

Time: 3 hours

Max. Marks: 80

N.B. (1) Question one is Compulsory.

(2) Attempt any 3 questions out of the remaining.

(3) Assume suitable data if required.

- Q. 1 a) Explain CLARANS extension in web mining 05
- b) Explain in detail the extract/transform/load (ETL) design of an automated warehouse. 05
- c) What is prediction? Explain about Linear regression method. 05
- d) Suppose data for clustering is {6,14,18,22,1,40,50,11,25}. Consider  $K=2$ , Cluster the given data using K-means algorithm 05
- Q. 2 a) Briefly outline with example, how to compute dissimilarity between the objects describe following
- i) Nominal attributes
- ii) Asymmetric binary attributes 10
- b) Discuss about a three-tier data warehouse architecture. 10
- Q. 3 a) Describe the various phases in knowledge discovery process with a neat diagram 10
- b) Explain Decision tree induction algorithm for classification. Discuss the usage of information gain in this.
- Q. 4 a) Following table gives fat & protein content of the items. Apply single linkage clustering & dendrogram 10

Food Item	Protein	Fat
1	1.1	60
2	8.2	20
3	4.2	35
4	1.5	21
5	7.6	15
6	2.0	55
7	3.9	39

b) Consider the following transactions

10

TID	ITEMS
01	1,3,4,6
02	2,3,5,7
03	1,2,3,5,8
04	2,5,9,10
05	1,4

Apply the Apriori algorithm with minimum support of 30% and minimum confidence of 75% and find large item set.

Q. 5 a) Explain Hyperlink Induced Topic Search Algorithm (HITS) Algorithm with example

10

b) What is market basket analysis? Explain with an example. State and explain with formula the meaning of the following term:

i) Support

ii) Confidence

10

Q. 6 a) Describe the working of K-medoid clustering with the help of sample dataset.

10

b) Define multidimensional and multilevel association mining

10

\*\*\*\*\*