Reg. No.														

B.Tech. DEGREE EXAMINATION, DECEMBER 2019
First to Eighth 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 45th 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. Data mining is defined as the process of
   (A) Deducing relationships in data (B) Simulating trends in data
   (C) Representing data (D) Identifying patterns in data

2. _____ is used to organize attributes or attribute values into different levels of abstraction.
   (A) Concept hierarchy (B) Decision tree
   (C) Subsequence (D) Frequent item set

3. The data ware house is _____
   (A) Read only (B) Write only
   (C) Read write only (D) Rewrite only

4. _____ is a summarization of the general characteristics or features of a target class of data
   (A) Data characterization (B) Data classification
   (C) Data discrimination (D) Data selection

5. The output of KDD is
   (A) Data (B) Information security
   (C) Query (D) Useful information

6. FP growth adopts _____strategy.
   (A) Pattern fragment (B) Apriori
   (C) Candidate generation (D) Divide and conquer

7. The basic idea of the apriori algorithm is to generate _____item sets of a particular size and scans the database.
   (A) Candidate (B) Primary
   (C) Secondary (D) Super key

8. _____and _____may be viewed as types of classification
   (A) Decision, verification (B) Estimation, prediction
   (C) Illustration, decision (D) Identification, clarification

9. Clustering is an example for
   (A) Supervised learning (B) Outlier
   (C) Unsupervised learning (D) Distance learning

10. _____ can be used to identify whether any two given attributes are statistically related
   (A) Relevance analysis (B) Regression analysis
   (C) Attribute subset selection (D) Correlation

11. Manhattan distance is also called as
   (A) City block distance (B) Euclidean distance
   (C) Minkowski distance (D) Similar distance

12. Which is not a characteristic of big data?
   (A) Volume (B) Variety
   (C) Visibility (D) Velocity

13. Learning from past experiences is
   (A) Machine learning (B) Improved learning
   (C) Predictive learning (D) Machine and improved learning

14. Pick out a K-method algorithm.
   (A) DBSCAN (B) Birch
   (C) Pam (D) Cure

15. The _____cloud infrastructure is operated for the exclusive use of organization.
   (A) Public (B) Hybrid
   (C) Private (D) Public and hybrid

16. The complete application running on someone else's system is
   (A) PAAS (B) SAAS
   (C) IAAS (D) CAAS

17. _____method is used to find the clusters of arbitrary shape
   (A) Grid-based (B) Partition-based
   (C) Density-based (D) Hierarchical methods

18. _____ predicts categorical labels
   (A) Prediction (B) Back propagation
   (C) Classification (D) Data trends

19. _____ is a statistical methodology that is most often used for numeric prediction.
   (A) Regression analysis (B) Classification
   (C) Class labels analysis (D) Decision tree classifiers

20. _____ refers to the ability to construct the classifier or predictor efficiently given large amounts of data.
   (A) Robustness (B) Scalability
   (C) Speed (D) Interpretability

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

21. Define data mining. Is the word "Data mining" a misnomer? Why?
22. Summarize any four techniques used for data reduction.
23. Compare supervised and unsupervised learning.
24. List the various efficient and scalable frequent item set mining methods.
25. What is dendrogram? How are the clusters merged?
26. When do you call the data as "Big"?
27. Define machine learning and its applications.

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

28. a. Outline the process of knowledge discovery from databases.

(OR)

b. Explain about the data preprocessing techniques with examples.

29. a. Illustrate the method for generating association rules from frequent item sets.

(OR)

b. Discuss the Apriori algorithm in detail.

30. a. Explain how linearly separable data is handled by SVM classification technique.

(OR)

b. Discuss the procedure of decision tree induction technique used for classification with an example.

31. a. Explain K-means algorithm. Illustrate the strength and weakness of K-means in comparison with K-mediods algorithm.

(OR)

b. Illustrate the DBSCAN algorithm with example.

32. a. Demonstrate in detail the application of data mining.

(OR)

b. Explain in detail about the types, services and benefits of cloud computing.

* * * * *