

Decision Tree analysis

For MIS2502 Section 004

Sajin Samuel

1. Description of the Data Set

- The four databases in this 1988 data set are Cleveland, Hungary, Switzerland, and Long Beach V. The "target" field references the patient's having heart illness. 0 means there is no disease, while 1 means there is a disease.

age	Age
sex	Sex
cp	Chest pain type (4 values)
trestbps	Resting blood pressure
chol	Serum cholesterol in mg/dl
fbs	Fasting blood sugar > 120 mg/dl
restecg	Resting electrocardiographic results (values 0,1,2)
thalach	Maximum heart rate achieved
exang	Exercise induced angina
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	Thalassemia 0 = normal; 1 = fixed defect; 2 = reversable defect
target	Patient's having heart issues, 0 means there is no issues and 1 means there is an issue

2. Best Value for Minimum Split

- The value for the minimum split is 60.
- The correct classification rate for the training set is 79.01%
- The correct classification rate for the validation set is 75.41%
- 60 splits would be better with high-performance validation and do not overfit on the training set compared to other possible values.

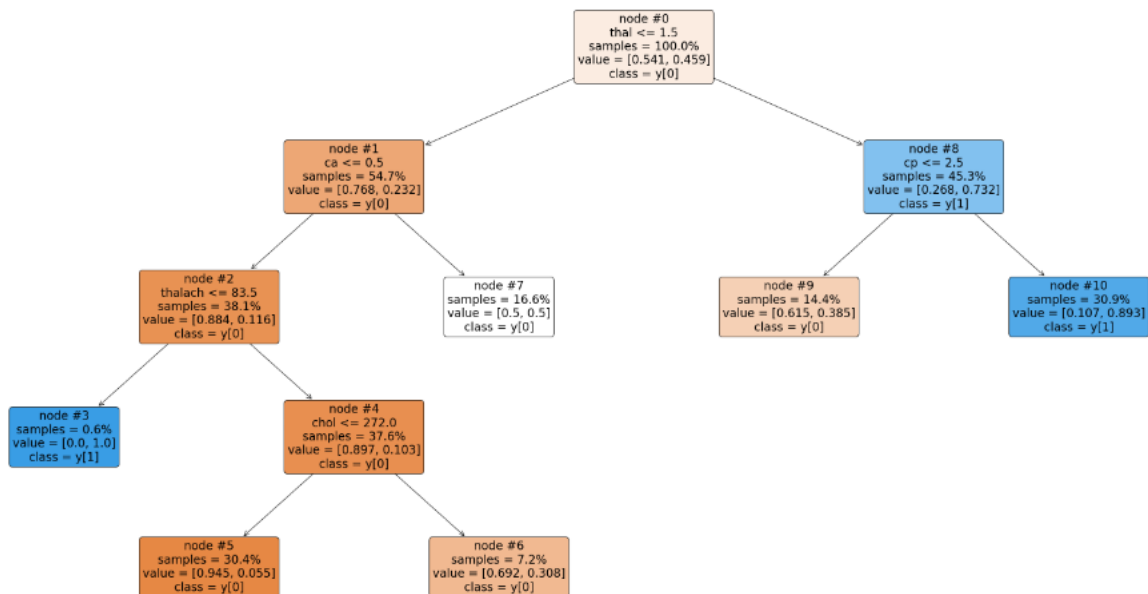
3. Nodes with the Highest and Lowest Probability

- **Node #3 has the highest probability.**
 - This implies that an individual who is analyzed and found to fit into this category will have a lower likelihood of having a heart issue.
 - Thalassemia is less than 1.5.
 - Major vessels colored fluoroscopy is less than 0.5.
 - The maximum heart rate achieved is less than/equal to 83.5.
- **Node #5 has the highest probability.**
 - This implies that an individual who is analyzed and found to fit into this category will have a higher likelihood of having a heart issue.
 - Thalassemia is less than/equal to 1.5.
 - Major vessels colored by fluoroscopy are less than/equal to 0.5.
 - The maximum heart rate achieved is greater than/equal to 83.5.
 - Cholesterol is less than/equal to 272.0.

4. Examples of data points

Individuals with normal thalassemia who has two major vessels colored by fluoroscopy	50%
Individuals with reversible defect of thalassemia who has chest pain of the value 3.	89.3%
Individuals with fixed defect who has 0 number of major vessels colored by fluoroscopy along with a maximum heart rate achieved is 71.	100%
Individuals with fixed defect who has 0 number of major vessels colored by fluoroscopy along with a maximum heart rate achieved is 120.	10.3%

Decision tree



Sources

<https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset>