizagem de máquina. Apesar de ser recente, já possui uma ampla utilização pelos programadores e tem uma constante evolução nas bibliotecas disponibilizadas. O principal objetivo dessa ferramenta é realizar o tratamento da inserção e manipulação dos dados na biblioteca de forma mais fácil (GOLEARN, *on-line*).

**5 - Weka:** é uma ferramenta que possui uma diversidade de algoritmos de aprendizado de máquina em Java, os quais são utilizados especificamente na mineração de dados. Está licenciada com GNU GPLv3 com um sistema de pacotes que podem estender suas funcionalidades por qualquer programador. Um diferencial dessa ferramenta é que ela disponibiliza um livro com o intuito de explicar o *software* e as técnicas utilizadas nas bibliotecas disponíveis (WEKA, *on-line*).

Por fim, temos uma pequena amostra das ferramentas que utilizam os algoritmos de aprendizado de máquina em diferentes linguagens para diversos contextos. Não existe a ferramenta correta, mas, sim, a ideal que vai se propor a solucionar da melhor maneira o problema que precisa ser resolvido dentro do aprendizado de máquina.

### **FIQUE POR DENTRO**

O uso da ferramenta e biblioteca adequada é importante para um projeto de aprendizado de máquina. Para tanto, é sugerida a busca de maiores detalhes sobre cada recurso disponível no mercado, para que comparações possam ser feitas tanto do ponto de vista técnico como no modelo de negócio.

Seguem os *links* com informações sobre as principais ferramentas e bibliotecas mencionadas nesta unidade:

- Scikit-learn: <https://scikit-learn.org/stable/>.
- Shogun: <https://www.shogun-toolbox.org/>.
- Accord Framework/AForge.net: <http://accord-framework.net/>.
- GoLearn: <https://www.techleer.com/articles/351-golearn-machine-learning-for-go/>.
- Weka: <https://www.cs.waikato.ac.nz/ml/weka/>.

## ATIVIDADE

7) É considerada uma subárea do aprendizado de máquina e tem como objetivo simular a rede neural do cérebro humano, por meio da utilização de algoritmos de alto nível. O contexto apresentado no enunciado da questão refere-se à(ao):

- a) inteligência artificial.
- b) *machine learning*.
- c) *deep learning*.
- d) ciência de dados.
- e) aprendizado por reforço.

## **Aprendizado de máquina - Exemplos de usos e cases**

Agora chegou o momento de apresentarmos alguns exemplos práticos utilizados por empresas, as quais estão se apropriando dos conhecimentos da área de *machine learning* para realizar melhorias ou, até mesmo, inovações em seus processos.

### **Case 1: General Eletric (GE, *on-line*)**

A General Eletric é uma empresa multinacional com sede global em Boston, nos EUA, tendo como objetivo desenvolver soluções para as áreas de: transporte, saúde, iluminação e energia.

Em seu processo produtivo automatizado por máquinas, as quais possuem gêmeas digitais que seriam as réplicas de forma virtual dos equipamentos físicos, o objetivo é simular o comportamento da máquina física e, assim, aprender os diferentes estados que as máquinas possuem mediante um contexto.

Se a máquina de verdade apresentar algo considerado fora do normal, seja um aumento de calor, emissão de ruídos e vibrações inconsistentes, isso é coletado por sensores na gêmea digital e armazenado na nuvem. A partir disso, o aprendizado de máquina é ativado, para determinar o que fazer mediante essa situação anormal.

Além disso, há uma simulação dessa situação anormal na gêmea digital para saber se a réplica induz que essa anormalidade pode acarretar alguma falha no processo produtivo ou até mesmo sugerir manutenções nas máquinas físicas.

### **Case 2: JPMorgan (JPMORGAN, *on-line*)**

O grupo JPMorgan é um banco americano e líder mundial em serviços financeiros e oferece esses serviços a governos, corporações e instituições em mais de 100 países. Os serviços oferecidos são: *investment bank*, *global corporate bank*, *asset management*, *private banking* e *treasury & securities services*. Para os clientes, o banco oferece uma

gama de serviços que se propõe a combinar conhecimento especializado com técnicas avançadas de liderança para os negócios do cliente.

Uma das grandes tarefas que não era automatizada nesse banco era a análise e interpretação de acordos realizados em empréstimos comerciais. Essa tarefa era realizada por advogados, pois era necessário analisar algumas cláusulas contratuais para que a empresa já se respaldasse mediante alguns processos realizados pelos clientes.

A empresa, então, resolveu criar um algoritmo baseado em *machine learning*, que realizou essa tarefa em segundos, quando os advogados humanos precisavam de mais ou menos 360 mil horas por ano. Esse foi um grande avanço de melhoria nessa rotina, e os advogados tendem a realizar outras rotinas de análise na empresa.

