ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE – 27
BCA(DATA ANALYTICS) – II SEMESTER
SEMESTER EXAMINATION – APRIL 2022
(EXAM CONDUCTED IN JULY - AUGUST 2022)
BCADA 2121: PRINCIPLES AND PRACTICES OF DATA SCIENCE USING R

Time: 2 hrs.   **Maximum marks: 60 marks**

This question paper contains THREE printed pages and THREE parts

PART A

Answer all questions from the following   **1 x 10 = 10 Marks**

1. Logistic regression is for predicting_____
   a) Continuous variable
   b) Categorical Variable
   c) Continuous and categorical
   d) None

2. R script files has an extension_____
   a) .RP
   b) .R
   c) .Rn
   d) None

3. Which one of the following function can be used to list the installed packages in R?
   a) library[]
   b) library()
   c) library{}
   d) None

4. Identify the term related to analytics.
   a) Analytics about explaining past events
   b) Analytics about explaining present events
   c) Analytics about predicting future events
   d) None

5. Which of the following is one of the key data science skills?
   a) Statistics
   b) Machine Learning
   c) Data Visualization
   d) All of the mentioned

6. _____uses data on some object to predict values for other object
   a) Inferential
   b) Exploratory
   c) Predictive
   d) None of the mentioned

7. Accurate prediction depends heavily on measuring the right variables.
   a) True
   b) False
   c) Can't say
   d) None

8. If P value is more than 0.05 then it is _____
   a) Null Hypothesis
   b) Alternate Hypothesis
   c) positive correlation
   d) No correlation

9. Correlation values will vary from _____
   a) 0 to 1
   b) -1 to 1
   c) infinity to 1
   d) none

10. Which of the following characteristic of big data is relatively more concerned to data science?
    a) Velocity
    b) Variety
    c) Volume
    d) None

PART B

Answer any six questions from the following   **6 x 5 = 30 Marks**

11. Write down the formula for precision, recall, specificity and F1 score.
12. How will you manage missing values for continuous and categorical variables?
13. Write short note on Natural language and Machine generated data.
14. How to enhance the value of data science for society?
15. Discuss about the validation techniques used in regression and classification problem.
16. List and explain the different types of machine learning.
17. Write down the formula for support, confidence and lift in apriori algorithm.
18. Explain about feature Engineering and model selection.

PART C

Answer any two questions from the following   **2 x 10 = 20 Marks**

19. Group the below data into two using k means clustering algorithm.

Income	Spending
40	20
50	15
60	25
20	35
65	35
80	19

20. Predict the output(y) for the following data using linear regression.

X	Y
1	3
2	4
3	2
4	4
5	5

21. Explain in detail about data science process.