QP CODE: 24803032

24803032

Reg No : .....

Name : .....

INTEGRATED MSC DEGREE EXAMINATION, MAY 2024

Seventh Semester

INTEGRATED MSC COMPUTER SCIENCE-ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

CORE - ICSA7CR4 - DATA SCIENCE AND ANALYTICS

2020 Admission Onwards

8EB0FB13

Time: 3 Hours Weightage: 30

Part A (Short Answer Questions)

*Answer any **eight** questions.*

*Weight **1** each.*

1. What you mean by Data Analytics?
2. List some benefits of using Tensorflow.
3. What importance does data quality hold in the field of data science, and what are some key factors that contribute to ensuring high-quality data for analysis?
4. What is primary data collection?
5. Name the types of sampling bias.
6. What makes nominal data different from ordinal data?
7. What are the different types of data visualization?
8. What is nominal and ordinal data? Explain with examples.
9. How to add titles to subplots in Matplotlib?
10. What is the diffrence between lists and tuples in python?

(8×1=8 weightage)

Part B (Short Essay/Problems)

*Answer any **six** questions.*

*Weight **2** each.*

11. What you mean by data ? What are the different types of data?

12. How does TensorFlow serve as a valuable tool in data science, and what are its primary applications in the field?

13. What are the 5 principles of data ethics in business professionals?

14. What are some techniques used for sampling? What is the main advantages of sampling?

15. What do you mean by data visualization?

16. Explain the conventional data visualization methods.

17. Explain the significance of utilizing a correlation matrix in data visualization, elaborating on how it aids in uncovering relationships between variables. Provide practical examples of scenarios where a correlation matrix proves valuable, and discuss its role in identifying patterns, dependencies, or trends within complex datasets.

18. What is Pandas in Python, and how does it serve as a powerful data manipulation and analysis library? Briefly explain its key features and functionalities.

(6×2=12 weightage)

Part C (Essay Type Questions)

*Answer any **two** questions.*

*Weight **5** each.*

19. What are the key terminologies used in the field of data science, and how do various methods of data repository, such as databases, data lakes, and distributed storage, contribute to the effective management and utilization of data in the context of data science projects?

20. What is the use of statistics in data science?

21. Design a step-by-step process for building a basic Naive Bayes classification model in Python, including data preprocessing, model training, and evaluation steps. Provide code snippets and explanations for each key stage, emphasizing the essential components and considerations in achieving a successful implementation.

22. What is Bokeh in Python, and how does it contribute to data visualization? Provide a concise overview of Bokeh's key features and explain how it enables the creation of interactive and aesthetically pleasing visualizations for effective data representation.

(2×5=10 weightage)