

Scheme	R2012
Semester	8
Course Code	CPC801
Course Name	Data Warehousing and Mining

Question No.	Question	a	b	c	d	Answer Key
1	A data warehouse can be used to analyze a particular _____	graph	chart	domain	subject	subject
2	which information is not provided by Information packages	Define the common subject areas	date of full refresh	Establish data granularity	Estimate data warehouse size	date of full refresh
3	Periodic Status is	data in which changes to existing records cause the previous version of the records to be eliminated	the value of the attribute is preserved as the status every time a change occurs	data that are never altered or deleted once they have been added	the value of the attribute at this moment of time.	the value of the attribute is preserved as the status every time a change occurs
4	Comparison of the general features of the target class data object against the general features of objects from one or multiple contrasting classes is a process of	Data Characterization	Data Classification	Data discrimination	Data selection	Data discrimination
5	After the initial load, the data warehouse is kept up-to-date by two actions: REFRESH and UPDATE. As the number of records increase in a Data Warehouse, cost of update operation _____ .	decreases	increases	remains constant	is same as cost of Refresh	increases
6	The values of an _____ attribute provide enough information to order objects.	ratio	Binary	Interval	ordinal	ordinal
7	As per the concept of KDD process, which of the following statement is valid ?	KDD and Data Mining have no connection at all	KDD is one of the steps in Data Mining	Data Mining is one of the steps in KDD process	KDD and Data Mining mean the same	Data Mining is one of the steps in KDD process
8	information stored in the data warehouse.	additive atleast over one dimension	Only numeric measures are used	All possible summaries are used	It is additive over every dimension of its dimensionality	It is additive over every dimension of its dimensionality
9	Converting data from different sources into a common format for processing is called as _____.	Selection	Preprocessing	Transformation	Interpretation	Transformation

10	Binary attribute are	This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit	This takes only three values.	This takes only four values.	It cant take any value.	This takes only two values. In general, these values will be 0 and 1 and they can be coded as one bit
11	It is measured on a scale of equal size units,these attributes allows us to compare such as temperature in C or F and thus values of attributes have order.	Interval Scaled attribute	Ratio scaled attribute	Binary attribute	Ternary attribute	Interval Scaled attribute
12	Which of the following is not a valid Visualization technique ?	Scatter plot	Decision Tree	Box plot	Histogram	Decision Tree
13	The _____numerical measure which tells that two objects are alike	dissimilarity	clarity	non clarity	similarity	similarity
14	Removing duplicate records is a data mining process called _____ .	Data isolation	Recovery	Data Cleaning	Data dredging	Data Cleaning
15	_____ is a process of taking operational data from one or more sources and mapping it, field by field, onto a new data structure in the data warehouse	Transformation	Cleansing	Integration	Scrubbing	Integration
16	_____ may be defined as the data objects that do not comply with the general behavior or model of the data available.	Evolution Analysis	Prediction	Classification	Outlier Analysis	Outlier Analysis
17	How many coefficients do you need to estimate in a simple linear regression model (One independent and one dependent variable)?	1	2	3	4	2
18	The mapping or classification of a class with some predefined group or class is known as?	Data Characterization	Data Discrimination	Data Subset	Data set	Data Characterization
19	To extract rules in supervised learning _____ is used	root node	sibling	decision trees	branches	decision trees
20	from the given options_____ is a predictive model	Clustering	Regression	Summarization	Association rules	Regression
21	Euclidean distance measure is	A stage of the KDD process in which new data is added to the existing selection.	The process of finding a solution for a problem simply by enumerating all possible solutions according to some pre-defined order and then testing them	The distance between two points as calculated using the Pythagoras theorem	The distance between two points as calculated using interval scale	The distance between two points as calculated using the Pythagoras theorem
22	Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8):Compute the Euclidean distance between the two objects.	6.32	6.71	6.15	6.22	6.71

23	The following rule is an example of which association rule. { age (X, "20.....29") ^ occupation(X, "student") → buys(X, "laptop") }.	multilevel association rules	interlevel association rules	multidimensional association rules	intralevel association rules	multidimensional association rules
24	Repeating the holdout many times is called _____	random subsampling	cross validation	bootstrap	bagging	random subsampling
25	Which algorithm requires fewer scans of data?	Apriori	FP growth	Apriori and FP Growth	decision	FP growth