QPC: RN23PHD396 AY 23 Reg. No | | | | | | | | | | |

GIET UNIVERSITY, GUNUPUR – 765022

Ph.D. (First Semester) Examinations, January – 2024

23SPPECS1011 - Data Mining and Data Warehousing

(CSE)

Time: 3 hrs Maximum: 70 Marks

The figures in the right hand margin indicate marks.

Answer ANY FIVE Questions **(14 x 5 = 70 Marks)**

		Marks
1.a.	What is metadata, and why is it essential in a Data Warehouse?	7
b.	Discuss the role of indexing in optimizing queries in a Data Warehouse.	7
2.a.	What is support and confidence in association rule mining? How are they used to evaluate the quality of rules?	7
b.	Describe the use of association rule mining in real-world applications. Provide examples.	7
3.a.	Discuss the scalability issues associated with sequential pattern mining algorithms.	7
b.	Describe the k-nearest neighbors (KNN) algorithm for classification. What are its strengths and weaknesses?	7
4.a.	How does logistic regression differ from decision trees in classification tasks?	7
b.	Discuss hierarchical methods in cluster analysis. How do agglomerative and divisive approaches differ?	7
5.a.	Define transactional patterns in the context of data mining.	7
b.	Define periodicity in time series data. How is it identified and analyzed?	7
6.a.	Discuss the importance of feature extraction in time series data analysis.	7
b.	Define data streams and explain the challenges associated with mining them.	7
7.a.	What is the role of incremental algorithms in sequential pattern mining for data streams?	7
b.	Explain the concept of graph pattern matching and its applications.	7
8.a.	Discuss the challenges associated with web mining, especially in dealing with large-scale and dynamic web data.	7
b.	Explain the challenges of mining multimedia data on the web, including issues related to scalability and diversity.	7

---End of Paper---