

MODELOS SEQUENCIAIS PARA REDUÇÃO DE VIÉS EM APRENDIZAGEM DE MÁQUINA

BRASIL, Roxana Macedo¹; BARRETO, Ana Cristina Lopes y Glória²; BRITO, Diogo de Freitas^{2:3:4}; EVERTON, Adriana Nunes da Fonseca⁵; CARVALHO JUNIOR, Sergio⁶; JUNIOR, Homero da Silva Nahum^{1:7}

438

Resumo

Norteados pelo objetivo de discutir modelos de *boosting*, apresentou-se a conceituação e o objetivo deles, comparando-os ao *bagging* e *stacking*. A fundamentação da Matemática fez-se necessária à compreensão do funcionamento dos pesos e da obtenção do modelo final. Posteriormente, discutiu-se as principais famílias de algoritmos, apresentando as adequações. A exposição dos conjuntos de vantagens e desvantagens ou limitações foram imperativos para ilustrar robustez, estabilidade e precisão das possibilidades de aplicações. Então, possível foi concluir que no domínio da aprendizagem de máquina, os modelos sequenciais seriam recursos significativos, em pese a necessidade de controlar a possibilidade de ajuste excessivo.

Palavras-chave: Inteligência artificial. Modelagem. Ciências de dados. Predição. Aprendizado de máquina.

Abstract

Guided by the objective of discussing boosting models, the conceptualization and purpose of these models were presented, comparing them to bagging and stacking. A mathematical foundation was necessary to understand the functioning of the weights and the obtaining of the final model. Subsequently, the main families of algorithms were discussed, presenting their suitability. The exposition of the sets of advantages and disadvantages or limitations was imperative to illustrate the robustness, stability, and precision of the application possibilities. Therefore, it was possible to conclude that in the domain of machine learning, sequential models would be significant resources, despite the need to control the possibility of overfitting.

Keywords: Artificial intelligence. Modelling. Data science. Prediction. Machine learning.

¹ Docente Ph.D. em Educação Física;

² Docentes do Curso de Educação Física do Centro Universitário Celso Lisboa;

³ Docente do Curso de Gestão Desportiva e do Lazer do Centro Universitário Celso Lisboa;

⁴ Consultor Iceberg Business Academy;

⁵ Profissional de Educação Física da Adriana Nunes Consultoria de Corrida;

⁶ Pesquisador convidado Biodesa;

⁷ Docente da Escola de Saúde da Universidade Cândido Mendes.

Introdução

Em aprendizagem de máquina possível seria combinar predições de diversos modelos isolados, originando um modelo final dotado de maior robustez (manutenção do desempenho na presença de perturbações, ruídos, entradas inválidas ou dados incompletos), estabilidade (referência à consistência e repetibilidade, dados parecidos deveriam ter resultados parecidos) e precisão (acurácia, quanto que predições indicariam os valores reais) no concernente à capacidade preditiva. Essa coadunação seria denominada aprendizado em conjunto ou método de *ensemble*, e alicerçada estaria na ideia de superação de desempenho de único modelo complexo (ou modelos fracos isolados) pelo grupo de modelos especializados, os quais individualmente poderiam apresentar desempenhos não satisfatórios, na reunião haveria a expectativa de redução da variância ou viés (Moreira *et al.*, 2021; Souza e Junior, 2023; Bezerra, 2024; Dâmaso *et al.*, 2025; Montenegro, 2025). Pelo exposto, o objetivo seria melhoria de desempenho preditivo, mesmo com elevado custo computacional.

As categorias principais de aprendizado conjunto seriam *bagging*, *boosting* e *stacking*. O *bagging* significaria *bootstrap aggregating* (Amostragem e Agregação, em tradução livre), cujos objetivos seriam a redução da variância e evitação do ajuste excessivo (*overfitting*). Para tanto, diversos modelos seriam treinados independente e paralelamente em subconjuntos distintos de dados. Esses obtidos por amostragem aleatória com reposição (*bootstrapping*). A agregação se apresentaria na obtenção da predição final, a qual poderia ser por votação majoritária ou média, quando empregada classificação ou regressão, respectivamente (Camargo, 2021; Schmoeller, Peron e Mesquita Filho, 2025; Villca, 2025; Centeno, 2025; Reis e Ruppert, 2026). Esse seria o caso clássico da Floresta Aleatória (*Random Forest*), que combinaria múltiplas árvores de decisão (Narasimhan e Victor, 2025; Yu *et al.*, 2025; Guzmán, Vivas e Morales, 2026).

O *boosting* (Impulso) objetivaria reduzir viés e variância, potencializando acurácia. Esses algoritmos transformariam determinado grupo de modelos de baixo desempenho (aprendizes fracos - *weak learners*) em um e somente um modelo de elevado desempenho (aprendiz forte - *strong learner*). Os resultados insatisfatórios daqueles modelos se substancializariam em predições errôneas ou ineficientes, o que indicaria falha na generalização do conhecimento adquirido na fase de treino, culminando em deficiência de precisão, confiabilidade ou eficiência. Materialmente, isso ocorreria em Taxas de Acerto \approx 50,00% para classificação ou predições \approx média aleatória ou constante para regressão

(Arruda *et al.*, 2025; Lopez, Lopez e Aguaded, 2025). Característica comum em árvores de decisão de baixa profundidade, $nós \leq 8$ (Veloso *et al.*, 2021; Távora, 2025). O funcionamento consistiria em treinar os modelos sequencialmente, buscando a correção dos erros (resíduos) do modelo corrente pelo posterior. O mecanismo empregado seria aumentar o peso das entradas classificadas incorretamente, o que elevaria a importância delas no modelo subsequente. A predição final seria a média ponderada de todos os modelos (Damke, Damke e Gregorini, 2025; Tertulino e Duarte, 2025; Pereira, 2025).

Na categoria *stacking* ou *stacked generalization* (Empilhamento ou Generalização Empilhada, em tradução livre) residiriam os algoritmos combinadores de distintos modelos, buscando o melhor desempenho, portanto o foco estaria em aprender como realizar a combinação, o que a distinguiria das anteriores. O mecanismo consistiria em treinar modelos de base, os quais, comumente, seriam algoritmos diferentes, por exemplo: floresta aleatória, rede neural, máquina de vetores de suporte e k-vizinhos mais próximos. As predições advindas seriam utilizadas para treinar o meta-classificador (metamodelo), o qual identificaria o modelo de base mais adequado (ou confiável) para cada parte dos dados. A agregação dessas predições resultaria na predição final, a qual tenderia a superar os resultados individualizados dos modelos iniciais (Ortiz *et al.*, 2024; Celeri, 2025; Silva e Ferreira, 2025; Pandolfo, 2025; Páez e Ramírez, 2026). Em síntese, possível seria a realização de analogias às categorias: *bagging* \approx reunião de condomínio, *boosting* \approx trabalho em equipe, e *stacking* \approx alocação de especialistas (Quadro I). O estudo corrente objetivou, então, discutir os modelos de *boosting*, fornecendo exemplos de aplicação.

