K. J. Somaiya Institute of Technology, Sion, Mumbai
(An Autonomous Institute Affiliated to the University of Mumbai)

End Semester Exam
April – May 2023

B.Tech. (Information Technology)

Examination: TY - Semester VI

Course Code: ITC601 **Course Name:** Data Mining and Business Intelligence
Date: May 12, 2023 **Duration:** 2.5 Hours **Max. Marks:** 60

Instructions:
(1) All questions are compulsory.
(2) Draw neat diagrams wherever applicable.
(3) Assume suitable data, if necessary.

Ques. No.	Question	Max. Marks	CO	BT Level
Q1.	**Solve any six questions out of eight:**	**12**		
i)	Explain Roll up operation with example.	2	CO1	U
ii)	Discretize the data of Sales Quantity: 42, 76, 405, 79, 89, 104, 35, 115, 85, 166, 51, 172, 190 into 04 bins using equal-frequency partitioning and perform smoothing by bin means.	2	CO2	A
iii)	Explain different methods for handling missing data.	2	CO2	U
iv)	Explain two-step process of Classification.	2	CO3	U
v)	Explain Holdout and Cross-validation approaches of evaluation.	2	CO3	U
vi)	Differentiate Classification and Clustering.	2	CO4	U
vii)	Explain Support and Confidence with example.	2	CO5	U
viii)	Explain the advantages of Business Intelligence systems.	2	CO6	U
Q2.	**Solve any four questions out of six:**	**16**		
i)	Compare OLAP and OLTP.	4	CO1	U
ii)	Suppose a hospital tested heart rates of 10 randomly selected adults after a cardio exercise, and recorded values: 79, 80, 85, 91, 92, 92, 94, 96, 98, 102. Calculate the 5-number summary.	4	CO2	A
iii)	Differentiate Classification and Prediction with examples. Explain in detail the assumption made by Naïve Bayes classifier.	4	CO3	U
iv)	Explain Core Points, Border Points, and Noise Points in DBSCAN algorithm using examples.	4	CO4	U
v)	Sketch a Concept Hierarchy for items in an Electronics Store and explain its use for Multi-level Association Rule Mining.	4	CO5	A
vi)	Explain phases of decision-making process.	4	CO6	U
Q3.	**Solve any two questions out of three:**	**16**		
i)	Sketch and explain the process of Knowledge Discovery from Data.	8	CO1	U
ii)	A survey was conducted to analyse if a restaurant's ambience has a correlation with the ratings given by customers. Apply Chi-square test to analyze whether the attributes *Restaurant_Ambience* & *Customer_Ratings* are correlated or not:	8	CO2	A

Restaurant_Ambience		*Customer_Ratings*		Total
		Good	Average	
	Yes	250	300	550
	No	50	1100	1150
Total		300	1400	1700

Consider that for 1 degrees of freedom, the chi-square value to reject the hypothesis at 0.001 significance level is 10.828.

iii)	Apply Decision Tree induction algorithm on the below data and find the best attribute to be selected as a root node. **Temperature** / **Humidity** / **Wind** / **Play** Hot / High / Weak / No Hot / High / Weak / Yes Mild / Normal / Strong / Yes Mild / High / Strong / Yes Mild / High / Strong / No Cool / Normal / Strong / No Mild / High / Weak / Yes Hot / High / Strong / No Hot / Normal / Weak / Yes Mild / High / Strong / No	8	CO3	A
Q4.	**Solve any two questions out of three:**	**16**		
i)	Apply agglomerative algorithm for Hierarchical Clustering of the below spatial data points using single link distance: **Data Point** / **X-Coordinate** / **Y-coordinate** A / 1 / 1 B / 2 / 2 C / 6 / 6 D / 4 / 4 E / 3 / 4 Sketch the Dendrogram tree and explain the same.	8	CO4	A
ii)	Consider a transactional dataset: **TID** / **Items** T1 / Bread, Cheese, Biscuits, Juice T2 / Bread, Cheese, Juice T3 / Bread, Milk, Yogurt T4 / Bread, Juice, Milk T5 / Cheese, Juice, Milk If minimum support = 50% and minimum confidence = 75%, find all possible association rules using Apriori algorithm. Elaborate on rule interestingness and use of lift measure to evaluate the rules..	8	CO5	A
iii)	Consider the case of Diabetes prediction and apply KDD process to derive Business Intelligence. Clearly explain each phase / operation in the KDD process with respect to the stated application.	8	CO6	A

****