DMA UNIT-1 QB - DMA QB

Data Mining And Analytics (SRM Institute of Science and Technology)

UNIT 1			
Why Data mining? What is Data mining? Kinds of data meant for mining, Kinds of patterns that can be mined, Applications suitable for data mining, Issues in Data mining, Data objects and Attribute types, Statistical descriptions of data, Need for data preprocessing and data Quality, Data cleaning, Data integration, Data reduction, Data transformation, Data cube and its usage.			
PART-A (Multiple Choice Questions)			
Q. No	**Questions**	**Course Outcome**	**Competence BT Level**
1	*An essential process used for applying intelligent methods to extract the data patterns is named as_____-* **a) data mining** *b) data analysis* *c) data implementation* *d) data computation*	*CO1*	1
2	_____ graphs the quantiles of one univariate distribution against the corresponding quantiles of another. A) Quantile plot **B) Quantile- Quantile plot** C) Histograms D) Scatter plots	*CO1*	2
3	*A _____ is a nominal attribute with only two categories or states.* *A) Discrete Attribute B) Nominal Attributes* **C) Binary attributes** *D) Continuous Attributes*	*CO1*	1
4	*_____ is the process, where the raw values of a numeric attribute are replaced by interval labels or conceptual labels.* *A) Smoothing B)Aggregation C)Normalization* *D)* **Discretization**	*CO1*	2
5	*Which of the following is not a data preprocessing technique?* *A)* Data cleaning *B)* Data integration **C) Data discrimination** *D)* Data Transformation	*CO1*	1
6	*_____ characterize properties of the data in a target data set.* **A) Descriptive mining task** *B) Predictive mining task* *C) Objective mining task D) Clustering mining task*	*CO1*	3
7	*Which of the following techniques combines multiple sources to form a coherent data store?* *A)* Data cleaning **B) Data integration** *C)* Data reduction *D)* Data Transformation	*CO1*	1
8	*Which of the following techniques obtain a reduced representation of the data while minimizing the loss of information content?* *A)* Data cleaning *B)* Data integration **C) Data reduction** *D)Data Transformation*	*CO1*	1
9	*Which of the following is not a data transformation strategy?* *A) Data discretization* **B) Data discrimination**	*CO1*	1

	C) Normalization D)Attribute construction		
10	*The process of removing the deficiencies and loop holes in the data is called as* *(a) Aggregation of data* *(b) Extracting of data* **(c) Cleaning up of data.** *(d) Loading of data*	*CO1*	*1*
11	*Which of the following process includes data cleaning, data integration, data selection, data transformation, data mining, pattern evolution and knowledge presentation?* **(a) KDD process** *(b) ETL process* *(c) KTL process* *(d) MDX process*	*CO1*	*1*
12	*8.Which of the following is general characteristics or features of a target class of data* *(a) Data selection* *(b) Data discrimination* *(c) Data Classification* **(d) Data Characterization**	*CO1*	*1*
13	*_____ performs a linear transformation on the original data.* **A) Min-max normalization** *B) z-score normalization* *C) Decimal Scaling* *D) Discretization*	*CO1*	*1*
14	*The mapping or classification of a class with some predefined group or class is known as?* *a. data characterization* **b. data discrimination** *c. data set* *d. data sub structure*	*CO1*	*2*
15	*The analysis performed to uncover interesting statistical correlations between associated-attribute-value pairs is called?* *a. Mining of association* *b. Mining of clusters* **c. Mining of correlations** *d. None of the above*	*CO1*	*2*
16	*_____ may be defined as the data objects that do not comply with the general behavior or model of the data available.* **a. Outlier Analysis** *b. Evolution Analysis* *c. Prediction* *d. Classification*	*CO1*	*1*
17	"Efficiency and scalability of data mining algorithms" issues comes under? a. Mining Methodology and User Interaction Issues **b. Performance Issues**	*CO1*	*1*

