Milestone 3: Efficiency

Data set :

Features (based on physicochemical tests):

1 - fixed acidity

2 - volatile acidity

3 - citric acid

4 - residual sugar

5 - chlorides

6 - free sulfur dioxide

7 - total sulfur dioxide

8 - density

9 - pH

10 - sulphates

11 - alcohol

We used a single dataset ‘winequality-red.csv’ to split it into train:test =7:3 to produce this report.

In this milestone, our goal is to use 10-re run Average runtime for traintime & testtime of all ML methods, and comparing them pair-wise.

As a suitable statistical test method, we use Independent Sample T test:

Student’s T test

Two-sample location test : such that means of two populations are equal

(variance of the two populations are assumed to be equal)

Welch’s T test

Two-sample location test : such that means of two populations are equal

Assumption dropped

We used these and conducted 2-independent sample difference of mean Welch’s T-test because:

1. we are not sure about the population for both samples.

2. we are not sure about population variances for both samples.

3. We are not sure whether two populations (for respective samples) have the same population variance.

We measured 10 re-run’s of Training time and Testing time for 7 milestone algorithms: Logistic Regression, Decision Tree, Adaboost, Random Forest, KernelSVM, Gaussian Processes, Neural Networks.

The following are the results of the different model evaluations.

**<Training Time T-test comparison>**

Comparing lr\_traintime and dtree\_traintime

New degree of freedom: 9.024333841492712

Test T-Score: 41.15836389300091

Comparable T-score: -3.2476188262299934

Significantly, dtree\_traintime is better than lr\_traintime / lr\_traintime mean train time: 0.38272788519999995 / dtree\_traintime mean train time: 0.008533930700000001

Comparing lr\_traintime and adaboost\_traintime

New degree of freedom: 9.831453763894432

Test T-Score: 33.28794258692248

Comparable T-score: -3.1814849392671953

Significantly, adaboost\_traintime is better than lr\_traintime / lr\_traintime mean train time: 0.38272788519999995 / adaboost\_traintime mean train time: 0.0733719349

Comparing lr\_traintime and rforest\_traintime

New degree of freedom: 9.429140710038748

Test T-Score: 37.23254176828583

Comparable T-score: -3.2127530042271832

Significantly, rforest\_traintime is better than lr\_traintime / lr\_traintime mean train time: 0.38272788519999995 / rforest\_traintime mean train time: 0.0404434443

Comparing lr\_traintime and kernelsvm\_traintime

New degree of freedom: 9.68841184920536

Test T-Score: 32.420233167629064

Comparable T-score: -3.192247229229565

Significantly, kernelsvm\_traintime is better than lr\_traintime / lr\_traintime mean train time: 0.38272788519999995 / kernelsvm\_traintime mean train time: 0.0825884103

Comparing lr\_traintime and gp\_traintime

New degree of freedom: 9.062823082587759

Test T-Score: -11.948609640990709

Comparable T-score: 3.244141945766533

Statistically, no difference detected. But in this sample, lr\_traintimeis slightly better.

lr\_traintime mean train time: 0.38272788519999995

Comparing lr\_traintime and nn\_traintime

New degree of freedom: 9.070111347899156

Test T-Score: -67.48362708259448

Comparable T-score: 3.2434875796024696

Statistically, no difference detected. But in this sample, lr\_traintimeis slightly better.

lr\_traintime mean train time: 0.38272788519999995

Comparing dtree\_traintime and adaboost\_traintime

New degree of freedom: 9.525225735237756

Test T-Score: -32.69534819858653

Comparable T-score: 3.204997803721538

Statistically, no difference detected. But in this sample, dtree\_traintimeis slightly better.

dtree\_traintime mean train time: 0.008533930700000001

Comparing dtree\_traintime and rforest\_traintime

New degree of freedom: 10.016821273271653

Test T-Score: -22.121661662975235

Comparable T-score: 3.1680806229037963

Statistically, no difference detected. But in this sample, dtree\_traintimeis slightly better.

dtree\_traintime mean train time: 0.008533930700000001

Comparing dtree\_traintime and kernelsvm\_traintime

New degree of freedom: 9.634538953422856

Test T-Score: -40.93242838624701

Comparable T-score: 3.1963995424623914

Statistically, no difference detected. But in this sample, dtree\_traintimeis slightly better.

