**20DS504**

**Hall Ticket Number:**

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

|  |  |  |  |
| --- | --- | --- | --- |
| **III/IV B.Tech (Regular/Supplementary) DEGREE EXAMINATION** | | | |
| **January, 2024** | **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 ONE question from each Unit.* | | | (4X14=56 Marks) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. | a) | Define Data Mining | CO1 | L1 | 1M |
|  | b) | Explain Data Integration | CO1 | L2 | 1M |
|  | c) | Infer about ordinal attribute with example | CO1 | L2 | 1M |
|  | d) | Explain star schema | CO2 | L2 | 1M |
|  | e) | Explain pivot operation | CO2 | L2 | 1M |
|  | f) | Define Data Warehouse | CO2 | L1 | 1M |
|  | g) | Explain Join in Apriori algorithm | CO3 | L2 | 1M |
|  | h) | Define Support | CO3 | L2 | 1M |
|  | i) | Illustrate the significance of an Association Rule | CO3 | L2 | 1M |
|  | j) | List a disadvantage of Apriori algorithm | CO3 | L1 | 1M |
|  | k) | Define Clustering | CO4 | L1 | 1M |
|  | l) | Illustrate how K-Means algorithm gets terminated | CO4 | L2 | 1M |
|  | m) | Compare Supervised and Unsupervised Learning | CO4 | L2 | 1M |
|  | n) | Infer an application for Outlier Analysis | CO4 | L2 | 1M |
| **Unit -I** | | | | | |
| 2. | a) | Explain the process of Knowledge Discovery of Data | CO1 | L2 | 7M |
|  | b) | Demonstrate the Data Mining functionalities. | CO1 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 3. | a) | Define Binning. Explain various types of binning | CO1 | L1 | 7M |
|  | b) | Demonstrate binning for the following data 2,4,6,8,9,10,11,13,17,20,22 with bin sizes 3 and 4 | CO1 | L2 | 7M |
|  |  | **Unit -II** |  |  |  |
| 4. | a) | Illustrate the differences between OLAP and OLTP | CO2 | L2 | 7M |
|  | b) | Demonstrate how data warehouse is constructed using stars, snowflakes and fact constellations in data cube. | CO2 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 5. | a) | Outline the measures included in a Data Cube with examples | CO2 | L2 | 7M |
|  | b) | Explain how concept hierarchy influences in construction of Data Warehouse | CO2 | L2 | 7M |
|  |  | **Unit -III** | |  |  |
| 6. | a) | Construct an algorithm Apriori for finding frequent patterns | CO3 | L3 | 7M |
|  | b) | Identify and extract frequent patterns using Apriori algorithm for the following data set: | CO3 | L3 | 7M |
| **P.T.O**  **20DS504**  **(OR)** | | | | | |
| 7. | a) | Identify and extract frequent patterns using FP Growth algorithm for the following data set: | CO3 | L3 | 7M |
|  | b) | Compare the Apriori and FP Growth algorithms | CO3 | L2 | 7M |
|  |  | **Unit -IV** |  |  |  |
| 8. | a) | Explain types of clusters | CO4 | L2 | 7M |
|  | b) | Explain the K-Means algorithm with an example. | CO4 | L2 | 7M |
|  |  | **(OR)** |  |  |  |
| 9. | a) | Construct clusters using k-means algorithm and Euclidean distance to cluster the following 8 points into 3 clusters: A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9). | CO4 | L3 | 7M |
|  | b) | Compare Agglomerative and Divisive Hierarchical clustering algorithms | CO4 | L2 | 7M |

****