	c. Diverse Data Types Issues d. None of the above		
18	_____ *is not a data mining functionality?* *(a) Clustering and Analysis* **(b)Selection and interpretation** *(c) Classification and regression* *(d)Characterization and Discrimination*	*CO1*	*1*
19	_____ *is the process of finding a model that describes and distinguishes data classes or concepts.* *(a) Regression* *(b) Clustering* *(c)* **Classification** *(d) Characterization*	*CO1*	*1*
20	*A _____ is a numeric attribute with an inherent zero point.* *A) Discrete Attributes* **B) Ratio-scaled Attributes** *C) Interval-scaled Attributes D) Continuous Attributes*	*CO1*	*1*
21	*Which of the following is the collection of data objects that are similar to one another within the same group? (A) Partitioning (B) Grid* **(C) Cluster** *(D) Table*	*CO1*	*1*
22	*In _____, the values for an attribute A, are normalized based on the mean and standard deviation of A.* *A) Min-max normalization* **B) z-score normalization** *C) Decimal Scaling* *D) Discretization*	*CO1*	*1*
23	_____ *are measured on a scale of equal-size units.* *A) Discrete Attributes B) Ratio-scaled Attributes* **C) Interval-scaled Attributes** *D) Continuous Attributes*	*CO1*	*1*
24	_____ *are points taken at regular intervals of a data distribution, dividing it into equal size consecutive sets.* *A) Interquartile range* **B) Quantiles** *B) Percentile* *C) Range*	*CO1*	*1*
25	The distance between the third quartile and the first quartile is called as _____. *A) Quantile B) Percentile C) Range* **D) Interquartile range**	*CO1*	*1*
26	*The _____is the data point dividing the lower and upper halves of the data distribution* *A)8-quantiles* *B) 100-quantiles* **C) 2-quantiles** *D) 4-quantiles*	*CO1*	*1*
27	*A_____ has a finite or countably infinite set of values, which may or may not be represented as integers.* **A) Discrete Attribute** *B) Ratio-scaled Attributes* *C) Interval-scaled Attributes D) Continuous Attributes*	*CO1*	*1*
28	*Which method reduces the number of random variables or*	*CO1*	*1*

	attributes under consideration?_____. *A)* **Dimensionality reduction** *B) Numerosity reduction* *C) Data Tranformation* *D) Data Compression*		
	PART B (5 Marks)		
1	*What is Data Integration? Explain in detail.*	*CO1*	*BT*
2	*Describe three challenges to data mining regarding data mining methodology and user interaction issues.*	*CO1*	*2*
3	*Demonstrate any 2 data mining applications.*	*CO1*	*3*
4	*Describe the following with examples.* (a) Ordinal Attributes and Numeric attributes. (b) Discrete versus Continuous Attributes	*CO1*	*2*
5	*Explain in detail about Statistical Descriptions of Data.*	*CO1*	*2*
6	*Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.* (a) Calculate the mean of the data? (b) Calculate the median? (c) Calculate the mode of the data? (d) Calculate the midrange of the data?	*CO1*	*3*
7	*Describe Binary and Nominal attributes. Give example.*	*CO1*	*2*
8	*Suppose that the data for analysis includes the attribute salary. The data values for the data tuples are (in increasing order) 30, 36, 47, 50, 52, 52, 56, 60, 63, 70, 70, 110.* (e) Calculate the mean of the data? (f) Calculate the median? (g) Calculate the mode of the data? *()Calculate the midrange of the data?*	*CO1*	3
9	*What is Data Cleaning? Demonstrate various methods of Data Cleaning.*	*CO1*	3
	PART C (12 Marks)		
1	*What is KDD? Explain about data mining as a step in the process of knowledge discovery.*	*CO1*	*1*
2	*What kinds of patterns can be mined? Explain with an example.*	*CO1*	*2*
3	*Describe various data reduction technology with suitable examples.*	*CO1*	*1*
4	*Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.* *(a) What is the mean of the data? What is the median?* *(b) What is the mode of the data? Comment on the data's modality (i.e., bimodal, trimodal, etc.).* *(c) What is the midrange of the data?* *Write short notes on following* *(d) What is Quartile and Quantile?*	*CO1*	*3*

	(e) What is five-number summary? *(f) Show how boxplot is drawn.* *(g) How is a quantile–quantile plot different from a quantile plot?*		
5	*Describe various data transformation technology with suitable examples.*	*CO1*	*2*