Please find the data set names, link for data set and the method used for each type below. Programs, Outputs and the data set used are placed in each folder separately.

1. Classification & Dimensionality Reduction

Data set name: Wine (Data set description is available in the ‘wine-dataset’ folder.)

Data set Link: <https://archive.ics.uci.edu/ml/datasets/Wine>

Decision Tree and Random forest techniques are used. The code and the output files are placed in separate folders. Code related for dimensionality reduction are placed in different folders.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Number of instances | Number of attributes | Number of attributes used in case of dimensionality reduction | Classification method | Train / Test ratio | Test Accuracy / Error | F1 | Weighted |
| 178 | 13 | - | Decision Tree | 70/30 | 0.0677 | 0.932 | 0.932 |
| 178 | - | 2 | Decision Tree | 70/30 | 0.333 | 0.665 | 0.664 |
| 178 | 13 | - | Random forest | 70/30 | 0.017 | 0.982 | 0.983 |
| 178 | - | 2 | Random forest | 70/30 | 0.190 | 0.812 | 0.817 |

1. Clustering

Dataset name: seeds

Data set link: <https://archive.ics.uci.edu/ml/datasets/seeds>

Method: k-means (k = 3)

1. Regression

Dataset name: airfoil self noice

Dataset link: <https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>

Method: Linear Regression

1. Collaborative filtering

http://grouplens.org/datasets/movielens/

Name: MovieLens 1M Dataset

<http://files.grouplens.org/datasets/movielens/ml-1m.zip>

1. Frequent Pattern matching

Dataset name: mushroom

Dataset link: http://fimi.ua.ac.be/data/mushroom.dat