dtree\_traintime mean train time: 0.008533930700000001

Comparing dtree\_traintime and gp\_traintime

New degree of freedom: 9.000084930463888

Test T-Score: -14.402607240847145

Comparable T-score: 3.2498277819849823

Statistically, no difference detected. But in this sample, dtree\_traintimeis slightly better.

dtree\_traintime mean train time: 0.008533930700000001

Comparing dtree\_traintime and nn\_traintime

New degree of freedom: 9.000094783746077

Test T-Score: -70.18521416887616

Comparable T-score: 3.249826881477689

Statistically, no difference detected. But in this sample, dtree\_traintimeis slightly better.

dtree\_traintime mean train time: 0.008533930700000001

Comparing adaboost\_traintime and rforest\_traintime

New degree of freedom: 16.329422717836668

Test T-Score: 13.684419492486557

Comparable T-score: -2.9130131678063544

Significantly, rforest\_traintime is better than adaboost\_traintime / adaboost\_traintime mean train time: 0.0733719349 / rforest\_traintime mean train time: 0.0404434443

Comparing adaboost\_traintime and kernelsvm\_traintime

New degree of freedom: 17.840848124976812

Test T-Score: -3.4878255042847277

Comparable T-score: 2.8814263265096405

Statistically, no difference detected. But in this sample, adaboost\_traintimeis slightly better.

adaboost\_traintime mean train time: 0.0733719349

Comparing adaboost\_traintime and gp\_traintime

New degree of freedom: 9.002908169728219

Test T-Score: -13.979902473075683

Comparable T-score: 3.249569859067281

Statistically, no difference detected. But in this sample, adaboost\_traintimeis slightly better.

adaboost\_traintime mean train time: 0.0733719349

Comparing adaboost\_traintime and nn\_traintime

New degree of freedom: 9.003245563566725

Test T-Score: -69.7337172283953

Comparable T-score: 3.249539048827241

Statistically, no difference detected. But in this sample, adaboost\_traintimeis slightly better.

adaboost\_traintime mean train time: 0.0733719349

Comparing rforest\_traintime and kernelsvm\_traintime

New degree of freedom: 17.07746462282735

Test T-Score: -18.6062621763615

Comparable T-score: 2.8966057130195977

Statistically, no difference detected. But in this sample, rforest\_traintimeis slightly better.

rforest\_traintime mean train time: 0.0404434443

Comparing rforest\_traintime and gp\_traintime

New degree of freedom: 9.001498645120677

Test T-Score: -14.19455854720382

Comparable T-score: 3.249698604686172

Statistically, no difference detected. But in this sample, rforest\_traintimeis slightly better.

rforest\_traintime mean train time: 0.0404434443

Comparing rforest\_traintime and nn\_traintime

New degree of freedom: 9.001672511744696

Test T-Score: -69.96295090474513

Comparable T-score: 3.2496827211180443

Statistically, no difference detected. But in this sample, rforest\_traintimeis slightly better.

rforest\_traintime mean train time: 0.0404434443

Comparing kernelsvm\_traintime and gp\_traintime

New degree of freedom: 9.002406229198515

Test T-Score: -13.920171190539653

Comparable T-score: 3.2496157006008293

Statistically, no difference detected. But in this sample, kernelsvm\_traintimeis slightly better.

kernelsvm\_traintime mean train time: 0.0825884103

Comparing kernelsvm\_traintime and nn\_traintime

New degree of freedom: 9.00268538997566

Test T-Score: -69.67149563142766

Comparable T-score: 3.249590204471039

Statistically, no difference detected. But in this sample, kernelsvm\_traintimeis slightly better.

kernelsvm\_traintime mean train time: 0.0825884103

Comparing gp\_traintime and nn\_traintime

New degree of freedom: 17.946053166832296

Test T-Score: -37.78922115921443

Comparable T-score: 2.8794460283593137

Statistically, no difference detected. But in this sample, gp\_traintimeis slightly better.

