

UNIVERSITY OF WOLVERHAMPTON
FACULTY OF SCIENCE AND ENGINEERING
7CS033 DATA MINING & INFORMATICS
WORKSHOP # 6

T6.1- Assume six data points with two binary attributes, X_1 and X_2 , are given as listed in Table 1. These data points belong to three classes, $Y \in \{1,2,3\}$, and our purpose is to classify these data points using a decision tree classifier with only one split.

Calculate the information gain values when the data points are split using X_1 and X_2 . Explain which split is better and why. Draw the decision tree using the best split, label the branches, and determine what the predicted class label in each leaf is. [10%]

Table 1

X_1	X_2	Y
1	1	1
1	1	1
1	1	2
1	0	3
0	0	2
0	0	3

T6.2 - Use the K-means clustering to find two different clusters in the following sequence of three-dimensional points:

$$X=[(1,9,14),(2,18,23),(3,30,30),(4,21,9),(5,9,17),(6,25,32),(7,36,25),(8,10,12),(9,38,45),(10,1,2)]$$

Choose two random centres for your clusters to start the algorithm. Include the centres of clusters and calculations for each iteration in your answer. You can use the Euclidean distance to calculate the distance between points [5%]

T6.3 – Download Workshop_6 notebook from Canvas and complete the task [10%]