



2 Agricultura digital: definições e tecnologias

Kleber Xavier Sampaio de Souza
Stanley Robson de Medeiros Oliveira
Carla Geovana do Nascimento Macário
Júlio César Dalla Mora Esquerdo
Maria Fernanda Moura
Maria Angelica de Andrade Leite
Helano Póvoas de Lima
Alexandre de Castro
Sônia Ternes
Inácio Henrique Yano
Edgard Henrique dos Santos

1 Introdução

Os avanços no processamento da informação e nas áreas de nanotecnologia, biotecnologia e ciência cognitiva estão promovendo uma convergência entre ciências chamada Nano-bio-info-cogno. O relatório *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science* (Roco; Bainbridge, 2003), encomendado da National Science Foundation dos Estados Unidos, foi elaborado por mais de 100 cientistas, que apontaram a sinergia entre nanotecnologia, biotecnologia, tecnologia da informação e ciência cognitiva como o segmento com maior potencial de avanço na inovação. Esse relatório destaca que as abordagens sistêmicas com o uso da matemática e da computação permitirão, pela primeira vez, entender o funcionamento de sistemas complexos existentes no mundo natural, tais como a mente humana, as explosões estelares, as interações sociais e os órgãos do corpo humano e os fenômenos naturais envolvidos com a agricultura.

A agricultura trabalha diretamente com três dessas áreas, a nanotecnologia, a biotecnologia e a tecnologia da informação. De fato, ela tem sido influenciada e alimentada pelo crescimento vertiginoso da capacidade de aquisição de dados, provenientes de diversas fontes, que vão desde a escala da célula, tais como as informações obtidas pela análise no campo das ciências “ômicas” (genômica, proteômica, transcriptômica e metabolômica), até a escala macroscópica, que inclui os dados socioeconômicos e os obtidos por meio de dispositivos de sensoriamento remoto, tais como satélites, passando pela escala da fazenda, com seus equipamentos agrícolas e sensores.

Atualmente tem-se a agricultura digital, que é uma agricultura cada vez mais conectada e remota que vai atuar no levantamento e no processamento dessa grande quantidade de dados coletados em todos os elos das cadeias produtivas, incluindo as fases de pré-produção, produção e pós-produção. Ela envolve, portanto, diferentes tipos de tecnologias digitais: sensores embarcados em plataformas orbitais, suborbitais, aerotransportados ou sistemas autônomos (drones, máquinas agrícolas), instalados diretamente no campo ou em diferentes ‘coisas’ (IoT) ao longo das cadeias produtivas, sistemas de telecomunicação, posicionamento global, software de controle, gestão e análise (*data analytics*) e atuadores.

Os dados provenientes dessas tecnologias passam a ser coletados não somente a partir de meios convencionais, mas também a partir de plataformas colaborativas ou mídias sociais (ciência do cidadão), dentre outros. Seu acúmulo representa um desafio para os sistemas de armazenamento, busca e recuperação, e também impactará os métodos de processamento e obtenção de informação.

Entretanto, se por um lado há uma quantidade de dados abundantes, por outro há grande defasagem no que tange à capacidade de gerência e análise desses mesmos dados e, conseqüentemente, da produção de conhecimento a partir deles. Configura-se, assim, um cenário complexo em que a transformação de dados em informações e conhecimento assume um papel estratégico em todos os setores da economia e na agropecuária, em particular, uma vez que esse setor é estratégico para o Brasil. Todos esses dados necessitam ser integrados, pré-processados e analisados para que deles se extraia conhecimento necessário ao estabelecimento da agricultura digital.

Este capítulo apresenta os conceitos relativos às tecnologias digitais que são utilizadas ao longo dos capítulos do livro, de forma consolidada, para facilitar o entendimento e o acesso por parte dos leitores.

2 Tecnologias digitais

As tecnologias digitais aqui apresentadas são divididas em cinco grupos. No primeiro grupo estão as tecnologias vinculadas à organização e à

representação da informação. No segundo estão as técnicas de modelagem matemática e estatística de processos envolvendo fenômenos biológicos, sociais e ambientais. No terceiro, a aplicação de inteligência artificial na agricultura. No quarto grupo, as tecnologias de sensores e robótica. Por fim, no quinto e último grupo, estão as tecnologias cujas aplicações possuem interação com a agricultura, como computação em nuvem e *blockchain*.

2.1 Organização, representação e acesso à informação

O volume de informação e a diversidade de formatos (dados de DNA, imagens de satélites, dados de sensores) em que essa informação é apresentada representam um desafio enorme à sua organização e reuso. Torna-se então necessário anotar, classificar, estruturar e prover mecanismos de acesso para que a informação possa ser encontrada e reutilizada no futuro, sendo este o propósito das tecnologias desta seção.

Tesouros - De acordo com a norma ANSI/NISO Z39.19-2005 (National Information Standards Organization, 2010), os tesouros são vocabulários controlados, arranjados de uma maneira tal que a relação entre seus termos é claramente identificada e padronizada. Os termos são compostos de uma ou mais palavras e selecionados da linguagem natural para serem incluídos no tesouro. No Thesaurus Agrícola Nacional (Thesagro), por exemplo, a palavra *Ácaro* está relacionada à palavra *Aracnídeo*, de forma que *Aracnídeo* é o termo mais abrangente (BT - *Broader Term* em inglês) e *Ácaro* é o termo mais específico (NT - *Narrower Term*). Existem também outros termos subordinados a *Aracnídeo* e que, portanto, são NT de *Aracnídeo*, tais como *Aranha* e *Escorpião*. BT e NT são formas de relacionamento verticais entre termos usadas nos tesouros; existem também as associações horizontais entre termos, expressas como termo relacionado (RT - *Related Term*). Para esse exemplo, existem no Thesagro os termos *Acaricida* e *Carrapato* associados como RT de *Ácaro*, sendo o primeiro um remédio para *Ácaros*, pertencente à hierarquia que começa com *Pesticida*, e o segundo um termo que pertence à hierarquia de *Parasito de Animal*. Os tesouros formam, portanto, uma teia de relacionamentos entre os termos, e essa teia ajuda a encontrar a informação que se procura. Os termos e a sua hierarquia podem ser usados para organizar o conteúdo de sites na Web e para expandir as buscas que são feitas nos conteúdos. Por exemplo, quando se busca *Ácaro*, podem-se recuperar também documentos que falem de *Carrapato* ou de *Acaricida*.

Ontologia - Uma ontologia define formalmente um vocabulário comum para que se compartilhe informação sobre um domínio de conhecimento. A ontologia inclui definições interpretáveis por máquina dos conceitos básicos desse domínio e as relações entre esses conceitos (Noy; McGuinness, 2001).

Nas ontologias, as relações entre os conceitos do domínio são explicitadas, de forma que consigam ser interpretadas por computadores. Cada conceito contém seus atributos, para os quais existem valores possíveis. Por exemplo, o conceito veículo, que é uma classe, contém as subclasses carro e motocicleta. O carro tem geralmente quatro rodas e a motocicleta, duas. Então, o atributo número de rodas seria quatro para veículo e duas para motocicletas. Tanto a motocicleta quanto o carro têm um atributo fabricante e tantos outros atributos quanto se queira usar para enriquecer a informação contida na ontologia e permitir o seu reuso. Como as ontologias disponibilizam uma linguagem comum, processável por máquina, um agente pode percorrer automaticamente vários sites Web que trabalham sobre um mesmo assunto, por exemplo peças de automóveis, e agregar a informação provida por eles, permitindo a comparação de preços. Isto é grandemente facilitado quando os vários sites usam a mesma ontologia para descrever suas peças.

