Classification:

Draw the Decision tree for the following data:

TID	Refund	Marital Status	Taxable Income	Class: Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	single	90K	Yes

Then test the following record

Test Data:

Refund	Marital Status	Taxable Income	Class: Cheat
No	Married	80K	?

Class: Cheat = No

Applying Weka, use the above dataset

a) What is the type of each attribute in this dataset?

b) What are the possible values for the attribute Refund, Marital Status, Taxable Income and Class: Cheat?

c) Construct the decision tree (DT) algorithm using weka.

d) Apply the testing data above on weka tool

Confusion Matrix

Four classifiers are generated for the same training set, which has 100 instances.
They have the following confusion matrices.

		Predicted class	
		+	−
Actual class	+	50	10
	−	10	30

		Predicted class	
		+	−
Actual class	+	55	5
	−	5	35

		Predicted class	
		+	−
Actual class	+	40	20
	−	1	39

		Predicted class	
		+	−
Actual class	+	60	0
	−	20	20

a) Calculate the values of true positive rate and false positive rate for each classifier and plot them on a ROC graph.
b) Calculate the value of the Euclidean distance measure Euc for each one.
c) Which classifier would you consider the best if you were equally concerned with avoiding false positive and false negative classifications?