



GIET UNIVERSITY, GUNUPUR - 765022
M. Tech (First Semester) Examinations, January- 2024
MPECS1045 - Data Mining and Data Warehousing
(CSE)

Time: 3Hrs

Maximum: 70 Marks

(The figures in the right hand margin indicate marks.)

PART – A**(2 x 10 = 20 Marks)**

Q.1. Answer all questions

	CO#	Blooms Level
a. How is data warehouse different from database? How are they similar?	CO1	K2
b. Explain the difference between OLAP and OLTP.	CO1	K2
c. Define frequent patterns in the context of data mining.	CO1	K2
d. Explain the ETL process in the context of data warehousing.	CO2	K2
e. What is the primary goal of classification in data mining?	CO3	K2
f. What is the fundamental idea behind hierarchical clustering?	CO2	K2
g. What is the role of training and testing datasets in classification?	CO2	K2
h. What are the challenges associated with mining time series data?	CO3	K2
i. What is web mining, and how does it contribute to data mining?	CO3	K2
j. Explain the challenges associated with mining unstructured data from the web.	CO4	K2

PART – B**(10 x 5=50 Marks)**Answer ANY FIVE questions

	Marks	CO#	Blooms Level
2. a. Explain the steps of KDD, with the help of a diagram.	5	CO1	K2
b. There are five transactions (T1, T2, T3, T4, T5) with items (A, B, C, D) purchased as T1(B, C), T2(A, C, D), T3(B, C), T4(A, B, C, D), T5(B, D). The min_sup=2. Show how Apriori Rule Mining Algorithm can generate the association rule for the above dataset.	5	CO2	K4
3.a. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):	5	CO2	K4
(a) Compute the Euclidean distance between the two objects.			
(b) Compute the Manhattan distance between the two objects.			
(c) Compute the Minkowski distance between the two objects, using q = 3.			
(d) Compute the supremum distance between the two objects			
b. Explain the importance of LIFT in association rule mining.	5	CO2	K2

4. a. Apply the DBSCAN algorithm with the similarity threshold 0.8 to the given data points and $\text{Minpoints} \geq 2$ (Minimum required point in the cluster). What are the boarder points, Nosie(outlier) in the set of point given in the table

	P1	P2	P3	P4	P5
P1	1.00	0.10	0.41	0.55	0.35
P2	0.10	1.00	0.64	0.47	0.98
P3	0.41	0.64	1.00	0.44	0.85
P4	0.55	0.47	0.44	1.00	0.76
P5	0.35	0.98	0.85	0.76	1.00

- b. Explain how outlier detection is performed in the context of time series data? 3 CO2 K4
- 5.a. Find the proximity measure of the following tabular data: - 10 CO2 K4

Text1 (Nominal)	Text2 (Ordinal)	Text3 (Numerical)
A	EXCELLENT	45
B	FAIR	22
C	GOOD	64
A	EXCELLENT	28

6. a. Write short note on: 10 CO2 K4
- (i) Issues regarding classification and prediction
- (ii) Outlier Analysis
- 7.a. Discuss the applications of social network analysis in different domains. 5 CO3 K3
- b. Describe the challenges associated with privacy and ethical considerations in social network analysis. 5 CO4 K2
8. a. Explain the differences between web content mining, web structure mining, and web usage mining. 5 CO4 K2
- b. Describe the applications of multimedia data mining in social media. 5 CO4 K2

--- End of Paper ---