

Enhancing the Interpretability of Cardiovascular Disease Classifiers using Born-Again Tree Ensembles

L. G. S. N. A. Magalhães¹, E. C. Gonçalves¹

Escola Nacional de Ciências Estatísticas (ENCE-IBGE), Brazil
luisg.santanna@gmail.com, eduardo.correa@ibge.gov.br

Abstract. According to the Pan American Health Organization, more people die each year from cardiovascular diseases than from any other cause. Due to this, ensemble classifiers such as Random Forest (RF) have been increasingly employed to build models targeted at the early prediction of such diseases. Nevertheless, one of the main disadvantages of the ensemble approaches lies in the fact that they cannot be applied when the goal is to build interpretable models (which are often desired or even required by both physicians and patients). To tackle this problem, in this work we evaluate the use of Born-Again Tree Ensembles (BA), a recently proposed technique that corresponds to the the first exact algorithm that transforms an RF into a single decision tree. Experiments carried out on a dataset containing data about 1,417 patients show that BA was able to produce a model that is directly interpretable, while at the same time keeping the same predictive power of an RF model.

CCS Concepts: • **Computing methodologies** → **Machine learning algorithms**.

Keywords: born-again tree, random forest, interpretability, cardiovascular disease classification, data mining

1. INTRODUÇÃO

De acordo com a Organização Pan-Americana da Saúde [OPAS 2024b], as doenças cardiovasculares (DCVs) são a principal causa de morte no mundo. DCV é um termo geral para condições que afetam o coração ou os vasos sanguíneos, usualmente associadas a um acúmulo de depósitos gordurosos dentro das artérias (aterosclerose) e um risco aumentado de coágulos sanguíneos. No Brasil, no ano de 2021, a taxa de prevalência de DCV foi de 7,6% para os homens e de 6,3% para as mulheres, enquanto a taxa de mortalidade padronizada por idade por 100 mil habitantes foi 348,5 [de Oliveira et al. 2024].

A identificação precoce do desenvolvimento de DCV reduz a probabilidade de complicações em pacientes de alto risco. Por essa razão, diferentes competições para promover a pesquisa sobre uso de algoritmos de mineração de dados e aprendizado de máquina na predição de DCV têm sido realizadas desde o início dos anos 2000 [Berka et al. 2007]. Por exemplo, no ano de 2004, os organizadores da competição ECML-PKDD Discovery Challenge [ECML-PKDD 2004] disponibilizaram a base de dados STULONG/ENTRY, que armazena os dados de um estudo sobre o desenvolvimento de fatores de risco de DCV em uma população de 1.417 homens de meia-idade. Esses pacientes preencheram um formulário com seus dados pessoais e hábitos gerais e também foram submetidos a exames físicos e bioquímicos. De acordo com essas informações, os pacientes foram classificados por especialistas em 3 diferentes grupos de risco: baixo, médio e alto. O objetivo da competição foi o de empregar algoritmos de mineração de dados para descobrir padrões escondidos associados a cada um dos grupos de risco na base de dados. Ao final do desafio, algoritmos baseados em regras de associação [Gonçalves and Plastino 2004] e padrões emergentes [Soulet and Hébert 2004] obtiveram os melhores resultados.

Ao longo dos 20 anos que se passaram desde a PKDD Discovery Challenge 2004, a previsão do risco

Copyright©2024 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

de DCV se manteve como um importante tópico de pesquisa na área de aprendizado de máquina [Berka et al. 2007; Bhatt et al. 2023]. Nos anos recentes, a maior parte dos trabalhos propostos na literatura envolve o uso de algoritmos de classificação [Asif et al. 2023; Bhatt et al. 2023; Sumwiza et al. 2023]. No cenário mais comum, tais algoritmos são aplicados sobre bases de dados em que cada exemplo (paciente) está associado a um rótulo de classe (ex.: paciente acometido ou não acometido por DCV) com o objetivo de realizar o aprendizado de um modelo preditivo capaz de inferir o rótulo de classe de novos exemplos (novos pacientes). Entretanto, nota-se que essas abordagens recentes são baseadas na construção de modelos de classificação do tipo caixa-preta, através do emprego de técnicas como redes neurais e *ensembles* [Bhatt et al. 2023]. Estas técnicas foram projetadas para maximizar a acurácia dos classificadores, mas possuem uma desvantagem relevante: não são adequadas quando o problema requer a construção de modelos de classificação interpretáveis. Um classificador interpretável possui a habilidade de “explicar” as suas classificações para os usuários através, por exemplo, da apresentação de regras de classificação no formato: SE <condição> ENTÃO <rótulo de classe> [da Silva and Gonçalves 2021]. No problema da previsão do risco de DCV, a habilidade de interpretar o resultado de uma classificação é considerada desejável por médicos e pacientes por diferentes razões:

- Entender um modelo de classificação é normalmente um pré-requisito para que os médicos e pacientes confiem nas predições realizadas pelo mesmo e para que os pacientes sigam as recomendações associadas a estas predições [Freitas 2014].
- Classificadores interpretáveis são capazes de apresentar novos *insights* a respeito de relações preditivas importantes na base de dados. Por exemplo, indicar quais combinações de atributos representam os mais fortes preditores para a predição de DCV [Berka et al. 2007].
- Razões econômicas. No campo da medicina, entender as predições de um modelo ajuda os profissionais a confiarem no modelo de classificação, aumentando as chances de que tempo e dinheiro sejam investidos na execução de exames médicos sofisticados que tentarão confirmar as predições [Freitas 2014].
- Razões legais e éticas. Por exemplo, desde 2018, uma lei da União Europeia exige que algoritmos que realizem predições capazes de “afetar significativamente as pessoas” devam fornecer o “direito à explicação” [Doshi-Velez and Kim 2017; Rudin 2019].

Sendo assim, o presente artigo tem por objetivo avaliar o desempenho preditivo e a interpretabilidade de um modelo para classificação de risco de DCV produzido através da técnica Born-Again Tree Ensembles (BA), recentemente proposta em [Vidal and Schiffer 2020]. O algoritmo BA é o primeiro algoritmo exato capaz de transformar um modelo de classificação Random Forest (*ensemble* de árvores de decisão) em uma única árvore de decisão com o mesmo poder preditivo, porém com a vantagem de ser diretamente interpretável pelos usuários. Os experimentos foram realizados na base de dados STULONG/ENTRY e os resultados da técnica BA foram comparados com os obtidos pelos tradicionais algoritmos CART [Breiman et al. 1984] e Random Forest (RF) [Breiman 2001]. Salvo melhor juízo, este é o primeiro estudo em que o algoritmo BA é aplicado no problema de classificar o risco de DCV em pacientes.

O restante do artigo está dividido da seguinte forma. A Seção 2 faz uma revisão sobre trabalhos dedicados à previsão do risco de DCV. A Seção 3 apresenta uma breve introdução ao algoritmo BA. Em seguida, a Seção 4 apresenta mais detalhes sobre a base de dados STULONG/ENTRY, utilizada neste estudo, enquanto a Seção 5 apresenta os resultados dos experimentos. Por fim, conclusões e ideias para trabalhos futuros são apresentadas na Seção 6.

2. TRABALHOS RELACIONADOS

Esta seção faz uma breve revisão sobre duas diferentes abordagens para previsão do risco de DCV: as que empregam equações para realizar estimativas e as que são baseadas na utilização de algoritmos de mineração de dados e aprendizado de máquina.

Os métodos baseados em escores e equações são os mais conhecidos e utilizados pela comunidade médica. Nesta categoria, destaca-se a equação SCORE2 [SCORE2 WORKING GROUP 2021], utilizada na Europa. Ela está disponível na internet para acesso público e utiliza poucos atributos para realizar a estimativa de risco de DCV: idade, sexo, pressão sanguínea sistólica, colesterol total, colesterol HDL (*high-density lipoprotein*) e se o paciente é ou não fumante. Além de realizar a estimativa, a SCORE2 também recomenda hábitos para melhorar a saúde e metas para alcançar com o objetivo de reduzir o risco de DCV. Outra ferramenta conhecida é a Calculadora de Risco Cardiovascular disponível no site da Organização Pan-Americana da Saúde [OPAS 2024a], que estima a possibilidade de ocorrência de infarto do miocárdio, acidente vascular cerebral ou morte de origem cardiovascular em dez anos. Esse escore de risco surgiu a partir de uma extensa revisão das coortes disponíveis e uma adaptação para 21 regiões mundiais com base na análise da carga de morbimortalidade cardiovascular. Para cada região, publicou-se uma estimativa que demanda o conhecimento do nível sanguíneo de colesterol. Ela tem como objetivo ajudar os profissionais de saúde a calcular com rapidez o risco cardiovascular e a conversar com os pacientes para analisar em que medida o risco pode ser modificado. Além disso, pretende ajudar pessoas preocupadas com a própria saúde, permitindo que avaliem a necessidade de uma consulta médica quando seu risco não for baixo.

