Heart Disease Prediction using Machine Learning

Maitreya Ganu, IISER Thiruvananthapuram

7^{th} December 2024

Abstract

Heart disease is a major health issue worldwide, and early detection is key to better treatment. In this project, I use machine learning to predict whether a person is at risk for heart disease based on health data like age, blood pressure, cholesterol, etc. I applied a logistic regression model to classify individuals into two categories: at risk and not at risk. The model was tested for accuracy and other performance measures, showing the potential of machine learning to support doctors in making faster and more accurate diagnoses.

1 Introduction:

In this project, we use machine learning to predict the risk of heart disease based on simple health data such as age, blood pressure, cholesterol, and other factors. By using a logistic regression model, we aim to build a tool that can quickly predict whether someone is at risk for heart disease. This tool can support doctors in making faster and more accurate decisions, ultimately improving patient care. The output is binary: 0 means "not at risk of heart disease," and 1 means "at risk of heart disease."

2 Overview of the Data:

The dataset used in this project contained 1025 data points, with no null or missing values. Out of these, 499 individuals were not at risk for heart disease, and 526 were at high risk.

3 Model Development:

I have used a logistic regression model, where 20% of the data was used to train the model, and 80% was used for testing. Using the 'sklearn' library, an accuracy score of approximately 81.46% was achieved. The model takes an input in the form of (age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target). The meaning of these variables can be understood from the dataset described in Section 4. The output will either be 0 or 1, indicating the risk level for heart disease.

4 Data and Code Availability:

- Access the data used in this project:
  Heart Disease Prediction Dataset
- Access the Python code:
  Code for Heart Disease Prediction