

Análise Comparativa de Algoritmos de Aprendizado de Máquina para Triagem de Vítimas Baseada em Sinais Vitais

Renzo Tognella De Rosa¹, Miguel Dos Anjos Brack¹

¹Universidade Tecnológica Federal do Paraná (UTFPR) - Campus Curitiba
Av. Sete de Setembro, 3165 - Rebouças, Curitiba - PR, 80230-901

renzotognella@alunos.utfpr.edu.br, miguelanjosbrack@alunos.utfpr.edu.br

Abstract. *This study comparatively evaluates Decision Tree (CART), Random Forest (RF), and Multilayer Perceptron (MLP) models for predicting victim severity scores and classifying health status using vital signs (s_{i3} , s_{i4} , s_{i5}) in emergency triage. Models underwent hyperparameter adjustment and were assessed on regression and classification tasks. Results suggest RF and MLP generally offer enhanced predictive performance over single CART models, highlighting the relevance of all vital signs. This research underscores machine learning's potential in critical care decision support, with specific author contributions detailed.*

Resumo. *Este estudo avalia comparativamente Árvores de Decisão (CART), Random Forest (RF) e Perceptron de Múltiplas Camadas (MLP) na predição de escores de gravidade e classificação do estado de saúde de vítimas, utilizando sinais vitais (s_{i3} , s_{i4} , s_{i5}) em cenários de triagem de emergência. Os modelos passaram por ajuste de hiperparâmetros e foram avaliados em tarefas de regressão e classificação. Os resultados indicam que RF e MLP geralmente apresentam desempenho preditivo superior ao CART, ressaltando a relevância dos sinais vitais. A pesquisa sublinha o potencial da aprendizagem de máquina no apoio à decisão em cuidados críticos.*

Palavras chave: Aprendizado de Máquina, Árvores de Decisão, CART, Random Forest, Redes Neurais MLP, Regressão, Classificação, Sinais Vitais, Ajuste de Hiperparâmetros, SHAP, Curvas de Aprendizado, Importância de Features, Matriz de Confusão.

1. Introdução

Este artigo apresentará a fundamentação teórica sobre algoritmos de aprendizagem de máquina [Russell and Norvig 2021], com foco nas modelagens empregadas para abordar o problema do cálculo do score de gravidade e da classificação do estado de saúde de vítimas em cenários de desastres, grandes acidentes e catástrofes naturais. A contribuição central deste estudo consiste na análise comparativa e na avaliação do impacto de hiperparâmetros nos modelos de Árvore de Decisão [Breiman et al. 1984], Random Forest [Breiman 2001] e Redes Neurais MLP [Goodfellow et al. 2016], especificamente aplicados à tarefa crítica de triagem de vítimas em contextos de emergência. Será detalhada a metodologia experimental, que inclui a análise do impacto de hiperparâmetros no desempenho dos modelos mencionados (detalhada na Seção 3.1). Finalmente, serão discutidos os resultados obtidos, suas implicações práticas para o resgate de vítimas e as conclusões do estudo.

1.1. Descrição do Conjunto de Dados Histórico

Para contextualizar, vamos explicar sobre o conjunto de dados que foi utilizado na modelagem do problema. Dentro do treinamento utilizamos dois arquivos que continham as seguintes colunas:

- **i (Identificação da Vítima):** Esta coluna é utilizada para indexação e rastreamento dos dados.
- **s_{i1} (Pressão Sistólica - pSist) e s_{i2} (Pressão Diastólica - pDiast):** s_{i1} representa a pressão sistólica, com valores no intervalo [5, 22], e s_{i2} a pressão diastólica, no intervalo [0, 15]. A informação contida nelas é indiretamente representada pela coluna s_{i3} (qPA), que é calculada a partir da relação entre a pressão sistólica e diastólica. Conforme as instruções do problema, s_{i1} e s_{i2} não foram diretamente utilizadas como features nos modelos.
- **s_{i3} (Qualidade da Pressão Arterial - qPA):** Esta variável quantifica a qualidade da pressão arterial, sendo derivada da avaliação da relação entre a pressão sistólica e a diastólica. Seus valores variam no intervalo [-10, 10], onde 0 indica a qualidade máxima. Valores de -10 representam a pior qualidade devido à pressão excessivamente baixa, enquanto +10 indicam a pior qualidade por pressão excessivamente alta. Esta é uma das features de entrada para os modelos.
- **s_{i4} (Pulso):** Refere-se à pulsação da vítima, medida em batimentos por minuto (BPM), com valores no intervalo [0, 200]. O pulso é um sinal vital clássico que reflete a resposta do sistema cardiovascular a traumas ou estresse fisiológico. Esta é uma das features de entrada para os modelos.
- **s_{i5} (Respiração - Frequência Respiratória):** Mede a frequência da respiração em respirações por minuto (FpM), contida no intervalo [0, 22]. É também referida como frequência respiratória. Esta é uma das features de entrada para os modelos.
- **g_i (Gravidade):** Um valor numérico real, previamente calculado pelo corpo médico em função dos sinais vitais s_{i3} , s_{i4} , e s_{i5} . Esta coluna representa a **variável alvo (target) para a tarefa de regressão**.
- **y_i (Classe de Saída):** Um rótulo categórico que representa o estado de saúde da vítima. São 4 classes (1: crítico, 2: instável, 3: potencialmente estável, 4: estável), definidas por intervalos da variável gravidade (g_i). Esta é a **variável alvo (target) para a tarefa de classificação**.

Compreender o significado e a função de cada uma dessas colunas é essencial para o desenvolvimento e a interpretação dos modelos de Aprendizagem de Máquina que buscam replicar a avaliação médica perdida. Os modelos utilizarão s_{i3} (qPA), s_{i4} (pulso) e s_{i5} (respiração) como entradas para, na regressão, prever g_i (gravidade) e, na classificação, prever y_i (classe de gravidade).

2. Metodologia

Nesta seção, detalhamos os algoritmos de aprendizado de máquina empregados neste estudo, com foco em suas características conceituais, e descrevemos o processo experimental geral. A tarefa consistia em duas frentes principais: (1) reconstituir o cálculo do valor da gravidade (g_i) por meio de regressão e (2) realizar a classificação das vítimas em quatro categorias de estado de saúde (y_i). Os dados de treinamento, como os arquivos textuais contendo os atributos s_{i3} (qualidade da pressão arterial), s_{i4} (pulso), s_{i5} (frequência respiratória) e os respectivos alvos g_i e y_i , foram utilizados como base para a modelagem.

2.1. Pré-processamento dos Dados

Antes de ajustar os modelos, aplicou-se um fluxo de pré-processamento em quatro etapas principais. Primeiro, os dados brutos foram carregados e as colunas renomeadas conforme a especificação do enunciado. Em seguida, tratou-se valores ausentes: entradas não numéricas ou faltantes foram convertidas para NaN e imputadas pela mediana de cada variável, de modo

a preservar a distribuição original sem introduzir viés. Depois, identificaram-se outliers via método do intervalo interquartil (IQR), limitando valores abaixo de $Q_1 - 1,5 \times \text{IQR}$ e acima de $Q_3 + 1,5 \times \text{IQR}$ aos respectivos limites para mitigar o efeito de valores extremos. Por fim, aplicou-se normalização z-score (média zero e desvio-padrão um) usando apenas estatísticas do conjunto de treino, garantindo independência e evitando vazamento de informação para validação/teste. Esses procedimentos asseguraram que os modelos recebessem entradas consistentes, livres de lacunas e de influências indevidas de outliers.

2.2. Árvores de Decisão (CART)

O algoritmo CART (Classification and Regression Trees) [Breiman et al. 1984] é um método de aprendizado supervisionado não paramétrico que constrói modelos de decisão com uma estrutura hierárquica semelhante a uma árvore. Este algoritmo foi selecionado por sua capacidade de tratar tanto tarefas de regressão quanto de classificação, sua relativa interpretabilidade e sua eficiência em identificar relações não lineares nos dados.

2.2.1. Estrutura e Funcionamento no CART

Uma árvore CART inicia-se com um **Nó Raiz**, representando o conjunto completo de dados. Este nó é dividido recursivamente em subconjuntos mais homogêneos com base em testes sobre os atributos de entrada (s_{i3}, s_{i4}, s_{i5}). Essas divisões geram **Nós Internos** (ou de decisão), e as ramificações representam os resultados dos testes. O processo continua até que **Nós Folha** (terminais) sejam alcançados, os quais contêm a predição final: um valor contínuo (média do alvo no nó) para regressão (g_i) ou um rótulo de classe para classificação (y_i). A predição para novas instâncias segue o caminho da raiz à folha correspondente.

2.2.2. Critérios de Divisão no CART

A eficácia do CART depende da seleção de um critério ótimo para dividir os nós.

