

# Inteligência artificial e ciência de dados para o desenvolvimento da agricultura inteligente

*Artificial intelligence and data science for the development of smart farming*

Ednaldo José Ferreira<sup>1</sup>, Alexandre Cláudio Botazzo Delbem<sup>2</sup>, Alex Bisetto Bertolla<sup>3</sup>, Marcos Cezar Visoli<sup>4</sup>, Cassio Antônio Andrade<sup>5</sup>, Suzana Lucy Nixdorf<sup>6</sup>, Edilene Cristina Ferreira<sup>7</sup>

<sup>1</sup>Analista, Pesquisador-Associado, Embrapa Instrumentação e Centro de Inteligência Artificial (C4AI - AgriBIO), São Carlos (SP), Brasil, ednaldo.ferreira@embrapa.br

<sup>2</sup>Professor, Pesquisador, Instituto de Ciências Matemáticas e de Computação (USP-ICMC), Centro de Inteligência Artificial (C4AI - AgriBIO), São Carlos (SP), Brasil, acbd@icmc.usp.br

<sup>3</sup>Analista, Embrapa Instrumentação, São Carlos (SP), Brasil, alex.bertolla@embrapa.br

<sup>4</sup>Pesquisador, Embrapa Agricultura Digital, Campinas (SP), Brasil, marcos.visoli@embrapa.br

<sup>5</sup>Mestrando, Instituto de Ciências Matemáticas e de Computação (USP-ICMC), São Carlos (SP), Brasil, andrade.cassio@usp.br

<sup>6</sup>Professora, Departamento de Química, Universidade Estadual de Londrina (UEL), Londrina (PR), Brasil, snixdorf@uel.br

<sup>7</sup>Professora, Universidade Estadual Paulista (UNESP), Instituto de Química, Araraquara (SP), Brasil, edilene.c.ferreira@unesp.br

## RESUMO

O desenvolvimento da agricultura digital e de precisão é uma das vertentes que prometem provocar uma nova revolução verde na agricultura. O campo produtivo deve se firmar como o principal segmento de desafio científico-tecnológico nas cadeias agroalimentares e, conseqüentemente, maior provedor de dados às tomadas de decisões das empresas rurais. O sensoriamento diversificado para aquisição de dados do espaço agrícola tem se tornado cada vez mais onipresente. No entanto, extrair valor de variadas e volumosas fontes de dados é uma tarefa complexa. Adicionalmente, a dinâmica e as interações multifatoriais dos agroecossistemas exibem alta complexidade, o que torna a extração de novos conhecimentos e a modelagem altamente susceptíveis aos *drifts* de dados e de conceito. Há enormes expectativas e evidências científicas de que o desafio será travado nas áreas de inteligência artificial e ciência de dados. O sucesso da relação de ambas as áreas com a agricultura deve promover mútuos avanços científicos a todas as envolvidas, assim como contribuir para o desenvolvimento tecnológico de uma *agricultura inteligente plenamente orientada por dados*. Nesse contexto, o objetivo deste capítulo é dispor uma revisão crítica pautada pela literatura, firmando conceitos, acautelando pesquisadores e profissionais quanto à contemporização de áreas precursoras e apontando um paradigma capaz de agilizar e transformar entregas científicas em inovações efetivas. As soluções oriundas da inteligência artificial e da ciência de dados precisam estar comprometidas e/ou alinhadas com áreas histórica e intrinsecamente ligadas às ciências agrárias, como a estatística experimental e a geoestatística, visando credibilizar as entregas de novos valores. As ferramentas de apoio às decisões operacionais podem aproximar as entregas científicas das efetivas inovações nas cadeias agroalimentares. Nesse aspecto, os sistemas de recomendação ao produtor merecem prioridade pela capacidade de melhorar a utilização dos dados, aumentar a acurácia das estimativas dos modelos, customizar recomendações conforme as singularidades do sensoriamento e peculiaridades do agroecossistema e, finalmente, promover maior eficiência nas inovações ante os esforços científicos envidados na solução de problemas agrícolas.

