STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI – 600 086
(For Candidates admitted during the academic year 2019-2020 and thereafter)

B.C.A. DEGREE EXAMINATION – NOVEMBER 2024
FIFTH SEMESTER

COURSE : MAJOR CORE
PAPER : DATA SCIENCE
SUBJECT CODE: 19CS/MC/DS54
TIME : 3 HOURS **MAX. MARKS: 100**

SECTION – A

ANSWER ALL THE QUESTIONS **(20 X 1 = 20)**

Choose the correct answer:

1. _____ is NOT a built-in data structure in Python.
   A) List B) Dictionary C) Tuple D) Matrix
2. The purpose of the universal functions (ufuncs) in NumPy is _____.
   A) To create custom data types
   B) To perform element-wise operations on arrays efficiently
   C) To handle file input and output operations
   D) To perform matrix multiplication
3. _____ function is used for creating simple line plots.
   A) scatter() B) plot() C) bar() D) contour ()
4. _____ method can be used to interact with a web API.
   A) pd.read_csv() B) requests.get() C) np.array() D) matplotlib.pyplot()
5. _____ method would remove leading and trailing whitespace from a string in Pandas.
   A) trim() B) clean() C) strip() D) remove ()
6. _____ method is used to set a hierarchical index in a Pandas DataFrame.
   A) set_index() B) multi_index() C) create_index() D) index_level()
7. _____ is the term used for the process of reducing the complexity of a decision tree to prevent overfitting.
   A) Pruning B) Clustering C) Normalization D) Feature Selection
8. _____ metric is commonly used to evaluate the performance of a classification model.
   A) Mean Absolute Error B) R-squared C) Accuracy D) Variance
9. _____ symbol is used to denote a hashtag.
   A) @ B) # C) & D) %
10. _____ is a common task in NLP.
    A) Image recognition B) Speech recognition
    C) Predictive modelling D) Data compression

Fill in the blanks:

11. _____ NumPy's primary feature supports multidimensional arrays and matrix operations.
12. The easiest way to store data efficiently in binary format is using Python's built-in _____ serialization.
13. A _____ is a plot that shows the density of points in a 2D space.

…2

14. The Matplotlib function _____ is used to visualize data with errors.
15. _____ function would replace occurrences of a substring within a string.
16. _____ is a primary data structure in Pandas represents one-dimensional labelled data.
17. _____ is a technique used to assess the performance of a machine learning model by partitioning the data into subsets.
18. Dimensionality reduction is a technique used to reduce the number of _____ in a dataset while preserving more information as possible.
19. A graph is said to be _____ if there is a path between any two vertices in the graph.
20. _____ technique is commonly used for sentiment analysis in NLP.

SECTION - B

ANSWER ALL THE QUESTIONS: **(5 X 2 = 10)**

21. What is a list in Python?
22. Give the description of the parsing function read_csv in Pandas.
23. State the use of regular expression.
24. What is regression in machine learning?
25. List the uses of graph algorithm in analysing the social networks.

SECTION – C

ANSWER ANY EIGHT QUESTIONS: **(8 X 5 = 40)**

26. Explain the data structure dict with examples.
27. Write and describe the basic array statistical methods.
28. Explain the three methods used to display three-dimensional data in two dimensions using contours.
29. Illustrate on customizing ticks with examples.
30. Write a note on handling missing data with its functions.
31. Elucidate on hierarchical indexing of Pandas.
32. Explain on Generalization, Overfitting and Underfitting with a graph.
33. Describe the k-Nearest Neighbour machine learning algorithm.
34. Explain the process of removing the twitter handles.
35. Explain the steps involved in performing Sentiment Analysis.

SECTION – D

ANSWER ANY THREE OF THE FOLLOWING: **(3 X 10 = 30)**

36. Elaborate on implementing functions in Python.
37. Explain the process of reading and writing data in text format.
38. Illustrate merge and concatenation functions of Pandas with examples.
39. Discuss on K-Means Clustering.
40. Explain the concepts of tokenization and stemming in Natural Language Processing (NLP).

****