Jared Roderick

Using the training and test sets specified in the syllabus, perform the following tasks:

a) On the madelon dataset, train decision trees of maximum depth 1, 2, ..., up to 12, for a total of 12 decision trees. If your package does not allow the max depth as a parameter, train trees with 2^1, 2^2, ..., 2^12 nodes, again a total of 12 trees. Use the trained trees to predict the class labels on the training and test sets, and obtain the training and test misclassification errors. Plot on the same graph the training and test misclassification errors vs tree depth (or log2 of nodes) as two separate curves. Report in a table the minimum test error and the tree depth (number of nodes or splits) for which the minimum was attained (1 point).

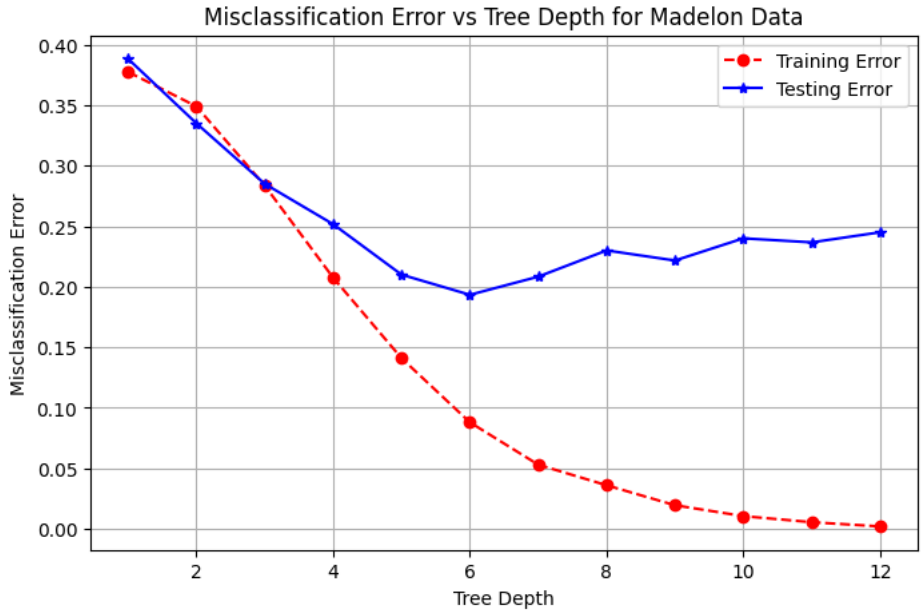
Obtain Training Misclassification Error:

```
The training misclassification error at depth 1 is: 0.37749999999999995
The training misclassification error at depth 2 is: 0.349
The training misclassification error at depth 3 is: 0.2835
The training misclassification error at depth 4 is: 0.20750000000000002
The training misclassification error at depth 5 is: 0.14100000000000001
The training misclassification error at depth 6 is: 0.08799999999999997
The training misclassification error at depth 7 is: 0.05300000000000005
The training misclassification error at depth 8 is: 0.03600000000000003
The training misclassification error at depth 9 is: 0.019499999999999962
The training misclassification error at depth 10 is: 0.010499999999999954
The training misclassification error at depth 11 is: 0.005499999999999949
The training misclassification error at depth 12 is: 0.002000000000000018
```

Obtain Testing Misclassification Error:

```
The testing misclassification error at depth 1 is: 0.3883333333333333
The testing misclassification error at depth 2 is: 0.33499999999999996
The testing misclassification error at depth 3 is: 0.28500000000000003
The testing misclassification error at depth 4 is: 0.2516666666666667
The testing misclassification error at depth 5 is: 0.20999999999999996
The testing misclassification error at depth 6 is: 0.19333333333333336
The testing misclassification error at depth 7 is: 0.20833333333333337
The testing misclassification error at depth 8 is: 0.22999999999999998
The testing misclassification error at depth 9 is: 0.22166666666666668
The testing misclassification error at depth 10 is: 0.24
The testing misclassification error at depth 11 is: 0.2366666666666667
The testing misclassification error at depth 12 is: 0.245
```

Plot the graph of Training and Testing Misclassification Errors vs Tree Depth:



Report in a table the minimum test error and the tree depth at which it was obtained:

	Category	Value
0	Minimum Testing Error	0.193333
1	Best Tree Depth	6.000000

b) Repeat point a) on the satimage, Gisette, hill-valley, and wilt datasets. (2 points)