Métodos como o SCORE2 e a Calculadora de Risco Cardiovascular são muito fáceis de serem utilizados (já que exigem poucos dados de entrada) e apresentam resultados (classificações de risco e suas causas) que podem ser facilmente interpretados por qualquer usuário, seja ele profissional da área da saúde ou um paciente. Entretanto, muitas combinações de fatores podem estar envolvidas no desenvolvimento de DCVs [Berka et al. 2007] e, por esse motivo, algoritmos de mineração de dados e aprendizado podem ser empregados para que melhores estimativas sejam obtidas. No desafio proposto pela PKDD Discovery Challenge 2004, algoritmos de mineração de dados foram utilizados para realizar a descoberta de associações escondidas, válidas e surpreendentes entre as características dos pacientes e o risco de desenvolver DCV. Dentre os vencedores do desafio, o trabalho de [Gonçalves and Plastino 2004] propõe um método para a mineração de exceções a partir da base de dados STULONG/ENTRY. Uma exceção é um tipo de regra de associação cujo suporte é significativamente superior ou inferior ao esperado em subconjuntos da base de dados que satisfazem condições específicas sobre atributos selecionados. Em uma linha semelhante, [Soulet and Hébert 2004] propõem combinar as técnicas de clusterização e padrões emergentes com o objetivo de identificar combinações de valores de atributos dos pacientes com distribuição significativamente diferente entre os diferentes clusters.

Nos anos recentes, o problema da predição de DCV tem sido tratado como um problema de classificação. Nesta abordagem, objetivo é aprender um modelo para classificar um paciente como propenso a desenvolver DCV ou não (problema binário). Embora muitos trabalhos reportem resultados de experimentos que produziram modelos com desempenho preditivo elevado, essas propostas possuem a desvantagem de utilizar classificadores do tipo caixa-preta que, ao contrário do método SCORE2 e da Calculadora de Risco Cardiovascular, não explicam as razões pela qual uma classificação é produzida. Por exemplo, [Sumwiza et al. 2023] avaliou o desempenho de diferentes algoritmos de classificação – alguns interpretáveis e outros do tipo caixa-preta – em uma base contendo dados de 769 pacientes. No experimento realizado, o algoritmo Random Forest obteve o melhor desempenho preditivo, apresentando acurácia 13% superior a obtida pela Regressão Logística, a técnica interpretável de melhor desempenho. De maneira similar, [Bhatt et al. 2023] comparou o desempenho de 4 diferentes algoritmos de classificação (Árvore de Decisão, Multi-Layer Perceptron, Random Forest e XGBoost) em uma base contendo 70.000 instâncias. No experimento realizado, o modelo de árvore de decisão (interpretável) obteve a menor acurácia e AUC, enquanto o melhor desempenho preditivo geral foi o do modelo Multi-Layer Perceptron (caixa-preta). Por fim, o trabalho de [Asif et al. 2023] faz uma revisão da literatura sobre trabalhos recentes a respeito de algoritmos de classificação para predição de DCV, identificando uma prevalência do uso de métodos de *ensemble*. Ainda no mesmo trabalho, são reportados os resultados de um experimento que comparou quatro técnicas de *ensemble* (Random Forest, Extra Trees, XGBoost e CatBoost), que geraram modelos com acurácia similar a partir de

uma base contendo dados de 389 pacientes.

Com o objetivo de produzir um modelo que seja interpretável e que ao mesmo tempo apresente desempenho preditivo tão elevado quanto o de modelos de redes neurais e *ensembles*, este artigo tem por objetivo avaliar a técnica Born-Again Tree Ensembles [Vidal and Schiffer 2020] no problema da previsão de risco de DCV.