Quadro I: Resumo e Analogia das Categorias de Aprendizado Conjunto.

Categoria	Objetivo	Treinamento	Combinação	Analogia
<i>Bagging</i>	↓variância	Paralelo	Votação ou média	Moradores independentes opinam e estima-se a média ou votos.
<i>Boosting</i>	↓viés e variância	Sequencial	Média	Cada membro da equipe aprende com o erro do anterior.
<i>Stacking</i>	↑Precisão	Paralelo	Metamodelo	Gerente aloca cada especialista na tarefa que melhor resolve.

Fonte: Os Autores (2026).

Considerando o conjunto de treino $\{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$, no qual $\forall i = 1, 2, \dots, n \mid y_i \in \{-1; 1\}$, os pesos iniciais foram estabelecidos como $w_i^{(1)} = \frac{1}{n}, \forall i = 1, 2, \dots, n$. Então, para cada iteração $t = 1, 2, \dots, T$, necessárias seriam as providências (Mantilla, 2004; Chaves, 2012; Pereira, 2024; Damke, Damke e Gregorini, 2025; Lopez, Lopez e Aguaded, 2025): 1) treinar o modelo fraco; 2) calcular o peso do modelo (classificador); e 3) atualizar os pesos dos dados. No treinamento de cada modelo, o objetivo residiria em minimizar o erro, portanto treinar-se-ia $h_t: X \rightarrow \{-1; 1\}$ utilizando os pesos $w^{(t)}$, o que resultaria em (I), tendo como função indicadora $f(y_{i \neq} h_t(x_i))$, a qual retornará 1,00 se o modelo acertar ou 0,00 se errar.

$$\varepsilon_t = \frac{\sum_{i=1}^n w_i^{(t)} \cdot f(y_{i \neq} h_t(x_i))}{\sum_{i=1}^n w_i^{(t)}} \quad (I)$$

O peso de cada classificador no modelo final guardará relação inversa com a respectiva precisão, logo quanto menor o erro, maior será o α_t (II), penalizando-se o acerto aleatório, então $\varepsilon_t = 0,50 \rightarrow \alpha_t = 0,00$ e $\varepsilon_t = 0,00 \rightarrow \alpha_t = \infty$. A atualização dos pesos consistiria em aumentá-los quando a classificação se mostrasse incorreta ($h_t(x_i) \neq y_i$) ou diminuí-los, caso contrário (III), o que exigiria a normalização para garantir que o somatório deles se igualasse à unidade (IV). O modelo final resultaria da função sinal da soma ponderada (*signum*, denotada por $\text{sgn}(x)$, $\text{sing}(x)$ ou *sinal*) dos modelos fracos (V). A utilização dessa função se justificaria por ela mapear o resultado contínuo (soma ponderada), tornando-o discreto (1, 0 ou -1), eliminando a magnitude, porém mantendo a direção, sim ou não, certo ou errado (Batista, 2024; Arruda *et al.*, 2025; Tertulino e Duarte, 2025; Pereira, 2025).

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_t}{\varepsilon_t} \right) \quad (II)$$

$$w_i^{(t+1)} = w_i^t \cdot \exp(-\alpha_t \cdot y_i \cdot h_t(x_i)) \quad (III)$$

$$w_i^{(t+1)} = \frac{w_i^{(t+1)}}{\sum_j w_j^{(t+1)}} \quad (IV)$$

$$H(x) = \text{sinal} \left(\sum_{t=1}^T \alpha_t \cdot h_t(x) \right) \quad (V)$$

Tipos de Algoritmos de *Boosting*

Originalmente, o *Adaptive Boosting*, primeiro algoritmo, realizava a reponderação de dados por métodos simples, o que foi superado pelo emprego de funções de perdas diferenciáveis e regularização (reduziria o ajuste excessivo), elevando velocidade, eficiência e precisão, sobretudo em grandes bancos de dados, variáveis categóricas e ausência de dados (Mayr *et al.*, 2014; Yu *et al.*, 2025; Nascimento Júnior, 2026).

O *Adaptive Boosting* (*AdaBoost* – Impulso Adaptativo ou Reforço Adaptativo), conforme a descrição matemática, realizaria a classificação binária, ajustando os pesos dos dados, priorizando as dificuldades (erros) a cada modelo subsequente. Conseqüentemente, dados ruidosos ou discrepantes poderiam dominar o processo, tornando o algoritmo suscetível a valores extremos, dados os maiores pesos atribuídos (Al-Stouhi e Reddy, 2011; Zheng e Yang, 2022).

O *Gradient Boosting Machine* (GBM – Máquina de Aumento de Gradiente ou Máquina de Impulso de Gradiente, em tradução livre) ou *Gradient Boosting* modelaria o problema como otimização numérica, logo a cada iteração (t), novo modelo (h_t) seria treinado para prever o resíduo da função de perda em relação ao modelo da iteração imediatamente anterior (h_{t-1}), objetivando diminuir o gradiente daquela função. Matematicamente, $F(x)$ seria a soma dos modelos anteriores, então o próximo modelo seria ajustado para se aproximar de $h_t \approx -\frac{\partial L(y, F(x))}{\partial F(x)}$. Em síntese, a evolução do GBM seria no sentido descendente do gradiente, minimizando os erros, o que se tornaria mais eficiente pela possibilidade de emprego de distintas funções de perda (Natekin e Knoll, 2013; Bentéjac, Csörgő e Martínez-Muñoz, 2021).

A melhora do GBM se concretizou no *eXtreme Gradient Boosting* (XGBoost – Extremo Aumento de Gradiente) que abarcou métodos de otimização como: 1) a regularização na função de custo para penalizar árvores complexas, reduzindo o ajuste excessivo; 2) processamento em paralelo de atributos e métodos de poda (*pruning*) para reduzir a profundidade de árvores de decisão; e 3) aproximação de segundo grau pela expansão de Taylor de segunda ordem (derivada segunda ou hessiana), o que possibilitaria a aproximação da função de perda e conseqüentemente a otimização eficiente do modelo, porque definiria a direção de crescimento da árvore, acelerando a convergência (Ogunleye e Wang, 2019; Niazkar *et al.*, 2024).