- **Classificação:** Foram considerados o **Índice de Gini** ($1 - \sum p_k^2$), que mede a probabilidade de classificação incorreta, e a **Entropia** ($-\sum p_k \log_2 p_k$), que mede a impureza do nó. Em ambos os casos, busca-se maximizar a redução da impureza (ou o Ganho de Informação, no caso da Entropia) após a divisão.
- **Regressão:** Utiliza-se a minimização do **Erro Quadrático Médio (MSE)** das amostras nos nós filhos, o que é equivalente a maximizar a redução da variância.

2.2.3. Controle de Complexidade: Overfitting e Poda

Árvores de decisão são propensas ao sobreajuste (*overfitting*). Para mitigar este problema, técnicas de poda são empregadas. A **pré-poda** (early stopping) interrompe o crescimento da árvore com base em critérios como profundidade máxima (`max_depth`), número mínimo de amostras para divisão (`min_samples_split`) ou por folha (`min_samples_leaf`). A **pós-poda** remove ramos de uma árvore completamente crescida que não melhoram significativamente o desempenho, como a poda por custo-complexidade (`ccp_alpha`). A análise detalhada do ajuste desses hiperparâmetros e seu impacto nos resultados é apresentada na Seção 3.1.

2.3. Random Forest

O Random Forest [Breiman 2001] é um algoritmo de aprendizado por conjunto (*ensemble learning*) que visa melhorar o desempenho e a robustez de árvores de decisão individuais, especialmente na redução do sobreajuste. Ele constrói múltiplas árvores CART durante o treinamento e combina suas predições.

2.3.1. Princípios Fundamentais e Funcionamento

Os principais mecanismos do Random Forest implementado são:

- **Amostragem Bootstrap (Bagging):** Cada árvore é treinada em uma subamostra do conjunto de dados original, gerada por amostragem aleatória com reposição. Isso introduz diversidade entre as árvores.
- **Subespaço Aleatório de Atributos:** Em cada nó de cada árvore, apenas um subconjunto aleatório dos atributos de entrada (s_{i3} , s_{i4} , s_{i5}) é considerado para a divisão, controlado pelo hiperparâmetro `n_features_per_tree`. Isso decorrelaciona as árvores.
- **Agregação de Predições:** Para regressão (g_i), a predição final é a média das predições das árvores. Para classificação (y_i), utiliza-se o voto majoritário (moda).

2.3.2. Hiperparâmetros e Construção das Árvores

As árvores individuais no Random Forest são árvores CART, utilizando os critérios de divisão (Gini para classificação, MSE para regressão) descritos anteriormente. Os hiperparâmetros chave do Random Forest que foram objeto de análise (detalhada na Seção 3.1) incluem o número de árvores (`n_trees`), além dos parâmetros de controle de complexidade de cada árvore individual (e.g., `max_depth`, `min_samples_leaf`, `ccp_alpha`) e o parâmetro `n_features_per_tree`.

2.4. Redes Neurais Artificiais (MLP)

As Redes Neurais Artificiais (RNA), especificamente a arquitetura Perceptron de Múltiplas Camadas (MLP), foram também investigadas devido à sua capacidade de modelar relações altamente complexas e não lineares [Goodfellow et al. 2016], sendo adequadas para as tarefas de regressão e classificação com os sinais vitais (s_{i3} , s_{i4} , s_{i5}).

2.4.1. Conceitos Fundamentais das Redes Neurais MLP

Uma MLP é uma rede neural *feedforward* composta por uma sequência de camadas de neurônios interconectados:

- **Neurônios (Nós):** São as unidades básicas de processamento. Cada neurônio recebe um conjunto de entradas, realiza uma combinação linear dessas entradas (soma ponderada), adiciona um valor de *bias* (viés), e aplica uma **Função de Ativação** não linear a este resultado para gerar sua saída. A função de ativação permite que a rede aprenda padrões complexos. Funções comuns incluem ReLU para camadas ocultas e funções lineares ou sigmóides/softmax para camadas de saída.
- **Camadas:** Os neurônios são organizados em camadas sequenciais:

- **Camada de Entrada (*Input Layer*):** Recebe os dados brutos (os atributos s_{i3}, s_{i4}, s_{i5}).
- **Camadas Ocultas (*Hidden Layers*):** Uma ou mais camadas que transformam os dados de entrada em representações progressivamente mais abstratas.
- **Camada de Saída (*Output Layer*):** Produz o resultado final. Para regressão (g_i), tipicamente um neurônio com ativação linear; para classificação, neurônios correspondentes às classes com ativação Softmax.
- **Pesos e Biases:** Parâmetros ajustáveis da rede (aprendidos durante o treinamento) que definem a força das conexões e os limiares de ativação.
- **Processo de Aprendizado (Treinamento):**
 - **Propagação Direta (*Forward Propagation*):** Os dados fluem da entrada para a saída, gerando uma predição.
 - **Função de Perda (*Loss Function*):** Mede a discrepância entre a predição e o valor real (e.g., MSE para regressão).
 - **Retropropagação (*Backpropagation*) e Otimizador:** O erro é propagado para trás, e um **otimizador** (e.g., Adam) ajusta pesos e biases para minimizar a perda.

2.4.2. Arquitetura e Configuração do Modelo MLP para Regressão

Para a tarefa de regressão da gravidade (g_i), foi projetado um modelo MLP utilizando TensorFlow com a API Keras. A arquitetura geral empregada incluiu uma camada inicial de normalização dos dados de entrada (s_{i3}, s_{i4}, s_{i5}); camadas ocultas densamente conectadas (`layers.Dense`) utilizando a função de ativação ReLU; e uma camada de saída com um único neurônio e função de ativação linear. O modelo foi compilado utilizando o otimizador Adam e a função de perda Erro Quadrático Médio (MSE). A especificação do número de neurônios em cada camada oculta, a taxa de aprendizado e outros hiperparâmetros relevantes foram determinados através do processo de ajuste descrito na Seção 3.1.

2.4.3. Processo de Treinamento e Classificação Indireta

O treinamento do modelo MLP de regressão foi conduzido utilizando os dados do arquivo de treinamento com rótulos, após a divisão em conjuntos de treinamento e validação. O processo iterativo ocorreu por um número definido de épocas e com um tamanho de lote específico. Para combater o sobreajuste e otimizar o tempo de treinamento, foi implementada a técnica de parada antecipada (*Early Stopping*), que monitora a performance no conjunto de validação. Neste estudo, a MLP foi primariamente designada para a tarefa de regressão para estimar g_i . A classificação do estado de saúde (y_i) nas quatro categorias foi então realizada como uma etapa de pós-processamento: os valores de gravidade g_i preditos pela MLP foram convertidos em rótulos de classe utilizando um conjunto de limiares (T_1, T_2, T_3). Estes limiares foram estabelecidos a partir da análise da distribuição dos valores de g_i para cada classe conhecida no conjunto de dados de treinamento.

2.4.4. Interpretabilidade do Modelo com SHAP

Considerando a natureza de "caixa-preta" das redes neurais, buscou-se obter alguma interpretabilidade sobre as predições do modelo MLP de regressão através da aplicação da técnica

SHAP (SHapley Additive exPlanations) [Lundberg and Lee 2017]. SHAP é um método baseado na teoria dos jogos que atribui um valor de importância para cada atributo de entrada para uma predição individual, permitindo uma visão sobre quais atributos (s_{i3} , s_{i4} , s_{i5}) mais influenciaram as predições de gravidade do modelo MLP.

2.5. Processo Experimental e Avaliação

Todos os modelos foram treinados e avaliados utilizando uma divisão dos dados de treinamento rotulados em conjuntos de treinamento e validação (80% para treino e 20% para validação). A robustez e a generalização foram avaliadas no conjunto de teste cego. As métricas de avaliação empregadas foram:

- **Para Regressão:** Raiz do Erro Quadrático Médio (RMSE) e Erro Absoluto Médio (MAE).
- **Para Classificação:** Acurácia, Precisão, Recall, Medida-F1 (F1-score) e Matriz de Confusão.

A seleção final dos hiperparâmetros para cada modelo foi guiada pelo desempenho no conjunto de validação e/ou por validação cruzada, conforme detalhado na análise de hiperparâmetros (Seção 3.1).

3. Resultados e Análise

Nesta seção, são apresentados e analisados os resultados obtidos pelos modelos de aprendizado de máquina. Iniciamos com a discussão sobre o ajuste e seleção dos hiperparâmetros para os algoritmos de Árvore de Decisão (CART), Random Forest e Redes Neurais MLP, fundamentando as escolhas que levaram às configurações finais dos modelos. Em seguida, apresentamos o desempenho comparativo dos modelos e uma discussão sobre os achados.

3.1. Análise e Ajuste de Hiperparâmetros

A performance dos modelos de aprendizado de máquina é sensível à configuração de seus hiperparâmetros. Portanto, uma exploração criteriosa destes parâmetros é fundamental para otimizar a capacidade de aprendizado do modelo e sua generalização para dados não vistos, mitigando o sobreajuste. Os resultados dessa exploração, detalhados nas tabelas do Apêndice A e visualizados nas figuras dos Apêndices B e C, foram a base para as configurações finais dos modelos.

