Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations –December 2025

Date: 10- 12- 2025 **Time:** 1.00pm to 04.00pm

School: SOCSE	**Program:** B.TECH-Computer science and engineering (CBD)	
Course Code : CBD2506	**Course Name :** Data Mining and Predictive analysis	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**
Marks	**24**	**24**	**26**	**26**

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks. 10Q x2M=20M

1	What are the steps in the data mining process?	**2 Marks**	**L1**	**CO1**
2	What is Data Characterization?	**2 Marks**	**L1**	**CO1**
3	Define association rules?	**2 Marks**	**L2**	**CO2**
4	What is linear regression?	**2 Marks**	**L2**	**CO2**
5	List out the classification of data mining system?	**2 Marks**	**L2**	**CO3**
6	What are Bayesian Classifiers?	**2 Marks**	**L2**	**CO3**
7	Where are Decision Trees mainly used?	**2 Marks**	**L1**	**CO4**
8	What is Apriori algorithm?	**2 Marks**	**L2**	**CO4**
9	Explain the various Time series analysis?	**2 Marks**	**L2**	**CO4**
10	State predictive analysis ?	**2 Marks**	**L1**	**CO4**

Part B

Answer the Questions : **Total: 80 Marks.**

11.	**a.**	Describe data mining and explain how it is related to Knowledge Discovery in Databases (KDD).	**10 Marks**	**L2**	**CO1**
	b.	Define normalization , derive in detail . Suppose that a group of 1500 people was surveyed. The gender of each person was noted. Each person was polled as to whether his or her preferred type of reading material was fiction or nonfiction. Thus, we have two attributes ,gender and preferred reading. Find the Corelation analysis on nominal and numerical data.	**10 Marks**	**L2**	**CO1**

	male	*female*	*Total*
fiction	250	200	450
non_fiction	50	1000	1050
Total	300	1200	1500

Or

12.	**a.**	Explain the process of data cleaning and its role in improving data quality. Solve Binning using the given data using bin size of 4. 4, 3 , __ , 5 , 6, __ , 21 , 24 , 26 , __.	**10 Marks**	**L2**	**CO1**
	b.	What is the need of Dimensionality Reduction? Explain any two techniques for dimensionality reduction with examples.	**10 Marks**	**L2**	**CO2**

13.	**a.**	Explain Logistic regression in detail.	**10 Marks**	**L2**	**CO2**
	b.	John is a student in computer science that loves listening to music. Sometimes, he has homework to do, it can be programming homework, or else. We have some examples of the type of music he listens to, according to some features. X={ Time = Morning , Homework =Yes , Prog = NO }	**10 Marks**	**L2**	**CO2**

Time of Day	**Homework Due**	**Programming**	**Music Type**
Morning	**Yes**	**No**	**Classical**
Morning	**No**	**No**	**Pop**
Morning	**No**	**Yes**	**Classical**
Morning	**Yes**	**No**	**Classical**
Afternoon	**Yes**	**Yes**	**Pop**
Afternoon	**No**	**No**	**Pop**
Evening	**No**	**Yes**	**Pop**
Evening	**Yes**	**Yes**	**Classical**

Or

14.	**a.**	Given a samples of S, where S={(5, Y), (2, N), (6, Y), (7, N), (9, N), (1, N), (10, Y), (15, Y)}. If S has to be partitioned into two intervals S1 and S2 using two possible split points 5 and 9. Compute the best split point.	**20 Marks**	**L2**	**CO3**

15.	**a.**	A database has five transactions. Let min sup = 3 and min confidence= 80%. TID | ITEMS T1 | {M,O,N,K,E,Y} T2 | {D,O,N,K,E,Y} T3 | {M,A,K,E} T4 | {M,U,C,K,Y} T5 | {C,O,O,K,I,E} Construct FP-growth tree and find Conditional Pattern Base and Conditional FPTree.	**20 Marks**	**L3**	**CO3**

Or

16.	**a.**	Apply Apriori Algorithm to generate strong association rules for the given data set. Assume Min Support Count=2 & Min Confidence=70% **Transaction ID** | **Items** T101 | I1,I2, I5 T102 | I2, I4 T103 | I2, I3 T104 | I1,I2,I4 T105 | I1, I3 T106 | I2, I3 T107 | I1,I3 T108 | I1, I2, I3, I5 T109 | I1, I2, I3	**20 Marks**	**L3**	**CO4**

17.	**a.**	Derive types of times series analysis in detail with suitable example.	**20 Marks**	**L3**	**CO4**

Or

18.	**a.**	Perform Hierarchical Agglomerative clustering by considering Complete linkage and average linkage for the given distance matrix. p1 p2 p3 p4 p5 p6 p1 0	**20 Marks**	**L3**	**CO5**

		p2 1 0 p3 2 3 0 p4 1 4 6 0 p5 11 8 2 3 0 p6 5 9 12 2 6 0 Also draw the dendrogram which shows all the steps.			

******* BEST WISHES *******