Também focado na eficiência, porém para volumosos bancos de dados, o *Light Gradient Boosting Machine* (LightGBM - Máquina Leve de Aumento de Gradiente, em

tradução livre) concentrou-se no crescimento de árvore baseado em folhas em detrimento dos níveis, resultando em treinamento mais rápido e menor perda de precisão. Para tanto, utilizaria *Gradient-based One-Side Sampling* (GOSS – Amostragem Unilateral Baseada em Gradiente, em tradução livre) e *Exclusive Feature Bundling* (EFB – Pacote de Características/Atributos/Variáveis Exclusivas, em tradução livre). O GOSS manteria todos os dados com altos erros (gradientes elevados), e amostraria aqueles com baixos resíduos, mantendo somente determinada parte, cujos pesos sofreriam efeito multiplicador para manter a distribuição de dados. O EFB agruparia variáveis que raramente assumiriam valores não nulos simultaneamente, o que reduziria a dimensionalidade, por conseguinte, a complexidade computacional sem atenuação da precisão (Wang, Li e Zhao, 2022; Lokker *et al.*, 2024).

Não raramente, variáveis categóricas requisitariam pré-processamento extenso como na aplicação de *One-Hot Encoding* (Codificação One-Hot ou Codificação Binária), o qual converteria cada categoria em variável binária (0 ou 1), exemplificando: a variável Escolaridade apresentando as categorias Superior, Médio e Fundamental, transformada seria nas variáveis correspondentes aos três níveis acadêmicos, então se originalmente, o indivíduo 10 tinha em Escolaridade = Superior, após a transformação, ele teria: Superior = 1, Médio = 0 e Fundamental = 0. Isso seria evitado pelo algoritmo *Categorical Boosting* (CatBoost – Impulsionamento Categórico, em tradução livre) empregando *Ordered Target Statistics* (Estatísticas de Alvo Ordenado), que consiste em estimar a média do valor alvo apenas com as observações anteriores do conjunto de dados ordenado aleatoriamente, durante a fase de treinamento. Isso evitaria que informações de dados de teste ou validação se tornassem conhecidas durante o treino (vazamento de dados), o que poderia ocorrer, o que poderia ocorrer, por exemplo, na normalização, padronização ou imputação de valores faltantes em todo banco de dados, comprometendo robustez e precisão (Cai, Yuan e Zhou, 2024; Fan, Gou e Weng, 2025).

Vantagens e Desvantagens

Os modelos sequenciais teriam por vantagens (Liska *et al.*, 2012; Silva, 2023; Silva e Sátiro, 2024; Baptista e Rodrigues, 2024; Schmoeller, Peron e Mesquita Filho, 2025): 1) alto desempenho preditivo, particularmente, no processamento de dados estruturados; 2) redução de viés, a combinação de diversos modelos possibilitaria representar matematicamente relações não-lineares e de elevada complexidade, as quais não seriam

alcançadas por um e somente um modelo; 3) função de perda variada, de forma ímpar no GBM, todo sortilégio de funções diferenciáveis poderia ser empregado (entropia cruzada binária ou categórica, logística binária, erro quadrático médio, erro absoluto médio, huber, quantil, Poisson, perda de articulação ou personalizada), extrapolando as aplicações em classificação não-balanceada, regressão robusta e predição de contagem; e 4) interpretação, possível seria determinar as respectivas importâncias dos atributos para eleger as variáveis que impactariam na predição.

Talvez, a principal desvantagem ou limitação residisse no custo computacional e sequencialidade, porque o *boosting* seria essencialmente sequencial, isto posto, a paralelização completa não seria alcançada, mesmo em aplicações XGBoost e LightGBM, ou seja, a iteração $t+1$ sempre dependerá da t . Com consequência, a demanda temporal de treinamento poderia inviabilizar implementações para bancos de dados extremamente grandes (Chaves, 2012; Silva R, 2023; Freitas, 2025), o que seria atenuado no emprego de modelos lineares, como regressão linear simples ou múltipla, regressão logística, e redes neurais simples como *perceptron* (Santos *et al.*, 2019; Du *et al.*, 2022 ; Silva *et al.*, 2023; Paixão, Lombardi e Barroso, 2025). Comparativamente, implementações de Floresta Aleatória, modelo não linear, conseguiria escalar linearmente com núcleos de CPU (Buschjäger e Morik, 2017; Al Qassem *et al.*, 2023)

No domínio do usuário, a dificuldade de interpretação seria o ponto sensível, dado que o modelo final seria a soma de diversos outros, comumente na ordem de grandeza de centena ou milhar, dificultando ou impossibilitando a explicação de alguma predição, mesmo com a utilização de ferramentas específicas para acompanhar a contribuição de cada variável ao modelo ou resultado final (Wojtas e Chen, 2020; Musolf *et al.*, 2022; Ponce-Bobadilla *et al.*, 2024, Viana *et al.*, 2024).

Modelos sequenciais seriam suscetíveis a decorar ruídos porventura existentes no banco de dados, como consequência da magnitude da redução de erros no treino, destacadamente, quando quantitativo de estimadores ou taxa de aprendizado for elevado, como consequência, o modelo apresentaria ajuste excessivo (Buhlmann e Hothorn, 2007; Dantas, Aranha e Silva, 2024). Esse seria evitado com o refinamento, o modelo treinado com grande banco de dados genérico seria submetido a novo treinamento, porém sobre conjunto de dados menor e mais específico, processo de ajuste fino (*tuning* ou *fine-tuning*) de *hiperparâmetros*, tais como profundidade da árvore de decisão e redução de pesos ou

coeficientes (*shrinkage*, encolhimento), por exemplo (Thomas *et al.*, 2018; Datta, Das e Kumar, 2022).

O efeito de valores extremos seria particularmente intenso em AdaBoost e GBM. O primeiro imporá pesos elevados, favorecendo a sensibilidade, o que mitigaria a robustez. Na segunda família de algoritmos, valores discrepantes exigiriam funções de perda insensíveis, como Huber, atenuando assim o crescimento de gradientes (Friedman, 2001; Ferraris e Liautaud, 2022).

