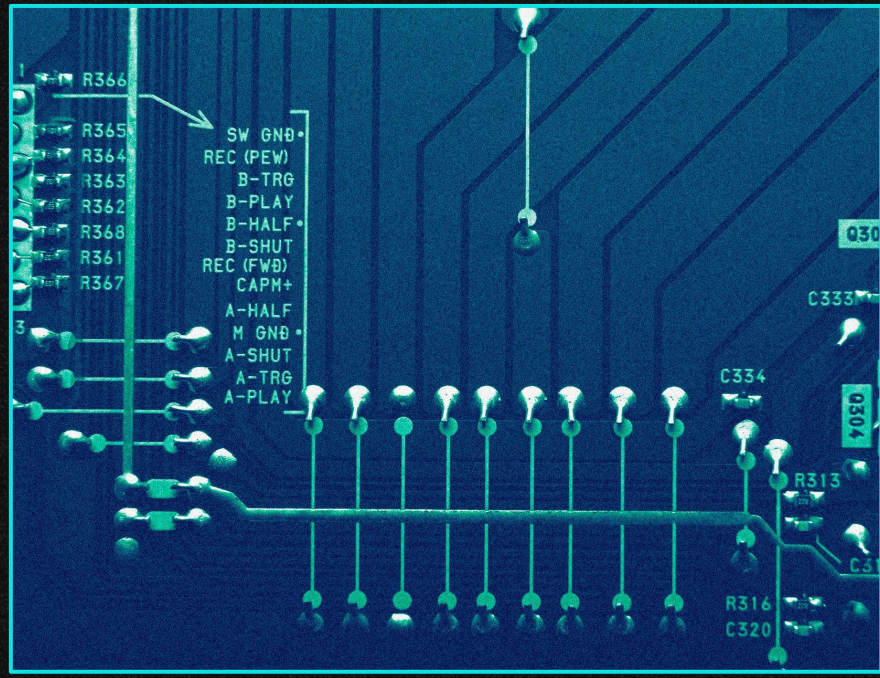
ILLINOIS DATA SCIENCE CLUB

Heart Disease Classification

DATA DIVE



By Team CWMDSJ

Table of contents

Problem Identification 01

Idk

Collecting Data 02

Idk

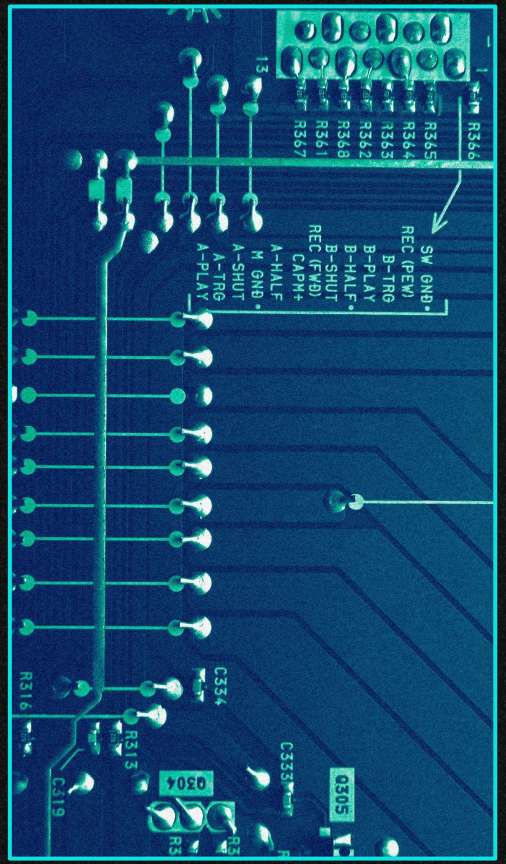
Processing Data 03

Idk

Modeling Data 04

idk

Data Analysis 05



Problem Identification

- Why heart disease ?
  - The leading cause of death for both men and women in U.S.
  - Is heart disease preventable?

- To solve the questions? We should know:
  - What features are correlated with heart disease ?
  - How to classify the features that are correlated with heart disease?

Collecting Data

- Dataset from UC Irvine
  - Cleveland Database
- 14 Columns (more detail on next slide)
  - 13 possible explanatory variables
  - 1 response variable
-

Column Descriptions

- Age - age in years
- Sex - (1 = male; 0 = female)
- CP - chest pain type
- Trestbps - resting blood pressure (in mm Hg on admission to the hospital)
- Chol - serum cholesterol in mg/dl
- FBS - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- Restecg - resting electrocardiographic results
- Thalach - maximum heart rate achieved
- Exang - exercise induced angina (1 = yes; 0 = no)
- Oldpeak - ST depression induced by exercise relative to rest
- Slope - the slope of the peak exercise ST segment
- Ca - number of major vessels (0-3) colored by fluoroscopy
- Thal - 1 = normal; 2 = fixed defect; 3 = reversible defect
- Num - artery diameter (0-4)

Data Cleaning

Our data was in the wrong:

- Double click on it
- Type your new text in the dialog box
- You can change the fill color in the top menu
- You can also modify the stroke, its color, thickness or type
- Press Enter

[insert pic of error]

Machine Learning Model

Exploratory Data Analysis

Feature Analysis & Statistical Analysis