

Program: B.E. Information Technology Engineering

Curriculum Scheme: Rev. 2016

Examination: Third Year - Semester VI

Course Code: ITC602

Course Name: Data Mining and Business Intelligence

Time: 1 hour

Max. Marks: 50

Note to the students: All the Questions are compulsory and carry equal marks.

Question No.	Question Statement	Options			
		A:	B:	C:	D:
1	Which data mining functionality describes and distinguishes classes for future prediction?	Association	Clustering	Outlier Analysis	Classification
2	For data D = {54, 59, 61, 61, 62, 63} find mean, mode and median?	54,61,63	60,61,61	61,61,62	61,61,61
3	What is popular way of visualizing a five-number summary of a distribution.	Star schema	ER diagram	Boxplot	Decision Tree
4	Friendship structure of users in a social networking site can be considered as an example of _____.	Record data	Ordered data	Graph data	Hierarchical data
5	Suppose that the minimum and maximum values for the feature income are 13 and 70 respectively. We would like to map income to the range 0.0,1.0. By min-max normalization, a value of 35 for income is transformed to _____.	0.386	1.225	0	1
6	Attribute Color with values {gray, pink, yellow, blue} is an example of: _____.	Nominal attribute	Continuous attribute	Ordinal attribute	Numeric attribute
7	A quiz question had names of 6 algorithms of which the student had to select only the classification algorithms. A student identified 3 of them as classification algorithms. During evaluation it was found that 2 of the algorithms identified by the student were indeed classification algorithms. The student was unable to identify 2 other classification algorithms in the list. The accuracy is: _____.	0.22	0.44	1	0.5
8	Assume there are 3 classes and 50% of the examples belong to class1, and 25% of the examples belong to class2 and class3, respectively. Compute the entropy of this class distribution.	1	1.5	0.5	0.75
9	Recall is defined as:	$(TP+TN)/(P+N)$	$TP/(TP+FP)$	$(FP+FN)/(P+N)$	TP/P
10	The information gain measure is biased toward _____	Tests with zero outcomes	Tests with one outcome	Tests with two outcomes	Tests with many outcomes
11	In order to evaluate the performance of a classifier, the given data is randomly partitioned into two independent sets wherein 2/3rd is used for model construction whereas 1/3rd is used for accuracy estimation. What is this evaluation method known as?	Holdout	Cross-Validation	Bootstrapping	Dissection
12	A good clustering is one having _____.	High inter-cluster distance and low intra-cluster distance	Low inter-cluster distance and low intra-cluster distance	Low inter-cluster distance and high intra-cluster distance	High inter-cluster distance and high intra-cluster distance
13	_____ is an exploratory data mining technique.	Classification	Clustering	Prediction	Regression

14	Which of the following is not an application of the outlier?	Security and surveillance	Fake news identification	Detection of fraud in financial data	Predicting prices of house
15	Which of the following is not true about the DBSCAN algorithm?	The number of clusters need to be specified in advance	It requires two parameters MinPts and epsilon	It can produce non-convex shaped clusters	It is a density based clustering algorithm
16	Agglomerative clustering works by _____.	Using the mean of the objects and then adding them into clusters hierarchically	Using Bottom-up approach; goes adding single elements into clusters and then process is repeated until we get bigger cluster	Uses median while adding the points into the clusters, similar like k-median	Dividing the data into two parts and then does the clustering in the similar way K-means does
17	How to calculate confidence (A-->B)?	$P(B A)$	$P(A \cap B)$	$P(A \cup B)$	$P(A B)$
18	Which of the following is the property of Apriori algorithm?	All itemsets must be frequent	All nonempty subsets of a frequent itemset must also be frequent	It mines all frequent patterns by constructing a FP tree	Based on the kinds of rules
19	Which of the following is not used in pruning strategies?	Item merging	Sub-itemset pruning	Item skipping	partitioning
20	The second phase of Apriori algorithm is _____.	Itemset eliminations	Pruning	Itemset generation	Partitioning
21	Information is transformed into _____ and then it is to be used for making decisions and develop the corresponding actions.	knowledge	data	report	file
22	Decisions are _____ when they affect the entire organization or at least a substantial part of it for a long period of time.	operational	technical	tactical	strategic
23	A/An _____ mathematical model is a material representation of a real system, whose behavior is imitated for the purpose of the analysis.	analogical	iconic	symbolic	stochastic
24	To develop business intelligence sytem for a credit card company that helps in identifying fraudulet transactions from a dataset having genuine as well as fraudulent transactions, which data mining task should be applied?	Association	Classification	Clustering	Feature Extraction
25	For developing a BI system that segments customers to better assign marketing campaigns using customer characteristics, which is the best suited data mining task?	Association	Classification	Clustering	Feature Extraction