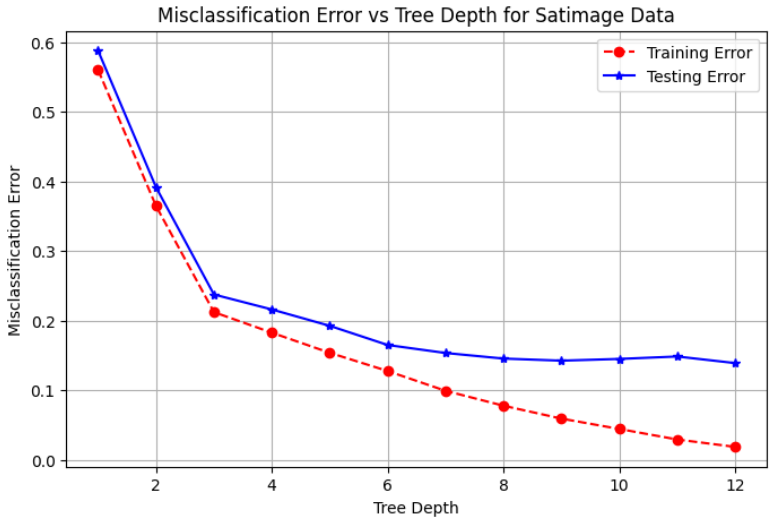
Obtain Training Misclassification Error (satimage):

```
The training misclassification error at depth 1 is: 0.5598647125140924
The training misclassification error at depth 2 is: 0.3645997745208568
The training misclassification error at depth 3 is: 0.2121758737316798
The training misclassification error at depth 4 is: 0.18263810597519725
The training misclassification error at depth 5 is: 0.15377677564825254
The training misclassification error at depth 6 is: 0.1276211950394589
The training misclassification error at depth 7 is: 0.09921082299887263
The training misclassification error at depth 8 is: 0.07779030439684331
The training misclassification error at depth 9 is: 0.059301014656144346
The training misclassification error at depth 10 is: 0.04441939120631344
The training misclassification error at depth 11 is: 0.02931228861330326
The training misclassification error at depth 12 is: 0.01871476888387824
```

Obtain Testing Misclassification Error (satimage):

```
The testing misclassification error at depth 1 is: 0.5875
The testing misclassification error at depth 2 is: 0.39149999999999996
The testing misclassification error at depth 3 is: 0.23750000000000004
The testing misclassification error at depth 4 is: 0.21599999999999997
The testing misclassification error at depth 5 is: 0.1925
The testing misclassification error at depth 6 is: 0.16500000000000004
The testing misclassification error at depth 7 is: 0.15349999999999997
The testing misclassification error at depth 8 is: 0.14549999999999996
The testing misclassification error at depth 9 is: 0.14249999999999996
The testing misclassification error at depth 10 is: 0.14500000000000002
The testing misclassification error at depth 11 is: 0.14849999999999997
The testing misclassification error at depth 12 is: 0.139
```

Plot the graph of Training and Testing Misclassification Errors vs Tree Depth (satimage):



Report in a table the minimum test error and the tree depth at which it was obtained (satimage):

