## TE/IT/SEMVI/GBCS

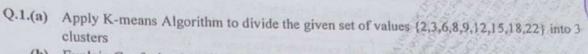
1 6 MAY 2019

NB: 1. Question no. 1 is compulsory.

(3 Hours)

[Total Marks:80]

- 2. Answer any three out of the remaining questions.
- 3. Assume data, if missing, with justification.



(b) Explain Confusion Matrix. Calculate Accuracy, Precision and Recall for the following Confusion Matrix

Cancer Classes	Yes	No	Total.
Yes	90	210	300
No	140	9560	9700
Total	230	9770	10000

(c) What are the major issues in data mining?

[05]

(d) What is noisy data? How to handle it?

[05]

[10]

Q.2.(a) Consider the transaction database given in table below. Apply Apriori Algorithm with minimum support of 50% and confidence of 50%. Find all frequent itemsets and all the

association rules.

THE RESERVE THE PARTY OF THE PA	
Tid	Items
100	1,3,4
200	2,3,5
300	1,2,3,5
400	2,5
500	1,2,3
600	28 3,5
700	1,2,3,5
800	1,5
900	1,3
- Allen	700000000

(b) Explain Regression. Explain linear regression with example.

[10]

Q.3.(a) Suppose we have five objects with name A, B, C, D and E. Apply single linkage clustering [10] and draw dendrogram for the given data.

X	Y
S S S S S S S S S S S S S S S S S S S	1
A S S S S S S S S S S S S S S S S S S S	1.5
000000000000000000000000000000000000000	5
3	4
800000	4
153200	3.5
S P S B 3	1 for outlie

(b) What is an outlier? Describe methods that are used for outlier analysis.

[10]

Page 1 of 2

69039

Q4.(a) Using the given training dataset classify the following tuple using Naïve Bayes Algorithm:

<Homeowner: No, Marital Status: Married, Job experience:3>

		(in years)	Defaulted
Homeowner	Marital Status	Job experience (in years)	No
Yes	Single	3 3 3 3 3 3 3 3	No
No	Married	453335	No
No	Single	1 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	No
Yes	Married	8 4 2 3 8 8 6	Yes
No	Divorced	38480	No
No	Married	4 3 3 3 3 3	No
Yes	Divorced	0086400808	Yes
No	Married	3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	No
No	Married	133350850	Yes
Yes	Single S	10 3 - 5 3 3 3 3 3 5 5 5 5 5 5 5 5 5 5 5 5	1000

- (b) Explain Business Intelligence issues.
- Q5. (a) What is data mining? Explain KDD process with diagram.
  - (b) Explain Market-Basket analysis with example.
- Q6. (a) What are multiple level and multidimensional association rules? Explain with suitable examples for each:
  - (b) Suppose that data for analysis includes the attribute age. The age values for data tuples are [10] (in increasing order):

[10]

[10]

[10]

- 13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,35,35,35,35,35,36,40,45,46,52,70
- i) What is mean of data? What is median of data?
- ii) What is mode of data? Comment on data's modality.
- iii) What is mid-range of data?
- iv) Give the five-point summary of the data.
- v) Show box plot of the data.

\*\*\*\*\*

69039

Page 2 of 2