### **Case 3: Tati (TATI..., on-line)**

O aplicativo para *smartphones* nomeado como Tati tem o objetivo de fornecer uma série de recursos de estudos para os alunos, mediante suas dificuldades apresentadas no aplicativo.

Tati foi desenvolvida pela empresa Conexia Educação, que se propõe a oferecer uma gama de soluções educacionais pertencentes ao grupo SEB. O aplicativo tem seu funcionamento realizado por meio de computação cognitiva, a qual permite a simulação de um atendimento virtual como se fosse um humano e ainda auxilia no aprendizado dos estudantes no aplicativo.

Adicionalmente, Tati utiliza *machine learning* quando observa e analisa as quantidades de dados que lhe são fornecidas pelo aplicativo e, após aprender com esses dados, transmite o conhecimento que aprendeu a partir da relação encontrada, por exemplo: dificuldade do aluno e rotina de estudos para aumentar o aprendizado.

Hoje em dia, Tati realiza os atendimentos aos alunos do ensino médio, nas dificuldades de disciplinas como história, geografia, matemática entre outras, mas futuramente, poderá

ser integrada para alunos do ensino superior, por meio da descrição das competências e habilidades que devem ser analisadas por Tati nesse contexto.

#### **Case 4: Plataforma SAGAH (SAGAH, on-line)**

A SAGAH é uma ferramenta de cunho educacional desenvolvida pelo Grupo A, a qual provê soluções educacionais customizadas para as empresas e instituições de ensino. Possui um portfólio de negócios com o foco na educação, fazendo a integração de conteúdos, tecnologias e serviços.

Por meio do *machine learning*, a SAGAH possui uma plataforma adaptável para realizar o nivelamento dos seus estudantes em disciplinas como português e matemática. A plataforma reconhece as dificuldades dos alunos sobre os exercícios ofertados e, assim, fornece algumas recomendações para sanar tais dificuldades.

E como funciona? A plataforma disponibiliza uma lista de exercícios com o objetivo de identificar as principais dificuldades do aluno assim que ele inicia a graduação. Com base na pontuação do aluno (dados de entrada), o sistema analisa e procura uma recomendação que pretende diminuir esse *gap* de dificuldade com a disciplina da lista de exercício.

Um exemplo que podemos citar é: para o nivelamento de matemática, há atividades referentes aos assuntos logaritmo, exponencial e regra de 3. O aluno realiza esses exercícios e, se a máquina perceber que o aluno não acertou nenhuma das questões ou se está abaixo da média que precisa para o exercício, a máquina sugere uma série de vídeos e outros materiais que auxiliam o aluno.

O propósito da ferramenta é simular um professor que percebe que o aluno tem dificuldade em tal assunto e, assim, sugere alguns recursos educacionais para melhorar seu aprendizado. Vale ressaltar que a plataforma não pretende substituir os professores humanos, pois estes são necessários para explorar o potencial da tecnologia e, assim, poder implementar para cada aluno que tem dificuldade. Isso faz com que a aula se torne mais dinâmica e faz uma aproximação entre docente e discente, em que ambos têm

melhorias em seus desempenhos, seja para uma aula mais atrativa (para o professor) ou melhoria do aprendizado da disciplina (para o aluno com a aula mais dinâmica).

### **Case 5: Waze (WAZE, on-line)**

O aplicativo Waze tem o propósito de indicar as melhores rotas de caminho terrestre a partir das informações dos seus usuários. Ele já é bastante popular entre as pessoas que dirigem carros e cada vez mais as pessoas o fazem de um recurso principal no momento de escolher uma rota, principalmente em horários de *rush*.

A essência do funcionamento do aplicativo consiste em receber as informações de tráfego dos usuários em uma mesma rota e, assim, o aplicativo recomenda ou não, que se utilize essa via para se chegar ao seu destino.

Os humanos não conseguiriam, em tempo real, processar toda essa informação e fazer as devidas sugestões de melhor caminho, porém, os algoritmos conseguem realizar isso em tempo hábil, 24 horas por dia.

O aprendizado de máquina no Waze está relacionado ao fato de que o aplicativo tem a capacidade de aprender por conta própria as informações retiradas das análises dos dados e, assim, prover informação útil aos seus usuários.

O resultado do *machine learning* fornece resultados cada vez mais precisos, rápidos e eficientes, mesmo que se tenha uma grande massa de dados para serem analisados; conforme os usuários inserem informações no aplicativo, os riscos são bem menores do que se fossem realizados exclusivamente por seres humanos.

Agora que já vimos alguns cases de sucesso, vejamos alguns exemplos de usos de aprendizado de máquina, que podem ser aplicados em diversas áreas.

**1. Jogos:** anteriormente, os jogos funcionavam de forma bidimensional, ou seja, os usuários realizam o mesmo caminho para conseguir chegar ao objetivo final do jogo.

