

**PARUL UNIVERSITY**  
**FACULTY OF ENGINEERING & TECHNOLOGY**  
**B.Tech. Summer 2023 – 24 Examination**

**Semester: 6**  
**Subject Code: 203108361**  
**Subject Name: Data Mining and Data Visualization**

**Date: 29/04/2024**  
**Time: 10.30 am to 1.00 pm**  
**Total Marks: 60**

**Instructions:**

1. All questions are compulsory.
2. Figures to the right indicate full marks.
3. Make suitable assumptions wherever necessary.
4. Start new question on new page.

**Q.1 Objective Type Questions -**

(15)	CO	PO	Bloom's Taxonomy
1. Which algorithm is commonly used for association rule mining in KDD? a) K-means b) Apriori c) Decision Trees d) Support Vector Machines (SVM)	3	2	L4
2. Which of the following is NOT a type of clustering algorithm? a) K-means b) Hierarchical clustering c) Decision trees d) DBSCAN	3	2	L4
3. Which data visualization technique is best suited for showing the distribution of a single numerical variable? a) Histogram b) Box plot c) Scatter plot d) Pie chart	3	2	L4
4. Which of the following algorithms is commonly used for classification in data mining? a) K-means b) Apriori c) Naive Bayes d) DBSCAN	3	2	L4
5. Which of the following visualization tools is specifically designed for creating interactive visualizations? a) Tableau b) Matplotlib c) Seaborn d) Pandas	2	2	L4
6. _____step of the KDD process involves the selection of relevant data from the database.	1	1	L2
7. _____ is the measure of Apriori algorithm, by which we can identify strong and weak association rules.	3	1	L2
8. In DIKW pyramid, W stands for _____	3	1	L2
9. _____ is the technique, in which class label is unknown.	3	1	L2
10. Multiple numbers of data sources get combined in_____ step of the Knowledge Discovery.	1	1	L2
11. What do you mean by association rule mining?	3	1	L1
12. Define outlier.	3	1	L1
13. Define MinPts(Minimum points) with respect to DBSCAN algorithm.	3	1	L1
14. What are the measures of data dispersion?	4	1	L1
15. What are the measures of central tendency?	4	1	L1

**Q.2 Answer the following questions. (Attempt any three)**

(15)			
A) Enlist data visualization techniques. Explain any one of it.	3	1	L2
B) Explain agglomerative and divisive method, under hierarchical clustering.	3	2	L2
C) What are the major issues encountered in data mining?	1	1	L2
D) Enlist five applications of data mining. Explain any one of it in detail.	1	1	L2

**Q.3** A) How can we apply Binning method for data smoothing? Explain Binning by mean, Binning by median and Binning by boundaries with proper set of data. (07) 1 3 L3

B) For the given transaction table, find the frequent item sets and association rules using Apriori algorithm. Consider minimum support=2, minimum confidence=50%. (08) 3 2 L3

TID	Itemsets
T1	A, B
T2	B, D
T3	B, C
T4	A, B, D
T5	A, C
T6	B, C
T7	A, C
T8	A, B, C, E
T9	A, B, C

**OR**

B) For the given transaction table, generate frequent pattern tree and frequent pattern rules, using FP Growth algorithm. Consider minimum support be 3. (08) 3 2 L3

Transaction ID	Items
T101	{E,K,M,N,O,Y}
T102	{D,E,K,N,O,Y}
T103	{A,E,K,M}
T104	{C,K,M,U,Y}
T105	{C,E,I,K,O,O}

**Q.4** A) Explain the working of K-Means clustering algorithm with proper example. (07) 3 3 L3

**OR**

A) Define Information gain. For the given dataset, calculate Information gain for 'Age' attribute. (07) 3 2 L3

Age	Income	Student	Credit_Rating	Class : buys_computer
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31..40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31..40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31..40	Medium	No	Excellent	Yes
31..40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

B) How can we extract the useful, previously unknown, and potentially valuable information from large datasets? Explain entire process with the neat sketch. (08) 1 3 L2