1. (10) Name two measures to examine node purity when using decision trees for classification?

\_\_\_\_\_\_gini coefficient\_\_\_ \_\_entropy\_

1. (10) (T/F) Decision Trees produces the optimal classification rules based on the training set. TRUE
2. (15) Explain the difference between clustering and classification in at most three sentences?

The main difference between them is dataset type which works on. While clustering works on unlabeled dataset, classification works on labeled dataset.

1. (15) What is the most important advantage of using a decision tree? Explain your answer in at most two sentences.

Because decision trees work with the basic strategy which choses the best option on current split, they provide us look at fastly whole dataset.

1. (15) If the training accuracy is \_90\_% and the test accuracy is \_30\_\_% , this would be an example of overfitting. Choose any two numbers that you find illustrative and explain your answer in at most three sentences.
2. (15) Explain the relationship between decision trees and random forests in at most two sentences.

Random trees contain a lot of decision trees and they calculate the total entropy and gini coefficients.

1. (10) Calculate the GINI Index for the following decision tree node: (for xx put the last two digits of your MEF ID)

Class 1: 10 observations Class 2: 21 observations

1 – (10/31)2 – (21/31)2 = 0.437

1. (10) Calculate the entropy for the following decision tree node: (for xx put the last two digits of your MEF ID)

Class 1: 10 observations Class 2: 21 observations