Agora, com o *machine learning*, são utilizados algoritmos de alta complexidade para compreender as preferências de cada jogador e propõem uma jornada exclusiva no jogo.

**2. Softwares e programas operacionais:** cada usuário utiliza o seu *smartphone* de uma forma diferenciada e tem sua própria experiência de uso. Com o *machine learning*, algumas personalizações já são oferecidas aos usuários, mas isso ainda está sendo aperfeiçoado, e o objetivo é que as preferências do usuário sejam realizadas pela máquina que entendem suas necessidades, de forma que o usuário não precisa realizar nenhum tipo de configuração.

**3. Finanças:** nesse contexto, o aprendizado de máquina utiliza os algoritmos de aprendizado para realizar previsões sobre cenários econômicos futuros para que instituições financeiras possam se adaptar rapidamente às mudanças do mercado financeiro.

**4. Segurança:** uma das dificuldades nesta área é o reconhecimento facial pelos vídeos realizados no monitoramento da segurança. Com os algoritmos do aprendizado de máquina, é possível reconhecer os rostos com base no padrão fornecido ou até mesmo reconhecer comportamentos e atividades ilícitas, conforme a lei.

**5. Carros autônomos:** a atividade de dirigir um carro exige uma série de ações e inteligência, sendo um processo de aprendizado contínuo. Para saber dirigir bem, é necessário que o condutor tenha a prática de executar as mesmas ações para diferentes contextos na direção. É com base nisso que os carros autônomos estão sendo programados, utilizando as abordagens de inteligência artificial para simular um condutor humano e o aprendizado de máquina para aprender as atividades inerentes ao ato de dirigir um carro com a utilização de diversos tipos de algoritmos e aprendizado por reforço.

Aqui, vimos muitos exemplos de cases e usos relacionados ao aprendizado de máquina. É importante frisar que o mundo está cada vez mais tecnológico, e aquele que detém a tecnologia a seu favor está dando um passo mais adiante no processo de inovação tecnológica e, por mais que as máquinas substituam muitas tarefas realizadas exclusivamente por seres humanos, serão necessários humanos que saibam descobrir as melhores oportunidades no benefício da tecnologia.

## ATIVIDADE

8) Em sistemas de recomendação de compras *on-line*, são utilizados mecanismos de aprendizado de máquina. É quase inacreditável o fato de que o programa sabe qual produto você tem mais chances de comprar! De que forma esses sistemas estão utilizando aprendizado de máquina?

- a) Por meio do histórico de compras anteriores e visualização de determinados produtos.
- b) Criando uma rede neural para entender o perfil do consumidor.
- c) Simulando um consumidor que realiza compras *on-line*.
- d) Analisando grande quantidade de dados de perfis de consumidores de produtos *on-line*.
- e) Com base na forma de pagamento do consumidor.

## **INDICAÇÃO DE LEITURA**

Nome do livro: Aprendizado de máquina para leigos

Editora: Alta Books

Autores: John Paul Mueller e Luca Massaron

ISBN: 8550802344

Comentário: Este livro é bem didático para quem está iniciando na área de aprendizado de máquina, com discussões explicadas sobre o funcionamento dos algoritmos básicos de aprendizado de máquina, utilizando as linguagens Python e R. O livro apresenta vários exemplos práticos que aplicam a utilização desses algoritmos.

## CONCLUSÃO DO LIVRO

Chegamos ao fim da disciplina Tópicos Especiais em Sistemas de Informação e esperamos que você tenha assimilado todos os conceitos e as temáticas trabalhados nas quatro unidades.

Na primeira unidade, vimos as definições, as características, as ferramentas e as aplicações na área de *Big Data*. Também, contextualizamos alguns problemas nessa área e possíveis soluções para eles, com bases nos algoritmos utilizados para análise de dados. Para o desenvolvimento das soluções, foram utilizadas as linguagens de programação R e *Python* e as ferramentas Excel, *Hadoop* e *Spark*.

Na unidade 2, foram abordados alguns exemplos práticos de aplicação de conhecimentos da área de IoT, a fim de proporcionar melhorias e inovações em seus processos. Nesse sentido, quanto mais conhecermos sobre esta área, mais isso ajudará no desenvolvimento de interatividade de dispositivos conectados. Para isso, também foram apresentados vários *cases* de aplicação e suas respectivas considerações.

Na unidade 3, por sua vez, abordamos importantes fatores da Inteligência Artificial e os resultados positivos nas organizações que a utilizam. Refletimos, também, acerca da agilidade, da eficiência, da qualidade e da rapidez nos processos em que a inteligência artificial está inserida, promovendo uma melhora significativa em empresas que a utilizam.

Por fim, na última unidade de nossa disciplina, apresentamos diversos exemplos de *cases* e refletimos a respeito dos benefícios das inovações tecnológicas, bem como do conhecimento humano atrelado a essa tecnologia.

