

Projeto eHealth

Análise de Dados para Aterosclerosis

Equipe: Lucas Ismaily B. Freitas - ra: 142696

Marcos Felipe de Menezes Mota - ra:211893

1. Introdução

Aterosclerose (coronary artery disease), é uma doença do coração acarretada, normalmente, pelo depósito de gordura nas paredes das principais artérias do coração. Atualmente, a forma mais precisa de detectar a aterosclerose é através é através de um angiograma, no entanto esse é um exame caro, em média US\$ 8.911, que é invasivo. Por isso, médicos vem registrando diferentes sintomas e características que indicam a doença sem ter que fazer um angiograma ou pelo menos quantificar qual seria o melhor caso para indicar o exame e diminuir custos.

Dessa forma, o projeto se propõe a usar redes Bayesianas tanto para aprendizado supervisionado, como para quantificar incerteza e dar suporte a decisão na prescrição de um exame. A ferramenta utilizada para realizar a análise proposta vai ser o Weka, pois já fornece uma grande quantidade de métodos de análise de dados implementado e capaz integrar com outros ambientes de programação, como por exemplo o Python.

2. Fonte dos Dados

Os dados a serem utilizados, são duas bases distintas com diferentes características que podem indicar a presença ou não de aterosclerose e os respectivos resultados do angiograma.

- *Extention of Z-Alizadeh sani dataset* : Possui 55 características, entre eles o resultado do angiograma e qual das 3 artérias foi encontrada o depósito de gordura. A base de dados possui 303 entradas. Url: https://www.researchgate.net/publication/311582821_extention_of_Z-Alizadeh_sani_dataset

- *Kaggle - Heart.csv* : Semelhante a primeira base de dados, no entanto possui apenas 14 características para 304 indivíduos. Url:
<https://www.kaggle.com/zhaoyingzhu/heartcsv/data>