

Trabalhos Científicos

Título: Modelagem Estatística E Resistência À Insulina: Aplicação De Algoritmos De Classificação Em Dados Clínicos De Adolescentes.

Autores: DANIEL INGRISANI BRANCO (CHC-UFPR), GIOVANNI ARAUJO BACOCCHINA (CHC-UFPR), BEATRIZ ELIZABETH BAGATIN VELEDA BERMUDEZ (CHC-UFPR)

Resumo: O uso recente do aprendizado de máquina nas áreas médicas possibilitou a melhora do fluxo de trabalho do clínico e a existência de uma nova ferramenta de auxílio médico. A aplicação desses modelos é emergente, com potencial de prever desfechos clínicos. O atual trabalho propõe a aplicação de algoritmos complexos de aprendizado de máquina em dados clínicos de exames laboratoriais de rotina para prever o desfecho resistência à insulina (RI) em adolescentes, auxiliando o médico pediatra em seu fluxo de trabalho e evitando exames específicos para determinar a RI, como a insulinemia, mandatório para o cálculo do HOMA-IR. Foram coletadas 105 observações, de adolescentes de um ambulatório de hebiatria, e 8 variáveis, sendo essas idade, glicemia em jejum, colesterol total, triglicérides, HDL, LDL, 25(OH)D e a variável alvo, HOMA-IR. Elas foram selecionadas por integrarem os exames de rotina realizados em pediatria ambulatorial. Foram excluídos do trabalho os valores de 25(OH)D, pela ausência de mais de 10% dos dados, sendo um consenso na literatura que análises estatísticas tendem ao viés quando atingem tal valor. Não há dados faltantes nos outros features utilizados. Para fins de modelagem estatística, foram utilizadas as funções Máquina de Vetores de Suporte (SVM), Elastic Net, K-vizinhos mais próximos (k-NN) e Naive Bayes. Para obtenção de estimativa sem viés do desempenho de generalização das funções, foi aplicada 5-Fold CV (validação cruzada). Para o modelo regularizado por Elastic Net, os valores de ROC-AUC médio das validações cruzadas (CV), exatidão média das CV e precisão média das CV, respectivamente com os devidos valores de desvio padrão (963,), foram de 0.73 (963,=0.12), 0.68 (963,=0.10) e 0.63 (963,=0.21). Para a função SVM, esses mesmos dados resultaram em 0.78 (963,=0.14), 0.72 (963,=0.07) e 0.79 (963,=0.17). Em seguida, o modelo k-NN performou 0.78 (963,=0.08), 0.7 (963,=0.05) e 0.74 (963,=0.14). Por fim, Bayes atingiu, para os valores acima e nesta ordem referidos, 0.67 (963,=0.08), 0.63 (963,=0.07) e 0.56 (963,=0.14). Dos modelos treinados, 50% apresentaram resultados favoráveis. A literatura sugere como valores clinicamente úteis os acima de 0.80, linha de corte próxima da alcançada pelo SVM e k-NN. No entanto, as funções Elastic Net e Bayes apresentaram um resultado não satisfatório, com ROC-AUC respectivamente pobre e ruim. Sugerimos como explicação a não pressuposição de linearidade que SVM e k-NN realizam, podendo aprender fronteiras de decisão mais complexas. Por outro lado, o classificador de Bayes assume independência condicional entre as variáveis e o modelo de regularização por Elastic Net pode depender de uma base de dados com um maior tamanho absoluto, não detectando com eficiência os padrões presentes. Os resultados são promissores, podendo se tornar instrumento auxiliar no atendimento pediátrico ambulatorial em pesquisas posteriores, com um número de observações maior.