a) Stats :-

1) Test1.txt

In case of unpruned tree

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

In case of pruned tree with pruning in forward direction

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

In case of pruned tree in backward direction

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

2) **Test2.txt**

In case of unpruned tree

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

In case of pruned tree with pruning in forward direction

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

In case of pruned tree with pruning in backward direction

Training data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

Test data results:
Precision: 1.0000
Recall: 1.0000
Accuracy: 1.0000

3) testSig.txt

In case of unpruned tree

Training data results:
Precision: 0.8182
Recall: 1.0000
Accuracy: 0.8750

Test data results:
Precision: 0.6667
Recall: 1.0000
Accuracy: 0.7500

In case of pruned tree with pruning in forward direction

Training data results:
Precision: 0.5625
Recall: 1.0000
Accuracy: 0.5625

Test data results:
Precision: 0.5000
Recall: 1.0000
Accuracy: 0.5000

4) **adult.data.csv**

In case of pruned tree with pruning in forward direction

Training data results:
Precision: 0.7962
Recall: 0.7349
Accuracy: 0.8872

Test data results:
Precision: 0.6983
Recall: 0.6462
Accuracy: 0.8445

b)

In case of training stats
After pruning , I think precision and recall value of decision tree should degrade as were are removing checks for most of the combination of features which we were able to do without pruning. But in the training case result, recall is not reducing which is unexpected.

In case of test case stats
After pruning, precision and recall value of decision tree should improve as compare to what we have without pruning. Because of pruning, tree is generalized and is good enough to work for unseen data. As expected we are getting better results in the stats atleast when pruning in backward direction and in forward direction it's atleast equal.

c) In the forward pruned tree, we are checking for marital-status and relationship, I think both are irrelevant as we are checking on number of hour a person works and a person earns on the basis of number of hours he has worked and not on the basis of marital-status or relationship.

d) Precision = true_positive /(true_positive + false_positive)
Recall = true_positive /(true_positive + false_negative)
Accuracy = (true_positive+ true_negative) /(true_positive + false_positive + true_negative + false_negative)

If we have got 90% accuracy and 0% precision, recall then in this case, true_positive is zero i.e our model marked all true examples as False and all false examples as False and this is why we getting such stats.