



Date:
Registration number:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27
BCA(DATA ANALYTICS) - I SEMESTER
SEMESTER EXAMINATION - OCTOBER 2021
(Examination conducted in January-March 2022)

OEDS 1: BASICS OF DATA SCIENCE

TIME: 3 HOURS

MAX MARKS: 100

This question paper contains FOUR printed pages and THREE parts

PART A

Answer ALL the questions

20 X 1 = 20

1. Which of the following is characteristic of Processed Data?
 - a) Data is not ready for analysis
 - b) All steps should be noted
 - c) Hard to use for data analysis
 - d) None of the mentioned
2. Which of the following characteristic of big data is relatively more concerned to data science?
 - a) Velocity
 - b) Variety
 - c) Volume
 - d) None
3. Which of the following step is performed by data scientist after acquiring the data?
 - a) Data Cleansing
 - b) Data Integration
 - c) Data replication
 - d) All the mentioned
4. Which of the following is performed by Data Scientist?
 - a) Define the question
 - b) Create reproducible code
 - c) Challenge results
 - d) All of the mentioned
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. Which of the following 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) Cant say
 - d) None
8. If P value is more than 0.05?
- 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. Logistic regression is for predicting
- a) Continuous variable
 - b) Categorical Variable
 - c) continuous and categorical
 - d) None
11. R script files has an extension_____
- a) .RP
 - b) .R
 - c) .Rn
 - d) None
12. Which one of the following function is can be used to list the installed packages in R?
- a) library[]
 - b) library()
 - c) library{}
 - d) None
13. Which of the following is related to the meaning of the term analytics?
- a) Analytics about explaining past events
 - b) Analytics about explaining present events
 - c) Analytics about predicting future events
 - d) None
14. Which of the following is not considered a data analytics activity?
- a) Case study
 - b) data reporting
 - c) Data cleaning
 - d) None
15. Which of the following is considered data science?
- a) Case study
 - b) Qualitative analysis
 - c) Qualitative analysis
 - d) Forecasting
16. Which of the following is not an example of where Machine Learning is being applied today?
- a) symbolic reasoning

- b) Forecasting
 - c) Client retention
 - d) Image recognition
17. Point out the correct statement.
- a) Raw data is original source of data
 - b) Pre processed data is original source of data
 - c) Raw data is the data obtained after processing steps
 - d) None
18. Point out the wrong statement
- a) Merging concerns combining datasets on the same observations to produce a result with more variables
 - b) Data visualization is the organization of information according to preset specifications
 - c) Sub setting can be used to select and exclude variables and observations
 - d) All the above
19. Which of the following is one of the key data science skills?
- a) Statistics
 - b) Machine learning
 - c) Visualization
 - d) All the above
20. Which of the following is characteristic of Processed Data?
- a) Data is not ready for analysis
 - b) Hard to use for data analysis
 - c) All the mentioned
 - d) None

PART - B

2. Answer Any Ten Questions

10x5 = 50

- 21. Discuss about Benefits and uses of data science and big data.
- 22. Discuss about retrieve data in the data science process
- 23. Write short notes on Distributed file systems
- 24. Discuss about the validation techniques used in regression and classification problem.
- 25. Write down the types of machine learning and explain the same.
- 26. Explain in detail about data preparation.
- 27. Discuss about Data visualization options.
- 28. Explain in detail about one time presentation and a real-time dashboard.
- 29. Write down the formula for precision, recall, specificity and F1 score.
- 30. How will you manage missing values for continuous and categorical variables.
- 31. Write about Natural language and Machine generated data.
- 32. Write down the Applications for machine learning in data science.

PART - C

Answer Any Three Questions

10x3 = 3

33. Discuss the data preparation phase in detail for assessing risk when loaning money using big data techniques
34. Explain in detail about data science process.
35. With relevant examples discuss the facets of data in data science and big data
36. Discuss about Dashboard development tools.
37. Explain in detail about Exploratory Data Analysis.