Predicting Heart Disease

Aim:

Build a model using the 'data.csv' employing either R or Python, with the goal of predicting if an individual has heart disease. During the interview, the panel will ask you to discuss how you choose to build this model. At this point you will be free to talk us through your code, choice of model and any conclusions you made regarding the data.

TIP: Save any visualisations you may use to aid the description of the data during this discussion.

Description of Data

The attached 'data.csv' file contains information on 1,000 patients who were screened for heart disease. In total, 13 variables were recorded in relation to their diagnosis (diseased or not diseased), the 14th variable is the response variable:

1. 'age'
   - An individuals recorded age in years.
2. 'sex'
   - Female = 0, Male = 1
3. 'chest_pain_type'
   - Typical Angina = 0, Atypical Angina = 1, Non-Anginal Pain = 2 and Asymptomatic = 3; 0-2 can be interpreted as some form of chest pain, 3, as no chest pain observed.
4. 'resting_blood_pressure'
   - Systolic Blood Pressure. Values of <= 120 can be considered normal/healthy and values > 120 are abnormal.
5. 'cholesterol'
   - Circulating blood cholesterol in mg/dl. Larger values can be considered unhealthy.
6. 'fasting_blood_sugar'
   - Normal = 0, Abnormally High = 1
7. 'ecg'
   - Results from an ECG test. Normal = 0, abnormality type A = 1, abnormality type B = 2
8. 'max_heart_rate'
   - Maximum resting heartrate in Beats Per Minute (BPM)
9. 'exercise_induced_chest_pain'
   - No Pain Observed = 0, Chest Pain Observed During Exercise = 1
10. 'st_depression'
    - ST Depression induced by exercise relative to rest. Higher values can be considered abnormal.
11. 'st_slope'
    - The slope of the peak exercise ST segment
12. 'stained_blood_vessels'
    - Number of blood vessels with observed plaque build-up. A larger number can be considered unhealthy.
13. 'blood_disorder'
    - Blood Disorder: Normal = 0, Irreversible Damage = 1, Reversable Damage = 2
14. 'heart_disease'
    - Response variable, No disease = 0, Heart Disease Present = 1