Intro to ML Coursework 1 – Part 4 Results

Confusion matrix are of the form:

Rows: {A (actual), C (actual), E (actual), G (actual), O (actual), Q (actual)}

Cols: { A (pred.), C (pred.), E (pred.), G (pred.), O (pred.), Q (pred.)}

Metrics results are of the form: Cols: { A , C , E , G , O , Q}

Result of drop_col_num = 1 on training_full.txt

```
Best columns to be dropped are: [10]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[33.  0.  0.  0.  1.  0.]
 [ 0. 35.  0.  2.  0.  0.]
 [ 0.  3. 23.  0.  0.  0.]
 [ 0.  1.  0. 26.  0.  0.]
 [ 0.  2.  0.  0. 28.  4.]
 [ 1.  0.  0.  0.  7. 34.]]
Accuracy: 0.895
Recall: [0.97058824 0.94594595 0.88461538 0.96296296 0.82352941 0.80952381]
Precision: [0.97058824 0.85365854 1.         0.92857143 0.77777778 0.89473684]
F1 score: [0.97058824 0.8974359  0.93877551 0.94545455 0.8        0.85      ]
Macro-averaged recall: 0.899527625017821
Macro-averaged precision: 0.9042221367223254
Macro-averaged F1 score: 0.9003756980647738
```

Result of drop_col_num = 2 on training_full.txt

```
Best columns to be dropped are: [8, 5]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[33.  0.  0.  0.  1.  0.]
 [ 0. 35.  2.  0.  0.  0.]
 [ 0.  0. 26.  0.  0.  0.]
 [ 0.  0.  2. 23.  0.  2.]
 [ 0.  0.  2.  2. 28.  2.]
 [ 0.  1.  1.  1.  3. 36.]]
Accuracy: 0.905
Recall: [0.97058824 0.94594595 1.         0.85185185 0.82352941 0.85714286]
Precision: [1.         0.97222222 0.78787879 0.88461538 0.875      0.9       ]
F1 score: [0.98507463 0.95890411 0.88135593 0.86792453 0.84848485 0.87804878]
Macro-averaged recall: 0.9081763836665796
Macro-averaged precision: 0.903286065786066
Macro-averaged F1 score: 0.9032988043221071
```

Results of drop_col_num = 3 on training_full.txt

```
Best columns to be dropped are: [7, 5, 1]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[33.  1.  0.  0.  0.  0.]
 [ 0. 34.  2.  1.  0.  0.]
 [ 0.  0. 25.  1.  0.  0.]
 [ 0.  1.  4. 17.  2.  3.]
 [ 0.  0.  0.  0. 33.  1.]
 [ 0.  0.  0.  0.  4. 38.]]
Accuracy: 0.9
Recall: [0.97058824 0.91891892 0.96153846 0.62962963 0.97058824 0.9047619 ]
Precision: [1.         0.94444444 0.80645161 0.89473684 0.84615385 0.9047619 ]
F1 score: [0.98507463 0.93150685 0.87719298 0.73913043 0.90410959 0.9047619 ]
Macro-averaged recall: 0.8926708975728584
Macro-averaged precision: 0.8994247750614474
Macro-averaged F1 score: 0.8902960645370817
```

Results of drop_col_num = 4 on training_full.txt

```
Best columns to be dropped are: [9, 7, 4, 1]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[32.  1.  0.  0.  0.  1.]
 [ 1. 32.  2.  2.  0.  0.]
 [ 0.  0. 26.  0.  0.  0.]
 [ 0.  0.  2. 23.  1.  1.]
 [ 1.  1.  0.  1. 31.  0.]
 [ 1.  0.  1.  0.  3. 37.]]
Accuracy: 0.905
Recall: [0.94117647 0.86486486 1.         0.85185185 0.91176471 0.88095238]
Precision: [0.91428571 0.94117647 0.83870968 0.88461538 0.88571429 0.94871795]
F1 score: [0.92753623 0.90140845 0.9122807  0.86792453 0.89855072 0.91358025]
Macro-averaged recall: 0.9084350456899477
Macro-averaged precision: 0.9022032468901541
Macro-averaged F1 score: 0.9035468140326363
```

