

DESAFIOS MATEMÁTICOS EM *BIG DATA*, MINERAÇÃO DE DADOS E INTERNET DAS COISAS EM ADMINISTRAÇÃO, SAÚDE E EDUCAÇÃO FÍSICA

BRITO, Diogo de Freitas^{1,2,3}; EVERTON, Adriana Nunes da Fonseca⁴; BRASIL, Roxana Macedo⁵; BARRETO, Ana Cristina Lopes y Glória¹; CARVALHO JUNIOR, Sergio⁶; JUNIOR, Homero da Silva Nahum^{1,7}

Resumo

O objetivo da pesquisa foi discutir os desafios matemáticos de Big Data, mineração de dados e internet das coisas. Essas consequências tecnológicas modificaram o processo de tomada de decisão e a personalização de serviços. Na Administração, favoreceram os planejamentos logístico, comunicação, operacional e estratégico, assim como promoveram modelos de negócio baseados em dados. Na Saúde, possibilitaram monitoramento remoto de pacientes, diagnóstico precoce e personalização de intervenções em diversas disciplinas. A Educação Física se beneficiou de sensores que coletam dados biomecânicos e fisiológicos em tempo real, potencializando prescrição de treinos, prevenção de lesões e análise de desempenho, bem como avaliação e monitoramento contínuos. Na interseção, a obtenção de conhecimento permite a transformação de setores em modelos eficientes, preventivos e centrados no indivíduo.

Palavras-chave: Negócio. Ciência de Dados. Tecnologia da Informação. Gerência de Dados. Computação Científica.

Abstract

The objective of this research was to discuss the mathematical challenges of Big Data, data mining, and the Internet of Things. These technological consequences have modified the decision-making process and the personalization of services. In Administration, they have favored logistical, communication, operational, and strategic planning, as well as promoting data-driven business models. In Healthcare, they have enabled remote patient monitoring, early diagnosis, and personalized interventions in various disciplines. Physical Education has benefited from sensors that collect biomechanical and physiological data in real time, enhancing training prescription, injury prevention, and performance analysis, as well as continuous evaluation and monitoring. At the intersection, the acquisition of knowledge allows the transformation of sectors into efficient, preventive, and individual-centered models.

Keywords: Business. Data Science. Information Technology. Data Management. Scientific Computing.

¹ Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

² Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

³ Consultor Iceberg Business Academy;

⁴ Profissional de Educação Física da Adriana Nunes Consultoria de Corrida;

⁵ Docente Ph.D. em Educação Física;

⁶ Pesquisador convidado Bidesa;

⁷ Docente da Escola de Saúde da Universidade Cândido Mendes.

Introdução

Aspectos como digitalização, políticas de dados abertos e expansão da pesquisa científica, particularmente ocorridos na última década, promoveram o crescimento acelerado e significativo do volume de dados produzidos em negócios, saúde e educação física. Essa, particularmente, tenderia à elevação da produção acadêmica e digitalização de dados nos domínios da formação profissional (Nascimento *et al.*, 2024), práticas pedagógicas (Duarte e Neira, 2022) e saúde corporal (Bezerra e Pinheiros, 2022).

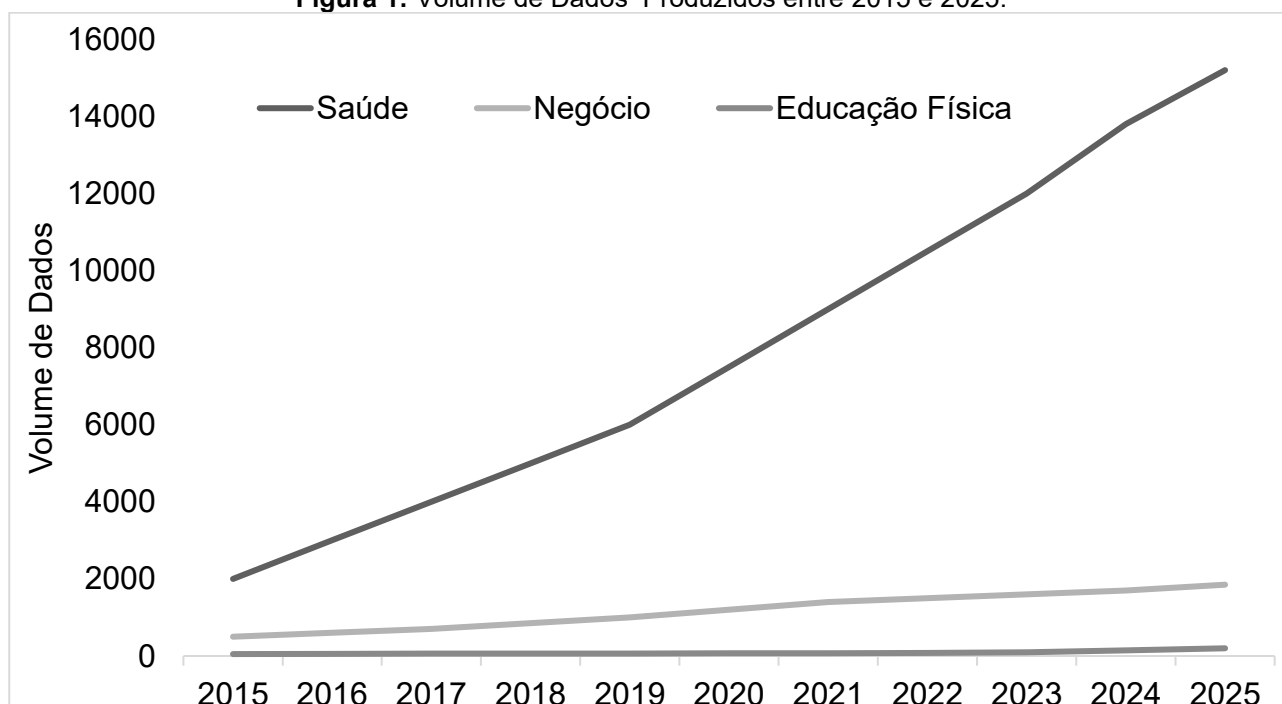
No contexto da saúde, a expectativa seria de crescimento exponencial de dados digitais, sobretudo após a pandemia de Covid-19, para vigilância epidemiológica conforme substancializado pela Pesquisa Nacional de Saúde realizada em 2013 e 2019 pelo Ministério da Saúde e Instituto Brasileiro de Geografia e Estatística, consolidou grandes volumes de dados populacionais sobre morbidade, estilos de vida e acesso a serviços (Ministério da Saúde, 2020; Stopa *et al.*, 2020). O desenvolvimento do Centro de Integração de Dados e Conhecimentos para Saúde - Cidacs (Fiocruz Bahia) em 2016, objetivando coadunar grandes bases de dados de saúde pública para investigações sobre desigualdades e apoio às políticas públicas (Barreto *et al.*, 2019; Barbosa *et al.*, 2020; Rocha *et al.*, 2022; Almeida *et al.*, 2024; Barreto *et al.*, 2025), ratificaria o exposto. Ainda na área de saúde, o Portal de Dados Abertos do SUS (OpenDataSUS) reuniu volumosos conjuntos de dados sobre assistência, e atenção primária e farmacêutica (Lima *et al.*, 2022; Reis-Santos, 2023; Boschiero *et al.*, 2024; Lira e Alchieri, 2024; Rangel *et al.*, 2025; Leandro *et al.*, 2025).

No condizente aos negócios, a produção, o armazenamento e processamento de bancos de dados densos não configurariam exceção, essa interseção se estenderia à necessidade de sistemas operarem conjuntamente de forma eficiente, segura e coordenada (interoperabilidade), utilizando inteligência artificial (Brito *et al.*, 2025; Barreto *et al.*, 2025) e análise preditiva (Junior, 2023; Junior, Brasil e Barreto, 2024; Brasil, Barreto e Junior, 2024; Barreto1 *et al.*, 2025; Junior *et al.*, 2025; Brasil *et al.*, 2025; Barreto2 *et al.*, 2025; Barreto3 *et al.*, 2025; Brasil2 *et al.*, 2025; Santos *et al.*, 2025; Brasil3 *et al.*, 2025; Brito2 *et al.*, 2025; Brito3 *et al.*, 2025; Brito4 *et al.*, 2025). Investigações sobre o gerenciamento de informações demonstraram o amadurecimento da discussão sobre informação e conhecimento, particularmente, no período de 1993 a 2023 (Castanha e Cazane, 2024), o que teria sido facilitado pelo Portal Brasileiro de Dados Abertos, que

disponibilizou diversas bases de dados relacionadas à Economia, finanças e gerência pública, com o favorecimento da interoperabilidade e facilidades de repositórios digitais (Silva e Oliveira, 2017; Moreira *et al.*, 2017; Silva *et al.*, 2020; Silva, Monteiro e Reis, 2020; Oliveira, Silva e Pereira, 2024; Oliveira e Silva, 2025).

Talvez, a digitalização tenha impulsionado com elevada magnitude a elaboração e disponibilização de conjuntos de dados, particularmente, na área de Saúde, a qual, também, foi impactada pelo elevado número de profissões pertencentes à esfera ora em tela, comparativamente à área de Negócios (Figura 1). Os bancos de dados relacionados à Educação Física seriam prioritariamente relacionados às políticas educacionais e formação acadêmica, o que explicaria, mesmo que parcialmente, a menor velocidade de crescimento. Na contemporaneidade, o sortilégio de interseções aceleraria a criação de outros conjuntos de dados e demandaria métodos, principalmente, matemáticos ao processamento deles para obtenção de informações. Nesse sentido, aplicações como *Big Data*, Mineração de Dados (*Data Mining*) e Internet das Coisas (IoT) se apresentariam como adequadas à realidade, transformando as etapas de coleta, análise e gerenciamento dos dados.

Figura 1: Volume de Dados Produzidos entre 2015 e 2025.



Fonte: Os Autores (2026).

Na saúde, o *Big Data* conquistaria destaque, particularmente, na Medicina de precisão e prontuários eletrônicos, enquanto a IoT aplicada seria ao monitoramento clínico (Rodrigues, Nóbrega e Dias, 2017; Saldanha, Barcellos e Pedroso, 2021). O Sistema Único de Saúde recepcionou a saúde digital como objetivo proposto pela Organização Mundial de Saúde, alicerçada na constatação de que “o campo do conhecimento e da prática associado ao desenvolvimento e ao uso de tecnologias digitais para melhorar a saúde” seria prerrogativa, tanto que no contexto nacional houve avanço da infraestrutura digital básica ao emprego de inteligência artificial e algoritmos preditivos (Niu *et al.*, 2019; Dong, Li e Nassif, 2021; Belseck, 2023). As práticas propedêuticas na Educação Física integrariam consequências tecnológicas digitais e *Big Data* para potencializar a prática educativa (Mendonça, 2024; Marques, 2025), o que envolveria a utilização de sensores para a obtenção de dados de desempenho, e o aprimoramento de métodos de ensino e dos processos de aprendizagem pela extração de informações em grandes massas de dados (Paiva, 2020; Pereira, 2022; Maia, 2024; Santos, 2025)

A *international Data Corporation* - IDC (2018) preconizou a criação mundial de mais de 175 zettabytes (10^{21} bytes = bilhão de terabytes) de dados relacionados aos negócios até 2025, o que demandaria profundidade na utilização de *Big Data* e mineração de dados (Galvão e Marin, 2009; Rodrigues e Sotto, 2022; Piccolo, 2025). Especificamente no caso brasileiro, as Organizações utilizariam diariamente 15 petabytes (10^{15} bytes = mil terabytes) de dados estruturados e não estruturados (DC, 2018). Todo o quadro com tendência de crescimento exponencial, demandando observações e discussões sobre conectividade e ética (Magrani, 2019; Silva, Bonacelli e Pacheco, 2020), planejamentos estratégicos público e privado (Santos e Freitas, 2016; Ferlin e Rezende, 2019) e equacionamento do tratamento de dados para a obtenção de informações e conquista de conhecimento (Santaella e Kaufman, 2021; Barreto, Brasil e Junior, 2024). Então, objetivou-se discutir os domínios matemáticos de *Big Data*, mineração de dados e internet das coisas, particularmente em negócios, saúde e Educação Física.

Big Data

Conceitualmente, *Big Data* seria conjuntos volumosos e complexos de dados, de tais magnitudes que métodos clássicos se tornaram inadequados ao armazenamento (guarda), processamento (transformação) e à análise (extração de padrões e significados). Isso decorreria das características (5V): 1) volume, a quantidade de dados alcançaria ordem de

zettabytes ou, pelo menos, exabytes (10^{18} bytes = milhão de terabytes); 2) variedade, reunindo dados estruturados (SQL e planilhas eletrônicas, por exemplo), cuja organização seria rígida e pré-definida, semiestruturados (Json e XML, dentre outros), carentes de estrutura rígida, mas organizados hierárquica e semanticamente, e não estruturados (textos, áudio, imagens e vídeos), quando inexistentes o modelo rígido e a organização interna, existindo em formato nativo; 3) velocidade, demandando geração e processamento em tempo real, como em cirurgia à distância, monitoramento fisiológico de atleta ou controle de processo, operação ou eficiência; 4) veracidade, o processamento deveria garantir a qualidade e confiabilidade, portanto, idealmente com nulidade de ruídos e inconsistências; e 5) valor, a transformação dos dados deveria favorecer a percepção ou ideia profunda sobre o fenômeno estudado (Camargo-Vega, Camargo-Ortega e Joyanes-Aguilar, 2015).

Pelo exposto, a história do *Big Data* estaria intrinsecamente associada à elevação de dados digitais e evolução da capacidade de armazenamento e processamento. Nas décadas de 1960 e 1970, o termo não havia sido cunhado, comumente a área de informática se referia à explosão de informações, todavia, o banco de dados relacional e os primeiros *data centers* eram realidade. Nos 20 anos seguintes, a utilização da internet se consolidou no público geral, favorecendo o comércio eletrônico (*e-commerce*), o que gerou maior volume de dados transacionais. Como consequência o armazenamento digital evoluiu, levando ao surgimento de armazém de dados (*data warehouse*), repositório central destinado à coleta, integração e guarda de grandes conjuntos de dados. Em 2001, Doug Laney formalizou as características de volume, velocidade e variedade do *Big Data*, que requisitava o desenvolvimento de código aberto para processamento distribuído. Em 2003, o sistema de arquivos distribuídos *Google File System* (GFS) seria lançado para gerenciamento de dados em larga escala, garantindo desempenho e confiabilidade. No ano seguinte, o modelo de programação *MapReduce* seria apresentado para processamento paralelo de volumes massivos de dados, dividindo-os em mapeamento (*Map*), transformação dos dados em pares chave-valor, e redução (*Reduce*), agregação e resumo daqueles pares para produção do resultado. Essas criações alicerçaram, em 2006, o desenvolvimento de estrutura de trabalho (*framework*) de código aberto para processamento e armazenamento distribuídos daqueles conjuntos de dados em *clusters* de *hardware* comuns, o *Hadoop*. Esse amadureceu, a partir da década de 2010, simultaneamente à computação em nuvem e captura de dados em redes sociais,

smartphones e sensores diversos, demandando o estabelecimento das características de veracidade e valor ao *Big Data* (Monleón-Getino, 2015; Pinto, 2023).

Contemporaneamente, o processamento distribuído e a infraestrutura em nuvem ainda seriam requisitos, desse conjunto participariam a utilização de Estatística avançada, aprendizagem de máquina e métodos de visualização, assim como o armazenamento escalável em bancos NoSQL (por exemplo, MongoDB e Cassandra), conjuntos não relacionais que utilizariam formatos flexíveis em detrimento de tabelas rígidas, potencializado escalabilidade e desenvolvimento ágil, e o *Hadoop Distributed File System* (HDFS), principal componente do Hadoop (Rautenberg e Carmo, 2019). Esse quadro permitiria predição de epidemias (Costa Filho, 2021), análise de imagens médicas (Santos *et al.*, 2024), monitoramento de desempenho atlético (Mănescu, 2025), melhora de treinos por análise biomecânica (Amadio, 2024; Pennone, 2025), recomendação personalizada no varejo (Félix, Tavares e Cavalcante, 2018; Maciel *et al.*, 2025) ou detecção de fraude (Rosini Filho, Rosini e Palmisano, 2020; Dias, Santos e Diniz, 2022). As tendências de pesquisa versariam sobre o processamento descentralizado, utilizando internet das coisas (Mohammadi *et al.*, 2018; Misra *et al.*, 2020), a capacidade de interpretação de modelos de inteligência artificial (*AI Explainability*), especificamente modelos de aprendizagem de máquina (Chandio *et al.*, 2024; Raz *et al.*, 2024) e a aceleração de algoritmos com o advento da computação quântica (Pacheco e Disconzi, 2019; Lima, Silva e Nascimento, 2025).

Para além do elevado custo na adequação de programas e equipamentos, as limitações do *Big Data* residiriam em desafios matemáticos em complexidade computacional, dimensionalidade dos problemas, otimização estocástica e privacidade diferencial. O primeiro aspecto faria referência à inadequação de algoritmos tradicionais à densidade dos bancos de dados, em razão do tempo demandado para completar o processo, exemplificando, a eficiência de: A) ordenação a qual poderia ter relação 1) linear ($O(n)$), para cada elemento n uma operação constante seria realizada; 2) logarítmico ($O(\log n)$), a cada passo, o problema seria reduzido à metade; 3) combinação dos anteriores ($O(n \cdot \log n)$), o banco seria dividido ao meio sucessivamente ($\log n$) até haver listas com um único elemento, as quais seriam reunidas em listas maiores e ordenadas (n); ou 4) quadrático ($O(n^2)$), o tempo seria proporcional ao quadrado da quantidade de registros; e B) agrupamento teria complexidade $O(n.k.d.i)$, ou seja, o tempo de execução seria o produto entre a quantidade de registros (n), número pré-definido de grupos (k), quantidade de características de cada dado (d) e total de iterações necessárias à convergência (i).

Então, esses algoritmos seriam mantidos em execução por longos períodos de tempo, o que tornaria as aplicações inviáveis. A solução residu nos métodos distribuídos (por exemplo, *MapReduce*), o qual pela paralelização reduziria o tempo de processamento, em razão da quantidade núcleos (p), para $T \approx \frac{T_{seq}}{p}$. Todavia, não ocorreria a anulação do dilema da escolha entre precisão (acurácia) e velocidade (rapidez ou latência), os quais apresentariam comportamentos inversos, assim, quanto maior a necessidade de acurácia, menor a velocidade de resposta (Fernandes, 2021).

A conhecida Maldição da Dimensionalidade (*Curse of Dimensionality*) significaria que em espaços de alta dimensão ($\gg 100$), a distância euclidiana seria desprovida de significado, pois a distribuição se tornaria esparsa, o que equivaleria a afirmar que todos os pontos estariam equidistantes ($\lim_{d \rightarrow \infty} \frac{Distância\ Máxima - Distância\ Mínima}{Distância\ Mínima} \rightarrow 0$). A solução poderia considerar a análise espectral, especificamente, pela aplicação de métodos baseados na teoria dos grafos (Aliferis, 2017), ou a redução do número de variáveis (dimensionalidade) por análise de componentes principais ou *t-distributed Stochastic Neighbor Embedding* - t-SNE (Incorporação estocástica de vizinhos com distribuição t), o primeiro seria determinístico (Abd Alwahab e Abd Alrazak, 2020), enquanto o segundo probabilístico (Kashyap, 2024).

