

9. _____ is a top down splitting technique based on a specified number or bins. 1 1 2 1
 (A) Normalization (B) Binning
 (C) Clustering (D) Classification
10. _____ is the process of finding a model that describes and distinguishes data classes. 1 1 2 1
 (A) Association rule mining (B) Data discretization
 (C) Clustering (D) Classification
11. If a substructure occurs frequently it is called _____. 1 1 3 1
 (A) Sequential pattern (B) Semi structured pattern
 (C) Frequent pattern (D) Structured pattern
12. The apriori algorithm is used for _____ data mining task. 1 1 3 1
 (A) Association (B) Clustering
 (C) Classification (D) Regression
13. _____ is not a supervised learning. 1 1 3 1
 (A) PCA (B) Naive Bayesian
 (C) Linear regression (D) Decision tree
14. Decision nodes are represented as _____. 1 1 3 1
 (A) Disks (B) Square
 (C) Circles (D) Triangles
15. Which of the following is a testable hypothesis? 1 1 3 1
 (A) Dogs are better pets than cats (B) Eating vegetables every day improves health
 (C) The moon is made of cheese (D) People prefer blue cars over red cars
16. _____ smoothing technique is a widely known smoothing model for forecasting data that has a trend. 1 1 4 1
 (A) Moving average (B) Holt
 (C) Winter (D) Exponential
17. _____ models use past values of the time series to predict future values. 1 1 4 1
 (A) Auto regressive (B) Moving average
 (C) Auto repetitive (D) Partial auto correlation function
18. _____ is association between a time series and its lagged values. 1 1 5 2
 (A) Auto regression (B) Auto correlation
 (C) Auto regularization (D) Regularization
19. In the regression equation, $Y = 24 - 3x$ the slope is _____. 1 1 5 2
 (A) 24 (B) -24
 (C) 3 (D) -3
20. _____ is used to perform inference on the current data to make predictions. 1 1 5 1
 (A) Data mining (B) Data pattern
 (C) Predictive (D) Descriptive

PART – B (5 × 4 = 20 Marks)

Answer ANY FIVE Questions

	Marks	BL	CO	PO
21. What are the major issues in data mining?	4	2	1	2
22. Differentiate supervised learning and unsupervised learning.	4	2	2	2
23. How data is transformed in data pre processing method?	4	2	2	1
24. Write steps involved in KNN algorithm.	4	1	3	1
25. How logistic regression is different form linear regression?	4	1	5	2
26. Annotate the different tests of hypotheses.	4	2	4	2
27. Write note on moving average.	4	2	5	1

PART – C (5 × 12 = 60 Marks)

Answer ALL Questions

	Marks	BL	CO	PO
28. a. Define machine learning. Discuss different types of algorithm in detail.	12	1	1	2
(OR)				
b. Explain various stages of KDD process in detail with neat diagram.	12	2	1	3
29. a. How can frequent patterns are classified? Explain.	12	2	3	1
(OR)				
b. Explain statistical Bayesian classification with real time data set.	12	3	3	2
30. a. Describe data mining knowledge representation with neat diagram.	12	3	2	1
(OR)				
b. Explain about need for data preprocessing and data quality in data mining operations.	12	4	2	1
31. a. Elucidate different test used to validate logistic regression.	12	2	4	1
(OR)				
b. Describe semi parametric and non parametric regression models.	12	2	4	1
32. a. Enumerate auto covariance and auto correlation in detail.	12	2	5	1
(OR)				
b. Discuss ARMA and ARIMA in detail.	12	2	5	1

* * * * *

