**There are a total of four (4) questions, with point values noted for each question. See the description in “Lecture” module.**

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

**Combine all your answers/files into a single zipped file and post the zipped file to “Final Submissions” in CANVAS**

**All submissions are due no later than 11:59 pm on Wednesday May 9 , 2018.**

**Solve problems 1 and 2 using the following table:**

**Using the “HELP” dataset, a researcher has prepared the following table of patient counts:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ethnicity** | **Age Category** | **Alcohol** | **Cocaine** | **Heroin** | **Row Total** |
| **Black** | **Old** | **30** | **48** | **17** | **95** |
|  | **Young** | **25** | **72** | **13** | **110** |
| **Hispanic** | **Old** | **7** | **0** | **5** | **12** |
|  | **Young** | **8** | **7** | **19** | **34** |
| **White** | **Old** | **60** | **2** | **17** | **79** |
|  | **Young** | **26** | **10** | **34** | **70** |
| **Column Total** |  | **156** | **139** | **105** | **400** |

**1) Use the table above and excel to classify patient addiction type (alcohol, cocaine, heroin) by constructing a classification and regression tree (CART) (two levels only), using Ethnicity and Age Category**. (20 points)

**2) Use the table above and excel to classify patient addiction type (alcohol, cocaine, heroin) by constructing a C4.5 tree (two levels only), using Ethnicity and Age Category.** (20 points)

**3) Use the “IBM Employee Attrition V2” dataset in CANVAS to uncover the features that can predict employee attrition. This is a subset of a fictional data set created by IBM data scientists. Do not normalize the data. Choose at least two of the following techniques:**

* **Naïve Bayes**
* **CART**
* **C5.0**
* **Random Forest**

(45 points)

**4) Using data in the table below, construct a Neural Network with one Output Layer (z) and one Hidden Layer (A and B). Calculate the predicted outcome if the inputs to the input nodes are (x=1, Node 1=.4, Node 2=.7 Node 3= .7 and Node 4=.2).** (15 points)

|  |  |  |
| --- | --- | --- |
| **From** | **To** | **Weight** |
| X | A | 0.5 |
| Node 1 | A | 0.6 |
| Node 2 | A | 0.8 |
| Node 3 | A | 0.6 |
| Node 4 | A | 0.2 |
| x | B | 0.7 |
| Node 1 | B | 0.9 |
| Node 2 | B | 0.8 |
| Node 3 | B | 0.4 |
| Node 4 | B | 0.2 |
| xx | z | 0.5 |
| A | z | 0.9 |
| B | z | 0.9 |