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: **Date:**	**End Semester Examination**	**Time :2 Hrs.** **Max Marks:50**

Question Paper

Instructions: 1) All Questions are compulsory.

			Marks	Bloom's Level	CO
Q.1	**Answer the following (any two)**–				
	a)	Discuss major tasks involved in data preprocessing with suitable examples.	**8**	L2	01
	b)	With neat diagrams, explain star, snowflake, and fact constellation schemas for multidimensional data modeling in data warehouse.	**8**	L2	03
	c)	Apply *k*-Means algorithm to cluster the data {2, 4, 10, 12, 3, 20, 30, 11, 25} for k = 2.	**8**	L2	03
Q.2	**Answer the following (any two)** –				
	a)	Describe three-tier data warehousing architecture.	**8**	L2	01
	b)	Discuss the working of web crawler.	**8**	L3	02
	c)	Demonstrate working of Naïve Bayes classifier to classify test case ***X*** = (*age = youth*, *income = medium*, *student = yes*, *credit rating = fair*) by using following labeled dataset.	**8**	L3	02

RID	age	income	student _	credit rating	Class: bu
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle aged	medium	no	excellent	yes
13	middle aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

Q.3 Answer the following (any two) –

a) Explain Apriori algorithm with suitable example. **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	I100, I200, I300
T2	I200, I300, I400
T3	I400, I500
T4	I100, I200, I400
T5	I100, I200, I300, I500
T6	I100, I200, I300, I400

Build FP Tree and mine the FP Tree to generate frequent itemset.

c) Differentiate between OLTP and OLAP systems. **9** L3 02

*****