O desafio da otimização estocástica consistiria em processar parâmetros como se variáveis aleatórias fossem, portanto, com comportamento descrito por alguma distribuição de probabilidade, o que potencializaria a toma de decisão em cenário caracteristicamente de incerteza. Objetivamente, consistiria em minimizar funções de perdas de dados massivos, o que ocorreria, por exemplo, em regressão logística sobre banco de dados com um terabyte (Lima, 2021; Araújo, 2024). Tradicionalmente, aplicado seria o Gradiente Descendente Estocástico (*Stochastic Gradient Descent* - SGD), algoritmo de otimização que ajustaria os parâmetros do modelo visando minimizar a função custo (Tang e Fong, 2018), para tanto, selecionaria aleatoriamente o mini-batch (Martins, 2024), pequeno subconjunto de dados para estimar o gradiente ($\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t; x_i, y_i)$).

O último problema faria morada na lacuna entre anonimato e não-identificação, na ciência da inexistência de implicação, logo, a privacidade diferencial seria o recurso matemático para extração de informações, mantendo a privatividade dos indivíduos, o que seria conquistado com a adição de ruído aleatório aos resultados ($\mathcal{M}(D) = f(D) + \text{Noise}(\Delta f/\epsilon)$; Δf = sensibilidade da consulta), o que, pelo menos, dificultaria a inferência

sobre a presença de determinado único indivíduo, portanto, atenuando a probabilidade de reidentificação, aspecto objetivado para dados sensíveis (Licks, 2016; Joaquim, 2022; Lopes, 2024; Silva, 2025). Comumente, empregados seriam o ruído gaussiano (I), seguido da distribuição Normal (a função densidade de probabilidade), ou seja, curva simétrica com concentração em torno da média, ou laplaciano (II), quando descrito pela distribuição de Laplace, pontiaguda ao centro e detentora de caudas densas (Meyer, 1987; Ross, 2010).

$$f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma}} \quad (\text{I})$$

Onde: x: variável aleatória; μ : média; σ : desvio padrão.

$$f(x; \mu, \beta) = \frac{e^{-\frac{|x-\mu|}{\beta}}}{2\beta} \quad (\text{II})$$

Onde: x: variável aleatória; μ : média; β : parâmetro de dispersão.

Mineração de Dados

Tratar-se-ia do processo de identificação de padrões, tendências, correlações e anomalias não explícitas sobre bancos de dados densos, possibilitando a predição de resultados e o apoio à tomada de decisões em diversas áreas (Coradine, Lopes e Maciel, 2011). Intuitivamente, poderia ser representada pela interseção entre Estatística, aprendizagem de máquina e ciência da computação em um Diagrama de Venn. Isso, porque, as raízes estariam fixadas na análise exploratória de dados (Fonseca e Namen, 2016) e nos métodos estatísticos (Silva *et al.*, 2019), especialmente com o desenvolvimento de algoritmos de classificação (Vieira *et al.*, 2024) e análise de grupos (Everton *et al.*, 2024) durante as décadas de 1960 e 1980. Em 1989, Gregory Piatetsky-Shapiro cunhou o termo *Knowledge Discovery in Databases* – KDD (Descoberta de Conhecimento em Bases de Dados), porém, a expressão *Data Mining* conquistou popularidade entre profissionais de bancos de dados no início dos idos de 1990. Tendo a área evoluído com as conferências em 1995 da *Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining* – ACM SIGKDD (Grupo de Interesse Especial da Associação para Computação focado em Descoberta de Conhecimento e Mineração de Dados), bem como pelo desenvolvimento de algoritmos de 1) árvores de decisão como *Classification and*

Regression Trees – CART (Árvores de Classificação e Regressão), *Iterative Dichotomiser* 3 – ID3 (Dicotomizador Iterativo) e C4.5, versão aprimorada do anterior; 2) regras de associação e conjuntos de itens frequentes, como o Apriori; 3) agrupamento de dados em quantidade determinada para minimizar as distâncias entre os pontos e o centróide do grupo, *k-means*; e 4) Máquina de Vetores de Suporte (*Support Vector Machine* – SVM) com amplo emprego em problemas complexos (dados não separáveis linearmente) de classificação e regressão, maximizando a distância entre grupos, por conseguinte garantindo solução robusta.

Nos anos 2000, ocorreu a integração com bancos de dados relacionais, armazéns de dados e ferramentas estatísticas (SPSS e SAS, por exemplo), o que favoreceu o desenvolvimento de aplicação em marketing (Cezar e Silva, 2022), bioinformática (Marmitt *et al.*, 2024), educação (Rigo *et al.*, 2014) e contabilidade (Pezzini, 2017), dentre outras. De 2010 em diante, intensificada foi a combinação com *Big Data* e aprendizagem de máquina, impulsionada pelo desenvolvimento de aprendizagem profunda (*Deep Learning*) para processamento de texto e imagem, e automação da construção de modelos de aprendizado de máquina (AutoML) para mineração em fluxo ou *streaming* (*Data Stream Mining*), aquela dedicada aos fenômenos com entrada contínua de dados, como transações de cartão de crédito ou tráfego de rede (Kumar, Steinbach e Tan, 2009). Essa entendida como tendência em pesquisa, assim como a mineração em grafos utilizada em redes sociais (Gomes, 2013) e epidemiologia (Alves, 2019).

Então, o foco da mineração de dados residiria em classificação, agrupamento e associação, para tanto, empregaria algoritmos supervisionados, quando o treino utilizar dados com a resposta correta, como nos casos de SVM e redes neurais, ou não supervisionados, não haveria fornecimento de resposta correta, por exemplo em *k-means* (Silva, Peres e Boscarioli, 2016), ou seja, em qualquer emprego, a escolha do algoritmo deveria priorizar a eficiência e no pré-processamento etapas de limpeza e normalização ($z = \frac{x-\mu}{\sigma}$) seriam necessárias. Esses detalhes seriam determinantes ao diagnóstico de câncer com emprego de SVM (Vieria Junior, 2013), à identificação de talento esportivo (Bonidia, Brancher e Busto, 2018) e previsão de perda de clientes (Martins e Silva, 2020).

Todavia, não se poderia negligenciar os desafios matemáticos, talvez aquele detentor de maior ocorrência fosse o equilíbrio entre *sobreajuste* (*overfitting*) e regularização. O primeiro seria o ajuste muito adequado aos dados de treinamento, mas a inadequação a novos dados, como consequência da memorização pelo modelo dos ruídos

e particularidades do grupo de treino em substituição ao aprendizado do padrão real. Efeito comum em modelos complexos, como redes neurais, entendido como compensação (conflito) entre complexidade e capacidade de generalização (Corcovia e Alves, 2019; Souza e Santos, 2021), portanto, denominando *tradeoff* viés-variância (*bias-variance tradeoff*). A regularização limitaria a liberdade do modelo, impondo penalidades à função de custo para evitar a complexidade excessiva, reduzindo a propensão à memorização e, conseqüentemente, elevando a capacidade de generalização, o que favoreceria a robustez e utilidade real do modelo. Classicamente, poderia ser aplicada a regularização L2 (Ridge ou Regressão de Ridge) ou L1 (*Least Absolute Shrinkage and Selection Operator* – Lasso (Operador de Seleção e Encolhimento Absoluto Mínimo) ou Regressão de Lasso). Esse utilizaria a soma dos valores absolutos (norma L1) para penalizar a função de custo pela possibilidade de zerar coeficientes, ao passo que a Ridge empregaria a soma dos quadrados dos coeficientes (norma L2). Como desfecho, esses raramente se tornariam nulos, mas reduzidos seriam a valores próximos de zero, por conseguinte, essa regularização não selecionaria automaticamente variáveis. Em suma, Lasso (III) seria adequada ao modelo cujo o resultado real seria conquistado com determinado subconjunto de variáveis ou na existência de demanda por modelos simples. A Regressão de Ridge (IV) deveria ser eleita quando a seleção de variáveis não for o objetivo ou, mesmo havendo *multicolinearidade* (muitas variáveis correlacionadas), a manutenção da influência de todas fosse necessária. Para bancos de dados demasiadamente complexos, necessário seria equilibrar seleção (L1) e redução de relevância (L2) de variáveis, essa combinação seria denominada *Elastic Net* (Fernandes e Chiavegatto Filho, 2019; Lima e Fagundes, 2020; Alves, 2022; Pereira, 2023; Rocha *et al.*, 2024).

$$\lambda \|\theta\|_1 \tag{III}$$

Onde: $\|\theta\|_1$: soma dos valores absolutos (penalidade); λ : magnitude da penalização (se elevado, maior a pena, forçando mais coeficientes a zero e aumentando esparsidade e regularização).

$$J(\theta) = \text{MSE}(\theta) + \lambda \|\theta\|_2^2 \tag{IV}$$

Onde: $J(\theta)$: função de custo total regularizada que o modelo buscaria minimizar; $MSE(\theta)$: erro quadrático médio (indicador da qualidade do ajuste); $\|\theta\|_2^2$: soma dos quadrados de todos os coeficientes; λ : magnitude da penalização (se elevado, coeficientes forçados a serem muito pequenos, aumentando o viés e diminuindo a variância; se zero, modelo se comportaria como regressão linear padrão; se pequeno, equilíbrio entre ajuste e manutenção de coeficientes pequenos).

