(6 pages)   Reg. No. : ....

Code No. : 10513 E   Sub. Code : CACS 41

B.Sc. (CBCS) DEGREE EXAMINATION,
APRIL 2024

Fourth Semester

Computer Science — Allied

MACHINE LEARNING TECHNIQUES

(For those who joined in July 2021–2022)

Time : Three hours   Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL the questions.

Choose the correct answer :

1. The term machine learning was coined in which year?

   (a) 1959   (b) 1960

   (c) 1961   (d) 1962

2. ———— is the machine learning algorithms that can be used with labeled data.

   (a) Regression algorithms

   (b) Clustering algorithms

   (c) Association algorithms

   (d) All of the above

3. ———— is a disadvantage of decision trees.

   (a) Decision trees are robust to outlier

   (b) Decision trees are prone to be overfit

   (c) Both (a) and (b)

   (d) None of these

4. Following are the types of supervised learning

   (a) Classification

   (b) Regression

   (c) Subgroup discovery

   (d) All of the above

5. ———— is a Python library for data wrangling and analysis.

   (a) Numpy   (b) Pandas

   (c) Jnode   (d) Jsnode

6. The ———— Notebook is an interactive environment for running code in the browser.

   (a) Jupyter   (b) Jnode

   (c) Jsnode   (d) Kyputer

7. Which of the following clustering requires merging approach?

   (a) Partitional

   (b) Hierarchical

   (c) Naive Bayes

   (d) None of the mentioned

8. Which of the following is not a type of association rule?

   (a) Positive rule   (b) Negative rule

   (c) Sequential rule   (d) Inverse rule

9. ———— is the task of assigning a label or class to a given text.

   (a) Text Classification   (b) Regression

   (c) Mining   (d) None of these

10. The Naive Bayes algorithm is a ———— machine learning algorithm.

    (a) Supervised   (b) Unsupervised

    (c) Both (a) and (b)   (d) None of these

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 250 words.

11. (a) Why Machine Learning?

    Or

    (b) How Python language used to develop machine learning algorithm?

12. (a) Define Simple Linear Regression.

    Or

    (b) Illustrate binary logistic regression.

13. (a) What is gradient based algorithm?

    Or

    (b) List the Scikit-Learn Library for machine learning.

14. (a) How does Clustering works?

    Or

    (b) Is collaborative filtering algorithm supervised or unsupervised? Explain.

15. (a) What is text analysis example?

Or

(b) Discuss about the Challenges in text analysis.

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Write about Framework for Developing Machine Learning Models.

Or

(b) What is data exploration in data visualization?

17. (a) Explain in detail about Multiple Linear Regression.

Or

(b) What is Credit Classification? Explain.

18. (a) Write and explain Gradient R Algorithm.

Or

(b) Why we need Advanced Regression Model?

19. (a) With example data explain K-Means clustering.

Or

(b) Discuss about Matrix factorization.

20. (a) Write in detail about Sentiment Classification.

Or

(b) Discuss text analysis with Tf-IDF Vectorization.

————