DATA MINING LAB

Part – A

1. Write R programs for

1.1 Data set import and export
1.2 Data exploration and visualization
1.3 Generate association rule using apriori and visualize them.
1.4 Construct decision tree and naïve Bayesian classifiers. Visualize and compare the results for accuracy.
1.5 Perform linear regression on a dataset and visualize the results.
1.6 Build clusters using K-means and Hierarchical clustering and visualize the results**.**

2. Use Rapidminer tool and do the following:

2.1 Import and Export data. Create data sets and import them.
2.2 The preprocessing techniques that can be applied are as follows:
   a. Normalization techniques
   b. Aggregation
   c. Data Cleansing
   d. Sampling
2.3 Perform the following on the preprocessed dataset:
   a. Association mining
   b. Decision Tree Classification
   c. Naïve Bayes Classification
   d. K-Means Clustering

Part - B

1. Write a java program to perform aggregation, discretization and stratified sampling on a given dataset.
2. Write a java program to handle missing values
   a. Replacing by the mean for numeric attributes.
   b. Replace by the value that occurs the maximum number of times for categorical attributes.
3. Write a java program to identify the frequent subsets, generate strong rules from a frequent 4-itemset given the confidence and support thresholds.
4. Write a java program to implement the information gain and gini index measures to identify the best attribute to split.
5. Write a java program to construct a Naïve Bayesian classifier for a given dataset.
6. Write a java program to construct a K-Nearest Neighbor classifier for a given dataset.
7. Write a java program to construct a single layer ANN perceptron for a given dataset.
8. Write a java program to perform linear regression on a given numeric dataset with numeric class attribute.
9. Write a java program to perform k-means clustering on numeric dataset.
10. Build a confusion matrix compute sensitivity, specificity, precision, recall, weighted accuracy.