Hall	Tick	et Ni	umbe	r:						
						-			Code No.: 17652 (A) N/0)

VASAVI COLLEGE OF ENGINEERING (AUTONOMOUS), HYDERABAD

Accredited by NAAC with A++ Grade

B.E. (I.T.) VII-Semester Main & Backlog Examinations, Dec.-23/Jan.-24 Data Mining (PE-I)

Time: 3 hours

Max. Marks: 60

Note: Answer all questions from Part-A and any FIVE from Part-B

Part-A $(10 \times 2 = 20 \text{ Marks})$

Q. No.	Stem of the question	M	L	СО	PO
1.	Discuss the applications of data mining.	2	1	1	1
2.	List and explain the OLAP operations.	2	1	1	1
3.	Define Apriori property with an example.	2	1	2	1
4.	Explain any two pattern evaluation measures.	2	2	2	1
5.	Describe clustering of graph data.	2	1	3	1
6.	How do you find an outlier by using the Boxplot method?	2	1	3	1
7.	What are the challenges involved in link mining?	2	1	4	1
8.	What is small world phenomenon?	2	1	4	1
9.	Write about the Twitter sentiment mining.	2	2	5	1
10.	How to mine complex data objects?	2	1	5	1
	Part-B ($5 \times 8 = 40 \text{ Marks}$)				
11. a)	Given dataset D= {98, 96, 88, 84, 31, 26, 22, 18, 11, 66, 61, 44, 43, 32}. Transform the value 45 onto the range [0.0, 1.0] using the min-max normalization.	4	3	1	2
b)	Suppose that a data warehouse for ABC University consists of four dimensions: student, course, semester, and instructor, and two measures count and avg grade. When at the lowest conceptual level (e.g., for a given student, course, semester, and instructor combination), the average grade measure stores the actual course grade of the student. At higher conceptual levels, average grade stores the average grade for the given combination. Draw a star schema diagram for the data warehouse.	4	3	1	2
12. a)	A database has six transactions. Let the minimum support is 25% and the minimum confidence is 45%. Find all frequent itemsets using Apriori.	4	3	2	2
	Tid Items				
	T10 H,D,R,K,S,X				
	T11 K,W,Q,A,Z,S,X				
	T12 W,H,Q,A,				
	T13 D,R,K,C,L				
	T14 Q,A,Z				
	T15 D,R,K,S,X				

Code No.: 17652 (A) N/O

.. 7 ..

b)	Suppose the data contain the frequent itemset FreqItem = I1, I3, I4, I7. What are the association rules that can be generated from FreqItem?	4	3	2	2
13. a)	Apply the K-Means clustering algorithm on the graph data.	4	2	3	1
b)	How do you find an outlier detection in high-dimensional data?	4	2	3	1
14. a)	Explain how does the CLIQUE algorithm work on the d-dimensional data? Give an example.	4	3	4	2
b)	Design the Apriori-based frequent substructure mining algorithm.	4	3	4	2
15. a)	Given two documents D1=" Data Mining", D2=" Text Data Mining". Find the similarity between these two documents.	4	3	5	2
b)	Explain Spatial Data Mining with an example.	4	2	5	1
16. a)	What is the importance of missing values in the data mining? Describe the methods used for filling in the missing values in a given dataset	4	2	1	1
b)	Compare the Pattern Growth with Apriori algorithm	4	2	2	1
17.	Answer any two of the following:	n An es			
a)	Compare and contrast classification & clustering tasks.	4	3	3	1
b)	Discuss the multi-relational data mining approach.	4	2	4	1
c)	Summarize the objectives of Twitter, Facebook, LinkedIn, Github data mining.	4	2	5	1

M : Marks; L: Bloom's Taxonomy Level; CO; Course Outcome; PO: Programme Outcome

i)	Blooms Taxonomy Level – 1	20%
ii)	Blooms Taxonomy Level – 2	40%
iii)	Blooms Taxonomy Level – 3 & 4	40%