Árvore de Decisão (CART). O ajuste do CART envolveu a análise do critério de divisão (criterion), profundidade máxima (max_depth), número mínimo de amostras por folha (min_samples_leaf) e o coeficiente de poda (ccp_alpha), cujos resultados detalhados podem ser consultados nas Tabelas 1 e 2 (Apêndice A). Para a **classificação**, a Figura 3 (Apêndice B) ilustra a comparação entre os critérios 'gini' e 'entropy', visando identificar aquele que proporcionava a maior acurácia (mean_test_accuracy). A profundidade da árvore (max_depth), conforme Figura 1, foi ajustada para evitar que a árvore se tornasse excessivamente complexa (risco de sobreajuste, observado por uma queda na acurácia de validação ou grande diferença para a acurácia de treino) ou demasiadamente simples (subajuste). O min_samples_leaf (Figura 2) atuou como um regularizador: valores maiores tendem a criar árvores mais genéricas, enquanto valores muito pequenos podem levar a folhas que capturam ruído. O ccp_alpha permitiu uma poda eficaz, simplificando a árvore ao remover ramos menos significativos, e o valor ótimo foi aquele que manteve alta performance com menor complexidade. A configuração final (ver Tabelas 1 e

2 no Apêndice A) foi escolhida buscando o melhor `rank_test_accuracy` e uma baixa variância (`std_test_accuracy`). Para a **regressão**, o objetivo foi minimizar o RMSE (`mean_test_neg_root_mean_squared_error`) e maximizar o R^2 Score. As Figuras 4, 6, 5, e 7 (Apêndice B) demonstram como `max_depth` e `min_samples_leaf` foram ajustados para otimizar essas métricas, resultando na configuração detalhada nas Tabelas 1 e 2.

Random Forest. Para os modelos Random Forest, cujos resultados do ajuste são apresentados nas Tabelas 3 e 4 (Apêndice A), foram ajustados `n_estimators`, `max_features`, e os parâmetros das árvores base (`max_depth`, `min_samples_leaf`, `ccp_alpha`). No **RandomForestClassifier**, o número de árvores (`n_estimators`) foi aumentado até que a acurácia (`test_score_accuracy`) e outras métricas como precisão, recall e F1-score (como exemplificado na análise da Figura 8 no Apêndice C) se estabilizassem, indicando que adicionar mais árvores não traria ganho significativo (e.g., a configuração de 20 árvores com `max_features='sqrt'`, `max_depth=8`, `min_samples_leaf=2`, `ccp_alpha=0.0030` e `criterion='gini'` mostrou-se uma combinação promissora). A escolha de `max_features='sqrt'` é uma heurística comum para garantir a diversidade entre as árvores, essencial para o bom desempenho do ensemble. Para o **RandomForestRegressor**, um número maior de árvores (`n_estimators=200`) contribuiu para a redução do RMSE (`test_score_rmse`). A configuração com `max_features='all'` (considerando todas as 3 features), `max_depth=None` (ou um valor alto), `min_samples_leaf=1`, `min_samples_split=3` e `ccp_alpha=0.0050` foi identificada como uma configuração de alto desempenho, onde o ensemble naturalmente lida com a complexidade das árvores individuais. A configuração final para ambos os modelos Random Forest é apresentada nas Tabelas 3 e 4 (Apêndice A).

Redes Neurais (MLP). O ajuste da MLP para regressão, com resultados sumarizados na Tabela 5 (Apêndice A), focou na arquitetura (número de neurônios em `NEURONS_LAYER_1` e `NEURONS_LAYER_2`) e parâmetros de treinamento como taxa de aprendizado (`LEARNING_RATE`) e tamanho do lote (`BATCH_SIZE`). O objetivo foi minimizar o `RMSE_Final_Validacao` e o `MAE_Final_Validacao`. As curvas de perda e MAE durante o treinamento (Figuras 16 e 17 no Apêndice C) foram cruciais para identificar o ponto ótimo de parada via `EarlyStopping` (coluna `Melhor_Epoca`), evitando sobreajuste. Uma configuração como 64 neurônios na primeira camada oculta, 32 na segunda e 16 na terceira, ativação ReLU, taxa de aprendizado de 0.001 e `BATCH_SIZE=32` emergiu como uma das mais promissoras, conforme sumarizado na Tabela 5 (Apêndice A).

3.2. Desempenho Comparativo dos Modelos

Com os hiperparâmetros ajustados, o desempenho dos modelos CART, Random Forest e MLP foi comparado. As curvas de aprendizado (Figuras 10 e 11 no Apêndice C) indicam a capacidade de generalização. As métricas finais obtidas no conjunto de teste para regressão e classificação são discutidas e analisadas nas subseções seguintes.

Tarefa de Regressão (Predição de g_i). Na predição do escore de gravidade g_i , a avaliação no conjunto de teste indicou que tanto o Random Forest Regressor quanto a MLP ajustada geralmente apresentaram RMSE e MAE inferiores aos da Árvore de Decisão única. Isso sugere que a agregação de múltiplas árvores no Random Forest e a capacidade da MLP de modelar

relações não lineares complexas foram vantajosas. A Figura 18 (Apêndice C) ilustra a dispersão das predições da MLP em relação aos valores reais, indicando a qualidade do ajuste.

Tarefa de Classificação (Predição de y_i). Para a classificação do estado de saúde y_i , de acordo com os resultados no conjunto de teste, o Random Forest Classifier geralmente exibiu maior acurácia e Medida-F1 em comparação com o CART. As matrizes de confusão para CART (Figura 15), Random Forest (Figura 9), e a derivada da MLP (a ser apresentada) detalham o desempenho por classe. A Figura 8 demonstra a performance detalhada do Random Forest, que tipicamente lida melhor com a variabilidade dos dados e apresenta um bom balanço entre precisão e recall nas diferentes classes. A classificação indireta da MLP, baseada nos limiares de g_i , obteve resultados competitivos, refletindo a qualidade do seu modelo de regressão subjacente.

3.3. Discussão

A análise comparativa dos modelos CART, Random Forest (RF) customizado e MLP permite discutir suas capacidades e adequações para a tarefa de triagem de vítimas baseada em sinais vitais. Os resultados do ajuste de hiperparâmetros (Seção 3.1) foram determinantes para extrair o potencial de cada abordagem, e o desempenho subsequente (Seção 3.2) permite uma discussão aprofundada.

Comparação Profunda do Desempenho dos Modelos. De maneira geral, os modelos de conjunto (Random Forest) e a Rede Neural MLP apresentaram tendência de superar a Árvore de Decisão (CART) individual em termos de métricas preditivas. A MLP, com sua arquitetura de múltiplas camadas e funções de ativação não lineares (ReLU, conforme Figura 21), possui uma estrutura que favorece a captura de relações complexas e não lineares nos dados [Goodfellow et al. 2016]. Essa capacidade pode justificar um desempenho potencialmente superior na tarefa de regressão para estimar g_i (Figura 18), especialmente se a relação subjacente entre os sinais vitais e a gravidade for intrincada. O Random Forest customizado, ao agregar as predições de múltiplas árvores de decisão, que individualmente já modelam não-linearidades através de partições no espaço de features, também demonstra essa capacidade. No entanto, a eficácia em capturar todas as nuances não-lineares pelo RF customizado pode ser dependente da profundidade e da otimalidade das árvores individuais, podendo, em alguns cenários, não alcançar a mesma flexibilidade de uma MLP bem ajustada.

Quanto à robustez a ruído, o Random Forest é conceitualmente superior a uma árvore de decisão única. A técnica de bagging (amostragem bootstrap) e a seleção aleatória de subconjuntos de features para cada árvore (descritas na Seção 2.3.1) contribuem para a construção de árvores decorrelacionadas. A média das predições (para regressão) ou o voto majoritário (para classificação) tende a suavizar o impacto de ruídos ou outliers que poderiam ser fortemente memorizados por uma única árvore de decisão, como o modelo CART. As curvas de aprendizado (Figuras 10 e 11) e a comparação das matrizes de confusão (e.g., Figura 9 versus Figura 15) frequentemente corroboram essa maior capacidade de generalização do RF.

Limitações do Estudo. Ao interpretar os resultados deste estudo, algumas limitações inerentes ao escopo e aos dados devem ser ponderadas. Um ponto relevante é o tamanho do conjunto de dados, constituído por aproximadamente 1500 amostras. Embora esta quantidade tenha permitido a modelagem e a extração de resultados significativos, um volume de dados mais expressivo poderia favorecer especialmente modelos mais complexos, como a MLP, no

que tange à capacidade de generalização e à estabilidade das estimativas de desempenho. Adicionalmente, a qualidade intrínseca dos dados de entrada é um fator determinante; eventuais imprecisões ou vieses na coleta dos sinais vitais ou na atribuição original da gravidade (g_i) e classe (y_i) seriam naturalmente propagados e aprendidos pelos algoritmos.