**Palavras-chave:** inteligência artificial; ciência de dados; agricultura inteligente; sistema de recomendação.

## ABSTRACT

The development of digital and precision agriculture is one of the main areas that have promised a green revolution in agriculture. The cultivation field is the main area presenting scientific-technological challenges for agri-food chains and, additionally, the biggest provider of data for decision-making in rural areas. However, while different sensing technologies have become increasingly widespread in agriculture, extracting data from

<https://doi.org/10.4322/978-65-86819-38-0.1000071>

 Este é um capítulo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution-NonCommercial-NoDerivatives, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que sem fins comerciais, sem alterações e que o trabalho original seja corretamente citado.

varied and voluminous data sources is a complex task. Furthermore, the dynamics and multifactorial interactions of agroecosystems are highly complex, with the development of new knowledge and models being extremely susceptible to data and concept drifts. Scientific findings and professional expectations have indicated that the challenge will be met by employing tools from artificial intelligence and data science. A successful relationship between both areas and agriculture should promote mutual scientific advances, as well as supporting the technological development of fully data-driven, smart farming. Thus, the objective of this chapter is to provide a critical review guided by the literature, outlining concepts, drawing researcher's and professional's attention to the adoption of innovative areas, and highlight a paradigm capable of streamlining and transforming scientific input towards effective innovation. The solutions arising from artificial intelligence and data science need to be integrated and/or aligned with areas historically and intrinsically linked to agricultural sciences, such as experimental statistics and geostatistics, in order to make future data more reliable. Tools dedicated to supporting operational decisions can provide greater scientific assistance for effective innovation in agrifood production chains. In this regard, recommender systems deserve priority due to their capacity for improving data application, increasing the accuracy of model estimates, customizing recommendations according to sensing and agroecosystem particularities and, finally, encouraging more efficient innovation through integrating scientific efforts attained to solve agricultural problems.

**Keywords:** artificial intelligence; data science; smart farming; recommender system.

## 1 INTRODUÇÃO

O célere desenvolvimento tecnológico ocorrido nas duas últimas décadas tornou a aquisição, o processamento e o armazenamento de dados praticamente ilimitados. A denominada “era digital”, iniciada no sobrevir do século XXI, trouxe rápidas e inexoráveis mudanças na organização da produção, na logística, na gestão de processos, na comunicação, nos negócios e em muitos outros campos. Um verdadeiro tsunami causado pela transformação digital tem invadido praticamente todos os setores, segmentos e nichos, assim como tem mudado o comportamento e as relações sociais. Essa força motriz já revolucionou as telecomunicações, bens de capital, serviços, saúde, setor automotivo, transportes e outros (Clay; Kitchen, 2018). A transformação digital também deu início a uma nova fase da agricultura, inaugurando a denominada agricultura digital (AD), encerrada na intuitivamente versionada Agricultura 4.0, a qual promete provocar uma nova revolução verde na agropecuária (Marçal et al., 2022).

Embora não haja unanimidade entre especialistas e pesquisadores, a AD é consensual e conceitualmente compreendida como uma extensão da agricultura de precisão (AP) (Marçal et al., 2022). A AD compreende a integração de tecnologias digitais e de comunicação ao processo produtivo agrícola (Marçal et al., 2022). Trata-se, portanto, de um conjunto de estratégias de gestão que se apoia em soluções digitais para adquirir, armazenar, processar, analisar, propiciar interatividade (muitas vezes remota e em tempo real) e integrar dados de contextos espaciais, temporais, pontuais, econômicos, de processos agroindustriais e logísticos, visando otimizar a produtividade, sustentabilidade, qualidade, lucratividade, rastreabi-

lidade, tomada de decisões (gestão) e as práticas de manejo que permeiam os sistemas agroalimentares.

No campo produtivo, o sensoriamento ubíquo do *espaço agrícola* deve representar o maior desafio e principal provedor de dados complexos às tomadas de decisões das empresas rurais. O *sensoriamento remoto* (SR), por exemplo, realizado por meio de plataformas orbitais, estacionárias, aeronaves tripuladas, balões, veículos autônomos não tripulados e outras, tem ampliado significativamente o horizonte de aplicações digitais que visam maximizar a assertividade de gestão permitindo, assim, maior precisão na avaliação do crescimento vegetativo e reprodutivo das culturas, identificação precoce de estresses bióticos e abióticos, entre inúmeras outras.

O *volume* e a *variedade* dos dados de campo já se consolidam como desafios expoentes na Agricultura 4.0. É notável, portanto, que as cadeias agroindustriais devem ser depreendidas no conceitual e tecnologicamente denominado big data. No entanto, este último desvela uma camada extra de complexidade, bem conhecida em outras áreas, trazendo desafios, bem conhecidos, relacionados à eficiência da extração de *insights* (Tantalaki et al., 2019). A capacidade de extração de *valor* dos dados, informações úteis ou mesmo novos conhecimentos dessa intrínca e exponencialmente crescente massa de dados ainda é de veras limitada. De acordo com Morota et al. (2018), tais desafios podem ser enfrentados por meio da aplicação de técnicas de *aprendizado de máquina* (AM) e *mineração de dados* (MD). No prospecto científico-tecnológico da extração de valores dos dados deposita-se, portanto, considerável expectativa de que o AM e a MD auferam vantagens do big data pa-

ra enfrentar os grandes desafios da agricultura global (Tantalaki et al., 2019).

