

**M.E DEGREE EXAMINATIONS: NOV/DEC 2012**

Third Semester

**COMPUTER SCIENCE AND ENGINEERING**

CSE587: Data Mining

**Time: Three Hours**

**Maximum Marks: 100**

**Answer all the Questions:-**

**PART A (10 x 2 = 20 Marks)**

1. List different attribute types of data related to data mining.
2. What is precision and bias in data quality?
3. Expand ACCENT.
4. What is conditional independence?
5. Write Apriori principle.
6. What is quantitative association rule?
7. List different types of clusters.
8. What is worst case time complexity of K-Means cluster algorithm with  $m$  points and  $n$  attributes?
9. What are anomalies/outliers?
10. Distinguish between proximity based office detection and density based officer detection and density based officer detection.

**PART B (5 x 16 = 80 Marks)**

11. a) Explain Feature subset selection process of dimensionality reduction with neat flowchart.

**(OR)**

- b) (i) Find simple matching coefficient for the following two vectors using Cosine similarities, Extended Jaccard Coefficient and correlation. (8)

$$X = (3, 2, 0, 5, 0, 0, 0, 2, 0, 0)$$

$$Y = (1, 0, 0, 0, 0, 0, 0, 1, 0, 2)$$

- (ii) Explain the different data smoothing techniques and solve for the following data:

7 7 9 11 13 15 15 25 25 33 38 45

12. a) Explain the following visualization techniques with example

- (i) Histogram
- (ii) Two dimensional histogram
- (iii) Box plots
- (iv) Matrix Plots

**(OR)**

b) Explain Decision tree based induction classification method in detail with necessary algorithm and discuss in information gain and Gini index ratio.

13. a) (i) Explain construction of FP-tree for extracting association rule in transactional (10) databases using the following table.

TID	Items
1	{A,B}
2	{B,C,D}
3	{A,C,D,E}
4	{A,D,E}
5	{A,B,C}
6	{A,B,C,D}
7	{B,C}
8	{A,B,C}
9	{A,B,D}
10	{B,C,E}

(ii) List and explain properties of objective measures. (6)

**(OR)**

b) Briefly explain the how continuous attributes are handled in association rule mining.

14. a) Explain agglomerative hierarchical clustering algorithm and its limitations with suitable example.

**(OR)**

b) (i) Write the algorithm for Bisecting K-Means algorithm and explain. (10)

(ii) List and explain measures of cluster validity. (6)

15. a) Discuss Density based outlier detection with necessary algorithm.

**(OR)**

b) (i) List different approaches adopted for anomaly detection. (6)

(ii) List and discuss the issues need to be addressed during anomaly detection. (10)

\*\*\*\*\*