	Category	Value
0	Minimum Testing Error	0.139
1	Best Tree Depth	12.000

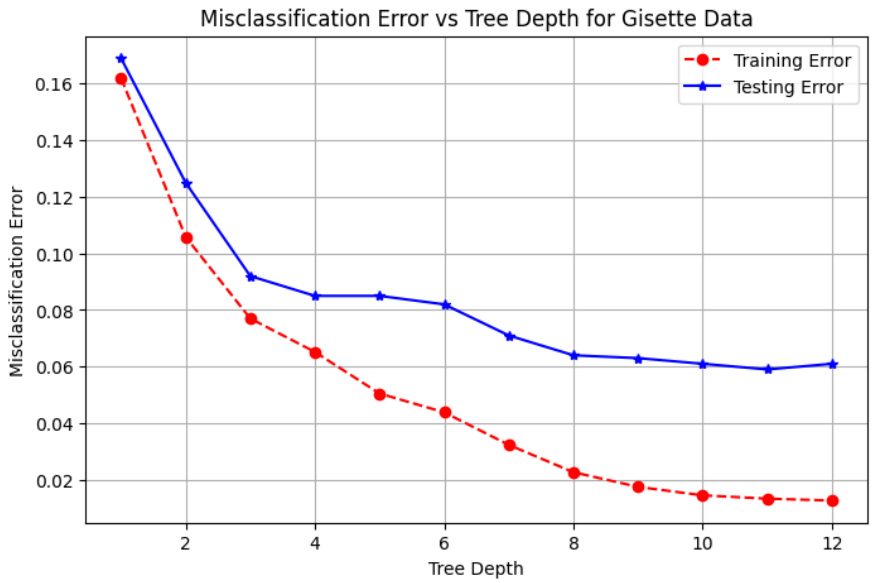
Obtain Training Misclassification Error (Gisette):

```
The training misclassification error at depth 1 is: 0.16200000000000003
The training misclassification error at depth 2 is: 0.10583333333333333
The training misclassification error at depth 3 is: 0.07699999999999996
The training misclassification error at depth 4 is: 0.0651666666666667
The training misclassification error at depth 5 is: 0.05049999999999999
The training misclassification error at depth 6 is: 0.04383333333333328
The training misclassification error at depth 7 is: 0.032333333333333325
The training misclassification error at depth 8 is: 0.022666666666666613
The training misclassification error at depth 9 is: 0.01749999999999996
The training misclassification error at depth 10 is: 0.014499999999999957
The training misclassification error at depth 11 is: 0.013333333333333308
The training misclassification error at depth 12 is: 0.012666666666666715
```

Obtain Testing Misclassification Error (Gisette):

```
The testing misclassification error at depth 1 is: 0.16900000000000004
The testing misclassification error at depth 2 is: 0.125
The testing misclassification error at depth 3 is: 0.09199999999999997
The testing misclassification error at depth 4 is: 0.08499999999999996
The testing misclassification error at depth 5 is: 0.08499999999999996
The testing misclassification error at depth 6 is: 0.08199999999999996
The testing misclassification error at depth 7 is: 0.07099999999999995
The testing misclassification error at depth 8 is: 0.06399999999999995
The testing misclassification error at depth 9 is: 0.06299999999999994
The testing misclassification error at depth 10 is: 0.061000000000000054
The testing misclassification error at depth 11 is: 0.05900000000000005
The testing misclassification error at depth 12 is: 0.061000000000000054
```

Plot the graph of Training and Testing Misclassification Errors vs Tree Depth (Gisette):



Report in a table the minimum test error and the tree depth at which it was obtained (Gisette):

	Category	Value
0	Minimum Testing Error	0.059
1	Best Tree Depth	11.000

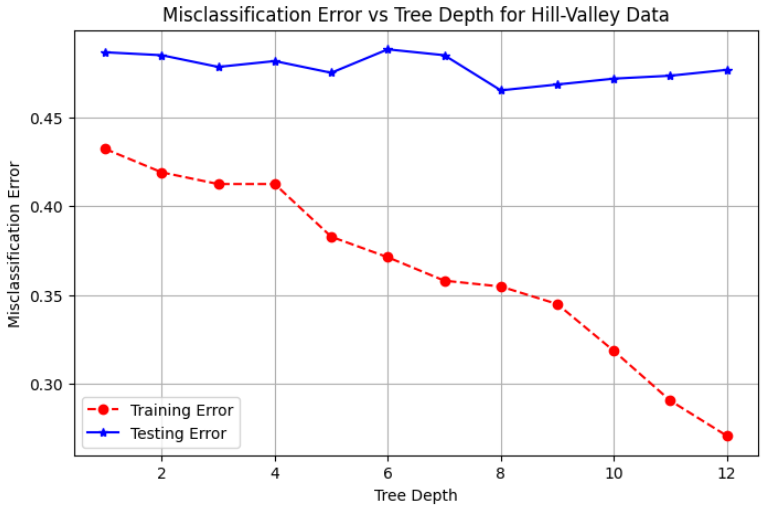
Obtain Training Misclassification Error (hill-valley):

```
The training misclassification error at depth 1 is: 0.43234323432343236
The training misclassification error at depth 2 is: 0.4191419141914191
The training misclassification error at depth 3 is: 0.41254125412541254
The training misclassification error at depth 4 is: 0.41254125412541254
The training misclassification error at depth 5 is: 0.3828382838283828
The training misclassification error at depth 6 is: 0.3712871287128713
The training misclassification error at depth 7 is: 0.35808580858085803
The training misclassification error at depth 8 is: 0.3547854785478548
The training misclassification error at depth 9 is: 0.3448844884488449
The training misclassification error at depth 10 is: 0.3184818481848185
The training misclassification error at depth 11 is: 0.2904290429042904
The training misclassification error at depth 12 is: 0.2706270627062707
```