A *inteligência artificial* (IA) e a *ciência de dados* (CD) têm papéis imprescindíveis ao fortalecimento da AP e à consolidação de uma *agricultura inteligente* (SF, do inglês: *Smart Farming*) orientada por dados. Adicionalmente, a natureza altamente intrincada e intrínseca aos desafios agropecuários não pode ser concebida como mero destino de aplicação de técnicas e conhecimentos oriundos de IA e CD, mas assimilada como uma fonte de complexidade ímpar ao avanço científico de ambas as áreas. Inovações, decisões orientadas por dados e novos conhecimentos agropecuários dar-se-ão, fundamentalmente, pela concomitância do desenvolvimento da IA e da CD debruçadas nas particularidades dos desafios agropecuários. Todavia, o sucesso dos esforços científicos para alavancar as inovações na SF orientada por dados requer um *denominador comum*. Nesse contexto, o objetivo deste capítulo é firmar conceitos, apresentar uma revisão crítica com base na literatura e apontar um caminho para um denominador comum que fortaleça o elo entre ciência e inovação efetiva, ampliando assim a adesão sistemática da IA e da CD na SF plenamente orientada por dados.

## 2 CONCEITUALIZAÇÃO DA IA E DA CD

É bastante comum que IA e CD sejam tratadas como sinônimos nas literaturas técnicas e científicas, algo que decorre da forte interação e mútua troca de conhecimentos entre ambas as áreas. Para efeitos midiáticos e de aplicação, tal indistinção não gera desconcerto. O discernimento se faz mister no âmbito científico, ante o propósito de generalizar as contribuições oriundas da SF a outras áreas do conhecimento.

### 2.1 Inteligência artificial e aprendizado de máquina

A definição de IA é mais bem compreendida quando da descrição de seus propósitos. De acordo com Grosan e Abraham (2011), o propósito da IA é “criar máquinas que resolvem problemas que, quando solucionados por humanos, requerem inteligência”. Trata-se de uma área de pesquisa que se ocupa com o desenvolvimento de algoritmos que, programados em computadores, aprendem contextos particulares e são capazes de imitar e até superar a inteligência humana no desempenho de tarefas específicas.

O AM é uma subárea de pesquisa da IA, focada no desenvolvimento e aplicação de algoritmos dotados de capacidade de aprender e reconhecer padrões automaticamente a partir dos dados, utilizando o mo-

delo aprendido para generalizar decisões diante de novos casos.

Os modelos de aprendizado profundo (*Deep Learning* – DL) são constituintes de uma subárea particular de AM, com promissoras e intensas pesquisas. Derivam um arcabouço de técnicas que geralmente compartilham estruturas em camadas de algoritmos (unidades de computação) que extraem automaticamente características relevantes de dados não estruturados (ou semiestruturados) como, por exemplo, imagens, vídeos e áudios. Em geral, requerem grandes quantidades de dados para exibirem desempenhos satisfatórios nas suas aplicações. O diagrama estruturado na Figura 1 ilustra a grande área de IA e subáreas nela contidas.

### 2.2 Ciência de dados

De acordo com Kelleher e Tierney (2018), a CD abrange um conjunto de princípios, definições de problemas, algoritmos e processos para extração de padrões úteis e não óbvios a partir de grandes bases de dados. Tal definição a torna bastante similar ao *processo de descoberta de conhecimento em bases de dados* (*Knowledge Discovery in Databases* – KDD) – popular e sinteticamente referenciado apenas como DM –, mas que se singulariza por absorver desafios ainda mais intrincados como, por exemplo, aquisição, limpeza, transformação, fusão e extração de conhecimentos e *insights* de dados não estruturados e semiestruturados, embrenhando-se nas tecnologias de big data e em questões regulamentares e éticas. A CD se solidifica na nobre missão de recrutar as ciências exatas dedicadas aos dados nas tarefas intensivas de extração de *insights*, informações e novos conhecimentos.

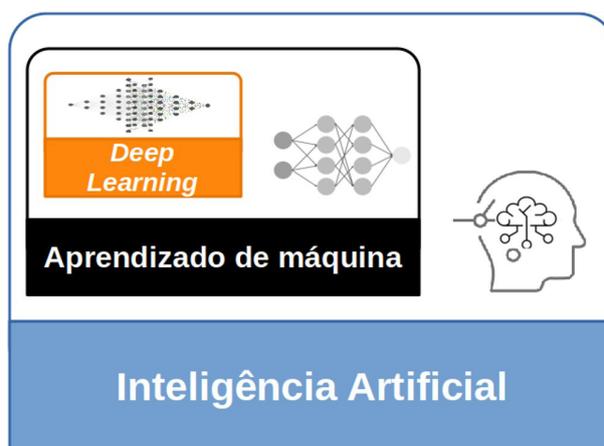


Figura 1. Inteligência artificial, aprendizado de máquina e Deep Learning.

