Heart Disease Risk Assessment

Discover your risk of developing heart disease through our comprehensive data-driven analysis. This project aims to empower you to make informed decisions about your health and preventive measures.

by Adla Abou Steiti

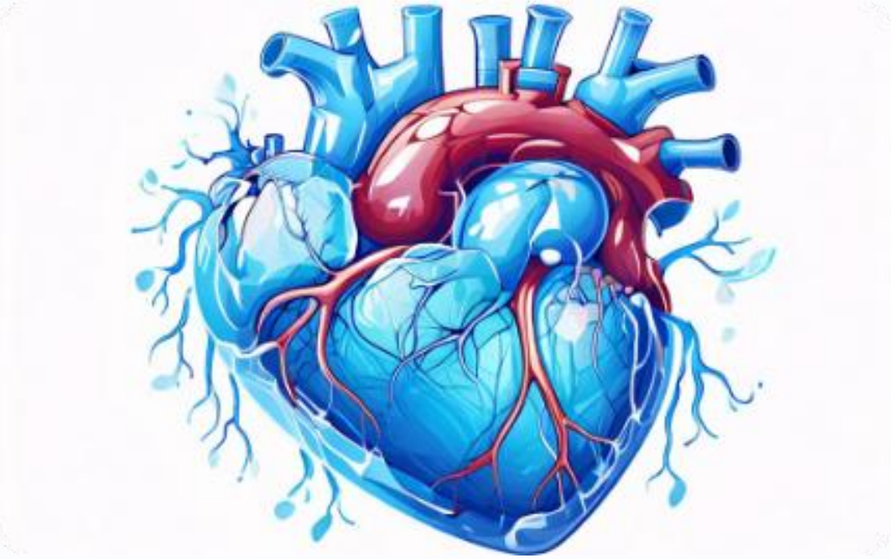
Project Overview

Research Question:

Can machine learning models accurately predict the risk of heart disease based on key health indicators?

Objective:

To develop a machine learning model to predict the risk of heart disease and assist in early detection and prevention



Importance:

Heart disease is a leading cause of mortality worldwide, early detection and prevention are critical

Data Collection

❑ Description of the dataset:

Contains anonymized information on patient demographics, medical history, and health metrics including the health indicators of interest: Blood Pressure, Cholesterol Level, Body Mass Index (Calculated)

❑ Data cleaning and preprocessing:

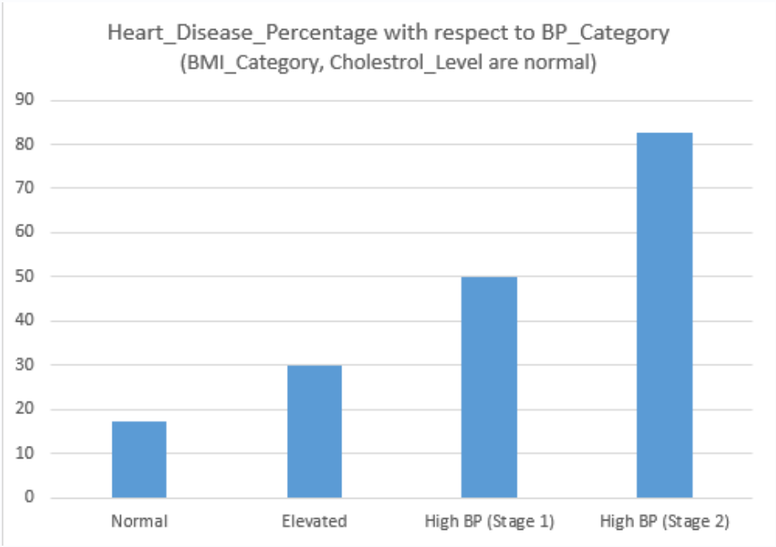
Ensuring no missing values, or outliers.
Calculating essential values based on given data
Encoding categorical variables for BMI, and BP

❑ Source of the dataset:

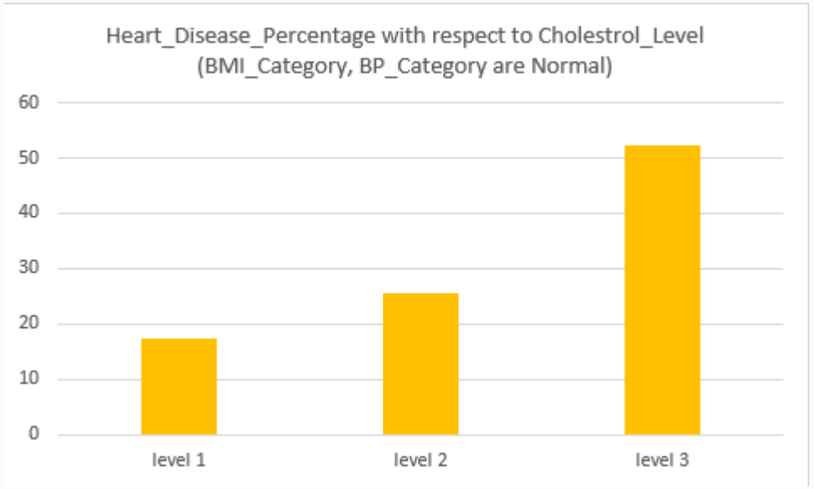
Kaggle

Data Analysis

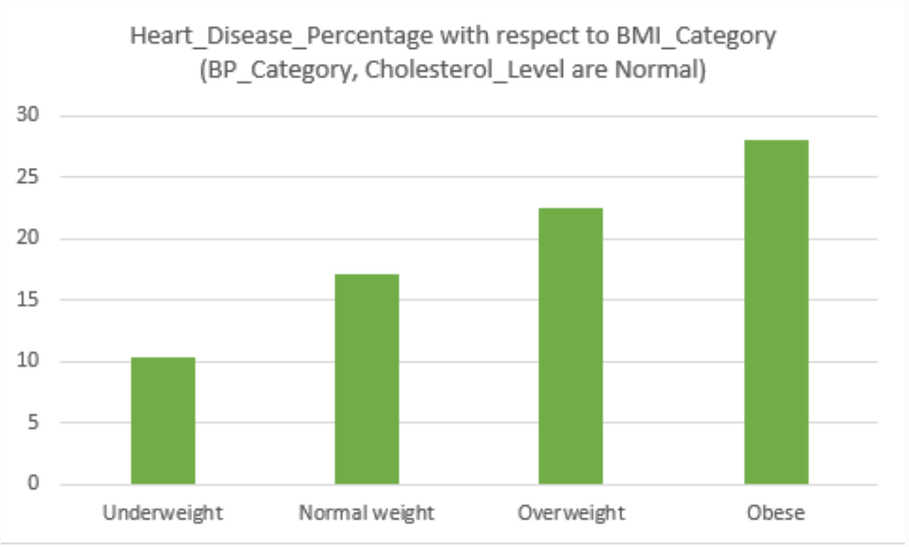
Each factor was analyzed to check correlation with heart disease risk percentage while the other factors are normal.
The result showed a positive correlation between each factor and heart disease risk percentage.



Blood Pressure



Cholesterol Levels



Body Mass Index (BMI)

Interesting Findings

- The highest risk percentage of developing heart disease was found unexpectedly in underweight individuals with high blood pressure (Stage 2) and cholesterol level 3 (90.9%).
- This risk was even higher than that in obese individuals with the same high blood pressure and cholesterol level (83.17%).

! Amplified Risk for Underweight Individuals:

- Underweight individuals showed a higher-than-expected risk when accompanied by high blood pressure (Stage 1 or 2).
- A possible interpretation is that being underweight amplifies the impact of high blood pressure, leading to a higher risk of developing heart disease.

❑ Need for Further Investigation

needed to better understand the reasons behind the amplified impact of high blood pressure in underweight individuals.

Decision Tree Model

Purpose

To predict the risk of heart disease based on health indicators (BMI, Blood Pressure, Cholesterol Level).

Training Process

Data Splitting: Dataset split into 80% training and 20% testing.
Feature Selection: Features: BMI, BP_Category, Cholesterol_Level
Target: Presence of heart disease (Cardio)

Evaluation Metrics

```
Accuracy: 0.73
Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.77      0.74     10461
           1       0.75      0.69      0.72     10539

    accuracy                           0.73     21000
   macro avg       0.73      0.73      0.73     21000
weighted avg       0.73      0.73      0.73     21000

Confusion Matrix:
[[8022 2439]
 [3295 7244]]
```

Lifestyle Modifications for Risk Reduction

1

Dietary Changes

Adopt a heart-healthy diet low in saturated and trans fats, with more fruits, vegetables, whole grains, and lean proteins to improve cholesterol and blood pressure levels.

2

Regular Exercise

Engage in at least 150 minutes of moderate-intensity aerobic activity per week to strengthen the cardiovascular system and manage weight.

3

Stress Management

Incorporate stress-reducing techniques like meditation, yoga, or deep breathing to lower blood pressure and reduce the strain on the heart.

Conclusion and Next Steps

- Summary of key findings: Successful development of a decision tree model for heart disease risk prediction
- Importance of the project: Contribution to early detection and prevention of heart disease, potential to save lives
- Opportunities for improvement: Incorporating additional features, refining model hyper parameters, exploring ensemble methods
- Potential applications: Integration with healthcare systems for early risk assessment, personalized interventions
- Limitations and challenges: However further studies and data are needed to improve the overall result of the model