3. BORN-AGAIN TREE ENSEMBLES

Nas áreas de mineração de dados e aprendizado de máquina, a técnica de Árvore de Decisão (AD) é provavelmente a mais conhecida e utilizada para o aprendizado de classificadores interpretáveis. Um algoritmo de AD constrói um modelo de classificação na forma de uma árvore, onde cada ramo representa um teste sobre um atributo preditivo (ex.: o paciente fuma mais de 20 cigarros por dia?) e cada folha consiste em um rótulo de classe (ex.: possui DCV ou não possui DCV). Um novo objeto é classificado através da aplicação de sucessivos testes que encontrarão um caminho na árvore, desde o nó raiz até um nó folha. A Figura 1(a) mostra um exemplo que classifica se um paciente possui DCV (Sim ou Não) em função da quantidade de cigarros que fuma por dia e de sua taxa de colesterol.

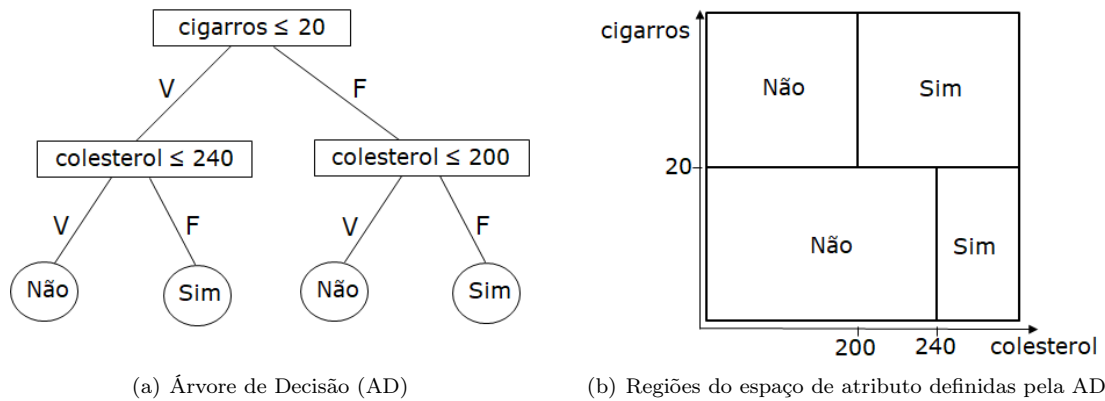


Fig. 1. (a) AD que classifica se paciente possui DCV em função da quantidade de cigarros que fuma por dia e de sua taxa de colesterol e (b) regiões do espaço de atributos que definem as classificações

Cada caminho da raiz para uma determinada folha representa uma regra de classificação específica que pode ser expressa em linguagem natural. Isso torna a AD tão simples de ser interpretada quanto os métodos tradicionais da área médica SCORE2 e Calculadora de Risco de DCV, introduzidos na seção anterior. Por exemplo, a AD da Figura 1(a) é composta pelas quatro regras listadas abaixo. A Figura 1(b) mostra que essas regras dividem o espaço de atributos em diferentes regiões ou células.

SE (cigarros \leq 20) E (colesterol \leq 240) ENTÃO (DCV = “Não”)

SE (cigarros \leq 20) E (colesterol $>$ 240) ENTÃO (DCV = “Sim”)

SE (cigarros $>$ 20) E (colesterol \leq 200) ENTÃO (DCV = “Não”)

SE (cigarros $>$ 20) E (colesterol $>$ 200) ENTÃO (DCV = “Sim”)

O algoritmo Random Forest (RF) [Breiman 2001] combina as decisões de múltiplos modelos de AD com o objetivo de aumentar a eficácia do processo de classificação. Mais especificamente, o algoritmo RF constrói um grupo (*ensemble*) de k árvores de decisão a partir da execução de k iterações sobre a base de dados de treinamento. Cada AD do grupo é gerada com o uso de subconjuntos aleatórios de atributos e instâncias através do processo de *bootstrapping*. Para prever a classe de um novo objeto, cada classificador individual vota e a classe mais popular é devolvida ao usuário.

Um modelo RF tende a possuir acurácia muito superior a uma única AD. Esta melhora na acurácia ocorre principalmente porque o modelo composto reduz a variância dos classificadores individuais [Breiman 2001]. Porém, uma importante desvantagem das RFs em relação às ADs – e que é relevante para problemas na área médica – é o fato de que uma RF não é diretamente interpretável, ao contrário do que ocorre com uma AD. Recentemente proposta em [Vidal and Schiffer 2020], a técnica Born-Again Tree Ensembles (BA) foi criada com o objetivo de resolver essa questão: trata-se do primeiro algoritmo exato capaz de transformar uma RF em uma única AD com o mesmo poder preditivo e fiel ao modelo RF em todo o seu espaço de atributos. A árvore gerada é chamada de árvore *born-again*.