A perda de desempenho ao processar dados não-estruturados, entre os quais texto puro, áudio e imagem, conquistou evidência. A possibilidade de utilizá-los como variável ou atributo não seria capaz de superar as aplicações de redes neurais profundas para os mesmos dados. Em síntese, métodos sequenciais seriam adequados a dados estruturados, configurando tal característica limitação significativa (Feliciano, 2018; Cabalheiro, 2023; Gonçalves, 2024)

Possibilidades de Aplicação

No domínio esportivo, plataformas e aplicativos de treinamento para corrida, ciclismo ou triatlo utilizariam modelos sequenciais para recomendar o tempo para percorrer um quilômetro (*pace*) e prever o desempenho em provas profissionais e amadoras. Esse seria, caracteristicamente, problema de regressão, envolvendo a distância específica, condições climáticas, tipo de piso e elevação, por exemplo. Podendo considerar, também, desempenho no treino, volume e intensidade treinados, características antropométricas e biomecânicas, volume máximo de oxigênio, frequência cardíaca e variabilidade da frequência cardíaca (Rodrigues, 2019; Borges, 2022; Barreto, 2023; Everton *et al.*, 2024; Barreto *et al.*, 2025). A utilização de XGBoost ou LightGBM, com função de perda quantílica, permitiria a predição do tempo de prova com intervalo de confiança. A adequação aos modelos de *boosting* residiria no conjunto de variáveis distintas e na existência de relações não-lineares, particularmente, no condizente à evolução do desempenho, esta seria a combinação de adaptações mecânicas, fisiológicas, psicológicas e nutricionais induzidas pelo treinamento, o que originaria o efeito platô, aspecto adequadamente capturado pelos modelos propostos, assim como o risco de treinamento excessivo (*overtraining*).

As estimativas de desempenho necessitariam da companhia da probabilidade de lesão por sobrecarga, possibilitando intervenções preventivas sobre fratura por estresse, por exemplo (Siciliano *et al.*, 2023; Carvalho Junior *et al.*, 2025). Para tanto, as variáveis

independentes poderiam ser, dentre outras (Pastre *et al.*, 2007; Hespanhol Junior *et al.*, 2012; Meneghetti *et al.*, 2021; Tiggemann, Gossmann e Cremonese, 2022; Correia *et al.*, 2025; Pinelli *et al.*, 2025; Qin, Li e Chen, 2025): histórico de lesões, regime de sono, calçado, fase do treinamento e carga aguda-crônica, essa a razão entre a carga de trabalho recente e a média da carga de trabalho acumulada (*Acute:Chronic Workload Ratio - ACWR*). A indicação à solução seria de classificação binária com janelas temporais deslizantes, a qual transformaria o problema contínuo em classificação sim ou não. Inicialmente, definido seria o subconjunto de dados de tamanho fixo, denominado janela de tamanho n . Essa capturaria a sequência de tempo de $t - n$ até t , predizendo o desfecho (sim ou não) no instante posterior, $t + 1$, ou dentro da janela. Algoritmos XGBoost e CatBoost poderiam ser adequados, utilizando o ajuste específico (aprendizado desbalanceado - *imbalanced learning*, por exemplo) de pesos em razão da magnitude de probabilidade de lesão (Martins, 2024). Os modelos sequenciais priorizariam a sensibilidade, corrigindo iterativamente os falsos negativos (ocorrência de lesão sem aviso ao atleta), e forneceria a importância das variáveis, favorecendo a identificação de fatores de risco.

Comumente a área de saúde demandaria generalizações, tratamento de dados faltantes e interpretação das predições, de forma ímpar para elevar a qualidade de vida do paciente e reduzir os custos hospitalares. Nesse norte, relevante seria prever a readmissão hospitalar após 30 dias da alta, propiciando ações como acompanhamento domiciliar e ajuste medicamentoso. O modelo seria de classificação binária com dados estruturados como idade, tempo de permanência no hospital, número de internações, fármacos, comorbidades e exames laboratoriais. Tal situação seria adequada ao XGBoost, o qual apresentaria desempenho elevado na ausência de dados, aprendendo a direção da árvore de decisão na falta de dados (Bezerra e Almeida, 2024; Scursone *et al.*, 2025).

Embora, o enquadramento do processamento de dados não-estruturados como limitação dos modelos sequenciais fosse verdadeiro, a extração de dados de imagens de médicas (*radiômica*) seria possível se o banco de dados possuir tamanho moderado (no máximo alguns milhares de pacientes, idealmente, centenas), mesmo na presença de muitas variáveis (alta dimensionalidade), sob tais condições, o desempenho poderia superar redes neurais simples e evitar o ajuste excessivo (Bastos e Mamede, 2025; Santos, 2025). Dados clínicos coadunados às características de lesões, como textura, intensidade, localização, padrão de calcificação e forma, determinariam nódulos como malignos ou

benignos, por exemplo, então tratar-se-ia de classificação (binária ou multiclasse), possibilitando a aplicação de LightGBM ou CatBoost.

A área de negócios seria amplamente demandadora de e abastecida por dados estruturados, com origem em Sistema Integrado de Gestão Empresarial ou Planejamento de Recursos Empresariais (*Enterprise Resource Planning* - ERP), Gerência de Relacionamento com o Cliente (*Customer Relationship Management* - CRM) ou arquivos de Linguagem de Consulta Estruturada (*Structured Query Language* – SQL). O primeiro integraria os processos internos da Organização (Caiçara Júnior, 2015; Singh e Lai, 2026), e o segundo auxiliaria a integração com o cliente externo (Zenone, 2019; Eng, Lin e Wu, 2026). O SQL seria arquivo de texto simples com comandos e roteiros para gerência de bancos de dados relacionais, portanto tratar-se-ia de linguagem padronizada para gerenciamento e recuperação de dados (Beaulieu, 2019; Li *et al.*, 2026). Em razão do imediatamente exposto, os modelos de *boosting* seriam amplamente aplicados, particularmente, em operações, finanças e marketing.

Exemplificando, a identificação de clientes externos com alta probabilidade de cancelamento permitiria implementar ações de retenção. Modelos sequenciais permitiriam ajustar pesos ou empregar a área sobre a curva (*Area Under the Curve* – AUC ou *Area Under the Curve* – *Receiver Operating Characteristic* – AUC-ROC) para validação, o que favoreceria a caracterização por classificação binária com dados transacionais, demográficos ou de uso (por exemplo: consumo de serviço). Assim entendido, o problema poderia ser exposto ao LightGBM, o qual processaria bancos de dados com diversas variáveis e milhões de clientes em velocidade elevada (Pinheiro *et al.*, 2024; Medeiros e Santos, 2024).

