Trey Hibbard

Executive Summary – Cardiovascular Heart Disease Classifier Models

---

Research Objective:

Using data from a thousand patients with varying conditions, the aim of this research was to build the best classifier model for predicting cardiovascular heart disease with both full medical information, as well as with the limited information a person could acquire at home. Additionally, a preference for any error to be type one is preferred, as the consequences for improperly treating heart disease can be fatal.

Findings:

Research finds that slope, chest pain, and resting blood pressure are two of the most important predictors in identifying cardiovascular heart disease. Luckily, two of these predictors, chest pain and resting blood pressure, are measurable at home, and being able to leverage this can inform a person if their risk of heart disease is as high as they may suspect.

When using the full list of variables available in a medical environment, a standard logistic regression model can carry high predictive accuracy of ~0.97%. An at home assessment reliant mostly on chest pain and resting blood pressure is much more accurate using a random forest model.

Takeaways:

While both models show high predictive accuracy, medical advice should be well informed and data driven. An at home assessment may provide a person with confidence that their condition is worth checking with a doctor, but the additional 7% accuracy from the medical assessment is a strong lead in predictive accuracy that can appropriately inform patients in need of help.

Data Source:

Data is sourced from:
https://data.mendeley.com/datasets/dzz48mvjht/1

Resources include a csv file and a pdf detailing all of the variable types and labeling.