Decision Tree Assignment

Problem Statement:

1. Pokémon is a group of adorable creatures peacefully colonizing a planet until humans come along and make them combat each other in order to get shiny badges and we can call them Pokémon masters.
2. In this universe, there exists a group of rare and often strong Pokémon, known as Legendary Pokémon. Unfortunately, there are no detailed criteria that define these Pokémon.
3. The only way to recognize a Legendary Pokémon is through information from official media, such as the game or anime.
4. This data set includes 721 Pokemon, including their number, name, first and second type, and basic stats: HP, Attack, Defense, Special Attack, Special Defense, and Speed. The legend of a pokemon cannot be suspected only by its Attack and Defense. It would be worth finding which variables can define the legend of a pokemon. The strategy is to analyze the data and perform a predictive task of classification to predict the legend of a pokemon using a decision tree algorithm.

Data Description:

Column Names	Description
ID	ID of each Pokemon
Name	Name of each Pokemon
Type 1	Each Pokemon has a type, this determines weakness/resistance to attacks
Type 2	Some Pokemon are dual type and have 2
Total	sum of all stats that come after this, a general guide to how strong a Pokemon is
HP	hit points, or health, defines how much damage a Pokemon can withstand before fainting

Attack	the base modifier for normal attacks (eg. Scratch, Punch)
Defense	the base damage resistance against normal attacks
SP Atk	special attack, the base modifier for special attacks (e.g. fire blast, bubble beam)
SP Def	the base damage resistance against special attacks
Speed	determines which Pokemon attacks first each round
Generation	type of generation of the Pokemon
Legendary	Rare Pokemon

1. How many pokemon are from the 5th generation?
   a. 178
   b. 165
   c. 150
   d. 170
2. How many pokemon have the highest defense score?
   a. 10
   b. 7
   c. 3
   d. 2
3. How you will be handling missing values in this dataset:
   a. Fill up the null values with the median.
   b. Fill up the null values with standard deviation.
   c. Fill up the null values with the mean.
   d. Fill up the null values with None.
4. Which columns are not having any kind of relationship with the generation column?
   a. Attack
   b. Speed

c. <mark>Both of the above</mark>
d. None of the above

5. Which of the following model is the best fit for predicting the legendary of the pokemon based on the below parameters:
1. Handle the missing values.
2. Split the dataset into a 70:30 ratio with random_state as 1.
   a. Linear Regression
   b. <mark>Logistic Regression</mark>
   c. <mark>Decision Tree Model</mark>
   d. <mark>Random Forest Model</mark>

6. What is the precision of the Decision Tree model when the target is False?
   a. <mark>0.90 to 0.1</mark>
   b. 0.80 to 0.90
   c. 1.0 to 2.0
   d. 0.50 to 0.60

7. What is the sensitivity of the above model when the target is True?
   a. <mark>0.90 to 1.0</mark>
   b. 0.50 to 0.60
   c. 0.60 to 0.70
   d. 0.30 to 0.40

8. How much correctly classified data has been retrieved from the above model?
   a. Between 15 to 20
   b. Between 7 to 10
   c. Between 30 to 45
   d. Between 50 to 70

9. Decision tree models might create some biased trees if some classes dominate. From the below options which action is best to take so that it won't create biased trees:
   a. <mark>balance the dataset prior to fitting</mark>
   b. imbalance the dataset prior to fitting
   c. balance the dataset after fitting
   d. None of the above

10. Suppose, you have to work with an ML problem, where you have to predict the number of oxygen tanks needed to be shipped from Indonesia. Which of the following ML algorithm you will choose:
    a. Logistic regression
    b. <mark>Decision Tree</mark>
    c. Both of the above

d. None of the above

11. Which of the following is true for the Decision Tree?
    a. The model can able to generate understandable rules
    b. The model can able to handle both continuous and categorical variables
    c. It can able to perform classification without requiring much computation
    d. All of the above

12. The total gain is computed by adding the expected value of each outcome and deducting the costs associated with the decision.
    a. True
    b. False

13. How we can avoid the overfitting in Decision Tree
    a. Stopping the Tree Growth
    b. Pruning the Full Grown Tree
    c. Both of above
    d. None of the Above