O processo de transformação de uma RF em uma única árvore *born-again* é ilustrado na Figura 2 (adaptada de [Vidal et al. 2020]). Neste exemplo, é considerada uma RF composta por um conjunto de 3 árvores. As classes são representadas pelos círculos (preto e branco). Cada uma delas gera uma região que particiona o conjunto de dados em células e classifica as observações. Em seu primeiro passo, o BA processa essas regiões e define uma região única final a partir das classes majoritárias das células do conjunto de árvores. No passo seguinte, com o uso de técnicas de programação dinâmica, é construída uma região que preserva as células do conjunto de árvores, transformando-o em uma única AD. Detalhes completos podem ser obtidos em [Vidal and Schiffer 2020; Vidal et al. 2020].

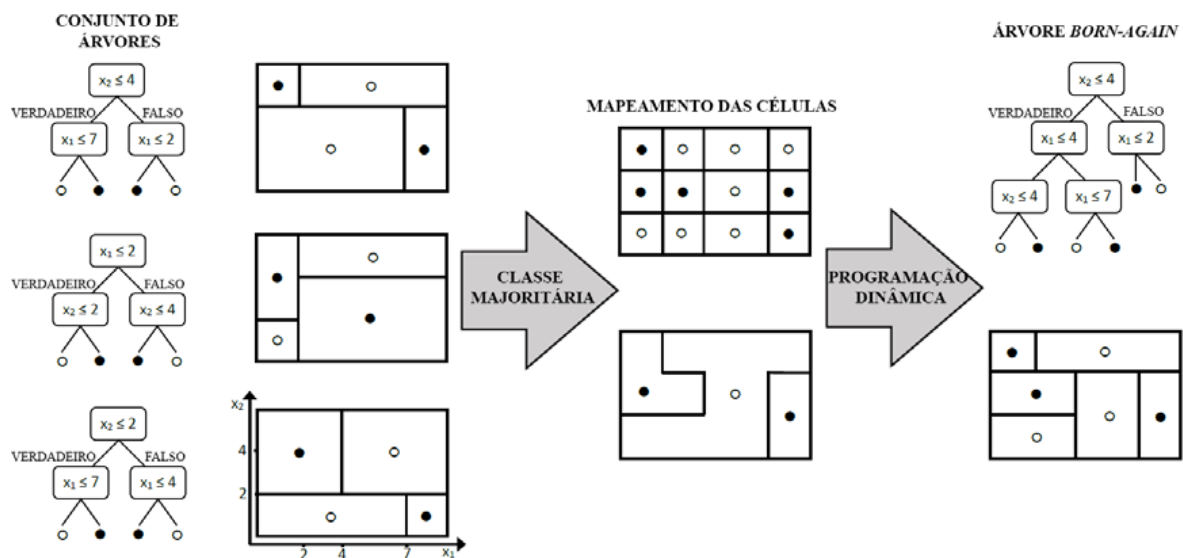


Fig. 2. Processo de transformação de uma Random Forest em uma única Born-Again Tree

4. BASE DE DADOS

A base de dados utilizada neste trabalho é a STULONG/ENTRY [Berka et al. 2007] que reúne informações de 1.249 pacientes de meia idade do sexo masculino provenientes de um estudo longitudinal sobre o desenvolvimento de fatores de risco de DCV. Esses pacientes preencheram um formulário com seus dados pessoais e hábitos gerais e também foram submetidos a exames físicos e bioquímicos. De acordo com essas informações, os pacientes foram classificados por especialistas em 3 diferentes grupos de risco: baixo (grupo 0), médio (grupo 1) e alto (grupo 2). O primeiro grupo é formado por homens sem a presença de qualquer fator de risco. O segundo, por homens com a presença de um ou mais fatores de risco, mas sem a manifestação de qualquer doença cardiovascular. E o último, é formado por homens com uma doença cardiovascular identificada.

