D3BMC2301 (PAGES: 2) **Reg. No.....**

Name: .....

THIRD SEMESTER B.Sc. DEGREE EXAMINATION, NOVEMBER 2024

(Regular/Improvement/Supplementary)

COMPUTER SCIENCE & MATHEMATICS (DOUBLE MAIN)
GDCS3A01T: INTRODUCTION TO DATA SCIENCE

Time: 2 ½ Hours **Maximum Marks: 80**

SECTION A: Answer the following questions. Each carries *two* marks.
(Ceiling 25 marks)

1. How is a software developer different from a data scientist?
2. What is Neo4j?
3. What do you mean by distributed file system?
4. Briefly explain the data science process.
5. What is the relevance of software security in the field of data science?
6. Describe the facets of data with appropriate examples.
7. Discuss about oath and regulations related to data science.
8. What is Big data? Bring out the usefulness of data science and big data.
9. Briefly discuss about the major challenges faced while handling large data and possible solutions.
10. Differentiate between variety and veracity of data.
11. Explain about the ACM code of ethics.
12. What is structured data? Give an example.
13. Explain the CAP theorem.
14. Discuss the different data structures that could be used in big data applications.
15. What is the need of regulations in the data science industry?

SECTION B: Answer the following questions. Each carries *five* marks.

(Ceiling 35 marks)

16. How can the five Cs be effectively implemented?
17. What are outliers? Explain in detail how data cleansing could be performed.
18. Explain how integration and transformation are performed on large data.
19. What is connected data? Bring out the differences between ACID and BASE model.
20. With the help of case study on ‘risk assessment of loaning money’, explain in detail, the data processing framework used in Big data.
21. Describe the challenges of conventional systems with respect to Big data. Briefly discuss about the approaches to overcome the same.

(PTO)

22. Briefly describe the different sources of data.

23. Discuss about the challenges in building ethics into a data driven culture.

SECTION C: Answer any *two* questions. Each carries *ten* marks.

24. With the help of a case study on disease prediction explain the data science process.

25. Describe the characteristics of Big data. Explain the 6V's in detail.

26. Enumerate the base principles of NoSQL databases. Describe in detail about graph databases.

27. Explain in detail the five framing guidelines for building data products.

(2 × 10 = 20 Marks)