Assim, esperamos que você possa ter aprofundado seus conhecimentos sobre as temáticas que estudamos nesta disciplina e que possa ter refletido acerca desses tópicos dos sistemas de informação.

## REFERÊNCIAS UNIDADE 1

AMARAL, F. **Introdução à Ciência de Dados: mineração de dados e big data**. Rio de Janeiro: Alta Books, 2016.

BISPO, C. A. F.; CAZARINI, E. W. **Conceitos básicos e a elaboração do projeto lógico de um data warehouse**. Niterói: UFF/TEP, 1998. [CD-ROM].

CALDAS, M. S.; SILVA, E. C. C. Fundamentos e Aplicações do Big Data: como tratar informações em uma sociedade de yottabytes. **Bibliotecas Universitárias: pesquisas, experiências e perspectivas**, Belo Horizonte, v. 3, n.1, p. 65-85, jan./jun. 2016. Disponível em: <https://periodicos.ufmg.br/index.php/revistarbu/article/view/3086/1886>. Acesso em: 10 dez. 2019.

DAVENPORT, T. H. **Ecologia da Informação: por que só a tecnologia não basta para o sucesso na era da informação**. Tradução de Bernadette Siqueira Abrão. São Paulo: Futura, 1998. Disponível em: <https://ppgic.files.wordpress.com/2018/07/davenport-t-h-2002.pdf>. Acesso em: 10 dez. 2019.

DIANA, M. D.; GEROSA, M. A. NOSQL na Web 2.0: um estudo comparativo de Bancos não-relacionais para gerenciamento de dados na Web 2.0. *In: WORKSHOP DE TESES E DISSERTAÇÕES EM BANCO DE DADOS*, 9., 2010, Belo Horizonte. **Anais [...]**. Belo Horizonte: UFMG, 2010. p. 68-75. Disponível em: [http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2010/sbbd\\_wtd\\_12.pdf](http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2010/sbbd_wtd_12.pdf). Acesso em: 10 dez. 2019.

ELIAS, D. A abordagem Top-Down e Bottom-Up no Data Warehouse. **Canaltech**, 2014. Disponível em: <https://canaltech.com.br/infra/a-abordagem-top-down-e-bottom-up-no-data-warehouse-21108/>. Acesso em: 23 Ago. 2019.

FAWCETT, T.; PROVOST, F. **Data Science para negócios**: O que você precisa saber sobre mineração de dados e pensamento analítico de dados. Rio de Janeiro: Alta Books, 2016.

GALDINO, N. Big Data: Ferramentas e Aplicabilidade. *In*: SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA, 13., 2016, Resende-RJ. **Anais** [...]. Resende-RJ: Associação Educacional Dom Bosco, 2016, p. 1-12. Disponível em: <https://www.aedb.br/seget/arquivos/artigos16/472427.pdf>. Acesso em: 10 dez. 2019.

LUCAS, A.; VIEIRA, A. F. G. Business Intelligence e a Ciência da Informação na Era dos Dados Abertos e Big Data. *In*: ENCONTRO IBÉRICO EDICIC, 9., 2019, Barcelona - Espanha. **Anais** [...]. Barcelona - Espanha: Faculdade de Biblioteconomia e Documentação da Universidade de Barcelona, 2019.

MATOS, D. Data Lake, a fonte do Big Data. **Ciência e Dados**, 14 nov. 2015. Disponível em: <http://www.cienciaedados.com/data-lake-a-fonte-do-big-data/>. Acesso em: 24 Agosto 2019.

MATOS, D. Top 6 NoSQL Database. **Ciência e Dados**, 08 fev. 2019. Disponível em: <http://www.cienciaedados.com/top-6-nosql-databases/>. Acesso em: 27 Agosto 2019.

MOURA, K. V. **Data Science**: Um estudo dos métodos no mercado e na academia. 2018. Trabalho de conclusão de curso (Graduação em Administração) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018. Disponível em: <https://lume.ufrgs.br/handle/10183/195011>. Acesso em: 4 set. 2009.

OLIVEIRA, J. P. M. D. Dados, informação e conhecimento. **Site do Prof. Palazzo**: Comunicação, Tecnologia e Humanismo, 2016. Disponível em: <https://www.palazzo.pro.br/Wordpress/?p=156>. Acesso em: 12 ago. 2019.

PIRES, F. M. M. **Data Lake em Viticultura**: big data management na agricultura. 2017. Dissertação (Mestrado em Gestão da Informação) - Instituto Superior de Estatística e

Gestão da Informação, Universidade Nova de Lisboa, Lisboa - Portugal, 2017. Disponível em: <https://run.unl.pt/bitstream/10362/33951/1/TGI0138.pdf>. Acesso em: 4 set. 2019.