A natureza da variável alvo g_i (gravidade), descrita como "previamente calculado pelo corpo médico" sem o conhecimento da fórmula exata, representa uma limitação significativa. Os modelos estão, em essência, tentando replicar uma função desconhecida, que pode ela mesma ser uma aproximação ou heurística. Isso dificulta a avaliação contra uma verdade fundamental absoluta e significa que os modelos estão limitados pela qualidade e pela natureza dessa estimativa original. Finalmente, a implementação customizada do Random Forest, embora valiosa do ponto de vista didático, pode apresentar simplificações em comparação com implementações de bibliotecas consolidadas. Por exemplo, a ausência de um mecanismo de poda por complexidade de custo (`ccp_alpha`) plenamente funcional nas árvores individuais do RF customizado, conforme detalhado no código-fonte, pode restringir seu desempenho ótimo ou sua capacidade de generalização em comparação com versões mais otimizadas.

Interpretabilidade versus Performance. A escolha entre interpretabilidade e performance é uma consideração central na seleção de modelos de aprendizado de máquina, sobretudo em domínios críticos como o da saúde. A Árvore de Decisão (CART) destaca-se pela sua alta interpretabilidade; suas regras de decisão podem ser visualizadas e compreendidas por especialistas do domínio, o que facilita a validação e a confiança no modelo. A Figura 14 já indica a importância de features, e a estrutura da árvore em si oferece um fluxograma decisório claro. Contudo, essa transparência pode vir ao custo de um poder preditivo inferior em comparação com modelos mais complexos.

O Random Forest customizado, sendo um ensemble, já é menos diretamente interpretável que uma única árvore, embora ainda forneça métricas de importância de features agregadas (Figuras 12 e 13) que oferecem alguma visão sobre o processo. A MLP é tradicionalmente vista como o modelo mais "caixa-preta", devido à sua complexa rede de neurônios interconectados e transformações não lineares (Seção 2.4.1). Embora frequentemente alcance alta performance, a dificuldade em entender *como* ela chega a uma predição específica pode ser uma barreira para sua adoção. Neste contexto, técnicas como SHAP (SHapley Additive exPlanations) [Lundberg and Lee 2017] são cruciais. Como demonstrado pela Figura 20 e discutido na Seção 2.4.4, o SHAP provê explicações para as saídas da MLP, atribuindo valores de importância para cada feature em predições individuais e de forma agregada. Isso aumenta a transparência da MLP, permitindo análises mais profundas sobre seu comportamento e contribuindo para mitigar o problema da interpretabilidade, tornando-a uma opção mais viável mesmo em cenários que demandam explicação.

4. Conclusões

Este estudo realizou uma análise comparativa entre Árvores de Decisão (CART), um modelo Random Forest customizado e Redes Neurais MLP para a tarefa de estimar o escore de gravidade e classificar o estado de saúde de vítimas com base em um conjunto de três sinais vitais: qualidade da pressão arterial (s_{i3}), pulso (s_{i4}) e frequência respiratória (s_{i5}). O processo de ajuste de hiperparâmetros, detalhado na Seção 3.1 e sumarizado no Apêndice A, demonstrou ser uma etapa essencial para otimizar o desempenho de cada um dos algoritmos investigados.

Os resultados apresentados na Seção 3.2 indicam que, de modo geral, tanto o Random Forest customizado quanto a Rede Neural MLP ofereceram um desempenho preditivo superior

em comparação com o modelo de Árvore de Decisão (CART) individual. O Random Forest destacou-se pela sua robustez e pela capacidade de alcançar um bom equilíbrio entre as métricas de regressão e classificação, beneficiando-se da agregação de múltiplas árvores. A MLP, por sua vez, demonstrou uma forte capacidade na tarefa de regressão para predição do escore de gravidade g_i , e sua abordagem de classificação indireta a partir dos escores de gravidade preditos também se mostrou competitiva.

A análise de importância de features, realizada através de métodos intrínsecos aos modelos baseados em árvore (Figuras 14, 12, 13) e via SHAP para a MLP (Figura 20), consistentemente apontou para a relevância dos três sinais vitais (s_{i3} , s_{i4} , s_{i5}) como preditores importantes para as tarefas propostas, embora a ordem de magnitude de sua influência pudesse variar entre os diferentes modelos e tarefas (regressão vs. classificação). Por exemplo, o pulso (s_{i4}) frequentemente figurou entre os atributos de maior impacto.

5. Contribuição dos Autores e Código Fonte

- Renzo Tognella De Rosa (30 horas): Responsável pela modelagem com Árvores de Decisão (CART), e ajustes no algoritmo de Random Forest incluindo o ajuste de seus hiperparâmetros e análise de resultados. Contribuiu com a estruturação geral do artigo e a elaboração dos gráficos apresentados.
- Miguel Dos Anjos Brack (34 horas): Encarregado da implementação e análise dos modelos Random Forest e Redes Neurais MLP, incluindo o ajuste de seus respectivos hiperparâmetros. Colaborou no desenvolvimento do artigo e na discussão dos resultados.

O código fonte das implementações está disponível em: https://github.com/Renzo-Tognella/trabalho_2_SI

Referências

- [Breiman 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- [Breiman et al. 1984] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. *Wadsworth International Group*.
- [Goodfellow et al. 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- [Lundberg and Lee 2017] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 4765–4774.
- [Russell and Norvig 2021] Russell, S. J. and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition.

A. Tabelas de Resumo do Ajuste de Hiperparâmetros

Nesta seção, são apresentadas tabelas resumindo uma seleção dos melhores resultados obtidos durante o ajuste de hiperparâmetros para cada algoritmo, ordenados pela métrica de desempenho principal. Os dados completos da experimentação foram utilizados para gerar estes sumários.

Tabela 1. Resultados Seleccionados do Ajuste da Árvore de Decisão (CART) - Classificação.

Critério	Max Depth	Min Samples Leaf	CCP Alpha	Acurácia Média (Teste)	Std (Acurácia)
gini	10	1	0.0005	0.9044	N/A
entropy	10	1	0.0005	0.9022	N/A
gini	7	5	0.0010	0.8978	N/A
gini	10	5	0.0010	0.8978	N/A
entropy	7	1	0.0010	0.8956	N/A

Seleção de resultados do ajuste da Árvore de Decisão para classificação. Std (Acurácia) não disponível nesta súmula.

Tabela 2. Resultados Seleccionados do Ajuste da Árvore de Decisão (CART) - Regressão.

Max Depth	Min Samples Leaf	CCP Alpha	RMSE Médio (Teste)	R ² Score (Teste)
12	6	0.0010	4.5186	0.9250
9	4	0.0020	4.6977	0.9190
8	1	0.0010	4.8698	0.9129
10	1	0.0008	5.0122	0.9077
10	2	0.0008	5.0122	0.9077

Seleção de resultados do ajuste da Árvore de Decisão para regressão, ordenados por RMSE.

Tabela 3. Resultados Seleccionados do Ajuste do Random Forest - Classificador.

N Estimators	Max Depth	Min Samples Leaf	Max Features	CCP Alpha	Acurácia (Teste)
200	10	1	all	0.0001	0.9113
20	10	2	sqrt	0.0030	0.9087
100	12	1	all	0.0005	0.9087
15	10	1	sqrt	0.0010	0.9073
200	7	1	log2	0.0008	0.9067

Seleção de resultados do ajuste do Random Forest para classificação, ordenados por Acurácia Média.

Tabela 4. Resultados Seleccionados do Ajuste do Random Forest - Regressor.

N Estimators	Max Depth	Min Samples Leaf	Max Features	CCP Alpha	RMSE (Teste)
200	N/A (None)	1	all	0.0050	2.9056
100	N/A (None)	1	all	0.0030	2.9989
200	N/A (None)	3	all	0.0050	3.0267
100	N/A (None)	1	sqrt	0.0010	3.0708
20	9	4	all	0.0008	3.2186

Seleção de resultados do ajuste do Random Forest para regressão, ordenados por RMSE Médio.

Tabela 5. Resultados Seleccionados da Rede Neural MLP.

Neurons L1,L2 e L3	Learning Rate	Batch Size	Melhor Época	Acurácia (Teste)
64, 32, 16	0.0010	8	200	0.9400
128, 64, 32	0.0050	16	200	0.9393
128, 64, 32	0.0005	8	178	0.9387
128, 64, 32	0.0005	16	200	0.9380
128, 64, 32	0.0050	64	200	0.9367

Seleção de resultados do ajuste da Rede Neural MLP, ordenados pela métrica de desempenho principal.

B. Figuras da Análise de Hiperparâmetros CART

Nesta seção, são apresentados gráficos que ilustram a análise de hiperparâmetros para os modelos de Árvore de Decisão (CART).

