

Reg. No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**B.Tech. DEGREE EXAMINATION, NOVEMBER 2018**  
3<sup>rd</sup> to 7<sup>th</sup> Semester

**15CS331E – DATA MINING AND ANALYTICS**  
(For the candidates admitted during the academic year 2015-2016 to 2017-2018)

**Note:**

- (i) **Part - A** should be answered in OMR sheet within first 45 minutes and OMR sheet should be handed over to hall invigilator at the end of 45<sup>th</sup> minute.
- (ii) **Part - B** and **Part - C** should be answered in answer booklet.

Time: Three Hours

Max. Marks: 100

**PART – A (20 × 1 = 20 Marks)**  
Answer ALL Questions

1. A data mining query is defined in terms of data mining \_\_\_\_\_.  
(A) Query language (B) SQL query  
(C) Task primitives (D) Query primitives
2. Many applications involve the generation and analysis of a new kind of data called \_\_\_\_\_.  
(A) Data repository (B) Data flow  
(C) Data storage (D) Stream data
3. \_\_\_\_\_ is a statistical methodology that is used for numeric prediction.  
(A) Regression analysis (B) Covariance  
(C) Correlation (D) Standard deviation
4. \_\_\_\_\_ is the process of finding a model that describes and distinguishes data classes.  
(A) Association rule mining (B) Data discretization  
(C) Clustering (D) Classification
5. \_\_\_\_\_ obtains a reduced representation of the data set.  
(A) Data cleaning (B) Data transformation  
(C) Data reduction (D) Data integration
6. A \_\_\_\_\_ is a measure that must be computed on the entire data set as a whole.  
(A) Algebraic measure (B) Holistic measure  
(C) Interquartile range (D) Distributive measure
7. \_\_\_\_\_ reduces the data set size by removing irrelevant or redundant attributes.  
(A) Attribute subset selection (B) Data cube aggregation  
(C) Dimensionality reduction (D) Numerosity reduction
8. \_\_\_\_\_ mining searches for frequent sub-structures in a structured data set  
(A) Structured pattern (B) Sequential pattern  
(C) Association rule (D) Frequent itemset
9. \_\_\_\_\_ is the learning of decision trees from class-labeled training tuples.  
(A) Laplacian correction (B) Belief network  
(C) Probabilistic network (D) Decision tree induction

10. \_\_\_\_\_ is an extension of straight line regression so as to involve more than one predictor variable.  
 (A) Linear regression (B) Non-linear regression  
 (C) Multiple linear regression (D) Predictor regression
11. \_\_\_\_\_ algorithm attempt to improve accuracy by removing tree branches reflecting noise in the data.  
 (A) Apriori (B) Pruning  
 (C) Bootstrap (D) ID5
12. \_\_\_\_\_ uses a set of if-then rules for classification.  
 (A) Genetic algorithm (B) Rough set theory  
 (C) Rule based classifier (D) Regression models
13. \_\_\_\_\_ show interesting relationships between attribute-value pairs that occurs frequently in a given  
 (A) Association rules (B) Accusative classification  
 (C) Sequential patterns (D) Frequent patterns
14. \_\_\_\_\_ methods can divide a set of objects into multiple exclusive cluster.  
 (A) Density-based (B) Hierarchical  
 (C) Agglomerative (D) Grid based
15. The process of grouping a set of physical or abstract objects into classes of similar objects is called as \_\_\_\_\_.  
 (A) Prediction (B) Association  
 (C) Correlation (D) Clustering
16. \_\_\_\_\_ variable is generalization of the binary variable in that it can take on more than two states.  
 (A) Ratio-scaled (B) Ordinal  
 (C) Categorical (D) Discrete
17. \_\_\_\_\_ is a transformative computing that involves delivering applications and services over the internet.  
 (A) Cloud computing (B) Green computing  
 (C) Bio-inspired computing (D) Pervasive computing
18. \_\_\_\_\_ a hierarchal clustering algorithm that uses dynamic modeling to determine the similarity between the pairs of clusters.  
 (A) Machine learning (B) Chameton  
 (C) Performance computing (D) Data processing.
19. The \_\_\_\_\_ cloud infrastructure is operated for the exclusive use of organization.  
 (A) Public (B) Private  
 (C) Hybrid (D) Public and hybrid
20. Which of the following data mining application is used to allow the retailer to understand the purchase behavior of a buyer?  
 (A) Manufacturing engineering (B) Fraud detection  
 (C) Corporate surveillance (D) Market basket analysis

**PART – B (5 × 4 = 20 Marks)**  
 Answer ANY FIVE Questions

21. Define 'Data cleaning'.  
 22. What is outlier analysis?  
 23. Define frequent patterns.  
 24. What is 'Lift'?  
 25. Explain in brief about information gain.  
 26. What do you mean by ratio-scaled variables?  
 27. Explain 'Web mining'.

**PART – C (5 × 12 = 60 Marks)**  
 Answer ALL Questions

28. a. Explain the steps involved in knowledge discovery process with diagram.

(OR)

- b. Describe the architecture of a data mining system with diagram.

29. a. How can frequent patterns be classified in different ways? Explain.

(OR)

- b. Explain Apriori algorithm to find frequent item sets using candidate generation.

30. a. Explain the basic steps involved in decision tree algorithm.

(OR)

- b. Discuss SVM classification method.

31. a. Write about hierarchical clustering methods.

(OR)

- b. Discuss any two approaches for outlier detection.

32. a. Analyze how data mining can be used to improve telecommunication services.

(OR)

- b. How the machine learning used in solving emerging scientific applications? Discuss the way cloud computing is used in solving emerging scientific applications.

\* \* \* \* \*