

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER-VI EXAMINATION – WINTER 2025

Subject Code:3160714

Date:21-11-2025

Subject Name:Data Mining

Time:02:30 PM TO 05:00 PM

Total Marks:70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

		MARKS
Q.1	(a) What is data mining? List out the data mining functionality.	03
	(b) Briefly explain the issues in data mining.	04
	(c) Explain frequent item-set mining algorithm with suitable example.	07
Q.2	(a) Explain k-fold cross validation.	03
	(b) Discuss dimensionality reduction techniques.	04
	(c) Explain KDD process.	07
OR		
	(c) Explain data pre-processing functionalities.	07
Q.3	(a) What is spatial data mining? Explain with its application.	03
	(b) Discuss support, confidence, frequent item-set and strong association rules.	04
	(c) Explain decision tree algorithm with suitable example.	07
OR		
Q.3	(a) Explain classification Vs. Prediction.	03
	(b) Explain proximity measures used in distance based classifiers.	04
	(c) Explain ANN.	07
Q.4	(a) What is missing values and noise? How to handle them?	03
	(b) Differentiate linear and logistic regression.	04
	(c) Explain hierarchical and agglomerative method of clustering.	07
OR		
Q.4	(a) What is closed item-set and maximal frequent item-set?	03
	(b) Define accuracy, precision, recall and G-mean.	04
	(c) Discuss spatial Data Mining, Temporal Mining, and Multimedia Mining with their application.	07
Q.5	(a) Briefly explain outlier analysis.	03
	(b) Differentiate clustering and classification.	04
	(c) Explain K-means algorithm.	07
OR		
Q.5	(a) How to evaluate clustering algorithm?	03
	(b) Discuss application of web mining.	04
	(c) Explain PAM algorithm.	07

GUJARAT TECHNOLOGICAL UNIVERSITY

BE- SEMESTER-VI (NEW) EXAMINATION – WINTER 2024

Subject Code:3160714

Date:02-12-2024

Subject Name:Data Mining

Time:02:30 PM TO 05:00 PM

Total Marks:70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

- Q.1** (a) Define each of the following data mining functionalities: characterization, discrimination, regression. **03**
- (b) How is a data warehouse different from database? **04**
- (c) Explain KDD process. **07**

- Q.2** (a) How to handle missing values? Explain. **03**
- (b) How to handle noisy data? **04**
- (c) Consider a database, **D**, consisting of 9 transactions. **07**
 Suppose min. support count required is **2**.
 Let minimum confidence required is **70%**.
 Find out the frequent itemset using Apriori algorithm.

TID	List of Items
T100	I1, I2, I5
T100	I2, I4
T100	I2, I3
T100	I1, I2, I4
T100	I1, I3
T100	I2, I3
T100	I1, I3
T100	I1, I2, I3, I5
T100	I1, I2, I3

OR

- (c) A database has five transactions. Let min sup=60% and min conf =80%. **07**

Tid	Item brought
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}

Find all frequent item sets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.

- Q.3** (a) Explain market basket analysis. **03**
- (b) Explain Linear regression. **04**
- (c) Explain decision tree algorithm. **07**

OR

- Q.3** (a) Explain WEKA tool. **03**
(b) Explain logistic regression. **04**
(c) Explain CART Classification Method. **07**
- Q.4** (a) Compare classification and Clustering. **03**
(b) Which metrics used for evaluating classifier performance? **04**
(c) Explain Principal Component Analysis. **07**

OR

- Q.4** (a) Compare classification and prediction. **03**
(b) Explain outlier detection. **04**
(c) Explain Backpropagation algorithm. **07**
- Q.5** (a) Write applications of clustering graph and network data. **03**
(b) What is Web log structure? And discuss issues regarding web logs. **04**
(c) Explain PAM clustering Algorithm. **07**

OR

- Q.5** (a) Write similarity measures for clustering graph and network data. **03**
(b) Explain Web Structure mining. **04**
(c) Write Applications of Distributed and parallel Data Mining. **07**

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-VI (NEW) EXAMINATION – WINTER 2023****Subject Code:3160714****Date:11-12-2023****Subject Name:Data Mining****Time:02:30 PM TO 05:00 PM****Total Marks:70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

		Marks
Q.1	(a) Define Data Warehouse. State it's features.	03
	(b) Differentiate between OLAP and OLTP.	04
	(c) Explain in detail different steps of KDD process.	07
Q.2	(a) Why to preprocess the data in data Mining?	03
	(b) Explain Binning method with the help of example.	04
	(c) Explain following terms related to Association Rule Mining: Itemset, Support Count, support, and Association rule.	07

Transaction ID	Items
1	Bread, Milk
2	Bread, Chocolate, Pepsi, Eggs
3	Milk, Chocolate, Pepsi, Coke
4	Bread, Milk, Chocolate, Pepsi
5	Bread, Milk, Chocolate, Coke

For given example find support & confidence for

{Milk, Chocolate} \Rightarrow Pepsi.