SOBREIRO, S. A. R. **Estudo de tecnologias para sistemas de Big Data**. 2018. Dissertação (Mestrado em Engenharia Informática) – Instituto Superior de Engenharia do Porto, Porto - Portugal, 2018. Disponível em: [https://recipp.ipp.pt/bitstream/10400.22/11936/1/DM\\_SauloSobreiro\\_2018\\_MEI.pdf](https://recipp.ipp.pt/bitstream/10400.22/11936/1/DM_SauloSobreiro_2018_MEI.pdf). Acesso em: 11 dez. 2019.

TAURION, C. **Big Data**. Rio de Janeiro: Brasport, 2013.

WHITE, T. **Hadoop: The Definitive Guide**. 4. ed. Sebastopol: O'Reilly, 2015. Disponível em: <http://grut-computing.com/HadoopBook.pdf>. Acesso em: 11 dez. 2019.

## REFERÊNCIAS UNIDADE 2

ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: A survey. **Computer Networks**, v. 54, n. 15, 2010, p. 2787-2805.

BARROS, T. O que falta na internet para as coisas? **Cesar**, 2016. Disponível em: <https://www.cesar.org.br/index.php/2016/09/14/o-que-falta-na-internet-para-as-coisas/>. Acesso em: 27 dez. 2019.

BELLIAS, M. Connected trains: How IoT is driving the future of rail. **IBM**, 2016. Disponível em: <https://www.ibm.com/blogs/internet-of-things/connected-trains-rail-travel/>. Acesso em: 27 dez. 2019.

EUROPEAN COMMISSION. **What is Horizon 2020?** 2019. Disponível em: <http://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>. Acesso em: 10 set. 2019.

EVANS, D. A internet das coisas. Como a próxima evolução da Internet está. **Cisco**, 2011. Disponível em: [https://www.cisco.com/c/dam/global/pt\\_br/assets/executives/pdf/internet\\_of\\_things\\_iot\\_ibsg\\_0411final.pdf](https://www.cisco.com/c/dam/global/pt_br/assets/executives/pdf/internet_of_things_iot_ibsg_0411final.pdf). Acesso em: 27 dez. 2019.

IOT: Nokia e Altran apresentam solução para manutenção de trens. **TeleSÍntese**, 2018. Disponível em: <http://www.telesintese.com.br/iot-nokia-e-altran-apresentam-solucao-para-manutencao-de-trens/>. Acesso em: 27 dez. 2019.

RAM, P. LPWAN, LoRa, LoRaWAN and the Internet of Things. **Medium**, 2018. Disponível em: <https://medium.com/coinmonks/lpwan-lora-lorawan-and-the-internet-of-things-aed7d5975d5d>. Acesso em: 27 dez. 2019.

SINCLAIR, B. **IoT**: como usar a "Internet das Coisas" para alavancar seus negócios. São Paulo: Autêntica Business, 2018.

SANTOS, S. **Introdução à IoT**. Desvendando a internet das coisas. São Paulo: Clube de Autores, 2018.

SOUZA, R. G. **Sistema de automação de baixo custo**. Artigo Científico (Tecnologia em Mecatrônica Industrial) – Faculdade de Tecnologia de Garça, Garça, 2016. Disponível em: <http://fatecgarca.edu.br/uploads/documentos/tcc/monografias/mecatronica/2016-1/Ruan%20Giovani%20de%20Souza%3B%20Vinicius%20dos%20Santos%20-%20Automa%C3%A7%C3%A3o%20residencial%20de%20baixo%20custo.pdf>.

Acesso em: 19 fev. 2020.

VALENTE, F. J.; COLENCI NETO, A. Internet das coisas na manufatura avançada: caso da produção de mudas de cana de açúcar. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 37., 2017, Joinville. **Anais [...]**. Joinville: Enegep, 2017. Disponível em: [http://www.abepro.org.br/biblioteca/TN\\_STO\\_238\\_379\\_32741.pdf](http://www.abepro.org.br/biblioteca/TN_STO_238_379_32741.pdf). Acesso em: 19 fev. 2020.

### REFERÊNCIAS UNIDADE 3

ABDALLAH, C. Algoritmo genético prevê o crescimento vertical de áreas urbanas. **Archdaily**, 2018. Disponível em: <https://www.archdaily.com.br/br/897187/algoritmo-generico-preve-o-crescimento-vertical-de-areas-urbanas>. Acesso em: 19 fev. 2020.

ALAN Turing: **The Enigma**. Disponível em: <https://www.turing.org.uk/index.html>. Acesso em: 19 fev. 2020.

ALVES, C. Carreiras são impactadas pela inteligência artificial. **IP News**, 2019. Disponível em: <https://ipnews.com.br/carreiras-sao-impactadas-pela-inteligencia-artificial/>. Acesso em: 19 fev. 2020.

ASSIS, P. O que são Redes Neurais? **Tecmundo**, 2009. Disponível em: <https://www.tecmundo.com.br/programacao/2754-o-que-sao-redes-neurais-.htm>. Acesso em: 19 fev. 2020.

