

**Gandhi Institute of Engineering and Technology University, Odisha, Gunupur
(GIET University)**



M.Tech. (First Semester – Regular/Supplementary) Examinations, January – 2026
24MCSPE11021 – Data Mining and Data Warehousing

Time: 3 hrs

Maximum: 60 Marks

**Answer ALL questions
(The figures in the right hand margin indicate marks)**

PART – A **(2 x 5 = 10 Marks)**

- Q.1. Answer **ALL** questions CO # Blooms Level
- a. In a transaction database of 100 transactions:30 transactions contain item A,40 transactions contain item B,20 transactions contain both A and B. Compute: Support of the itemset {A, B}, Confidence of the rule $A \rightarrow B$,Lift of the rule $A \rightarrow B$ CO1 K3
 - b. Consider the candidate 3-itemset:{A,B,C},{A, B, C}{A,B,C} .Its 2-item subsets are: {A,B},{A,C},{B,C} .Given that the frequent 2-itemsets are: $L_2=\{ \{A,B\},\{B,C\} \}$.Using the Apriori pruning principle, determine whether the candidate itemset {A, B, C} should be pruned or retained. Justify your answer. CO1 K3
 - c. What are the limitations of k-means clustering algorithm? CO2 K2
 - d. Explain various issues related to data cleaning. CO1 K2
 - e. Write short notes on outlier detection CO2 K2

PART – B **(10 x 5 = 50 Marks)**

- Answer ALL the questions Marks CO # Blooms Level
- 2. a. Define data warehouse. Draw the architecture of data warehouse and explain the three tiers in details with a case study? 5 CO1 K2
 - b. Consider the following transactions: If **the** minimum support count = 2, identify the frequent item sets using the FP-Growth principle 5 CO1 K3

TID	Items
T1	{A, B, C}
T2	{A, B}
T3	{A, C}
T4	{B, C}

(OR)

- c. The following data represents the values of two attributes X **and** Y collected from a dataset: 5 CO1 K3

X	2	4	6	8	10
Y	5	9	13	17	21

- 1. Compute the Pearson correlation coefficient (r) between X and Y.
- 2. State whether the correlation is positive, negative, or zero.
- d. Explain the data preprocessing techniques in detail. 5 CO1 K2
- 3.a. Both *k-means* and *k-medoids* algorithms can perform effective Clustering. Illustrate the strength and weakness of *k-means* in comparison with the *k-medoids* algorithm 5 CO2 K4

b. Consider the following dataset:

5 CO2 K3

Point	Coordinates
A	(1, 2)
B	(2, 1)
C	(4, 3)
D	(5, 4)

Let $K=2$, Assume the initial medoids are A (1,2) and C (4,3). Using Euclidean distance, perform K-Medoids clustering and identify the final medoids.

(OR)

- c. Compare single-linkage, complete-linkage, and average-linkage clustering techniques. 5 CO2 K4
- d. What is decision tree induction? List down the attribute selection measures used by the ID3 algorithm to construct a Decision Tree. 5 CO2 K3
- 4.a. Discuss real-world applications of time-series mining in finance, healthcare, and sensor networks. 5 CO4 K2
- b. Explain the class imbalance problem statement? How do you know that whether the dataset is imbalance or balance? Justify it with suitable example 5 CO4 K2

(OR)

- c. A binary classification dataset contains 1,000 instances, out of which 920 belong to the majority class and 80 belong to the minority class. A classifier produces the following results: True Positives (TP) = 40, False Negatives (FN) = 40, True Negatives (TN) = 900, False Positives (FP) = 20. Compute the Accuracy of the classifier. Compute the Precision, Recall, and F1-Score for the minority class. Comment on why accuracy is misleading in this imbalanced dataset. 5 CO3 K3
- d. Discuss real-world applications of time-series mining in sensor networks. 5 CO3 K2
- 5.a. Define Web Mining and explain its significance in knowledge discovery 5 CO4 K2
- b. List the major preprocessing steps in web usage mining. 5 CO4 K2

(OR)

- c. Differentiate between Web Content Mining, Web Structure Mining, and Web Usage Mining. 5 CO4 K4
- d. Explain challenges in mining multimedia data on the web. 5 CO4 K2
- 6.a. Define Web Mining and explain its taxonomy: web content mining, web structure mining, and web usage mining 5 CO4 K2
- b. Explain challenges in mining multimedia data on the web. 5 CO4 K2

(OR)

- c. Define Distributed Data Mining (DDM) and explain its architecture. 5 CO5 K2
- d. Explain Social Network Analysis (SNA) metrics such as degree centrality, betweenness, and closeness. 5 CO5 K2

--- End of Paper ---