

Big Data e Aprendizado de Máquina na Agricultura

João Pedro dos Santos¹, Emili Everz Golombiési², Maria Salete Marcon Gomes Vaz³

¹Acadêmico do curso de Engenharia de Computação - UEPG, Ponta Grossa, Brasil
(21007326@uepg.br)

²Mestranda do Programa de Pós-Graduação em Computação Aplicada - UEPG, Ponta Grossa, Brasil

³Professora Permanente do Programa de Pós-Graduação em Computação Aplicada - UEPG, Ponta Grossa, Brasil

Resumo: O crescimento exponencial de dados impulsiona o Big Data. Esse estudo explora a integração e análise de dados nos contextos de Big Data, Machine Learning aplicados principalmente à agricultura, focando nos métodos de análise de dados, avanços nos últimos anos, além de estudos de caso relacionados aos desafios e oportunidades em diversas áreas. O objetivo é sintetizar o conhecimento e direcionar pesquisas e aplicações da análise de dados em larga escala.

Palavras-chave: análise de dados, big data, processamento de dados

INTRODUÇÃO

Nos últimos anos testemunhamos um aumento global na geração de dados nas mais diversas áreas. Bailly et al. (2018) apontou que 90% das informações e dados presentes na internet, desde seu surgimento, foram criados recentemente, o que aponta um crescimento exponencial para as próximas décadas.

Esse fenômeno resultou em um cenário caracterizado pelo alto volume, alta velocidade e alta variedade de dados, amplamente conhecido como *Big Data*, um conceito aprofundado por Davis et al. (2018), como um grande volume de dados que requer novas arquiteturas e tecnologias para o processamento e extração de resultados, usando o aprendizado de máquina. Nessa magnitude, a entrada e análise de dados manual não faz mais sentido, conforme Trnka et al. (2018).

Essa explosão de dados oferece oportunidades e aplicações em diversas áreas como, por exemplo, na agricultura, como demonstrado por Shastri et al. (2019). A análise de dados de larga escala possibilita transformar a agricultura, melhorando as previsões de colheita, agricultura de precisão, geração de sementes de alta qualidade, dentre outras.

Esses dados podem ser úteis nas engenharias, onde são incluídos estudos de caso, como elaborado por Lavasani et al. (2021), permitindo reconhecer padrões e tomar decisões rápidas e inteligentes.

Diante desse cenário, este estudo se propôs explorar o papel desempenhado pela integração e análise de dados em *Big Data*, investigando o estado atual dos métodos de análise em larga escala, os avanços nos últimos anos, bem como os desafios e as oportunidades na utilização de *Big Data* e o

aprendizado de máquina, elaborando uma síntese, apontando direções de pesquisas bem como para o desenvolvimento de aplicações de *Big Data*.

MATERIAL E MÉTODOS

Para atingir os objetivos mencionados, foi realizada uma revisão de literatura que permitiu diagnosticar qual o estado da análise e gerenciamento de dados de *Big Data*, as aplicações que tem apresentado maior potencial e quais foram os principais avanços na integração dessa tecnologia nas mais diversas áreas, especialmente na agricultura e agronomia.

Foram encontradas as mais variadas oportunidades de melhoria em processos como previsões agrícolas quanto aos grãos, cultivos, condições, diagnósticos de grãos em geral, um cultivo mais preciso e claro, proporcionando a melhoria na tomada de decisões.

Na revisão de literatura, foram consultados repositórios de artigos científicos como o Google Acadêmico e SciELO, buscando pelas palavras chaves: *Big Data*, *Machine Learning*, *Data Analysis*, *Large Scale Systems*, *Opportunities in Agronomy*.

Em seguida foi realizada a análise de estudos de caso com implementações dos conceitos de gerenciamento de dados em larga escala bem como a implementação de modelos de aprendizado de máquina para a análise desses dados, demonstrando na prática quais as aplicações que trouxeram resultados em seus campos.

Finalmente, com os resultados da revisão de literatura e das análises de caso, foi elaborada uma síntese sobre o estado atual das tecnologias envolvidas bem como quais os avanços da pesquisa sobre o assunto, identificação das idéias principais e a elaboração da síntese em forma de tabela.

RESULTADOS E DISCUSSÃO

Esta pesquisa evidenciou a forma que sistemas de larga escala de dados se tornaram fundamentais nas tecnologias presentes no dia a dia. Um estudo realizado por Alani (2020) demonstrou que no ano de 2020, a quantidade de usuários ativos da internet passou de 4.57 bilhões, mostrando uma alta adoção da internet para resolver problemas cotidianos.

