

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**Computer Science and Engineering****(2020 SCHEME)****Course Code : 20CST464****Course Name : Data Mining****Max. Marks : 100****Duration:3 Hours****PART A***(Answer all questions. Each question carries 3 marks)*

1. Explain roll-up operation with the help of a diagram.
2. What are the characteristics of OLAP?
3. What is data preprocessing? What are the different techniques for data preprocessing?
4. What is data reduction? What are the different data reduction strategies?
5. Explain the error measures.
6. Explain PAM.
7. How is dynamic itemset counting implemented?
8. Discuss two advantages of the FP-Growth algorithm compared to other frequent itemset mining algorithms like Apriori.
9. Write any three applications of text mining.
10. How does the Clever algorithm improve upon PageRank in terms of identifying communities on the web?

PART B*(Answer one full question from each module, each question carries 14 marks)***MODULE I**

11. What is data warehouse? Explain in detail, the differences between Operational Database Systems and Data Warehouses. 14

OR

12. Explain the Mining methodology and user interaction issues in data mining. 14

MODULE II

13. Explain the different steps in data integration. 14

OR

14. Explain with example dimensionality reduction? How is it different from numerosity reduction? 14

MODULE III

15. Explain hierarchial clustering algorithm. What are features of DBSCAN algorithm. 14

OR

16. With help of an example, explain in detail ID3 algorithm. 14

MODULE IV

17. Illustrate the apriori algorithm with an example. Explain the applications of Apriori Algorithm. 14

OR

18. a) Explain the working of Pincer Search Algorithm. 7

b) What are the applications of Pincer Search Algorithm. 7

MODULE V

19. List and explain the different data structures used for web usage mining? 14

OR

20. Describe the three primary categories of web mining and their respective applications in extracting valuable insights from the vast amount of online data. 14