Embora a base de dados possua 30 atributos, apenas os 6 apresentados na Tabela I foram selecionados para o experimento, uma vez que são eles os mais frequentemente apontados como fatores de risco para DCV [SCORE2 WORKING GROUP 2021; OPAS 2024a]. Observe que todos são categóricos. Com relação a distribuição de classes, a base é desbalanceada, contendo 22% de instâncias da classe 0, 69% da classe 1 e 9% da classe 2.

Atributo	Categorias
Fumante	não fumante; 1-4 cig./dia; 5-14 cig./dia; 15-20 cig./dia; 21 ou mais cig./dia
PressaoSanguinea	normal; normal/alta; alta
Colesterol	limítrofe; desejável; alto
Educacao	ensino fundamental; ensino medio; especializacao; superior
FaixaEtária	35-39; 40-44; 45-49; ≥ 50
FaixaIMC	baixo peso, normal; excesso de peso; obeso; obeso morbido
Grupo de Risco (classe)	baixo (classe 0); médio (classe 1); alto (classe 2)

Table I. Atributos da base STULONG/ENTRY utilizados no experimento

5. RESULTADOS

Nesta seção são apresentados e comentados os resultados obtidos no experimento que comparou as técnicas AD, RF e BA na base de dados STULONG/ENTRY. As implementações de AD e RF avaliadas no experimento são as disponibilizadas pelo pacote scikit-learn¹ da linguagem Python. Já o algoritmo BA está disponibilizado na página de seu projeto² e recebe como entrada uma RF gerada pela scikit-learn, produzindo como saída uma única árvore *born-again*. O algoritmo de AD implementado na scikit-learn é o CART [Breiman et al. 1984], que foi configurado com os seguintes parâmetros no experimento: máximo de oito nós folha e profundidade máxima de três níveis. Para o algoritmo de RF, foi definido o uso de dez árvores CART com a mesma configuração. Para a avaliação dos modelos foi utilizado o método *holdout*, com 90% dos dados usados para treinamento e 10% para teste. As seguintes medidas de desempenho foram utilizadas: Precisão, Revocação e F1.

5.1 Desempenho Preditivo

A Tabela II apresenta os resultados obtidos pelo modelo de árvore de decisão CART, com as medidas de desempenho indicadas para cada classe. Para o grupo 2 (risco alto), que é a classe minoritária, porém a mais importante neste problema, a Precisão obtida foi de 48% e a Revocação de 36,36%. O valor de F1 médio do modelo (média macro) é igual a 61,12%.

A Tabela III apresenta os resultados obtidos pelos modelos RF e BA. Recorde que a árvore *born-again* é uma representação diferente da mesma função de decisão da RF, por isso, ambos os modelos alcançaram o mesmo desempenho preditivo. Conforme mostra a tabela, os modelos RF e BA obtiveram resultados superiores aos do CART para todas as medidas de desempenho nas três classes. Observe que a Precisão da classe referente ao grupo de risco alto (grupo 2) aumentou em 22,5% em relação os resultados da árvore CART, enquanto a Revocação aumentou em 18,6%. A média macro da medida F1 foi de 72,42% (11% superior a que foi obtida pelo modelo CART).

Grupo de Risco	Precisão	Revocação	F1
baixo (classe 0)	0,6854	0,8500	0,7589
médio (classe 1)	0,7000	0,6300	0,6632
alto (classe 2)	0,4800	0,3636	0,4138

Table II. Desempenho por classe do algoritmo CART para as medidas de precisão, revocação e F1

¹<https://scikit-learn.org>

²<https://github.com/vidalt/BA-Trees>

Grupo de Risco	Precisão	Revocação	F1
baixo (classe 0)	0,7111	0,9600	0,8170
médio (classe 1)	0,7931	0,6900	0,7379
alto (classe 2)	0,7051	0,5500	0,6179

Table III. Desempenho por classe dos algoritmos RF e BA para as medidas de precisão, revocação e F1

5.2 Comparação das Árvores de Decisão Geradas

As ADs geradas modelos CART e BA são apresentadas nas Figuras 3 e 4, respectivamente. Observe que no modelo de AD gerado pelo CART, há somente duas regras que classificam pacientes do grupo 2 (alto risco de DCV). Por outro lado, a árvore *born-again* gerada pelo BA (proveniente da transformação de 10 árvores do *ensemble* RF) possui 34 nós folha, em que 14 representam regras que envolvem a classificação de pacientes do grupo 2. Desta forma, além de possuir melhor desempenho preditivo (igual ao do modelo RF), a árvore produzida pelo algoritmo BA é mais detalhada, tendendo a representar de forma mais fidedigna o fato de que múltiplas combinações de fatores podem estar envolvidas no desenvolvimento de DCVs.

