

UTS SEMESTER GENAP 2017/2018

Mata Kuliah : **Data Mining**
Hari/Tanggal : **Sabtu / 21 April 2018**
Waktu : **WIB (120 menit)**
Sifat : **Terbuka**

1. a. Explain briefly what is the dirty data and how can you handle it! (10 points)
 b. Explain five major task in data preprocessing! (10 points)
 c. Explain briefly six Data Mining tasks! (10 points)
2. Given the following data:
 62, 65, 108, 23, 82, 132, 50, 78, 56, 121, 39, 104, 143, 105, 115
 a. Partition into 3 Bins using equal-width (5 points)
 b. Smoothing by bin median (5 points)
 c. Smoothing by bin means (5 points)
 d. Smoothing by bin boundaries (5 points)

3. Given the following table:

Pregnant	Plasma-	
	Glucose	DiastolicBP
6	148	72
1	85	66
8	183	64
1	89	66
5	116	74
3	78	50
2	197	70
4	110	92
10	168	74

In the relation to Data Transformation, normalize Pregnant, Plasma-Glucose dan DiastolicBp using the following concepts:

- a. *max-min normalization* by new max-min of all attributes in 0 to 1, and use max-min of data as old max-min (10 points)
 - b. *z-score normalization* (10 points)
 - c. *decimal scaling by 10* (5 points)
4. By using Greedy Approach (Entropy and Information Gain), construct a decision tree from the attributes *Body Temperature*, *Give Birth*, *Four-Legged* and *Hibernates* to determine *Class Label*. (35 points)

Table 4.4. An example test set for classifying mammals.

Name	Body Temperature	Gives Birth	Four-legged	Hibernates	Class Label
human	warm-blooded	yes	no	no	yes
pigeon	warm-blooded	no	no	no	no
elephant	warm-blooded	yes	yes	no	yes
leopard shark	cold-blooded	yes	no	no	no
turtle	cold-blooded	no	yes	no	no
penguin	cold-blooded	no	no	no	no
eel	cold-blooded	no	no	no	no
dolphin	warm-blooded	yes	no	no	yes
spiny antenter	warm-blooded	no	yes	yes	yes
gila monster	cold-blooded	no	yes	yes	no