(6 pages)

Reg. No. : .....

Code No.: 6061 **Sub. Code: ZCAM 22**

M.C.A.(CBCS) DEGREE EXAMINATION, APRIL 2023.

Second Semester

Computer Applications – Core

MACHINE LEARNING USING PYTHON

(For those who joined in July 2021 onwards)

Time : 3 hours Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. The Scikit-learn package depends on ———— and ———— packages.
   (a) NumPy and Jupytor
   (b) NumPy and SciPy
   (c) DLL
   (d) COM

2. Which of the following is more appropriate to do feature selection?
   (a) Ridge (b) Lasso
   (c) Both (a) and (b) (d) Neither (a) nor (b)

3. PCA is ———— method.
   (a) Backward feature selection
   (b) Forward feature selection
   (c) Feature extraction
   (d) Data smoothing

4. In which of the following cases will K-means clustering fail to give good results?
   (a) Data points with outliers
   (b) Data points with different densities
   (c) Data points with non convex shapes
   (d) All

5. _____ refers to the process of converting or partitioning continuous attributes, features or variables to discretized or nominal attributes/features.
   (a) binning (b) discretization
   (c) trees (d) feature selection

6. We can find out which features have been selected using.
   (a) reshape ( ) (b) get-support ( )
   (c) hstack ( ) (d) All of the above

7. ———— is a statistical method of evaluating generalization performance that is more stable and thorough than using a split into a training and a test set
   (a) cross-validation (b) feature extraction
   (c) smoothing (d) k-fold

8. _____ is the number of correct predictions by the number of all samples.
   (a) false positive (b) accuracy
   (c) false negative (d) confusion matrix

9. The evaluation metrics which can be used to evaluate a model while modeling a continuous output variable is _____
   (a) AUC-ROC (b) Accuracy
   (c) recall (d) Mean-Squard-Error

10. The function _____ that will create a pipeline for us and automatically name each step based on its class.
    (a) pipeline( ) (b) create_pipeline( )
    (c) make_pipeline( ) (d) MinMaxScalar( )

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).
Each answer should not exceed 250 words.

11. (a) Recognize any five essential libraries and tools for python. Give examples.

    Or

    (b) Distinguish between overfitting and underfitting. How to avoid overfitting and underfitting?

12. (a) Hypothesize the types and challenges of unsupervised learning

    Or

    (b) Explain Non-Negative Matrix Factorization for feature extraction.

13. (a) Describe one-hot encoding with an example.

    Or

    (b) Infer the process of univariate nonlinear transformations.

14. (a) Discuss how cross validation is carried out in Scikit-learn.

    Or

    (b) Recall the usage of Precision-recall curves and ROC curves.

15. (a) Elaborate on building pipelines with sample code.

Or

(b) Describe the General Pipeline Interface.

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b)
Each answer should not exceed 600 words.

16. (a) Reframe the classification and regression algorithm with its types in machine learning.

Or

(b) Analyse NaïveBayes supervised machine learning algorithm.

17. (a) Elaborate on principal component analysis for feature extraction.

Or

(b) Establish K-means clustering in unsupervised machine learning.

18. (a) Indicate the three automatic feature selection methods.

Or

(b) How to enhance interactions and polynomial features to enrich feature representation.

19. (a) Interpret Stratified k-Fold Cross-Validation and Other Strategies.

Or

(b) Elaborate on Grid Search with Cross-Validation with the block representation.

20. (a) Propose detailed notes on Grid-Searching preprocessing steps and Model Parameters.

Or

(b) Illustrate Parameter Selection with Preprocessing.

———————