Obtain Testing Misclassification Error (hill-valley):

```
The testing misclassification error at depth 1 is: 0.48679867986798675
The testing misclassification error at depth 2 is: 0.48514851485148514
The testing misclassification error at depth 3 is: 0.47854785478547857
The testing misclassification error at depth 4 is: 0.4818481848184818
The testing misclassification error at depth 5 is: 0.4752475247524752
The testing misclassification error at depth 6 is: 0.4884488448844885
The testing misclassification error at depth 7 is: 0.48514851485148514
The testing misclassification error at depth 8 is: 0.4653465346534653
The testing misclassification error at depth 9 is: 0.46864686468646866
The testing misclassification error at depth 10 is: 0.471947194719472
The testing misclassification error at depth 11 is: 0.4735973597359736
The testing misclassification error at depth 12 is: 0.47689768976897695
```

Plot the graph of Training and Testing Misclassification Errors vs Tree Depth (hill-valley):



Report in a table the minimum test error and the tree depth at which it was obtained (hill-valley):

	Category	Value
0	Minimum Testing Error	0.465347
1	Best Tree Depth	8.000000

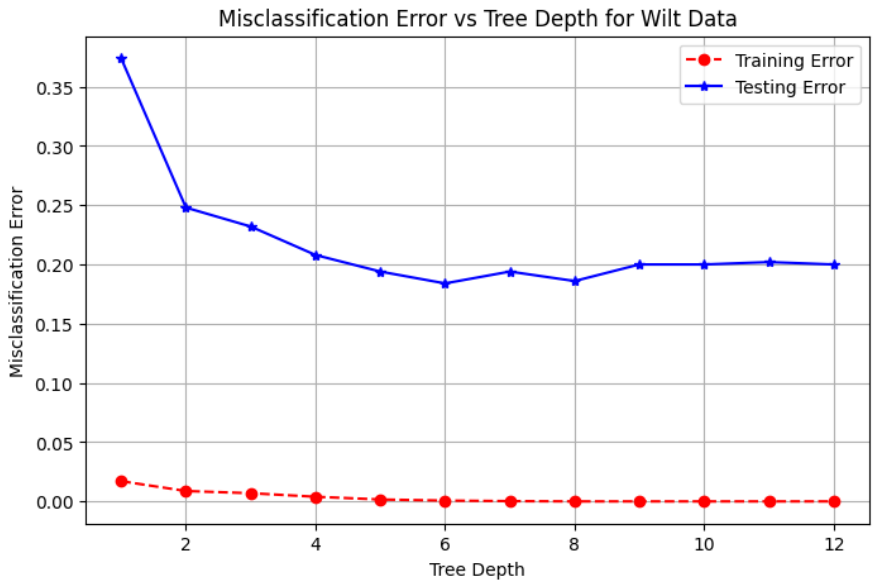
Obtain Training Misclassification Error (wilt):

```
The training misclassification error at depth 1 is: 0.017054620880387228
The training misclassification error at depth 2 is: 0.00875777828992852
The training misclassification error at depth 3 is: 0.006914035492048831
The training misclassification error at depth 4 is: 0.003917953445494393
The training misclassification error at depth 5 is: 0.0016132749481447828
The training misclassification error at depth 6 is: 0.0006914035492049386
The training misclassification error at depth 7 is: 0.00023046784973501655
The training misclassification error at depth 8 is: 0.0
The training misclassification error at depth 9 is: 0.0
The training misclassification error at depth 10 is: 0.0
The training misclassification error at depth 11 is: 0.0
The training misclassification error at depth 12 is: 0.0
```

Obtain Testing Misclassification Error (wilt):

```
The testing misclassification error at depth 1 is: 0.374
The testing misclassification error at depth 2 is: 0.248
The testing misclassification error at depth 3 is: 0.23199999999999998
The testing misclassification error at depth 4 is: 0.20799999999999996
The testing misclassification error at depth 5 is: 0.19399999999999995
The testing misclassification error at depth 6 is: 0.18400000000000005
The testing misclassification error at depth 7 is: 0.19399999999999995
The testing misclassification error at depth 8 is: 0.18600000000000005
The testing misclassification error at depth 9 is: 0.19999999999999996
The testing misclassification error at depth 10 is: 0.19999999999999996
The testing misclassification error at depth 11 is: 0.20199999999999996
The testing misclassification error at depth 12 is: 0.19999999999999996
```

