Project Report

Heart Attack Analysis & Prediction

GitHub URL

<https://github.com/ajaymalik8/UCDPA_AjayMalik.git>

Abstract

Cardiovascular diseases (CVDs) are the leading cause of death globally. The number of heart attacks is rising in India as well.

In this Study, we will explore...

* The heart disease dataset using exploratory data analysis (EDA)
* Exercise with classification algorithms for prediction (modelling)

I selected this data set to analyse the vital and sound an early warning to lessen its impact.

Refer project code file **GitHub >> Project\_Report\_AjayMalik\_UCDPA\_BatchC**

Analysis:

Data size: 303 rows and 14 columns (13 independent + one target variable)

* Found 1 Duplicate record
* Data has no missing values

Features (columns) data type:

Six features are numerical

The rest (seven features) are categorical variables

Target variable is fairly balanced, 54% no-disease to 46% has-disease

Correlations:

Correlation between features is weak at best

From the numerical features num\_major\_vessels, max\_heart\_rate\_achieved and st\_depression are reasonabily fairly correlated with the target variable at -0.47, 0.43 and -0.43 correlation coefficient respectively.

From the categorical features chest\_pain\_type, num\_major\_vessels, thalassemia, and exercise\_induced\_angina are better correlated with the target variable, thalassemia being the highest at 0.52.

Cholestrol (to my surprize, but what do I know?) has less correlation with heart desease.

Takeaway: features that have higher predictive power could be, chest\_pain\_type, num\_major\_vessels, thalassemia, exercise\_induced\_angina max\_heart\_rate\_achieved and st\_depression. We will see which features will appear as imporatnt by the classification models.

Introduction

The phrase "heart disease" is a general one that refers to various illnesses and ailments that affect the heart and circulatory system. They are also known as cardiovascular illnesses. It is a major cause of disability all around the world. Since the heart is one of the body's most important organs, ailments that affect it also impact other organs and body parts. Heart disorders come in a variety of shapes and sorts. The most frequent ones result in heart failure and heart attacks by narrowing or blocking the coronary arteries, altering the heart's valves, growing the size of the heart, among other effects."

Dataset

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

Data Dictionary

age - Age of the patient

sex - Sex of the patient

cp - Chest pain type ~ 0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic

trtbps - Resting blood pressure (in mm Hg)

chol - Cholestoral in mg/dl fetched via BMI sensor

fbs - (fasting blood sugar > 120 mg/dl) ~ 1 = True, 0 = False

restecg - Resting electrocardiographic results ~ 0 = Normal, 1 = ST-T wave normality, 2 = Left ventricular hypertrophy

thalachh - Maximum heart rate achieved

oldpeak - Previous peak

slp - Slope

caa - Number of major vessels

thall - Thalium Stress Test result ~ (0,3) (stresstest)

exng - Exercise induced angina ~ 1 = Yes, 0 = No

output - Target variable

Information About the data set

Implementation Process

Results

Insights

NOTE: the data set size was very small need to perform similar analysis on extended data set.

References

1. [An introduction to seaborn — seaborn 0.12.1 documentation (pydata.org)](https://seaborn.pydata.org/tutorial/introduction.html)
2. UCD Curriculum
3. Data Camp
4. Stackoverflow
5. GeeksforGeeks