STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI – 600 086
(For Candidates admitted during the academic year 2019-2020 and thereafter)
SUBJECT CODE: 19CS/MC/DS54

B.C.A. DEGREE EXAMINATION – NOVEMBER 2022
FIFTH SEMESTER

COURSE : MAJOR CORE
PAPER : DATA SCIENCE
TIME : 3 HOURS **MAX. MARKS: 100**

SECTION – A

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

Choose the best answer:

1. Which of the following is used to define a block of code in Python language?
   a) Indentation b) Key c) Brackets d) All of the mentioned

2. Which one of the following is the use of function in Python?
   a) Functions don't provide better modularity for your application
   b) User defined functions cannot be created
   c) Functions are reusable pieces of programs
   d) All of the mentioned

3. Data can be visualized using_____.
   a) Graphs b) Charts c) Maps d) All of the mentioned

4. Which of the following is not a part of data science process?
   a) Discovery b) Model Planning c) Communication Building d) Operationalization

5. A/An _____ can be used as a data in Pandas.
   a) python dict b) ndarray c) scalar value d) All of the mentioned

6. _____ is the machine learning algorithms that can be used with labelled data.
   a) Regression algorithm b) Clustering algorithm
   c) Association algorithm d) All of the mentioned

7._____is used as an input to the machine learning model for training and prediction purposes.
   a) Feature b) Feature vector c) Fitting d) None of the mentioned

8. The problem of finding hidden structures in unlabelled data is called as _____.
   a) supervised learning b) unsupervised learning
   c) reinforcement learning d) None of the mentioned

9._____is the process for reducing inflected words to their root forms.
   a) Rooting b) text proofing c) rooting d) stemming

10._____ summarization is a kind of text summarization.
   a) Topic based b) Extraction based c) History based d) All of the mentioned

…2

Fill in the blanks:

11. _____ function is used for merging data frames.
12. List, tuple, and range are the _____ data Types.
13._____is the purpose of NumPy in Python.
14._____can be used to find the dimension of the array.
15. A graphical representation of information and data is called as_____.
16. _____technique is used to identify pages commonly viewed during the same visit to the website.
17. _____ method shows hierarchical data in a nested format.
18. The process of removing noise and incorrect input from a database is called _____.
19. _____is used to study of construction of words from primitive meaningful units.
20._____helps to select the meaning which makes most sense in context.

SECTION - B

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

21. Define Data wrangling.
22. What is Clustering?
23. List the uses of Matplotlib.
24. What is sentimental analysis?
25. State the importance of social network analysis

SECTION – C

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

26. Elucidate the sequence datatypes in python using an example.
27. What is python interpreter? Explain the difference between interactive mode and script mode?
28. Justify the role of data science
29. Discuss the various string manipulations in Python using an example.
30. Examine the functionality of Pandas in data analysis.
31. Demonstrate the features of graph theory in social network analysis
32. Identify the features that are to be analysed in a histogram.
33. Compare supervised and unsupervised learning
34. What is stemming? Explain with an example.
35. What is overfitting? How can we avoid it?

SECTION - D

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

36. What is Natural language processing? Explain the various pre-processing steps used in extracting features from tweets.
37. Demonstrate the various evaluation metrics for classification with an example.
38. Highlight the importance of visualization in data analysis and explain the various techniques with suitable example.
39. Give a detailed description on the various functions used in data transformation.
40. Explain with an example how multidimensional array objects are implemented in NumPy.

****