

(6 pages)

Reg. No. :

Code No. : 5471

Sub. Code : ZCSE 23

M.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2023.

Second Semester

Computer Science

Elective — DATA SCIENCE AND BIG DATA
ANALYTICS

(For those who joined in July 2021 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. Which one of the following data that has no inherent structure, which may include text documents, PDFs, images. and video?
 - (a) Structured data
 - (b) Unstructured data
 - (c) Semi structured data
 - (d) Quasi-structured data

2. Which one of the following is develops data sets for testing, training, and production purposes in data analytics life cycle?
 - (a) data operationalize
 - (b) model planning
 - (c) model building
 - (d) communicate results
3. Which function is used to import the CSV file using R?
 - (a) read.csv ()
 - (b) get.csv ()
 - (c) get_csv()
 - (d) read_csv ()
4. Which function determines the type is stored in memory in R?
 - (a) typeof()
 - (b) infoof()
 - (c) dataof()
 - (d) memof()
5. The ———— function from the *arule* package implements to create frequent itemsets
 - (a) apriori()
 - (b) rule()
 - (c) arule()
 - (d) frequent()



6. Which one of the measure is used to discovering a rule?

- (a) confidence (b) itemset
- (c) support (d) transaction

7. Which one of the following is not a tree based classification?

- (a) ID3 (b) C4.5
- (c) CART (d) Naive Bayes

8. Which one of the following package is has a built-in naiveBayes function?

- (a) cart (b) c50
- (c) e1071 (d) arules

9. HDFS stands for _____

- (a) Hadoop Data File System
- (b) Hypertext Distributed File System
- (c) Hadoop Data Flow System
- (d) Hadoop Distributed File System

10. Which one of the following is read and write access to data sets with billions of rows and millions of columns?

- (a) Pig (b) Hive
- (c) Mahout (d) Hbase

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 250 words.

11. (a) Compare business intelligence and data science.

Or

(b) Examine the analyst perspective on data repositories with example.

12. (a) How do you read a data using R? Give example.

Or

(b) Explain the type – I and type – II errors in statistical methods.

13. (a) Illustrate the steps to determine the number of cluster in k-means.

Or

(b) Identify the various applications of association rules.

14. (a) Describe about the smoothing in classification.

Or

(b) Explain the text analysis with example.



15. (a) Discuss the concept of NoSQL with example.

Or

- (b) Discuss the joins in SQL for database analytics with example.

PART C — ($5 \times 8 = 40$ marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Describe about the various data structures used in big data analytics with example.

Or

- (b) Explain the data analytics life cycle with example.

17. (a) Illustrate the step to visualizing a data using R.

Or

- (b) Classify the various statistical methods for data analytics with example.

Page 5

Code No. : 5471

18. (a) Analyze the need of association rule mining in data science.

Or

- (b) How is the logistic regression used as a classifier? Examine it.

19. (a) Write a code to implement the ID3 classification using R.

Or

- (b) Write a code to implement text analysis using R.

20. (a) Discuss the process of map reduce in hadoop with example.

Or

- (b) Discuss the following for database analytics using SQL.

(i) UNION ALL

(ii) INTERSECT.

Page 6

Code No. : 5471

