v7.n6.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

ANÁLISE COMPARATIVA DE ALGORITMOS DE PREDIÇÃO EM DADOS TABULARES: REGRESSÃO LINEAR, RANDOM FOREST, GRADIENT BOOSTING E REDES NEURAIS MLP

COMPARATIVE ANALYSIS OF PREDICTION ALGORITHMS ON TABULAR DATA: LINEAR REGRESSION, RANDOM FOREST, GRADIENT BOOSTING, AND MLP NEURAL NETWORKS

ANÁLISIS COMPARATIVO DE ALGORITMOS DE PREDICCIÓN EN DATOS TABULARES: REGRESIÓN LINEAL, BOSQUE ALEATORIO, POTENCIACIÓN DE GRADIENTE Y REDES NEURONALES MLP

Evandro Preuss[1], Íthalo Paterno Argenta[2], Roberto Franciscatto[3], Solange de Lurdes Pertile[4]

e768276
https://doi.org/10.47820/recima21.v7i7.8276
PUBLICADO: 06/2026

RESUMO

Este trabalho apresenta uma análise comparativa entre quatro algoritmos de aprendizado de máquina, aplicados a um problema de regressão supervisionada sobre dados tabulares reais: Regressão Linear, Random Forest, *Gradient Boosting* e Rede Neural Perceptron Multicamadas (MLP). O estudo adota um protocolo metodológico, contemplando pré-processamento com isolamento estrito das partições, validação cruzada K-Fold e escalonamento de atributos diferenciado por modelo. Os modelos são avaliados pelo coeficiente de determinação (R^2) e pela raiz do erro quadrático médio (RMSE). Os resultados demonstram a superioridade dos métodos não lineares: Random Forest atingiu $R^2 \approx 0{,}71$, contra $R^2 \approx 0{,}16$ da Regressão Linear. Os achados fornecem subsídios empíricos para a escolha algorítmica em tarefas de regressão e apontam perspectivas para trabalhos futuros com otimização de hiperparâmetros e extensão a múltiplos conjuntos de dados.

PALAVRAS-CHAVE: Aprendizado de Máquina. Algoritmos de Regressão. Dados Tabulares.

ABSTRACT

This work presents a comparative analysis of four machine learning algorithms applied to a supervised regression problem on real tabular data: Linear Regression, Random Forest, Gradient Boosting, and Multilayer Perceptron (MLP) Neural Network. The study adopts a methodological protocol, including preprocessing with strict partition isolation, K-Fold cross-validation, and differentiated feature scaling per model. The models are evaluated by the coefficient of determination (R^2) and the root mean squared error (RMSE). The results demonstrate the superiority of the non-linear methods: Random Forest achieved $R^2 \approx 0.71$, compared to $R^2 \approx 0.16$ for Linear Regression. The findings provide empirical support for algorithmic selection in regression tasks and point to perspectives for future work with hyperparameter optimization and extension to multiple datasets.

[1] Doutor em Informática na Educação pela Universidade Federal do Rio Grande do Sul (PPGIE/UFRGS). Professor na Universidade Federal de Santa Maria - RS (UFSM).
[2] Bacharelado em Sistemas de Informação pela Universidade Federal de Santa Maria - RS (UFSM).
[3] Doutor em Informática na Educação pela Universidade Federal do Rio Grande do Sul (PPGIE/UFRGS). Professor na Universidade Federal de Santa Maria - RS (UFSM).
[4] Doutora em Computação pela Universidade Federal do Rio Grande do Sul (UFRGS).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

KEYWORDS: *Machine Learning. Regression Algorithms. Tabular Data.*

RESUMEN
Este trabajo presenta un análisis comparativo de cuatro algoritmos de aprendizaje automático aplicados a un problema de regresión supervisada en datos tabulares reales: Regresión Lineal, Bosque Aleatorio, Gradient Boosting y Red Neuronal Perceptrón Multicapa (MLP). El estudio adopta un protocolo metodológico que incluye preprocesamiento con aislamiento estricto de particiones, validación cruzada K-Fold y escalado de características diferenciado por modelo. Los modelos se evalúan mediante el coeficiente de determinación (R^2) y el raíz del error cuadrático medio (RMSE). Los resultados demuestran la superioridad de los métodos no lineales: Bosque Aleatorio alcanzó un $R^2 \approx 0,71$, en comparación con un $R^2 \approx 0,16$ para la Regresión Lineal. Los hallazgos proporcionan apoyo empírico para la selección de algoritmos en tareas de regresión y señalan perspectivas para trabajos futuros con optimización de hiperparámetros y extensión a múltiples conjuntos de datos.

PALABRAS CLAVE*: Aprendizaje automático. Algoritmos de regresión. Datos tabulares.*

1. INTRODUÇÃO

O avanço das técnicas de aprendizagem de máquina (*machine learning*) nas últimas décadas consolidou a predição sobre dados como uma das principais frentes de aplicação em ciência de dados. Tarefas preditivas permeiam domínios tão distintos quanto finanças, saúde, engenharia e ciências agrárias, tornando indispensável a compreensão aprofundada dos algoritmos disponíveis e de suas características intrínsecas. Nesse cenário, a capacidade de antecipar valores contínuos a partir de variáveis de entrada estruturadas, isto é, a regressão supervisionada, emerge como um dos pilares da inteligência analítica moderna, subsidiando decisões estratégicas em ambientes de alta complexidade e incerteza.

A relevância da predição não reside apenas na acurácia das estimativas produzidas, mas também na confiabilidade e na generalização dos modelos quando submetidos a dados inéditos. Sob a ótica da Teoria de Decisão Estatística, o objetivo central é encontrar uma função f(X) capaz de minimizar a função de perda esperada (como o Erro Quadrático Médio), garantindo que as previsões se aproximem com a maior fidelidade possível dos valores reais (Hastie; Tibshirani; Friedman, 2009). Entretanto, a qualidade desse mapeamento depende não somente da escolha do algoritmo, mas também do rigor metodológico aplicado nas etapas de pré-processamento, escalonamento e validação dos dados.

