

Hall Ticket Number:

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Code No. : 16605 N (A)

**VASAVI COLLEGE OF ENGINEERING (Autonomous), HYDERABAD**  
**B.E. (IT: CBCS) VI-Semester Main Examinations, May-2019**

**Data Mining**  
(Elective-I)

Time: 3 hours

Max. Marks: 70

*Note: Answer ALL questions in Part-A and any FIVE from Part-B*

Q.No.	Stem of the question	M	L	CO	PO
<b>Part-A (10 × 2 = 20 Marks)</b>					
1.	Given a group of 12 sales price records has been sorted as follows: 5, 10, 11, 13, 15, 35, 50, 55, 72, 92, 204, 215 Partition them into three bins by equal-width partitioning.	2	2	1	2
2.	In a supermarket dataset the customer income attribute values are missing. How these missing values of this attribute are filled?	2	2	1	1
3.	Compare Data Warehouse and Heterogeneous DBMS.	2	2	2	1
4.	Give a short example to show that items in a strong association rule actually may be negatively correlated.	2	3	2	1
5.	Why is tree pruning useful in decision tree induction?	2	1	3	1
6.	In an Electronics Customer Database (D), The class label attribute, buys computer, has two distinct values {Yes, No}. There are nine tuples of class Yes and five tuples of class No. Compute information gain needed to classify a tuple in D.	2	2	3	2
7.	What are the drawbacks of K-means clustering method?	2	2	4	1
8.	Compare agglomerative and divisive hierarchical clustering.	2	2	4	1
9.	List any two applications of Web Mining.	2	1	5	2
10.	What are n-grams? How are they used in text mining?	2	1	5	2
<b>Part-B (5 × 10 = 50) Marks</b>					
11.a)	Describe the steps involved in data mining when viewed as a process of knowledge discovery.	4	2	1	1
b)	Given the attribute age has the following values: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. i) Plot an equal-width histogram of width 10. ii) Apply smoothing by bin means to smooth these data, using a bin depth of 3. Comment on the effect of this technique for the given data.	6	3	1	2

12.a) Compare Snowflake schema, fact constellation, star schema model with an example.

4 3 2 1

b) A database has four transactions. Let min sup = 60% and min conf = 80%.

6 3 2 2

TID	items_bought (in the form of brand-item_category)
T100	{King's-Crab, Sunset-Milk, Dairyland-Cheese, Best-Bread}
T200	{Best-Cheese, Dairyland-Milk, Goldenfarm-Apple, Tasty-Pie, Wonder-Bread}
T300	{Westcoast-Apple, Dairyland-Milk, Wonder-Bread, Tasty-Pie}
T400	{Wonder-Bread, Sunset-Milk, Dairyland-Cheese}

Find all frequent itemsets by applying Apriori Algorithm.

13.a) Demonstrate how Bayesian classification helps in predicting class membership probabilities.

4 2 3 1

b) Construct a decision tree from the below data. A, B, C, D attributes can be considered as predictors and E column as target variable having class labels.

6 3 3 2

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

14.a) Analyze under which conditions the density-based clustering is more suitable than partitioning-based clustering and hierarchical clustering.

4 4 4 1

b) Suppose that the data mining task is to cluster points (with (x, y) representing location) into three clusters, where the points are:

6 3 4 2

A1(2,10),A2(2,5),A3(8,4),B1(5,8),B2(7,5),B3(6,4),C1(1,2),C2(4,9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Apply the k-means algorithm to show only

- i) The three cluster centers after the first round of execution.
- ii) The final three clusters.

15.a) What techniques are used for mining the Web? Illustrate with examples.

5 1 5 2

b) Discuss the method of mining spatial databases in detail.

5 2 5 1

16.a) Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):

5 2 1 2

- i) Compute the Euclidean distance between the two objects.
- ii) Compute the Manhattan distance between the two objects.
- iii) Compute the Minkowski distance between the two objects.

b) Compare and contrast the OLAP and OLTP Systems with various features.

5 2 2 1