Big Data - O termo *Big Data* inclui conjuntos de dados, cujos tamanhos vão além da capacidade que os sistemas gerenciadores de dados possuem de processá-los. Geralmente, são dados provenientes de várias fontes, tais como dispositivos móveis, sensores corporais, mídias sociais, e-mails, registros médicos eletrônicos, dados de genômica e de sensores geoespaciais, entre muitas outras. Essa variedade de fontes, a quantidade de dados e a velocidade com que os dados chegam para processamento geram o que se chama de “os três Vs” do *Big Data*, que são: volume, velocidade e variedade, aos quais às vezes ainda se adicionam “veracidade” e “valor”. A definição engloba dados estruturados, semiestruturados e não estruturados, embora seja muito mais comum o tratamento de dados não estruturados pelos sistemas que processam *Big Data* (Dedić; Stanier, 2017). Aplicações de *Big Data* surgem a todo momento: quando se analisam postagens em redes sociais sobre determinado assunto para ver a sua repercussão; quando se analisam buscas que são feitas no Google para identificar surtos de pandemias de gripe. Diante da inadequação dos sistemas tradicionais gerenciadores de bancos de dados em processar *Big Data*, foram desenvolvidas soluções pelas empresas que tradicionalmente sempre operaram com grandes volumes de dados, como é o caso do Google e da Cloudera, que desenvolveram o MapReduce, o Flume e o Sqoop. O MapReduce (Dean; Ghemawat, 2008) é um algoritmo desenvolvido pelo Google e que possui uma implementação livre desenvolvida pela Fundação Apache, chamada Hadoop (White, 2012). Essa implementação opera distribuindo grandes conjuntos de dados para serem processados em vários computadores em paralelo (possivelmente milhares de computadores) e depois

consolidando as respostas. O Apache Flume¹ foi originalmente desenvolvido pela Cloudera para gerenciar grandes volumes de dados de arquivos de log, mas foi estendido para processar eventos de fontes na Web como Twitter e Facebook. O Apache Sqoop² é uma ferramenta para transferir, de modo eficiente, dados entre fontes de dados estruturados, semiestruturados e não estruturados. É uma ferramenta interessante para trazer dados de fontes externas, tais como os bancos de dados relacionais, para dentro do sistema de arquivos distribuídos do Hadoop. MapReduce, Flume e Sqoop são apenas exemplos de sistemas que foram desenvolvidos para tratar *Big Data*, não sendo os únicos sistemas capazes de tratar dados de grande volume, variedade e velocidade.

API - Uma API (*Application Programming Interface*), ou interface de programação de aplicativos, em português, é uma forma de duas aplicações conversarem entre si, em que uma aplicação requisitante dispara a execução de uma outra para que sua própria tarefa seja concluída, ou seja, a aplicação requisitante necessita da segunda como provedora para seu funcionamento. O intermediário da comunicação entre as duas aplicações é a API, que define protocolos, rotinas e ferramentas para que a mensagem seja entregue à aplicação provedora e a resposta retorne para a aplicação requisitante. Uma API Web opera na internet usando os protocolos usuais para troca de informações, tais como HTML, XML ou JSON. Como exemplo de API aplicada à área agrícola pode-se citar a plataforma AgroAPI da Embrapa Informática Agropecuária (2020), que disponibiliza uma série de informações e modelos que podem ser usados por empresas, incluindo as startups, e instituições públicas e privadas para serem acopladas aos seus próprios softwares, sistemas Web ou aplicativos móveis. Sua utilização é gratuita para até 1.000 requisições por mês para cada API. Fazem parte da AgroAPI a API Agritec, que reúne informações sobre época de plantio, adubação, produtividade, zoneamentos agrícolas e cultivares para cinco culturas agrícolas; e a API SATVeg, que gera a visualização da evolução no tempo dos índices vegetativos NDVI e EVI para toda a América do Sul, a partir de dados de satélite. Esses índices possibilitam a observação das variações da biomassa verde na superfície terrestre, podendo auxiliar, por exemplo, na implementação do Código Florestal, ou no acompanhamento do ciclo de uma cultura agrícola, dentre outras dinâmicas da cobertura terrestre.

¹ Disponível em: <http://flume.apache.org/>

² Disponível em: <https://sqoop.apache.org/>

2.2 Modelagem matemática e estatística

A representação dos fenômenos da natureza por meio de modelos é parte integrante do método científico. Esta seção dedica-se a conceituar as categorias de representação usadas no método científico por meio de seus modelos. Conceitua também a Ciência de Dados, que surgiu de uma confluência de vários ramos do conhecimento para extrair conhecimento de massas de dados cada vez mais abundantes.

Modelo matemático - Um modelo surge da necessidade de se entender um fenômeno do mundo físico e prever o seu comportamento em determinada situação. Um modelo é sempre uma abstração do que ocorre no mundo real, uma simplificação do que ocorre no mundo real, de forma que se possa entender e quantificar um sistema (Torres; Santos, 2015). Em Bassanezi (2002), um modelo matemático consiste na transformação da realidade em problemas matemáticos, que são resolvidos e interpretados à luz do que ocorre no mundo real. A construção de um modelo matemático envolve várias etapas: (a) **conceituação**, que ocorre após observações iniciais sobre o problema, formulação de hipóteses para explicar seu funcionamento e uma primeira seleção de quais variáveis, processos e interações são considerados relevantes. Durante a conceituação ocorre uma importante tarefa, que é a simplificação do modelo em termos de variáveis e interações que são essenciais para a representação do problema, uma vez que os fenômenos do mundo natural, especialmente os biológicos, são excessivamente complexos; (b) **formalização matemática**, que é a tradução do problema em linguagem matemática. Existem muitas abordagens diferentes para se realizar essa tradução, tais como equações diferenciais, equações bayesianas, sistemas estocásticos, equações de diferenças finitas e sistemas baseados em agentes, cada uma com suas vantagens e limitações. Sua escolha depende da natureza do problema que se está modelando; (c) **estimação de parâmetros**, que envolve a descoberta de quais valores numéricos estão guiando a formulação matemática elaborada. A obtenção desses parâmetros pode ser realizada por meio de medidas experimentais, e a adoção de técnicas da estatística experimental lhes agrega maior confiabilidade; (d) **simulação e predição**, que é o momento em que se resolve o sistema de equações analiticamente ou se executa o modelo no computador. Como os problemas biológicos geralmente envolvem mecanismos de controle e regulação, a solução analítica dos modelos é quase sempre impossível, o que faz com que a abordagem computacional seja a mais frequente para a resolução do modelo; (e) **validação do modelo**, na qual se verifica a resposta do sistema para cada cenário dos valores de entrada das variáveis do modelo. Essa resposta tem de coincidir tanto em termos de trajetória do sistema quanto com os valores obtidos nas medições experimentais. Portanto, é nesse momento que se avalia quão próximo o

modelo está representando a realidade e que se mede a sua acurácia. Outra característica desejada do modelo é que ele seja capaz de prever novos fatos e relações não conhecidas que possam ser verificadas no mundo real; (f) **refinamento do modelo**, em que se critica a validade dos resultados do modelo, em termos das trajetórias do sistema sendo modelado, quando confrontado com o mundo real, e avalia-se a sua acurácia. Quando os modelos desviam do esperado, pode ser que alguma hipótese não tenha sido considerada ou seja falsa. Também pode ter havido um erro na obtenção dos dados que alimentaram a construção do modelo ou uma incorreção em sua formulação matemática. Nesse caso, podem ser propostas novas hipóteses e/ou novas variáveis e uma reavaliação do modelo matemático.

