Roll No. | | | | | | | | | |

(178) SET-II

COER University

END SEMESTER EXAMINATION, EVEN SEMESTER, 2023-24

Time : 3-hour **Semester** : VI **Total Marks : 100**
Program Name : B.Tech CSE **Branch/Specialization : CSE**
Course Code : SOC314 **Course Name : Data Mining and Warehousing**

Note: All questions are compulsory. No student is allowed to leave the examination hall before the completion of the time.

Q. No 1	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Explain the term Data Mining with example.	CO 1	2
(b)	Apply the two methods below to normalize the following group of data: 200, 300, 400, 600, and 1000. a) Use min-max normalization by setting min=0 and max=1.0 b) Z-score normalization	CO 1	3
(c)	Define data integration and transformation in data mining.	CO 1	2
(d)	Explain the Data Discretization and Concept hierarchy generation with suitable example.	CO 1	2
(e)	Define data characterization and data discrimination with example.	CO 1	2

Q. No 2	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Consider the data Set D with following transactions. **Transaction ID** — **Items** 100 — A, C, D 200 — B, C, E 300 — A, B, C, E 400 — B, E Using the threshold values support = 2 and confidence=50%. Evaluate the following. a) All Frequent item sets in database D using Apriori Algorithm. b) Strong Association rules for database D.	CO 2	5
(b)	Define frequent sets, Confidence and support countwith suitable example.	CO 2	2
(c)	Explain Association mining to correlation analysis with example.	CO 2	2
(d)	Explain multilevel and multidimensional association rule with suitable example.	CO 2	2
(e)	Discuss in detail about the quantitative association rule mining.	CO 2	2

Q. No 3	Attempt Any Four Parts. Each Question Carries 5 Marks.	CO	BL
(a)	Define classification with an example. Also, discuss different challenges occur during classification process?	CO 3	2
(b)	Define backward and forward propagation in neural network with neat and clean diagram.	CO 3	2
(c)	Define confusion matrix with an example. Also discuss precision, recall, accuracy and F1-score performance measures.	CO 3	2
(d)	How Bayesian classification is based on Bayes' theorem and probabilistic reasoning? Discuss the application of Bayesian classification in text classification and spam filtering.	CO 3	3
(e)	Explain density-based clustering method based on connected regions (DBSCAN) algorithm with an example.	CO 3	2

Q. No 4	Attempt Any Two Parts. Each Question Carries 10 Marks.	CO	BL
(a)	Define logistic regression with example? How logistic regression method is different from linear regression. Also, explain the loss function used in logistic regression.	CO 4	2
(b)	Define exploratory data analysis. Also, define the steps involved in exploratory data analysis.	CO 4	2
(c)	Elaborate the following: a) Web Mining b) Text Mining c) Multimedia Data Mining	CO 4	2

Q. No 5	Attempt Any Two Parts. Each Question Carries 10 Marks.	CO	BL
(a)	Describe Data warehouse. Also, define the architecture of data warehouse along with its with advantages and disadvantages.	CO 5	
(b)	Elaborate the following OLAP function: a) Roll-up b) Drill-down c) Slicing d) Dicing	CO 5	
(c)	Explain online analytical processing (OLAP) tool. How OLAP technology works and also discuss its benefits.	CO 5	

----------End of Paper----------