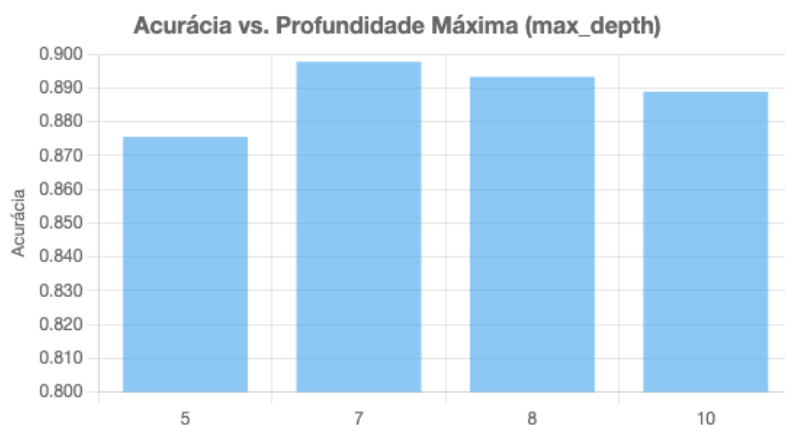


Figura 1. Impacto da Profundidade Máxima (max_depth) na Acurácia do modelo de classificação CART.

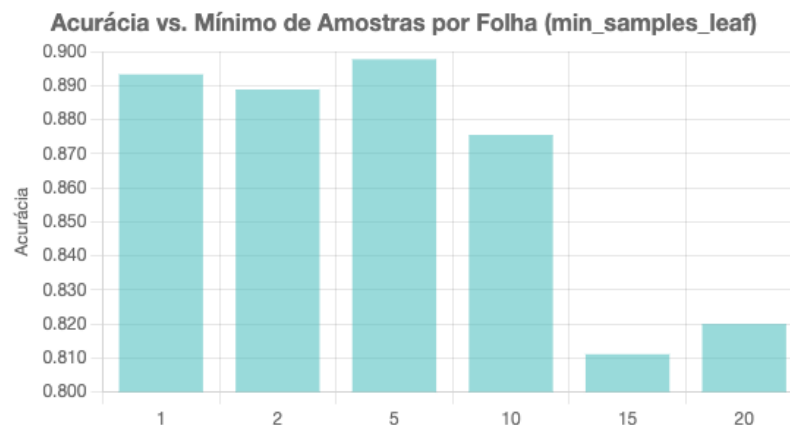


Figura 2. Impacto do Mínimo de Amostras por Folha (min_samples_leaf) na Acurácia do modelo de classificação CART.

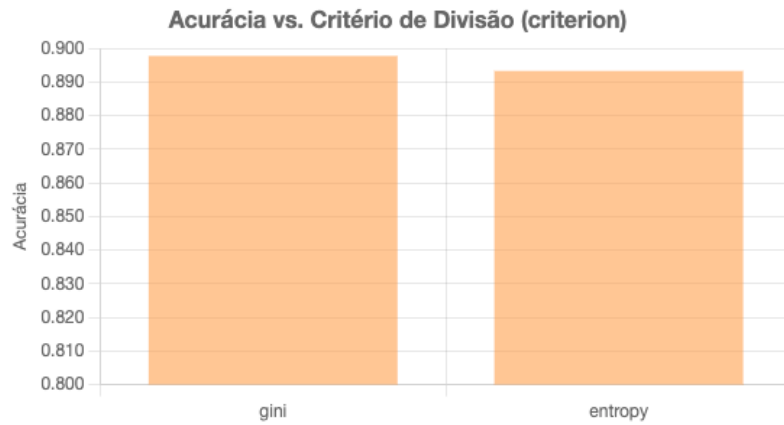


Figura 3. Comparação da Acurácia entre os Critérios de Divisão 'gini' e 'entropy' para o modelo de classificação CART.

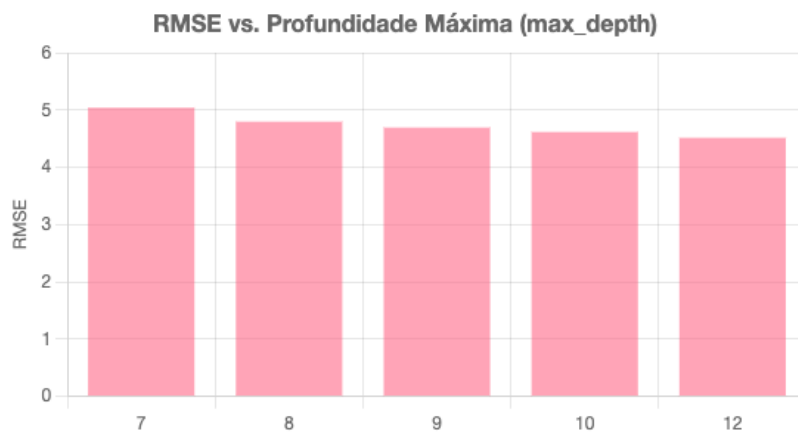


Figura 4. Impacto da Profundidade Máxima (max_depth) no RMSE do modelo de regressão CART.

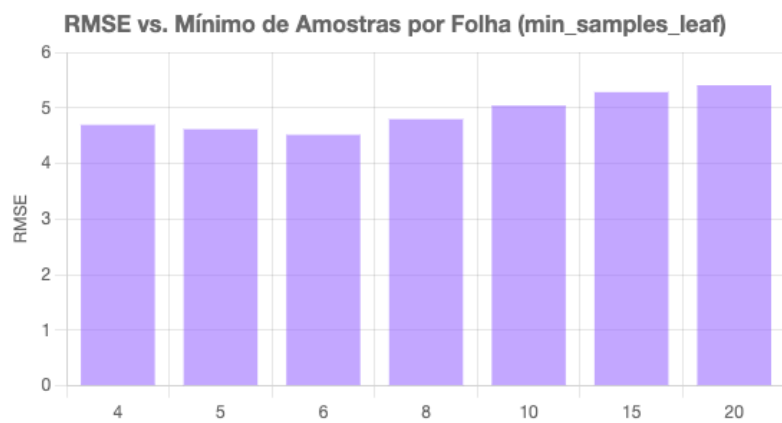


Figura 5. Impacto do Mínimo de Amostras por Folha (min_samples_leaf) no RMSE do modelo de regressão CART.

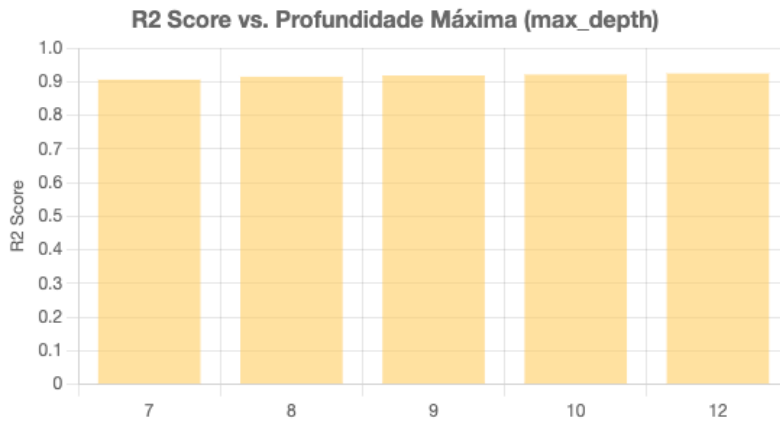


Figura 6. Impacto da Profundidade Máxima (`max_depth`) no R^2 Score do modelo de regressão CART.

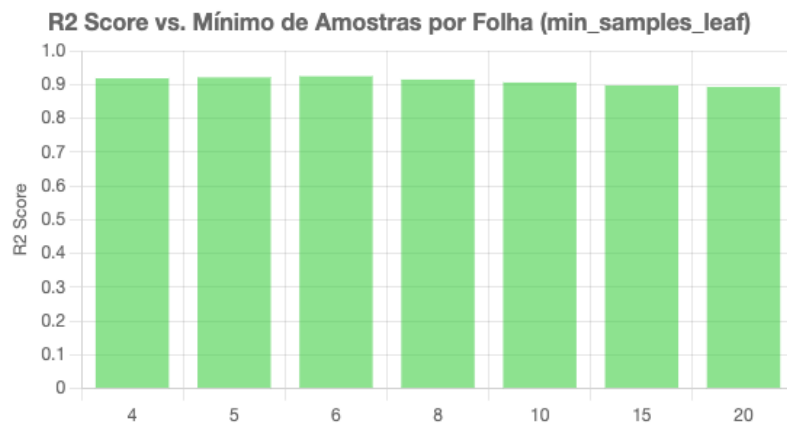


Figura 7. Impacto do Mínimo de Amostras por Folha (`min_samples_leaf`) no R^2 Score do modelo de regressão CART.

C. Figuras de Desempenho e Interpretabilidade dos Modelos

Esta seção contém figuras relativas ao desempenho dos modelos Random Forest e MLP, curvas de aprendizado e análise de importância de features.

