Daffodil International University

Department of Computer Science and Engineering

Faculty of Science and Information Technology

Final Examination SUMMER 2018

Course Code: CSE450 Course Title: Data Mining

Level : 4 Term: 1 Section: ALL

Instructor: ALL

Time: 2 hrs **Full Marks: 40**

PART-A: MCQ [select the right one(s) and write justification in the answer booklet] 8X2=16

1. **What do you mean by support(A)?**
   a. Total number of transactions containing A
   b. Total Number of transactions not containing A
   c. Number of transactions containing A / Total number of transactions
   d. Number of transactions not containing A / Total number of transactions

2. **What techniques can be used to improve the efficiency of apriori algorithm?**
   a. Hash-based techniques
   b. Transaction Reduction
   c. Partitioning
   d. All of the above

3. **What does FP growth algorithm do?**
   a. It mines all frequent patterns through pruning rules with lesser support
   b. It mines all frequent patterns through pruning rules with higher support
   c. It mines all frequent patterns by constructing a FP tree
   d. All of the above

4. **Which of the following is direct application of frequent itemset mining?**
   a. Social Network Analysis
   b. Market Basket Analysis
   c. Outlier Detection
   d. Intrusion Detection

5. **What is not true about FP growth algorithms?**
   a. It mines frequent itemsets without candidate generation.
   b. There are chances that FP trees may not fit in the memory
   c. FP trees are very expensive to build.
   d. It expands the original database to build FP trees.

6. **What is the relation between candidate and frequent itemsets?**
   a. A candidate itemset is always a frequent itemset
   b. A frequent itemset must be a candidate itemset
   c. No relation between the two
   d. Both are same

7. **What is frequent pattern growth?**
   a. Same as frequent itemset mining
   b. Use of hashing to make discovery of frequent itemsets more efficient
   c. Mining of frequent itemsets without candidate generation
   d. None of the above

8. **Which of these is not a frequent pattern mining algorithm?**
   a. Apriori
   b. FP growth
   c. Decision trees
   d. Eclat

PART-B: Association Mining and Clustering

24

1. Consider the following transaction data table: 6+4+4+5+5 = 24

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

Answer the following questions:

a. What is the maximum number of association rules that can be extracted from the data that includes zero support.

b. What is the maximum size of frequent itemsets that can be extracted considering the minsup > 0.

c. Find an itemset of size 2 or larger that has the largest support.

d. Find a pair of items, *a* and *b*, such that the rules {a} → {b} and {b} → {a} have the same confidence.

e. What are challenges in k-means clustering and how do you propose to overcome them.

---------- Good luck ----------