Esse aumento fez com que a quantidade de informações geradas e armazenadas aumentasse da mesma forma, onde também no ano de 2020, mais de $2,5 \times 10^{12}$ bytes foram gerados diariamente, com isso, se fez necessária a adoção de sistemas que sejam capazes de analisar todos esses dados e tirar conclusões úteis para as organizações que os implementaram, considerando a inviabilidade da análise manual ou funcional nesses casos.

A principal estratégia adotada por organizações que enfrentaram esse desafio foi a implementação de algoritmos de aprendizado de máquina, os quais foram capazes de analisar esses dados em larga escala, como apresentado por Lwakatarea et al. (2020), que demonstraram diversas áreas com integração de *Big Data* com o aprendizado de máquina, com a síntese apresentada na Tabela 1.

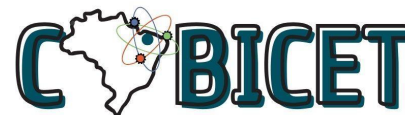
Tabela 1. Implementações de Sucesso de *Big Data*

Área de Trabalho	Implementação
Agricultura	Os avanços na capacidade computacional em termos de velocidade e volume permite novas técnicas de análise em grandes volumes de dados, como dados climáticos, dados coletados por máquinas e outros sensores agrícolas. Isso implica na análise dos dados em diversas circunstâncias, no gerenciamento de plantios e culturas, apontado por Coble et al. (2018). A previsão com maior precisão do valor agregado ao plantio, com o desenvolvimento dos grãos, considerando dados climáticos da estação, sistemas de irrigação inteligentes com dados coletados por sensores e detecção de doenças nas colheitas, dentre outras.
Serviços Online	Sistemas colecionam quantias de dados, de forma contínua, como é o caso dos produtos Google e Facebook, usando algoritmos de aprendizado de máquina num conjunto de

	dados em larga escala, coletados na forma que os usuários interagem com suas plataformas, permitindo a recomendação de anúncios e produtos de interesse. O mesmo pode ser aplicado da forma que as recomendações de vídeos no YouTube e Netflix funcionam.
Indústria Automotiva	Os dados obtidos pelos sistemas de análise de dados em larga escala têm sido usados no desenvolvimento de tecnologias de assistência ao motorista, os quais têm sido implementados cada vez mais em nível de consumidores.
Aviação	As aeronaves contam com sistemas autônomos, onde tanto o <i>hardware</i> quanto o <i>software</i> utilizam algoritmos de aprendizado de máquina em grandes conjuntos de dados obtidos pelos sensores do avião.
Telecomunicações	Técnicas de aprendizado de máquina são usadas sobre conjuntos de dados relacionados ao controle da qualidade da experiência dos usuários.
Medicina	Os sistemas avançados de imagem e diagnóstico tem a capacidade de coletar um conjunto maior de dados, processados e utilizando sistemas de aprendizado de máquina, conforme Carvalho et al. (2020), onde a combinação de <i>Big Data</i> de aprendizado de máquina tem revolucionado o diagnóstico e o tratamento de pacientes, melhorando o diagnóstico precoce, otimizando terapias e melhorando ensaios clínicos.

Alguns estudos, como de Giri et al. (2022) e Mathur et al. (2020) apresentaram referências ao Apache Hadoop e Spark como uma solução para a análise de dados em larga escala.

Contudo, a pesquisa elaborada por Divya et al. (2018) explica que essas tecnologias são programas de código aberto, escrito em Java, e que fornece uma plataforma para armazenar, processar e analisar grandes volumes



de dados de forma escalável e distribuída, podendo ser usada para treinar modelos de linguagem de aprendizado de máquina juntamente com o Apache Spark.

Na Tabela 2 é possível verificar a síntese acerca das tecnologias de integração e aprendizado de máquina no contexto de *Big Data*.

Tabela 2. Integração do Aprendizado de Máquina em *Big Data*.

Tópico	Síntese
Crescimento dos Dados	A quantidade de usuários da internet e de dados gerados cresce exponencialmente, tornando inviável a análise manual.
Dados na Internet	Mais de $2,5 \times 10^{12}$ bytes de dados foram gerados diariamente, na internet, em 2020.
Algoritmos de Aprendizado de Máquina	O aprendizado de máquina permite analisar dados em larga escala e chegar em conclusões úteis sobre eles.
Implementações de Sucesso	Agricultura, serviços online, indústria automotiva, aviação, telecomunicações e medicina são exemplos.
Apache Hadoop e Spark	Permitem o armazenamento, processamento e análise de grandes volumes de dados.
Benefício	A implementação dessas tecnologias leva a melhor tomada de decisões, maior eficiência e competitividade.