Nesse contexto, a seleção do algoritmo preditivo mais adequado para um determinado problema constitui uma decisão não trivial. O teorema do *No Free Lunch* (Wolpert, 1996) estabelece que, na ausência de conhecimento prévio sobre a estrutura dos dados, nenhum algoritmo pode ser considerado universalmente superior a outro. Essa premissa reforça a necessidade de estudos comparativos rigorosos que avaliem o comportamento de diferentes

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

modelos sob condições controladas e sobre conjuntos de dados reais, evitando generalizações precipitadas a partir de benchmarks sintéticos ou de domínios específicos.

Entre os algoritmos amplamente empregados para problemas de regressão, destacam-se a Regressão Linear, o Random Forest, o *Gradient Boosting* e as Redes Neurais do tipo Perceptron Multicamadas (MLP). Cada um desses métodos representa uma abordagem distinta ao problema de aprendizagem supervisionada: enquanto a Regressão Linear oferece alta interpretabilidade ao custo de suposições restritivas de linearidade, os métodos de ensemble (Random Forest e Gradient Boosting) exploram a combinação de múltiplos estimadores para capturar padrões não lineares com maior robustez. As Redes Neurais MLP, por sua vez, ampliam esse potencial por meio de arquiteturas multicamadas e funções de ativação não lineares, capazes de aproximar relações de alta complexidade (Hastie; Tibshirani; Friedman, 2009). A despeito da disponibilidade dessas abordagens, análises comparativas conduzidas sobre *datasets* reais (considerando simultaneamente desempenho preditivo, estabilidade e aplicabilidade prática) ainda representam uma contribuição relevante para a comunidade científica.

Motivado por essa lacuna, o presente trabalho realiza uma análise comparativa entre os quatro algoritmos mencionados, aplicados a um conjunto de dados tabulares reais em um problema de regressão. A escolha de um *dataset* do mundo real permite avaliar os modelos diante de desafios concretos, como presença de valores ausentes, outliers, variáveis categóricas e distribuições assimétricas, tornando os resultados mais representativos das condições encontradas em aplicações práticas. Para tanto, adota-se um protocolo rigoroso de pré-processamento (incluindo o isolamento correto das partições de treinamento e teste, o emprego de validação cruzada K-Fold e a aplicação de técnicas de escalonamento adequadas a cada modelo), conforme preconizado por Géron (2025) e Kuhn e Johnson (2013).

Dentro dessa perspectiva, o presente estudo tem por objetivo geral medir e comparar o desempenho preditivo, a confiabilidade estatística e a aplicabilidade prática dos quatro algoritmos selecionados: Regressão Linear, Random Forest, *Gradient Boosting* e MLP, por meio de métricas quantitativas obtidas em condições metodologicamente equivalentes. Como objetivos específicos, destacam-se: (a) implementar um protocolo metodológico de pré-processamento com isolamento estrito das partições de treinamento e teste; (b) avaliar os modelos pelos coeficientes R^2 e RMSE em condições de validação cruzada K-Fold; e (c) oferecer subsídios empíricos que orientem a escolha algorítmica em problemas de regressão sobre dados tabulares. A justificativa do estudo fundamenta-se no problema da ausência de análises comparativas conduzidas sobre dados reais que avaliem simultaneamente desempenho,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

estabilidade e aplicabilidade prática, contemplando desafios concretos como valores ausentes, outliers, variáveis categóricas e distribuições assimétricas.

O artigo em questão está estruturado da seguinte forma: a Seção 2 apresenta a fundamentação teórica dos conceitos de predição, pré-processamento e dos algoritmos estudados; a Seção 3 descreve a metodologia adotada; a Seção 4 expõe os resultados e a discussão comparativa; e a Seção 5 traz as conclusões e perspectivas futuras.

2. PREDIÇÃO DE DADOS

O processo de aprendizagem supervisionada é um método no qual se utiliza um conjunto de variáveis de entrada, medidas ou predeterminadas, com o objetivo de prever valores de saída. Sob o viés matemático, o problema da predição consiste em buscar uma função f(X) que seja capaz de estimar a saída Y ao ser exposta a novos dados. Para alcançar esse objetivo é essencial ter disponível um conjunto de treinamento que segue denotado por (xi, yi) para i=1,…, N, em que o modelo é treinado com base nos exemplos disponíveis.

Em vista disso, sob a ótica da Teoria de Decisão Estatística, a predição é tratada como um processo analítico que busca encontrar uma função capaz de minimizar uma Função de Perda Esperada, como, por exemplo, o erro quadrático médio, garantindo que as previsões geradas pelo modelo se aproximem com a maior fidelidade possível dos dados reais (Hastie; Tibshirani; Friedman, 2009).

Pelo ponto de vista estatístico, assume-se que os dados são uma derivação do modelo aditivo que possui a seguinte formulação: $Y=f(X)+\epsilon$. Nessa equação, o f(X) simboliza a função a ser mapeada, ao mesmo tempo que ϵ simboliza o erro irredutível. Esse erro é que determina limites intransponíveis à acurácia preditiva dos modelos, mesmo quando a verdadeira função f(X) seja determinada perfeitamente. Determinada tal formulação, o que define o tipo da tarefa de aprendizado é exclusivamente a variável de saída, podendo o problema ser caracterizado como Regressão ou Classificação. Em se tratando de uma resposta Y, com valores quantitativos e contínuos, o problema será definido como Regressão. De outro lado, caso a resposta seja expressa por G, a tarefa caracteriza-se como uma Classificação, na qual a variável assumirá valores qualitativos.

Diante disso, é necessário aprofundar a diferença fundamental entre as duas tarefas. Na Regressão, os valores de saída assumem números contínuos, nos quais valores numericamente próximos representam grandezas ordenadas e semelhantes. Por outro lado, a Classificação atua com valores em um conjunto discreto, de modo que a meta do modelo consiste em calcular as

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

chances de uma observação pertencer a uma classe específica e estabelecer fronteiras que separem as categorias, ao contrário de prever um valor numérico exato.

