# AUTUMN MID SEMESTER EXAMINATION-2018
## School of Computer Engineering
## KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY
## DEEMED TO BE UNIVERSITY, BHUBANESWAR-24

### DATA WAREHOUSING AND DATA MINING
### [CS-6301]

**Time: 1½ Hours** **Full Mark: 20**

***Answer any four questions including question No.1 which is compulsory. The figures in the margin indicate full marks. Candidates are required to give their answers in their own words as far as practicable and all parts of a question should be answered at one place only.***

Q.1. [5x1]

(a) What are the steps involved in KDD process?
(b) What condition makes association rules are interesting?
(c) Define anti-monotone property.
(d) What is Attribute Selection Measure in Decision Tree?
(e) What condition two item sets A and B will have "No Correlation between them"?

Q.2.

(a) What is Interquartile range (IQR)? Draw an IQR with an odd sample size as ***70, 64,77, 63, 81, 72, 76, 64, 81.*** [2.5]

(b) Calculate both Eigen vectors & Eigen values for $A=\begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ [2.5]

Q.3.

(a) What are the rule strength measures used in any association rule model? Discuss with their probability formula. [2.5]

(b) What is the Apriori property? Using Apriori Algorithm find the final item set for the following dataset S. [2.5]
(*Where minimum support is 0.5*)

**S=**

| TID | Items |
|---|---|
| 101 | Milk,Bread,Butter |
| 102 | Jam,Bread,Egg |
| 103 | Milk,Jam,Bread,Egg |
| 104 | Jam,Egg |

Q.4.

(a) Measure your dependent/correlated events LIFT computation for the following contingency table, [2.5]
*play basketball⇒eat cereal* [40%, 66.7%], *play basketball⇒not eat cereal* [20%, 33.3%]

| | Basketball | Not basketball | Sum (row) |
|---|---|---|---|
| Cereal | 2000 | 1750 | **3750** |
| Not cereal | 1000 | 250 | **1250** |
| Sum(col.) | **3000** | **2000** | **5000** |

(b) What are the information gains of $a_1$ and $a_2$ relative to these training examples? [***Entropy of the training set is 0.9911***]. [2.5]

| TID | $a_1$ | $a_2$ | CLASS |
|---|---|---|---|
| 1 | T | T | YES |
| 2 | T | T | YES |
| 3 | T | F | NO |
| 4 | F | F | YES |
| 5 | F | T | NO |
| 6 | F | T | NO |
| 7 | F | F | NO |
| 8 | T | F | YES |
| 9 | F | T | NO |

Q.5.

(a) What is Bessel's correction? Calculate and Draw a box plot for the data set ***3, 3, 7, 8, 7, 4, 4, 10, 1, 5, 1, 7, 2, 7, 9.*** [2.5]

(b) What is Posterior Probability? Use Bayes Theorem to find out "***If a patient has fever, what's the probability he/she has Typhoid?***" (i). Normally, Typhoid causes high fever 60% of the time. (ii). Prior probability of any patient having Typhoid is 5/10000, (iii). Prior probability of any patient having fever is 1/50 [2.5]

**********