gp\_traintime mean train time: 2.2234712123

Comparing lr\_testtime and dtree\_testtime

New degree of freedom: 9.067273311447666

Test T-Score: 82.92676472579211

Comparable T-score: -3.243742237558302

Significantly, dtree\_testtime is better than lr\_testtime / lr\_testtime mean test time: 0.45868430149999995 / dtree\_testtime mean test time: 0.0012292863000000001

Comparing lr\_testtime and adaboost\_testtime

New degree of freedom: 9.078355301152882

Test T-Score: 81.77891768778332

Comparable T-score: -3.242748936719746

Significantly, adaboost\_testtime is better than lr\_testtime / lr\_testtime mean test time: 0.45868430149999995 / adaboost\_testtime mean test time: 0.0074229002999999984

Comparing lr\_testtime and rforest\_testtime

New degree of freedom: 9.006319239688464

Test T-Score: 82.85264579499179

Comparable T-score: -3.249258494353792

Significantly, rforest\_testtime is better than lr\_testtime / lr\_testtime mean test time: 0.45868430149999995 / rforest\_testtime mean test time: 0.002409792

Comparing lr\_testtime and kernelsvm\_testtime

New degree of freedom: 9.09211446182014

Test T-Score: 76.77552604992763

Comparable T-score: -3.241519741012434

Significantly, kernelsvm\_testtime is better than lr\_testtime / lr\_testtime mean test time: 0.45868430149999995 / kernelsvm\_testtime mean test time: 0.0348707438

Comparing lr\_testtime and gp\_testtime

New degree of freedom: 9.000775218140717

Test T-Score: -21.37284689157829

Comparable T-score: 3.249764701249508

Statistically, no difference detected. But in this sample, lr\_testtime is slightly better.

lr\_testtime mean test time: 0.45868430149999995

Comparing lr\_testtime and nn\_testtime

New degree of freedom: 10.34965642257543

Test T-Score: -14.06682264649891

Comparable T-score: 3.1454345483450727

Statistically, no difference detected. But in this sample, lr\_testtime is slightly better.

lr\_testtime mean test time: 0.45868430149999995

Comparing dtree\_testtime and adaboost\_testtime

New degree of freedom: 17.89635819031156

Test T-Score: -12.505751999477427

Comparable T-score: 2.880378247697618

Statistically, no difference detected. But in this sample, dtree\_testtime is slightly better.

dtree\_testtime mean test time: 0.0012292863000000001

Comparing dtree\_testtime and rforest\_testtime

New degree of freedom: 10.67599763726382

Test T-Score: -3.3530577371014583

Comparable T-score: 3.124840151527637

Statistically, no difference detected. But in this sample, dtree\_testtime is slightly better.

dtree\_testtime mean test time: 0.0012292863000000001

Comparing dtree\_testtime and kernelsvm\_testtime

New degree of freedom: 17.573110697046104

Test T-Score: -64.92850184301071

Comparable T-score: 2.8865841999692687

Statistically, no difference detected. But in this sample, dtree\_testtime is slightly better.

dtree\_testtime mean test time: 0.0012292863000000001

Comparing dtree\_testtime and gp\_testtime

New degree of freedom: 9.000002897345556

Test T-Score: -21.918536310759137

Comparable T-score: 3.249835279216015

Statistically, no difference detected. But in this sample, dtree\_testtime is slightly better.

dtree\_testtime mean test time: 0.0012292863000000001

Comparing dtree\_testtime and nn\_testtime

New degree of freedom: 9.005072967426358

Test T-Score: -37.39680892764113

Comparable T-score: 3.2493722218539562

Statistically, no difference detected. But in this sample, dtree\_testtime is slightly better.

dtree\_testtime mean test time: 0.0012292863000000001

Comparing adaboost\_testtime and rforest\_testtime

New degree of freedom: 10.442265761678023

Test T-Score: 13.274540581839481

Comparable T-score: -3.1394359995355314

Significantly, rforest\_testtime is better than adaboost\_testtime / adaboost\_testtime mean test time: 0.0074229002999999984 / rforest\_testtime mean test time: 0.002409792

Comparing adaboost\_testtime and kernelsvm\_testtime

New degree of freedom: 17.883485483780614

Test T-Score: -51.223851919936166

Comparable T-score: 2.8806206564952075

Statistically, no difference detected. But in this sample, adaboost\_testtime is slightly better.