CARVALHO, A. P. L. F. **Algoritmos genéticos**. Disponível em: <http://conteudo.icmc.usp.br/pessoas/andre/research/genetic/>. Acesso em: 19 fev. 2020.

DEEP LEARNING. **Capítulo 2** – Uma breve história das redes neurais artificiais. Disponível em: [deeplearningbook.com.br/uma-breve-historia-das-redes-neurais-artificiais/](https://deeplearningbook.com.br/uma-breve-historia-das-redes-neurais-artificiais/). Acesso em: 19 fev. 2020.

FERNANDES, A. M. R. **Inteligência artificial**. Florianópolis: Visual Books, 2003.

GILLIARD, L.; PACHECO, M. **Evoluindo o comportamento de agentes inteligentes em jogos eletrônicos**. Disponível em: <https://docplayer.com.br/31589789-Evoluindo-o-comportamento-de-agentes-inteligentes-em-jogos-eletronicos.html>. Acesso em: 19 fev. 2020.

GOMES, D. S. M.; RODRIGUES, M. C. **Introdução à Lógica Fuzzy com Java**. Disponível em: <https://www.devmedia.com.br/introducao-a-logica-fuzzy-com-java/32444>. Acesso em: 19 fev. 2020.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

IBM WATSON. **IBM Watson for Oncology**. Disponível em: <https://www.ibm.com/br-pt/marketplace/clinical-decision-support-oncology>. Acesso em: 19 fev. 2020.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação gerenciais**. São Paulo: Pearson, 2015.

LUGER, G. F. **Inteligência artificial**. 6. ed. São Paulo: Pearson, 2013.

MARRO, A. A. *et al.* **Lógica fuzzy: conceitos e aplicações**. Departamento de Informática e Matemática Aplicada. Natal: Universidade Federal do Rio Grande do Norte, 2010. Disponível em: [http://aquilesburlamaqui.wdfiles.com/local--files/logica-aplicada-a-computacao/texto\\_fuzzy.pdf](http://aquilesburlamaqui.wdfiles.com/local--files/logica-aplicada-a-computacao/texto_fuzzy.pdf). Acesso em: 19 fev. 2020.

MARTINS, A. L. Potenciais aplicações da Inteligência Artificial na Ciência da Informação. **Informação & Informação**, Londrina, v. 15, n. 1, 2010, p. 1-16. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/3882/5876>. Acesso em: 20 fev. 2020.

MEDEIROS, L. F. **Inteligência artificial: uma abordagem introdutória**. Curitiba: Intersaberes, 2018.

MELO NETO, A. **Lógica fuzzy**. Disponível em: <https://www.ime.usp.br/~adao/LOGICAFUZZY2017F.pdf>. Acesso em: 19 fev. 2020.

MIRANDA, M. N. **Algoritmos genéticos: fundamentos e aplicações**. UFRJ, 2019. Disponível em: <http://www.nce.ufrj.br/GINAPE/VIDA/alggenet.htm>. Acesso em: 19 fev. 2020.

NOSÉ, W. Como a inteligência artificial pode melhorar nossa saúde. **Revista Saúde**, 2019. Disponível em: <https://saude.abril.com.br/blog/com-a-palavra/como-a-inteligencia-artificial-pode-melhorar-nossa-saude/>. Acesso em: 19 fev. 2020.

O'BRIEN, A. J.; MARAKAS, G. M. **Administração de sistemas de informação: uma introdução**. São Paulo: Saraiva, 2008.

OS BENEFÍCIOS da inteligência artificial. **TOTVS**, 17 fev. 2020. Disponível em: <https://www.totvs.com/blog/inteligencia-artificial/>. Acesso em: 19 fev. 2020.

RAHDE, A. Assistentes virtuais: confira índices de acertos e de erros. **Vivo Tech**, 2019. Disponível em: <https://www.vivotech.com.br/assistentes-virtuais-acertos-erros/>. Acesso em: 19 fev. 2020.

REDES neurais: o que são e qual sua importância? **SAS**, 2019. Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/neural-networks.html](https://www.sas.com/pt_br/insights/analytics/neural-networks.html). Acesso em: 19 fev. 2020.

ROSS Intelligence. Disponível em: <https://rossintelligence.com/>. Acesso em: 19 fev. 2020.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013.

SELVAMANIKKAM, M. **Introduction to Artificial Intelligence**. 2018. Disponível em: <https://becominghuman.ai/introduction-to-artificial-intelligence-5fba0148ec99>. Acesso em: 19 fev. 2020.

SIMÕES, M. G. **Controle e modelagem fuzzy**. 2. ed. São Paulo: Blucher, 2007.

STEFANINI. Como usar inteligência artificial na educação? **Stefanini Group**, 2019. Disponível em: <https://stefanini.com/pt-br/trends/cases/como-usar-inteligencia-artificial-na-educacao->. Acesso em: 19 fev. 2020.