Modelo estatístico - A Estatística é a base do método científico, que assim pode ser resumido: i) definição do problema a ser estudado; ii) formulação de uma ou mais hipóteses a serem testadas; iii) condução de experimentos para testar as hipóteses formuladas; iv) análise estatística dos dados obtidos; v) interpretação dos resultados e obtenção das conclusões, isto é, obtenção de um modelo estatístico descritivo ou inferencial que comprove ou não as hipóteses originais. Tomemos um exemplo simples (Snedecor; Cochran, 1967): i) o problema a ser estudado era a variabilidade da concentração de cálcio em nabos; ii) as hipóteses diziam respeito ao comportamento da variabilidade do cálcio nas plantas e, especificamente, nas folhas de cada planta; iii) no experimento, quatro plantas foram escolhidas ao acaso e, em seguida, três folhas de cada planta foram selecionadas aleatoriamente e duas amostras de 100 mg foram tomadas de cada folha, determinando-se a quantidade de cálcio de cada amostra através de processos microquímicos; iv) os dados foram submetidos a uma análise de variância de acordo com o modelo das hipóteses formuladas; v) a análise concluiu que, estatisticamente a um nível de significância de 5%, a variabilidade nas folhas de cada planta mostra-se mais importante que a variabilidade em toda a planta, e que o modelo idealizado (as hipóteses colocadas) representa estatisticamente a realidade. Cada um desses efeitos de variabilidade é estimado de acordo com as hipóteses iniciais, as estimativas calculadas mostram se o modelo está adequado ou não às hipóteses formuladas a partir de uma margem de erro aceita, que no caso foi de 5%. Em geral, os processos biológicos são inerentemente complexos e a variabilidade de cada fator observado precisa ser estimada, isto é, o número de variáveis observado é gigantesco e, por vezes, nem todas as variáveis são conhecidas; as não conhecidas introduzirão um maior erro ao modelo estimado; lembrando que o modelo é aceito após as estimativas serem estatisticamente comprovadas e que o modelo como um todo possui um erro, também estimado. Nesse cenário, e em muitos outros, como a teoria

geral dos gases ou a teoria da seleção natural (Fisher, 1934), os argumentos são construídos sobre bases estatísticas.

Ciência de dados - Ciência de Dados é um campo interdisciplinar sobre processos e sistemas para extrair conhecimento ou insights de dados em várias formas, estruturados ou não. Incorpora técnicas e teorias das mais diversas áreas de conhecimento como computação, engenharia, matemática, estatística, economia, mineração de dados e inteligência artificial, com o objetivo de coletar dados, processá-los, integrá-los e analisá-los visando à criação de produtos e serviços de dados (Amaral, 2016). A Ciência de Dados não está restrita apenas à análise de grandes volumes de dados (*Big Data analytics*). Pequenos (*Small Data*) e grandes (*Big Data*) repositórios de dados são aspectos importantes desta área de pesquisa. *Small Data* contempla informações simples, que estão no banco de dados de qualquer empresa ou de pequenas propriedades rurais. *Small Data* inclui resultados de pesquisas, dados de consumidores ou produtores rurais, dados sobre propriedades agrícolas, e-mails com informações sobre práticas de manejo, dados contendo volume de produção agrícola por período, entre outros. Em geral, é constituído de dados estruturados, prontos para a análise. Já o *Big Data* refere-se a dados (principalmente) não estruturados, oriundos de múltiplas fontes e que deverão ser coletados, agregados e analisados no intuito de gerar informações de cunho gerencial. Dentre as aplicações de Ciência de Dados estão o marketing digital, que elabora anúncios personalizados a partir de informações obtidas por meio dos perfis de usuário e seu histórico de navegação nas empresas; os sistemas de recomendação, que se baseiam no padrão de páginas visitadas ou produtos comprados para sugerir novos produtos; e os sistemas de avaliação de crédito de clientes bancários, que, baseados em seu histórico e escores existentes em empresas de proteção ao crédito, calculam a probabilidade de o cliente se tornar inadimplente.

2.3 Inteligência artificial

As tecnologias de reconhecimento de padrões e aprendizado de máquina, incluindo *deep learning*, são parte integrante de muitos sistemas existentes na atualidade, tais como os carros autônomos e os sistemas de reconhecimento de voz. Trata-se de um rol de tecnologias que surgiram para analisar grandes conjuntos de dados e deles aprender padrões que possibilitem, por exemplo, identificar objetos ou antecipar a próxima palavra a ser falada em uma frase. Por outro lado, quando se deseja explicitar as regras de um sistema diretamente, sem o uso de aprendizado de máquina, mas ainda assim permitindo um certo grau de imprecisão, emprega-se a lógica nebulosa.

Reconhecimento de padrões - Um padrão, na forma como é entendido dentro do conceito de reconhecimento de padrões, pode ser a representação de um número escrito à mão, um número escrito em uma casa, uma laranja, um carro, uma palavra pronunciada, sequências de medições de temperatura, pressão e chuva, sequências de valores de ações da bolsa, enfim, muitas outras coisas que queremos que um sistema computacional aprenda a reconhecer. É por essa razão que muitos dos problemas importantes de reconhecimento de padrões podem ser caracterizados seja como classificações de forma de onda (sons, medições de temperatura, valores de ações etc.), seja como classificação de figuras geométricas, como ocorre com as imagens (Fukunaga, 1990). Nosso cérebro é especialmente projetado para reconhecer padrões. Logo nos primeiros anos da nossa existência, aprendemos a diferenciar os sons, as palavras, o que é um gato e o que ele tem de diferente de um cachorro e tantas outras coisas. O que se deseja com o reconhecimento de padrões por um sistema computacional é que ele aprenda a diferenciar os dados que se lhe apresentam, uma atividade que é computacionalmente conhecida como classificação. Para o processamento desses dados, existem várias técnicas que são usadas para reconhecimento de padrões, tais como árvores de decisão, florestas aleatórias, k-vizinhos mais próximos, máquinas de vetores de suporte e redes neurais (Bishop, 2006). A aplicação de técnicas de reconhecimento de padrões pode indicar, por exemplo, que uma dada sequência de valores de temperatura está dentro da normalidade, que uma ação na bolsa de valores está em trajetória de queda, que o número manuscrito em um papel é o 3, ou que o objeto que está em determinada posição em uma imagem é uma laranja.

