THAKUR COLLEGE OF ENGINEERING & TECHNOLOGY

IN-SEMESTER EXAMINATION-I 05 MAR 2026

TE (Semester-VI)

SUBJECT – Data Mining and Business Intelligence

Date: 05/03/2026
Timing: 02:00 pm to 03:00 pm
Maximum Marks: 20

Branch: IT
Div.: ALL
Duration: 60 Minutes
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	CO	Learning Levels
	a.	Describe any two types of patterns that can be discovered using data mining techniques.	2	CO1	U
	b.	Outline the concept of characterization and discrimination in data mining.	2	CO1	R
	c.	Distinguish between OLTP and OLAP System. (Any 2)	2	CO1	U
	d.	Define data exploration and state one objective of data exploration in data mining.	2	CO2	R
	e.	Illustrate the statement: Data can be dirty.	2	CO2	AN
	f.	Describe vertical data format with example.	2	CO3	U
	g.	In a transaction database of 100 transactions: Support(A) = 40 Support(B) = 50 Support(A ∩ B) = 30 Calculate the Lift of the rule A → B.	2	CO3	A
Q.2	a.	Use the two methods below to normalize the following group of data:200, 300, 400, 600, 1000 (a) min-max normalization by setting min = 0 and max = 1 (b) z-score normalization"	5	CO2	A
		OR			
	b.	The data set is given as: {5, 40, 42, 46, 48, 49, 50, 50, 52, 53, 55, 56, 58, 75, 102}; a) What is mean and median of the data? b) What is the mode of the data? Comment on data's modality. c) What is the midrange of the data? d) Find the Q1 and Q3 of the data e) Give the five-number summary of the data and also show a boxplot of the data.	5	CO2	A
	a.	Compare and contrast between FP tree Growth Algorithm and Apriori algorithm	5	CO3	AN
		OR			
	b.	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	CO3	A

Transaction ID	Items
T1	Milk, Bread, Butter
T2	Bread, Diaper, Beer, Eggs
T3	Milk, Diaper, Beer, Coke
T4	Bread, Milk, Diaper, Beer
T5	Bread, Milk, Diaper, Coke

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

Branch: IT
Div.: ALL
Duration: 60 Minutes
Date: 21/04/2026
Timing: 12:00 PM to 1:00 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.

<table>
<tr><th>Q.1</th><th></th><th>Answer any 5 of the following questions</th><th>Marks</th><th>Course Outcomes</th><th>Learning Levels</th></tr>
<tr><td></td><td>a.</td><td>Contrast pre-pruning and post-pruning techniques in decision tree learning.</td><td>2</td><td>CO4</td><td>AN</td></tr>
<tr><td></td><td>b.</td><td>Specify the concepts of coverage and accuracy of a classification rule.</td><td>2</td><td>CO4</td><td>U</td></tr>
<tr><td></td><td>c.</td><td>Contrast partitioning-based and hierarchical clustering methods.</td><td>2</td><td>CO5</td><td>AN</td></tr>
<tr><td></td><td>d.</td><td>Analyze the computational efficiency of K-Means and K-Medoids.</td><td>2</td><td>CO5</td><td>AN</td></tr>
<tr><td></td><td>e.</td><td>Clarify the role of a Decision Support System in Business Intelligence.</td><td>2</td><td>CO6</td><td>U</td></tr>
<tr><td></td><td>f.</td><td>Examine the importance of BI in market segmentation for the retail industry.</td><td>2</td><td>CO6</td><td>AN</td></tr>
<tr><td></td><td>g.</td><td>Investigate how BI supports decision-making in banking and finance applications.</td><td>2</td><td>CO6</td><td>AN</td></tr>
<tr><td>Q.2</td><td>a.</td><td>Illustrate the different types of classification methods in data mining.</td><td>5</td><td>CO4</td><td>AN</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>b.</td><td>Interpret the concepts of accuracy and coverage in a rule-based classifier using the given purchase prediction dataset and rule."If the customer is female, aged between 25-35, and has made more than 3 purchases in the last month, predict a purchase."
<table>
<tr><th>Customer ID</th><th>Gender</th><th>Age</th><th>Purchases Last Month</th><th>Purchase Prediction</th></tr>
<tr><td>1</td><td>Male</td><td>30</td><td>2</td><td>No</td></tr>
<tr><td>2</td><td>Female</td><td>25</td><td>4</td><td>Yes</td></tr>
<tr><td>3</td><td>Male</td><td>40</td><td>1</td><td>No</td></tr>
<tr><td>4</td><td>Female</td><td>30</td><td>5</td><td>Yes</td></tr>
<tr><td>5</td><td>Female</td><td>35</td><td>3</td><td>Yes</td></tr>
<tr><td>6</td><td>Male</td><td>28</td><td>0</td><td>No</td></tr>
<tr><td>7</td><td>Female</td><td>22</td><td>2</td><td>No</td></tr>
<tr><td>8</td><td>Male</td><td>45</td><td>3</td><td>Yes</td></tr>
<tr><td>9</td><td>Female</td><td>32</td><td>6</td><td>Yes</td></tr>
<tr><td>10</td><td>Male</td><td>38</td><td>4</td><td>Yes</td></tr>
</table></td><td>5</td><td>CO4</td><td>AN</td></tr>
<tr><td>Q.3</td><td>a.</td><td>Implement K-Medoids clustering for K = 2 on the given dataset D={2,4,5,7,9,12}</td><td>5</td><td>CO5</td><td>A</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>b.</td><td>Identify different types of outliers and provide two examples for each</td><td>5</td><td>CO5</td><td>A</td></tr>
</table>

