B **302B4** **Total pages: 2**

Register No: …… Name: .....

SAINTGITS COLLEGE OF ENGINEERING (AUTONOMOUS)

(AFFILIATED TO APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY, THIRUVANANTHAPURAM)

EIGHTH SEMESTER B.TECH. DEGREE EXAMINATION(R), MAY 2024

Mechanical Engineering

(2020 SCHEME)

Course Code : 20MET442

Course Name : Data Analytics for Engineers

Max. Marks : 100 **Duration:3 Hours**

Scientific calculator and statistical table is allowed in the examination hall.

PART A

(Answer all questions. Each question carries 3 marks)

1. What does ETL stand for in data preparation? Describe at least two ETL steps.
2. What are the three common measures of central tendency?
3. Define dimensionality reduction and explain its significance in machine learning and data analysis.
4. Name two applications of Principal Component Analysis (PCA) in real-world data analysis.
5. Define descriptive analytics and explain how it contrasts with predictive analytics
6. Define association rule mining and explain its significance in data mining.
7. List at least three applications for big data analytics
8. Define a big data framework and explain its significance in handling large volumes of data.
9. What is anomaly detection and how is it related to fraud detection
10. What is meant by Sentiment Analysis? How is it utilized in Social Media Analytics?

PART B

(Answer one full question from each module, each question carries 14 marks)

MODULE I

11. Discuss the factors that organizations should consider when selecting data analytical tools for their business needs. How can the choice of tools impact the efficiency and effectiveness of data analysis processes? 14

OR

12. Evaluate the factors that contribute to prediction error in predictive modeling. How can model performance be improved by reducing prediction error? 14

MODULE II

13. a. Evaluate the advantages and disadvantages of decision tree algorithms compared to other machine learning techniques. Provide examples of scenarios where decision trees are particularly useful or unsuitable. 4
    b. Breifly explain the process of constructing a decision tree for classification tasks. 10

OR

14. The experience in years and monthly income (in 1000s) for members of a team is given below.

Experience	Monthly Income

Experience	Monthly Income
11	10
7	8
9	9
5	5
8	9
6	7
10	11

(a) Using the method of least squares, calculate the regression equation for income vs years of service. 10

(b) For a person with experience of 13 years, what should be the expected income? 4

MODULE III

15. Describe the concept of clustering in unsupervised learning, explaining at least two approaches in detail. 14

OR

16. For the given data, compute two clusters (using centrod method) using K-means algorithm for clustering where initial cluster centers are (1.0, 1.0) and (5.0, 7.0). Execute for two iterations. 14

x	y
1	1
1.5	2
3	4
5	7
3.5	5
4.5	5
3.5	4.5

MODULE IV

17. Discuss how big data is different from small data. Use the 6V's of big data to explain. 14

OR

18. Discuss the role of data governance in big data management and analytics. How does effective data governance contribute to data quality, privacy, and compliance in large-scale data environments? 14

MODULE V

19. a. Discuss the role of Hadoop in Big Data processing and analytics. 5

b. Explore the limitations of the MapReduce paradigm in handling complex data processing tasks. Discuss how newer frameworks like Apache Spark address these limitations and provide advantages over traditional MapReduce implementations. 9

OR

20. Describe at least two types of Big Data storage methods, its characteristics and tools. 14

****