Portanto, após definidas as distinções sobre os dois principais paradigmas, cabe destacar que o escopo deste trabalho consiste em resolver problemas de Regressão. Desse modo, os algoritmos preditivos e as métricas de avaliação a serem estudados na sequência do trabalho estão ligados à previsão de variáveis de saída quantitativas e contínuas.

2.1. Pré-processamento de Dados

A etapa de pré-processamento de dados é o ponto de partida basilar no ambiente de desenvolvimento de modelos preditivos. Essa importância decorre da realidade dos conjuntos de dados existentes, caracterizados por apresentarem inconsistências e serem suscetíveis a imperfeições em suas estruturas. Segundo Kuhn e Johnson (2013, p. 27), "[...] técnicas de pré-processamento de dados geralmente referem-se à adição, exclusão ou transformação de dados do conjunto de treinamento". Os autores salientam a importância dessa etapa ao reforçarem que "[...] a preparação de dados pode fazer ou desfazer a capacidade preditiva de um modelo" (Kuhn; Johnson, 2013, p. 27).

Nesta etapa inicial, um dos principais desafios consiste na presença de outliers, definidos como valores irregulares que não harmonizam com a distribuição padrão das variáveis. Assim, faz-se necessária a identificação exata desses valores anômalos e o seu devido tratamento. Tal procedimento é indispensável para garantir a estabilidade do modelo diante das anomalias e proporcionar confiança nas previsões, visto que técnicas estatísticas e de regressão são extremamente sensíveis a esses valores discrepantes.

Além dos outliers, a ocorrência de valores ausentes (*missing values*) é outro desafio frequente em bases de dados brutas. Tal desafio exige elaborar uma análise criteriosa para verificar se o padrão de ausência possui correlação com a variável de saída; em caso positivo, configura-se o fenômeno chamado ausência informativa, que pode introduzir vieses no algoritmo. A solução para isso consiste em utilizar técnicas de imputação estatística, nas quais serão usados os dados presentes no conjunto de treinamento para estimar as lacunas, atuando como "um modelo preditivo dentro de um modelo preditivo" (Kuhn; Johnson, 2013). Para isso, destaca-se o uso de algoritmos como o K-Nearest Neighbors, que preenche o dado ausente a partir do cálculo da média das amostras mais similares e espacialmente próximas disponíveis no conjunto de treinamento.

Além do tratamento de anomalias e de lacunas na estrutura dos dados, a fase de pré-processamento exige a transformação de variáveis qualitativas. Visto que os algoritmos

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

computacionais operam apenas com valores numéricos, a presença de variáveis categóricas demanda uma etapa obrigatória de conversão. Segundo Kuhn e Johnson (2013), essas categorias são transformadas em indicadores binários (0 ou 1), denominados *dummy variables*. Porém, tal transformação pode ocasionar problemas numéricos em modelos com intercepto, como a regressão linear, pois a inclusão simultânea de todas as variáveis indicadoras induz à colinearidade perfeita, tornando difícil isolar o efeito individual das variáveis preditoras no algoritmo. Para solucionar essa colinearidade, exclui-se uma das categorias, a qual atuará de forma implícita como base de referência do modelo.

Por fim, para garantir a validade científica do algoritmo, deve-se definir uma divisão rigorosa dos dados antes de iniciar qualquer transformação. O protocolo a seguir consiste em separar a base de dados em duas partes, conjunto de treinamento e de testes. No conjunto de treinamento, encontram-se os dados usados na criação do modelo, enquanto no conjunto de testes estão presentes os dados voltados a medir a sua capacidade preditiva. Visando alcançar a máxima confiabilidade, realizam-se técnicas de reamostragem iterativa, como, por exemplo, a Validação Cruzada (*K-Fold Cross-Validation*).

Durante todo esse processo, é necessário que seja evitado o vazamento de dados (*data leakage*). Tal problema ocorre quando são aplicadas técnicas de pré-processamento, como as mencionadas anteriormente, sobre a base completa antes da divisão dos dados. Com isso, a fim de evitar a geração de previsões artificialmente otimistas, Kuhn e Johnson (2013, p. 42) afirmam que “[...] a imputação deve ser incorporada dentro da reamostragem. Isso aumentará o tempo computacional para a construção de modelos, mas também fornecerá estimativas honestas do desempenho do modelo”. Tal condição garante que os dados de teste fiquem completamente isolados durante a fase de aprendizagem do modelo.

2.2. Normalização e padronização

A etapa de escalonamento de atributos (*feature scaling*) é a fase em que se realiza o tratamento da dimensão das variáveis, essencial para assegurar a eficiência do modelo preditivo. Com algumas exceções, a maioria dos modelos preditivos sofrem uma perda considerável de desempenho em virtude do enviesamento gerado por escalas discrepantes evidenciadas nas variáveis numéricas de entrada. Conforme Géron (2025), encontramos tal fragilidade quando nos deparamos com valores discrepantes, como, por exemplo, um atributo que se encontre entre 6 a 39.320 e outro de 0 a 10, em que o algoritmo confere maior importância aos valores maiores, acabando por discriminar os valores menores e ofuscar suas características, levando os algoritmos a um enviesamento matemático.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Este problema fica nítido em algoritmos de otimização como o Gradiente Descendente (*Gradient Descent*), visto que a disparidade das escalas confunde o algoritmo, tornando a predição mais demorada e mais cara em razão das oscilações durante o treinamento. O escalonamento vai permitir que o modelo alcance o menor erro de forma mais ágil e eficaz, e mesmo que modelos como as Árvores de Decisão (*Decision Trees*) não necessitem do ajuste de escalas por serem baseados em regras simples de decisão, o escalonamento por proporcionar o nivelamento de informações sempre será uma etapa fundamental para a grande maioria dos modelos de aprendizagem.

