

Heart Disease Prediction Using Decision Tree

This report documents the process of using decision tree classification to predict heart disease based on clinical data. It includes the methodology, tools, results, decision tree visualization, and key takeaways.

Problem Statement:

Heart disease is a major health challenge globally. Early prediction using machine learning models can help reduce risks and aid diagnosis. This project aims to build a decision tree model using clinical indicators to classify whether a person is likely to have heart disease.

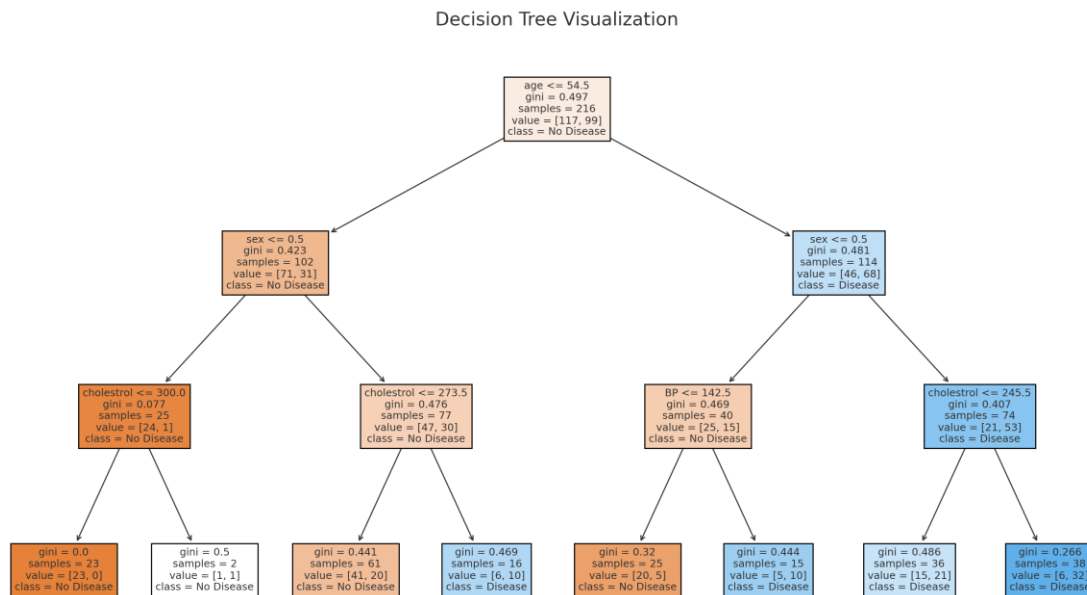
Tools & Technologies:

- Python (Jupyter Notebook)
- pandas & NumPy for data manipulation
- matplotlib for plotting
- scikit-learn for machine learning and evaluation

Approach:

1. Load and inspect dataset (`heart_v2.csv`)
2. Split data into train/test sets
3. Train a decision tree classifier
4. Evaluate model performance using classification metrics
5. Visualize the decision-making process of the trained tree

Decision Tree Visualization:



The decision tree above provides a visual representation of the model's logic. Each internal node tests a feature (like age or cholesterol), and branches represent the outcome of that test. Leaf nodes represent the predicted class: either 'No Disease' or 'Disease'. The colors and proportions help identify which paths were more common. For example, patients with lower cholesterol and non-typical chest pain often fall into the 'No Disease' category. This tree helps doctors and stakeholders interpret the model's logic transparently.

Results Summary:

- Accuracy: 0.63
- Precision: 0.52
- Recall: 0.57
- F1 Score: 0.55

Key Takeaways:

- Decision Trees are intuitive and easy to interpret.
- Accuracy on test data was moderate, showing room for improvement.
- Feature selection and hyperparameter tuning (like ``max_depth``) significantly affect performance.
- Chest pain type and cholesterol levels played notable roles in classification.