Sanjay Ghodawat University, Kolhapur

2020-21

Established as State Private University under Govt. of Maharashtra. Act No XL, 2017

EXM/P/09/00

Year and Program: Third Year, B. Tech	**School of Technology**	**Department: CSE**
Course Code: CST313.3	**Course Title: Program Vertical I: Data Mining & Warehousing**	**Semester : V**
Day: Wednesday **Date: 23/12/2020**	**End Semester Examination**	**Time : 2 Hrs.** **Max Marks: 50**

Instructions: 1) All Questions are compulsory.

			Marks	Bloom's Level	CO
Q.1	**Answer the following (any Two)**–				
	a)	Discuss data cleaning approaches for handling missing values and noisy data.	8	L2	01
	b)	Elaborate on data-warehouse architecture with suitable diagram.	8	L2	03
	c)	Differentiate between multidimensional data model schemas.	8	L2	03

Q.2 **Answer the following (any two)** –

a) Comment on indexing of OLAP data using: 8 L2 01

i. Bitmap indexing
ii. Join indexing

b) Demonstrate working of Naïve Bayes classifier for given labeled dataset. 8 L3 02

No	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Classify test case: (color: Red Type: SUV Origin: Domestic).

c) Use hierarchical clustering technique on given distance matrix to perform single link and complete link clustering. 8 L3 02

	A	**B**	**C**	**D**
A	0	1	4	5
B	1	0	2	6
C	4	2	0	3
D	5	6	3	0

Q.3 Answer the following (any two) –

a) Differentiate between web content mining, web structure mining and web usage mining. 9 L2 03

b) Apply Frequent Pattern (FP) Growth algorithm on given transactional itemset with Support =50%, Confidence= 60% 9 L3 02

Transaction	**List of items**
T1	I1, I2, I3
T2	I2, I3, I4
T3	I4, I5
T4	I1, I2, I4
T5	I1, I2, I3, I5
T6	I1, I2, I3, I4

i. Build FP Tree.
ii. Mine the FP Tree to generate frequent itemset.

c) Use Apriori algorithm to find frequent itemset from given transactional dataset with minimum support of 2 ($Support_{min}$=2) 9 L3 02

Transaction	**Items purchased**
T1	{I1, I3, I4}
T2	{I2, I3, I5}
T3	{I1, I2, I3, I5}
T4	{I2, I5}

..

*****