

(6 pages)

Reg. No. : .....

Code No.: 20112 E Sub. Code: SECS 6 B/  
SESE 6 B

B.Sc. (CBCS) DEGREE EXAMINATION,  
NOVEMBER 2023.

Sixth Semester

Computer Science / Software Engineering

Major Elective – BIG DATA ANALYTICS

(For those who joined in July 2017–2019)

Time : Three hours Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. What is the term used for a collection of large, complex data sets that cannot be processed using traditional data processing tools?  
(a) Big Data (b) Small Data  
(c) Medium Data (d) Mini Data

6. Which of the following is a technique for grouping similar data points together?  
(a) Classification  
(b) Regression  
(c) Clustering  
(d) Dimensionality Reduction
7. Which of the following is not a common data preprocessing technique?  
(a) Normalization  
(b) One-Hot Encoding  
(c) Dimensionality Reduction  
(d) Regression
8. Which of the following is a measure of the relationship between two variables?  
(a) Correlation (b) Covariance  
(c) Standard Deviation (d) Mean
9. Which of the following is a measure of how much a dependent variable changes when an independent variable changes?  
(a) Covariance (b) Correlation  
(c) Slope (d) Intercept

Page 3 Code No. : 20112 E

2. Which of the following is not one of the four V's of Big Data?  
(a) Velocity (b) Volume  
(c) Variety (d) Value
3. Which of the following is a tool used for processing and analyzing Big Data?  
(a) Hadoop (b) MySQL  
(c) PostgreSQL (d) Oracle
4. What is the process of examining large and varied data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences, and other useful information?  
(a) Data Mining (b) Data Warehousing  
(c) Data Integration (d) Data Processing.
5. Which of the following is a technique used to extract meaningful insights from data sets that are too large or complex to be processed by traditional data processing tools?  
(a) Business Intelligence  
(b) Machine Learning  
(c) Artificial Intelligence  
(d) Data Science

Page 2 Code No. : 20112 E

10. Which of the following is a measure of how much a model's predictions vary for different input values?  
(a) Bias (b) Variance  
(c) Precision (d) Recall

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).  
Each answer should not exceed 250 words.

11. (a) What is Bigdata? Describe the main features of a big data in detail.  
Or  
(b) Explain Extracting information from Unstructured Data.
12. (a) Briefly discuss about MapReduce and YARN.  
Or  
(b) Explain about k-means Clustering in detail.
13. (a) What is Hadoop? Explain its components.  
Or  
(b) Which benefits can you obtain from implementing fraud detection using Big Data analytics?

Page 4 Code No. : 20112 E

[P.T.O.]

14. (a) Explain the architecture of big data foundation.

Or

(b) How to Choosing an Analytics Tools?

15. (a) Briefly describe about the rise of the data Scientist.

Or

(b) Explain when data starts making decisions.

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b)  
Each answer should not exceed 600 words.

16. (a) Given in detail about Nature of Data and its applications.

Or

(b) Briefly describe about Storage Considerations in Big Data.

17. (a) Explain about Classification of Decision trees in detail.

Or

(b) Discuss in detail about Naïve Bayes Classification.

Page 5 Code No. : 20112 E

18. (a) Briefly give difference between Big Data and Cloud Computing.

Or

(b) Explain various types and uses of R and D.

19. (a) Give detailed note on data foundation layers.

Or

(b) Derive the creation and consumption of analytics.

20. (a) Briefly give details about grab data.

Or

(b) Explain the importance of Privacy-Enhancing Technologies.

Page 6 Code No. : 20112 E