

**Course Code** : 2101CS521**Date** : 23-05-2024**Course Name** : Data Mining**Duration** : 150 Minutes**Total Marks** : 70**Instructions:**

1. Attempt all the questions.
2. Figures to the right indicates maximum marks.
3. Make suitable assumptions wherever necessary.

**Q.1 (A)** Explain methods to find dissimilarity of numerical data with example. **4**

**(B)** Explain quantitative attributes with example. **3**

**OR**

Explain Qualitative attributes with example.

**(C)** Explain KDD process in detail. **7**

**OR**

Explain any seven issues of data mining.

**Q.2 (A)** What is data cleaning? Explain data cleaning steps in detail. **4**

**(B)** Explain binning method with an example. **3**

**OR**

Explain technique used for data Discretization.

**(C)** Explain data transformation with an example. **7**

**OR**

Explain data reduction with an example.

**Q.3 (A)** What is confidence, support and lift? Explain with an example. **4**

**(B)** Explain Maximal Frequent Itemsets with an example. **3**

**OR**

Explain Closed Frequent Itemsets with an example.

**(C)** Consider a transactional database. **7**

TID	Items
T1	A, B, C
T2	D, B, E
T3	A, D, B, E
T4	D, E

Suppose the minimum support count is 2 and minimum confidence threshold is 90%. Find all frequent itemsets using Apriori Algorithm and generate association rules.

**OR**

For the given transactional database

TID	Items
T1	1 2 3
T2	2 3 4
T3	4 5
T4	1 2 4
T5	1 2 3 5
T6	1 2 3 4

Find the frequent itemsets using FP-Growth algorithm. Assume that minimum support threshold=50% and Confidence= 60%.

**Q.4 (A)** Explain Naive Bayesian classification with an example. **4**

**(B)** Explain Accuracy, Error rate and Precision for classification model. **3**

**OR**

Explain Class imbalance problem with an example.

**(C)** Explain Decision Tree algorithm with an example. **7**

**OR**

Explain rule-based classification with an example.

**Q.5 (A)** What is clustering? Explain applications of clustering. **4**

**(B)** What is Outlier? Explain Challenges of Outlier Detection. **3**

**OR**

Explain Dendrogram with an example.

**(C)** Explain K-Mean with algorithmic steps and example. **7**

**OR**

Explain DBSCAN with algorithmic steps and example.

**\*\*\***