Avinashilingam Institute for Home Science and Higher Education for Women

(Deemed to be University under Category 'A' by MHRD, Estd. u/s 3 of UGC Act 1956)
Re-accredited with 'A+' Grade by NAAC. Recognised by UGC Under Section 12B
Coimbatore - 641 043, Tamil Nadu, India

Bachelor's Degree Examination – August 2020
VI Semester

Class : III UG **Time : 2 Hours**
Major : Information Technology **Max. Marks : 50**

15BITC29 Data Mining

Part - A **10 x 1 = 10**
Choose the Correct Answer

1. Which is not a data mining functionality?
   a. Clustering and Analysis b. Selection and interpretation
   c. Classification and regression d. Characterization and Discrimination

2. Which of the following is required by K-means clustering?
   a. Defined distance metric b. Number of clusters
   c. Initial guess as to cluster centroids d. All of the mentioned

3. Which of the following are general characteristics or features of a target class of data?
   a. Data selection b. Data discrimination
   c. Data Classification d. Data Characterization

4. In KDD and data mining, noise is referred to as
   a. repeated data b. random errors in database
   c. meta data d. complex data

5. OLAP is used to explore the _____ knowledge.
   a. shallow b. deep
   c. multidimensional d. hidden

6. In which approach data warehouse is build first and all information needed is selected
   a. top-down b. client/server
   c. bottom-up d. DSS

7. The set of attribute in a database that refers to data in another table is called
   a. primary key b. candidate key
   c. foreign key d. super key

8. Strategic value of data mining is
   a. cost-sensitive b. work-sensitive
   c. time-sensitive d. technical-sensitive

9. The average positive difference between computed and desired outcome values
   a. mean positive error b. mean squared error
   c. mean absolute error d. root mean squared error

10. Simple regression assumes a _____ relationship between the input attribute and output attribute.
   a. linear b. inverse
   c. quadratic d. reciprocal

Part B 3 x 6 = 18

Answer any **Three** questions

Each answer should not exceed 400 words or two pages

11. How to integrate data mining system with a database or data warehouse?

12. List the differences between operational database system and data ware house.

13. Describe the different steps involved in data transformation.

14. Illustrate K-means clustering algorithm.

15. What is outlier? Give an application for outlier analysis.

16. Describe the role of attribute selection in association rule mining.

17. Consider a student database and explain each step involved in KDD process with a neat diagram. In your answer emphasise that data mining is not equivalent to KDD.

18. What are principal components? How they are useful in dimensionality reduction?

19. Discuss various types of data to be handled in Cluster Analysis.

20. Deliberate the process involved in mining the web page layout structure .

Part C 2 x 11 = 22

Answer any **Two** questions

Each answer should not exceed 800 words or four pages

21. With a neat pictorial representation, illustrate the three-tier architecture of data warehousing model.

22. List the distinct features of OLTP and OLAP. Discuss the types of OLAP servers with example.

23. How to merge the data from multiple data stores in data mining? Elucidate the method used for merging the data.

24. Discuss in detail the process involved in market based analysis and its impact on businesses.

25. Apply Apriori algorithm to the following data set and find the frequent patterns and generate association rules. Use 50% as minimal support value. Minimum confidence value is 60%.

Transaction ID	Items Purchased
01	2 3
02	1 3 5
03	1 2 4
04	2 3

26. Discuss the features and functionalities of centroid based techniques.

27. Write the pseudo code of Pincer-search algorithm for frequent itemset mining and explain with an example.

28. Discuss the Hierarchical based partitioning algorithm with an example.

29. Explain about Web usage mining.

30. Illustrate with suitable examples the procedure to identify authoritative web pages by analysing link structures.

****