Comumente Organizações de transporte aéreo ou por aplicativo, hospedagem e comércio eletrônico implementariam modelos de *boosting* (regressão ou incremento de conversão - *uplift*) na precificação em tempo real. A adoção de preços dinâmicos objetivaria maximizar receita ou lucro, tendo por variáveis: sazonalidade, elasticidade de demanda, estoque e preços da concorrência, dentre outras. O GBM seria utilizado para predição de probabilidade de aquisição por classe de preço, porém com função de perda personalizada. O XGBoost modelaria, de forma não paramétrica, a curva de demanda, pois capturaria adequadamente interações complexas, por exemplo: atenuação da conversão pela combinação entre alta concorrência, preço alto e feriado (Haery, Mahpour e Vafaeinejad, 2024; Kumar e Kannan, 2025).

A combinação de modelos sequenciais e redes neurais seria, não raramente, utilizada, dadas à elevada precisão e baixa latência, na identificação de fraudes ou transações suspeitas em seguros, transferências bancárias e cartões de crédito. Caracteristicamente, o problema seria de classificação binária, conseqüentemente, adequada seria a implementação de XGBoost com parada antecipada (*early stopping*) para evitar o ajuste excessivo, e subamostragem por linha e coluna, uma vez que imperativo seria o fornecimento de explicação a órgãos reguladores. Aquela família de algoritmos daria elevada capacidade de processar fenômenos demasiadamente minoritários (*extreme class imbalance* – Desequilíbrio Extremo de Classes), aspecto relevante, dado que as fraudes representariam menos de 0,10% das transações (Alfocea, 2021; Sánchez-Mojica e López-Nuñez, 2025).

Considerações Finais

O estudo objetivou discutir os modelos de *boosting*, fornecendo exemplos de aplicação. Demonstrou-se que os modelos sequenciais seriam recursos significativos em aprendizado supervisionado, particularmente, por equilibrar precisão e capacidade de generalização, considerando dados estruturados. A fragilidade residiria, especialmente, na tendência de ajuste excessivo, o que poderia ser evitado com o desenvolvimento sistematizado de validação cruzada e regularização.

Aos estudos futuros, recomenda-se a implementação de algoritmos para predição de desempenho esportivo, o qual seria, essencialmente, resultado de treinos sequenciais. Relevante seria demonstrar a robustez em problema de taxa de cancelamento em escola esportiva ou instituição de ensino. No condizente à saúde, o desenvolvimento de pesquisa que favoreça a redução de dimensionalidade no diagnóstico de câncer, depressão ou ansiedade tenderia a favorecer a redução da lacuna de tempo à intervenção adequada.

Referências

- AL QASSEM, LM *et al.* Proactive random-forest autoscaler for microservice resource allocation. **IEEe Access**, v. 11, p. 2570-2585, 2023.
- ALFOCEA, A. Comment le machine learning permet de détecter la fraude bancaire?. **Management & Datascience**, v. 5, n. 4, 2021. DOI: 10.36863/mds.a.16671
- AL-STOUHI, S; REDDY, CK. Adaptive boosting for transfer learning using dynamic updates. In: **Joint European Conference on Machine Learning and Knowledge Discovery in Databases**. Berlin, Heidelberg (Deutschland): Springer Berlin Heidelberg, 2011. p. 60-75.
- ARRUDA, VF *et al.* Algoritmos de reforço na previsão da força máxima de aderência aço-concreto em barras finas. **Ambiente Construído**, v. 25, e149626, 2025.

BAPTISTA, GC; RODRIGUES, CA. Previsão das movimentações de ativos financeiros usando ensemble learning para implementação de um portfólio descorrelacionado. *In: Escola Regional de Computação Bahia, Alagoas e Sergipe (ERBASE)*. SBC, 2024. p. 149-158.

BARRETO, ACLG *et al.* Idade como moderadora da variabilidade da frequência cardíaca no treinamento intervalado de alta intensidade. *Revista Presença*, v. 11, n. 26, p. 204-216, 2025.

BARRETO, ACLG. **Respostas fisiológicas e bioquímicas em indivíduos praticantes de Crossfiit® submetidos a protocolos de treinamento físico agudos e crônicos**. Tese (Doutorado em Educação Física) – Programa de Pós-graduação Stricto-Sensu em Educação Física. Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2023.

BASTOS, AL; MAMEDE, M. Radiomics in PET/CT and HRCT for systemic sclerosis-associated interstitial lung disease: breakthroughs and future directions. *Radiologia Brasileira*, v. 58, e20250021, 2025.

BATISTA, JJS. **Recomendação de produtos financeiros utilizando aprendizado de máquina**. Dissertação (Mestrado em Tecnologia da Informação) – Programa de Pós-graduação em Tecnologia da Informação. Instituto MetrÓpole Digital. Universidade Federal do Rio Grande do Norte. Natal (RN), 2024.

BEAULIEU, A. **Aprendendo SQL: dominando os fundamentos de SQL**. São Paulo: Novatec, 2019.

BENTÉJAC, C; CSÖRGŐ, A; MARTÍNEZ-MUÑOZ, G. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, v. 54, n. 3, p. 1937-1967, 2021.

BEZERRA, FDV. **Avaliação de métodos de ensemble dinâmico em aprendizado de máquina para previsões de irradiância solar e velocidade do vento**. Dissertação (Mestrado em Engenharia Mecânica) – Programa de Pós-graduação em Engenharia Mecânica. Departamento de Engenharia Mecânica. Universidade Federal do Ceará. Fortaleza (CE), 2024.

BEZERRA, JHS; ALMEIDA, FM. Desenvolvimento de modelos preditivos com machine learning - análise de dados para saúde de gestantes e puérperas. *InterSciencePlace*, v. 19, a. 16, p. 304-335, 2024.

BORGES, FFR. **Variabilidade da frequência cardíaca de corredores durante teste incremental máximo em esteira ergométrica**. Dissertação (Mestrado em Engenharia Biomédica) – Programa de pós-graduação em Engenharia Biomédica. Faculdade de Engenharia Elétrica. Universidade Federal de Uberlândia. Uberlândia (MG), 2022.

BUHLMANN, P; HOTHORN, T. Boosting algorithms: regularization, prediction and model fitting. *Statistical Science*, v. 22, n. 4, p. 477-505, 2007.

BUSCHJÄGER, S; MORIK, K. Decision tree and random forest implementations for fast filtering of sensor data. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 65, n. 1, p. 209-222, 2017.

