DEEP LOGIC

Advanced Python Programing Class

***Decision Trees* Assignment**

Instructions: *Try and answer all the following questions to the best of your ability, all by yourself. If you need any assistance, please reach out to me so I can help you decipher the fair details or interact with fellow classmates for group discussions. Your development environment of choice should be any IDE you would prefer, but the best IDE for this work would be Jupyter Notebook which will allow you to download and save the completed work as an HTML file which you will eventually share with me via your GitHub account. You also can use Google Colab and share your work and this would be ONLY good during the development stages.*

Background: In this assignment, you will classify heart disease diagnosed versus no heart disease diagnosed. You will be using supervised machine learning with a tool called decision trees. Decision trees are predictive models that use a set of binary rules to calculate a target value. Each individual tree has a root node, decision nodes, branches and leaves.

Data was collected by the Cleveland Clinic Foundation. Data are deidentified and publicly available. The variables are:

- chol — serum cholesterol in mg/dl
- restecg — resting electrocardiographic results
  - 0: normal
  - 1: having ST-T wave abnormality
  - Value 2: showing probable or definite left ventricular hypertrophy
- target — have disease or not (1 = yes, 0=no)

You will submit your Python code, all graphs or tables created, and this document with the answers. Paste code in directions below as appropriate.

Python: Using Python, load the data set and complete the work below.

Directions: With the assignment data do the following;

1. Load the data into Python
2. Select the features and the target.
3. Create a countplot of the target
4. Create one graph for each of the features.
5. Create a DecisionTreeClassifier model
6. Fit your model
7. Predict with your model
8. Create a confusion matrix using pd.crosstab
9. Print a classification report of the data
10. Graph the decision tree