I. ZeroR, OneR and Decision Tree Classifiers

1. Consider the below dataset on "Buys_Computer" with 12 observations.

Train #	Income	Student	Credit Rating	Buys_Computer
1	High	No	Fair	Yes
2	High	No	Excellent	Yes
3	High	Yes	Fair	No
4	Medium	No	Excellent	Yes
5	Low	Yes	Fair	No
6	Low	Yes	Excellent	No
7	Medium	Yes	Excellent	No
8	High	Yes	Excellent	No
9	Medium	Yes	Fair	No
10	Medium	No	Fair	Yes
11	Medium	No	Fair	Yes
12	Low	No	Fair	No

a. Apply *ZeroR* classification algorithm and predict the baseline performance.

Zero R

58.33% 7 no > 5 yes

b. Apply OneR classification algorithm, determine the best predictor and calculate its accuracy

Income	Yes	No
High	**2**	**2**
Medium	**3**	2
Low	**0**	3

Student	Yes	No
Yes	**0**	**6**
No	**5**	1

Credit Rating	Yes	No
Fair	**3**	4
Excellent	**2**	3

Student will be best predictor; accuracy: **91.67%**

c. Calculate Information gain for all the predictors Income, Student, Credit Rating and construct the decision tree

Decision Tree
Info Gain:
Ranking 1. Student 2. Income 3.Credit Rating

Info(D given Student) = 6/12 * I (0,6) + 6/12*I(5,1) = 0.325

I(0,6) = - (0/6) *log2(0/6)-((6/6)*log2(6/6)) = 0

I(5,1) = -5/6*log2(5/6) – ((1/6)*log2(1/6)) = 0.66

InfoGain(Student) = Info(D) – Info(D given Student) = 0.99 – 0.33 = 0.66

d. Derive and write down all the classification rules

If Student = "Yes" then Buys_Computer = "No"

If Student = "No" and Income = "Low" Buys_Computer = No

If Student = "No" and Income = "High" Buys_Computer = Yes

If Student = "No" and Income = "Medium" Buys_Computer = Yes

e. Use the following test dataset and predict the class "Buys_Computer" based on the constructed model.

Test #	Income	Student	Credit Rating	Actual Buys_Computer	Predicted Buys_Computer
1	High	No	Excellent	No	Yes
2	High	Yes	Fair	Yes	No
3	Medium	No	Excellent	Yes	Yes
4	Medium	Yes	Fair	No	No
5	Low	Yes	Fair	Yes	No
6	Low	No	Fair	No	No

f. Create the confusion matrix for the model.

		Predicted	
		Yes	No
Actual	Yes	1	2 (FN)
	No	1 (FP)	2

g. From the Confusion Matrix, calculate
   a. Accuracy
   b. Error Rate.
   c. True Positive Rate
   d. False Positive Rate

Accuracy : 50%

Error Rate : 50%

TPR:	TP/(TP+FN)	33.3333333%
FPR	FP/(FP+TN)	33.3333333%