

N.B.

- 1. Attempt any two questions from question numbers 1, 2, 3 and any two questions from question numbers 4, 5, 6.**
- 2. Figures to the right indicate full marks**
- 3. Simple non-programmable calculator is allowed.**

- 1 a. Define Data mining. Differentiate between data mining processing and database processing with respect to query, data and output of a query. (05)
- b. Explain different steps of KDD process. Is data mining a part of this process? Explain. (05)
- c. Write a short note on text mining. (05)
- 2 a. Discuss what is meant by the following terms while describing the characteristics of the data in a data warehouse: (i) subject oriented, (ii) integrated, (iii) time-variant, (iv) non-updatable. (05)
- b. What are the different data integration problems faced by user while loading data into data warehouse? Justify your answer. (05)
- c. Explain discretization process to handle noisy data in preprocessing of data. Discretize the set of attribute age: {4, 8, 16, 20, 21, 30, 22, 27, 32, 34} using equi-width binning method by taking bin width = 16. (05)
- 3 a. Define surrogate key in designing data model. Why there is a need of this key? Explain with an example. (05)
- b. Define snowflake schema. Consider single source facts about a retailer's sales activity and shipments. Form a fact table and hence identifying at least two dimensions and create snowflake schema. (05)
- c. Define a term OLAP and explain OLAP operations: roll-up, drill-down, slicing and dicing. (05)
- 4 a. Derive the control limits for acceptance process level and α risk for modifications of control charts. (08)
- b. Briefly describe the concept of Latent structure methods. (07)
- 5 a. Obtain the expression for mean μ and variance covariance matrix Σ in multivariate process monitoring. (08)
- b. For a process with USL = 62, LSL = 38, $\sigma = 2.0$ and $\mu = 50$. Compute PCR and PCRk. Comment on the same for $\mu = 56, 60, 65$ respectively. (07)
- 6 a. Explain multiple stream processes with special reference to a Group Control charts. (08)
- b. Describe sample and population autocorrelation function. Also specify the need for ARIMA modeling. (07)