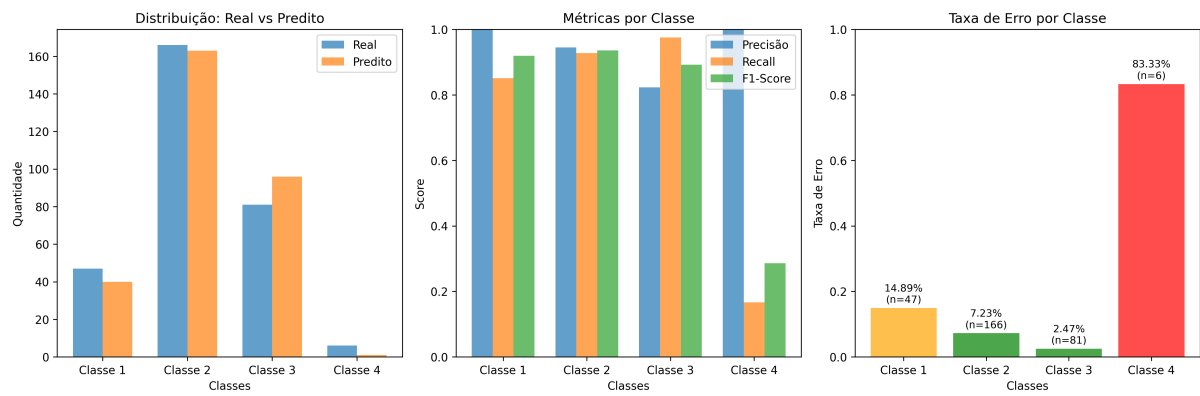


Figura 8. Análise de desempenho do Random Forest Classifier: (a) Distribuição de classes reais vs. preditas, (b) Métricas de Precisão, Recall e F1-Score por classe, e (c) Taxa de erro por classe.

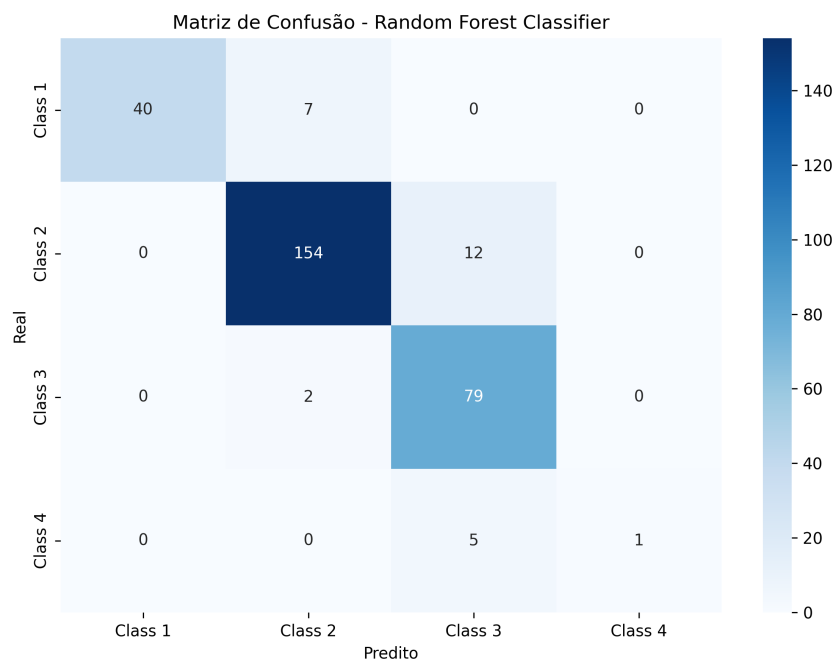


Figura 9. Matriz de confusão para o modelo Random Forest Classifier.

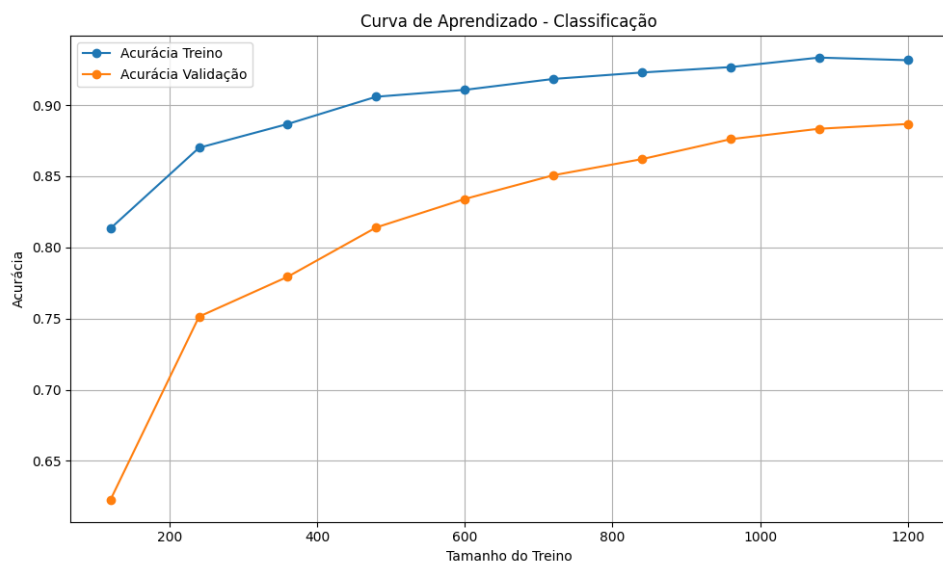


Figura 10. Curva de aprendizado ilustrativa para um modelo de classificação.

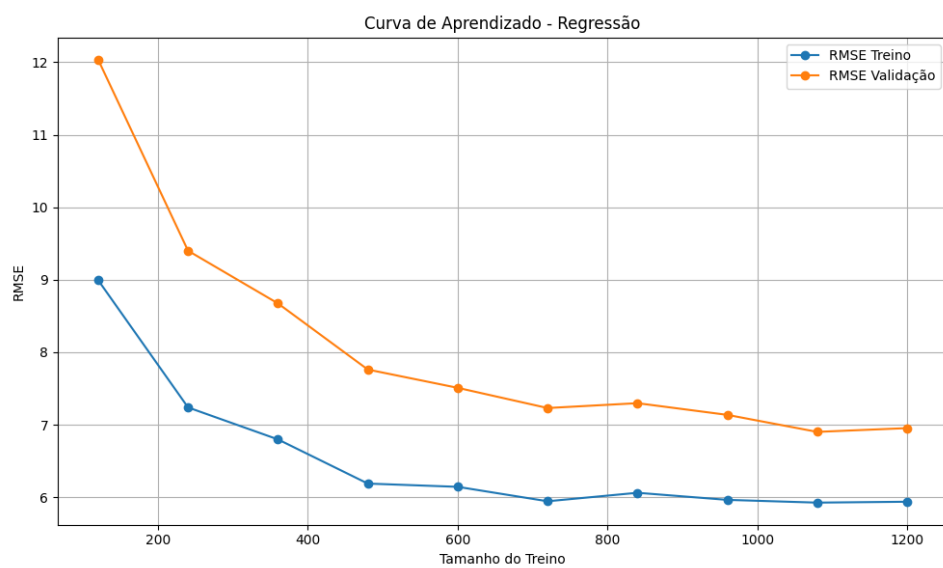


Figura 11. Curva de aprendizado ilustrativa para um modelo de regressão.

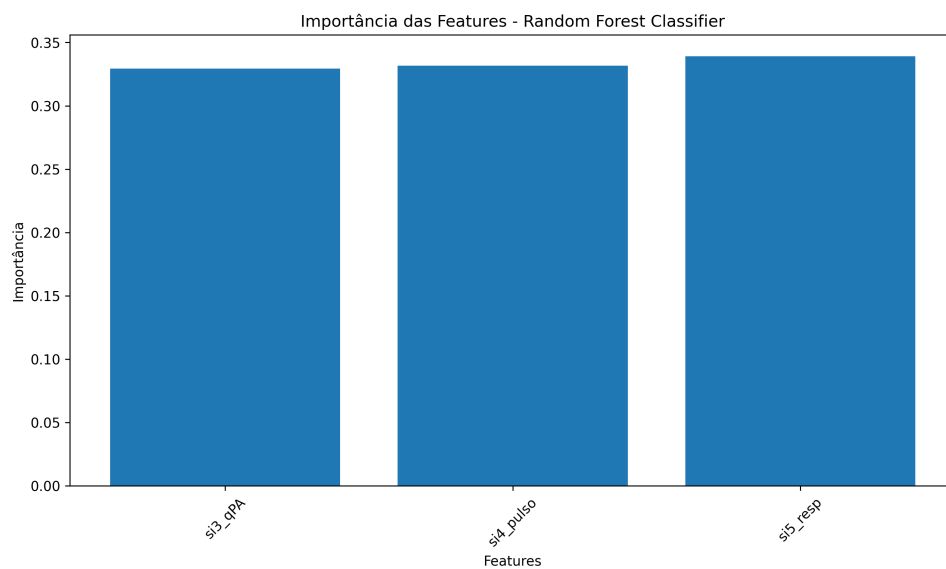


Figura 12. Importância das features para o Random Forest Classifier.

