PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

MCA DEGREE EXAMINATION DECEMBER 2023
(Third Semester)

Branch – COMPUTER APPLICATIONS

PYTHON FOR MACHINE LEARNING

Time: Three Hours   Maximum: 50 Marks

SECTION-A (5 Marks)
Answer ALL questions
ALL questions carry EQUAL marks   (5 x 1 = 5)

1   Machine learning is a subset of which of the following _____?
     (i)   Artificial Intelligence   (ii) Deep Learning
     (iii) Data Learning   (iv) Data Wrangling

2   Which Python library is used for creating static and interactive visualizations?
     (i)   Pandas   (ii) NumPy
     (iii) Scikit-learn   (iv) Matplotlib

3   _____ refers to how well your model can represent all possible outcomes.
     (i)   bias   (ii) variance
     (iii) standard deviation   (iv) regression

4   Which of the following is required by K- means clustering?
     (i)   defined distance metric   (ii) number of clusters
     (iii) initial guess as to cluster centroids   (iv) all of the mentioned

5   Which of the following ensemble model helps in reducing variance?
     (i)   Boosting   (ii) Bagging
     (iii) Stacking   (iv) Voting

SECTION - B (15 Marks)
Answer ALL Questions
ALL Questions Carry EQUAL Marks   (5 x 3 = 15)

6   a   Bring out the behavioral characteristics of Data scientist.
                              OR
     b   Justify the role of a machine learning engineer in the data science workflow.

7   a   Compare List and Tuples.
                              OR
     b   What is NumPy and why it is important in Python data analysis?

8   a   Distinguish between Regression and Classification tasks in machine learning.
                              OR
     b   Write a note on Supervised, Unsupervised and Reinforcement learning.

9   a   Illustrate K-Means clustering with example.
                              OR
     b   Discuss the main assumption of the Naïve Bayes classifier.

10 a   Explain the basic concept of kernel trick in support vector machines.

OR

   b   State the main advantages of deep learning.

SECTION -C (30 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** Marks   (5 x 6 = 30)

11 a   Formulate the interdisciplinary knowledge and skills required to a data scientist.

OR

   b   Interpret Data Munging.

12 a   Classify the various data types in Python with suitable example.

OR

   b   State the following Iteration loop with example.
     i) For...Loop
     ii) Write a Python program to find the number is odd or even using for loop.

13 a   Justify the common techniques for feature selection.

OR

   b   Construct three machine learning tasks or techniques supported by Sci-kit-Learn.

14 a   Design ROC curve and calculate the AUC for a binary classification problem.

OR

   b   Elucidate the impact of noisy data on the performance of KNN algorithm.

15 a   Analyze the impact of varying ensemble size on the performance of a boosting algorithm.

OR

   b   Design a hierarchical clustering algorithm from scratch and explain each step in detail.

Z-Z-Z   END