CABALHEIRO, TF. **Comparação de arquiteturas de Word2Vec na análise de textos curtos**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Departamento de Estatística. Instituto de Matemática e Estatística. Universidade Federal do Rio Grande do Sul. Porto Alegre (RS), 2023.

CAI, Y; YUAN, Y; ZHOU, A. Predictive slope stability early warning model based on CatBoost. *Scientific reports*, v. 14, n. 1, a. 25727, 2024.

CAIÇARA JÚNIOR, C. **Sistemas integrados de gestão – ERP: Uma abordagem gerencial**. Curitiba (PR): Intersaberes, 2015.

CAMARGO, JS. **Método bagging para aprimoramento de previsões de séries temporais.** Dissertação (Mestrado em Estatística) – Programa Interinstitucional de Pós-graduação em Estatística. Departamento de Estatística. Instituto de Ciências Matemáticas e de Computação. Universidade de São Paulo. São Carlos (SP), 2021.

CARVALHO JUNIOR, S *et al.* Análise de sobrevivência e regime de lesões em maratonistas. **Revista Presença**, v. 11, n. 26, p. 169-183, 2025.

CELERI, MO. **Stacking Ensemble Learning para seleção genômica em características complexas.** Tese (Doutorado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa. Viçosa (MG), 2025.

CENTENO, NR.V. Aplicación del modelo de machine learning (Random Forest) para la identificación de los indicadores financieros más influyentes en la predicción del desempeño de la industria minera peruana. **Económicas Habla**, n. 6, p. 49-52, 2025.

CHAVES, BB. **Estudo do algoritmo adaboost de aprendizagem de máquina aplicado a sensores e sistemas embarcados.** Dissertação (Mestrado em Engenharia Mecânica) – Escola Politécnica. Universidade de São Paulo. São Paulo, 2012.

CORREIA, CK *et al.* Fatores de risco biomecânicos para lesões relacionadas a corrida: uma revisão guarda-chuva. **Brazilian Journal of Motor Behavior**, v. 19, n. S2, p. 93-94, 2025.

DÂMASO, FT *et al.* Aplicação das técnicas de machine learning na categorização de despesas de fluxo de caixa: uma pesquisa-ação. **Revista Inovação, Projetos e Tecnologias**, v. 13, n. 1, 2025. DOI: <https://doi.org/10.5585/2025.27432>.

DAMKE, GT; DAMKE, MT; GREGORINI, DM. Revisão de abordagens baseadas em aprendizado de máquina para o diagnóstico de diabetes mellitus tipo 2. **Revista Caribeña**, v. 14, n. 2, e4449, 2025.

DANTAS, JCS; ARANHA, EHS; SILVA, TR. Predição de alunos em risco de reprovação: uma comparação do impacto de diferentes técnicas de amostragem. **RENOTE**, v. 22, n. 3, p. 230-239, 2024.

DATTA, P; DAS, P; KUMAR, A. Hyper parameter tuning based gradient boosting algorithm for detection of diabetic retinopathy: an analytical review. **Bulletin of Electrical Engineering and Informatics**, v. 11, n. 2, p. 814-824, 2022.

DU, KL *et al.* Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era. **Mathematics**, v. 10, n. 24, a. 4730, 2022.

ENG, TY; LIN, F; WU, LC. Strengthening CRM performance through customer knowledge: the role of technological capability in SMEs. **European Business Review**, v. 38, n. 3, p. 403-428, 2026.

EVERTON, ANF *et al.* Caracterização do treinamento de praticantes de ultramaratona. **Revista Presença**, v. 10, n. 22, p. 41-63, 2024.

FAN, Z; GOU, J; WENG, S. Complementary CatBoost based on residual error for student performance prediction. **Pattern Recognition**, v. 161, a. 111265, 2025.

FELICIANO, MBD. **Classificação e predição de preços de imóveis a partir de dados estruturados e não estruturados.** Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Centro de Informática. Universidade Federal de Pernambuco. Recife (PE), 2018.

FERRARIS, L; LIAUTAUD, P. **Method of gradient boosting.** Paris (France): Sorbonne Université, 2022.

FREITAS, FW. **Estudo comparativo de modelos de Machine Learning na predição de um Behaviour Score**. Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Departamento de Estatística. Universidade Federal de São Carlos. São Carlos (SP), 2025.

FRIEDMAN, J. Greedy function approximation: A gradient boosting machine. **The Annals of Statistics**, v. 29, p. 1189-1232, 2001.

GONÇALVES, JL. **Um processo para construção de datasets que viabilizem a aplicação de diferentes métodos de detecção de fake news e bots sociais**. Dissertação (Mestrado em Ciências de Sistemas e Computação) - Programa De Pós-graduação em Ciência Sistemas e Computação. Departamento de Ciência e Tecnologia. Instituto Militar de Engenharia. Rio de Janeiro, 2024.

GUZMÁN, DT; VIVAS, JDP; MORALES, EB. Tortuosity and discrete compactness biomarkers for machine learning-based classification of mild cognitive impairment. **Biomedical Signal Processing and Control**, v. 112, a. 108848, 2026.

HAERY, S; MAHPOUR, A; VAFAEINEJAD, A. Forecasting urban travel demand with geo-AI: a combination of GIS and machine learning techniques utilizing uber data in New York City. **Environmental Earth Sciences**, v. 83, n. 20, a. 594, 2024.

HESPANHOL JUNIOR, LC *et al.* Perfil das características do treinamento e associação com lesões musculoesqueléticas prévias em corredores recreacionais: um estudo transversal. **Brazilian Journal of Physical Therapy**, v. 16, p. 46-53, 2012.

KUMAR, IV; KANNAN, KN. Optimizing Uber Data Analysis Using Decision Tree and Random Forest. In KALITA, K *et al.* **Decision-Making Techniques and Methods for Sustainable Technological Innovation: Strategies and Applications in Industry 5.0**. Beverly (USA): Scrivener Publishing LLC, 2025, p. 183-201.

LI, C *et al.* Sea-sql: Semantic-enhanced text-to-sql with adaptive refinement. **Frontiers of Computer Science**, v. 20, n. 3, p. 2003602, 2026.

LISKA, GR *et al.* Classificação de dados em modelos com resposta binária via algoritmo boosting e regressão logística. **Revista Matemática e Estatística em Foco-ISSN**, v. 2318, p. 0552, 2012.

LOKKER, C *et al.* Boosting efficiency in a clinical literature surveillance system with LightGBM. **PLOS Digital Health**, v. 3, n. 9, e0000299, 2024.