Na prática da etapa de escalonamento de atributos, a escolha metodológica concentra-se entre os métodos de Padronização e Normalização. A padronização (z-score), expressa pela equação $x' = (x - \mu) / \sigma$, garante que a distribuição dos valores apresente média 0 e desvio padrão igual a 1. Já a normalização (Min-Max), obtida pela equação $x' = (x - x_{min}) / (x_{max} - x_{min})$, atua comprimindo todos os valores dentro da faixa fixa [0, 1]. Por isso, deve ser analisado o uso caso a caso das duas técnicas. Algoritmos preditivos como a Regressão Linear e o *Gradient Boosting* beneficiam-se da estabilidade conferida pela Padronização quando se deparam com dados ruidosos. De outro lado, a Normalização atende melhor algoritmos como os das Redes Neurais (MLP), que atingem melhores resultados ao receberem entradas em faixas preestabelecidas.

Por fim, independentemente do método a ser utilizado, é imprescindível realizar o isolamento rigoroso das partições para evitar a contaminação cruzada das informações, fenômeno definido como data leakage (vazamento de dados). Segundo Géron (2025), as extrações das médias, assim como valores máximos e mínimos usados para calibrar o escalonador, devem acontecer exclusivamente nos dados de treinamento. Se os dados de teste forem expostos à padronização ou normalização, o algoritmo acabará decorando o padrão da base completa, retornando medidas de validação artificialmente otimistas. Para garantir a efetividade e evitar esse vazamento, deve-se adotar um protocolo seguro em que o modelo aprende as escalas apenas com os dados disponíveis no conjunto de treinamento, garantindo que as estimativas sejam honestas e assertivas quanto à verdadeira performance do modelo perante novos dados.

2.3. Algoritmos de Predição

Na criação de sistemas preditivos, partimos da origem de que não há um algoritmo superior que se encaixe em qualquer conjunto de dados, o que requer uma análise criteriosa para a escolha. Tal princípio é fundamentado no teorema do *No Free Lunch* (Wolpert, 1996),

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

segundo o qual, na ausência de suposições prévias sobre a natureza dos dados, não se deve presumir a supremacia universal de um algoritmo perante os outros (Géron, 2025). Assim, como pressupostos básicos para a construção do modelo, deve-se primeiramente definir um equilíbrio entre transparência e precisão. Modelos menos complexos tornam suas decisões mais claras, aumentando a interpretabilidade dos resultados, de outra maneira, os modelos mais robustos, ao custo de uma interpretabilidade limitada, possuem maior precisão no mapeamento de padrões ocultos nos dados.

2.3.1. Regressão Linear

A Regressão Linear é um dos métodos preditivos mais consolidados no campo da estatística e da aprendizagem de máquina. Seu princípio fundamental consiste em modelar a relação entre uma variável dependente contínua Y e um conjunto de variáveis independentes X_1, X_2, ..., X_p por meio de uma função aditiva linear, expressa pela equação: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_pX_p + \varepsilon$

Nessa formulação, os termos β_0, β_1, ..., β_p representam os coeficientes do modelo (sendo β_0 o intercepto e os demais os pesos atribuídos a cada preditor), enquanto ε denota o erro irredutível, isto é, a parcela da variação em Y que não é explicada pelas variáveis de entrada (Hastie; Tibshirani; Friedman, 2009).

A estimação dos coeficientes se dá pelo Método dos Mínimos Quadrados Ordinários (MQO), que busca minimizar a soma dos Quadrados dos Resíduos (RSS - Residual Sum of Squares). O MQO possui solução analítica fechada, o que confere ao modelo alta eficiência computacional. Além disso, cada coeficiente β_j possui interpretação direta: representa a variação esperada em Y para um acréscimo unitário em X_j, mantendo os demais preditores constantes. Essa transparência torna a Regressão Linear um modelo altamente interpretável, o que justifica sua ampla adoção em contextos em que a aplicabilidade das previsões é prioritária (Géron, 2022).

Contudo, a Regressão Linear impõe pressupostos restritivos que podem comprometer seu desempenho em cenários práticos. A suposição de linearidade entre preditores e resposta impede que o modelo capture relações curvilíneas ou interações complexas entre variáveis. Além disso, o estimador MQO é sensível à presença de outliers, pois esses valores extremos exercem influência desproporcional sobre os coeficientes estimados, deslocando a reta de predição de forma a prejudicar as previsões para o conjunto majoritário dos dados. Tais limitações fundamentam a necessidade de explorar métodos não lineares com maior capacidade de generalização.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2.3.2. Random Forest

Random Forest é um algoritmo do tipo ensemble (categoria de métodos que combinam múltiplos modelos preditivos para alcançar desempenho superior ao de qualquer estimador individual). O princípio por trás dos métodos de ensemble está fundamentado na ideia de que um conjunto diversificado de modelos, mesmo que individualmente imperfeitos, produz previsões coletivas mais robustas e precisas (Breiman, 2001).

A base do Random Forest é o método de Bagging (*Bootstrap Aggregating*), introduzido por Breiman (1996 apud Hastie; Tibshirani; Friedman, 2009), que visa reduzir a variância do modelo sem aumentar seu viés. O procedimento consiste em: (1) gerar B subconjuntos de treinamento por reamostragem com reposição (bootstrap) a partir do conjunto original; (2) treinar uma árvore de decisão independente sobre cada subconjunto; e (3) agregar as previsões das B árvores, calculando a média aritmética no caso de regressão. Como cada árvore é construída sobre uma amostra distinta, os erros individuais tendem a ser não correlacionados entre si, e sua média converge para um estimador com variância significativamente reduzida.

O Random Forest acrescenta ao *Bagging* uma segunda fonte de aleatoriedade: em cada nó de divisão de cada árvore, apenas um subconjunto aleatório de m atributos (com $m < p$, sendo p o número total de preditores) é considerado para a escolha do melhor critério de partição. Essa estratégia reduz a correlação entre as árvores e aumenta a diversidade do conjunto, tornando o modelo mais robusto diante de outliers e ruído nos dados. O teorema central do Random Forest demonstra que o erro de generalização converge para um limite finito à medida que o número de árvores cresce, garantindo que o modelo não sofra o *overfitting* com o aumento de B (Breiman, 2001).

