Homework 4
Due on 22h00, 18/07/2021

Rule
- File header: Fullname and student-ID
- File name: StudentID_HW04.pdf

Question 1

Given a training set as follows,

No.	Shape	Country	Status	Group
1	Small	Germany	Single	A
2	Big	French	Single	A
3	Big	Germany	Single	A
4	Small	Italy	Single	B
5	Big	Germany	Married	B
6	Big	Italy	Single	B
7	Big	Italy	Married	B
8	Small	Germany	Married	B

a) Create a decision tree by using the Information Gain measure.

b) Retrieve classification rules for Group A and Group B from the tree of a).

c) Create a decision tree by using the Gain Ratio measure.

d) Evaluate the tree of a) and c).

Question 2

Given a database as follows,

Customer	Article1	Article2	Article3	Article4	Article5	Artical6
1	0	1	0	1	1	0
2	0	1	1	1	0	1
3	1	0	1	0	1	0
4	1	0	0	1	0	1
5	1	0	0	0	1	0

a) Use the K-mean algorithm and Cosine distance to cluster 6 articles into 2 clusters with initial centroids article1 and article3.

The Cosine distance of two vectors A and B is represented as follows,

$$\frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

b) Evaluate the quality of clustering by using BetaCV measure.

The end.