31. a.	Describe logistic regression in detail.	12	1	4	2
	(OR)				
b.	Interpret regression model in detail.	12	1	4	2
32. a.	Explain auto covariance in detail.	12	2	5	3
	(OR)				
b.	Describe exploratory time series analysis and how smoothing is handled in an analysis.	12	2	5	2

* * * * *

Reg. No.															

B.Tech. DEGREE EXAMINATION, JUNE 2023

Sixth Semester

18CSE366J – DATA MINING AND ANALYTICS

(For the candidates admitted during the academic year 2018-2019 to 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 & 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	PO
1.	Which of the following refers to the problem of finding abstracted structures in the unlabeled data? (A) Supervised learning (B) Unsupervised learning (C) Hybrid learning (D) Reinforcement learning	1	1	1	1
2.	_____ is not an operation of OLAP. (A) Drill up (B) Roll up (C) Flip up (D) Pivot	1	1	1	1
3.	_____ is not a data mining function. (A) Classification (B) Clustering (C) Selection and interpretation (D) Characterization and discrimination	1	1	1	1
4.	_____ used to measure the linear relationship between the objects. (A) Covariance (B) Mean value (C) Standard deviation (D) Correlation	1	1	1	1
5.	_____ used to minimize the misclassification error in decision tree. (A) Boosting (B) Over fitting (C) Pruning (D) Bagging	1	1	1	1
6.	_____ is not an other name for data mining. (A) KDD (B) Data archeology (C) Data analysis (D) Data warehouse	1	1	2	1
7.	Visualization technique used for representation of categorical data is _____. (A) Pixel oriented (B) Tag cloud (C) Parallel coordinates (D) Tree map	1	1	2	1
8.	OLAP is sued to _____ the data. (A) Analyze (B) Manipulate (C) Reduce (D) Transform	1	1	2	1

9. _____ is an application of reinforcement learning. 1 1 2 2
(A) Topic modeling (B) Recommendation system
(C) Pattern recognition (D) Image classification

10. _____ studies the collection, analysis interpretation and presentation of data. 1 1 2 2
(A) Statistics (B) Visualization
(C) Data mining (D) Clustering

11. _____ is the main technique employed for data selection. 1 1 3 2
(A) Noise (B) Clustering
(C) Histogram (D) Sampling

12. In _____ the attribute data are scaled so as to fall within a smaller range, such as –1.0 to 1.0. 1 1 3 2
(A) Aggregation (B) Binning
(C) Clustering (D) Normalization

13. Normalization by _____ normalizes by moving the decimal point of value of attribute. 1 1 3 2
(A) Z score (B) Z index
(C) Decimal scaling (D) Min max normalization

14. What do you mean by support (A)? 1 1 3 2
(A) Total number of transactions containing A (B) Total number of transactions not containing A
(C) Number of transaction containing A/ total number of transaction (D) Number of transaction not containing A / total number of transaction

15. _____ is a first order iterative optimization algorithm for finding a local minimum of a differential function. 1 1 3 1
(A) Steepest descent (B) Stochastic descent
(C) Minim descent (D) Batch descent

16. Auto correlation refers to _____. 1 1 4 1
(A) Correlation between two different variables (B) Correlation between the same variables over time
(C) Correlation between two different samples (D) Correlation between two different populations

17. _____ statistical tests is used to test for auto correlation. 1 1 4 2
(A) Pearson correlation test (B) T-test
(C) Watson test (D) Anova

18. Stochastic modeling refers to _____. 1 1 5 2
(A) Modeling deterministic system (B) Modeling non-deterministic system
(C) Modeling linear systems (D) Modeling non linear systems

19. _____ is a measures of uncertainty in stochastic modeling 1 1 5 2
(A) Mean (B) Variance
(C) Standard deviation (D) Mode

20. _____ involves a graphical representation of a sequence of decision and their possible consequences. 1 1 5 2
(A) Clustering (B) Decision tree
(C) KNN (D) K means

PART – B (5 × 4 = 20 Marks) Marks BL CO PO
Answer **ANY FIVE** Questions

21. Annotate reinforcement with an example. 4 1 1 2

22. What are the different languages in DBMS? 4 1 1 1

23. Elaborate any two types in data reduction. 4 2 2 2

24. Illustrate different visualization techniques with diagram. 4 2 3 2

25. Define frequent patterns. How rules are generated by item sets? 4 1 4 2

26. Write note on LR test and score test. 4 2 4 2

27. Describe different smoothing techniques. 4 3 5 2

PART – C (5 × 12 = 60 Marks) Marks BL CO PO
Answer **ALL** Questions

28. a. Elaborate the basic statistical descriptions of data. 12 2 1 1

(OR)

b. Explain OLAP and its types in detail. 12 2 1 2

29. a. State machine learning. Discuss various supervised machine learning algorithm. 12 1 2 3

(OR)

b. Explain statistical Bayesian classification with real time data set. 12 1 2 3

30. a. Illustrate with a real time dataset for construction of decision tree and generate rule to classify the dataset. 12 2 3 3

(OR)

b. What are the limitations of KNN? Why it is called lazy learner? Explain with an algorithm. 12 2 3 1