

COMPARAÇÃO DE 2 MODELOS DE RISCO CARDIOVASCULAR BASEADOS EM MÉTODOS CONVENCIONAIS COM 5 BASEADOS EM APRENDIZADO DE MÁQUINA

Joaquim Guanaes Gomes¹; Dr^a Luciana Campos Baltatu²; Dr Ovidiu Constantin Baltatu³

Universidade Anhembi Morumbi

Engenharia Biomédica, São José dos Campos, SP

joaquimguanaes@yahoo.com

Introdução

Doenças cardiovasculares (DCV) continuam sendo uma das principais causas de morbidade e a principal causa de mortalidade em todo o mundo, contribuindo significativamente para a perda de saúde e acréscimo dos custos dos sistemas de saúde (1). A avaliação precisa e a estratificação do risco cardiovascular são essenciais para estratégias eficazes de sua prevenção e tratamento. O desenvolvimento e amplo uso de escores para avaliação do risco de DCV é uma das principais recomendações das instituições geradoras de diretrizes de combate as DCV (2). No entanto, esses modelos geralmente desenvolvidos com metodologias convencionais baseadas em técnicas de regressão se concentram em um conjunto limitado de fatores de risco e mostraram precisão preditiva variável muitas vezes e, portanto, com confiabilidade questionável e abaixo do desejável.

Objetivos

O presente trabalho tem objetivo de comparar a performance de dois modelos de risco cardiovascular baseados em metodologias estatísticas convencionais de regressão (Regressão Logística e Regressão Logística de Lasso) e 5 modelos de risco cardiovascular baseados em algoritmos de aprendizado de máquina.

Metodologia

Os dados usados neste estudo foram criados para abranger um variedade bastante ampla de condições de saúde cardiovascular (DCV) e compreende 100.000 registros com fatores de riscos tais como: eventos fatais e não fatais de DCV, gênero, idade, se é fumante ou não, pressão arterial sistólica típica, nível de colesterol total, colesterol HDL e histórico familiar de DCV.

Resultados

Desempenho de Classificação dos Algoritmos

Algoritmo	Sensibilidade	Especificidade	Precisão
Regressão Logística	0,820	0,904	70,1
Regressão Logística de Lasso	0,845	0,909	77,2
Decisão de Arvore	0,940	0,935	87,5
Random Forest	0,949	0,942	89,2
Gradient Boosted Machine	0,968	0,950	91,3
Extreme Gradient Boosted	0,970	0,958	92,5
KNN	0,975	0,960	92,8

AUC e Estatística C dos Algoritmos

Algoritmo	AUC	95% IC AUC	Índice C
Regressão Logística	0,820	0,915	0,748
Regressão Logística de Lasso	0,845	0,929	0,756
Decisão de Arvore	0,940	0,935	0,801
Random Forest	0,949	0,942	0,870
Gradient Boosted Machine	0,968	0,950	0,891
Extreme Gradient Boosted	0,970	0,958	0,915
KNN	0,975	0,960	0,926

Conclusões

Todos cinco modelos de risco desenvolvidos por algoritmos de aprendizado de máquina tiveram melhor capacidade preditiva para eventos cardiovasculares e melhores indicadores de classificação do que os desenvolvidos por métodos regressão. Quando o objetivo é gerar um modelo que previsão com mais precisão, algoritmos de aprendizado de máquina pode ser vantajoso em relação aos métodos de regressão tradicionais. Tais algoritmos podem ser empregados para ajudar a confrontar múltiplas questões que afetam adversamente desenvolvimento de previsão de risco cardiovascular como fatores de riscos correlacionados, com relacionamentos não lineares e com interações entre os fatores de riscos e desfechos.

Apoio Financeiro: O presente trabalho conta com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Os autores agradecem ao Instituto Ânima e ao CITÉ pelo apoio concedido ao projeto de pesquisa.

