**Enrollment No: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**PARUL UNIVERSITY**

**FACULTY OF INSTITUTE OF ENGINEERING & TECHNOLOGY**

**B.Tech Mid-Sem Exam 2024-25**

**Semester: 5 Date: (dd/mm/yyyy)**

**Subject Code: (303108303) Time: (1hr: 30min)**

**Subject Name: (DMDV) Total Marks: 40**

|  |  |  |
| --- | --- | --- |
| Sr. No. |  | Marks |
| Q.1 | (A) Five Multiple Choice Questions  (1) Which of the following represents the correct sequence of steps in the KDD process?  (a) Data mining → data cleaning → data integration → data transformation  (b) Data cleaning → data mining → data integration → data transformation  (c) Data integration → data mining → data cleaning → data transformation  (d) Data transformation → data mining → data integration → data cleaning  (2) Identify the correct sequence in ascending order.  (a) Knowledge-> Information-> Data  (b) Information-> Knowledge-> Data  (c) Data-> Information-> Knowledge  (d) Data-> Knowledge-> Information  (3) Data means \_\_\_\_\_\_\_\_\_ & information means \_\_\_\_\_\_\_\_\_\_\_\_\_.  (a) raw facts, meaningful data  (b) meaningful data, raw facts  (c) meaningful information, unmeaningful data  (d) none of the above  (4) What is the purpose of dimensionality reduction in data mining?  (a) Increasing data redundancy  (b) Reducing the number of features or attributes  (c) Enhancing data visualization  (d) Aggregating data from multiple sources  (5) Which method/learning is used for prediction on unlabeled data?  (a) Un-supervised learning  (b) Reinforcement learning  (c) Supervised learning  (d) none of the above | 05 |
|  | (B) Five Fill in the blanks or One-line questions  (1) Define Outliers.  (2) \_\_\_\_\_\_\_\_\_\_\_\_\_\_ step of the KDD process involves the selection of relevant data from the database.  (3) \_\_\_\_\_\_\_\_\_\_\_\_\_\_ is the method by which we can identify strong and weak association rules.  (4) \_\_\_\_\_\_\_\_\_\_\_\_\_\_ is the technique in which class label is known. (Classification or Clustering)  (5) What do you mean by support and confidence of given itemset? What is the formula for confidence? | 05 |
| Q.2 | Attempt any four (Short Questions) | 12 |
|  | (1) What do you mean by data, information and knowledge. |  |
|  | (2) Enlist and explain in brief, different data pre- processing techniques. |  |
|  | (3) Explain data integration in detail along with its problems and solutions. |  |
|  | (4) What are the major issues encountered in data mining? |  |
|  | (5) Explain in brief the concept of market- basket analysis. What is the use of association rule mining? |  |
| Q.3 | Attempt any two questions | 08 |
|  | (1) Enlist 5 applications in data mining. |  |
|  | (2) Explain in brief the different types of normalization methods used for data transformation.  Suppose that the minimum and maximum values for an attribute income are $12,000 and $98,000, respectively. We would like to map income to range [0,1]. What will be the mapped value for income of $73,600? |  |
|  | (3) List down and explain the disadvantages of data mining. |  |
| Q.4 | (A) Explain in detail with a diagram- Knowledge discovery process from a database. | 05 |
|  | (B) For the given transaction table, find the frequent item sets and association rules using Apriori algorithm. Consider minimum support=2, minimum confidence=50%.   |  |  | | --- | --- | | TID | Itemsets | | T1 | A, B | | T2 | B, D | | T3 | B, C | | T4 | A, B, D | | T5 | A, C | | T6 | B, C | | T7 | A, C | | T8 | A, B, C, E | | T9 | A, B, C | | 05 |
|  | OR |  |
|  | (B) For the given transaction table, generate frequent pattern tree and frequent pattern rules, using FP Growth algorithm. Consider minimum support be 3.   |  |  | | --- | --- | | Transaction ID | Items | | T101 | {E, K, M, N, O, Y} | | T102 | {D, E, K, N, O, Y} | | T103 | {A, E, K, M} | | T104 | {C, K, M, U, Y} | | T105 | {C, E, I, K, O, O} | | 05 |