SUBJECT CODE: 19BI/PC/BD44

M. Sc. DEGREE EXAMINATION, APRIL 2023
BIOINFORMATICS
FOURTH SEMESTER

COURSE   :   **CORE**
PAPER   :   **BIG DATA ANALYSIS**
TIME   :   **3 HOURS**   **MAX. MARKS: 100**

SECTION – A

ANSWER ALL QUESTIONS   **(20 X 1=20)**

1. What is data?
2. What is unstructured data?
3. What is the role of data science in drug discovery?
4. What is big data?
5. What techniques are commonly used in medical image analysis to segment brain tumors?
6. What are the benefits of using electronic health records (EHRs) for healthcare data management.
7. How can big data analytics be used to improve patient outcomes in healthcare?
8. How can big data be used to drive decisions and gain a competitive advantage in today's digital age?
9. What are the three V's of big data and what do they represent?
10. What are some common sources of big data in today's digital landscape?
11. How does big data differ from traditional data in terms of volume, velocity, and variety?
12. What are the five steps involved in the process of structuring big data analysis to extract value from it?
13. How can healthcare organizations effectively apply the five-step process to gain insights and make informed decisions from their big data?
14. What are the key components of a successful big data strategy, and how can organizations ensure they align with their business goals?
15. How can genomics leverage big data to drive innovation, increase operational efficiency, and gain a competitive advantage in their industry?
16. What are the first and last steps in the data science process?
17. What is the purpose of exploratory data analysis in the data science process?
18. What is the difference between supervised and unsupervised learning in machine learning?
19. What is Hadoop, and what is its primary function?
20. What are the two core components of Hadoop, and what are their respective functions?

SECTION – B

ANSWER ANY FOUR QUESTIONS..   **(4 x 10 = 40)**

21. Write notes on the human genome project as a typical case study for biomedical big data.
22. Discuss What are some ethical considerations when using patient data for healthcare research?
23. What is probabilistic modeling, and what is its primary goal?

24. What are some of the main challenges associated with processing and analyzing big data, and how are these challenges typically addressed?
25. What are the different steps involved in the data science process, and how do they relate to each other?
26. How does a distributed file system handle data storage and retrieval across multiple nodes, and what are some of the key challenges associated with this approach?
27. What are some of the key benefits of cloud computing, and how do they enable organizations to scale their computing resources more effectively?

SECTION – C

ANSWER ANY TWO QUESTIONS..   **(2 x 20 = 40)**

28. What is big data analytics, and how is it used to extract insights and value from large and complex data sets? Describe some of the key technologies and tools that are commonly used in big data analytics, and explain how they enable data scientists to process, analyze, and visualize massive amounts of data.

29. How do the four Vs of big data - volume, velocity, variety, and veracity - impact the analysis and interpretation of large and complex data sets, and what are some of the key challenges associated with each of these characteristics?

30. How can data visualization techniques be effectively utilized to enhance the insights gained from data analysis, and what are the key considerations and best practices for designing and implementing effective visualizations that aid in decision-making processes?

31. What is Hadoop, and what are some of its key features and components? Describe how Hadoop is used to process and analyze large volumes of data, and discuss the benefits and limitations of this approach.

****