Aprendizado de máquina - Trata-se de um processo intimamente relacionado ao reconhecimento de padrões (tópico anterior), pois o que se deseja durante o aprendizado de máquina é que o computador aprenda com os padrões que lhe foram apresentados. Segundo Bishop (2006), aprendizado de máquina e reconhecimento de padrões são duas facetas de um mesmo campo do conhecimento, tendo o reconhecimento de padrões originado da engenharia e o aprendizado de máquina da computação. Por essa razão, também são compartilhados os algoritmos entre reconhecimento de padrões e aprendizado de máquina. Em geral, é possível dividir o aprendizado de máquina em **supervisionado**, quando a partir de um conjunto previamente definido de dados rotulados deseja-se encontrar uma função que seja capaz de prever rótulos desconhecidos; e **não supervisionado**, em que se busca identificar grupos ou padrões a partir dos dados, sem um objetivo específico a ser alcançado (Russel; Norvig, 2020). Esses dois conceitos estão definidos na sequência.

Aprendizado não supervisionado - Nesse tipo de aprendizagem o conjunto de dados utilizado não possui nenhum tipo de rótulo. O objetivo desse tipo de aprendizagem é detectar similaridades e anomalias entre os objetos analisados. O processo de agrupamento de objetos em classes similares é denominado **clusterização**. Esse procedimento é também conhecido como segmentação de dados, pois particiona grandes conjuntos de dados de acordo com a similaridade entre subconjuntos. Os objetos mais similares com relação às características impostas pelo domínio devem ser alocados em um mesmo grupo, ao passo que aqueles menos similares devem ser alocados em grupos distintos. A similaridade entre os objetos deve ser obtida por medidas algébricas, como a distância euclidiana, para valores reais; ou por correspondência simples, para valores nominais. Esses algoritmos podem ser divididos em duas classes mais gerais, de acordo com a heurística empregada para a construção dos grupos.

A primeira classe são os **algoritmos particionais**, que, normalmente com custo computacional de execução linear, operam de maneira iterativa a partir da definição prévia da quantidade de grupos desejada e da definição de objetos representativos de cada grupo, conhecidos como centroides. Em cada iteração, cada objeto é associado ao centroide e, conseqüentemente, ao seu grupo mais similar. Os centroides dos grupos são então recalculados para a próxima iteração. O algoritmo atinge o seu ponto de convergência quando os centroides não são mais alterados entre uma iteração e outra, ou seja, quando os grupos estão bem definidos, considerando a medida de similaridade utilizada. Nessa subclasse de algoritmos encontra-se o *k-means* (Macqueen et al., 1967), considerado um dos dez algoritmos mais influentes em mineração de dados (Wu, 2008).

A segunda classe são os **algoritmos hierárquicos**, que possuem custo computacional de execução normalmente quadrático e não exigem, por sua vez, a identificação de representantes iniciais e nem da quantidade de grupos desejada. Assim, em uma única execução, podem ser gerados n particionamentos aninhados para o mesmo conjunto de n instâncias, contendo de 1 até n grupos cada, constituindo uma hierarquia de agrupamentos (Han; Kamber, 2006). Duas estratégias distintas podem ser utilizadas para a construção dessa hierarquia: a aglomerativa, que considera inicialmente cada instância do conjunto de dados como um grupo, fundindo pares de grupos em cada iteração; e a divisiva, que considera inicialmente todas as amostras pertencentes a um único grupo, dividindo-as em grupos menores em cada iteração (Hastie et al., 2009).

Aprendizado supervisionado - O processo de aprendizado supervisionado de máquina (supervised machine learning, em inglês) consiste em apresentar uma grande quantidade de dados previamente classificados a um

computador e fazer com que ele aprenda a partir desses dados. O aprendizado acontece pela modificação dos parâmetros do sistema à medida que mais e mais exemplos lhe são apresentados. Esses parâmetros são números para os quais não se sabe quais deveriam ser seus valores. Então, a tarefa do aprendizado é descobrir quais valores fazem com que o sistema acerte mais vezes. A cada exemplo, verifica-se se o sistema aprendeu a classificar corretamente aquele exemplo. Em caso positivo, o sistema reforça os parâmetros que permitiram essa classificação correta por meio dos seus pesos. Caso contrário, calcula-se qual a correção que o sistema deve sofrer para não cometer esse erro e ajustam-se negativamente os pesos que levaram àquela resposta errada. Pode-se, então, imaginar um sistema com muitos e muitos botões que têm de ser girados na medida certa para que esse sistema acerte a resposta no final. Entretanto, em vez de girarmos nós mesmos os botões, temos algoritmos que o fazem de forma controlada para que o aprendizado aconteça. Os principais paradigmas de aprendizado podem ser enumerados como a seguir:

- a) **simbólico** (árvores de decisão): uma árvore de decisão é uma estrutura semelhante a um fluxograma na qual cada nó interno representa um “teste” em um atributo, cada ramo representa o resultado do teste e cada nó folha representa um rótulo de classe (decisão tomada após computar todos os atributos). Os caminhos da raiz para a folha representam as regras de classificação (Quinlan, 1986).
- b) **baseado em instâncias** (k -NN ou k vizinhos mais próximos): a ideia principal do k -NN é determinar o rótulo de classificação de uma amostra baseado nas k amostras vizinhas advindas de um conjunto de treinamento. Dentre os k exemplos, verifica-se a classe mais frequente. Essa classe é atribuída ao novo exemplo (Fukunaga; Narendra, 1975).
- c) **baseado em aprendizado estatístico** (SVM - *Support Vector Machines*): a forma mais simples de particionar um espaço euclidiano de n dimensões é através de hiperplanos. O classificador SVM baseia-se também nessa estratégia, porém, utiliza um tipo especial, o hiperplano de separação ótima. Trata-se de um hiperplano que divide as classes maximizando a margem de separação entre elas (Vapnik, 1995, 1998).
- d) **baseado em comitê**: é o campo do aprendizado de máquina que constrói um grupo de classificadores, denominados classificadores-base, com o objetivo de ser mais preciso que o melhor dos elementos do grupo. A abordagem mais simples baseada nesse algoritmo é o voto da maioria simples, em que diversos classificadores são combinados em uma estratégia de voto. Como resultado, a resposta que receber o maior número de votos é considerada a resposta do comitê (Han; Kamber, 2006). Um exemplo desse tipo de abordagem é o Random Forest. Trata-se de uma técnica de

classificação e regressão que consiste num conjunto de árvores de decisão combinadas para solucionar problemas de classificação (Breiman, 2001).

- e) **conexionista** (Redes Neurais Artificiais - RNA): são modelos computacionais, inspirados pelo sistema nervoso central (em particular o cérebro), capazes de realizar o aprendizado de máquina, bem como o reconhecimento de padrões. Um exemplo de modelo conexonista é a técnica de *deep learning*, detalhada a seguir.