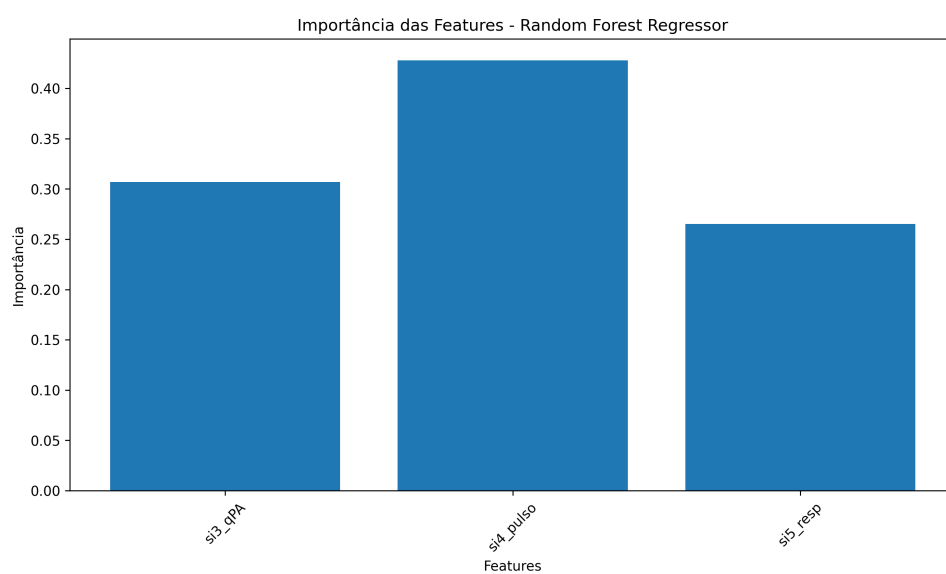


Figura 13. Importância das features para o Random Forest Regressor.

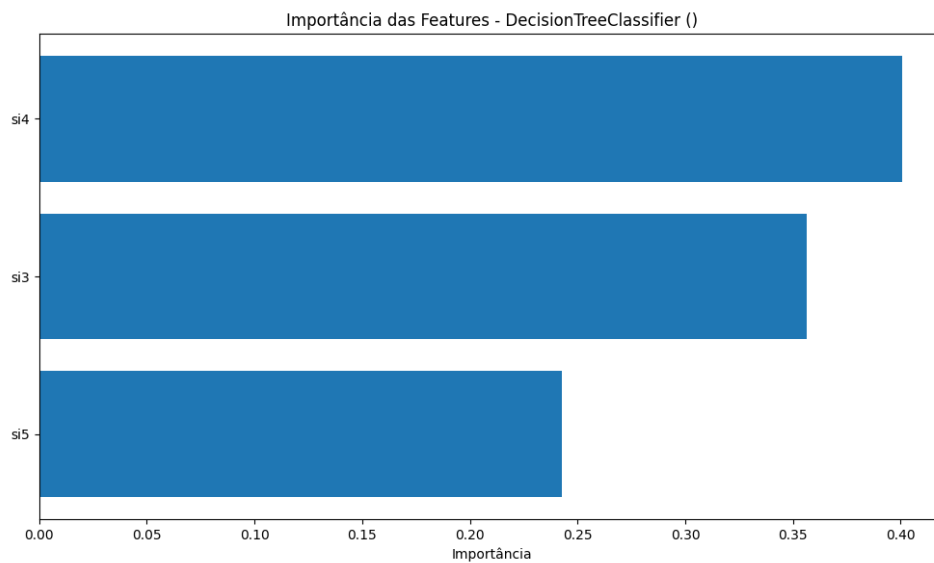


Figura 14. Importância das features para o modelo de Árvore de Decisão (CART) de classificação.

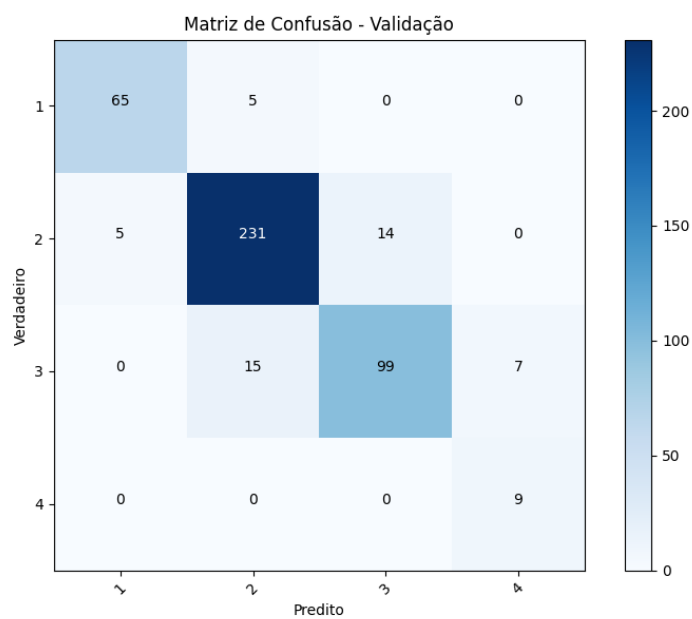


Figura 15. Matriz de confusão do modelo de Árvore de Decisão (CART) de classificação (validação).

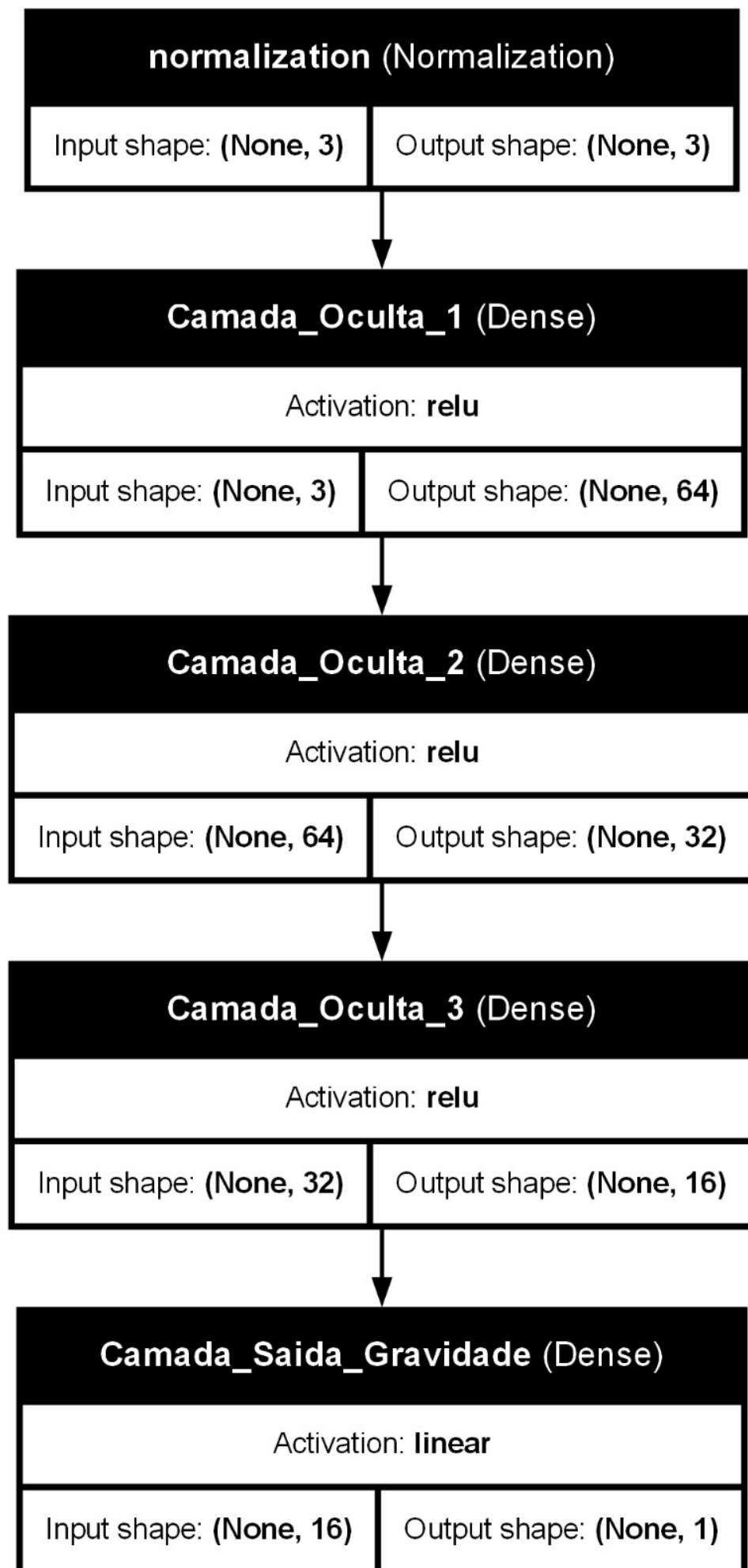


Figura 16. Diagrama da arquitetura da Rede Neural ML P utilizada para regressão)