2.3.3. Gradient Boosting

Enquanto o Random Forest reduz a variância por meio de paralelismo e aleatoriedade, o *Gradient Boosting* atua sobre o viés do modelo a partir de uma estratégia sequencial e aditiva. O método, formalizado por Friedman (2001), pertence à família dos algoritmos de Boosting, nos quais cada novo estimador é construído com o objetivo explícito de corrigir os erros cometidos pelo estimador anterior.

O algoritmo opera por meio de otimização numérica no espaço funcional. A cada iteração m, calcula-se o gradiente negativo da função de perda em relação às previsões acumuladas (no caso da regressão com erro quadrático, esse gradiente corresponde diretamente aos resíduos $r_i = y_i - F_{m-1}(x_i)$). Uma árvore de decisão rasa é então ajustada sobre esses resíduos, e sua

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

contribuição é incorporada ao modelo por meio de uma taxa de aprendizado η (learning rate), que controla o passo de atualização e atua como mecanismo de regularização: F_m(x) = F_{m-1}(x) + η · h_m(x).

Esse processo iterativo faz com que o modelo concentre progressivamente sua capacidade de aprendizagem nas regiões de maior erro, reduzindo o viés de forma gradual e controlada. A profundidade reduzida das árvores individuais (estimadores fracos) garante que cada componente capture apenas padrões locais, enquanto o conjunto da expansão aditiva captura estruturas de alta complexidade (Friedman, 2001; Hastie; Tibshirani; Friedman, 2009).

A principal vantagem do *Gradient Boosting* reside em sua capacidade de modelar relações altamente não lineares com grande precisão preditiva. Como contrapartida, o modelo é mais suscetível a *overfitting* do que o Random Forest quando a taxa de aprendizado ou o número de iterações não são devidamente ajustados. Técnicas como subamostragem estocástica *(stochastic gradient boosting),* penalização da complexidade das árvores e parada antecipada *(early stopping)* são frequentemente empregadas para mitigar esse risco (Géron, 2022).

2.3.4. Redes Neurais - Perceptron Multicamadas (MLP)

O Perceptron Multicamadas (MLP - *Multilayer Perceptron)* é uma arquitetura de rede neural artificial composta por múltiplas camadas de neurônios artificiais totalmente conectadas *(fully connected).* A estrutura básica de um MLP compreende três tipos de camadas: (1) a camada de entrada, responsável por receber as variáveis preditoras; (2) uma ou mais camadas ocultas *(hidden layers),* nas quais ocorre a transformação não linear dos dados; e (3) a camada de saída, que produz a previsão final. A profundidade e a largura da rede (número de camadas e de neurônios por camada, respectivamente) são hiperparâmetros que definem a capacidade expressiva do modelo (Géron, 2022).

O elemento central que confere poder representacional ao MLP são as funções de ativação não lineares, aplicadas na saída de cada neurônio das camadas ocultas. Funções como a Sigmoide, a Tangente Hiperbólica (tanh) e, especialmente, a Unidade Linear Retificada (ReLU - Rectified Linear Unit, definida por f(z) = max(0, z)) permitem que a rede aprenda representações hierárquicas de crescente complexidade, superando a limitação das transformações puramente lineares. Sem tais não linearidades, uma rede com múltiplas camadas seria matematicamente equivalente a uma simples regressão linear, independentemente de sua profundidade (Hastie; Tibshirani; Friedman, 2009).

O treinamento do MLP é conduzido pelo algoritmo de *Backpropagation* (retropropagação do erro), formalizado por Rumelhart, Hinton e Williams (1986). O processo ocorre em duas

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

etapas: na propagação direta *(forward pass),* os dados percorrem a rede camada a camada até gerar uma previsão; na retropropagação *(backward pass),* o erro entre a previsão e o valor real é calculado e propagado de volta pela rede por meio da regra da cadeia do cálculo diferencial, obtendo-se o gradiente da função de perda em relação a cada peso. Esses gradientes são então utilizados por um algoritmo de otimização (tipicamente o Gradiente Descendente Estocástico (SGD) ou variantes adaptativas como Adam e RMSProp) para atualizar os pesos iterativamente, minimizando a função de perda.

A capacidade do MLP de aproximar funções arbitrariamente complexas, garantida pelo Teorema da Aproximação Universal, confere ao modelo um potencial preditivo que supera os demais algoritmos estudados em problemas de alta dimensionalidade e com padrões latentes sofisticados. Em contrapartida, esse poder vem acompanhado de maior exigência computacional, maior número de hiperparâmetros a ajustar e maior sensibilidade à escala das variáveis de entrada, reforçando a importância das etapas de pré-processamento e normalização discutidas anteriormente (Géron, 2025).

3. METODOLOGIA

Este capítulo apresenta a metodologia utilizada no desenvolvimento do sistema preditivo de faturamento comercial, abrangendo a caracterização dos dados, as etapas de pré-processamento e a implementação e validação dos modelos de aprendizado de máquina empregados no estudo.

3.1. Conjunto de dados

O conjunto de dados utilizado neste trabalho é composto por informações reais empregadas para fins de análise preditiva, abrangendo mais de 3.000 pontos de venda monitorados ao longo dos últimos cinco anos e consolidados com indicadores socioeconômicos de seus respectivos municípios. Durante a etapa de preparação dos dados, foi realizada uma análise exploratória com o objetivo de identificar inconsistências, atributos irrelevantes e registros incompletos ou redundantes. Após esse processo, foram removidas instâncias consideradas inadequadas para a modelagem, além da exclusão de atributos que apresentavam baixa relevância para a predição.

