Darshan Institute of Engineering & Technology
B.Tech. | Sem-5 | Winter-2024

Course Code	: 2101CS521	**Date**	: 22-10-2024
Course Name	: Data Mining	**Duration**	: 150 Minutes
		Total Marks	: 70

Instructions:
1. Attempt all the questions.
2. Figures to the right indicates maximum marks.
3. Make suitable assumptions wherever necessary.

Q.1	(A)	Explain five numbers summaries with a boxplot.	4
	(B)	Which Kinds of Applications Are targeted by data mining?	3

OR

Differentiate Symmetric vs Skewed Data.

	(C)	Explain issues of data mining.	7

OR

What is an attribute? Explain different types of attributes with example.

Q.2	(A)	Explain ways to fill missing values in data.	4
	(B)	Explain different types of bin Operations used in Binning method.	3

OR

Explain technique used for data Discretization.

	(C)	Explain min-max, z-score and decimal scaling with example.	7

OR

Explain PCA, Attribute Subset Selection and Sampling techniques of data reduction.

Q.3	(A)	Differentiate Maximal Frequent Itemsets vs Closed Frequent Itemsets.	4
	(B)	Explain Apriori Property.	3

OR

Explain Market Basket Analysis with an example.

	(C)	Consider a transactional database.	7

TID	Items
T1	Laptop, Mouse, Keyboard
T2	Laptop, Mouse
T3	Laptop, Phone, Headphones
T4	Headphones, Phone

| T5 | Headphones, Charger |
| T6 | Laptop, Phone, Headphones |

Suppose the minimum support count is 2 and minimum confidence threshold is 90%. Find all frequent itemsets using Apriori Algorithm and generate association rules.

OR

Consider a transactional database.

TID	Items
100	Bread, Milk, Eggs, Butter, Jam, Juice
200	Cereal, Milk, Eggs, Butter, Jam, Juice
300	Bread, Apples, Eggs, Butter
400	Bread, Cheese, Butter, Jam
500	Cheese, Milk, Milk, Butter, Ice Cream, Jam

Suppose the minimum support count is 3. Find all frequent itemsets using FP-growth algorithm.

Q.4	**(A)**	Explain confusion matrix with an example.	**4**
	(B)	Explain F1 and Fβ Measure with example.	**3**

OR

Explain Accuracy and Precision of classification model.

	(C)	Explain Naive Bayesian classification with an example.	**7**

OR

What is gini index? Write down steps of decision tree algorithm.

Q.5	**(A)**	Explain supervised vs Unsupervised Learning.	**4**
	(B)	Explain requirements for Cluster Analysis algorithm.	**3**

OR

What is outlier? Explain different types of outliers with example.

	(C)	Explain K-Medoids with algorithmic steps and example.	**7**

OR

Explain DBSCAN with algorithmic steps and example.

***