Problem Statement Worksheet

By December 2022, what can be developed that informs Great Lake Hospital whether a person will have a cardiovascular disease based on their characteristics, habits, or other factors?

H

1 Context

"Cardiovascular diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year…Identifying those at highest risk of CVDs and ensuring they receive appropriate treatment can prevent premature deaths" (WHO, 2022). Great Lakes Hospital is starting a new campaign to prevent heart disease by identifying those with the greatest risk and intervening before a CVD occurs.

2 Criteria for success

By December 2022, Great Lakes Hospital will have a model that predicts whether a person will have a cardiovascular disease.

Deliverables: A GitHub repo containing each step of the project, a slide deck, and a project report.

3 Scope of solution space

We will find the best model (multiple linear regression, logistic regression, SVM, random forest, etc.) to predict heart disease through training and testing and cross-validating, including feature selection and principal component analysis.

4 Constraints within solution space

- The data may not be a representative sample of the target population
- People may not have given honest answers, intentionally or unintentionally
- There may be factors that cannot be changed (height, gender, age, etc)

5 Stakeholders to provide key insight

- Great Lakes President
- Great Lakes Vice-President
- Marketing Director
- Analytics Director

6 Key data sources

Dataset from the CDC from 2020 taken from a nationwide telephone survey about US residents' health status

H D E I P