TEIXEIRA, J. F. **Inteligência artificial**. Recife: Paulus, 2009.

WAR of Secrets: Cryptology in WWII. **National Museum of the United States Air Force**, 2015. Disponível em: <https://www.nationalmuseum.af.mil/Visit/Museum-Exhibits/Fact-Sheets/Display/Article/196193/war-of-secrets-cryptology-in-wwii/>. Acesso em: 19 fev. 2020.

YU, K.; BEAM, A.; KOHANE, I. Artificial intelligence in healthcare. **Nature Biomed. Eng.**, v. 2, p. 719-731, 2018.

YUGE, C. Assistente digital Alexa está rindo sozinha e aterrorizando os usuários. **Tecmundo**, 2018. Disponível em: <https://www.tecmundo.com.br/produto/127948-assistente-digital-alexa-rindo-sozinha-ateterrorizando-usuarios.htm>. Acesso em: 19 fev. 2020.

ZADEH, L. A. Fuzzy Sets as a Basis for a Theory of Possibility. **Fuzzy Sets and Systems**, v. 1. 1976. p. 3-28.

## REFERÊNCIAS UNIDADE 4

3 EXEMPLOS do uso da data science para resolver problemas. **MasterTech**, 2019. Disponível em: <https://blog.mastertech.com.br/tecnologia/3-exemplos-do-uso-da-data-science-para-resolver-problemas/>. Acesso em: 20 fev. 2020.

ACCORD.NET. **Machine learning made in a minute**. Disponível em: <http://accord-framework.net/>. Acesso em: 20 fev. 2020.

AEROSOLVE - A machine learning package built for humans. **Airbnb**. Disponível em: <https://airbnb.io/aerosolve/>. Acesso em: 20 fev. 2020.

AIRBNB. **Detalhes da empresa**. Disponível em: <https://www.airbnb.com.br/about/company-details>. Acesso em: 20 fev. 2020.

AMARAL, F. **Introdução à ciência de dados - Mineração de dados e Big Data**. Rio de Janeiro: Alta Books, 2016.

AZAM, A. The first rule of data science. **Berkeley**, 2014. Disponível em: <http://berkeleysciencereview.com/article/first-rule-data-science/>. Acesso em: 20 fev. 2020.

BAUDISCH, A. R. Ciência de Dados é explorar Big Data para fazer perguntas para prever o futuro. **Medium**, 2016. Disponível em: <https://medium.com/@AlfredBaudisch/o-que-%C3%A9-ci%C3%Aancia-de-dados-data-science-7af5bdac101a>. Acesso em: 20 fev. 2020.

COELHO, L. Ciência de dados: o que é, conceito e definição. **Cetax**, 2019. Disponível em: <https://www.cetax.com.br/blog/data-science-ou-ciencia-de-dados/>. Acesso em: 20 fev. 2020.

COMO a análise de dados na Netflix construiu o sucesso da empresa? **TecMundo**, 2018. Disponível em: <https://www.tecmundo.com.br/internet/131307-analise-dados-netflix-construiu-sucesso-empresa.htm>. Acesso em: 20 fev. 2020.

DEEP LEARNING. **Capítulo 3** – O que são redes neurais artificiais profundas ou deep learning? Disponível em: <http://deeplearningbook.com.br/o-que-sao-redes-neurais-artificiais-profundas/>. Acesso em: 20 fev. 2020.

EDUREKA. **Pig Vs Hive**. 2019. Disponível em: <https://www.edureka.co/blog/pig-vs-hive/>. Acesso em: 20 fev. 2020.

ELIZA. **Masswerk**. Disponível em: <https://www.masswerk.at/elizabot/>. Acesso em: 20 fev. 2020.

GALLI, E. Netflix: a união do Big Data e da criatividade. **Canal Tech**, 2017. Disponível em: <https://canaltech.com.br/big-data/netflix-a-uniao-do-big-data-e-da-criatividade-101536>. Acesso em: 20 fev. 2020.

GE DIGITAL. **Machine Learning and Analytics**. Disponível em: <https://www.ge.com/digital/iiot-platform/machine-learning-analytics>. Acesso em: 20 fev. 2020.

GOLEARN. Disponível em: <https://github.com/sjwhitworth/golearn>. Acesso em: 20 fev. 2020.

GOUVEIA, L. B. **A linguagem R**: um ambiente para explorar dados e aprender com eles. Conferência Hello World, Universidade Fernando Pessoa, 2017. Disponível em: [https://bdigital.ufp.pt/bitstream/10284/5937/1/ufp\\_hello3maio2017.pdf](https://bdigital.ufp.pt/bitstream/10284/5937/1/ufp_hello3maio2017.pdf). Acesso em: 20 fev. 2020.

