

**BACHELOR OF COMPUTER APPLICATION
SIXTH SEMESTER
DATA MINING
BCA-602**

**SET
B**

[USE OMR FOR OBJECTIVE PART]

Duration: 1.30 hrs.

Full Marks: 35

Time: 15 min.

(Objective)

Marks: 10

Choose the correct answer from the following:

1X10=10

1. A decision tree is a flowchart like structure and has nodes and branches. What does branches and nodes represent, respectively?
 - a. Attribute value and outcome of the test
 - b. Outcome of the test and attribute value
 - c. Classes and Class Description
 - d. Class Description and Classes
2. Given below are the descriptive data mining functionalities. Pick the functionality which does not come under the predictive task of the data mining.
 - a. Associations
 - b. Classification
 - c. Cluster Analysis
 - d. Outlier Analysis
3. What are the objective interesting measures for Classification (If Then) rules?
 - a. Support and Confidence
 - b. Coverage and Confidence
 - c. Support and Accuracy
 - d. Accuracy and Coverage
4. There are many ways in which various types of data can be stored. Ex: 0,1 as binary data and names as nominal data. Choose the attribute which is not relevant to store data.
 - a. Binary Attributes
 - b. Nominal Attributes
 - c. Executable Attributes
 - d. Ordinal Attributes
5. Sometimes the data under analysis has too many attributes, some of which may not be useful as per the task being performed. This problem is referred to as
 - a. Overfitting
 - b. Dimensionality curse
 - c. Outliers in data
 - d. Missing data
6. is a technique that predict data values using results derived from different data.
 - a. Regression
 - b. Summarization
 - c. Sequence discovery
 - d. Association rules
7. Which of the following is true about bottom up discretization?
 - a. Some the values are treated as potential split points
 - b. All the values are treated as potential split points

- c. Split points are not considered d. Only one value are treated as potential split points
8. is not a proximity measure.
- a. Dissimilarity measures b. Similarity measures
c. Probability measures d. Distance measures
9. Which of the following statement about knowledge and data discovery management system (KDDMS) is false?
- a. It will provide concurrency features b. It will provide recovery features
c. It will include data mining tools and data management tools d. It will include data mining tools but not data management tools
10. Which of the following is not a basic data mining task?
- a. Spooling b. Prediction
c. Classification d. Clustering

(Descriptive)

Time : 1 hrs. 15 min.

Marks : 25

[Answer question no.1 & any two (2) from the rest]

1. Give a brief comparison among clustering, nearest neighbor classification and anomaly detection. 5
2. Consider the following transactions: 10

Transaction Id	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diapers, Beer, Cola
4	Bread, Milk, Diapers, Beer
5	Bread, Milk, Diapers, Cola

-At a minimum support of 60%, using a vertical layout, find all frequent item sets.

3. Explain DBSCAN algorithm with its two important parameters? Write down the applications of CLIQUE. 10
4. What do you understand by instance based learning? Show that if a candidate k-itemset X has a subset of sized less than k-1 that is infrequent, then at least one of the (k-1) size subset of X is necessarily infrequent. 10
5. Write short notes on (any two) 10
(a) SVM (b) NLP (c) Web mining

== *** ==