**Instructions:**

**You can use Word, Excel, Power Point, R and/or Python to answer the questions in this exam. There are a total of Seven (7) multi-part questions, with point values noted for each question. You must use Excel if it is specified in the problem statement.**

**Please show your calculations, or the details of your program(s) for each problem. You must supply the R/Python programs, and the programs should be commented so that each step is clearly explained.**

**Combine all of your answers/files into a single zipped file and post**

1. A CART tree will be used to classify Blood Pressure using the features 'Gender' and 'Smoker'. Using the training data provided in the 'BP\_cat\_train' sheet of the 'BP\_train\_test' Excel file and 'BP\_Status' (target):
   1. Utilize Excel to calculate the value of the Q(s/t) function for Gender=’Male’ vs. Gender=’Female’.
   2. Identify all possible first level splits of the classification and regression tree. (do not calculate the Q(s/t))

(20 points)

1. Using the training data provided in the 'BP\_cat\_train' sheet of the 'BP\_train\_test' Excel file. Utilize Excel to build a categorical Naïve Bayes model and score the ‘BP\_cat\_test’ using the features 'Gender' and 'Smoker' as predictors and ‘BP\_Status' as the target variable. (20 points)
2. Using the training data provided in the 'BP-Num\_train’ sheet of the 'BP\_train\_test' Excel file, use Excel to build a knn model and score the ‘BP\_Num\_test’ using the features ‘Gender’, ‘Age’, ‘height’ and ‘weight’ as predictors and ‘BP\_Status' as the target variable. Use: K=2, weighted, Euclidian distance. (20 points)

|  |  |  |
| --- | --- | --- |
|  | **Min** | **Max** |
| Age (years) | 25 | 65 |
| Weight (lb) | 50 | 300 |
| Height (inch) | 50 | 80 |

Refer to the 'BP.CSV' file for the following questions. In these questions, 'BP\_status' is the target variable and the remaining features are predictors. Also, treat the ‘Smoker’ feature as an ordinal variable.

1. Show box and pair plots for the “Age’, ‘Height’ and ‘Weight’ features. (10 Points)
2. Use the CART methodology to develop a classification model for BP.CSV using 30% of the data as training data. Show your model metrics (accuracy, F1 and Recall). (10 points)
3. Use the Gaussian NB methodology to develop a classification model for BP.CSV using 30% of the data as training data. Show your model metrics (accuracy, F1 and Recall). (10 points)
4. Use the knn methodology to develop a classification model for BP.CSV using 30% of the data as training data. Show your model metrics (accuracy, F1 and Recall). (10 points)

Data dependency: BP\_Train\_Test Excel file

BP.CSV