Progressive Education Society's
Modern College of Arts, Science & Commerce Ganeshkhind, Pune – 16
(Autonomous)
End Semester Examination: OCT / NOV 2024
Faculty: Science and Technology

Program: BCA Code: BCASc08 **Semester: V** **SET: A**
Program (Specific): BCA (Science) **Course Type:DSE**
Class: T.Y.B.C.A. **Max.Marks: 70**
Name of the Course: Data Mining & Data Science
Course Code: 24-BCA-352 **Time: 3Hr**
Paper: -

Instructions to the candidate:

1) *There are 4 sections in the question paper. Write each section on separate page.*
2) *All Sections are compulsory.*
3) *Figures to the right indicate full marks.*
4) *Draw a well labelled diagram wherever necessary.*

SECTION: A

Q1) Attempt the following.

a) Multiple choice question **[5x1=5]**

I. Which of the following is not a data preprocessing method
   a) Data Visualization b) Data Discretization
   c) Data Cleaning d) Data Reduction

II. ---------- is subject oriented, integrated, time variant, nonvolatile collection of data in support of management decisions.
   a) Data Mining b) Data Structures
   c) Data Warehousing d) Web mining

III. What is Meta Data
   a) Information about data b) Data about data
   c) Attribute information about data d) All of the above

IV. The clustering technique starts with as many clusters as there are records.
   a) Agglomerative b) Divisive
   c) Partition d) None of the above

V. The number of student in a class is an example of ---------
   a) Discrete data b) Ordinal data
   c) Nominal data d) Summary data

b) Answer the following in one or two sentences. **[5x1=5]**

I. Write the equation to define the relationship between dependent and independent variable
II. Define Reinforcement learning
III. Define Noisy Data
IV. State the formula to calculate Euclidian Distance
V. Define Outlier

SECTION: B

Q2) Short answer questions (Attempt any 5) **[5x3=15]**

I. List any three probability distributions
II. Define - a) Information gain
b) Entropy
III. Define Linear Regression with a suitable diagram
IV. Write the difference between OLTP and OLAP
V. What is quantitative data? Explain its types
VI. Explain tree pruning with example
VII. Write steps in data preprocessing

SECTION: C

Q3) Short answer questions (Attempt any 5) **[5x4=20]**

I. Explain the Decision Tree with its advantages and disadvantages.
II. Explain KDD process in detail
III. What is probability distribution? Explain its types
IV. What is data mart? Explain its category
V. Explain k - Nearest neighbor classifiers with example
VI. Differentiate between star schema and snowflakes schema
VII. Write a note on SVM classifier

SECTION: D

Q4) Long answer type questions (Attempt any 5) **[5x5=25]**

I. Solve Using Naïve Bayes Classification method for the tuple
Data sample X = (age <=30, Income = medium, Student = yes Credit_rating = Fair)

AGE	INCOME	STUDENT	CREDIT_RATING	BUYS_COMP
<=30	High	No	Fair	No
<=30	High	No	Excellent	No
31…40	High	No	Fair	Yes
>40	Medium	No	Fair	Yes
>40	Low	Yes	Fair	Yes
>40	Low	Yes	Excellent	No
31…40	Low	Yes	Excellent	Yes
<=30	Medium	No	Fair	No
<=30	Low	Yes	Fair	Yes
>40	Medium	Yes	Fair	Yes
<=30	Medium	Yes	Excellent	Yes
31…40	Medium	No	Excellent	Yes
31…40	High	Yes	Fair	Yes
>40	Medium	No	Excellent	No

II. Consider the single variable cluster {2, 4, 10, 12, 3, 20, 30, 11, 25} Solve using the K-Means clustering method. Assuming number of clusters k=2
III. What is machine learning? explain its type with a tree diagram
IV. What is Data Warehouse? Explain its architecture and modelling with suitable diagram.
V. Explain Data preprocessing in detail.
VI. What is data analytics? Explain each types in data analytics
VII. Which are the components of data science? Explain data science process