Fonte: Autores.



**Figura 2.** Diagrama adaptado de Palmer (2015).  
**Fonte original:** Data Science for the C-Suite (Digital Living Press, 2015 print).

Em termos didáticos, o diagrama de Venn apresentado na Figura 2 ilustra as interseções entre a matemática, a computação e o domínio do conhecimento, confluência de grandes áreas que particulariza a CD como área que compartilha múltiplos conhecimentos para extração de valor dos dados de um domínio.

A evolução da matemática atada aos múltiplos desafios de aprendizagem a partir de dados deu origem à *estatística*, ciência pautada na extração sistemática de conhecimento e medições das incertezas. A estatística, como área *mater* e virtuosa na compreensão dos significados dos dados, baliza o principal argumento de uma corrente de pesquisadores que defende que a CD é um mito enquanto ciência (ou mesmo uma “*buzzword*”), um ideário sintetizado na interpretação de que a referida área é tão somente corpo metodológico e dados, os quais representariam apenas dois dos quatro elementos que definem uma ciência como tal (Learner; Phillips, 1993; Phillips, 2017). Por outro lado, uma corrente contrária acredita na formação de cientistas de dados sem a necessidade dos embasamentos teóricos atinentes à área de estatística (Hassani et al., 2021; Davison, 2018). O fato incontestado é que o cenário do big data trouxe desafios inéditos, particularmente quanto ao volume, velocidade transacional e, principalmente, variedade de dados não estruturados, aos quais apenas as contribuições pretéritas e atuais edificadas no campo da

estatística não catalisam os avanços rápidos e significativos ao contexto; não obstante, subestimá-los é tornar insólita a fundamentação de inúmeras técnicas da IA e da própria CD.

Não há como relegar os avanços significativos de modelagem oriundos da área de computação. Breiman (2001) já havia constatado essas duas frentes de modelagem de problemas, as quais definiu como *cultura de modelagem de dados*, oriunda da estatística/matemática, e *cultura de modelagem algorítmica*, descendida da pesquisa em computação. Logo, a compreensão da CD como ciência reside exatamente na dimensão compartilhada dessas duas culturas de pesquisa (vide interseção central no diagrama da Figura 2). De acordo com Saltz e Stanton (2017), a CD está profunda e equitativamente ligada à matemática, à estatística e à ciência da computação. Os avanços significativos na capacidade de lidar com problemas complexos são contribuições que procedem da CD (MacGillivray, 2021). Conforme reforça Hassani et al. (2021), estatística e CD são complementares e os profissionais envolvidos em cada uma das áreas devem buscar a ampliação da colaboração mútua.

O balanceamento dos conhecimentos de áreas envolvidas na CD merece ênfase especial na formação do cientista de dados dedicado ao contexto agrícola. O elo histórico e visceral entre estatística e agricultura para compreensão de fenômenos e extração do

conhecimento praticável é devidamente apreendido pelas ciências agrárias. Qualquer corrente de pensamento que possa desguarnecer o arcabouço estatístico-agronômico consolidado, em favor de uma CD exclusivamente fincada nas vertentes do *processamento de dados* ou do AM, será incompleta e pode suscetibilizá-la ao ceticismo enquanto área científica. As habilidades com estatística são importantes para os cientistas de dados (Ribeiro et al., 2017), mas ainda mais imprescindíveis àqueles comprometidos com os avanços da SF orientada por dados.

### 3 COMPLEXIDADE BIOGEOFÍSICA À MODELAGEM

As interações que ocorrem em agroecossistemas são altamente complexas e, muitas vezes, desconhecidas nos impactos pormenorizados. É de domínio geral que um agroecossistema está condicionado por uma rede complexa de múltiplos fatores macro, sendo mais conhecidos aqueles atrelados aos aspectos climáticos, da planta, do solo (pedológicos e edafológicos), dos recursos e disponibilidade hídrica, dos elementos e estresses bióticos (pragas e doenças), das demandas nutricionais, das singularidades espaciais e temporais, do manejo de intervenção, dentre outros. À vista destes, derivar modelos genéricos, confiáveis e reutilizáveis em agroecossistemas, altamente heterogêneos e variáveis no espaço-tempo, constitui um desafio sem igual quando da comparação com a intensiva aplicação de IA e CD em outros setores. A singularidade idiossincrática de um agrossistema dificulta a portabilidade de técnicas e de modelos previamente treinados, a despeito de condições que aparentam similaridades. Os denominados *drifts de dados e de conceito* tendem a ser consideravelmente mais severos nos desafios do sensoriamento agrícola, na comparação com outros segmentos bem-sucedidos na implementação de uma gestão plenamente orientada por dados.

