Lagdu Singh Charitable Trust's (Regd.)
THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY
Autonomous College Affiliated to University of Mumbai
Approved by All India Council for Technical Education(AICTE) and Government of Maharashtra(GoM)
Conferred Autonomous Status by University Grants Commission (UGC) for 10 years w.e.f. A.Y 2019-20
Among Top 250 Colleges in the Country where NIRF India Ranking 2020 in Engineering College category
• ISO 9001:2015, 14001:2015, 50001:2018 Certified • Programmes Accredited by National Board of Accre
• Institute Accredited by National Assessment and Accreditation Council (NAAC), Bangalore
Website : www.tctmumbai.in
tcet Estd. 2001

IN-SEMESTER EXAMINATION-I

T.E. (Semester-VI)

SUBJECT – Data Mining and Business Intelligence

Branch: Information Technology
Div.: ALL
Duration: 60 Minutes

Date:10/03/2025
Timing: 4:00 TO 5:00 PM
Maximum Marks: 20

Instructions –

1. All questions are compulsory.
2. Assume suitable data wherever necessary and state the assumptions made.
3. Diagrams / sketches should be given wherever necessary.
4. Use of logarithmic table, drawing instruments and non-programmable calculators is permitted.
5. Figures to the right indicate full marks.

Q.1		Answer any 5 of the following questions	Marks	Course Outcomes	Learning Levels
	a.	How do data warehouses differ from transactional databases in the context of data mining?	2	1	U
	b.	Explain the concept of data mart and its role in a data warehousing environment.	2	1	E
	c.	Define the term Data Mining	2	1	R
	d.	Define the following terms: Support, Confidence	2	2	R
	e.	Justify the statement: Data can be dirty.	2	2	E
	f.	Illustrate vertical data format?	2	3	U
	g.	What is FP-Growth? Where is it applied?	2	3	U
Q.2	a.	Suppose a group of 12 sales price records has been sorted as follows:5; 10; 11; 13; 15; 35; 50; 55; 72; 92; 204; 215 Partition them into three bins by each of the following methods (a)equal-width partitioning (b) equal-frequency partitioning	5	2	U
		OR			
	b.	Describe the major tasks of data pre-processing	5	2	U
Q.3	a.	Consider the given transactions for the items purchased. If minimum support is 50% then find the itemsets which satisfy the min support. If minimum confidence is 50% then which rules are generated?	5	3	AN

Transaction ID	Items Bought
1	Shoes, Shirt, Jacket
2	Shoes,Jacket
3	Shoes, Jeans
4	Shirt, Sweatshirt

		OR			
	b.	Describe the various applications of frequent pattern mining	5	3	U

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

IN-SEMESTER EXAMINATION-II
T.E. (Semester-VI)
SUBJECT – Data Mining and Business Intelligence

Branch: Information Technology
Div.: ALL
Duration: 60 Minutes

Date:15/04/2025
Timing: 2:00 TO 3:00 PM
Maximum Marks: 20

Instructions –

1. All questions are compulsory.
2. Assume suitable data wherever necessary and state the assumptions made.
3. Diagrams / sketches should be given wherever necessary.
4. Use of logarithmic table, drawing instruments and non-programmable calculators is permitted.
5. Figures to the right indicate full marks.

Q.1		Answer any 5 of the following questions	Marks	Course Outcomes	Learning Levels
	a.	Define classification in data mining and describe its primary objective.	2	4	R
	b.	Give any two difference between Classification and Regression	2	4	AN
	c.	Provide the applications of Clustering techniques.	2	5	R
	d.	Describe silhouette coefficient method in clustering algorithms	2	5	U
	e.	Explain the role of Mathematical Model in Business Intelligence.	2	6	U
	f.	Define decision support system.	2	6	R
	g.	List down different components of Business Intelligence System.	2	6	R
Q.2	a.	Explain the accuracy and coverage in Rule based Classifier.use the following rule for purchase prediction:"If the customer is female, aged between 25-35, and has made more than 3 purchases in the last month, predict a purchase." | Customer ID | Gender | Age | Purchases Last Month | Purchase Prediction | |---|---|---|---|---| | 1 | Male | 30 | 2 | No | | 2 | Female | 25 | 4 | Yes | | 3 | Male | 40 | 1 | No | | 4 | Female | 30 | 5 | Yes | | 5 | Female | 35 | 3 | Yes | | 6 | Male | 28 | 0 | No | | 7 | Female | 22 | 2 | No | | 8 | Male | 45 | 3 | Yes | | 9 | Female | 32 | 6 | Yes | | 10 | Male | 38 | 4 | Yes |	5	4	AN
		OR			
	b.	Elaborate the different Types of Classification algorithms.	5	4	R
Q.3	a.	Discuss the advantage and disadvantage of K-Medoid Algorithm	5	5	R
		OR			
	b.	Differentiate between Agglomerative and Divisive appraches with diagram and example.	5	5	AN

THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

END SEMESTER EXAMINATION, MAY 2025
T.E. SEMESTER VI (CBCGS-HME 2023)

Branch:	**Information Technology (IT)**	**Q.P. Code:**	**E665RM252-1**
Subject:	**Data Mining and Business Intelligence**	**Duration:**	**2 hours**
Subject Code:	**PCC-IT 601**	**Max. Marks:**	**60**

Instructions: 1. All sections are compulsory
2. Figures to the right indicate full marks.
3. Assume suitable data if necessary and state the assumptions clearly.

Section-I **Short Answer Questions (Answer any 05 questions out of 06)** **(10 Marks)**
(Fundamental, Core Types)

Q. No.		Marks	CO	RBT Level	PI
1	Describe the concept of characterization and discrimination in data mining.	2	1	U	2.7.1
2	Explain data correlation? Name the method to test the correlation analysis.	2	2	U	3.6.1
3	Define frequent pattern	2	3	R	1.6.1
4	Explain the importance of cross-validation in evaluating classification models and provide an example of its application.	2	4	U	3.6.1
5	Compare the computational efficiency of K-Means and K-Medoids.	2	5	R	2.6.5
6	Explain the phases involved in the Decision-Making process.	2	6	A	1.2.1

Section-II **Descriptive Answer Questions (Answer any 04 out of 06)** **(20 Marks)**
(Descriptive, Comprehension Types)

1	Differentiate between structured, semi-structured, and unstructured data in the context of data mining.	5	1	AN	4.4.1
2	Suppose a group of 12 sales price records has been sorted as follows:5; 10; 11; 13; 15; 35; 50; 55; 72; 92; 204; 215 Partition them into three bins by each of the following methods (a)equal-width partitioning (b) equal-frequency partitioning	5	2	A	2.5.3
3	Design FP-Tree for following dataset. Assume support count = 2.	5	3	AN	2.8.1
4	Consider the scenario of classifying emails as either spam or not spam (ham) based on the presence of certain keywords. Use the Naive Bayes algorithm to calculate the prior and posterior probabilities for the following training data: Total emails: 1000,	5	4	A	2.8.1

Transaction ID	Items
T1	Blood Test, X-Ray, ECG
T2	Blood Test, Urine Test
T3	X-Ray, ECG, Blood Pressure
T4	Blood Test, X-Ray
T5	Urine Test, Blood Test, Blood Pressure
T6	ECG, Blood Test

	Spam emails: 300,Ham emails: 700, Keyword "offer" appears in 100 spam emails and 20 ham emails,Keyword "discount" appears in 200 spam emails and 50 ham emails. Classify a new email that contains the words "offer discount" as spam or ham.				
5	For K=2, build 2 cluster using Kmeans for 2 D data Object: / Attribute 1(x): / Attribute 2(y): A: 2, 2 B: 3, 2 C: 1, 1 D: 3, 1 E: 1.5, 0.5	5	5	A	2.5.3
6	Analyze the different BI tools and strategies contribute to organizational growth. Support your answer with two real-world examples.	5	6	U	5.5.1

Section-III **Long Answer Question (Answer any 03 out of 05)** **(30 Marks)**
(Application, Analytical, Evaluation, Design Type)

1	Illustrate the different types of attributes in data mining with suitable examples. How do they impact statistical analysis and data preprocessing?	10	2	A	1.2.1,2.5.1
2	Design a FP tree for the following dataset and also find frequent itemsets with support count = 2. TID / List of Item_IDs T10: I1, I2, I5 T20: I2, I4 T30: I2, I3 T40: I1, I2, I4 T50: I1, I3 T60: I2, I3 T70: I1, I3 T80: I1, I2, I3, I5 T90: I1, I2, I3	10	3	AN	2.8.1
3	Explain the concept of classification in Data Mining. Describe the process of building a classification model and discuss its key components.	10	4	U	3.6.1
4	Explain Density-Based Clustering in detail with following parameters: a. ε-Neighborhood b. Minimum Points c. Core Point d. Border Points e. Outlier Ponts	10	5	U	1.7.1

5	Analyze the development and implementation of a Business Intelligence (BI) system in a Banking & Finance industry. Your analysis should cover system architecture, data integration, visualization tools, challenges encountered, and the impact on organizational decision-making and performance.	10	6	AN	5.4.2