PSG COLLEGE OF ARTS & SCIENCE
(AUTONOMOUS)

MCA DEGREE EXAMINATION MAY 2022
(Third Semester)

Branch – **COMPUTER APPLICATIONS**

PYTHON FOR MACHINE LEARNING

Time: Three Hours Maximum: 75 Marks

SECTION-A (10 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** marks (10 x 1 = 10)

1. Which open source tools used for clustering-computing framework for data analytics?
   a. R b. Julia
   c. Spark d. Apache Mahout

2. Scipy stands for
   a. Standard python b. Scientific Python
   c. Standalone Python d. Source Python

3. A List is denoted by _____
   a. ( ) b. { }
   c. [ ] d. " "

4. Exception handling in Python can be done with the _____structure.
   a. try... except b. try…catch
   c. try….later d. catch…throw

5. _____refers to the performance of a model against unseen data.
   a. Generalisation b. regularisation
   c. Variance d. Bias

6. Classification is a form of _____
   a. Supervised learning b. Unsupervised learning
   c. Reinforcement learning d. None of the given

7. _____ is an important part of the process to determine the effectiveness of the algorithm.
   a. Cluster validation b. Cohesion
   c. Separation d. k-means

8. _____ is used for continuous variables.
   a. Classification b. Regression
   c. Logistic Regression d. Separation

9. Hierarchical clustering is an _____ task.
   a. Supervised learning b. Unsupervised learning
   c. Reinforcement learning d. None of the given

10. Bagging stands for ____
    a. Bootstrap aggregation. b. Bootstrap degradation
    c. Bootstrap generalisation d. Bootstrap separation

SECTION - B (25 Marks)
Answer **ALL** questions
ALL questions carry **EQUAL** Marks (5 x 5 = 25)

11. a. Determine three pillars in the data science triumvirate

OR

b. Justify the open source tools that can be readily used in the data science workflow.

12. a. Python program to check whether a number entered by the user is positive, negative or 0.

OR

b. Python program to reverse a number using string slicing.

Cont…

13. a. Explain the impact of Artificial Intelligence in Machine Learning

OR

b. Describe about Euclidean distance & Manhattan distance.

14. a. What is meant by Confusion Matrix? Explain.

OR

b. Differentiate Cohesion and Separation in clustering.

15. a. How hierarchical clustering can be implemented using Scipy. Explain with the code.

OR

b. Justify about bagging & boosting in the ensemble techniques.

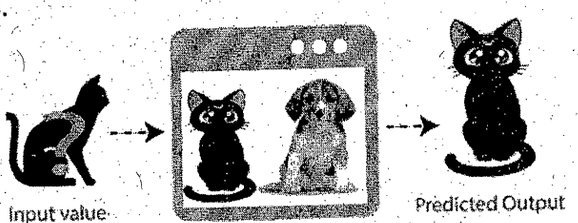
SECTION -C (40 Marks)

Answer **ALL** questions

ALL questions carry **EQUAL** Marks (5 x 8 = 40)

Question No.16 is Compulsory

16. Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, which algorithm, as it works on a similarity measure.



17. a. Design the steps in the data science workflow

OR

b. List the roles played by a data scientists? Give the characteristics of a Data scientist and a Data Science team.

18. a. Python program to sort alphabetically the words form a string provided by the user.

OR

b. Python program to display all the prime numbers within an interval.

19. a. Explain in details, how k-means clustering is carried out and implemented using python?

OR

b. Elaborate with the example for the naive Bayes classifier is based on text classification.

20. a. What is the use of Support Vector Machine? Write the code for implementing the same.

OR

b. Implement the various ensemble techniques using scikit-learn.

Z-Z-Z END