Apoiar-se somente em modelos paramétricos representa considerável restrição no processo decisório de uma SF orientada por dados. De acordo com Verrelst et al. (2015), os modelos paramétricos são inadequados para derivar produtos biogeofísicos generalizáveis ou operacionais. Os modelos paramétricos pautam sua capacidade de generalização em premissas conceituais para induzir regressões válidas a partir dos dados experimentais. Entretanto, os dados do sensoriamento de campo muitas vezes se apresentam em configurações experimentais significativamente desiguais, com escalas distintas (níveis de folha, de planta ou de copa) ou por características dissemelhantes intrínsecas às tecnologias emprega-

das e/ou, ainda, sob condições ambientais bastante singulares (Verrelst et al., 2015). Em contrapartida, os modelos não paramétricos não estabelecem pressuposições severamente restritivas, sendo mais adequados e acurados em predições do rendimento da produção agrícola (Gonzalez-Sanchez et al., 2014). Há, portanto, indícios gerais em favor das técnicas de AM (não paramétricas), embora a suscetibilidade à obtenção de *modelos sensor-específicos* ainda imponha baixa transferibilidade ante os dispositivos usados no SR (Verrelst et al., 2015).

Atidos às aplicações de AM, Tantalaki et al. (2019) realizaram uma revisão crítica dos trabalhos científicos que empregaram AM na solução de problemas oriundos da agricultura. Dessa revisão, os autores aglutinaram as técnicas e concluíram o seguinte:

- As *redes neurais artificiais* (RNAs) encontram aplicações mais relacionadas ao SR dos campos agrícolas, sendo mais acuradas e adaptadas do que, por exemplo, as *máquinas de aprendizado extremo* (*Extreme Learning Machines – ELM*). Todavia, por serem dotadas de muitos parâmetros e elevada complexidade, tendem ao *overfitting*. Em adição, demandam grandes conjuntos de dados e métodos otimizados para regularização dos parâmetros ajustáveis do modelo;
- ELM e outras técnicas de neuro-fuzzy são vantajosas na capacidade de generalização do modelo gerado e apresentam desempenhos adequados, mas constituem modelos de “caixa preta” com baixa explicabilidade, demandando pesquisas para comprovação da confiabilidade;
- As árvores de decisão e de regressão são dotadas de maior compreensibilidade, mas são muito sensíveis às pequenas mudanças (adições, modificações ou remoções de instâncias) no conjunto de treinamento, o que as tornam instáveis e suscetíveis ao *overfitting*;
- As SVMs e os métodos de florestas aleatórias (*Random Forest – RF*) apresentam grande potencial para aplicações agrícolas, uma vez que são fáceis de implementar, rápidas na indução e robustas aos ruídos presentes nos dados de treinamento, além de apresentarem menor susceptibilidade ao *overfitting* na comparação com outros métodos. Essas têm sido amplamente utilizadas e consideravelmente populares na pesquisa agrícola, especialmente em problemas de classificação;
- As técnicas e arquiteturas de aprendizado profundo (*Deep Learning – DL*) têm sido promissoras no reconhecimento de padrões em imagens digitais de diferentes naturezas; por outro lado, demandam grande quantidade de dados para obtenção de precisão

e acurácia, algo que aumenta significativamente o custo computacional envolvido com a estrutura e o treinamento do modelo. Além disso, tendem fortemente ao *overfitting*, exigindo a aplicação e o ajuste cuidadoso das técnicas de regularização.

Os apontamentos de Verrelst et al. (2015) e Tantalaki et al. (2019) não intentam o estabelecimento de adequações ou inadequações de aplicações específicas. Ao contrário, são balizadores e indicadores de potencialidades, mantendo abertas as pesquisas em IA e CD nos agroecossistemas.

#### 4 CONTEMPORIZAÇÃO DE ÁREAS PRECURSORAS E CENÁRIO ATUAL

O enveredar da IA e da CD na agricultura encontra áreas precursoras e basilares, dentre elas a *estatística experimental* e a *geoestatística*. Ferreira (2018) ressalta a importância e obrigatoriedade do domínio dos princípios básicos da *experimentação estatística* para salvar a aplicação do método científico na elucidação dos princípios biológicos e solução dos problemas agropecuários. Ele destaca a importância da estatística experimental no melhoramento genético de plantas o que, por extensão, deve reger a fenotipagem de plantas. Andriotti (2003) apresenta, didaticamente, a importância do domínio da geoestatística para desenvolvimento das ciências da terra e do uso das ferramentas de SR. Ele condiciona, ainda, a compreensão da geoestatística a um domínio prévio da estatística. Os avanços da IA e, principalmente, da CD no desenvolvimento da SF precisam abarcar ambas as áreas a fim de maximizar as contribuições efetivas. A CD, como geradora de *insights* que podem ampliar o domínio do conhecimento e as inovações nas cadeias agroindustriais, deve incorporar o rigor da experimentação estatística para compreensão da causa-efeito.

