Reg. No															

B.Tech DEGREE EXAMINATION, NOVEMBER 2023

Fifth to Seventh Semester

18CSE355T - DATA MINING AND ANALYTICS

(For the candidates admitted during the academic year 2020 - 2021 & 2021 - 2022)

Note:

i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours **Max. Marks: 100**

PART - A (20 × 1 = 20 Marks) — Answer **all** Questions

		Marks	BL	CO
1.	Which is the following is an essential process in which the intelligent methods are applied to extract data pattern? (A) Data Mining (B) Data transformation (C) Data Selection (D) Data Warehousing	1	1	1
2.	A Graph that uses vertical bars to represent data is called (A) Bar Graph (B) Line Graph (C) Scatter Plot (D) Pie Chart	1	1	1
3.	What is KDD in data mining? (A) Knowledge discovery database (B) Knowledge discovery data (C) Knowledge data definition (D) Knowledge data detection	1	1	1
4.	In some cases, telecommunication companies desire to segment their clients into distinct groups in order to send suitable and related subscription offer. This can be considered as an example of which of the following methods. (A) Supervised learning (B) Unsupervised learning (C) Segregation (D) Data extraction	1	4	1
5.	Which one of the following is the main reason for the pruning a decision tree (A) To save computing during testing (B) To save space for storing the decision tree (C) To make the training set error smaller (D) To avoid over fitting the training set	1	2	3
6.	The initial steps concerned in the process of knowledge discovery is (A) Data Selection (B) Data Integration (C) Data Cleaning (D) Data Transformation	1	1	1
7.	Suppose your classification model predicted true for a class which actual value was false, then this is a (A) False Positive (B) False Negative (C) True Positive (D) True Negative	1	2	3
8.	_____ models Continuous Valued Functions. (A) Prediction (B) Back Propagation (C) Classification (D) Data Trends	1	2	3
9.	Which algorithm requires fewer scans of data in association rule mining? (A) FP Growth (B) Naive Bayes (C) Apriori (D) Decision Tree	1	2	2
10.	What technique can be used to improve the efficiency of Apriori algorithm? (A) Transition Technique (B) Hash based Technique. (C) Sampling (D) Cleaning	1	2	2

11.	Frequency of occurrence of an itemset is called as (A) Support (B) Support Count (C) Confidence (D) Rules	1	1	2
12.	Classification rules are extracted from _____ (A) Root Node (B) Decision Tree (C) Leaf Nodes (D) Branches	1	2	3
13.	K-means is a _____ algorithm (A) Supervised Learning (B) Unsupervised Learning (C) Supervised Learning and Unsupervised Learning (D) Semi Supervised Learning	1	2	4
14.	Which of the following clustering requires merging approach (A) Partitioning (B) Hierarchical (C) Naive Bayes (D) Semi-Partitional	1	2	4
15.	Which of the following is finally produced by hierarchical clustering? (A) Final estimate of cluster centroids (B) Tree showing how close things are to each other. (C) Assigned of each point to cluster (D) No cluster	1	4	4
16.	Which of the following is not a type of decision tree node? (A) Root Node (B) Leaf node (C) Decision node (D) Branch Node	1	1	4
17.	Under which category does sensor data belongs to (A) Time Series (B) Data Stream (C) Facial (D) Networked	1	4	5
18.	The data objects that do not comply with the general model or behavior of available data are (A) Evolution analysis (B) Outlier Analysis (C) Classification (D) Prediction	1	2	5
19.	Which of the following is the direct application of frequent itemset mining? (A) Social Network Analysis (B) Market Basket Analysis (C) Outlier Detection (D) Intrusion Detection	1	2	5
20.	To detect fraudulent usage of credit cards, the following data mining task should be used (A) Feature Selection (B) Prediction (C) Outlier Analysis (D) Data Transformation	1	4	5

PART - B (5 × 4 = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	Briefly discuss about the Quantile plot and Scatter plot with example.	4	1	1
22.	Differentiate between KDD and data mining in terms of objective and workflow.	4	4	1
23.	Describe the various metrics used for evaluating the performance of a classifier.	4	2	3
24.	What is Frequent set Mining? Give an example	4	2	2
25.	Illustrate Agglomerative algorithm with example.	4	3	4
26.	Explain outlier analysis with example.	4	3	5
27.	Write a note on Intrusion Detection and Prevention.	4	4	5

PART - C (5 × 12 = 60 Marks)

Answer **all** Questions

Marks	BL	CO

28. (a) Explain the steps in the process of knowledge discovery in database? 12 3 1

(OR)

(b) Analyse the sequence of operations on calculating the following for the given dataset x = 2, 7, 3,3,12, 9.
Method:
1. Mean
2. Median
3.mode
4.Standard Deviation

29. (a)

TID	List of Items
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1.I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

12 6 2

Find the frequent itemset with Apriori Algorithm in the given transaction table and explain the steps.

(OR)

(b) Construct a Tree and illustrate FP Growth Algorithm.

30. (a) Explain about Naive Bayes Classification algorithm with example 12 2 3

(OR)

(b) Write in detail about Random Forest and write how they fit with ensembling , bagging and bootstrapping concepts.

31. (a) Illustrate DBSCAN algorithm with suitable example. 12 5 4

(OR)

(b) Illustrate hierarchical Clustering with suitable example.

32. (a) Explain Global, Contextual and Collective outliers with example. 12 4 5

(OR)

(b) How Data mining is applied in Financial Data Analysis?

* * * * *