

THAKUR COLLEGE OF ENGINEERING AND TECHNOLOGY

S. T. SEMESTER IV (CBCGS-HME 2020) MAY 2022

Subject: Data Mining and Business Visualzation

Branch: AI&ML

Q.P. Code: B-46004

Total Marks: 60

(Time: 2 hours)

- Instructions: 1. All Sections are compulsory.
2. Figures to the right indicate full marks.
3. Assume suitable data if necessary and state the assumptions clearly.

Section-I (08 Marks)		
Q.No.	Multiple choice question(Answer any 08 questions out of 10)	Marks
1	The term Information Gain is related to which classification algorithm? A. ID3 B. NN C. KNN D. K-Mean	1
2	Which is the following area is affected by Business Intelligence? A. Revenue B. Sales C. CRM D. Manager	1
3	Find the median of the given set <u>2,6,6,8,4,2,7,9</u> 2, 2, 4, <u>6</u> , 6, 7, 8, 9 A. 2 B. 6 C. 8 D. 4	1
4	What do you mean by support(A)? A. Total number of transactions containing A B. Total Number of transactions not containing A C. Number of transactions containing A / Total number of transactions D. Number of transactions not containing A / Total number of transactions	1
5	Which of the following is use in <u>classification</u> ? A. Taylor series B. Laurent series C. Bayes theorem D. k-mean	1
6	Amongst which of the following step is performed by data scientist after acquiring the data? A. Deletion B. Data Replication C. Data Integration D. Data Cleansing	1

Page - 01

7	<p>Point out the wrong statement.</p> <p><input checked="" type="checkbox"/> A. Data is factual information for analysis</p> <p>B. BI is a category of database software that provides an interface to help users quickly and interactively scrutinize the results in a variety of dimensions of the data</p> <p>C. Customer relationship management (CRM) entails all aspects of interaction that a company has with its customer</p> <p>D. Business intelligence equips enterprises to gain business advantage from data</p>	1
8	<p>What is not require for clustering?</p> <p><input checked="" type="checkbox"/> A. Labeled data</p> <p>B. Unlabeled data</p> <p>C. Numerical data</p> <p>D. Categorical data</p>	1
9	<p>Mode refers to the value within a series that occurs how many number of times.</p> <p><input checked="" type="checkbox"/> A. Maximum</p> <p>B. Minimum</p> <p>C. Zero</p> <p>D. Infinite</p>	1
10	<p>The earliest step in the data mining process is usually?</p> <p>A. Visualization</p> <p><input checked="" type="checkbox"/> B. Preprocessing</p> <p>C. Modelling</p> <p>D. Deployment</p>	1
Section-II (12 Marks)		
	Objective questions(Answer any 4)	
1	Illustrate Major Issues in Data Mining.	3
2	Illustrate data, information and knowledge	3
3	Analyze why data preprocessing is required?	3
4	Illustrate Outlier with their method.	3
5	Point out Apriori advantage and Disadvantage.	3
6	Write Business Intelligence Issues	3
Section-III (20 Marks)		
	Short Answer question(Answer any 04)	
1	List Five number summary and explain with example.	5
2	Explain different types of Attributes with examples?	5
3	Illustrate Data mining and which kind of pattern to be mined?	5
4	Illustrate K-Mean with example.	5
5	Illustrate Agglomerative clustering with example.	5
6	Summarize Confusion Matrix with example.	5

Page 02

Section-IV (20 Marks)																						
	Long Answer question (answer any 2)																					
1	<p>Suppose we have the following dataset that has various transactions, and from this dataset, we need to find the frequent item sets and generate the association rules using the Apriori algorithm: Minimum Support =2, Minimum Confidence =50%</p> <table><thead><tr><th>TID</th><th>ITEMSETS</th></tr></thead><tbody><tr><td>T1</td><td>A, B</td></tr><tr><td>T2</td><td>B, D</td></tr><tr><td>T3</td><td>B, C</td></tr><tr><td>T4</td><td>A, B, D</td></tr><tr><td>T5</td><td>A, C</td></tr><tr><td>T6</td><td>B, C</td></tr><tr><td>T7</td><td>A, C</td></tr><tr><td>T8</td><td>A, B, C, E</td></tr><tr><td>T9</td><td>A, B, C</td></tr></tbody></table> <p>$ABC = 2$ $AB = 50\%$ $BC = 50\%$ $AC = 50\%$</p>	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	10
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																					
2	Discuss Business Intelligence with example .Write the phases in the development of BI system in details.	10																				
3	Summarize classification and its importance? Explain Decision Tree with example.	10																				