A especificação da quantidade adequada de grupos (k) ao algoritmo *k-means* seria desafio, porque, inexisteria solução única, a inspeção visual não seria reveladora em alta dimensionalidade e o conhecimento sobre o fenômeno poderia ser insuficiente. Pragmaticamente, realizada seria a comparação das soluções para diversos k , considerando simplicidade do modelo e compacidade dos grupos, o que poderia ser feito pelo Método do Cotovelo (*Elbow Method*). Esse consistiria em estimar a Soma dos Quadrados das Distâncias Dentro dos Grupos (*Within-Cluster Sum of Squared Errors - WCSS*) para cada k , quanto menor o WCSS, mais compacto seria o grupo. Logo, a redução da estimativa acompanharia o aumento no número de grupos. A plotagem dos resultados permitiria identificar o ponto no qual WCSS mudaria abruptamente, o valor correspondente de k considerado seria ótimo (Monteiro e Araujo, 2023; Alves *et al.*, 2024; Luz e Camossi, 2025). Possível seria comparar a WCSS estimada com aquela oriunda de um conjunto de dados gerado aleatoriamente, isso seria denominado Estatística de GAP (*Gap Statistics Method*), cujo o maior valor elegeria o k (Maltauro, Guedes e Uribe-Opazo, 2025). Outra possibilidade seria o emprego do Índice de Silhueta (*Silhouette Analysis*), o qual estabeleceria o quanto cada ponto estaria adequadamente estabelecido em determinado grupo em relação ao *cluster* mais próximo, estabelecendo pontuação entre -1,0 (agrupamento inadequado) e +1,0 (agrupamento ótimo). A quantidade ideal de grupos seria aquela com maior pontuação média, indicando adequadas separação e compacidade (Semaan *et al.*, 2012; Silva, Brito e Oliveira, 2019; Vargas *et al.*, 2021).

Regras de Associação seriam padrões de relações entre itens transacionais (Junior, Brasil e Barreto, 2024), por exemplo, indivíduos cardiopatas tenderiam a apresentar comprometimento renal ou a compra de arroz estaria relacionada à aquisição de feijão. A identificação das ocorrências poderia orientar a indicação de terapias e cuidados, assim como a recomendação de aquisição de produtos ou serviços. Isso, porque, as regras seriam expressas como implicações lógicas (se x , então y), o que permitiria estimar a ocorrência do par (x, y); a probabilidade de y ocorrer, dado que x ocorreu; e a magnitude da ocorrência

de x sobre a ocorrência de y , essas métricas seriam denominadas, respectivamente Suporte, Confiança e Lift. O problema estaria na complexidade ($O(2^n)$), ou seja, a quantidade de regras se comportaria exponencialmente com o número de itens únicos (n), simplificada, a cada item, o tempo de execução do algoritmo duplicaria. Então, possível seria predefinir determinado valor Mínimo e eliminar associações sob a condição Suporte < Mínimo, processo denominado podar combinações (Carvalho *et al.*, 2012). Dentre os algoritmos utilizados estariam: 1) Apriori (Librelotto e Mozzaquatro, 2014), consideraria que se determinado conjunto de itens fosse infrequente, então os respectivos superconjuntos, também, o seriam (propriedade *anti-monótona* do Suporte); e 2) *Frequent Pattern Growth* - FP-Growth (Crescimento de Padrões Frequentes), distribuindo os dados em árvore (*Frequent Pattern Tree* – FP-Tree), identificaria os padrões frequentes em duas varreduras do banco de dados (Garcia *et al.*, 2025).

O último desafio matemático da Mineração de Dados se manifestaria na existência de classes desbalanceadas (*class imbalance*), caracterizada pela desproporção quantitativa, a partir de 90,00% para 10,00%, como rotineiramente se manifestaria em doenças raras, diagnóstico de saúde, transações fraudulentas ou falhas raras. Algoritmos de aprendizagem de máquina se tornariam tendenciosos em relação à classe majoritária, dado que assumiriam a existência de balanceamento (Barreto e Loula, 2019), resultando em solução enviesada. A solução passaria, por exemplo, pela geração de registros relativos à classe minoritária (*sobreamostragem* – *oversampling*), o que poderia ser realizado pelo algoritmo *Synthetic Minority Over-sampling Technique* – SMOTE (Rufino, Veiga e Nakamoto, 2016). Outra possibilidade seria aplicar a perda ponderada (*weighted loss*), a qual atribuiria distintas importâncias às diferentes classes ou observações (V), durante o treinamento do modelo (Scaioni, 2025).

$$L(y, \hat{y}) = -\sum w_i \cdot y_i \log(\hat{y}_i) \quad (V)$$

Onde: L : valor da perda; y_i : valor observado para a classe i ; \hat{y}_i : probabilidade prevista para a classe i ; w_i : importância da classe i .

Internet das Coisas

O conceito de IoT seria a conexão de objetos físicos com a internet, permitindo a coleta e troca de dados autônomas por meio de sensores, programas ou outra

consequência tecnológica. Logo, a integração entre os mundos físico e digital se substancializaria, favorecendo processos e ofertando conveniência pela automação e pelo controle remoto. Isso somente funcionaria, porque, dentre os objetos físicos haveria eletrônicos (assistente de voz, sistema de segurança e iluminação residencial, por exemplo), eletrodomésticos (geladeira, ar-condicionado e lavadoras, dentre outros), maquinário industrial e comercial (sensor de manutenção preventiva, medidores inteligentes de consumo de energia, controladores de demanda, analisadores de qualidade energética e inversores solares inteligentes, por exemplo), dispositivos eletrônicos vestíveis – *wearables* (pulseiras *fitness*, óculos inteligentes, *smartwatches* e roupas com sensores, dentre outros) e carros (direção autônoma, controle via aplicativo, controle de cruzeiro adaptativo, alerta de ponto cego e comunicação com outros veículos e infraestrutura de trânsito, por exemplo). Em suma, ampla rede artefatos cotidianamente utilizados por expressivo quantitativo de indivíduos (Sinclair, 2018; Chuquimarca e Maita, 2022; Santos, Faria e Vasconcelos, 2024).

Esse desenvolvimento se iniciou nas décadas de 1980 e 1990, precisamente em 1982, quando uma máquina de refrigerantes instalada no campus da *Carnegie Mellon University* (Pittsburgh, Pensilvânia, EUA) e conectada à *Advanced Research Projects Agency Network* – ARPANET (Rede da Agência para Projetos de Pesquisa Avançada), rede criada pelo Departamento de Defesa dos EUA para conectar universidades e centros de pesquisa, fornecia relatórios sobre o estoque. Em 1999, a expressão *Internet of Things* (Internet das Coisas) foi criada por Kevin Ashton, contudo, no domínio da cadeia suprimentos, utilizando *Radio-Frequency Identification* – RFID, ou seja, ondas de rádio para ler e armazenar dados em etiquetas, o que possibilitava rastreamento e gerenciamento remotos de itens, permitindo aplicações em logística, controle de acesso, inventários e segurança. Na década seguinte, a expansão da RFID ocorreu, levando ao desenvolvimento de sensores sem fio, atuadores conectados (dispositivo mecânico que converte energia elétrica, hidráulica ou pneumática em movimento) e padrões de comunicação como Zigbee (protocolo para redes em malha) e LoWPAN (arquitetura que possibilita o funcionamento do protocolo de internet em redes de sensores de baixo consumo). Isso embasou, a partir da década de 2010, diversas aplicações como cidades inteligentes (Martins *et al.*, 2024) e indústria 4.0 (Colombo e Lucca Filho, 2018), mais intensamente em saúde (Silva e Oliveira, 2017) e dispositivos vestíveis. Esse quadro manteria a relação de retroalimentação com o desenvolvimento de padrões e plataformas de comunicação (MQTT, CoAP, AWS IoT e

Google Cloud IoT, por exemplo), redução de custos de sensores e microcontroladores (Arduino e Raspberry Pi, por exemplo) e popularização de banda larga móvel e *smartphones* (Lima *et al.*, 2019; Mendieta, Herrera e Peña, 2019).

Principalmente, a conectividade omnipresente possibilitada pelos padrões de comunicação e o processamento de dados em tempo real estabeleceram tendências de investigação em integração com modelo de inteligência artificial para decisões autônomas e associação com redes 6G, as quais exigiriam a produção de chips para velocidades ultrarrápidas, latência em microssegundos e faixas de terahertz, além de protocolos de segurança por criptografia (Chowdhury *et al.*, 2019; Nguyen *et al.*, 2021; Guo *et al.*, 2021; Chen *et al.*, 2022; Ferrag *et al.*, 2023; Emek, Adilbekova e Kaya, 2025). Dessarte, possível foi o monitoramento remoto de eletrocardiograma (Morello *et al.*, 2022; Ikeda *et al.*, 2025), a execução de estádios inteligentes (Van Heck, Valks e Den Heijer, 2021; Mahdi, Aljuboori e Hussein, 2021) e inovação em negócios (Chauhan e Popat, 2025; Parra-Sánchez, 2025).