Deep learning - A técnica de *deep learning*, ou redes neurais profundas, é uma técnica de aprendizado de máquina na qual o modelo escolhido para o algoritmo de aprendizado é uma rede neural artificial com muitas camadas. As redes neurais foram inspiradas pela forma como os neurônios funcionam em sistemas biológicos, em que operam de modo paralelo e descentralizado (Marblestone et al., 2016). Tipicamente, uma rede neural pode conter mais de 100 camadas, dispostas uma após a outra ou em paralelo. Cada uma dessas camadas é composta por um ou mais neurônios, interligados entre si de forma que o resultado dos neurônios que estão em uma camada alimenta a entrada dos neurônios que estão na camada posterior. O método de treinamento de redes neurais emprega frequentemente um algoritmo chamado de *backpropagation*. Como associado a cada neurônio existe um peso que aquele neurônio representa na resposta, esse algoritmo compara a resposta do sistema com o valor que deveria ter sido e distribui o erro recalculando os valores dos pesos dos neurônios para trás. Existem muitas arquiteturas de redes neurais disponíveis, como as redes conectadas para frente (*feedforward*), as redes convolutivas, as redes recorrentes e as máquinas de Boltzmann restritas, entre muitas outras (Goodfellow et al., 2016). A arquitetura escolhida para a rede é dependente do problema que se pretende resolver: as redes conectadas para frente são usadas tanto em problemas de classificação quanto de regressão; as redes convolutivas, para problemas de classificação em imagens; as redes recorrentes, para problemas que envolvem sequências, tais como o processamento de linguagem natural; e as máquinas de Boltzmann restritas são aplicadas para redução de dimensionalidade, uma tarefa na qual a quantidade de variáveis de entrada é grande e se busca identificar as mais significativas. Essa lista de problemas para cada rede não é excludente, mas apenas para servir de exemplo, pois uma máquina de Boltzmann restrita pode ser usada para resolver outros problemas, tais como regressão e classificação, assim como também ocorre com as demais arquiteturas de rede neural. A área de *deep learning* tem muito de arte envolvida na seleção de uma dada arquitetura para um problema, bem como na parametrização dos modelos.

Conjuntos nebulosos e lógica nebulosa: A teoria de conjuntos clássica define como conjunto uma classe de objetos com pertinência binária, ou seja,

cada elemento pode pertencer ou não ao conjunto (\in ou \notin). Zadeh (1965) fundamentou o conceito de conjuntos *fuzzy* (CF) como sendo uma classe de objetos em que cada elemento possui um grau de pertinência contínuo, admitindo qualquer valor entre zero e um. Tal conceito permite que sejam tratados problemas do mundo real onde os critérios de pertinência e as fronteiras entre classes não são precisamente definidos (ou seja, são nebulosos ou *fuzzy*, em inglês). Um elemento pode ter “graus de pertinência” diferentes para vários conjuntos. Analogamente à teoria de conjuntos clássica, toda uma classe de operações lógicas é derivada dos conjuntos nebulosos, denominada de lógica nebulosa. Os que operam com lógica nebulosa são denominados Sistemas Baseados em Regras Fuzzy (SBRF). São sistemas de inferência cujos componentes lógicos são expressos por meio de CF. Tipicamente é composto por uma base de dados *fuzzy* (variáveis de entrada e saída), um mecanismo de inferência e uma base de regras *fuzzy* do tipo “SE A então B”, cujos termos linguísticos são CF (Klir; Yuan, 1995).

2.4 Sensores e estudo da terra

Os sensores e atuadores estão no cerne da agricultura digital, pois por meio deles pode-se perceber o que está ocorrendo no meio ambiente e tomar as ações adequadas. Os sensores podem ser orbitais, como os satélites, que permitem a coleta de dados geoespaciais, ou proximais, como os sensores instalados nas propriedades rurais e interligados aos dispositivos da internet das coisas. Quando a computação está plenamente integrada aos sensores de um ambiente e distribuída nesse ambiente, surge a Computação Ubíqua.

Computação ubíqua - O termo Computação Ubíqua foi proposto pelo cientista Mark Weiser (1991), do Centro de Pesquisa de Palo Alto da Xerox (PARC), para denominar um paradigma de computação proposto para o século 21. Nesse paradigma a computação deveria estar em todos os lugares, daí o termo ubíqua, e ser invisível para seu usuário. Para explicar o conceito, Weiser considera que a linguagem escrita foi a primeira tecnologia ubíqua, pois antes dela a informação estava restrita à memória das pessoas. Com sua invenção, qualquer um que saiba ler é capaz de entender o que está escrito, estando, portanto, independente da memória de quem escreveu. O conceito de computação em todos os lugares é diferente da ideia de se levar um notebook a qualquer lugar, pois mesmo nesse caso, o que se leva consigo é o poder computacional e o foco, portanto, continua sendo o computador. Com a computação ubíqua, os computadores operam a distância, sem contato físico com os usuários. A interação com esses computadores que estão no ambiente poderia ser feita por reconhecimento de presença, voz e gestos por sensores instalados no ambiente, displays e projetores. A computação ubíqua também implica mais inteligência por parte do computador, pois seus sensores teriam

de perceber o que está acontecendo no ambiente e tomar as ações para facilitar a tarefa dos usuários que nele estão, ativando inclusive serviços, sem que o usuário os tenha demandado. Por exemplo, ao entrar em seu escritório e procurar por determinado documento guardado em papel, o sistema apontaria onde você guardou esse documento no passado. O sistema também poderia trazer para uma sala de reuniões o projeto em que você estava trabalhando para que pudesse ser apresentado. É claro que a computação ubíqua apresentaria novos desafios em termos de privacidade e segurança, pois o primeiro exemplo significa que o sistema estava observando todos os seus passos quando guardou aquele documento no passado, enquanto o segundo quer dizer que o sistema teria acesso a todos os seus arquivos e transferiria apenas os arquivos necessários para a apresentação. Além de privacidade e segurança, também existem outros desafios, tais como a junção de partes de hardware e software de vários fabricantes, cujos softwares teriam de ser integrados e comandados por um sistema maior. Embora não exista um sistema que implemente completamente a ideia da computação ubíqua, algumas tecnologias tentam se aproximar desse ideal, como os sistemas de caixas de som que ouvem o que se está dizendo e executam tarefas, tais como ajustar a iluminação, tocar uma música preferida ou realizar uma busca na internet. Na agricultura, o conceito de computação ubíqua tem sido usado na aplicação de agroquímicos. Nessa aplicação, sensores existentes próximos às folhas orientariam a eletrônica embarcada nos pulverizadores para se obter uma administração com a maior cobertura possível, gastando o mínimo de líquido.

