

(6 pages)

Reg. No. :

Code No. : 20323 E Sub. Code : ASCS 31

B.Sc. (CBCS) DEGREE EXAMINATION,
NOVEMBER 2022.

Third Semester

Computer Science

Skill Based Subject — INTRODUCTION TO BIG DATA
ANALYTICS

(For those who joined in July 2020 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. One terabyte is equals to _____ bytes.
- (a) 10^{12} (b) 10^{15}
(c) 10^{21} (d) 10^{17}

2. Sensors placed on the objects provide information about their environment, the conditions of use and the objects operations
- (a) autonomy (b) monitoring
(c) control (d) data collection
3. The analytics which focuses on reporting on what happened in the past
- (a) *data collection*
(b) *modeling and analysis*
(c) *interpretation*
(d) *descriptive*
4. An algorithm is a _____; a user can introduce data (inputs) and they will obtain the results (outputs).
- (a) procedure (b) instructions
(c) black box (d) goal
5. _____ is the analysis of data contained in natural language text.
- (a) Text mining
(b) Machine learning
(c) Artificial intelligence
(d) None of the above

6. The collected, cleaned and prepared data can now be explored in _____ phase.
(a) second (b) third
(c) fourth (d) fifth
7. A supervised learning task is called "classification" if the outputs are discrete or _____ if the outputs are continuous.
(a) Cluster (b) Modeling
(c) Regression (d) None of the above
8. _____ are graph structures where each potential decision creates a new node, resulting in a tree-like graph.
(a) Random forest (b) Decision trees
(c) Neural networks (d) Classification
9. _____ goal coupling big data and ML is to enable customization, merchandising and A/B testing of new features to enhance the user experience.
(a) eBay's (b) spotify
(c) IBM (d) None of the above

10. _____ extract knowledge from a set of data containing input-output pairs.
(a) Unsupervised algorithms
(b) Supervised algorithms
(c) Clustering
(d) Text mining

PART B — (5 × 5 = 25 marks)

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

11. (a) Write the definition of big data.
Or
(b) Write notes on Beyond 3v's.
12. (a) What is data analytics?
Or
(b) Explain the analytics culture within companies.
13. (a) What is machine learning?
Or
(b) Write notes on
(i) *Defining the tasks to be accomplished*
(ii) *Which technology to adopt?*

14. (a) Describe supervised learning.

Or

(b) Write notes on neural network.

15. (a) Write about algorithm selection in detail.

Or

(b) Write notes on

(i) Amazon

(ii) NetFlix.

PART C — (5 × 8 = 40 marks)

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

16. (a) What are the challenges the companies has to overcome in bigdata?

Or

(b) Explain about the data revolution.

17. (a) Explain

(i) Advanced analytics in new paradigm

(ii) Statistical and computational paradigm.

Or

(b) Explain the data analytics address.

Page 5 Code No. : 20323 E

18. (a) Mention the disciplines that support the big data analytics process.

Or

(b) How the fifth phase works for transform data into actionable knowledge?

19. (a) What is classifications? Write any two classification algorithm.

Or

(b) Explain about principles of clustering algorithm.

20. (a) Differentiate supervised or unsupervised algorithm : in which case do we use each one?

Or

(b) Discuss about other ML algorithm.

Page 6 Code No. : 20323 E