Em resumo, na base de dados do experimento, o modelo BA obteve o mesmo poder preditivo de um modelo *ensemble* (classificador caixa-preta). Assim como ocorre com o método SCORE2 e outros tradicionalmente utilizados na área de medicina, o classificador gerado pelo BA (árvore *born-again*) produz classificações diretamente interpretáveis. No experimento realizado, a técnica BA ofereceu ainda a vantagem de ter sido capaz de realizar o aprendizado de uma AD com maior poder descritivo do que a AD produzida pelo algoritmo CART.

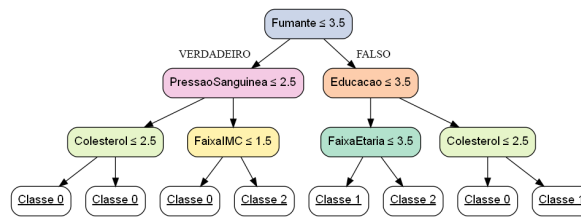


Fig. 3. Árvore de decisão gerada pelo algoritmo CART

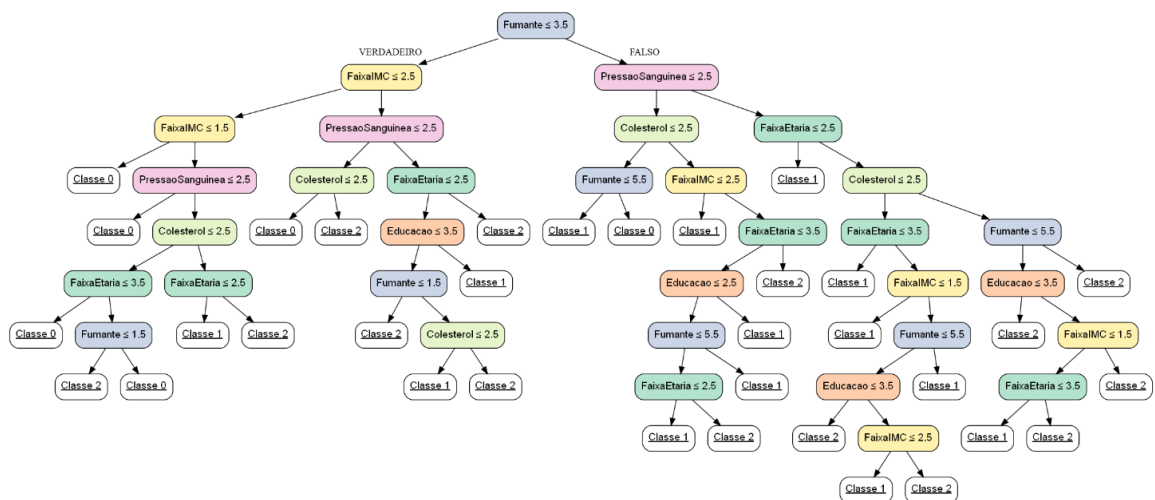


Fig. 4. Árvore de decisão gerada pelo algoritmo Born-Again Tree

6. CONCLUSÕES

Este artigo avaliou o uso do algoritmo Born-Again Tree Ensembles no problema da previsão de risco de DCV. Foi utilizada uma base de dados contendo informações a respeito de 1.417 pacientes. A árvore *born-again* produzida pelo algoritmo BA possui o mesmo poder preditivo de uma Random Forest, porém gera classificações diretamente interpretáveis, da mesma forma que uma AD comum e que os métodos baseados em escores e equações amplamente difundidos na comunidade médica. Além disso, a árvore *born-again* gerada pelo BA tende a ser mais descritiva (possuir mais regras de classificação) do que uma árvore de decisão gerada pelo algoritmo CART. Como trabalho futuro, pretende-se avaliar o BA em bases de dados mais volumosas, levando em conta não apenas o desempenho preditivo e interpretabilidade do modelo gerado, mas também questões relacionadas a eficiência do processo de transformação da RF em uma única árvore *born-again*.