IoT - A internet das coisas (*internet of things* - IoT, em inglês) é definida pela International Telecommunication Union - ITU (2012) como sendo uma infraestrutura global para a sociedade da informação, possibilitando serviços avançados pela interconexão de coisas (físicas e virtuais), baseadas em tecnologias da informação e comunicação interoperáveis, quer sejam essas estruturas existentes ou em evolução. Do ponto de vista da internet das coisas, a ITU define que as coisas são objetos do mundo físico ou do mundo virtual que são capazes de serem identificados e integrados às redes de comunicação. Os objetos virtuais são incluídos à IoT por meio de coisas físicas ligadas a dispositivos, que por sua vez possuem capacidade mandatória de comunicação. A comunicação entre dispositivos pode ser realizada via rede de comunicação (com ou sem *gateway* intermediário) ou diretamente entre os dispositivos, sem uma rede de comunicação, sendo necessário, neste último caso, que haja comunicação direta entre os dispositivos. Quando a comunicação entre dispositivos ocorre via um *gateway*, este deve prover no mínimo duas tecnologias de rede, seja para integrar dispositivos, como ZigBee, Bluetooth, Wi-Fi ou LoRa, seja para integrar dispositivos à rede de comunicação, como as redes 2G, 3G, LTE, satelitais ou outras. Os dispositivos podem ainda possuir

a capacidade de entrar em repouso e retornar automaticamente para economizar energia. Essa capacidade é especialmente importante para sensores que ficam instalados em locais remotos, que não possuem ligação direta à energia elétrica, como ocorre com alguns sensores de monitoramento agrícola. Os objetos conectados à rede IoT podem ser desde pessoas ou animais com etiquetas RFID até marcapassos e outros dispositivos hospitalares de uso individual, implementos agrícolas, celulares, câmeras de vigilância, sensores de umidade e pressão atmosférica, pluviômetros, automóveis com sensores embarcados e muitos outros. O requerido para todas as coisas conectadas à IoT é que possuam endereço na internet, ou seja, um endereço IP. Tendo esse endereço, as coisas podem ser acessadas por quaisquer máquinas ligadas à internet. Esse acesso a qualquer tempo faz com que as coisas ligadas à IoT estejam vulneráveis em dois pontos: segurança e privacidade. A preocupação com segurança gera a necessidade da implementação de requisitos que procurem assegurar a confidencialidade e a integridade das informações, tanto nos dados quanto nos serviços que processam esses dados. A questão da privacidade também precisa ser suportada pela IoT, pois os dados que trafegam no sistema IoT podem transitar informações sensíveis vinculadas aos proprietários ou usuários das coisas conectadas. A proteção à privacidade desses dados deve ocorrer durante a transmissão dos dados, sua agregação, armazenamento, processamento e mineração. Na agricultura, sensores RFID têm sido usados para a identificação e o rastreamento de animais no campo.

Robótica - O termo robô vem da palavra *robot*, que significa servo na língua tcheca. Foi proposta por Josef Čapek ao seu irmão Karel para ser usada na peça de ficção Rossum's Universal Robots, publicada em 1920 (Szabolcsi, 2014). Nessa peça, máquinas com comportamento e aparência humana executam trabalhos. Nos dias atuais, os robôs assumem as mais variadas formas e funções. Na indústria, assumem a forma de braços para executar tarefas repetitivas, como soldagem, ou tarefas perigosas, como descontaminação em instalações nucleares. Os robôs militares, agrícolas e para exploração espacial frequentemente são veículos com rodas ou asas. A robótica é uma área de pesquisa que conjuga esforços de múltiplas áreas, tais como engenharia de computação, engenharia da informação, engenharia mecânica, engenharia eletrônica, biologia, indo até as ciências sociais à medida que os robôs devem assumir comportamentos adequados à interação humana. O grau de autonomia de um robô pode variar do controle remoto à operação totalmente autônoma. Dependendo da tarefa ou do grau de autonomia, o robô precisa: ser dotado de visão computacional para construir uma representação global do ambiente em que se encontra e dos objetos dentro do campo de visão; possuir um sistema de controle para realizar a tarefa desejada, podendo ou não incluir inteligência artificial; possuir atuadores que irão mover as partes de acordo

com o controle e implementar uma interface com o usuário. Pode ainda necessitar de dispositivos que implementem os sentidos de tato, audição e olfato. Existem vários robôs avançados atualmente: o Asimo, desenvolvido pela Honda, é um dos robôs com aparência humanoide mais evoluídos. Ele consegue andar sobre superfícies irregulares, conversar com várias pessoas ao mesmo tempo, abrir garrafas e colocar líquido em copo, além de dominar várias conversas simultâneas com pessoas diferentes. A NASA criou o Robonaut2 e o enviou à estação espacial internacional para auxiliar na realização de tarefas perigosas ou mesmo corriqueiras. Na agricultura, os robôs frequentemente possuem o formato de um veículo para todo-o-terreno (*off-road*), como o robô *See and Spray*, desenvolvido pela Blue River para detectar ervas daninhas e aplicar defensivos agrícolas de forma seletiva, apenas sobre essas ervas, evitando a cultura plantada.

Dados geoespaciais - Também denominados dados geográficos, pertencem a uma classe particular de dados que descrevem fatos, objetos e fenômenos do globo terrestre, associados à sua localização sobre a superfície terrestre, num certo instante ou período (Câmara et al., 1996). Os dados geoespaciais distinguem-se essencialmente dos demais por sua componente espacial, que associa a cada entidade ou fenômeno uma localização traduzida por um sistema geodésico de referência terrestre. Geotecnologias é o nome dado a uma categoria especial de tecnologias utilizadas para o processo de aquisição, visualização, processamento, análise e/ou disponibilização de dados geoespaciais. Nesse contexto, tecnologias como o sensoriamento remoto, o *Global Positioning System* (GPS), a topografia, os Sistemas de Informações Geográficas (SIGs), os bancos de dados geográficos, entre outras, são classificadas como geotecnologias. Quando uma informação geoespacial é derivada de uma ou mais geotecnologias, ela é denominada geoinformação ou geodado. Por fim, o processo de aplicação de uma ou mais geotecnologias para adquirir, processar, visualizar, analisar e/ou disponibilizar dados espacialmente referenciados, com o intuito de gerar geoinformação, é denominado geoprocessamento. Os dados geoespaciais são usados na agricultura, por exemplo, no monitoramento da safra de uma determinada *commodity*, em que uma sequência de imagens de satélite é analisada ao longo do tempo em uma região para se determinar quanto será produzido.

SIGs - Sistemas de Informações Geográficas (SIG) ou GIS (*Geographic Information System*), em inglês, é uma das principais tecnologias para visualização, análise e tratamento dos dados geográficos. Existem diversas definições sobre o que são SIGs, desde as mais complexas às mais simples. Pires et al. (1994) definem SIG como um sistema que realiza o tratamento computacional de dados geoespaciais, armazenando, gerenciando e recuperando

informações. Esses sistemas são muito utilizados em ambientes de decisão, provendo aos usuários facilidades de combinar as informações de uma determinada região. A principal diferença entre um SIG e um sistema de informação convencional é a capacidade do SIG de armazenar tanto os atributos descritivos dos dados quanto as geometrias dos diferentes tipos de dados geográficos. As principais características de SIGs são: inserir e integrar, numa única base de dados, informações espaciais textuais e outras fontes de dados, como imagens de satélite e dados de GPS; e oferecer mecanismos para combinar as várias informações, por meio de algoritmos de manipulação e análise, bem como para consultar, recuperar e visualizar o conteúdo da base de dados geográficos. A abordagem tradicionalmente utilizada para a organização de dados geoespaciais em um SIG é a sua distribuição em camadas, também denominadas *layers* ou planos de informação, em que cada uma aborda um tema distinto para uma dada região geográfica. Por exemplo, uma imagem de satélite de uma região é uma camada, assim como os municípios dessa região, sua geomorfologia e sua hidrologia. Cada camada é representada internamente usando estruturas lógicas próprias de cada SIG e é armazenada em arquivos distintos, de acordo com o formato do sistema utilizado. Na agricultura, um SIG pode ser usado para criar um modelo digital de uma propriedade rural, a partir de medições feitas usando-se um GPS em vários pontos da propriedade.

