

INFS4203/INFS7203 Data Mining (2009)

Assignment 2 - Classification

Due date: 11 September, 2009.

The goal of this individual written assignment is to explore data classification. This assignment counts for 5% of the total unit assessment. Submit either a hardcopy or a soft copy of your assignment through email to the course instructor before or during the Tutorial time 2pm Friday 11 Sep. 09. Show details of your answers and calculations step-by-step, otherwise your marks will be deducted.

Problem

Perform ID3 Algorithm to derive a decision tree and a set of production rules for job application decisions. The attribute TO-APPLY is the outcome (class) of the training set.

Salary	Profession	Work-Stress	TO-APPLY
Good	Different	normal	Yes
Low	Different	high	No
High	Different	High	No
Low	Same	normal	Yes
Good	Different	high	No
High	Same	high	Yes
Low	Different	normal	No
High	Same	normal	Yes

The computations may be recursive. You need show all the results of computations. The following template can be used to show the results.

	Entropy	TO-APPLY = Yes	TO-APPLY = No
H(C Salary=Good)			
H(C Salary=Low)			
H(C Salary=High)			
H(C Salary)			
H(C Profession=Same)			
H(C Profession=Different)			
H(C Profession)			
H(C Work_Stress=Normal)			
H(C Work_Stress=High)			
H(C Work_Stress)			

- Q1:** Show the computation of the Entropies of the attributes.
Q2: Show the building of the decision tree.