CONCLUSÃO

Neste estudo, foi elaborada uma síntese baseada na revisão de literatura sobre o tema de análise e gerenciamento de dados em larga escala (*Big Data*) em conjunção com aprendizado de máquina, com o objetivo de identificar tendências, desafios e oportunidades na agricultura.

Por meio dessa análise, foi sintetizado o estado atual do conhecimento, os principais avanços na área nos últimos anos e uma abordagem que tem sido utilizada com sucesso, contribuindo para apontar direções para pesquisas futuras sobre o assunto.

A análise de tendências, desafios e oportunidades nesse campo requer uma compreensão abrangente da interseção entre sistemas de larga escala de dados e as demandas tecnológicas do mundo contemporâneo.

É evidente que a proliferação dos usuários da internet, conforme destacado por diversos estudos, têm impulsionado uma explosão na geração e

armazenamento de informações. Este fenômeno, por sua vez, tem desafiado as organizações a lidar com volumes massivos de dados de forma eficiente, levando à adoção crescente de algoritmos de aprendizado de máquina.

Tais algoritmos, conforme demonstrado em várias pesquisas, têm sido implementados com sucesso em uma variedade de setores, desde serviços online até a medicina, levando a capacidade de extração de informações úteis a partir de conjuntos de dados em larga escala. Além disso, esses estudos mostram oportunidades na agricultura, com a coleta e análise de dados em larga escala possibilitando uma melhora na eficiência geral e tomada de decisões de coleta, plantio e gerenciamento das fazendas.

Em última análise, a síntese dessas tecnologias destaca não apenas os desafios enfrentados, mas também as oportunidades para aprimorar a tomada de decisões e impulsionar a inovação em um mundo movido por dados.

AGRADECIMENTOS

Os autores agradecem a UEPG e CNPQ pelo apoio concedido neste trabalho.

REFERÊNCIAS

- ALANI, M.; Big data in cybersecurity: a survey of applications and future trends. *Journal of Reliable Intelligent Environments*, v7. 2021.
- BAILLY, S.; MEYFROIDT, G.; TIMSIT, J. What's new in ICU in 2050: big data and machine learning. *Intensive Care Med*, v44. 2018.
- CARVALHO, D.; CRUZ, R.; Big data and machine learning in health. *European Journal of Public Health*, v30. 2020.
- COBLE, K.; MISHRA A.; FERRELL S.; GRIFFIN T.; Big Data in Agriculture: A Challenge for the Future. *Applied Economic Perspectives and Policy*, v40. 2018.
- DAVIS, A.; DIVYA, M.; VAKAYIL, T.; Addressing Big data. *International journal of engineering research and technology*, v3. 2018.
- DIVYA, K.; BHARGAVI, P.; JYOTHI, S.; *International Journal of Computer Sciences and Engineering*, v6. 2018.
- GIRI, P.; SHARMA G.; Apache Hadoop Architecture, Applications, and Hadoop Distributed File System. *Semiconductor Science and Information Devices*, v4. 2022.
- HWANG, H.; RYAN, L.; Statistical strategies for the analysis of massive data sets, v62. 2019.
- LAVASANI, M.; RAEISI, N.; GHAREBAGH, R.; ZARGHAMI, R.; Big data analytics opportunities



for applications in process engineering. Reviews in Chemical Engineering, v39. 2021.

LWAKATAREA, L.; RAJA, A.; CRNKOVICA, I.; BOSCHA, J.; OLSSON, H.; Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. Information and Software Technology, v127. 2020.

MATHUR, N.; JAIN M.; A Distribution of Nodes in Big Data using Hadoop Open Source System, v9. 2020.

SCHOLAR, Google. Google Scholar. Disponível em: <https://scholar.google.com/>. Acesso em: 22 mar. 2024.

SciELO. Disponível em: <https://www.scielo.org/>. Acesso em: 22 mar. 2024.

SHASTRY, K.; SANJAY, H.; Data Analysis and Prediction Using Big Data Analytics in Agriculture. Internet of Things and Analytics for Agriculture, v2. 2019.

TRNKA, A.; Big Data Analysis. European Journal of Science and Technology, v14. 2018.