Cohort 10

C3 PGP AI/ML

Assignment 3 – Decision Tree

CASE:

People with cardiovascular disease or at high cardiovascular risk (due to factors like hypertension, diabetes, hyperlipidemia, or existing disease) require early detection and management, where a machine learning model can be immensely beneficial.

The objective of this assignment is to fit the model using Decision Tree Classifier and evaluate the same.
You are free to use **scikit-learn** implementation directly for this assignment.

Detailed instructions are as below:

Dataset: Attached with assignment in Canvas, please download the CSV from there

[HeartDiseasePrediction.csv]

Tasks:

1. Load the dataset [0.5]

2. Preprocess the data and perform encoding if required with proper explanation. [2]

3. Split the dataset with 30% test size [0.5]

4. Build the tree and print training and testing accuracy [2]

5. Hyperparameter tuning along with printing of best parameters and also print accuracy. Consider the following parameters:
   - criterion can be either "gini" or "entropy"
   - max_depth where max_depth in [2,3,4,5,6]--integer
   - min_samples_leaf where min_samples_leaf in [5, 10, 20, 30] -- the amount of patients each leaf of the tree has to hold

   The last two are both regularizers and limiting them is a way to avoid overfitting. [2.5]

6. Visualize/Plot the tree with class name as “Heart Disease” and “No Heart Disease”. [1.5]

ALL THE BEST!