Ao final do tratamento, a base de dados consolidada, contendo os valores de venda mensal de cada ponto de venda, passou a contar com 11.833 instâncias e 12 atributos. Essa redução contribuiu para a obtenção de um conjunto de dados mais consistente e adequado à

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

aplicação de algoritmos de aprendizado de máquina, reduzindo ruídos e melhorando a eficiência computacional durante o treinamento dos modelos preditivos.

3.2. Pré-processamento Aplicado

Nessa etapa, foram aplicadas técnicas de pré-processamento com o objetivo de preparar os dados para utilização nos modelos de predição, garantindo maior qualidade, consistência e confiabilidade das informações analisadas.

Inicialmente, foi realizada a limpeza dos dados, removendo registros duplicados, inconsistentes ou incompletos que poderiam comprometer os resultados dos algoritmos. Em seguida, as instâncias foram agrupadas por período de tempo (ano), buscando possibilitar uma análise temporal dos dados e a identificação de padrões ao longo do tempo.

Além do agrupamento das instâncias, foram removidos valores discrepantes (*outliers*), reduzindo a influência de dados anômalos no treinamento dos modelos preditivos. Durante o processo, foram utilizados atributos numéricos e categóricos, permitindo representar tanto características quantitativas quanto qualitativas presentes no conjunto de dados.

Também foi realizada a seleção dos atributos mais relevantes para o problema estudado, eliminando variáveis redundantes ou com baixa contribuição para o processo de predição. Por fim, aplicou-se o ajuste de escala dos atributos numéricos, com o objetivo de padronizar os valores e melhorar o desempenho dos algoritmos de aprendizado de máquina.

Essas etapas de pré-processamento foram fundamentais para tornar o conjunto de dados mais adequado às análises preditivas e contribuir para a obtenção de resultados mais precisos e confiáveis.

3.3. Implementação dos algoritmos

Neste artigo, foram aplicados os algoritmos de Regressão Linear, Random Forest, *Gradient Boosting* e Rede Neural MLP *(Multilayer Perceptron),* implementados através da linguagem de programação Python, devido à sua ampla aplicação em projetos de ciência de dados e aprendizado de máquina. Para o desenvolvimento dos modelos, foram utilizadas bibliotecas específicas para manipulação de dados, treinamento e avaliação dos algoritmos. A escolha desses algoritmos visa cobrir um espectro representativo de abordagens de aprendizado supervisionado para regressão: de modelos paramétricos lineares a métodos de ensemble e redes neurais, conforme fundamentado na Seção 2. Os hiperparâmetros foram mantidos em seus valores padrão para garantir comparabilidade entre os modelos, sem viés de ajuste individual.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A Regressão Linear foi utilizada como modelo estatístico para identificar relações entre variáveis independentes e a variável alvo, servindo como uma abordagem inicial de modelagem preditiva. O algoritmo Random Forest foi empregado na abordagem de regressão por sua capacidade de construir múltiplas árvores de decisão, aumentando a robustez do modelo e reduzindo o *overfitting*.

O algoritmo *Gradient Boosting* foi utilizado devido à sua eficiência em tarefas preditivas, combinando diversos modelos fracos para formar um modelo mais preciso e otimizado. Já as Redes Neurais MLP foram aplicadas por sua capacidade de aprendizado não linear, permitindo identificar padrões complexos presentes no conjunto de dados.

Os modelos foram treinados e avaliados utilizando os dados previamente processados. Durante os experimentos, foram realizadas etapas de treinamento, validação e teste, visando comparar o desempenho dos algoritmos por meio de métricas de avaliação adequadas ao problema estudado. Dessa forma, foi possível analisar a eficiência de cada técnica na tarefa de predição proposta.

3.4. Avaliação dos Modelos

Para a avaliação dos modelos preditivos, o conjunto de dados históricos foi submetido a um protocolo de isolamento de partições para simular a predição de faturamento de um novo estabelecimento comercial. Esse fluxo foi estruturado por meio de um pré-processamento, encarregado de aplicar as técnicas de escalonamento de atributos de forma estritamente isolada nos subconjuntos de treinamento, impedindo o vazamento de dados (*data leakage*) para os subconjuntos de teste, os quais foram empregados exclusivamente para validar o desempenho final dos modelos em dados inéditos.

Além disso, o desenho metodológico empregou a técnica de validação cruzada (K-Fold Cross Validation) com 5 partições ótimas (*folds*), realizando múltiplos ciclos síncronos de treinamento e teste para garantir que todas as instâncias participassem dos dois processos em etapas distintas, permitindo uma mensuração estável do R^2 e do RMSE generalizáveis. Para o algoritmo de *Random Forest*, a avaliação foi complementada com a validação nativa baseada no escore *Out-of-Bag* (OOB), que aproveita as amostras deixadas fora do treinamento de cada árvore individual durante o processo de amostragem por reposição. Com isso, o modelo obteve uma estimativa de erro e de capacidade de generalização estatisticamente blindada contra o superajuste (*overfitting*), otimizando o tempo de resposta computacional sem renunciar ao rigor científico.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Como métricas de avaliação e validação, foi utilizado o Coeficiente de Determinação R^2, que mede a proporção da variância da variável resposta explicada pelo modelo (em que $R^2 = 1$ representa o ajuste perfeito e $R^2 \leq 0$ indica um desempenho estritamente inferior ao uso da média histórica), calculado de forma generalizável por meio de validação cruzada K-Fold e pela métrica nativa Out-of-Bag (R^2 OOB). Complementarmente, empregou-se a Raiz do Erro Quadrático Médio (RMSE - *Root Mean Squared Error*), computada sobre as predições de teste do K-Fold e sobre as predições acumuladas Out-of-Bag, a qual calcula a raiz quadrada da média dos quadrados das diferenças entre os faturamentos preditos e os reais. Essa abordagem permitiu penalizar erros de grande magnitude de forma mais contundente e expressar a margem média de desvio diretamente na unidade original dos dados (em Reais - R$), servindo de base para a determinação do intervalo de confiança das projeções.

4. RESULTADOS E DISCUSSÃO

Foram aplicados quatro modelos de regressão com o objetivo de avaliar o desempenho preditivo sobre o conjunto de dados: Regressão Linear, Random Forest Regressor, *Gradient Boosting* Regressor e Rede Neural MLP.

Os modelos foram avaliados utilizando o coeficiente de determinação (R^2), o valor de venda previsto para um novo ponto de venda em um determinado município e a raiz do erro quadrático médio (RMSE), permitindo uma análise comparativa entre diferentes abordagens de aprendizado de máquina somente com o algoritmo, com normalização e com o isolamento correto das partições de treinamento e teste, o emprego de validação cruzada K-Fold e as técnicas de escalonamento. Os resultados obtidos são apresentados na Tabela 1.

Tabela 1. Desempenho dos modelos

Modelo	R^2	Valor Previsto	RMSE
Regressão Linear	0,2169	-1457,43	2633,35
Regressão Linear com Normalização	0,2169	-1457,43	2633,35
Regressão Linear com Normalização e K-Fold	0,1613	-1457,43	2677,87
Gradient Boosting Regressor	0,7363	759,08	1528,01
Gradient Boosting Regressor com Normalização	0,7363	759,08	1528,01
Gradient Boosting Regressor com Normalização e K-Fold	0,4835	759,08	1959,65
MLPRegressor (Multi-Layer Perceptron – rede neural)	-0,1479	1355,10	3188,16
MLPRegressor com Normalização	0,7094	405,04	1604,25

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

MLPRegressor com Normalização e K-Fold	0,4705	405,04	1975,74
Random Forest	0,9624	894,33	577,36
Random Forest com Normalização	0,9627	906,35	574,65
Random Forest com Normalização e K-Fold	0,6181	906,35	1737,18
Random Forest com Normalização e Out-of-Bag Score	0,7103	906,35	1601,03

Fonte: Os autores, 2026.

Os resultados indicam que os modelos não lineares apresentaram desempenho superior em relação à Regressão Linear, evidenciando a existência de relações complexas entre as variáveis do conjunto de dados. A comparação dos resultados confirma, na prática, as características teóricas dos algoritmos. Os modelos de Regressão Linear, *Gradient Boosting* e Random Forest mostraram ser pouco afetados pela normalização dos dados. Isso ocorre porque a Regressão Linear ajusta seus coeficientes ao novo escalonamento dos dados, enquanto os algoritmos baseados em árvores utilizam regras de divisão que dependem apenas da ordem dos valores. Assim, a normalização altera apenas os valores numéricos dos limites de corte, sem modificar a estrutura das divisões dos dados de faturamento.

A Regressão Linear apresentou o pior desempenho ($R^2 \approx 0,16$ a $0,21$), indicando baixa capacidade de captura das relações presentes nos dados. Esse resultado sugere que a relação entre variáveis não pode ser adequadamente representada por um modelo linear simples.

O *Gradient Boosting* Regressor apresentou desempenho intermediário ($R^2 \approx 0,48$), demonstrando capacidade de modelar padrões não lineares por meio da estratégia sequencial e aditiva de redução do viés proposta por Friedman (2001). Verificou-se uma diferença de aproximadamente 13 pontos percentuais em relação ao Random Forest. Essa diferença indica que, com os hiperparâmetros padrão, o *Gradient Boosting* não atingiu seu potencial preditivo máximo. Esse comportamento é esperado, pois o algoritmo é particularmente sensível ao ajuste da taxa de aprendizado (η) e à profundidade das árvores individuais (parâmetros que não foram otimizados neste estudo). A literatura aponta que, com otimização adequada (ex.: *Grid Search* ou *Bayesian Optimization),* o *Gradient Boosting* frequentemente iguala ou supera o Random Forest em tarefas de regressão (Hastie; Tibshirani; Friedman, 2009).

O Random Forest Regressor apresentou desempenho superior ($R^2 \approx 0,61$ a $0,71$), destacando-se pela robustez e estabilidade, além de lidar bem com variáveis complexas e interações não lineares. Esse resultado é explicado pelo mecanismo de correlação entre árvores. Ao selecionar aleatoriamente $m < p$ atributos em cada nó, o ensemble gera estimadores independentes cujos erros tendem a se cancelar na média, reduzindo a variância sem aumentar o viés (Breiman, 2001). Para dados tabulares de dimensão moderada, como o conjunto utilizado

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

neste estudo (11.833 instâncias, 12 atributos), o Random Forest representa uma escolha robusta e de baixo custo de ajuste.

A Rede Neural MLP mostrou forte dependência da normalização dos dados, evidenciando que o balanceamento das escalas é essencial para seu adequado funcionamento. Com a normalização, seu desempenho melhorou significativamente, porém permaneceu inferior ao do Random Forest. Esse resultado sugere que, nas condições deste experimento, a arquitetura utilizada não foi capaz de superar os modelos baseados em árvores, mesmo contando com camadas ocultas e funções de ativação não lineares e com a normalização dos dados, etapa fundamental para a convergência do algoritmo de retropropagação (*backpropagation*) (Géron, 2025). Contudo, a diferença em relação ao *Gradient Boosting* foi pequena, indicando que ambos os modelos apresentam capacidade preditiva semelhante. Essa proximidade reforça que, em conjuntos de dados tabulares de dimensão moderada, métodos de *ensemble* baseados em árvores podem alcançar desempenho comparável ou superior ao de redes neurais, com menor custo computacional e menor necessidade de ajuste de hiperparâmetros.

O principal resultado da análise está relacionado ao impacto da validação cruzada K-Fold na avaliação da capacidade de generalização dos modelos. Nos testes sem K-Fold, o Random Forest e o *Gradient Boosting* apresentaram desempenhos elevados, com destaque para o *Random Forest*, que alcançou um coeficiente de determinação R^2 de 0,9624 e RMSE de 577,36. Entretanto, ao utilizar a validação cruzada com 5 *folds*, na qual os modelos são avaliados em subconjuntos não utilizados durante o treinamento, observou-se uma redução significativa do desempenho. O R^2 do *Random Forest* caiu para 0,6181, enquanto o do *Gradient Boosting* reduziu-se para 0,4835, acompanhado de aumento nos erros médios de predição. Esses resultados evidenciam que parte do desempenho observado inicialmente estava associada ao ajuste excessivo aos dados de treinamento (*overfitting*), enquanto a validação cruzada forneceu uma estimativa mais realista da capacidade de generalização dos modelos.