adaboost\_testtime mean test time: 0.0074229002999999984

Comparing adaboost\_testtime and gp\_testtime

New degree of freedom: 9.00000337464456

Test T-Score: -21.91115398343426

Comparable T-score: 3.2498352355938764

Statistically, no difference detected. But in this sample, adaboost\_testtime is slightly better.

adaboost\_testtime mean test time: 0.0074229002999999984

Comparing adaboost\_testtime and nn\_testtime

New degree of freedom: 9.005908670922405

Test T-Score: -37.08709941209162

Comparable T-score: 3.249295956250449

Statistically, no difference detected. But in this sample, adaboost\_testtime is slightly better.

adaboost\_testtime mean test time: 0.0074229002999999984

Comparing rforest\_testtime and kernelsvm\_testtime

New degree of freedom: 10.229020708848193

Test T-Score: -79.72172098022362

Comparable T-score: 3.1534411833547438

Statistically, no difference detected. But in this sample, rforest\_testtime is slightly better.

rforest\_testtime mean test time: 0.002409792

Comparing rforest\_testtime and gp\_testtime

New degree of freedom: 9.000000272154994

Test T-Score: -21.91713088935856

Comparable T-score: 3.249835519142087

Statistically, no difference detected. But in this sample, rforest\_testtime is slightly better.

rforest\_testtime mean test time: 0.002409792

Comparing rforest\_testtime and nn\_testtime

New degree of freedom: 9.00047651669131

Test T-Score: -37.34270904577926

Comparable T-score: 3.2497919961250465

Statistically, no difference detected. But in this sample, rforest\_testtime is slightly better.

rforest\_testtime mean test time: 0.002409792

Comparing kernelsvm\_testtime and gp\_testtime

New degree of freedom: 9.000003967259566

Test T-Score: -21.878439122722952

Comparable T-score: 3.2498351814325837

Statistically, no difference detected. But in this sample, kernelsvm\_testtime is slightly better.

kernelsvm\_testtime mean test time: 0.0348707438

Comparing kernelsvm\_testtime and nn\_testtime

New degree of freedom: 9.0069462812611

Test T-Score: -35.71739643937128

Comparable T-score: 3.2492012885991928

Statistically, no difference detected. But in this sample, kernelsvm\_testtime is slightly better.

kernelsvm\_testtime mean test time: 0.0348707438

Comparing gp\_testtime and nn\_testtime

New degree of freedom: 9.010280412717599

Test T-Score: 21.018686306024303

Comparable T-score: -3.2488972733929056

Significantly, nn\_testtime is better than gp\_testtime / gp\_testtime mean test time: 18.391133951 / nn\_testtime mean test time: 0.75118

From the results, we can observe the relative efficiency between each pair of models. Here, we have 7 models and we are doing a T test for all possible pairs, both for training and test set. Based on the result, we conclude that strictly comparing the runtime of each algorithm, models can be ordered from least time to longer time:

Decision Tree < Random Forest < Adaboost < Kernel SVM < Logistic Regression(Multi) < Gaussian Processes Neural Networks < Deep Neural Networks

Some facts to note from the analysis:

1. Ensemble Learning methods(Decision Tree, Random Forest, Adaboost) are significantly faster, both to train and test, than any other algorithms. Slight differences in time were detected within the 3 ensemble methods, and Decision Tree was the fastest to run.
2. Kernel SVM using RBF kernel(C=1, gamma=0.1) was surprisingly faster than Linear Classifier(Multinomial Logistic Regression). Possible cause is due to the fact that Multinomial Logistic Regression minimizes the multinomial loss fit across the entire probability distribution, and uses a gradient descent solver that takes a long time.
3. GP and NN are significantly slower than all other ML algorithms. GP is faster than NN when training, and the opposite when testing. GP integrates a kernel, and solving the kernel to make predictions will obviously take a long time. For Neural Networks, training time takes longer than GP due to forward and back propagation.
4. Deep Neural Networks has much more layers and nodes than the default NN model, and therefore it is the slowest in training.

In conclusion, by strictly comparing the models with respect to the efficiency of training and test time, ensemble learning is the most efficient followed by Kernel SVM with a small difference. Logistic Regression, Gaussian Processes, and Neural Network is relatively inefficient by purely comparing the runtimes.