B.Sc. (Data Science)
Discipline Specific Course (DSC)
Semester III
BSDB32302P: Data Mining and Visualization Lab

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 2
Pass Percentage: 40%

Besides below given practical, List of Lab Assignments – Session wise will be given to student

1. Explore WEKA Data Mining/Machine Learning Toolkit

   1. Downloading and/or installation of WEKA data mining toolkit,
   2. Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface,
   3. Experimenter, command-line interface.
   4 Navigate the options available in the WEKA
   5. (ex. Select attributes panel, Preprocess panel, Classify panel, Cluster panel, Associate panel and Visualize panel)
   6. Study the arff file format
   7. Explore the available data sets in WEKA.
   8. Load a data set (ex. Weather dataset, Iris dataset, etc.)
   9. Load each dataset and observe the following:

   1. List the attribute names and they types

   2. Number of records in each dataset

   3. Identify the class attribute (if any)

   4. Plot Histogram

   5. Determine the number of records for each class.

   6. Visualize the data in various dimensions

2. Perform data pre-processing tasks and Demonstrate performing association rule mining on data sets.
3. Demonstrate performing classification on data sets

   1. Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier
   2. output. Compute entropy values, Kappa statistic.
   3. Extract if-then rules from the decision tree generated by the classifier, Observe the confusion matrix.
   4. Load each dataset into Weka and perform Naïve-bayes classification and k-Nearest Neighbour classification. Interpret the results obtained.
   5. Plot RoC CurvesCompare classification results of ID3, J48, Naïve-Bayes and k-NN classifiers for each dataset, and deduce which classifier is performing best and poor for each dataset and justify.

4. Demonstrate performing clustering of data sets

5. Load each dataset into Weka and run simple k-means clustering algorithm with different values of k (number of desired clusters). Study the clusters formed. Observe the sum of squared errors and centroids, and derive insight

6. Explore other clustering techniques available in Weka.

7. Explore visualization features of Weka to visualize the clusters. Derive interesting insights and explain.