b.	Illustrate different types of visualization techniques with neat diagram.	10	4	2	1
28. a.	Elaborate how classification is different from clustering.	10	3	3	3
	(OR)				
b.	Discuss different data mining algorithms in detail.	10	3	3	3
29. a.	What are the type of hypothesis test? Explain.	10	4	4	1
	(OR)				
b.	Explain non parametric regression method in detail.	10	4	4	1
30. a.	Define auto-correlation and explain its properties.	10	3	5	1
	(OR)				
b.	Explain decision and risk analysis in detail.	10	4	5	1

* * * * *

Reg. No.														

B.Tech. DEGREE EXAMINATION, MAY 2022
Sixth Semester

18CSE366J – DATA MINING AND ANALYTICS
(For the candidates admitted from the academic year 2018-2019 to 2019-2020)

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** should be answered in answer booklet.

Time: 2½ Hours Max. Marks: 75

PART – A (25 × 1 = 25 Marks)
Answer **ALL** Questions

		Marks	BL	CO	PO
1.	Which of the following is an essential process in which the intelligent methods are applied to exact data patterns? (A) Warehousing (B) Data mining (C) Text mining (D) Data selection	1	1	1	1
2.	_____ is a repository of multiple heterogeneous data sources organized under a unified schema at a single site to facilitate management decision making. (A) Data warehouse (B) Data mining (C) DBMS (D) KDD	1	1	1	1
3.	_____ is a dataset which contain samples that to not share with the common characteristics or model of the dataset. (A) Outliers (B) Data discrimination (C) Data characterization (D) Bounding box	1	2	1	1
4.	Which of the following is not an operation of OLAP? (A) Drill up (B) Roll up (C) Flip up (D) Pivot	1	1	1	1
5.	_____ is not a data mining function. (A) Classification (B) Selection and interpretation (C) Characterization and discrimination (D) Clustering				
6.	_____ studies the collection, analysis, interpretation or explanation and presentation of data. (A) Statistics (B) Visualization (C) Data mining (D) Clustering	1	1	2	1
7.	When performing PCA we want to (A) Find orthogonal vectors (B) Estimate the number of dimensions (C) Find the most meaningful basis (D) Find the components of the dataset	1	1	2	3

8. The initial steps concerned in the process of knowledge discovery is — 1 1 2 4
   (A) Data selection (B) Data integration
   (C) Data cleaning (D) Data transformation

9. It is the main technique employed for data selection — 1 1 2 4
   (A) Noise (B) Clustering
   (C) Histogram (D) Sampling

10. _____ partitions the objects into different groups. — 1 1 2 1
   (A) Mapping (B) Clustering
   (C) Classification (D) Prediction

11. Which of the following properties is false in the case of Bayesian network? — 1 1 3 1
   (A) The edges are directed
   (B) Contains cycle
   (C) Represents conditional relations among random variables
   (D) Contains edges

12. A collection of one or more items is called as — 1 2 3 1
   (A) Itemset (B) Support
   (C) Confidence (D) Support count

13. An itemset whose support is greater than or equal to minimum support threshold is _____. — 1 2 3 1
   (A) Itemset (B) Frequent itemset
   (C) Intrequent items (D) Threshold values

14. What does FP growth algorithm do? — 1 1 3 1
   (A) It mines all frequent patterns through pruning rules with lesser support (B) It mines all frequent patterns through pruning rules with higher support
   (C) It mines all frequent patterns by constructing a FP tree (D) It mines all frequent patterns by constructing an itemset

15. Which of the following is the direct application of frequent itemset mining? — 1 1 3 1
   (A) Social network analysis (B) Market basket analysis
   (C) Outliner detection (D) Intrusion detection

16. In the regression equation, $y = 24 - 3x$ the slope is — 1 1 4 1
   (A) 24 (B) –24
   (C) 3 (D) –3

17. In binary logistic regression — 1 2 4 1
   (A) The dependent variable is continuous (B) The dependent variable consist of two categories
   (C) There is no dependent variable (D) The dependent variable is divided into two equal sub categories

18. _____ is a first order iterative optimization algorithm for finding a local minimum of a differential function. — 1 2 4 1
   (A) Steepest descent (B) Stochastic descent
   (C) Mini descent (D) Batch descent

19. _____ is a forecasting method where historical evidence is unavailable — 1 2 4 1
   (A) Quantitative method (B) Average method
   (C) Qualitative method (D) Naive method

20. _____ is a statistical measure used to know how data is divided across a range. — 1 1 4 1
   (A) Central tendency (B) Measures of variability
   (C) Measures of frequence (D) Measures of dispersion

21. The _____ function is defined as the sequence of covariances of a stationary process. — 1 1 5 1
   (A) Auto correlation (B) Auto covariance
   (C) Partial auto correlation (D) Partial auto covariance

22. _____ smoothing technique is a widely known smoothing model for forecasting data that has a trend. — 1 1 5 1
   (A) Moving average (B) Holt
   (C) Winter (D) Exponential

23. ARIMA stands for _____. — 1 1 5 1
   (A) Autoregressive integrated moving average (B) Autoregressive integrated mean average
   (C) Auto reactive integrated moving average (D) Auto reactive integrated mean average

24. Decision nodes are represented as _____ in decision tree. — 1 1 5 1
   (A) Disks (B) Squares
   (C) Circles (D) Triangles

25. _____ optimization is said to be conflicting objectives. — 1 1 5 3
   (A) Stochastic (B) Multi-objective
   (C) Non-convex (D) Spark

PART – B (5 × 10 = 50 Marks) — Marks BL CO PO
Answer **ALL** Questions

26. a. State machine learning. Discuss various learning algorithms. — 10 3 1 1

(OR)

   b. Elaborate the applications of data mining in different fields. — 10 3 1 1

27. a. Discus the various process used in data reduction. — 10 3 2 1

(OR)