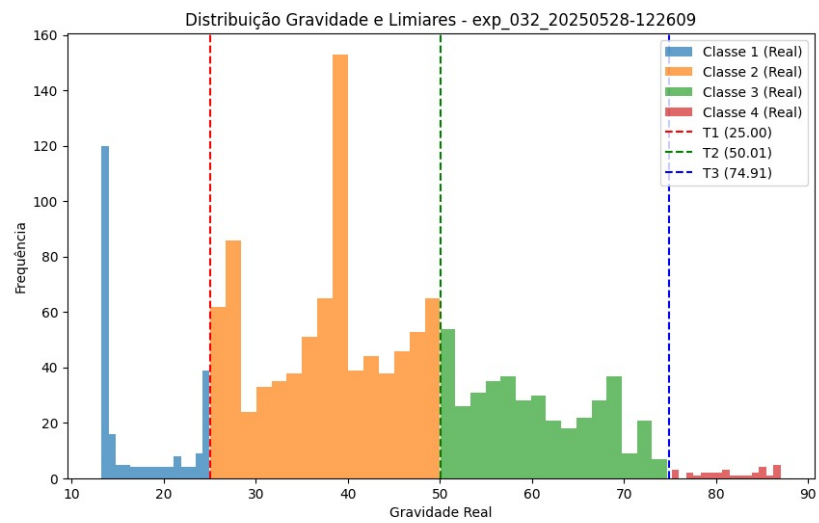


Figura 17. Distribuição da gravidade e Limiares

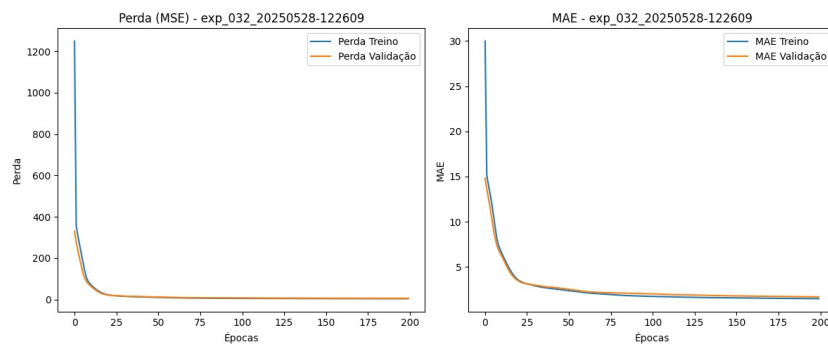


Figura 18. Histórico de Perda (MSE) durante o treinamento do modelo MLP de regressão (Treino vs. Validação e Histórico do Erro Absoluto Médio (MAE) durante o treinamento do modelo MLP de regressão (Treino vs. Validação).

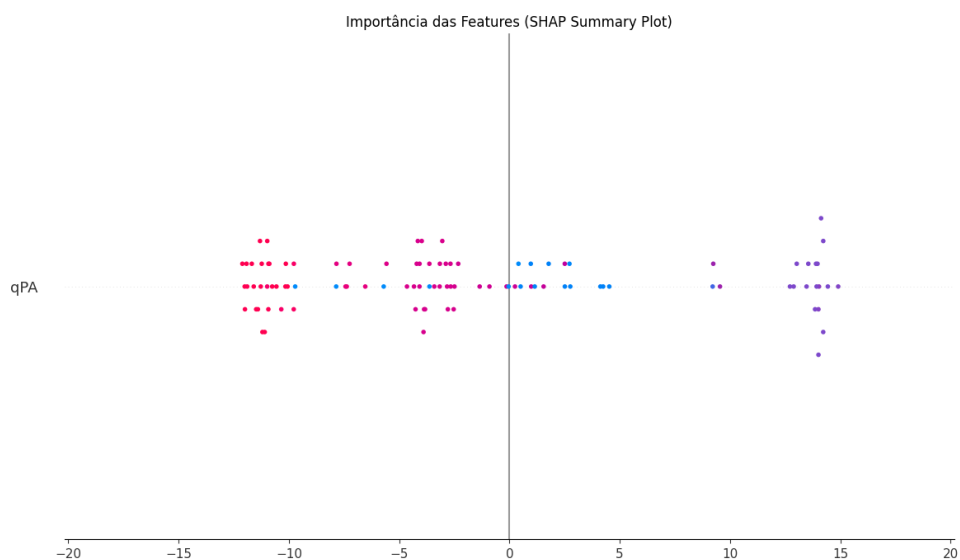


Figura 19. Importância das features (SHAP Summary Plot) para o modelo MLP de regressão.

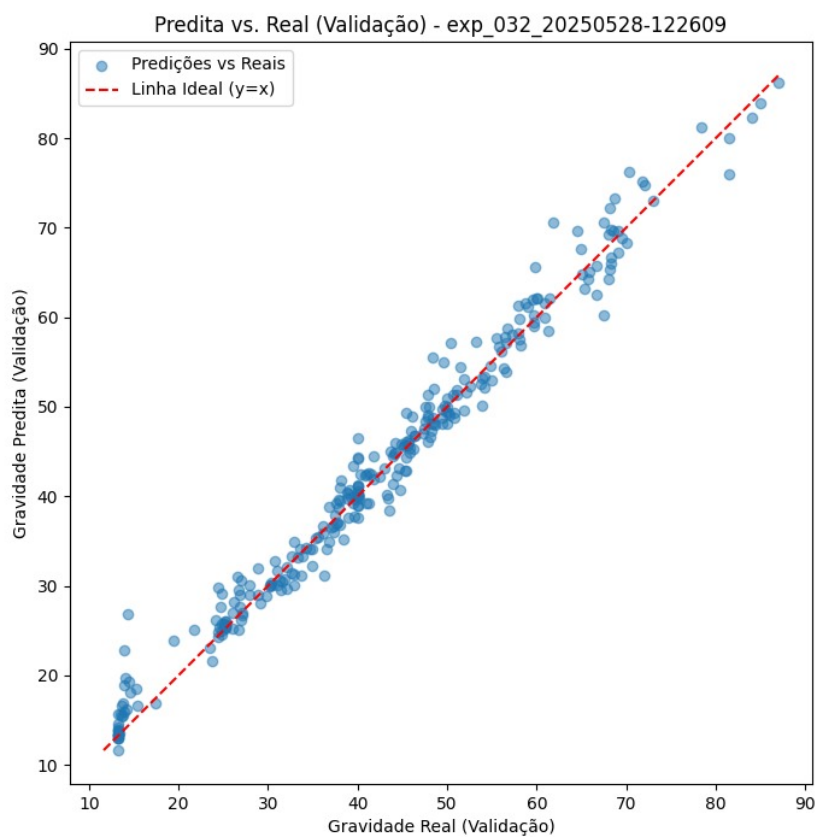


Figura 20. Comparação entre Gravidade Predita vs. Real para o modelo MLP de regressão no conjunto de validação

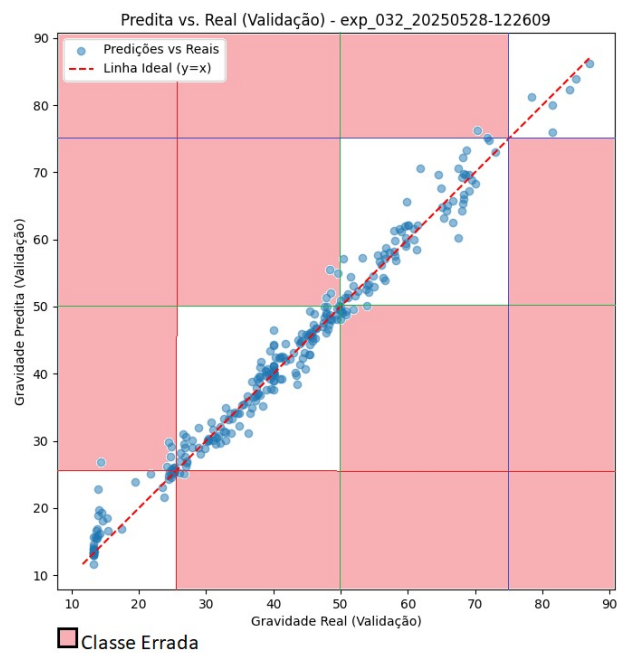


Figura 21. Comparação entre Gravidade Predita vs. Real para o modelo MLP de regressão no conjunto de validação (Divisão por Classes)