

PROJECT 1 for AI-Bootcamp MSU-edX – October 23rd 2023 to April 22nd 2024

Team Name:

PROJECT_1_Team_DecisionMakers

Team Members:

Ø Jesse Kranyak

Ø Jeffery Boczkaja

Ø Mohamed Altoobli

Ø Siriesha Mandava

Project Theme: Heart Disease or Cardiovascular Disease

Heart Attack Prediction Metrics

This project focuses on the factors that predict heart attacks in adults. To determine the factors that predict heart attacks in adults several data sets will be utilized and cleaned in this project.

Questions we hope to answer in this project;

1. What is the most important factor in determining heart disease?
2. What are the highest correlated factors for heart disease?
3. What groups are the most at risk for heart attacks?
4. What could be the most important choices made to reverse the potential negative outcomes of heart disease?
5. How will this data help hospitals to make informed decisions and develop effective business plans and strategies?

Who is the project for?

Insurance company

Hospital Doctors, Researchers, Business executive team

About our Data

1. * Framingham Data

<https://www.kaggle.com/datasets/aasheesh200/framingham-heart-study-dataset/data>

"The Framingham heart disease dataset includes over 4,240 records, 16 columns and 15 attributes.

The goal of the dataset is to predict whether the patient has 10-year risk of future (CHD) coronary heart disease"

2. * CDC Data

Our CDC dataset

(<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>) comes from kaggle as well. This is a partially cleaned dataset where the author has reduced the data points from over 300 to the 40 most applicable for heart disease.

Here is a link to the repository of the original data and how it was cleaned

(https://github.com/kamilpytlak/data-science-projects/blob/main/heart-disease-prediction/2022/notebooks/data_processing.ipynb)

****Licensing of Data****

* This specific sampling of the Framingham dataset has ambiguous licensing. The licensing when downloaded from kaggle is stated as 'unknown', full use of the Framingham dataset requires an application and approval of research. The full data set covers a much larger dataset than the sampling we have accessed off kaggle.com.

* CDC dataset has CC0: Public Domain licensing

****References****

This data set dates from 1988 and consists of four databases: Cleveland, Hungary, Switzerland, and Long Beach V. It contains 76 attributes, including the predicted attribute, but all published experiments refer to using a subset of 14 of them. The "target" field refers to the presence of heart disease in the patient. It is integer valued 0 = no disease and 1 = disease.

<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset?resource=download>

[Heart Disease Dataset](#)