R18

Code No: 157BC

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD B. Tech IV Year I Semester Examinations, February/March - 2022 DATA MINING

(Common to CSE, IT)

Time: 3 Hours Max. Marks: 75

Answer any Five Questions All Questions Carry Equal Marks

- - -

- 1. Explain the need of data preprocessing and various forms of preprocessing. [15]
- 2. What is a data warehouse? Demonstrate integrating data mining system with a data warehouse with a neat diagram. [15]
- 3. Apply FP-Growth algorithm to the following data for finding frequent item sets, consider support threshold as 30%. [15]

TID	List of ItemIDs
1	I1, i2, i4, i5
2	I2, i4, i7
3	I2,i3,i4,i5
4	I1,i3,i4,i7
5	I1,i2,i3,i4,i5
6	I3,i4,i5,i6

4.a) How to identify sub graphs in a graph?

b) Give an overview of correlation analysis.

[8+7]

5.a) Explain classification as a two step process.

b) State Bayes theorem. How this concept is used in classification.

[8+7]

6. What is a decision tree? Explain decision tree induction algorithm.

[15]

7.a) Contrast k-means clustering with k-medoids clustering approach.

b) Discuss the merits and demerits of hierarchical approaches for clustering.

[8+7]

8. How to apply mining techniques to unstructured text database? Explain with example.

[15]

---ooOoo---