



Register Number:

DATE:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27  
BCA( DATA ANALYTICS) – III SEMESTER  
SEMESTER EXAMINATION: OCTOBER 2021  
(Examination to be conducted January – March 2022)  
**BCADA 3421: DATA WAREHOUSING AND MINING**

**TIME 2.5 HOURS**

**MAXIMUM MARKS 70**

**This Question Paper Contains FOUR Printed Papers and THREE Parts**

**PART A**

**ANSWER ALL QUESTIONS (MCQs)**

**20 X1 = 20**

1. Expand OLAP
  - a) Online Analytical Pre-Processing
  - b) Online Analytical Process
  - c) Online Analytical Programming
  - d) Online Analytical Processing
  
2. Datawarehouse is
  - i)Subject oriented    ii)volatile    iii)time variant    iv)non integrated
  
  - a) All of the above
  - b) None of the Above
  - c) i,iii
  - d) i,ii,iii
  
3. The start schema follows \_\_\_\_\_ relationship from fact to dimension table
  - a) One to many
  - b) Many to many
  - c) One to one
  - d) None of the above
  
4. \_\_\_\_\_ selects a sub-cube from the OLAP cube by selecting two or more dimensions
  - a) Slice
  - b) Dice
  - c) Roll up
  - d) Drill down
  
5. Dimensional table and a fact table can be connected with the following database keys:
  - a) Foreign key
  - b) Surrogate key
  - c) Candidate key
  - d) All of the above

6. Select the appropriate steps of data analysis
  - a) Feature Selection, Data Mining, Pattern Interpretation
  - b) Pattern Interpretation, Data Mining, Feature Selection
  - c) Data Mining, Data Visualization, Feature Extraction
  
7. Identify the dependent variable (DV) and independent variable(IV).  
Running will increases the heart rate of a person.
  - a) Running (IV),Hear Rate(DV)
  - b) Running (DV),Hear Rate(IV)
  - c) Running (IV),Hear Rate(IV)
  - d) No dependent variable
  
8. In an experiment the researcher manipulates the \_\_\_\_\_ variable and measure the \_\_\_\_\_ variable.
  - a) Independent, Dependent
  - b) Dependent, Independent
  - c) Casual, Dependent
  - d) Dependent, Casual
  
9. The creation of a new set of features from the original raw data is known
  - a) Binarization
  - b) Discretization
  - c) Dimension reduction
  - d) Feature extraction.
  
10. \_\_\_\_\_ is the method of transforming a continuous attribute into a categorical attribute
  - a) Feature extraction
  - b) Feature selection
  - c) Data Conversion
  - d) Discretization
  
11. \_\_\_\_\_ will have known class labels and in \_\_\_\_\_ will have unknown class labels .
  - a) Training data, test data
  - b) Test data, Training data
  - c) Error data, Accurate data
  - d) None of the above.
  
12. \_\_\_\_\_ will have zero incoming edges and zero or more outgoing edges.
  - a) Leaf note
  - b) Root node
  - c) Terminal node
  - d) Internal node
  
13. In decision tree the border between two neighbouring regions of different class label is known as
  - a) Distance boundary
  - b) Decision boundary
  - c) Class label boundary
  - d) None of the above

14. \_\_\_\_\_, which corresponds to the number of negative examples wrongly predicted as positive by the classification model
- False Positive
  - False Negative
  - True Positive
  - True Negative
15. Identify the formula for recall
- $FP/(FP+FN)$
  - $TP/(TP+FN)$
  - $TP/(TP+TP)$
  - $(TP+TN)/(FP+FN)$
16. What is association rule mining?
- Same as frequent itemset mining
  - Finding of strong association rules using frequent itemsets
  - Using association to analyse correlation rules
  - Finding Itemsets for future trends
17. How frequently the item appears in data is known as \_\_\_\_\_
- Support
  - Confidence
  - Lift
  - None
18. An association rule is an implication expression  $X \rightarrow Y$ . Where  $X \cap Y$  is
- Zero
  - All the values from X and Y
  - Only the value from X
  - 1
19. The number of iterations in apriori \_\_\_\_\_
- increases with the size of the data
  - decreases with the increase in size of the data
  - increases with the size of the maximum frequent set
  - decreases with increase in size of the maximum frequent set
20. To determine association rules from frequent item sets
- Only minimum confidence needed
  - Neither support nor confidence needed
  - Both minimum support and confidence are needed
  - Minimum support is needed

## PART B

**ANSWER ANY SIX QUESTIONS**

**6 X 5 = 30**

21. What do you mean by star schema? Draw a star schema for sales management.  
Fact table--sales, dimension tables –Product, Order, Customer, Employee.  
Measures are total\_sale, Quantity.
22. Explain the different data Preprocessing techniques.
23. Compare Descriptive analysis with Predictive Analysis

24. Construct a decision tree using hunt's Algorithm.
25. With a diagram explain the architecture of Data Warehouse.
26. Define attribute. Explain the different types of attribute with example.
27. Assume that the outcome of some classification results in 6 TPs, 4 FNs, 8 TNs, and 2 FPs. Calculate
  - a) Draw confusion matrix
  - b) Error rate
  - c) Accuracy
  - d) Precision
28. Mention the impact of Model Overfitting on Classifier.

### PART C

#### ANSWER ANY TWO QUESTIONS

**2 X10 = 20**

29. Distinguish between OLAP and OLTP. Explain the different types OLAP operation on data cube of vehicle details. Three Dimensions are vehicle type, state and year\_of\_purchasing
30. With neat diagram explain the various stages of KDD. Mention the advantages of showing the final result of analysis using visualization tools.
31. Consider the given example

Transaction ID	items
T1	A,B,C,D
T2	A,B,D
T3	A,B,D,E
T4	A,B,C
T5	A,E,C
T6	C,D
T7	A,D

- a) List the steps of apriori algorithm
- b) Identify the frequent item set with support 4 and confidence 75%
- c) Discover all the association rules with confidence above 75%