Plot the graph of Training and Testing Misclassification Errors vs Tree Depth (wilt):

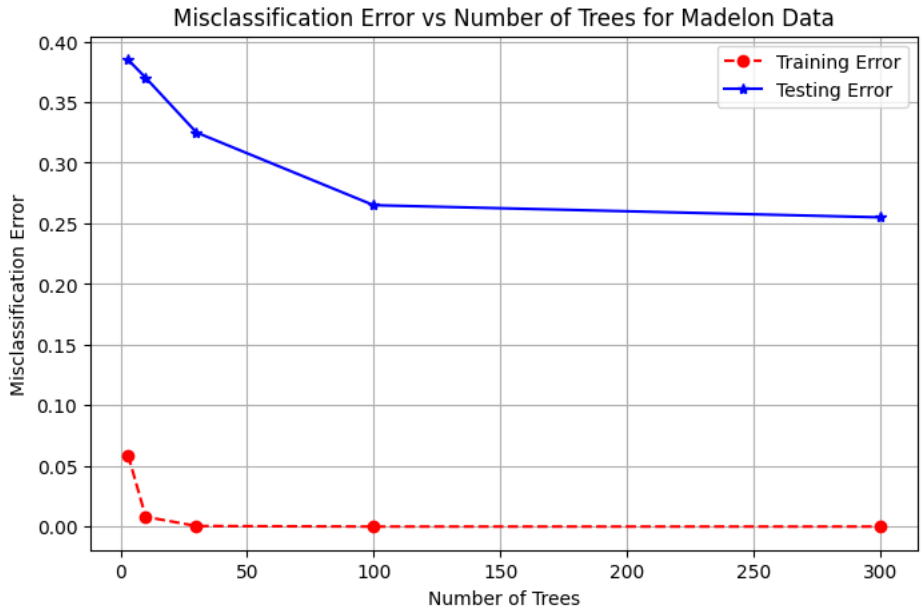


Report in a table the minimum test error and the tree depth at which it was obtained (wilt):

	Category	Value
0	Minimum Testing Error	0.184
1	Best Tree Depth	6.000

c) On the madelon dataset, for each of k in {3, 10, 30, 100, 300} train a random forest with k trees where the split attribute at each node is chosen from a random subset of 22 = sqrt(500) features. Use the trained trees to predict the class labels on the training and test sets, and obtain the training and test misclassification errors. Plot on the same graph the training and test errors vs the number of trees k as two separate curves. Report the training and test misclassification errors in a table.

Plot the graph of Training and Testing Misclassification Errors vs Number of Trees (madelon):

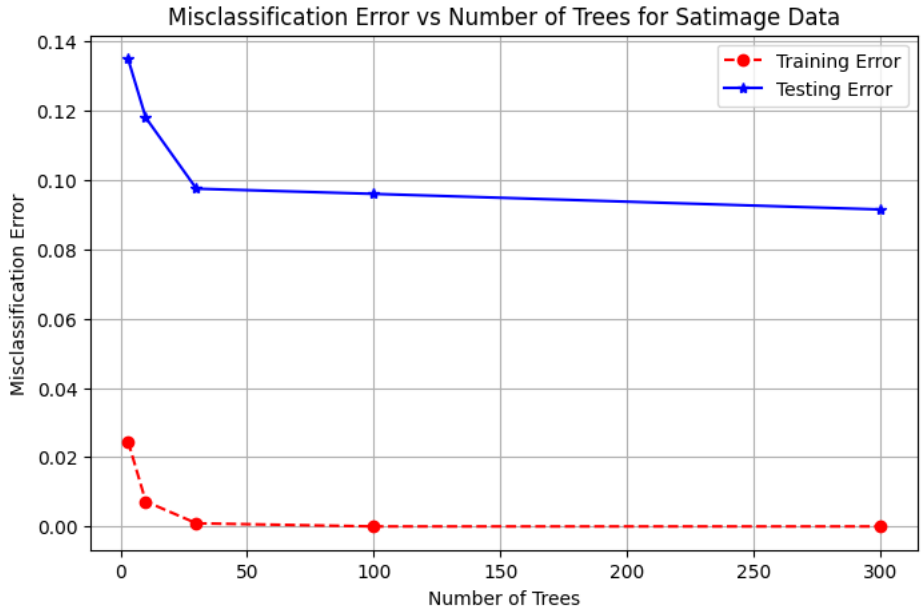


Report the Training and Test Misclassification Errors in a table (madelon):