PORTAL GSTI. **O que é Ciência de Dados (Data Science)**. Disponível em: <https://www.portalgsti.com.br/ciencia-de-dados/sobre/>, Acesso em: 26.02.2020.

HONDA, H. **Introdução básica à clusterização**. 2017. Disponível em: [https://lamfo-unb.github.io/2017/10/05/Introducao\\_basica\\_a\\_clusterizacao/](https://lamfo-unb.github.io/2017/10/05/Introducao_basica_a_clusterizacao/). Acesso em: 20 fev. 2020.

IMPACT on solidary micro-credit. **Bayes**. Disponível em: [https://www.bayesimpact.org/en/focus/micro\\_credit](https://www.bayesimpact.org/en/focus/micro_credit). Acesso em: 20 fev. 2020.

ILUMEO. **Como funciona o departamento de Data Science do Airbnb?** Disponível em: <https://www.ilumeo.com.br/todos-posts/2018/11/5/como-funciona-o-departamento-de-data-science-do-airbnb>. Acesso em: 20 fev. 2020.

JPMORGAN. **Innovations in Finance with Machine Learning, Big Data and Artificial Intelligence**. Disponível em: <https://www.jpmorgan.com/global/research/machine-learning>. Acesso em: 20 fev. 2020.

MAGNUS, T. O que é machine learning e como funciona? **Transformação Digital**, 2018. Disponível em: <https://transformacaodigital.com/o-que-e-machine-learning-e-como-funciona/>. Acesso em: 20 fev. 2020.

MATOS, D. R ou Python para Análise de Dados? **Ciência e Dados**, 2018. Disponível em: <http://www.cienciaedados.com/r-ou-python-para-analise-de-dados/>. Acesso em: 20 fev. 2020.

MERCKGROUP. **Quem somos**. Disponível em: <https://www.merckgroup.com/br-pt/company/who-we-are.html>. Acesso em: 20 fev. 2020.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of machine learning**. Cambridge, Massachusetts: The MIT Press, 2012.

MSPERLIN. **Introdução**. Disponível em: <https://www.msperlin.com/padfeR/introducao.html#o-que-e-o-r>. Acesso em: 20 fev. 2020.

OLIVEIRA, P. F.; GUERRA, S.; MCDONNELL, R. **Ciência de dados com R**. Brasília: IBPAD, 2018.

PRASAD, H. Interesting Deep Learning Techniques — Transfer Learning with Resnet and Differential Learning rates. **Medium**, 2018. Disponível em: <https://medium.com/@hariprasad0994/interesting-deep-learning-techniques-transfer-learning-with-resnet-and-differential-learning-b5e6616a94f>. Acesso em: 20 fev. 2020.

PROVOST, F. **Data Science para negócios**. Rio de Janeiro: Alta Books, 2016.

PYTHON SOFTWARE FOUNDATION. **Python Programming Language**. 2010. Disponível em: <http://www.python.org/>. Acesso em: 20 fev. 2020.

SAGAH. Disponível em: <https://sagah.grupoa.com.br/>. Acesso em: 20 fev. 2020.

SAS. **Machine Learning**. O que é e qual sua importância? Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/machine-learning.html](https://www.sas.com/pt_br/insights/analytics/machine-learning.html). Acesso em: 20 fev. 2020.

SHOGUN. Disponível em: <http://shogun-toolbox.org/>. Acesso em: 20 fev. 2020.

SILVA, A. M. L. *et al.* Descoberta de conhecimento através de métodos de aprendizagem de máquina supervisionados aplicados ao SIGAA/UFPI. **Revista de Sistemas e Computação**, Salvador, v. 7, n. 1, p. 68-78, jan./jun. 2007.

SQL. Disponível em: <https://www.w3schools.com/sql/>. Acesso em: 20 fev. 2020.

SCIKIT-LEARN. Disponível em: <http://scikit-learn.org/>. Acesso em: 20 fev. 2020.

TATI ENEM: a nova assistente virtual inteligente do AZ Escolas em Rede. **Conexia Educação**. Disponível em: <https://conexiaeducacao.com.br/wp/blog/tati-enem-a-nova-assistente-virtual-inteligente-do-az-escolas-em-rede/>. Acesso em: 20 fev. 2020.

VICTÓRIA, P. Banco de dados NoSQL: um manual prático e didático. **iMasters**, 2019. Disponível em: <https://imasters.com.br/banco-de-dados/banco-de-dados-nosql-um-manual-pratico-e-didatico>. Acesso em: 20 fev. 2020.

WAZE. **Como o Waze funciona?** Disponível em: <https://support.google.com/waze/answer/6078702?hl=pt-BR>. Acesso em: 20 fev. 2020.

WEKA. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>. Acesso em: 20 fev. 2020.

WICKHAM, H.; GROLEMUND, G. **R for data science**. Beijing [etc.]: O'Reilly, 2017.