REFERENCES

- ASIF, D., BIBI, M., ARIF, M. S., AND MUKHEIMER, A. Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. *Algorithms* 16 (6), 2023.
- BERKA, P., RAUCH, J., AND TOMĚČKOVÁ, M. Lessons Learned from the ECML/PKDD Discovery Challenge on the Atherosclerosis Risk Factors Data. *Computing and Informatics* 26 (3): 329–344, 2007.
- BHATT, C. M., PATEL, P., GHETIA, T., AND MAZZEO, P. L. Effective heart disease prediction using machine learning techniques. *Algorithms* 16 (88), 2023.
- BREIMAN, L. Random Forests. *Machine Learning* vol. 45, pp. 5–32, 2001.
- BREIMAN, L., FRIEDMAN, J., STONE, C. J., AND OLSHEN, R. A. *Classification and Regression Trees*. Taylor & Francis, 1984.
- DA SILVA, V. R. AND GONÇALVES, E. C. Análise Comparativa de Implementações de Algoritmos de Árvores de Decisão para Aplicações no Serviço Público. In *Anais da Escola Regional de Computação Bahia, Alagoas e Sergipe (ERBASE)*. Maceió, Brazil, pp. 10–19, 2021.
- DE OLIVEIRA, G. M. M., BRANT, L. C. C., POLANCZYK, C. A., MALTA, D. C., BIOLO, A., NASCIMENTO, B. R., DE FATIMA MARINHO DE SOUZA, M., LORENZO, A. R. D., DE PAIVA FAGUNDES JÚNIOR, A. A., SCHAAN, B. D., DE SOUZA E SILVA, C. G., DE CASTILHO, F. M., CESENA, F. H. Y., SOARES, G. P., JUNIOR, G. F. X., FILHO, J. A. S. B., PASSAGLIA, L. G., FILHO, M. M. P., MACHLINE-CARRION, M. J., BITTENCOURT, M. S., NETO, O. M. P., VILLELA, P. B., TEIXEIRA, R. A., STEIN, R., SAMPAIO, R. O., GAZIANO, T. A., PEREL, P., ROTH, G. A., AND RIBEIRO, A. L. P. Estatística Cardiovascular - Brasil. *Arq. Bras. Cardiol.* 121 (2): 1–131, 2024.
- DOSHI-VELEZ, F. AND KIM, B. Towards A Rigorous Science of Interpretable Machine Learning. *arXiv: Machine Learning*, 2017.
- ECML-PKDD. ECML/PKDD2004 Discovery Challenge. <https://sorry.vse.cz/berka/challenge/ecmlpkdd2004/Chall04-prog.htm>, 2004.
- FREITAS, A. A. Comprehensible Classification Models: A Position Paper. *SIGKDD Explor. Newsl.* 15 (1), mar, 2014.
- GONÇALVES, E. C. AND PLASTINO, A. Mining Strong Associations and Exceptions in the Stulong Data Set. In *Proceedings of the ECML/PKDD Discovery Challenge*. Pisa, Italy, pp. 44–55, 2004.
- OPAS. Calculadora de risco cardiovascular. <https://www.paho.org/pt/hearts-nas-americas/calculadora-risco-cardiovascular>, 2024a.
- OPAS. Organiza ção Pan-Americana da Saúde: Doen cas Cardiovasculares. <https://www.paho.org/pt/topicos/doencas-cardiovasculares>, 2024b.
- RUDIN, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence* vol. 1, pp. 206–215, 2019.
- SCORE2 WORKING GROUP. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *European Heart Journal* 42 (25): 2439–2454, 2021.
- SOULET, A. AND HÉBERT, C. Using Emerging Patterns from Clusters to Characterize Social Subgroups of Patients Affected by Atherosclerosis. In *Proceedings of the ECML/PKDD Discovery Challenge*. Pisa, Italy, pp. 104–115, 2004.
- SUMWIZA, K., TWIZERE, C., RUSHINGABIGWI, G., BAKUNZIBAKE, P., AND BAMURIGIRE, P. Enhanced Cardiovascular Disease Prediction Model Using Random Forest Algorithm. *Informatics in Medicine Unlocked* vol. 41, pp. 101316, 2023.
- VIDAL, T., PACHECO, T., AND SCHIFFER, M. Born-Again Tree Ensembles. *arXiv* vol. arXiv:2003.11132, 2020.
- VIDAL, T. AND SCHIFFER, M. Born-Again Tree Ensembles. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.