B.C.A. DEGREE EXAMINATION – NOVEMBER 2023
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. The maximum possible length of an identifier in Python is _____.
   a. 70 characters   b. 45 characters
   c. 65 characters   d. None of the above
2. The output of A [3] if A = {1,3,5,7,9,11,13,15} _____.
   a. 9   b. 7   c.5   d. None of the above
3. _____ is not an application for Data Science.
   a. Image and speech recognition   b. Privacy checker
   c. Recommendation systems   d. Online price comparison
4. _____ visually show the distribution of numerical data and skewness through displaying the data quartiles and averages.
   a. Scatter plot   b. Box plot
   c. Histogram   d. Work plot
5. Multiple data sources may be combined; it is referred as _____.
   a. Data Cleaning   b. Data Blending
   c. Data Integration   d. Data Transformation
6. Data transformation includes _____.
   a. A process to change data from a detailed level to a summary level
   b. A process to change data from a summary level to a detailed level
   c. Joining data from one source into various sources of data
   d. Separating data from one source into various sources of data
7. The model will have a _____ accuracy if it is overfitting.
   a. Even   b. High   c. Odd   d. Low
8. _____ is NOT supervised learning.
   a. Decision Tree   b. Linear regression
   c. Naïve Bayesian   d. PCA
9. _____ are the input and output of an NLP system.
   a. Speech and Noise   b. Noise and Written text
   c. Noise and value   d. Speech and written text
10. SNA can be described as _____.
    a. Study of graph theory   b. Study of relationships
    c. Study of human relationships by   d. All of the above
       means of graph theory

Fill in the blanks:

11. Full form of PEP is _____
12. The method used to sets the position of a file pointer _____.
13. _____ is defined as data files in text form or binary form with a structure that can be easily extracted by data mining algorithms.
14. A graphical representation of a data set is referred to as a _____.
15. _____ method returns a Boolean series indicating whether each row is a duplicate.
16. A _____ refers to an arrangement of elements.
17. _____language is best for Machine learning.
18. The unsupervised learning problems can be grouped as _____.
19. _____ is the act of splitting the text into words.
20. Data structure format in which graph can be represented _____.

SECTION - B

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

21. Write down the type conversions in Python.
22. Define Data Science.
23. List down any five vectorized string methods.
24. What is Naïve Bayes?
25. What is stemming?

SECTION – C

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

26. Explain the non-value returning function in Python.
27. Write a python program using shuffle () method.
28. How to read data from the text file? Explain.
29. Write python code to read and write in the database.
30. Describe about filtering the missing data.
31. Write a short note on index objects of pandas.
32. Explain K-Nearest Neighbors classification method.
33. Explain the types of unsupervised learning.
34. Write a note on graph dumping and sampling.
35. How to extract features from tweets? Explain.

SECTION – D

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

36. Explain the process of opening and closing a file in Python.
37. Describe about different types of plots in visualization.
38. Explain reshaping and pivoting operations briefly.
39. Distinguish between overfitting and underfitting in supervised learning.
40. Describe tweets preprocessing and cleaning of twitter handles.

****