M.Sc. (Computer Science)
Semester-3
MSCS-3-02P: Data Mining & Visualization Lab

Total Marks: 50
External Marks: 15
Internal Marks: 35
Credits: 2
Pass Percentage: 40%

Course: Data Mining & Visualization Lab	
Course Code: MSCS-3-02P	
Course Outcomes (COs)	
After the completion of this course, the students will be able to:	
CO1	Explore WEKA Data Mining/Machine Learning Toolkit.
CO2	Perform data pre-processing tasks and Demonstrate performing association rule mining on data sets.
CO3	Demonstrate the performance of Naïve-Bayes and K-Nearest Neighbor classifiers on data sets.
CO4	Evaluate the performance of Naïve-Bayes and k-Nearest Neighbor classifiers through ROC Curves
CO5	Explore visualization features of Weka to visualize the clusters.

Exp1. Explore WEKA Data Mining/Machine Learning Toolkit

- Downloading and/or installation of WEKA data mining toolkit,

- Understand the features of WEKA toolkit such as Explorer, Knowledge Flow interface,

- Experimenter, command-line interface.

- Navigate the options available in the WEKA (ex. Select attributes panel, Preprocess panel, classify panel, Cluster panel, Associate panel and Visualize panel)

- Study the arff file format

- Explore the available data sets in WEKA.

- Load a data set (ex. Weather dataset, Iris dataset, etc.)

- Load each dataset and observe the following:

  - ➢ List the attribute names and they types

  - ➢ Number of records in each dataset

  - ➢ Identify the class attribute (if any)

- ➢ Plot Histogram
- ➢ Determine the number of records for each class.
- ➢ Visualize the data in various dimensions

Exp2. Perform data pre-processing tasks and Demonstrate performing association rule mining on data sets.

Exp3. Demonstrate performing classification on data sets:

- Load each dataset into Weka and run Id3, J48 classification algorithm. Study the classifier output. Compute entropy values, Kappa statistic. Extract if-then rules from the decision tree generated by the classifier, Observe the confusion matrix.

- Load each dataset into Weka and perform Naïve-Bayes classification and k-Nearest Neighbor classification. Interpret the results obtained.

- Plot ROC Curves and Compare 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.

Exp4. Demonstrate performing clustering of data sets:

- 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 insights.

- Explore other clustering techniques available in Weka

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