2.5 Tecnologias convergentes

A agricultura digital incorpora conceitos que foram originalmente desenvolvidos para outras áreas, como o *blockchain* e a computação em nuvem, que convergem para a solução dos problemas agrícolas. A reutilização dessas tecnologias surgiu da necessidade de armazenamento de dados de forma remota, para que sejam processados com maior resiliência, e também para o atendimento de uma demanda recorrente na agricultura, que é a rastreabilidade de seus produtos e processos.

Blockchain (cadeia de blocos, em português) - É um tipo de banco de dados distribuído cujo modelo de armazenamento permite a guarda de registros de modo permanente e inviolável. É mundialmente conhecido por ser a tecnologia sobre a qual se desenvolveu a criptomoeda bitcoin, sendo sua origem datada de 2008, quando seu autor, sob o pseudônimo de Satoshi Nakamoto, publicou um artigo na internet (Nakamoto, 2008) sobre a criação de um sistema de pagamento eletrônico descentralizado, seguro e baseado em uma rede do tipo *peer-to-peer* (p2p). O *blockchain* permite codificar o conteúdo de uma mensagem de comprimento variável para dados de comprimento fixo via protocolos de integridade e autenticação baseados em cifras de uso único, ou função *hash* de mão única, (Castro, 2017; Ethereum, 2019).

Cada transação pode ser entendida como uma ação passível de rastreabilidade, e que é certificada pelos nós da rede, podendo haver sigilo de parte ou de todo o seu conteúdo. Essas transações são agrupadas de maneira semelhante a um livro razão, também utilizado em operações contábeis, e, por essa característica, o conjunto é chamado de *ledger*. Os *ledgers* são a base, dentro de uma estrutura de ferramentas computacionais, para implementação de sistemas de transações com a tecnologia *blockchain* em ambientes corporativos.

Sistemas de rastreabilidade via *blockchain* proporcionam uma forma segura e distribuída para fornecer informações no âmbito de uma cadeia produtiva agrícola, ou de quaisquer outros processos agroindustriais, permitindo rastrear informações como a origem do produto e seus insumos, o uso de agrotóxicos na lavoura, entre outras.

Computação em nuvem (*cloud computing*, em inglês) - Refere-se a uma tecnologia que permite o acesso a programas, arquivos e serviços por meio da internet, sem a necessidade de instalação de programas ou armazenamento de dados – daí vem a alusão à “nuvem”. O termo geralmente é usado para descrever centros de dados disponíveis para muitos utilizadores pela internet (Hayes, 2008). Uma vez devidamente conectado ao serviço on-line, é possível desfrutar de suas ferramentas e salvar todo o trabalho que for feito para acessá-lo depois, de qualquer lugar, a partir de qualquer computador que tenha acesso à internet, independentemente de plataforma. O requisito mínimo é um computador compatível com os recursos disponíveis na internet. Por exemplo, um computador pessoal torna-se apenas um *chip* ligado à internet, que neste caso representaria a “grande nuvem” de computadores, sendo necessários somente os dispositivos de entrada teclado, *mouse* e monitor.

A computação em nuvem pode ser vista como um paradigma de infraestrutura que permite o estabelecimento de software como serviço, sendo um grande conjunto de serviços baseados na Web, com o objetivo de fornecer funcionalidades que, até então, necessitavam de grandes investimentos em hardware e software, e que funciona através de um modelo de pagamento pelo uso (Buyya et al., 2009). Um exemplo típico de computação em nuvem são os serviços de sincronização de arquivos, como o Dropbox. Ao copiar ou mover um arquivo nesse espaço, ele será duplicado no servidor do aplicativo e em outros computadores que tenham o programa instalado e nos quais um usuário acesse a sua conta.

A computação em nuvens oferece diversos benefícios, como: 1) **redução de custos**: seja pela diminuição nos gastos com energia, *no-break* ou gerador, ar-condicionado e segurança física dos equipamentos, seja na aquisição de softwares e hardwares; 2) **economia de espaço**: a partir do momento em que se adere aos serviços em nuvem, o armazenamento será totalmente virtual; 3) **flexibilidade**: os serviços são perfeitamente adaptáveis aos diferentes tipos de

uma empresa. Se essa previsão se mostrar subestimada, facilmente pode-se incrementar o serviço, reajustando-o à demanda real; 4) **atualização constante**: a tecnologia avança e igualmente depressa os hardwares tornam-se defasados. Ao migrar para a computação em nuvem, acompanhar os passos do desenvolvimento tecnológico torna-se uma tarefa bem menos exaustiva e dispendiosa, já que serviços contratados são atualizados constantemente; 5) **capacidade de armazenamento**: a possibilidade de realizar backup de uma enorme quantidade de dados, de maneira instantânea, é tão importante quanto a facilidade de recuperação desses dados a qualquer momento, por um custo consideravelmente baixo; 6) **aumento da colaboração**: por permitir o acesso remoto de várias pessoas a um mesmo arquivo, a computação em nuvem estimula o trabalho colaborativo. Como as atualizações acontecem em tempo real, a troca entre membros de uma mesma equipe se dá de maneira muito mais rápida.

No entanto, o armazenamento em nuvens pode gerar desconfiança, principalmente no que se refere à segurança. Afinal, a proposta é manter informações importantes em um ambiente virtual, e não são todas as empresas e pessoas que se sentem à vontade com essa abordagem.

3 Considerações finais

Este capítulo apresentou os principais conceitos usados em gestão, processamento e visualização de dados da agricultura digital. Foram apresentadas tecnologias digitais vinculadas à organização e à representação da informação, modelagem matemática e estatística, inteligência artificial, sensores e robótica e tecnologias convergentes, como a computação em nuvem e o *blockchain*. Nos próximos capítulos, essas tecnologias são exploradas nas muitas aplicações, construídas pela Embrapa Informática Agropecuária e seus parceiros, com o objetivo de fornecer soluções para uma agricultura cada vez mais dinâmica e integrada, como é a agricultura digital. Como se pode perceber, pelo rol de tecnologias aqui conceitualizadas, o ferramental utilizado para a resolução de problemas agrícolas situa-se na fronteira do conhecimento tecnológico.

4 Referências

- AMARAL, F. **Introdução à ciência de dados**: mineração de dados e big data. Rio de Janeiro: Alta Brooks, 2016.
- BASSANEZI, R. **Ensino-aprendizagem com modelagem matemática**. São Paulo: Contexto, 2002.
- BISHOP, C. M. **Pattern recognition and machine learning**. Singapore: Springer Science+Business Media, 2006.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, p. 5-32, 2001. DOI: 10.1023/A:1010933404324.

BUYA, R.; YEO, C. S.; VENUGOPAL, S.; BROBERG, J.; BRANDIC, I. Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. **Future Generation Computer Systems**, v. 25, n. 6, p. 599-616, June 2009. DOI: 10.1016/j.future.2008.12.001.

