

Predicting Risk for Cardiovascular Disease

PROBLEM STATEMENT

Cardiovascular disease is the leading cause of death annually worldwide. Cardiovascular diseases can, however, be managed if caught early and simple lifestyle changes are made. This project would explore a set of data for patients measuring known factors for heart disease to develop a machine learning model to predict risk of developing heart disease within the next ten years.

THE CLIENT

A small clinic in Framingham, Massachusetts has seen a decided spike in cases of cardiovascular disease in their patients. Rather than sit back and wait for illness to strike, the lead cardiologist has decided to be proactive. They've reached out to us to review their clinical data and develop a model to predict their patient's risk of developing cardiovascular disease in the next 10 years. The physician has two goals in mind: help identify those at risk and start acting before things get worse, and determine which patients are statistically more likely to develop problems and provide them with educational material before things become a problem. The model developed from their patient dataset will be used to do exactly that. A statistical model will be used to develop a patient education program, targeted at specific potential risk patients, and to help bring awareness surrounding cardiovascular disease before it becomes a problem. A machine learning model will also be used to target identified "at risk" patients and help them combat more severe cardiovascular disease by targeting treatment and making recommended life-style changes for those patients who are identified as being at risk.

QUESTIONS ADDRESSED

1. What is the most significant predictor of risk for cardiovascular disease
2. How much is the risk of developing heart disease increases if you are both diabetic and a smoker
3. Does this demographic show a higher risk of heart disease as related to gender
4. Generate data visualization to gain actionable insights about the topic
5. How does this clinic's patient set compare to the known statistics regarding cardiovascular disease

DATASET

Data will be pulled from the HME Workshop in Oct 3, 2020 located on Kaggle:

- <https://www.kaggle.com/christofel04/cardiovascular-study-dataset-predict-heart-disea>
- Consists of both a train and a test file, 15 attributes, 4000+ records

APPROACH

- Clean the data to review any potential incomplete entries
- Complete EDA to establish a clear understanding of the dataset
- Generate figures and visualize details to determine there are actionable items
- Perform statistical analysis to determine significant features in this dataset
- Research and compare statistical facts for cardiovascular disease to this clinic's patients
- Develop model to predict cardiovascular disease
- Develop recommendation for preemptive patient education plan

DELIVERABLES

1. Report and presentation with findings from analysis and comparison to known Cardiovascular Disease Statistics
2. Code for predictive model