LOPEZ, JWM; LOPEZ, JMB; AGUADED, I. Predictive model for beneficiary households in cash transfer programs: a comparison of machine learning techniques. **RISTI (Revista Iberica de Sistemas e Tecnologias de Informacao)**, n. 57, p. 3-19, 2025.

MANTILLA, JFN. **Aplicación de boosting en la clasificación de poblaciones: un estudio práctico**. Proyecto de Grado (Título de Ingeniero Electrónico) – Departamento de Ingeniería Eléctrica y Electrónica. Facultad de Ingeniería. Universidad de los Andes. Bogotá, D.C. (Colombia), 2004.

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

MAYR, A *et al.* The evolution of boosting algorithms. **Methods of information in medicine**, v. 53, n. 06, p. 419-427, 2014.

MEDEIROS, JS; SANTOS, JEP. **O relacionamento com o cliente como estratégia de relações públicas e ferramenta de crescimento comercial: a aplicação do departamento de sucesso do aluno na Faculdade Raimundo Marinho**. Trabalho de Conclusão de Curso (Bacharelado em Relações Públicas) – Instituto de Ciências Humanas, Comunicação e Artes. Universidade Federal de Alagoas. Maceió (AL), 2024.

MENEGHETTI, HG *et al.* O uso da análise da variabilidade da frequência cardíaca no monitoramento de lesões esportivas e sua influência sobre o balanço autonômico: uma revisão sistemática. **Fisioterapia e Pesquisa**, v. 28, n. 3, p. 291-298, 2021.

MONTENEGRO, TC. **Uma arquitetura de referência para explicabilidade como serviço na saúde H-XAIaaS: Health-eXplainable Artificial Intelligence as a Service**. Dissertação (Mestrado em Informática) – Centro de Informática. Universidade Federal da Paraíba. João Pessoa (PB), 2025.

MOREIRA, AA *et al.* Técnicas de ensemble learning para sistema de detecção de intrusão no contexto da cibersegurança. **Revista de Segurança da Informação e Comunicação**, v. 10, n. 1, p. 1-15, 2021.

MUSOLF, AM *et al.* What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. **Human Genetics**, v. 141, n. 9, p. 1515-1528, 2022.

NARASIMHAN, G; VICTOR, A. A hybrid approach with metaheuristic optimization and random forest in improving heart disease prediction. **Scientific Reports**, v. 15, n. 1, a. 10971, 2025.

NASCIMENTO JÚNIOR, JFS. **Avaliação funcional da interpretabilidade de modelos de classificação *boosting* para previsão de inadimplência de crédito**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal do Ceará. Russas (CE), 2026.

NATEKIN, A; KNOLL, A. Gradient boosting machines, a tutorial. **Frontiers in Neurorobotics**, v. 7, a. 63623, 2013. DOI: <https://doi.org/10.3389/fnbot.2013.00021>

NIAZKAR, M *et al.* Applications of XGBoost in water resources engineering: A systematic literature review (Dec 2018–May 2023). **Environmental Modelling & Software**, v. 174, p. 105971, 2024.

OGUNLEYE, A; WANG, QG. XGBoost model for chronic kidney disease diagnosis. **IEEE/ACM transactions on computational biology and bioinformatics**, v. 17, n. 6, p. 2131-2140, 2019.

ORTIZ, SG *et al.* Una arquitectura de análisis de imágenes seriadas con la tomografía por emisión de positrones mediante la aplicación de machine learning combinado para la detección del cáncer de pulmón. **Revista Española de Medicina Nuclear e Imagen Molecular**, v. 43, n. 3, a. 500003, 2024.

PÁEZ, AR; RAMÍREZ, NDG. Ejercicios escolares para la predicción del rendimiento académico de estudiantes universitarios con técnicas de aprendizaje automático combinadas. **RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo**, v. 16, n. 32, e1040, 2026.

PAIXÃO, KWM; LOMBARDI, G; BARROSO, PRA. Machine Learning na prevenção de perdas no e-commerce: o uso da regressão logística para identificação de pedidos fraudulentos. **REFAS: Revista FATEC Zona Sul**, v. 11, n. 4, p. 1-17, 2025.

PANDOLFO, PT. **Machine learning na predição de desembolsos financeiros das manutenções de máquinas na colheita de madeira**. Dissertação (Mestrado em Ciência Florestal) – Faculdade de Ciências Agrônômicas. Universidade Estadual Paulista. Botucatu (SP), 2025.

PASTRE, CM *et al.* Exploração de fatores de risco para lesões no atletismo de alta performance. **Revista Brasileira de Medicina do Esporte**, v. 13, n. 3, p. 200-204, 2007.

PEREIRA, MM. **Árvore de Regressão aplicada à identificação e priorização de perdas não técnicas**. Dissertação (Mestrado em Engenharia Elétrica) - Programa de Pós-graduação em Engenharia de Eletricidade/CCET. Universidade Federal do Maranhão. São Luís (MA), 2025.

PEREIRA, RP. **Modelagem preditiva da produtividade da soja utilizando aprendizado de máquina**. Tese (Doutorado em Engenharia Agrícola) - Programa de Pós-graduação em Engenharia Agrícola). Universidade Estadual do Oeste do Paraná. Cascavel (PR), 2024.

PINELLI, S *et al.* Exploring the Relationship Between the Acute: Chronic Workload Ratio and Running Parameters in Elite Football Athletes. **Applied Sciences**, v. 15, n. 3, a. 1659, 2025.

PINHEIRO, LDPS *et al.* Rates and causes of outpatient surgical cancellations: a cross-sectional study. **Texto & Contexto-Enfermagem**, v. 33, e20240026, 2024.

PONCE-BOBADILLA, AV *et al.* Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. **Clinical and Translational Science**, v. 17, n. 11, e70056, 2024.

QIN, W; LI, R; CHEN, L. Acute to chronic workload ratio (ACWR) for predicting sports injury risk: a systematic review and meta-analysis. **BMC Sports Science, Medicine and Rehabilitation**, v. 17, n. 1, a. 285, 2025.

REIS, G; RUPPERT, G. Análise comparativa de algoritmos de aprendizado de máquina para predição de recidiva de câncer de próstata com dados da Fundação Oncocentro de São Paulo. **Revista Eletrônica de Iniciação Científica em Computação**, v. 24, n. 1, p. 192-197, 2026.