CÂMARA, G.; CASANOVA, M.; MEDEIROS, C. B.; MAGALHÃES, G.; HEMERLY, A. **Anatomia de Sistemas de Informação Geográfica**. Campinas: Ed. Unicamp, 1996. 193 p.

CASTRO, A. de. Quantum one-way permutation over the finite field of two elements. **Quantum Information Processing**, v. 16, article number 149, 2017. DOI: 10.1007/s11128-017-1599-6.

DEAN, J.; GHEMAWAT, S. MapReduce: simplified data processing on large clusters. **Communications of the ACM**, v. 51, n. 1, p. 107-113, 2008. DOI: 10.1145/1327452.1327492.

DEDIĆ, N.; STANIER, C. Towards differentiating business intelligence, big data, data analytics and knowledge discovery. In: PIAZOLO, F.; GEIST, V.; BREHM, L.; SCHMIDT, R. (ed.). **Innovations in enterprise information systems management and engineering**. Berlin; Heidelberg: Springer, 2017. p. 114-122. (Lecture Notes in Business Information Processing, n. 285). DOI: 10.1007/978-3-319-58801-8_10.

EMBRAPA INFORMÁTICA AGROPECUÁRIA. **AgroAPI**. Disponível em: <https://www.agroapi.cnptia.embrapa.br>. Acesso em: 18 jun. 2020.

ETHEREUM. 2019. Disponível em: <https://www.ethereum.org/>. Acesso em: 22 maio 2020.

FISHER, R. A. **Statistical methods for research workers**. 5th ed. Tweeddale Court: Oliver and Boyd, 1934.

FUKUNAGA, K. **Introduction to statistical pattern recognition**. 2nd ed. Boston: Academic Press, 1990. DOI: 10.1016/B978-0-08-047865-4.50007-7.

FUKUNAGA, K.; NARENDRA, P. M. A branch and bound algorithm for computing k-nearest neighbors. **IEEE Transactions on Computers**, v. 100, n. 7, p. 750-753, 1975. DOI: 10.1109/T-C.1975.224297.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge: The MIT Press, 2016.

HAN, J.; KAMBER, M. **Data mining: concepts and techniques**. 2nd ed. San Francisco: Morgan Kaufmann, 2006. 770 p.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning**. New York: Springer, 2009. DOI: 10.1007/978-0-387-84858-7.

HAYES, B. Cloud computing. **Communication of the ACM**, v. 51, n. 7, p. 9-11, Jul 2008. DOI: 10.1145/1364782.1364786.

INTERNATIONAL TELECOMMUNICATION UNION. **ITU-T Y.2060**. Series y: global information infrastructure, internet protocol aspects and next-generation networks: next generation networks – frameworks and functional architecture models: overview of the Internet of things. 2012. Former ITU-T Y.2060 renumbered as ITU-T Y.4000 on 2016-02-05 without further modification and without being republished. Disponível em: <https://www.itu.int/rec/T-REC-Y.2060-201206-I>. Acesso em: 17 abr. 2020.

KLIR, G. J.; YUAN, B. **Fuzzy sets and fuzzy logic: theory and applications**. Upper Saddle River: Prentice Hall, 1995. 574 p.

MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 5., 1967, Oakland.

Proceedings. Berkeley: University of California Press, 1967. v. 1, p. 281-297. Disponível em: <https://projecteuclid.org/euclid.bsmmsp/1200512992>. Acesso em: 17 abr. 2020.

MARBLESTONE, A. H.; WAYNE, G.; KORDING, K. P. Toward an Integration of deep learning and neuroscience. **Frontiers in Computational Neuroscience**, v. 10, n. 94, Sept 2016. DOI: [10.3389/fncom.2016.00094](https://doi.org/10.3389/fncom.2016.00094).

NAKAMOTO, S. **Bitcoin**: a peer-to-peer electronic cash system. 2008. Disponível em: <https://bitcoin.org/bitcoin.pdf>. Acesso em: 22 jan. 2020.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005 (R2010)**: guidelines for the construction, format, and management of monolingual controlled vocabularies. 2010. Disponível em: https://groups.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf. Acesso em: 18 jun. 2020.

NOY, N. F.; MCGUINNESS, D. L. **Ontology development 101**: a guide to creating your first ontology. 2001. Disponível em: http://protege.stanford.edu/publications/ontology_development/ontology101.pdf. Acesso em: 17 abr. 2020.

PIRES, M. F.; MEDEIROS, C. M. B.; SILVA, A. B. Modelling geographic information systems using an object oriented framework. In: BAEZA-YATES, R. (ed.). **Computer science 2**. Boston: Springer, 1994. DOI: [10.1007/978-1-4757-9805-0_18](https://doi.org/10.1007/978-1-4757-9805-0_18).

QUINLAN, J. R. Induction of decision trees. **Machine Learning**, v. 1, p. 81-106, 1986. DOI: [10.1007/BF00116251](https://doi.org/10.1007/BF00116251).

ROCO, M. C.; BAINBRIDGE, W. S. Overview converging technologies for improving human performance: nanotecnologia, biotecnologia, information technology and cognitive science. In: ROCO, M. C.; BAINBRIDGE, W. S. (ed.). **Converging technologies for improving human performance**. Dordrecht: Springer, 2003. p. 1-27. DOI: [10.1007/978-94-017-0359-8_1](https://doi.org/10.1007/978-94-017-0359-8_1).

RUSSEL, S.; NORVIG, P. **Artificial intelligence**: a modern approach. 4th ed. New Jersey: Prentice Hall, 2020.

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical methods**. 6th ed. Ames: The Iowa State University Press, 1967. 286 p.

SZABOLCSI, R. The birth of the term Robot. **Advances in Military Technology**, v. 9, n. 1, jun. 2014.

TORRES, N. V.; SANTOS, G. The (mathematical) modeling process in biosciences. **Frontiers in Genetics**, v. 6, n. 354, Dec 2015. DOI: [10.3389/fgene.2015.00354](https://doi.org/10.3389/fgene.2015.00354).

VAPNIK, V. N. **Statistical learning theory**. New York: John Wiley and Sons, 1998.

VAPNIK, V. N. **The nature of statistical learning theory**. New York: Springer-Verlag, 1995. DOI: [10.1007/978-1-4757-2440-0](https://doi.org/10.1007/978-1-4757-2440-0).

WEISER, M. The computer of the 21st Century. **Scientific American**, v. 265, n. 3, Sept. 1991. DOI: [10.1038/scientificamerican0991-94](https://doi.org/10.1038/scientificamerican0991-94).

WHITE, T. **Hadoop**: the definitive guide. Sebastopol: O'Reilly Media, 2012.

WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H.; MCLACHLAN, G. J.; NG, A.; LIU, B.; YU, P. S.; ZHOU, Z.-H.; STEINBACH, M.; HAND, D. J.; STEINBERG, D. Top 10 algorithms in data mining. **Knowledge and Information Systems**, v. 14, n. 1, p. 1-37, Jan 2008. DOI: [10.1007/s10115-007-0114-2](https://doi.org/10.1007/s10115-007-0114-2).

ZADEH, L. A. Fuzzy sets. **Information and Control**, v. 8, n. 3, p. 338-353, June 1965. DOI: [10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X).