Todas as aplicações de IoT exigiriam a baixa latência, mais claramente, a troca de dados entre dispositivos em tempo real, idealmente sem atrasos, exemplificando, esperar 500 ms poderia ser excessivo, ou proporcionar a elevação de riscos, em carros autônomos, automação industrial e cirurgias assistidas remotamente. Esse problema seria originado na centralização do processamento em nuvem, o qual poderia ser realizado 1) em roteadores ou *gateways*, dispositivos físicos que serviriam de ponte central com a nuvem, coletando e pré-processando dados (*Fog Computing*); ou 2) no próprio dispositivo, independentemente, de ser câmera, relógio, *smatphone*, vestíveis ou semáforos (*Edge Computing*). Isso poderia reduzir a latência de mais de 100 ms para valores inferiores aos 10 ms (Alameddine *et al.*, 2019; Tran, Hoang e Bui, 2023). Outro desafio consistiria na fragilidade de segurança de dispositivos IoT, os quais seriam passíveis de invasão por *botnet Mirai*, rede de dispositivos infectados por *malware* homônimo, empregado em ataques distribuídos de negação de serviço (DDoS). A utilização do AES-256 (*Advanced Encryption Standard 256 bits*), algoritmo de criptografia simétrica que dividiria os dados em blocos de 128 bits, tornando-o ilegível, cuja a reversão demandaria a chave de 256 bits, o que significaria que haveria 2^{256} combinações possíveis. Em termos simples, o algoritmo realizaria 14 rodadas de substituição, transposição e combinação dos dados originais pelo emprego de operações matemáticas complexas. O trabalho de processamento de cada unidade de informação (byte) exigiria ≈ 3.000 ciclos, representando elevada eficiência (ou baixo custo computacional) para os dispositivos em questão (Ahamed *et al.*, 2019; Okpu e Taylor,

2025). Atualmente, o padrão de segurança na internet seria o *Transport Layer Security - TLS 1.3*, considerado robusto, sem comprometer a latência e garantindo a privacidade por ocultar informações do cliente e servidor (Perugini e Vesco, 2024).

A escalabilidade da IoT seria limitada pelo consumo de energia, não raramente os dispositivos utilizariam baterias, exigindo substituições frequentes, o que oneraria as aplicações, sensores se classificariam como exceção. Além disso, a vida útil de bateria dependeria da Lei de Joule, especialmente expressa pela equação de dissipação de calor ($E = I^2 \cdot R \cdot t$), a qual relacionaria corrente (i), resistência (R) e tempo (t). A partir dela, possível seria compreender que o aquecimento gerado pelo Efeito Joule durante o uso (carga e descarga) e armazenamento da bateria aceleraria a degradação química dos componentes, logo, reduzindo a longevidade, o que poderia ser atenuado pelo emprego de protocolos de baixo consumo (LoRaWAN e Zigbee, por exemplo) e gerenciamento do ciclo de trabalho (*Duty Cycling*), proporção de tempo em que o dispositivo estaria ligado (Santilio *et al.*, 2023).

Considerações Finais

Abstraindo-se da Administração, saúde e Educação Física, a Internet das Coisas produziria fluxos contínuos de dados. Esses conjuntos ofertados ao *Big Data* seriam coadunados a outros, armazenados e processados. Dessa forma, a Mineração poderia analisados para identificar padrões, desmascarar tendências e relevar associações. Como resultados a prospecção de cenários, predição de ocorrências e otimização de processos se materializariam, aspectos essenciais à tomada de decisão. Porém, demandadores dos desenvolvimentos de algoritmos, equipamentos e instrumentos, os quais retroalimentariam o quadro, enriquecendo e enriquecidos sendo pelas circunstancialidades das aplicações, as quais tenderiam a maiores complexidades no domínio do tempo.

Referências

- ABD ALWAHAB, OA; ABD ALRAZAK, MS. Using nonlinear dimensionality reduction techniques in big data analysis. **Periodicals of Engineering and Natural Sciences (PEN)**, v. 8, n. 1, p. 142-155, 2020.
- AHAMED, J *et al.* AES and MQTT based security system in the internet of things. **Journal of Discrete Mathematical Sciences and Cryptography**, v. 22, n. 8, p. 1589-1598, 2019.
- ALAMEDDINE, HA *et al.* Dynamic task offloading and scheduling for low-latency IoT services in multi-access edge computing. **IEEE Journal on Selected Areas in Communications**, v. 37, n. 3, p. 668-682, 2019.

ALIFERIS, CF. State of the science in big data analytics. In DELANEY, C *et al.* (eds). **Big data-enabled nursing**. Health informatics. New York (USA): Springer, 2017.

ALMEIDA, BA *et al.* CIDACS'efforts towards an inclusive and dialogic data governance in Brazil: a focused literature review. **International Journal of Population Data Science**, v. 9, n. 1, p. 2163, 2024. doi: 10.23889/ijpds.v9i1.2163.

ALVES, CL. **Diagnóstico de doenças mentais baseado em mineração de dados e redes complexas**. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo. São Carlos (SP), 2019.

ALVES, KA *et al.* Análise de dados dos Planos de Desenvolvimento Institucional do Instituto Federal do Pará (2009-2023) utilizando o algoritmo de aprendizado de máquina K-means. **Cuadernos de Educación y Desarrollo**, v. 16, n. 11, e6259, 2024.

ALVES, RD. **Predição da força de reação do solo durante a caminhada e corrida na água utilizando técnicas de mineração de dados**. Dissertação (Mestrado em Tecnologias da Informação e Comunicação) - Programa de Pós-graduação em Tecnologias da Informação e Comunicação. Universidade Federal de Santa Catarina. Araranguá (SC), 2022.

AMADIO, AC. Análise do movimento humano. **Revista Portuguesa de Ciências do Desporto**, p. 9-12, 2024.

ARAÚJO, VF. **Algoritmo colônia artificial de abelhas**: um estudo comparativo. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Computação) – Departamento de Computação e Tecnologia. Universidade Federal do Rio Grande do Norte. Caicó (RN), 2024.

BARBOSA, GCG *et al.* CIDACS-RL: a novel indexing search and scoring-based record linkage system for huge datasets with high accuracy and scalability. **BMC medical informatics and decision making**, v. 20, n. 1, p. 289, 2020. <https://doi.org/10.1186/s12911-020-01285-w>.

BARRETO, ACLG; BRASIL, RM; JUNIOR, HSN. O paradoxo de Moravec e a educação física. **Revista Presença**, v. 10, n. 22, p.273-289. 2024.

BARRETO, JSAM; LOULA, AC. Aplicação de mineração de dados para predição de mortalidade em UTI: balanceamento, dados ausentes e classificadores. *In*: BRAZILIAN E-SCIENCE WORKSHOP (BRESOI), 13, 2019, Belém. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2019. DOI: <https://doi.org/10.5753/bresoi.2019.10024>.

BARRETO, ML *et al.* (Org.). **Desigualdades sociais e saúde**: abordagens inovadoras para avaliar seus efeitos na população brasileira. Salvador (BA): UFBA, 2025.

BARRETO, ML *et al.* The centre for data and knowledge integration for health (CIDACS): linking health and social data in Brazil. **International Journal of Population Data Science**, v. 4, n. 2, p. 1140, 2019. doi: 10.23889/ijpds.v4i2.1140.

BARRETO1, ACLG *et al.* Genetischer Algorithmus für die Athletenauswahl: der Moneyballeffekt. **Revista Presença**, v. 11, n. 26, p. 382-395, 2025.

BARRETO2, ACLG *et al.* Planification de l'offre de cours sur une plateforme numérique à l'aide de la programmation linéaire em nombres entiers et d'heuristiques. **Revista Presença**, v. 11, n. 26, p. 333-344, 2025.

BARRETO3, ACLG *et al.* Cadeia de Markov a tempo discreto para predição de faixas de receita em escola de futevôlei. **Revista Presença**, v. 11, n. 26, p. 127-140, 2025.

BELSECK, N. Tech challenges & solutions in SA healthcare. **Medical Chronicle**, v. 2023, n. 9, p. 10-10, 2023.

BEZERRA, MAA. PINHEIRO, JK. (Org.). **Produções científicas em Educação Física**. Iguatu (CE): Quipá, 2022.

BONIDIA, RP; BRANCHER, JD; BUSTO, RM. Data mining in sports: a systematic review. **IEEE Latin America Transactions**, v. 16, n. 1, p. 232-239, 2018.

BOSCHIERO, MN *et al.* EP-247 - perfil clínico de pacientes hospitalizados devido a infecção pelo vírus influenza com co-deteção viral no Brasil. **The Brazilian Journal of Infectious Diseases**, v. 28, p. 104159, 2024. <https://doi.org/10.1016/j.bjid.2024.104159>

BRASIL, RM *et al.* Maximização da receita no negócio fitness. **Revista Presença**, v. 11, n. 25, p. 345-355, 2025.

BRASIL, RM; BARRETO, ACLG; JUNIOR, HSN. Modelagem temporal de clientes externos e ganhos líquidos no negócio personal trainer. **Revista Presença**, v. 10, n. 22, p. 253-272, 2024.

BRASIL2, RM *et al.* Análise multicritério na priorização de investimento esportivo. **Revista Presença**, v. 11, n. 26, p. 184-203, 2025.

BRASIL3, RM *et al.* Modelagem de Markowitz para determinar portfólio de investimento em academias de ginástica. **Revista Presença**, v. 11, n. 26, p. 247-261, 2025.

BRITO, DF *et al.* Processamento em linguagem natural para análise de sentimentos de adolescentes de escola de futebol. **Revista Presença**, v. 11, n. 26, p. 315-332, 2025.

BRITO2, DF *et al.* Regressão de Poisson para estimar o comparecimento às intervenções pelo treinamento físico ou à escola de futebol. **Revista Presença**, v. 11, n. 26, p. 286-299, 2025.

BRITO3, DF *et al.* Eficiência do negócio personal trainer sob o domínio da análise de envoltória de dados. **Revista Presença**, v. 11, n. 26, p. 262-275, 2025.

BRITO4, DF *et al.* Probabilidades ao planejamento de escolas esportivas: unidades e matrículas. **Revista Presença**, v. 11, n. 26, p. 229-246, 2025.

