CO327 Machine Learning Practical List

Datasets to be used:

1. Iris dataset
2. Adult
3. Swarm Behavior
4. Breast Cancer Wisconsin (Diagnostic)
5. HCV Dataset
6. Bank Marketing Dataset
7. Wine Quality dataset

Datasets can be downloaded from UCI Machine Learning Repository http://archive.ics.uci.edu/ml/index.php

Note: Choose any one of the above datasets to perform the following experiments.

EXPERIMENTS

1. Exploring and demonstrating Python.
2. Perform Data Preprocessing like outlier detection, handling missing value, analyzing redundancy and normalization on different datasets.
3. Write a program to exhibit the working of the decision tree based ID3 algorithm. With the help of appropriate data set build the decision tree and classify a new sample.
4. Write a program to demonstrate the working of the decision tree based C4.5 algorithm. With the help of data set used in above experiment build the decision tree and classify a new sample.
5. Write a program to demonstrate the working of decision tree based CART algorithm. Build the decision tree and classify a new sample using suitable dataset. Compare the performance with that of ID, C4.5, and CART in terms of accuracy, recall, precision and sensitivity.
6. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
7. Write a program to implement the Naïve Bayesian classifier for appropriate dataset and compute the performance measures of the model.
8. Write a program to implement k-Nearest Neighbor algorithm to classify any dataset of your choice. Print both correct and wrong predictions.
9. Apply k-Means clustering algorithm on suitable datasets and comment on the quality of clustering.
10. Write a program to implement Linear Regression using any appropriate dataset.

11. Write a program to implement ensemble algorithms- AdaBoost and Bagging using the appropriate dataset and evaluate their performance on that dataset.

12. Select any two datasets based on their statistics and perform comparison among all the implemented algorithms using them.

13. Conduct survey (of at least five) different machine learning tools available.