As estatísticas que medem as efetivas contribuições da IA e da CD no balizamento dos processos decisórios dos sistemas agroalimentares são escassas. Ainda que as tecnologias digitais tenham viabilizado um notório aumento no volume e na diversidade de dados, pouco se sabe sobre a adesão e a sustentabilidade de suas aplicações para tomada de decisões nos ciclos de produção das cadeias. De acordo com um levantamento (amostral) realizado pela *Agricultural Resource Management Survey* (ARMS) do Departamento de Agricultura dos Estados Unidos (USDA), pouco mais de 65% dos produtores de milho que efetivamente adotaram os recursos do maquinário de monitoramento para coletar dados de colheita não os utilizaram sequer para gerar os mapas de produtividade (DeLay et al., 2020). No Brasil, as estatís-

ticas gerais quanto ao uso sistemático (não eventual) dos mapeamentos de solos, culturas e produtividade para tomada de decisões em ciclos subsequentes são desconhecidas ou pouco representativas das cadeias agroindustriais. É notável, portanto, o atual descompasso entre o universo que se desenha nas novidades científicas e o das inovações tecnológicas incorporadas de fato nas práticas agrícolas.

Quais são os caminhos para consolidação de uma agricultura plenamente orientada pela extração sistemática de valores dos dados (*data-driven agriculture*)? Esta é uma pergunta fundamental cuja resposta deve consolidar não apenas as áreas de IA e CD na agricultura digital inteligente mas agilizar e dinamizar as inovações no setor agrícola.

#### 5 TRILHA À INOVAÇÃO SISTEMÁTICA EM AGROECOSSISTEMAS

O principal desafio na AD é a agregação de valor a partir de diferentes fontes de dados (de campo, de safras, de maquinário agrícola, de aspectos econômicos e de outros) para extração de conhecimento (Molin et al., 2020). Nesse sentido, alguns trabalhos têm preconizado o desenvolvimento de *ferramentas de apoio à decisão* que visem os níveis operacionais (Yazdani et al., 2017; Bonfante et al., 2019). O propósito de tais ferramentas é, geralmente, aproximar as contribuições científicas das efetivas inovações nas cadeias agroalimentares.

Peters et al. (2020) apresentam uma abordagem inspirada em *sistemas de recomendação* ao produtor para maximizar a utilização de dados, melhorar as estimativas e a eficiência do esforço científico empenhado na solução de problemas agrícolas. Os sistemas de recomendação são amplamente utilizados nos mais variados tipos de negócios digitais, com inúmeros *cases* de sucesso. Trata-se de uma abordagem poderosa que viabiliza a utilização do conhecimento científico traduzido nos algoritmos inteligentes de recomendação, sendo ainda capazes de lidar com grandes volumes de dados agrícolas para prover recursos para antecipar a dinâmica dos agroecossistemas em constantes mudanças.

Os sistemas de recomendação são dotados de algoritmos que personalizam sugestões oferecidas ao usuário à medida que aprende particularidades extraídas do histórico de dados. Em geral, agregam técnicas/algoritmos de IA, AM e DL para melhorar a precisão das recomendações. Empresas como Google (YouTube), Amazon e Netflix estão entre os *cases* bem-sucedidos que mantêm investimentos contínuos em IA e CD para alavancar sistemas de recomendação assertivos e efi-

