

**R16**

Code No: 137BQ

**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD**

**B. Tech IV Year I Semester Examinations, January/February - 2023**

**DATA MINING**  
**(Common to CSE, IT)**

**Time: 3 Hours**

**Max. Marks: 75**

**Note:** i) Question paper consists of Part A, Part B.

ii) Part A is compulsory, which carries 25 marks. In Part A, Answer all questions.

iii) In Part B, Answer any one question from each unit. Each question carries 10 marks and may have a, b as sub questions.

**PART - A**

**(25 Marks)**

- 1.a) Define data mining. [2]
- b) What is meant by outlier analysis? [3]
- c) Define maximal frequent item set. [2]
- d) How to compute confidence of an association rule? Give example. [3]
- e) What is meant by test data? [2]
- f) What is the significance of information gain? [3]
- g) What is cluster analysis? [2]
- h) What are the drawbacks of single linkage clustering? [3]
- i) Give examples for un structured text. [2]
- j) List the applications of web usage mining. [3]

**PART - B**

**(50 Marks)**

2. Discuss the steps in knowledge discovery process and compare it with data access and information retrieval. [10]

**OR**

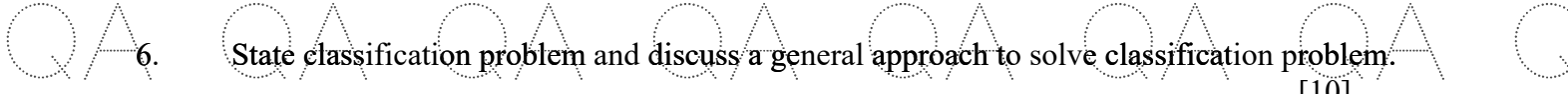
- 3.a) Appraise usage of smoothing in data transformation.
- b) Evaluate distance measures for dissimilarity computation. [5+5]
4. Apply FP –Growth algorithm to the following transactional database to find frequent item sets. [10]

	List of items
001	I1,I3,I5,I7
002	I1,I5,I6,I7
003	I6,I7
004	I2,I3,I6,I7
005	I8,I1,I6
006	I2,I5,I8



**OR**

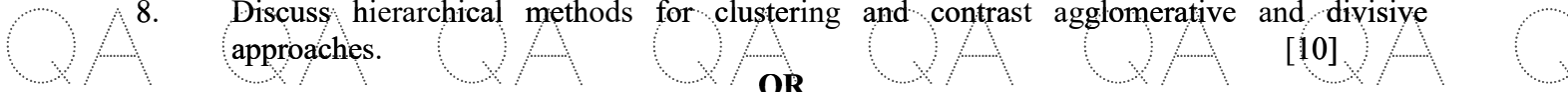
- 5.a) Appraise the limitations of Apriori and suggest mechanisms to improve it.
- b) Explain item merging concept for mining closed frequent item sets. [5+5]



- 6. State classification problem and discuss a general approach to solve classification problem. [10]

**OR**

- 7.a) Discuss decision tree over fitting and pruning techniques.
- b) Justify the selection of k value for kNN classifier. [5+5]



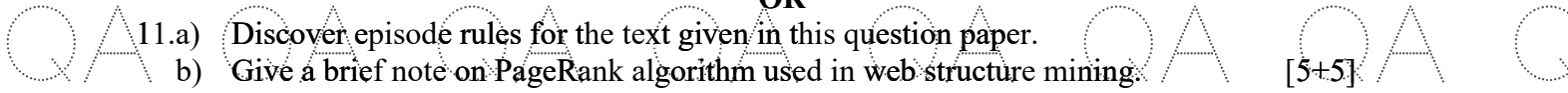
- 8. Discuss hierarchical methods for clustering and contrast agglomerative and divisive approaches. [10]

**OR**

- 9.a) With suitable data explain statistical based outlier detection.
- b) Criticize the evaluation metrics used for clusters. [5+5]

- 10. Discuss the data mining tasks applicable to text databases. [10]

**OR**



- 11.a) Discover episode rules for the text given in this question paper.
- b) Give a brief note on PageRank algorithm used in web structure mining. [5+5]

**---ooOoo---**

