**20DS504**

**Hall Ticket Number:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **III/IV B.Tech (Regular) DEGREE EXAMINATION** | | | |
| **February,2023** | **Data Science** | | |
| **Fifth Semester** | **Data Warehousing and Data Mining** | | |
| **Time:** Three Hours | | **Maximum: 7**0 Marks | |
| *Answer Question No. 1 Compulsorily.* | | | (14X1 = 14 Marks) |
| *Answer* ***ANY ONE*** *question from each Unit.* | | | (4X14=56 Marks) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. | a) | Define Data Discretization | CO1 | L1 | 1M |
|  | b) | Illustrate the difference between discrete and continuous attribute | CO1 | L2 | 1M |
|  | c) | Infer about Data Visualization | CO1 | L2 | 1M |
|  | d) | Explain snow flake schema | CO2 | L2 | 1M |
|  | e) | Explain slice operation | CO2 | L2 | 1M |
|  | f) | Define Data Warehouse | CO2 | L1 | 1M |
|  | g) | Explain Prune in Apriori algorithm | CO3 | L2 | 1M |
|  | h) | Define Confidence | CO3 | L2 | 1M |
|  | i) | Illustrate the significance of an Association Rule | CO3 | L2 | 1M |
|  | j) | List an advantage of FP Growth algorithm | CO3 | L1 | 1M |
|  | k) | Define Cluster | CO4 | L1 | 1M |
|  | l) | Illustrate how K-Means algorithm gets terminated | CO4 | L2 | 1M |
|  | m) | Illustrate how distance is calculated between two clusters | CO4 | L2 | 1M |
|  | n) | Infer an application for Outlier Analysis | CO4 | L2 | 1M |
| **Unit -I** | | | | | |
| 2. | a) | Define Data Mining. Give various synonyms of Data Mining | CO1 | L1 | 7M |
|  | b) | Illustrate how missing values are handled in data cleaning | CO1 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 3. | a) | Explain in detail about various data normalization techniques | CO1 | L2 | 7M |
|  | b) | Demonstrate all normalization techniques for the data 2,6,8,11,13 | CO1 | L2 | 7M |
|  |  | **Unit -II** |  |  |  |
| 4. | a) | Demonstrate how data warehouse is constructed using FACT tables and Data Cube | CO2 | L2 | 7M |
|  | b) | Explain any two Data Cube operations | CO2 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 5. | a) | Explain the architecture of Data Warehouse | CO2 | L2 | 7M |
|  | b) | Infer about Fact Constellation with example | CO2 | L2 | 7M |
|  |  | **Unit -III** | |  |  |
| 6. | a) | Identify and extract frequent patterns using Apriori algorithm for the following data set: | CO3 | L3 | 7M |
|  | b) | **P T O**  **20DS504**  Identify and extract frequent patterns using FP Growth algorithm for the following data set: | CO3 | L3 | 7M |
|  |  | **(OR)** |  |  |  |
| 7. | a) | Identify and extract frequent patterns using Apriori algorithm for the following data set: | CO3 | L3 | 7M |
|  | b) | Identify and extract frequent patterns using FP Growth algorithm for the following data set: | CO3 | L3 | 7M |
|  |  | **Unit -IV** |  |  |  |
| 8. | a) | Explain types of Clusters with neat diagrams | CO4 | L1 | 7M |
|  | b) | Construct clusters using k-means algorithm and Euclidean distance to cluster the following 5 points into 2 clusters: X1= (0.5, 2.5); X2= (0, 0); X3= (1.5, 1); X4= (5, 1); X5= (6, 2) | CO4 | L3 | 7M |
|  |  | **(OR)** |  |  |  |
| 9. |  | Construct clusters using Agglomerative Hierarchical Clustering to cluster the following points: A(1,1), B(1.5, 1.5), C(5, 5), D(3, 4), E(4, 4) and F(3, 3.5) | CO4 | L3 | 14M |

****