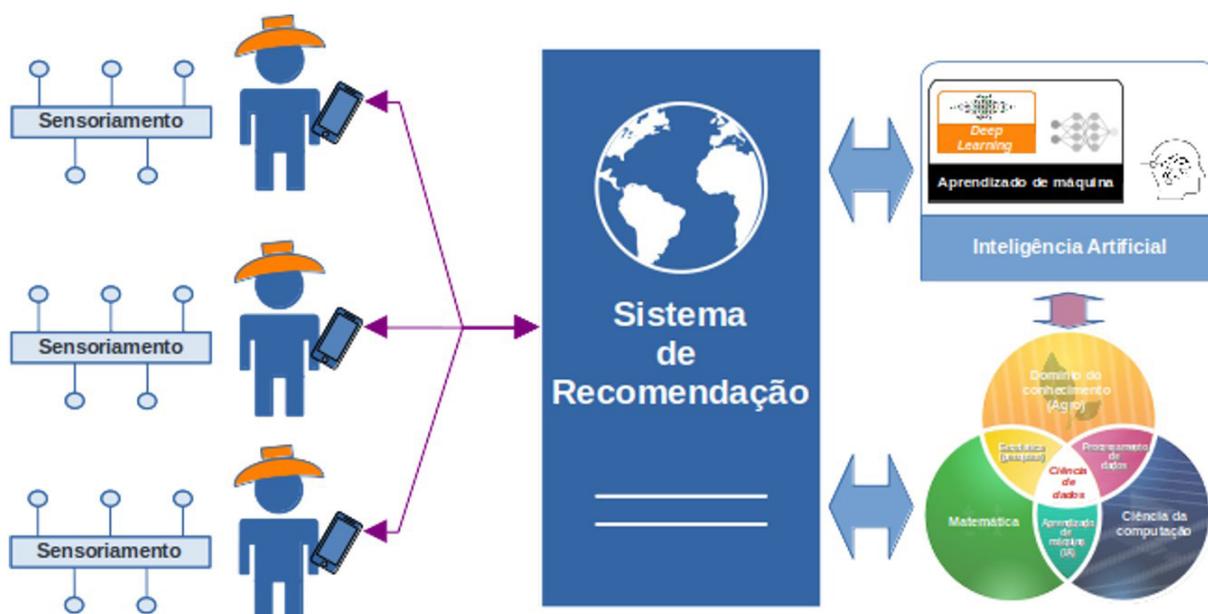


Figura 3. Sistema de recomendação na SF orientada por dados.  
Fonte: Autores.

cientistas. Na Figura 3 é apresentado um diagrama simplificado de um sistema de recomendação que customiza alertas e sugestões conforme especificidades do sensoriamento e características do sistema produtivo.

## 6 CONSIDERAÇÕES FINAIS

O grande desafio ao desenvolvimento de uma SF plenamente orientada por dados é a extração de valor de dados diversificados e de grandes volumes. As áreas de pesquisa em IA e CD merecem tônica na AP pelo potencial de impulsionar esse desenvolvimento. A complexidade inerente aos agroecossistemas demandam soluções e inovações que devem ampliar os conhecimentos do domínio de aplicação, bem como da autoevolução científica nela envolvida. A consolidação de uma SF orientada por dados não se resume no esforço de uma engenharia de aplicação atenta ao big data agrícola. Requer contemporização de áreas clássicas e um olhar focado na oferta imediata de valores extraídos de dados. O desenvolvimento de plataformas multifuncionais dotadas de sistemas de recomendação ao produtor é a ponte mais ágil entre pesquisa e inovação.

## AGRADECIMENTOS

Às unidades da Embrapa Instrumentação e da Agricultura Digital; ao Instituto de Química da Unesp de Araraquara; à UEL e ao Centro de Inteligência Artificial

(IBM-USP-Fapesp), particularmente aos colaboradores do desafio AgriBIO (C4AI-AgriBIO).

## REFERÊNCIAS

- ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. São Leopoldo: Editora Unisinos, 2003.
- BONFANTE, A.; MONACO, E.; MANNA, P.; DE MASCELLIS, R.; BASILE, A.; BUONANNO, M.; CANTILENA, G.; ESPOSITO, A.; TEDESCHI, A.; DE MICHELE, C.; BELFIORE, O.; CATAPANO, I.; LUDENO, G.; SALINAS, K.; BROOK, A. LCIS DSS-An irrigation supporting system for water use efficiency improvement in precision agriculture: A maize case study. **Agricultural Systems**, v. 176, p. 102646, 2019. DOI: <http://doi.org/10.1016/j.agry.2019.102646>.
- BREIMAN, L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). **Statistical Science**, v. 16, n. 3, p. 199-231, 2001. DOI: <http://doi.org/10.1214/ss/1009213726>.
- CLAY, D.; KITCHEN, N. **Precision agriculture basics**. New York: American Society of Agronomy, 2018. 287 p.
- DAVISON, J. No, machine learning is not just glorified statistics. Medium, 2018. Disponível em: <https://towardsdatascience.com/no-machine-learning-is-not-just-glorified-statistics-26d3952234e3>. Acesso em: 28 nov. 2022.
- DELAY, N.; Thompson, N.; Mintert, J. **Farm data usage in commercial agriculture**. Purdue: Purdue University, 2020. Disponível em: <https://ag.purdue.edu/commercialag/home/resource/2020/01/farm-data-usage-in-commercial-agriculture/>. Acesso em: 28 nov. 2022.

