



**PRESIDENCY UNIVERSITY  
BENGALURU**

**SCHOOL OF INFORMATION SCIENCE  
MID TERM EXAMINATION - APR 2023**

**Semester :** Semester IV - 2021

**Course Code :** CSA2007

**Course Name :** Sem IV - CSA2007 - Data Mining

**Program :** BCA

**Date :** 13-APR-2023

**Time :** 2PM - 3.30PM

**Max Marks :** 50

**Weightage :** 25%

**Instructions:**

- (i) Read all questions carefully and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and non-programmable calculator are permitted.
- (iv) Do not write any information on the question paper other than Roll Number.

**PART A**

**ANSWER ALL THE QUESTIONS**

**(10 X 1 = 10M)**

1. Some telecommunication company wants to segment their customers into distinct groups in order to send appropriate subscription offers, this is an example of
  - a) Supervised learning
  - b) Data extraction (CO1) [Knowledge]
  - c) Aggregation
  - d) Unsupervised Learning
2. Which statement given below closely defines the term cluster?
  - a) The group of the same objects that differ majorly from the other objects. (CO1) [Knowledge]
  - b) Symbolic representation of facts and ideas from which information can be extracted using the data mining process
  - c) It is simply an operation performed on databases to simplify the information so that it can be further transformed into a machine learning algorithm.
  - d) All of the above
3. In which step of Knowledge Discovery, multiple data sources are combined?
  - a) Data Cleaning
  - b) Data Integration. (CO1) [Knowledge]
  - c) Data Selection
  - d) Data Transformation
4. Which of the following is an essential process in which the intelligent methods are applied to extract data patterns?
  - a) Warehousing
  - b) Data Mining (CO1) [Knowledge]
  - c) Text Mining
  - d) Data Selection
5. Data Mining is a tool for allowing users to....
  - a) Find the hidden relationship in data.
  - b) Find the relationships in data. (CO1) [Knowledge]
  - c) Find the visible relationships in data.
  - d) None of the above
6. Data set {brown, black, blue, green , red} is an example of
  - a) Continuous attribute
  - b) Ordinal attributec) (CO2) [Knowledge]
  - Numeric attribute
  - d) Nominal attribute

7. What is the use of data cleaning?  
 a) To remove the noisy data. (CO2) [Knowledge]  
 b) Correct the inconsistencies in data.  
 c) Transformations to correct the wrong data.  
 d) All of the above
8. Which of the following is true about outliers – (CO2) [Knowledge]  
 a) Data points that deviate a lot from normal observations  
 b) Can reduce the accuracy of the model  
 c) Both A and B  
 d) None
9. Dimensionality reduction reduces the data set size by removing \_\_\_\_\_ (CO2) [Knowledge]  
 a) Composite attributes                                        b) Derived attributes  
 c) Relevant attributes    d) Irrelevant attributes
10. Given two objects represented by tuples (3, 5) and (2, 0). What is the Manhattan distance between two objects? (CO2) [Knowledge]  
 a) 5.1    b) 6  
 c) 5    d) None of the above

**PART B**

**ANSWER ALL THE QUESTIONS**

**(4 X 5 = 20M)**

11. Discuss major issues in data mining. (CO1) [Comprehension]
12. With a neat diagram explain data mining as a step in the process of knowledge discovery. (CO1) [Comprehension]
13. Suppose two stocks A and B have the following values in one week: (2, 5), (3, 8), (5, 10), (4, 11), (6, 14). If the stocks are affected by the same industry trends, will their prices rise or fall together? (CO2) [Comprehension]
14. Compute the similarity between the given binary vectors.  $p = 1000000010$ ,  $q = 0001001010$ . using SMC and JC. (CO2) [Comprehension]

**PART C**

**ANSWER ALL THE QUESTIONS**

**(2 X 10 = 20M)**

15. List the major applications of data mining. Explain any three in detail. (CO1) [Application]
16. Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to find out which political party they prefer. The results of the survey are shown in the table below:

|        | Republican | Democrat | Independent | Total |
|--------|------------|----------|-------------|-------|
| Male   | 100        | 70       | 30          | 200   |
| Female | 140        | 60       | 20          | 220   |
| Total  | 240        | 130      | 50          | 440   |

To see if gender is linked to political party preference, perform a Chi-Square test of independence. [consider level of significance=0.05, dof=2, chi-square static level=5.991]

(CO2) [Application]