	Number of Trees	Training Misclassification Error	Testing Misclassification Error
0	3	0.0590	0.385
1	10	0.0080	0.370
2	30	0.0005	0.325
3	100	0.0000	0.265
4	300	0.0000	0.255

d) Repeat point c) on the satimage, Gisette, and hill-valley datasets. (3 points)

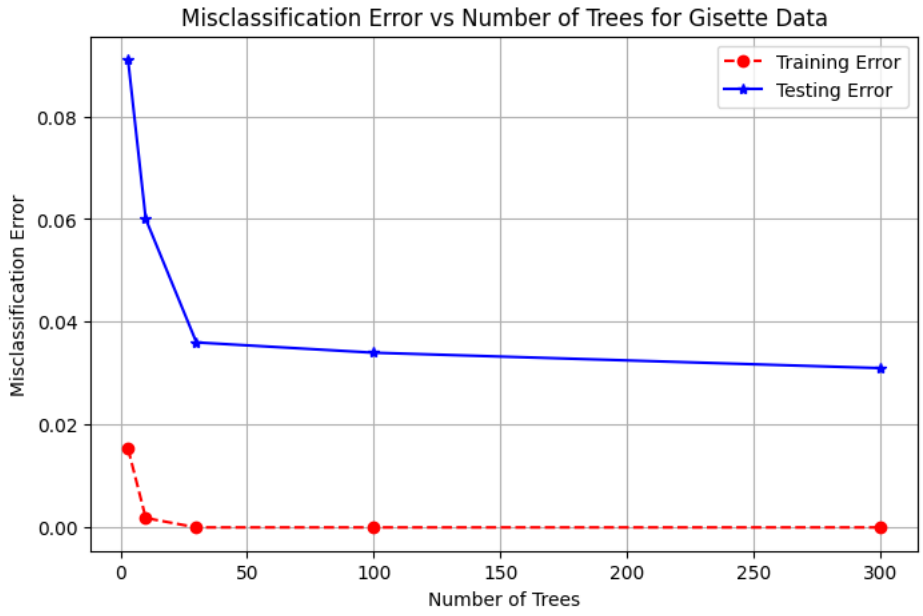
Plot the graph of Training and Testing Misclassification Errors vs Number of Trees (satimage):



Report the Training and Test Misclassification Errors in a table (satimage):

	Number of Trees	Training Misclassification Error	Testing Misclassification Error
0	3	0.024577	0.1350
1	10	0.007215	0.1180
2	30	0.000902	0.0975
3	100	0.000000	0.0960
4	300	0.000000	0.0915

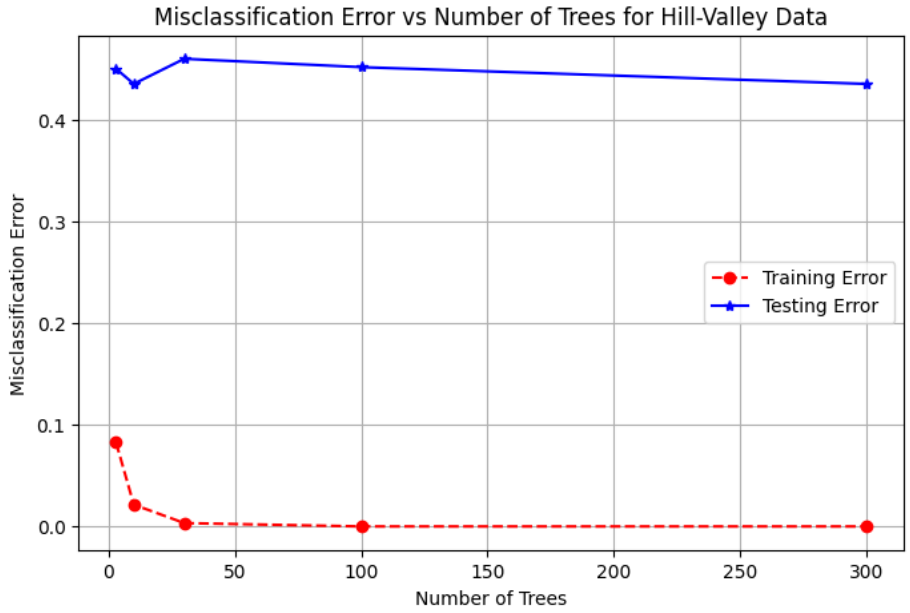
Plot the graph of Training and Testing Misclassification Errors vs Number of Trees (Gisette):



