Seat No.:		: Enrolment No		
		GUJARAT TECHNOLOGICAL UNIVERSITY ME – SEMESTER II (OLD) – • EXAMINATION – SUMMER 2016		
			/05/2016	
Ti	me: 1 truction 1.	t Name: Data Mining & Data Warehousing 0:30 am to 01:00 pm Total Marks: ons: . Attempt all questions Make suitable assumptions wherever necessary Figures to the right indicate full marks.	70	
Q.1	(a) (b)	<ol> <li>Discuss various methods to handle missing values in data mining.</li> <li>Differentiate OLAP and OLTP</li> <li>Explain various data repositories on which data mining techniques are applied.</li> </ol>	04 03 07	
Q.2	(a)	What are the key differences in snowflake and star schema? Where should they be applied?  Draw and explain the three-tier data warehousing architecture.	07 07	
	<b>(b)</b>	OR	U/	
	<b>(b)</b>	Describe multilayer feed-forward neural network.	07	
Q.3	(a)	Why is outlier mining important? Briefly describe the different approaches behind statistical-based outlier detection, distance-based outlier detection and deviation-based outlier detection.	07	
	(b)	Describe the difference between the following approaches for the integration of a data mining system with a database or data warehouse system: no coupling, loose coupling, semi tight coupling and tight coupling. State which approach you think is most popular and why?	07	
0.2		OR	0=	
Q.3	(a)	A list contains attribute Age values as {34, 23, 56, 67, 89, 19, 56, 45, 12, 76, 23, 37}. Normalize value for age=47 using min-max [0, 1], z-score and decimal normalization methods. Which one you think to be more accurate?	07	
	<b>(b)</b>	Explain K-Mean clustering algorithm and state its limitations.	<b>07</b>	
Q.4	(a)	What is association rule mining in KDD? Explain multi-level association rule mining with suitable example.	07	
	<b>(b)</b>	Explain InfoGain and Gini Index attribute selection measures with example.	<b>07</b>	

OR

Explain how data mining is useful in detecting fraud in online financial

OR

(b) Why naïve Bayesian classification is called "naïve"? State and explain the

\*\*\*\*\*\*\*

(b) Differentiate between partitioning and hierarchical methods for clustering.

(a) What is density-based method in clustering? Explain DBSCAN in detail.

Explain Agglomerative Hierarchical clustering algorithm.

(b) Explain common web log format? Discuss web usage mining.

(a) Explain any two methods to improve efficiency of Apriori algorithm.

Bayesian classification.

transaction.

**Q.4** 

**Q.5** 

Q.5

**07** 

**07** 

**07** 

**07** 

**07** 

**07**