Result of drop_col_num = 1 on training_noisy.txt

```
Best columns to be dropped are: [15]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[29.  0.  0.  2.  0.  3.]
 [ 0. 31.  1.  4.  1.  0.]
 [ 0.  2. 24.  0.  0.  0.]
 [ 0.  3.  1. 14.  1.  8.]
 [ 0.  0.  1.  3. 28.  2.]
 [ 1.  0.  3.  4.  5. 29.]]
Accuracy: 0.775
Recall: [0.85294118 0.83783784 0.92307692 0.51851852 0.82352941 0.69047619]
Precision: [0.96666667 0.86111111 0.8        0.51851852 0.8        0.69047619]
F1 score: [0.90625    0.84931507 0.85714286 0.51851852 0.8115942  0.69047619]
Macro-averaged recall: 0.7743966763574607
Macro-averaged precision: 0.7727954144620813
Macro-averaged F1 score: 0.7722161395882113
```

Result of drop_col_num = 2 on training_noisy.txt

```
Best columns to be dropped are: [11, 5]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[31.  0.  0.  1.  0.  2.]
 [ 0. 34.  1.  1.  1.  0.]
 [ 0.  2. 23.  1.  0.  0.]
 [ 2.  4.  1. 16.  1.  3.]
 [ 1.  0.  0.  2. 28.  3.]
 [ 1.  0.  2.  1.  6. 32.]]
Accuracy: 0.82
Recall: [0.91176471 0.91891892 0.88461538 0.59259259 0.82352941 0.76190476]
Precision: [0.88571429 0.85       0.85185185 0.72727273 0.77777778 0.8       ]
F1 score: [0.89855072 0.88311688 0.86792453 0.65306122 0.8        0.7804878 ]
Macro-averaged recall: 0.8155542959464528
Macro-averaged precision: 0.8154361071027738
Macro-averaged F1 score: 0.8138568609040492
```

Results of drop_col_num = 3 on training_noisy.txt

```
Best columns to be dropped are: [11, 5, 0]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[31.  0.  0.  0.  0.  3.]
 [ 0. 35.  0.  1.  1.  0.]
 [ 0.  2. 22.  2.  0.  0.]
 [ 2.  4.  1. 15.  1.  4.]
 [ 1.  0.  0.  2. 28.  3.]
 [ 1.  0.  2.  2.  6. 31.]]
Accuracy: 0.81
Recall: [0.91176471 0.94594595 0.84615385 0.55555556 0.82352941 0.73809524]
Precision: [0.88571429 0.85365854 0.88       0.68181818 0.77777778 0.75609756]
F1 score: [0.89855072 0.8974359  0.8627451  0.6122449  0.8        0.74698795]
Macro-averaged recall: 0.8035074505662742
Macro-averaged precision: 0.8058443904785367
Macro-averaged F1 score: 0.8029940949798678
```

Result of drop_col_num = 4 on training_noisy.txt

```
Best columns to be dropped are: [11, 5, 2, 0]
###Test results on decision tree trained by train_predict_method:###

Confusion Matrix:
[[30.  0.  0.  0.  1.  3.]
 [ 0. 35.  0.  1.  1.  0.]
 [ 0.  2. 22.  1.  0.  1.]
 [ 1.  4.  1. 17.  1.  3.]
 [ 1.  0.  0.  2. 29.  2.]
 [ 1.  0.  2.  2.  6. 31.]]
Accuracy: 0.82
Recall: [0.88235294 0.94594595 0.84615385 0.62962963 0.85294118 0.73809524]
Precision: [0.90909091 0.85365854 0.88       0.73913043 0.76315789 0.775     ]
F1 score: [0.89552239 0.8974359  0.8627451  0.68       0.80555556 0.75609756]
Macro-averaged recall: 0.8158531295786199
Macro-averaged precision: 0.8200062958659543
Macro-averaged F1 score: 0.81622608334433
```