CAMARGO-VEGA, JJ; CAMARGO-ORTEGA, JF; JOYANES-AGUILAR, L. Conociendo big data. **Revista Facultad de Ingeniería**, v. 24, n. 38, p. 63-77, 2015.

CARVALHO, DR *et al.* Mineração de dados aplicada à fisioterapia. **Fisioterapia em Movimento**, v. 25, n. 3, p. 595-605, 2012.

CASTANHA, RG; CAZANE, AL. A evolução das tendências de pesquisa em gestão da informação e gestão do conhecimento: uma análise de 1993 a 2023. **REAd : Revista Eletrônica de Administração**, v. 30, n. 3, p. 1439-1461, 2024.

CEZAR, DAL; SILVA, EAA. Mineração de dados: uma estratégia para melhorar o negócio. **Revista Cogitare**, v. 5, n. 2, p. 1-15, 2022.

CHANDIO, SA *et al.* Enhancing trust in healthcare: The role of AI explainability and professional familiarity. **The Asian Bulletin of Big Data Management**, v. 4, n. 1, p. Science 4 (1)-21, 2024.

CHAUHAN, S; POPAT, K. The role of IoT and M2M communication in fostering business innovation and driving organizational growth. **Discover Internet of Things**, v. 5, n. 1, a. 132, 2025. <https://doi.org/10.1007/s43926-025-00231-4>

CHEN, R *et al.* Reconfigurable intelligent surfaces for 6G IoT wireless positioning: A contemporary survey. **IEEE Internet of Things Journal**, v. 9, n. 23, p. 23570-23582, 2022.

CHOWDHURY, MZ *et al.* The role of optical wireless communication technologies in 5G/6G and IoT solutions: Prospects, directions, and challenges. **Applied Sciences**, v. 9, n. 20, p. 4367, 2019. <https://doi.org/10.3390/app9204367>

CHUQUIMARCA, CET; MAITA, SS. Análisis comparativo entre arquitecturas de sistemas IoT. **Revista de Investigación en Tecnologías de la Información: RITI**, v. 10, n. 21, p. 55-70, 2022.

COLOMBO, JF; LUCCA FILHO, J. Internet das coisas (IoT) e indústria 4.0: revolucionando o mundo dos negócios. **Revista Interface Tecnológica**, v. 15, n. 2, p. 72-85, 2018.

CORADINE, LC; LOPES, RVV; MACIEL, AF. Mineração de dados: Uma introdução. **Journal of the Brazilian Neural Network Society**, v. 9, n. 3, p. 168-184, 2011.

CORCOVIA, LO; ALVES, RS. Aprendizagem de máquina e mineração de dados: avaliação de métodos de aprendizagem. **Revista Interface Tecnológica**, v. 16, n. 1, p. 90-101, 2019.

COSTA FILHO, RV. **Smart-GISSA**: um sistema para governança em saúde digital baseado em aprendizado de máquina. Tese (Doutorado em Engenharia de Teleinformática) – Programa de Pós-Graduação em Engenharia de Teleinformática. Universidade Federal do Ceará. Fortaleza (CE), 2021.

DIAS, DR; SANTOS, ELA; DINIZ, LM. A contribuição do big data, gerenciamento de risco e gestão estratégica com ênfase em auditoria. **Revista Ibero-Americana de Humanidades, Ciências e Educação**, v. 8, n. 10, p. 3968-3988, 2022.

DONG, M; LI, W; NASSIF, AB. Long-term health index prediction for power asset classes based on sequence learning. **IEEE transactions on power delivery**, v. 37, n. 1, p. 197-207, 2021.

DUARTE, LC; NEIRA, MG. Educação Física na educação infantil: um balanço das dissertações e teses da última década. **Revista Didática Sistemática**, v. 23, n. 1, p. 16–34, 2022.

EMEK, ML; ADILBEKOVA, K; KAYA, MG. Sustainable 6g technologies developing energy-efficient networks for global connectivity supporting IOT ai in future digital ecosystems. **AI-6Gscape**, v. 1, n. 1, p. 15-29, 2025.

EVERTON, ANF *et al.* Agrupamento de variáveis de treinamento de ultramaratonistas. **Revista Presença**, v. 10, n. 22, p. 4-20, 2024.

FÉLIX, BM; TAVARES, E; CAVALCANTE, NWF. Fatores críticos de sucesso para adoção de Big Data no varejo virtual: estudo de caso do Magazine Luiza. **Revista Brasileira de Gestão de Negócios**, v. 20, p. 112-126, 2018.

FERLIN, EP; REZENDE, DA. Big Data aplicado à cidade digital estratégica: estudo sobre o volume de dados das aplicações Smart City. **Revista Gestão & Tecnologia**, v. 19, n. 2, p. 175-194, 2019.

FERNANDES, FT. **Machine learning em saúde e segurança do trabalhador**: perspectivas, desafios e aplicações. Tese (Doutorado em Saúde Pública) – Programa de Pós-graduação em Saúde Pública. Faculdade de Saúde Pública. Universidade de São Paulo. São Paulo, 2021.

FERNANDES, FT; CHIAVEGATTO FILHO, ADP. Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. **Revista Brasileira de Saúde Ocupacional**, v. 44, p. e13, 2019.

FERRAG, MA *et al.* Edge learning for 6G-enabled Internet of Things: A comprehensive survey of vulnerabilities, datasets, and defenses. **IEEE Communications Surveys & Tutorials**, v. 25, n. 4, p. 2654-2713, 2023.

FONSECA, SO; NAMEN, AA. Mineração em bases de dados do Inep: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133-157, 2016.

GALVÃO, ND; MARIN, HF. Técnica de mineração de dados: uma revisão da literatura. **Acta Paulista de Enfermagem**, v. 22, n. 5, p. 686-690, 2009.

GARCIA, LM *et al.* Avaliação de desempenho de algoritmos de mineração de dados e simulação de Monte Carlo na descoberta de tendências no *app hambre delivery*. **RECIMA21 - Revista Científica Multidisciplinar-ISSN 2675-6218**, v. 6, n. 8, e686669, 2025.

GOMES, AK. **Representação, extração e avaliação de interações entre usuários de redes sociais online**. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo. São Carlos (SP), 2013.

GUO, F *et al.* Enabling massive IoT toward 6G: A comprehensive survey. **IEEE Internet of Things Journal**, v. 8, n. 15, p. 11891-11915, 2021.

IKEDA, T *et al.* 2025 Japanese Heart Rhythm Society/Japanese Circulation Society Consensus Statement on the Appropriate Use of Ambulatory and Wearable Electrocardiographs. **Circulation Journal**, v. 89, n. 6, p. 850-876, 2025.

INTERNATIONAL DATA CORPORATION. **Data Age 2025**: the digitization of the world from edge to core. Framingham (USA): IDC/Seagate, 2018

JOAQUIM, JLM. **Arquitetura para data lakes adequada à privacidade de dados no contexto da GDPR**. Dissertação (Mestrado em Ciência da Computação) - Programa de Pós-Graduação em Ciência da Computação. Universidade Federal de Santa Catarina. Florianópolis (SC), 2022.

JUNIOR, HSN *et al.* Atendimento em saúde à luz da teoria das filas. **Revista Presença**, v. 11, n. 25, p. 113-126, 2025.

JUNIOR, HSN. Modelagem AR(1) para descrição e predição de receitas de uma pequena confecção. **Revista Presença**, v. 9, n. 21, p. 46-59, 2023.

JUNIOR, HSN; BRASIL, RM; BARRETO, ACLG. Regras de associação para vendas de modalidades de exercícios em academia de ginástica. **Revista Presença**, v. 10, n. 22, p. 354-370, 2024.

KASHYAP, G. Unsupervised learning for high-dimensional data: Advancements in unsupervised learning techniques like clustering, anomaly detection, and dimensionality reduction. **International Journal Leading Research Publication**, v. 5, n. 12, p. 1-10, 2024.

KUMAR, V; STEINBACH, M; TAN, PN. **Introdução ao data mining**. Mineração de dados. Rio de Janeiro: Ciência Moderna, 2009.

LEANDRO, GCW *et al.* Avaliação do sistema de vigilância de infecções respiratórias agudas e fatores associados: estudo transversal, Brasil, 2009-2021. **Epidemiologia e Serviços de Saúde**, v. 34, e20240555, 2025.

LIBRELOTTO, SR; MOZZAQUATRO, PM. Análise dos algoritmos de mineração J48 e Apriori aplicados na detecção de indicadores da qualidade de vida e saúde. **Revista Interdisciplinar de Ensino, Pesquisa e Extensão - RevInt**, v. 1, n. 1, p. 26-37, 2014.

LICKS, RAS. **Big Data**: diretrizes e técnicas para preservação da privacidade. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação) - Universidade Católica de Brasília. Brasília (DF), 2016.

LIMA, BMC *et al.* Perfil das Internações e Óbitos por COVID-19 no Nordeste do Brasil em 2020. **Research, Society and Development**, v. 11, n. 13, e474111334604, 2022.

LIMA, JES *et al.* Sistema industrial IOT (internet das coisas) e sua integração sociopolítica. **Brasil Para Todos-Revista Internacional**, v. 7, n. 1, p. 109-116, 2019.

LIMA, JMS. **Estudo e desenho de abordagem auto-organizada para sistemas de controle distribuídos**. Dissertação (Mestrado em Engenharia Eletrotécnica e de Computadores) –

Departamento de Engenharia Eletrotécnica e de Computadores. Universidade Nova de Lisboa. Lisboa (Portugal), 2021.