END SEMESTER EXAMINATION, May 2026
T.E. SEMESTER VI (CBCGS-HME 2023)

Branch:	**INFORMATION TECHNOLOGY**	**Q.P. Code:**	**E665RM262 -1**
Subject:	**DATA MINING & 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.	Illustrate how data mining techniques can be applied in different real-world scenario	2	1	A	5.6.1
2.	Apply min–max normalization to normalize the value 25 when minimum = 10 and maximum = 50.	2	2	A	1.6.1
3.	Analyze the Apriori property and its role in reducing the search space.	2	3	AN	2.6.3
4.	Let's look at the numbers for the hospital's cancer screening model tested on 1,000 patients. Because the cancer is rare, the dataset is highly imbalanced: 10 patients actually have cancer.990 patients are perfectly healthy. Compute Accuracy.	2	4	A	4.6.4
5.	Examine the key challenges in detecting outliers in high-dimensional data.	2	5	AN	1.7.1
6.	Identify any two components of a Business Intelligence architecture.	2	6	A	1.6.1

	Predicted: Cancer (Positive)	Predicted: Healthy (Negative)	Total
Actual: Cancer (Positive)	7 (True Positives)	3 (False Negatives)	10
Actual: Healthy (Negative)	5 (False Positives)	985 (True Negatives)	990
Total	12	988	1000

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

1.	You are the data analyst on the project team building a data warehouse for an insurance company. List the possible data sources from which you will bring the data into your data warehouse. State your assumptions. Draw Star Schema for the same.	5	1	A	5.6.1
2.	For the given data {2,6,7,8,8,11,12,13,14,15,22,23} find the following: a) What is mean and median of the data? b) What is the mode of the data? Comment on data's modality. c) What is the midrange of the data? d) Find the Q1 and Q3 of the data	5	2	A	2.5.3

	e) Give the five-number summary of the data and also show a boxplot of the data.				
3.	Differentiate between Frequent Patter Tree (FP)-Growth and Apriori Algorithm	5	3	AN	2.5.1
4.	A rule-based classification model has been developed to detect fraudulent credit card transactions. Out of 10,000 tested transactions, the model correctly identified 150 fraudulent transactions and correctly flagged 9,700 legitimate transactions. However, it missed 50 fraudulent transactions and falsely accused 100 legitimate transactions of being fraud. Analyze the given credit card fraud detection scenario by constructing the confusion matrix and calculating Accuracy, Precision, Recall, and F1-Score.	5	4	A	3.6.1
5.	Construct clusters using the DBSCAN algorithm for the given parameters (MinPoints = 4, Epsilon = 1.9).	5	5	A	2.5.3
6.	Justify the significance of historical data analysis in Business Intelligence applications.	5	6	AN	5.5.1

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
P1	0	K	K	K	K	K	K	K	K	K	K	K
P2	1.4	0	K	K	K	K	K	K	K	K	K	K
P3	2.8	1.4	0	K	K	K	K	K	K	K	K	K
P4	4.2	2.8	1.4	0	K	K	K	K	K	K	K	K
P5	5.7	4.2	2.8	1.4	0	K	K	K	K	K	K	K
P6	5.8	4.5	3.2	2	1.4	0	K	K	K	K	K	K
P7	6.4	5	3.6	2.2	1	1	0	K	K	K	K	K
P8	5.8	4.5	3.2	2	1.4	2.8	2.2	0	K	K	K	K
P9	4	3.2	2.8	3.2	4	3.2	4.1	5.1	0	K	K	K
P10	1.4	2	3.2	4.5	5.8	5.7	6.4	6.3	3.2	0	K	K
P11	2	1.4	2	3.2	4.5	4.2	5	5.1	2	1.4	0	K
P12	3.2	2.8	3.2	4	5.1	4.5	5.4	6	1.4	2	1.4	0

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

1.	Briefly outline how to compute the dissimilarity between objects described by the following: a) Nominal attributes b) Symmetric and Asymmetric binary attributes c) Ordinal attributes	10	2	AN	2.8.1
2.	Design a FP tree for the following dataset and also find frequent itemsets with support count = 2.	10	3	A	2.8.1
3.	Develop a decision tree classification model for the given dataset and classify the tuple: Refund: No, Marital Status: Married, Taxable Income: 80K, Cheat=?	10	4	A	2.5.3

TID	*items_bought*
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I ,E}

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

4.	Consider the following dataset with 6 points: P1(1,2),P2(2,3),P3(5,4),P4(6,7),P5(8,8),P6(9,5) Implement the K-Medoids algorithm with k = 2 on the given dataset and determine updated medoids after three iterations.	10	5	A	2.5.3
5.	Correlate how Business Intelligence supports customer relationship management with improved customer retention.	10	6	AN	5.4.1