VASAVI COLLEGE OF ENGINEERING (AUTONOMOUS), HYDERABAD
Accredited by NAAC with A++ Grade
B.E. (I.T.) VII-Semester Supplementary Examinations, July-2022
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 × 2 = 20 Marks)

Q. No.	Stem of the question	M	L	CO	PO
1.	What is normalization? Explain z-score and min-max normalization?	2	1	1	1
2.	What are the functionalities of Data Mining?	2	1	1	1
3.	Describe star schema, snowflake schema and fact constellation schemas.	2	1	1	1
4.	Define frequent itemset, closed itemsets, and association rule.	2	1	2	1
5.	Define sensitivity and specificity.	2	1	3	1
6.	Explain how the if-then rules can be extracted from a decision tree with suitable example.	2	1	3	1
7.	Define unsupervised learning? Explain the broad categorization of unsupervised learning methods.	2	2	4	1
8.	Compare single-link, complete-link and average-link clustering methods.	2	2	4	1
9.	Explain any application of Association Rule mining on spatial data. Provide an example.	2	2	5	1, 2
10.	How can we model a document to facilitate information retrieval? Explain TF-IDF measure.	2	2	5	1, 2

Part-B (5 × 8 = 40 Marks)

11. a)	What is Data preprocessing and how it is important issue for both data warehousing and data mining?	4	2	1	1, 2
b)	Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 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. (a) Can you find (roughly) the first quartile (Q1) and the third quartile (Q3) of the data? (b) Give the five-number summary of the data. (c) Show a boxplot of the data. (d) How is a quantile-quantile plot different from a quantile plot?	4	3	1	1, 2
12. a)	Compare Enterprise warehouse, data mart, virtual warehouse.	4	3	2	1, 2

			M	L	CO	PO
	b)	A database has four transactions. Let min sup = 60% and min conf = 80%.	4	3	2	1, 2

CID TID items bought (in the form of brand-item category)
01 T100 {King's-Crab, Sunset-Milk, Dairyland-Cheese, Best-Bread}
02 T200 {Best-Cheese, Dairyland-Milk, Goldenfarm-Apple, Tasty-Pie,
       Wonder-Bread}
01 T300 {Westcoast-Apple, Dairyland-Milk, Wonder-Bread, Tasty-Pie}
03 T400 {Wonder-Bread, Sunset-Milk, Dairyland-Cheese}

At the granularity of item category (e.g., item_i could be "Milk"), for the following rule template,

$\forall X \in$ transaction, buys(X, item1) $\wedge$ buys(X, item2)$\Rightarrow$buys(X, item3) [s, c]

list the association rules using frequent 3-itemset.

			M	L	CO	PO
13.	a)	Explain the major steps of decision tree classification. Why is tree pruning useful in decision tree induction? Analyze the drawback of using a separate set of tuples to evaluate pruning?	5	3	3	1, 2
	b)	*(table below)*	3	3	3	1, 2

Class	Buys computer = Yes	Buys computer = NO	Total
Buys computer = Yes	6954	46	7000
Buys computer = NO	412	2588	3000
Total	7366	2634	10000

Calculate the precision, recall and F-Score using the above matrix.

			M	L	CO	PO
14.	a)	What is outlier? Categorize the techniques for outlier detection.	5	4	4	1, 2
	b)	Outline BIRCH algorithm. Suppose that there are three points, (2, 5), (3, 2), and (4, 3), in a cluster, C1. Calculate the clustering feature of C1.	3	3	4	1, 2
15.	a)	What is web mining? Explain Web linkage, structures, Web contents, and Web access patterns.	4	1	5	1, 2
	b)	Identify any two applications of data mining in Retail Industry.	4	3	5	1, 2
16.	a)	Explain KDD process in detail.	4	2	1	1
	b)	Outline FP-Growth algorithm and analyze its merits and demerits.	4	3	2	1, 2
17.		Answer any *two* of the following:				
	a)	Briefly describe the following classifiers. i) K-NN ii) Support Vector Machines	4	2	3	1
	b)	Explain DBSCAN clustering algorithm.	4	2	4	1
	c)	Identify an application of data mining in business analysis.	4	3	5	1, 2

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

i)	Blooms Taxonomy Level – 1	20%
ii)	Blooms Taxonomy Level – 2	30%
iii)	Blooms Taxonomy Level – 3 & 4	50%

*****