

SRI KRISHNA INSTITUTE OF TECHNOLOGY

Bangalore

Department of Computer Science & Engineering

Subject: Data Mining and Data Warehousing

Sub Code: 18CS641

QUESTION BANK – 2

1. What is data mining ? what are the applications of data mining.
2. Explain Knowledge data discovery KDD with a neat diagram. .
3. Discuss the challenges that motivate the development of Data Mining. .
4. Explain the origin of data mining .
5. What is data mining? Explain various data mining task with examples .
6. What are data and data attributes ? Explain the types and properties of attributes. .
7. Differentiate between discrete and continuous attributes.
8. Distinguish between categorical and numerical attributes. .
9. Explain the types of data sets.
10. List and explain general characteristics of data sets.
11. What is data quality? What are the dimension that asses the data quality.
12. Describe any five data preprocessing approaches. .
13. What is sampling? Explain simple random sampling v/s stratified sampling v/s progressive sampling.
14. Describe the various approaches for feature selection. .
15. What is curse of dimensionality? Explain .
16. What is similarity and dissimilarity? Explain similarity and dissimilarity measures between simple attributes based on different types of attributes. .
17. Discuss the measures of proximity between objects that involve multiple attribute.
18. Explain the cosine similarity for calculating the similarity of two documents with an example.
19. Explain the difference between ROLAP and MOLAP
20. Define data mining. Explain the process of knowledge discovery in databases
21. Explain indexing OLAP data. Bit map index with example
22. Explain any 3 motivating challenges of Data Mining
23. Explain various tasks of data mining with example for each
24. Consider the following vectors. Find a) Simple Matching Co-efficient b) Jaccard Co-efficient c) Hamming Distance .

i) X: 0101010001 Y: 0100011000

25. Give the equation for defining Euclidean, Minkowski distance. Consider the following vectors. Find Hamming distance X=0101010001 Y=0100011000

8.