- FERREIRA, P. V. **Estatística experimental aplicada às ciências agrárias**. Viçosa: Editora UFV, 2018.
- GONZALEZ-SANCHEZ, A.; FRAUSTO-SOLIS, J.; OJEDABUSTAMANTE, W. Predictive ability of machine learning methods for massive crop yield prediction. **Spanish Journal of Agricultural Research**, v. 12, n. 2, p. 313-328, 2014. DOI: <http://doi.org/10.5424/sjar/2014122-4439>.
- GROSAN, C.; ABRAHAM, A. **Intelligent systems**. A modern approach. 17. ed. Berlin: Springer-Verlag, 2011.. DOI: <http://doi.org/10.1007/978-3-642-21004-4>.
- HASSANI, H.; BENEKI, C.; SILVA, E. S.; VANDEPUT, N.; MADSEN, D. O. The science of statistics versus data science: what is the future? **Technological Forecasting and Social Change**, v. 173, p. 121111, 2021. DOI: <http://doi.org/10.1016/j.techfore.2021.121111>.
- KELLEHER, J. D.; TIERNEY, B. **Data Science**. Massachusetts: MIT Press, 2018.. DOI: <http://doi.org/10.7551/mitpress/11140.001.0001>.
- LEARNER, D. B.; PHILLIPS, F. Y. Method and progress in management science. **Socio-Economic Planning Sciences**, v. 27, n. 1, p. 9-24, 1993. DOI: [http://doi.org/10.1016/0038-0121\(93\)90025-E](http://doi.org/10.1016/0038-0121(93)90025-E).
- MACGILLIVRAY, H. Statistics and data science must speak together. **Teaching Statistics**, v. 43, n. S1, 2021. DOI: <http://doi.org/10.1111/test.12281>.
- MARÇAL, D.; VALENTE, D. S. M.; CARVALHO PINTO, F. A.; BORÉM, A. **Agricultura digital**. São Paulo: Oficina de Textos, 2022. DOI: <http://doi.org/10.1007/978-3-031-14533-9>.
- MOLIN, J. P.; BAZAME, H. C.; MALDANER, L.; CORREDO, L. P.; MARTELLO, M.; CANATA, T. F. Precision agriculture and the digital contributions for site-specific management of the fields. **Revista Ciência Agronômica**, v. 51, n. 5, 2020. DOI: <http://doi.org/10.5935/1806-6690.20200088>.
- MOROTA, G.; VENTURA, R. V.; SILVA, F. F.; KOYAMA, M.; FERNANDO, S. C. Big Data Analytics And Precision Animal Agriculture Symposium: machine learning and data mining advance predictive big data analysis in precision animal agriculture1. **Journal of Animal Science**, v. 96, n. 4, p. 1540-1550, 2018. DOI: <http://doi.org/10.1093/jas/sky014>.
- PALMER, S. **Data Science for the C-Suite**. New York: Digital Living Press, 2015.
- PETERS, D. P. C.; SAVOY, H. M.; RAMIREZ, G. A.; HUANG, H. AI recommender system with ML for agricultural research. **IT Professional**, v. 22, n. 3, p. 30-32, 2020. DOI: <http://doi.org/10.1109/MITP.2020.2986125>.
- PHILLIPS, F. A perspective on 'Big Data. **Science & Public Policy**, v. 44, n. 5, p. 730-737, 2017. DOI: <http://doi.org/10.1093/scipol/scx012>.
- RIBEIRO, V.; ROCHA, A.; PEIXOTO, R.; PORTELA, F.; SANTOS, M. F. Importance of statistics for data mining and data science. In: INTERNATIONAL CONFERENCE ON FUTURE INTERNET OF THINGS AND CLOUD WORKSHOPS (FiCloudW), 5., 2017, Prague. **Proceedings [...]**. Prague: IEEE, 2017. p. 156-163. DOI: <http://doi.org/10.1109/FiCloudW.2017.86>.
- SALTZ, J. S.; STANTON, J. M. **An Introduction to data science**. Los Angeles: SAGE Publications, 2017
- TANTALAKI, N.; SOURAVLAS, S.; ROUMELIOTIS, M. Data-driven decision making in precision agriculture: the rise of big data in agricultural systems. **Journal of Agricultural & Food Information**, v. 20, n. 4, p. 344-380, 2019. DOI: <http://doi.org/10.1080/10496505.2019.1638264>.
- VERRELST, J.; CAMPS-VALLS, G.; MUÑOZ-MARÍ, J. RIVERA, J. P.; VEROUSTRAETE, F.; CLEVERS, J. G. P. W; MORENO, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 108, p. 273-290, 2015. DOI: <http://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- YAZDANI, M., ZARATE, P.; COULIBALY, A.; ZAVADSKAS, E. K. A group decision making support system in logistics and supply chain management. **Expert Systems with Applications**, v. 88, p. 376-392, 2017. DOI: <http://doi.org/10.1016/j.eswa.2017.07.014>.