**A CONTRIBUIÇÃO DE FRAMEWORKS ESTATÍSTICOS PARA O DIAGNÓSTICO MÉDICO**

¹Dionisio Alves da Silva Neto; ²Paulo Cerqueira dos Santos Júnior

1Acadêmico de Estatística, Universidade Federal do Pará (UFPA), Belém, Pará, Brasil; 2Professor Doutor em Estatística, Universidade Federal do Pará (UFPA), Belém, Pará, Brasil.

**Eixo Temático:** Empreendedorismo na Medicina.

**E-mail do Autor Principal:** dionisioneto899@gmail.com

**Introdução:** Modelos de aprendizado de máquina supervisionado são métodos inferenciais que podem prever um valor numérico ou dicotômico baseando-se em padrões já observados. Porém, ainda é escassa a existência de *softwares* livres, os quais auxiliem médicos e demais profissionais da saúde no diagnóstico de doenças. **Objetivo**: O projeto tem por objetivo oferecer uma plataforma gratuita, embasando-se no processo de aprendizado estatístico, para a classificação das doenças: diabetes, hipertensão e ataque cardíaco, por intermédio da informação de covariáveis explicativas. **Metodologia:** O banco de dados utilizado coleciona três tabelas a respeito da avaliação de médicos especialistas para o diagnóstico da hipertensão, do ataque cardíaco e da diabetes para 26.082, 26.082 e 70.691 pacientes, nesta respectiva ordem, com a informação de covariáveis que estão relacionadas às contrações das doenças. Para avaliar os doentes e sadios, utilizou-se do algoritmo supervisionado denominado *LIGHTGBM*. Ademais, para encontrar as configurações que auxiliaram na previsão adequada do modelo escolhido, utilizou-se do processo de otimização bayesiana. Por último, elaborou-se um autêntico *framework* estatístico, por meio da linguagem Python. **Resultados e Discussão:**  Primeiramente, para a construção do diagnóstico da hipertensão, o classificador foi eficaz em atingir uma acurácia de 100% após utilizamos do método bayesiano para encontrar a melhor configuração da predição. Com isso, ao inserirmos as covariáveis relacionadas aos biotipos dos pacientes e resultados de exames de rotina, conseguimos evitar a presença de falsos positivos ou falsos negativos dentro da matriz de confusão dos dados de teste do modelo. Em adição, para o diagnóstico do ataque cardíaco, também foi possível uma acurácia de 100% nas classificações ao utilizarmos das covariáveis relacionadas aos perfis dos pacientes, localidade e histórico familiar. Por consequência, a matriz de confusão evidenciou uma perfeita predição dos pacientes sadios e doentes na amostra de teste do modelo. Em contraste, para o cenário de predição da diabetes, o processo retornou uma acurácia de 75,17%, com o suporte das covariáveis relacionadas aos perfis dos pacientes e histórico familiar. A matriz de confusão nos dados de teste informou que, do total dos pacientes no conjunto de avaliação, cerca de 10,41% foram falsos positivos e 14,42% foram falsos negativos. Por fim, todas as variáveis utilizadas foram inseridas em uma estrutura do tipo formulário e, após a inserção dos valores, foi possível obter uma classificação para cada doença acompanhada do valor de acurácia como suporte para a tomada de decisão. **Considerações Finais:** Neste trabalho foram abordados três casos de diagnóstico de doenças pertinentes na saúde humana. Para a avaliação da hipertensão e ataque cardíaco, foi alcançado uma acurácia de excelência e nenhum erro na matriz de confusão. Para a diabetes, o classificador retornou uma acurácia menor, porém com resultados ainda satisfatórios. Por fim, por meio de códigos e estruturas interativas, construiu-se uma plataforma gratuita através da inserção de dados e previsão instantânea da ocorrência de cada uma das enfermidades.

**Palavras-chave:** Aprendizado de Máquina; Aprendizado Supervisionado; Tecnologia; Otimização Bayesiana; Website.

**Referências**

FERNEDA, E. Redes neurais e sua aplicação em sistemas de recuperação de informação. **Ciência da Informação**, v. 35, n. 1, p. 25-30, 2006.

KE, G.; MENG, Q.; FINLEY, T.; WANG, T. CHEN, W.; MA, W.; YE, Q.; TIE-YAN, L. Lightgbm: A highly efficient gradient boosting decision tree. In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31., 2017. **Anais**… Long Beach: Curran Associates, 2017. p. 1-9.

SNOEK, J.; LAROCHELLE, H.; ADAMS, R. P. Practical bayesian optimization of machine learning algorithms. In: ADVANCES IN NEURAL INFORMATION, 25., 2007. **Anais**… London: Curran Associates, 2007. p. 1-9.

TAKAKURA, A. M.; PERREIRA, D. R.; SILVA, F. A.; PAZOTI, M. A.; ALMEIDA, L. L.; SAPIA, H. M. PA Uso do aprendizado de máquina no diagnóstico médico de patologias. **Colloquium Exactarum**, v. 10, n. 1, p. 78-89, 2018.