LIMA, MNCA; FAGUNDES, RAA. Educational datamining: a study of the factors that cause school dropout in higher education institutions in brazil. **Renote**, v. 18, n. 1, 2020. DOI: 10.22456/1679-1916.105950.

LIMA, MNS; SILVA, LS; NASCIMENTO, FA. Computação quântica: desafios e suas potenciais aplicações práticas. **Revista Contemporânea**, v. 5, n. 9, e9023, 2025.

LIRA, KA; ALCHIERI, JC. Análise exploratória com python em base de dados de óbitos no Brasil no período entre 2019 e 2022. **RETEC-Revista de Tecnologias**, v. 17, n. 2, p. 72-88, 2024.

LOPES, B. **Pentágono da privacidade no big data privacy analytics**: Proposta de modelo multifacetado de garantia da privacidade e do valor analítico. Tese (Doutorado em Ciência da Informação) – Programa de Pós-graduação em Gestão & Organização do Conhecimento. Escola de Ciência da Informação. Universidade Federal de Minas Gerais. Belo Horizonte (MG), 2024.

LUZ, GB; CAMOSSO, G. Segmentação de clientes por meio de clusterização: uma análise com o algoritmo k-means aplicado a atrasos e saldos vencidos. **REGRAD-Revista Eletrônica de Graduação do UNIVEM-ISSN 1984-7866**, v. 18, n. 1, p. 16-28, 2025.

MACIEL, JVP *et al.* Transformação digital em pequenos negócios: contribuições dos sistemas de recomendação para visibilidade e competitividade. **Cadernos Cajuína**, v. 10, n. 6, e1537, 2025.

MAGRANI, E. **Entre dados e robôs**. Ética e privacidade na era da hiperconectividade. Porto Alegre (RS): Arquipélago Editorial, 2019.

MAHDI, MJ; ALJUBOORI, AF; HUSSEIN, AM. Smart stadium using cloud computing and Internet of Things (IoT): Existing and new models. **International Journal of Computer Applications Technology and Research**, v. 10, n. 5, p. 111-118, 2021.

MAIA, ERA. **Os conhecimentos em informática como mediadores do processo de ensino e aprendizagem no ensino médio nos últimos vinte anos**: uma revisão integrativa. Dissertação (Mestrado em Educação Tecnológica e Profissional) - Programa de Pós-graduação em Educação Profissional e Tecnológica. Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano. Salgueiro (PE), 2024.

MALTAURO, TC; GUEDES, LPC; URIBE-OPAZO, MA. Optimization algorithms for multivariate sampling reduction using spatial-temporal data. **Revista Brasileira de Ciência do Solo**, v. 49, p. e0240097, 2025.

MĂNESCU, DC. Big data analytics framework for decision-making in sports performance optimization. **Data**, v. 10, n. 7, p. 116, 2025. <https://doi.org/10.3390/data10070116>

MARMITT, N *et al.* Mineração de dados em saúde: novas possibilidades agregando cuidado no cenário oncohematológico. **Hematology, Transfusion and Cell Therapy**, v. 46, p. S1255, 2024.

MARQUES, TSS. **A reconfiguração do trabalho docente frente à cultura digital**: uma análise de escolas públicas municipais de ensino fundamental. Dissertação (Mestrado em Educação e Docência) – Programa de Pós-graduação em Educação e Docência. Faculdade de Educação. Universidade Federal de Minas Gerais. Belo Horizonte (MG), 2025.

MARTINS, DWP. **Inteligência artificial aplicada em séries temporais**. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística. Universidade de São Paulo. São Paulo, 2024.

MARTINS, GD; SILVA, D. Comparação de algoritmos de aprendizado de máquina na previsão do *churn* em uma empresa de tecnologia na área de educação continuada a distância. **Inteligencia Artificial**, v. 20, n. 59, p. 123-137, 2020.

MARTINS, JSB *et al.* Cidade inteligente na contemporaneidade com internet das coisas e inteligência artificial. **Gestão & Planejamento**, v. 25, n. 1, p. 277-296, 2024.

MENDIETA, TP; HERRERA, J; PEÑA, AJ. La Capacidad del IOT de Transformar el Futuro. **Revista Avenir**, v. 3, n. 1, p. 15-18, 2019.

MENDONÇA, HA. **A (in) visibilidade nos processos de colonialismo e decolonialidade digital na educação online e remota no Ensino Superior**. Tese (Doutorado em Letras) – Programa de Pós-graduação em Estudos Linguísticos e Literários em Língua Inglesa. Faculdade de Filosofia, Letras e Ciências Humanas. Universidade de São Paulo. São Paulo, 2024.

MEYER, PL. **Probabilidade**. Rio de Janeiro: LTC, 1987.

MINISTÉRIO DA SAÚDE. SECRETARIA DE VIGILÂNCIA EM SAÚDE. DEPARTAMENTO DE ANÁLISE EM SAÚDE E VIGILÂNCIA DE DOENÇAS NÃO TRANSMISSÍVEIS. **Vigitel Brasil 2019: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico [Internet]**. Brasília (DF): Ministério da Saúde; 2020.

MISRA, NN *et al.* IoT, big data, and artificial intelligence in agriculture and food industry. **IEEE Internet of things Journal**, v. 9, n. 9, p. 6305-6324, 2020.

MOHAMMADI, M *et al.* Deep learning for IoT big data and streaming analytics: A survey. **IEEE Communications Surveys & Tutorials**, v. 20, n. 4, p. 2923-2960, 2018.

MONLEÓN-GETINO, A. El impacto del Big-data en la Sociedad de la Información. Significado y utilidad. **Historia y comunicación social**, v. 20, n. 2, p. 427-445, 2015.

MONTEIRO, PHA; ARAUJO, MF. **Caracterização de projetos na gestão de recursos humanos de uma empresa farmacêutica utilizando a mineração de dados**. Monografia (Graduação em Engenharia de Produção) - Instituto de Ciências Exatas e Aplicadas. Universidade Federal de Ouro Preto. João Monlevade (MG), 2023.

MOREIRA, FM *et al.* Metadados para descrição de datasets e recursos informacionais do “Portal Brasileiro de Dados Abertos”. **Perspectivas em Ciência da Informação**, v. 22, n. 03, p. 158-185, 2017.

MORELLO, R *et al.* An IoT based ECG system to diagnose cardiac pathologies for healthcare applications in smart cities. **Measurement**, v. 190, p. 110685, 2022.

NASCIMENTO, OAS *et al.* Cursos de Educação Física no Brasil: consolidação de dados de 1995 a 2020. **Avaliação**, v. 29, e024004, 2024.

NGUYEN, DC *et al.* 6G Internet of Things: A comprehensive survey. **IEEE Internet of Things Journal**, v. 9, n. 1, p. 359-383, 2021.

NIU, G *et al.* Application of AHP and EIE in reliability analysis of complex production lines systems. **Mathematical Problems in Engineering**, v. 2019, n. 1, p. 7238785, 2019.

OKPU, EO; TAYLOR, OE. Analysing the Integration of AES-256 Encryption and HMAC Hashing in IoT Smart Healthcare Systems. **Ci-STEM Journal of Digital Technologies and Expert Systems**, v. 2, n. 1, p. 18-24, 2025.

OLIVEIRA, DT; SILVA, PN. Metaopengov: modelo de metadados para o portal brasileiro de dados abertos. **Encontros Bibli**, v. 30, e105323, 2025.

OLIVEIRA, DT; SILVA, PN; PEREIRA, FCM. Inteligência artificial para recuperação de dados abertos reflexões e proposições para a experiência do usuário no Portal Brasileiro de Dados Abertos. **Brazilian Journal of Information Science**, v. 18, e024016, 2024. DOI: 10.36311/1981-1640.2024.v18.e024016.

PACHECO, BBM; DISCONZI, MS. Ciência de dados: enfoque no desafio do processamento. **Research, Society and Development**, v. 8, n. 11, e128111444, 2019.

PAIVA, LFR. **A prática docente e as mediações didáticas na educação superior pela incorporação das TDIC: um estudo de caso**. Tese (Doutorado em Educação) - Programa de Pós-graduação em Educação. Universidade de Uberaba. Uberaba (MG), 2020.

PARRA-SÁNCHEZ, DT. Exploring the Internet of Things adoption in the Fourth Industrial Revolution: a comprehensive scientometric analysis. **Journal of Innovative Digital Transformation**, v. 2, n. 1, p. 1-18, 2025.

PENNONE, J. **Comparação de modelos estatísticos para prever o risco de lesões de membros inferiores em jogadores de futebol**. Tese (Doutorado em Ciências) – Faculdade de Medicina. Universidade de São Paulo. São Paulo, 2025.

PEREIRA, PEA. **Big data analytics e data mining para avaliação e previsão da eficiência financeira do setor bancário brasileiro: um estudo de 2011 a 2022**. 2023. Dissertação (Mestrado em Administração) – Programa de Pós-graduação em Administração. Faculdade de Economia, Administração, Contabilidade e Gestão de Políticas Públicas. Universidade de Brasília. Brasília (DF), 2023.

PEREIRA, WR. **Educação 4.0: os desafios na utilização das metodologias ativas e inserção das tecnologias digitais de informação e comunicação no ensino superior**. Dissertação (Mestrado em Educação) – Programa de Pós-graduação *Stricto Sensu* em Educação. Universidade Católica de Brasília. Brasília (DF), 2022.

PERUGINI, L; VESCO, A. On the integration of Self-Sovereign Identity with TLS 1.3 handshake to build trust in IoT systems. **Internet of Things**, v. 25, p. 101103, 2024.

PEZZINI, A. Mineração de textos: conceito, processo e aplicações. **Revista Brasileira De Contabilidade E Gestão**, v. 5, n. 10, p. 58-61, 2017.

