LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **COMPUTER SCIENCE**

FIRST SEMESTER – **NOVEMBER 2019**

PCS 1501 – DATA MINING

Date: 30-10-2019   Dept. No.   Max. : 100 Marks
Time: 01:00-04:00

Part – A

Answer ALL the questions   **10 x 2 = 20 marks**

1. State any four CRAN Task Views related to Data Mining.

2. Mention the function used for obtaining dimensionality and structure of data.

3. Differentiate between predictors and response.

4. What is the use of Logistic Regression?

5. Why CLARA is better than PAM?

6. Mention the two key parameters of DBSCAN.

7. What is Time Series Decomposition?

8. State any four measures of distance or dissimilarity.

9. Give formulae to determine support and confidence.

10. Stress the need for "Pruning" in Association Rule Mining.

Part – B

Answer ALL the questions   **5x 8 = 40 marks**

11. a) Explain the features available in R for the basic statistics and charts **(Or)**

    b) Describehow to check the relationship between two variables with an example.

12. a) Explain how to build a Decision Treewith an example**(Or)**

    b) Illustrate the steps on building a Linear Regression Model for CPI data.

13. a)How to cluster the Items {2,4,10,12,3.20.30,11,25} using K-Mean Clustering algorithm when K=2, M1 = 3 and M2 =18 and Explain**(Or)**

    b)Explain about Outlier detection using Local Outlier Factor algorithm.

14. a) Explain Time Series Decomposition with an example **(Or)**

    b)Illustrate how to extract features through Time Series Classification with an example.

15. a) Explain Apriori algorithm for the western sales transactions t1={Bread, Jelly,

   PeanutButter}, t2={Bread, PeanutButter} t3={Bread, Milk, PeanutButter}, t4= {Beer, Bread} and

   t5= {Beer, Milk}   **(Or)**

   b)Illustrate the steps for visualizing Association Rules on Titanic Dataset and explain.

Part – C

Answer ANY TWO questions   **2 x 20 = 40 marks**

16. a) Explain how to import data from .CSV files, Excel files and ODBC databases into R.

   b)Illustrate the steps involved in building a predictive model for iris data using Random Forest.

17. a) Describe about Hierarchical clustering with an example.

   b)Differentiate between Time Series Classification and Time Series Clustering.

18. a)Explain about titanic dataset and how association rules mining helps to identify survival rate.

   b)Illustratehow a boxplot can be used in Univariate outlier detection approach with an example.

****