{Milk, Pepsi} \rightarrow {Chocolate}

{Chocolate, Pepsi} \rightarrow {Milk}

OR

- (c) Solve the following problem using Apriori algorithm. **07**
 Find the frequent itemsets and generate association rules on this.
 Assume that minimum support threshold ($s = 33.33\%$), minimum confident threshold ($c = 60\%$), minimum support count=2.

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

- Q.3** (a) Define the following terms in Data Transformation: **03**
- i. Smoothing
 - ii. Normalization

	iii. Discretization	
	(b) Differentiate between Classification and Prediction.	04
	(c) Explain Decision Tree Classification algorithm with the help of example.	07
	OR	
Q.3	(a) Differentiate between supervised learning and unsupervised learning.	03
	(b) What is Regression? Explain Linear Regression in short.	04
	(c) Explain Naïve Bayes Classifier with example.	07
Q.4	(a) What do you mean by Tree Pruning? Explain with example.	03
	(b) Explain the following as attribute selection measure:	04
	(i) Information Gain	
	(ii) Gain Ratio	
	(c) What do you mean by learning-by-observation? Explain k-Means clustering algorithm in detail.	07
	OR	
Q.4	(a) Define Data Cube. Explain any two operations on it.	03
	(b) Differentiate between Partition method and Hierarchical method of Clustering.	04
	(c) What are the requirements of Clustering in Data Mining?	07
Q.5	(a) How K-Mean clustering method differs from K-Medoid clustering method?	03
	(b) Draw and explain the topology of a multilayer, feed-forward Neural Network.	04
	(c) Explain the major issues in data mining.	07
	OR	
Q.5	(a) Give difference between text mining and web mining.	03
	(b) Why Hadoop is important?	04
	(c) What is web log? Explain web structure mining and web usage mining in detail.	07

GUJARAT TECHNOLOGICAL UNIVERSITY**BE - SEMESTER-VI(NEW) EXAMINATION – WINTER 2022****Subject Code:3160714****Date:16-12-2022****Subject Name:Data Mining****Time:02:30 PM TO 05:00 PM****Total Marks:70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Simple and non-programmable scientific calculators are allowed.

		Marks
Q.1	(a) Compare descriptive and predictive data mining.	03
	(b) Explain the data mining functionalities.	04
	(c) Explain major requirements and challenges in data mining.	07
Q.2	(a) What do you mean by concept hierarchy?	03
	(b) Explain the smoothing techniques.	04
	(c) What is Data Cleaning? Describe various methods of Data Cleaning.	07
OR		
	(c) Explain about the different Data Reduction techniques.	07
Q.3	(a) What are the techniques to improve the efficiency of Apriori algorithm?	03
	(b) What is an Itemset? What is a Frequent Itemset?	04
	(c) For the given data	07

Transaction ID	Items
T1	Hot Dogs, Buns, Ketchup
T2	Hot Dogs, Buns
T3	Hot Dogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	Hot Dogs, Coke, Chips

Find the frequent itemsets and generate association rules on this. Assume that minimum support threshold ($s = 33.33\%$) and minimum confident threshold ($c = 60\%$).

OR

Q.3	(a) Describe the different classifications of Association rule mining.	03
	(b) What is meant by Reduced Minimum Support?	04
	(c) Explain the steps of the “Apriori Algorithm” for mining frequent itemsets with suitable example.	07
Q.4	(a) What are Bayesian Classifiers?	03
	(b) What are the hierarchical methods used in classification?	04
	(c) Describe in detail about Rule based Classification.	07

OR

- Q.4** (a) What is attribute selection measure? **03**
(b) What is the difference between supervised and unsupervised learning scheme. **04**
(c) Describe the issues regarding classification and prediction. Write an algorithm for decision tree. **07**

- Q.5** (a) List the requirements of clustering in data mining. **03**
(b) Differentiate Agglomerative and Divisive Hierarchical Clustering? **04**
(c) Write a short note: Web content mining. **07**

OR

- Q.5** (a) What is meant by hierarchical clustering? **03**
(b) Illustrate strength and weakness of k-mean in comparison with medoid algorithm. k- **04**
(c) Write a short note: Web usage mining. **07**