PICCOLO, FRR. Do Big Data ao conhecimento organizacional: Limites e possibilidades da inteligência artificial na governança corporativa. **Research, Society and Development**, v. 14, n. 9, e6714949560, 2025.

PINTO, F. Trilha histórica sobre prática da publicação de dados de pesquisa. **Perspectivas em Ciência da Informação**, v. 28, p. e-45978, 2023.

RANGEL, NF *et al.* Análise espacial no Brasil: Impactos econômicos das políticas de desinformação no sus (2020 a 2021). **Semestre econômico**, v. 28, n. 65, p. 1-27, 2025.

RAUTENBERG, S; CARMO, PRV. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science**, v. 13, n. 1, p. 56-67, 2019.

RAZ, A *et al.* Prediction and explainability in AI: Striking a new balance? **Big Data & Society**, v. 11, n. 1, p. 20539517241235871, 2024.

REIS-SANTOS, B. Sistemas de Informação em Saúde: o quanto estamos avançando? **Epidemiologia e Serviços de Saúde**, v. 32, n. 2, p. e2022433, 2023.

RIGO, SJ *et al.* Aplicações de mineração de dados educacionais e learning analytics com foco na evasão escolar: oportunidades e desafios. **Revista Brasileira de Informática na Educação**, v. 22, n. 01, p. 132-146, 2014.

ROCHA, AS *et al.* Differences in risk factors for incident and recurrent preterm birth: a population-based linkage of 3.5 million births from the CIDACS birth cohort. **BMC medicine**, v. 20, n. 1, p. 111, 2022. <https://doi.org/10.1186/s12916-022-02313-4>.

- ROCHA, ES *et al.* Algoritmos de aprendizado de máquina no diagnóstico de doenças cardiovasculares. **PET Matemática**, p. 37-43, 2024.
- RODRIGUES, AA; NÓBREGA, E; DIAS, GA. Desafios da gestão de dados na era do big data: perspectivas profissionais. **Informação & Tecnologia**, v. 4, n. 2, p. 63-79, 2017.
- RODRIGUES, CR; SOTTO, ECS. Big data e business intelligence: suas diferenças e importância para as organizações. **Revista Interface Tecnológica**, v. 19, n. 2, p. 43-54, 2022.
- ROSINI FILHO, AM; ROSINI, AM; PALMISANO, A. A era do big data: principais implicações sobre segurança e privacidade e as novas tecnologias capazes de auxiliar processos investigativos e detecção de fraudes em tempo real. **Journal on Innovation and Sustainability RISUS**, v. 11, n. 3, p. 13-34, 2020.
- ROSS, S. **Probabilidade**: um curso moderno com aplicações. Porto Alegre (RS): Bookman, 2010.
- RUFINO, HLP; VEIGA, ACP; NAKAMOTO, PT. Smote_easy: um algoritmo para tratar o problema de classificação em bases de dados reais. **JISTEM - Journal of Information Systems and Technology Management**, v. 13, n. 1, p. 61-80, 2016.
- SALDANHA, RF; BARCELLOS, C; PEDROSO, MM. Ciência de dados e big data: o que isso significa para estudos populacionais e da saúde? **Cadernos Saúde Coletiva**, v. 29, n. esp., p. 51-58, 2021.
- SANTAELLA, L; KAUFMAN, D. Os dados estão nos engolindo? **Civitas-Revista de Ciências Sociais**, v. 21, n. 2, p. 214-223, 2021.
- SANTILIO, FP *et al.* Análise do consumo de energia elétrica dos equipamentos IoT utilizados na automação residencial: um estudo de caso para redução do consumo de energia elétrica. **RECIMA21 - Revista Científica Multidisciplinar-ISSN 2675-6218**, v. 4, n. 10, e4104172, 2023.
- SANTOS, AHC *et al.* A dicotomia sexual como marcadora do gerenciamento de academia de ginástica. **Revista Presença**, v. 11, n. 26, p. 78-112, 2025.
- SANTOS, DO; FREITAS, EB. A internet das coisas e o big data inovando os negócios. **Refas-Revista Fatec Zona Sul**, v. 3, n. 1, p. 1-18, 2016.
- SANTOS, GO; FARIA, RC; VASCONCELOS, RO. O impacto das características da rede no desempenho de protocolos IoT. **Revista de Gestão e Secretariado**, v. 15, n. 5, e3773, 2024.
- SANTOS, W *et al.* Inteligência artificial na saúde: otimizando a análise de imagens médicas para diagnósticos mais precisos e humanizados. **ARACÊ**, v. 6, n. 4, p. 18671-18687, 2024.
- SANTOS, WL. **Itinerâncias docentes em cenários virtuais de aprendizagem**: práticas e experiências com dispositivos móveis. Tese (Doutorado em Educação) – Programa de Pós-graduação em Educação. Universidade Federal de Sergipe. São Cristóvão (SE), 2025.
- SCAIONI, AG. **Policiamento inteligente**: aplicação de ontologia, análise preditiva e redes neurais convolucionais no apoio ao planejamento operacional e roteirização. Dissertação (Mestrado em Ciências da Computação) – Faculdade de Ciências e Tecnologia. Universidade Estadual Paulista. Presidente Prudente (SP), 2025.
- SEMAAN, GS *et al.* Proposta de um método de classificação baseado em densidade para a determinação do número ideal de grupos em problemas de clusterização. **Journal of the Brazilian Computational Intelligence Society**, v. 10, n. 4, p. 242-262, 2012.
- SILVA, AAP; MONTEIRO, DAA; REIS, AO. Qualidade da Informação dos dados governamentais abertos: análise do portal de dados abertos brasileiro. **Revista Gestão em Análise**, v. 9, n. 1, p. 31-47, 2020.

SILVA, DBY. **Aplicação de privacidade diferencial na detecção de ataques em internet das coisas**. Trabalho de Conclusão de Curso (Tecnólogo em Redes de Computadores) - Universidade Federal do Ceará. Quixadá (CE), 2025.

SILVA, DM; BRITO, JAM; OLIVEIRA, CS. Um Estudo Computacional Comparativo entre Algoritmos de Agrupamento e de Detecção de Comunidades. **SPOLM**, v. 3, n. 1, p. 2808-2822, 2019.

SILVA, KR; OLIVEIRA, PCT. ORGANIZANDO DADOS ABERTOS: um estudo sobre métodos que aprimorem a interface do Portal Brasileiro de Dados Abertos acessibilizando a informação pública. **Revista Ágora: políticas públicas, comunicação e governança informacional**, v. 2, n. especial, p. 96-110, 2017.

SILVA, LA; PERES, SM; BOSCARIOLI, C. **Introdução à mineração de dados** - com aplicações em R. Rio de Janeiro: GEN LTC, 2016.

SILVA, PB *et al.* Identificação de outliers em dados de acidentes de trânsito no brasil: análise de cluster versus métodos estatísticos. **Revista de Engenharia e Tecnologia**, v. 11, n. 1, p. 103-114, 2019.

SILVA, RO; OLIVEIRA, JLS. A internet das coisas (IOT) com enfoque na saúde. **Tecnologias Em Projeção**, v. 8, n. 1, p. 77-85, 2017.

SILVA, SP *et al.* Indicadores para avaliação qualitativa de Dados Abertos: Inteligibilidade, operacionalidade e interatividade nos datasets do Governo Federal no Portal Brasileiro de Dados Abertos. **Informação & Sociedade: Estudos**, v. 30, n. 3, p. 1-19, 2020.

SILVA, VJ; BONACELLI, MBM; PACHECO, CA. O sistema tecnológico digital: inteligência artificial, computação em nuvem e Big Data. **Revista Brasileira de Inovação**, v. 19, p. e0200024, 2020.

SINCLAIR, B. **IoT: como usar a "Internet das Coisas" para alavancar seus negócios**. Belo Horizonte (MG): Autêntica Business, 2018.

SOUZA, VF; SANTOS, TCB. Processo de mineração de dados educacionais aplicado na previsão do desempenho de alunos: Uma comparação entre as técnicas de aprendizagem de máquina e aprendizagem profunda. **Revista Brasileira de Informática na Educação**, v. 29, p. 519-546, 2021.

STOPA, SR *et al.* Pesquisa Nacional de Saúde 2019: histórico, métodos e perspectivas. **Epidemiologia e Serviços de Saúde**, v. 29, n. 5, e2020315, 2020.

TANG, R; FONG, S. Clustering big IoT data by metaheuristic optimized mini-batch and parallel partition-based DGC in Hadoop. **Future Generation Computer Systems**, v. 86, p. 1395-1412, 2018.

TRAN, SN; HOANG, VT; BUI, DH. A hardware architecture of NIST lightweight cryptography applied in IPsec to secure high-throughput low-latency IoT networks. **IEEE Access**, v. 11, p. 89240-89248, 2023.

VAN HECK, S; VALKS, B; DEN HEIJER, A. The added value of smart stadiums: a case study at Johan Crujff Arena. **Journal of Corporate Real Estate**, v. 23, n. 2, p. 130-148, 2021.

VARGAS, V *et al.* Variantes do Índice Silhueta para Validação de Agrupamentos. **Cadernos do IME-Série Informática**, v. 46, p. 118-127, 2021.

VIEIRA JUNIOR, V. **Classificação de dados usando técnicas de datamining e aprendizado de máquina**. Monografia (Bacharelado em Ciência da Computação) – Departamento de Informática. Universidade Federal do Maranhão. São Luís (MA), 2013.

VIEIRA, AR *et al.* Classificação de indivíduos ansiosos por árvore de decisão. **Revista Presença**, v. 10, n. 22, p. 64-75, 2024.