Reg. No.																

Question Paper Code	12742

B.E. / B.Tech. - DEGREE EXAMINATIONS, APRIL / MAY 2024

Sixth Semester

Computer Science and Business Systems

20CBEL601 - DATA MINING AND ANALYTICS WITH LABORATORY

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (10 × 2 = 20 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	What is data mining?	2	K1	CO1
2.	Show the relationship of data mining to other areas/discipline.	2	K2	CO1
3.	Define itemset.	2	K1	CO1
4.	Why data preprocessing is important?	2	K2	CO2
5.	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 using equal-frequency (equal-depth) partitioning.	2	K2	CO2
6.	What is decision tree? Give example.	2	K2	CO2
7.	What are the benefits of trend analysis?	2	K1	CO5
8.	Differentiate between parametric regression and nonparametric regression method.	2	K2	CO5
9.	In ARIMA(p,d,q) model , what the parameters p,d,and q refers?	2	K1	CO6
10.	List the properties of auto-covariance.	2	K1	CO6

PART - B (5 × 13 = 65 Marks)

Answer ALL Questions

			Marks	K – Level	CO
11.	a) i)	Analyze different visualization techniques that can be used in data mining.	7	K4	CO1
	ii)	Describe the major issues in Data mining.	6	K4	CO1
		OR			
	b)	Explain the core data mining tasks. Discuss whether or not each of the following activities is a data mining task. Dividing the customers of a company according to their profitability. Computing the total sales of a company. Sorting a student database based on student identification numbers. Predicting the outcomes of tossing a (fair) pair of dice. Predicting the future stock price of a company using historical records. Monitoring the heart rate of a patient for abnormalities. Monitoring seismic waves for earthquake activities. Extracting the frequencies of a sound wave	13	K4	CO1

12. a) i) Why data might need to be cleaned in the data mining process. Describe the data cleaning techniques with example. *7 K2 CO2*

ii) Provide list of strategies associated with data reduction. Explain one of them with an example. *6 K2 CO2*

OR

b) Discuss about Data Mining Task Primitives with examples. *13 K2 CO2*

13. a) Write the k-nearest neighbor classification algorithm. Give an example. *13 K3 CO3*

OR

b) Consider the Data set D. Given the minimum support 2, apply Apriori algorithm on this dataset. *13 K3 CO3*

Transaction ID	Items
100	A,C,D
200	B,C,E
300	A,B,C,E
400	B,E

14. a) i) What is logistic regression? Provide some of the use cases of logistic regression. *7 K2 CO4*

ii) Explain the mathematics behind Wald test. *6 K2 CO4*

OR

b) Illustrate the data mining process followed in the linear models with suitable example *13 K2 CO4*

15. a) i) Outline the following non-linear least square method (NLS) Newton-Raphson method *7 K2 CO5*

ii) Explain Test for trend and seasonality. *6 K2 CO6*

OR

b) i) Outline the following non-linear least square method (NLS) Levenberg-Marquardt's method *7 K2 CO5*

ii) Explain time series analysis in detail. *6 K2 CO6*

PART - C (1× 15 = 15 Marks)

16. a) i) Elaborate on a typical architecture you would use to develop a data mining system for an organization of your choice. *8 K6 CO1*

ii) A data set for analysis includes only one attribute X: X={ 7,12,5,8,5,9,13,12,19,7,12,12,13,3,4,5,13,8,7,6} *7 K2 CO2*

(a) What is the mean of the data set X?

(b) What is the median?

(c) Find the standard deviation for X.

OR

b) i) Show how what kind of data can be inferred in data mining. *7 K2 CO1*

ii) Give Brief description of following: (a) Binning (b) regression (c) Smoothing (d) Generalization *8 K2 CO2*