Adicionalmente, a utilização da validação Out-of-Bag (OOB) no Random Forest com normalização apresentou resultados consistentes, alcançando $R^2 = 0,71$ e RMSE = 1601,03. Como o método OOB utiliza amostras não selecionadas durante a construção de cada árvore para realizar a validação interna, ele oferece uma estimativa eficiente do erro de generalização sem a necessidade de múltiplos ciclos de treinamento exigidos pelo K-Fold.

Considerando os resultados obtidos, o Random Forest destacou-se como o modelo mais estável e confiável para o problema analisado. Tanto na avaliação por K-Fold quanto por OOB, apresentou maior capacidade de explicação da variabilidade do faturamento (R^2 entre 0,61 e

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

0,71) e menores erros médios de predição quando comparado aos demais modelos. Dessa forma, o algoritmo fornece estimativas mais robustas para apoiar a tomada de decisão.

5. CONSIDERAÇÕES FINAIS

Este trabalho realizou uma análise comparativa entre quatro algoritmos de predição aplicados a um conjunto de dados tabulares reais: Regressão Linear, Random Forest, *Gradient Boosting* e Rede Neural MLP. Os experimentos foram conduzidos sob um protocolo metodológico que contemplou pré-processamento adequado, isolamento das partições de treinamento e teste, validação cruzada K-Fold e escalonamento de atributos diferenciado por modelo, garantindo a validade e a comparabilidade dos resultados obtidos.

Os resultados evidenciaram que a Regressão Linear apresentou desempenho significativamente inferior aos demais modelos ($R^2 \approx 0,16$), confirmando que as relações presentes no conjunto de dados extrapolam a capacidade representacional de um modelo linear simples. O *Gradient Boosting* Regressor e a Rede Neural MLP obtiveram desempenho intermediário ($R^2 \approx 0,48$ e $0,47$), demonstrando suas capacidades de capturar padrões não lineares por meio da estratégia sequencial e aditiva (Friedman, 2001). O Random Forest alcançou os melhores resultados, com $R^2 \approx 0,71$. Esse desempenho superior dos modelos não lineares corrobora o pressuposto teórico de que relações complexas entre variáveis exigem algoritmos com maior capacidade expressiva (Hastie; Tibshirani; Friedman, 2009).

Do ponto de vista prático, os resultados oferecem subsídios empíricos relevantes para a escolha algorítmica em problemas de regressão sobre dados tabulares. Em cenários em que a interpretabilidade é prioritária e a relação entre as variáveis é aproximadamente linear, a Regressão Linear permanece uma escolha adequada pela sua transparência e eficiência computacional. Para contextos que demandam robustez e tolerância a ruído, o Random Forest apresenta-se como uma alternativa confiável, ao passo que o *Gradient Boosting* é indicado quando o ajuste fino das previsões é imperativo e há recursos para a calibração adequada dos hiperparâmetros. A Rede Neural MLP, por sua vez, destaca-se quando a complexidade dos padrões latentes no conjunto de dados justifica o custo computacional e o esforço de normalização e otimização necessários ao seu treinamento, em consonância com o observado por Géron (2025) e Rumelhart, Hinton e Williams (1986).

Os resultados demonstram que não existe um algoritmo de aprendizado de máquina universalmente superior para a predição. O desempenho de cada modelo depende das características dos dados e do tratamento aplicado às variáveis. Modelos lineares e redes neurais são mais sensíveis à escala dos atributos, enquanto métodos baseados em árvores

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

apresentam maior robustez, mas exigem técnicas de validação adequadas para evitar o superajuste. Dessa forma, a escolha do modelo mais apropriado deve ser baseada em testes comparativos e processos de validação consistentes, garantindo aderência às características do conjunto de dados analisado.

Como limitações do presente estudo, destacam-se a restrição a um único conjunto de dados e a ausência de otimização sistemática de hiperparâmetros por meio de técnicas como *Grid Search* ou *Bayesian Optimization*, o que pode ter subestimado o potencial preditivo de alguns dos modelos avaliados, em especial o *Gradient Boosting* e a MLP. Para trabalhos futuros, propõe-se a extensão da análise a múltiplos conjuntos de dados de domínios distintos, a incorporação de estratégias de seleção e engenharia de atributos, bem como a investigação de variantes modernas dos algoritmos estudados, tais como XGBoost, LightGBM e arquiteturas de aprendizado profundo, visando ampliar a abrangência e a generalidade das conclusões aqui apresentadas.

REFERÊNCIAS

BREIMAN, Leo. Random forests. Machine Learning, New York, v. 45, n. 1, p. 5-32, out. 2001. DOI: 10.1023/A:1010933404324.

BREIMAN, Leo. Bagging predictors. Machine Learning, New York, v. 24, n. 2, p. 123-140, ago. 1996. DOI: 10.1007/BF00058655.

FRIEDMAN, Jerome H. Greedy function approximation: a gradient boosting machine. The Annals of Statistics, Beachwood, v. 29, n. 5, p. 1189-1232, out. 2001. DOI: 10.1214/aos/1013203451.

GÉRON, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 3. ed. Sebastopol: O'Reilly Media, 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2. ed. New York: Springer, 2009.

KUHN, Max; JOHNSON, Kjell. Applied Predictive Modeling. New York: Springer, 2013.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. Nature, London, v. 323, n. 6088, p. 533-536, out. 1986. DOI: 10.1038/323533a0.

WOLPERT, David H. The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation, Cambridge, v. 8, n. 7, p. 1341-1390, 1996.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.