RODRIGUES, MES. **Análise da técnica de corrida por meio de variáveis biomecânicas em corredores de rua de diferentes faixas etárias**. Dissertação (Mestrado em Engenharia Biomédica) – Programa de Pós-graduação em Engenharia Biomédica. Faculdade de Engenharia Elétrica. Universidade Federal de Uberlândia. Uberlândia (MG), 2019.

SÁNCHEZ-MOJICA, KY; LÓPEZ-NUÑEZ, M. Modelo de inteligencia artificial hibrido para el perfilamiento de usuarios como estrategia de detección de fraudes en redes de fibra óptica. **AiBi Revista de Investigación, Administración e Ingeniería**, v. 13, n. 1, p. 159-164, 2025.

SANTOS, HG *et al.* Machine learning para análises preditivas em saúde: exemplo de aplicação para prever óbito em idosos de São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 35, n. 7, e00050818, 2019.

SANTOS, LA. **O papel do [18F] FDG PET/CT na predição histológica do câncer de pulmão de não pequenas células: uma abordagem de aprendizado de máquina baseada em radiômica**. Tese (Doutorado em Bioengenharia) – Programa Interunidades em Bioengenharia. Escola de Engenharia de São Carlos. Faculdade de Medicina de Ribeirão Preto. Instituto de Química de São Carlos. Universidade de São Paulo. São Carlos (SP), 2025.

SCHMOELLER, RP; PERON, VD; MESQUITA FILHO, W. Desempenho acadêmico na previsão da evasão no ensino superior: comparação entre modelos bagging, boosting e ensemble de votação majoritária. **Revista Pleiade**, v. 19, n. 49, p. 42-55, 2025.

SCURSONE, GF *et al.* Predictive modeling with XGBOOST for respiratory hospitalizations in São Paulo: performance and decision support in the public health system. **Revista Políticas Públicas & Cidades**, v. 14, n. 9, e2705, 2025.

SICILIANO, JPAA *et al.* Eventos lesivos em categorias de base no futebol à luz da epidemiologia. **Revista Presença**, v. 9, n. 20, p. 67-88, 2023.

SILVA R, AD. **Machine learning para a predição da carga de treinamento de um atleta da Seleção Brasileira de Taekwondo**. Monografia (Especialização em Estatística) – Programa de Pós-graduação em Estatística. Instituto de Ciências Exatas. Universidade Federal de Minas Gerais. Belo Horizonte (MG), 2023.

SILVA, CB *et al.* Uma análise comparativa das técnicas de machine learning: regressão logística, árvores de decisão, random forest e SVM. **Apoena**, v. 7, p. 501-511, 2023.

SILVA, EG; FERREIRA, KC. Gestão de rejeitos de minério de ferro: implementação de filtro prensa como estratégia para sustentabilidade ambiental. **Revista Foco**, v. 18, n. 7, e9171, 2025.

SILVA, EVA. **Estudo de desempenho de modelos de IA aplicados ao churn de clientes.** Trabalho de Conclusão de Curso (Bacharelado em Estatística) - Instituto de Ciências Exatas e Biológicas. Universidade Federal de Ouro Preto. Ouro Preto (MG), 2023.

SILVA, JE; SÁTIRO, RM. O poder preditivo dos modelos boosting de machine learning no mercado brasileiro de ações. **Contabilometria**, v. 11, n. 1, p. 52-68, 2024.

SINGH, N; LAI, KH. Enterprise resource planning implementation in large-scale operations: a case study in enterprise information systems. **Enterprise Information Systems**, v. 20, n. 2, a. 2624368, 2026.

SOUZA, DHM; JÚNIOR, CJB. Novo algoritmo ensemble para detecção de fraude em transações de cartão de crédito. **Revista Tecnologia e Sociedade**, v. 19, n. 56, p. 128-145, 2023.

TÁVORA, VJ. **Machine Learning interpretável na avaliação de imóveis em massa:** aplicação do SHAP. Dissertação (Mestrado Profissional em Economia do Setor Público) - Faculdade de Economia, Administração, Atuária e Contabilidade. Universidade Federal do Ceará. Fortaleza (CE), 2025.

TERTULINO, R; DUARTE, R. Inteligência Artificial Explicável (XAI) Aplicada ao SAEB: uma análise preditiva dos fatores que influenciam o desempenho estudantil. **RENOTE**, v. 24, n. 2, p. 255-266, 2025.

THOMAS, J *et al.* Gradient boosting for distributional regression: faster tuning and improved variable selection via noncyclical updates. **Statistics and Computing**, v. 28, n. 3, p. 673-687, 2018.

TIGGEMANN, CL; GOSSMANN, J; CREMONESE, C. Perfil, prevalência e fatores de risco a lesões em corredores amadores do Rio Grande do Sul. **Arquivos de Ciências da Saúde da UNIPAR**, v. 26, n. 3, p. 845-861, 2022.

VELOSO, MF *et al.* Utilização do algoritmo de árvore de decisão para mapeamento do uso e cobertura da terra da área de contribuição de uma barragem localizada na Bacia Hidrográfica do São Francisco. **Caderno de Ciências Agrárias**, v. 13, p. 01-09, 2021. DOI: <https://doi.org/10.35699/2447-6218.2021.290392021>.

VIANA, LAL *et al.* Aplicação de machine learning na previsão de demandas: otimização de recursos no setor de e-commerce. **Revista Contemporânea**, v. 4, n. 11, p. e6545-e6545, 2024.

VILLCA, JEC. Predicción de la pobreza en Bolivia usando Machine Learning. **Cuadernos de Investigación**, v. 8, p. 1, n. 3-20, 2025.

WANG, D; LI, L; ZHAO, D. Corporate finance risk prediction based on LightGBM. **Information Sciences**, v. 602, p. 259-268, 2022.

WOJTAS, M; CHEN, K. Feature importance ranking for deep learning. **Advances in Neural Information Processing Systems**, v. 33, p. 5105-5114, 2020.

YU, TK *et al.* Predicting potential soil and groundwater contamination risks from gas stations using three machine learning models (XGBoost, LightGBM, and Random Forest). **Process Safety and Environmental Protection**, v. 199, a. 107249, 2025.

ZENONE, LC. **CRM (Customer Relationship Management):** marketing de relacionamento, fidelização de clientes e pós-venda. Lisboa (Portugal): Actual, 2019.

ZHENG, Z; YANG, Y. Adaptive boosting for domain adaptation: toward robust predictions in scene segmentation. **IEEE Transactions on Image Processing**, v. 31, p. 5371-5382, 2022.