Report the Training and Test Misclassification Errors in a table (Gisette):

	Number of Trees	Training Misclassification Error	Testing Misclassification Error
0	3	0.015333	0.091
1	10	0.001833	0.060
2	30	0.000000	0.036
3	100	0.000000	0.034
4	300	0.000000	0.031

Plot the graph of Training and Testing Misclassification Errors vs Number of Trees (Hill-Valley):

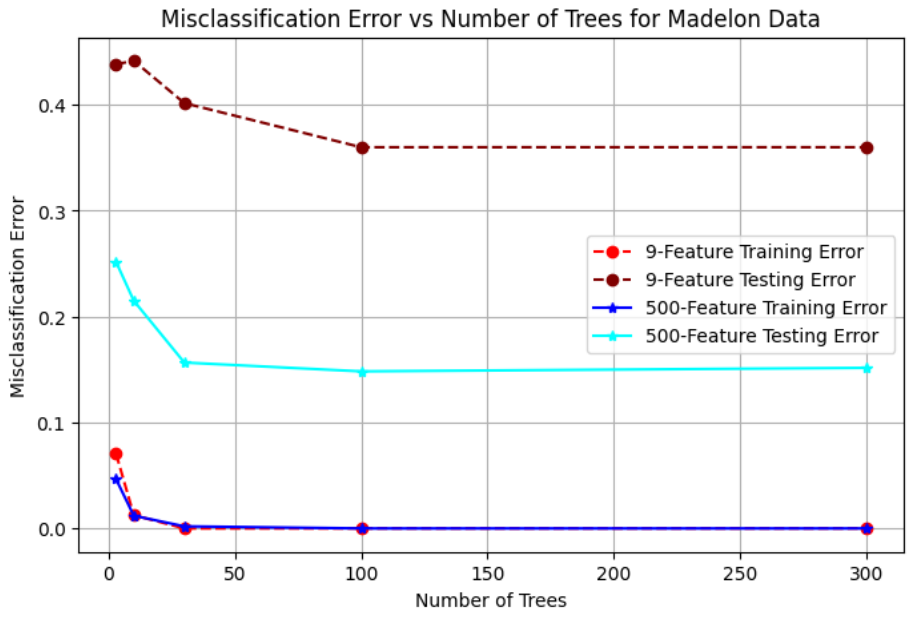


Report the Training and Test Misclassification Errors in a table (Hill-Valley):

	Number of Trees	Training Misclassification Error	Testing Misclassification Error
0	3	0.082508	0.450495
1	10	0.021452	0.435644
2	30	0.003300	0.460396
3	100	0.000000	0.452145
4	300	0.000000	0.435644

e) Repeat point c) on the madelon dataset where the split attribute at each node is chosen from a random subset of 9 ≈ log2(500) features, and where the split attribute at each node is chosen from all 500 features, plotting all four curves (2 train and 2 test misclassification errors) on the same graph.

Plot the graph of Training and Testing Misclassification Errors (9 Features and 500 Features) vs Number of Trees (madelon):




```
The training misclassification error at 3 trees when considering all 500 features is: 0.046499999999999986
The training misclassification error at 10 trees when considering all 500 features is: 0.01200000000000001
The training misclassification error at 30 trees when considering all 500 features is: 0.0020000000000000018
The training misclassification error at 100 trees when considering all 500 features is: 0.0
The training misclassification error at 300 trees when considering all 500 features is: 0.0
```

```
The testing misclassification error at 3 trees when considering all 500 features is: 0.2516666666666667
The testing misclassification error at 10 trees when considering all 500 features is: 0.21499999999999997
The testing misclassification error at 30 trees when considering all 500 features is: 0.15666666666666662
The testing misclassification error at 100 trees when considering all 500 features is: 0.14833333333333332
The testing misclassification error at 300 trees when considering all 500 features is: 0.15166666666666662
```

	Category	Value
0	Minimum Testing Error (9 Features)	0.360000
1	Best Tree Depth (9 Features)	100.000000
2	Minimum Testing Error (500 Features)	0.148333
3	Best Tree Depth (500 Features)	100.000000