SRI KRISHNA COLLEGE OF TECHNOLOGY

An Autonomous College, Accredited by NAAC with "A" grade

Coimbatore, Tamil Nadu

Academic Year 2019-2020 (ODD SEM)

B.E. Computer Science & Engineering

CONTINUOUS INTERNAL ASSESSMENT

Class	Course Code	Course Title	Date
III Year CSE	*19CSE14*	*Data Warehousing And Mining*	

Duration: 90 min **Max: 50 Marks**

Course Outcomes:		
CO1	Design a Data warehouse system and perform business analysis with OLAP tools.	**[AP]**
CO2	Understand the basics of Data mining and Knowledge Discovery process.	**[U]**
CO3	Apply suitable pre-processing and visualization techniques for data analysis.	**[AP]**
CO4	Apply frequent pattern and association rule mining techniques for data analysis.	**[AP]**
CO5	Apply appropriate classification and clustering techniques for data analysis.	**[AP]**
CO6	Design a databases using Weka Tool	**[AP]**

Part – A (09 x 02 = 18 Marks) **Answer All Questions**		**BT**	**CO**	**Marks**
1	Identify the use of OLAP cube in a data warehouse.	U	1	2
2	Choose the cube materialization technique which uses partial Materialization in it with explanation. **a) Full Cube** **b) Iceberg Cube** **c) Shell Cube** **d) Closed Cube**	AP	2	2
3	How Data Mining differs from Data Warehouse?	U	1	2
4	For a Supermarket chain consider the following dimensions, namely Product,store,time,promotion. The schema contains a central fact table,sales facts with three measures unit_sales, dollar_sales and dollar_cost. Design star schema for this application.	AP	1	2

<table>
<tr><td>5</td><td>Construct the data cube for the data given below regarding sales statistics in the last year
<table>
<tr><th>Location</th><th>Item</th><th>Sales (in million dollars)</th><th>Count (in thousands)</th></tr>
<tr><td>Asia</td><td>TV</td><td>15</td><td>300</td></tr>
<tr><td>Europe</td><td>TV</td><td>12</td><td>250</td></tr>
<tr><td>North America</td><td>TV</td><td>28</td><td>450</td></tr>
<tr><td>Asia</td><td>Computer</td><td>120</td><td>1000</td></tr>
<tr><td>Europe</td><td>Computer</td><td>150</td><td>1200</td></tr>
<tr><td>North America</td><td>Computer</td><td>200</td><td>1800</td></tr>
</table></td><td>AP</td><td>2</td><td>2</td></tr>
<tr><td>6</td><td>Identify the different types of Binning techniques involved in data cleaning, along with examples.</td><td>U</td><td>1</td><td>2</td></tr>
<tr><td>7</td><td>Suppose the recorded value of “F” ranges from -785 to 736. Normalize by decimal scaling.</td><td>AP</td><td>1</td><td>2</td></tr>
<tr><td>8</td><td>Differentiate between Star and Snowflake Schema.</td><td>U</td><td>2</td><td>2</td></tr>
<tr><td>9</td><td>Why do we need data modelling in a data warehouse?</td><td>U</td><td>2</td><td>2</td></tr>
</table>

Part – B (02 x 16 = 32 Marks) **Answer 2 Questions**		**BT**	**CO**	**Marks**
10	To design an efficient data warehouse, we need to analyze the business needs and also we need to construct a business analysis framework. Each person has different views regarding the design of a data warehouse. Identify the different views involved in the business analysis framework.	U	1	16
11	Patient dataset from a hospital has been taken to Identify whether the patient has heart disease or not. Dataset contains noisy data and some outliers present in it, for that dataset choose any of the suitable data preprocessing tasks and also tell how outliers or noisy data removed from that dataset.	AP	2	16