

Show your working in all calculations.

1. a) Explain the concepts of **nominal** and **ratio** attributes. Give TWO examples of each type of attribute. (6 marks)
- b) Explain why it is important to evaluate the performance of classifiers on data that was not used for training the classification model. What would happen if the performance of a 1-nearest neighbour classifier was evaluated using the same data from the training set? (6 marks)
- c) Explain what is meant by a **confusion matrix**. Compute the values of the true positive rate and false positive rate for the following confusion matrix:

		Predicted	
		P	N
Actual	P	45	5
	N	10	20

where P stands for positive and N for negative cases. (5 marks)

- d) Describe briefly how **Regression** and **Clustering** differ from **classification**. Give TWO example applications of regression and TWO of clustering. (8 marks)

(Total: 25 Marks)