1. Accuracy using two heuristics: information gain and variance impurity (table 1)

Heuristic	Accuracy(%)	
	Data set 1	Data set 2
Information Gain	49.700	47.336
Variance Impurity	49.900	47.000

Table 1: Accuracy for various heuristics.

Discussion: No heuristic is doing good, less than 50%. Moreover, no heuristic perform well for both data sets: information gain based heuristic does slightly better than variance impurity based heuristic for data set 1 and for data set 2, variance impurity based one does opposite (as we see table 1). Some reasons for these results: 1) features selected by these heuristics are not effected for represent the test data set. 2) test data set is completely different from training and validation data set.

2. Accuracy for pruned decision tree using information gain heuristic for various L and K values (table 2)

L	K	Accuracy(%)	
		Data set 1	Data set 2
5	5	49.850	47.836
5	20	51.400	47.500
20	5	49.850	47.836
20	20	49.850	48.000
20	50	51.400	49.500
50	20	49.850	47.836
50	50	49.850	49.500
100	50	49.850	47.836
50	100	49.850	49.500
100	100	49.850	48.000

Table 2: Accuracy for pruned decision tree generated by information gain heurisitc with various values of L and K.

Discussion: By using pruning on decision tree generated by information gain heuristic, still we are not receiving higher accuracy. For example, for data set 1, we receive 51.4 % as best result and 49.5 % for data set 2.

3. Accuracy for pruned decision tree using variance impurity heuristic for various L and K values (table 3)

Discussion: We see, almost same result as like information gain heuristic for variance impurity heuristic. With these heuristics, decision tree acts like a random classifier for thses two data sets. The class is labeled equaly that means one example is labled as 1 and next example it is labeled as 0. This makes hard to learn from data sets.

4. **Snap Shots:**

L	K	Accuracy(%)	
		Data set 1	Data set 2
5	5	51.400	47.500
5	20	49.850	49.500
20	5	49.850	47.500
20	20	49.850	47.500
20	50	49.850	47.500
50	20	49.850	47.667
50	50	49.850	47.836
100	50	49.850	47.500
50	100	49.850	47.000
100	100	51.400	47.000

Table 3: Accuracy for pruned decision tree generated by variance impurity heurisitc with various values of L and K.

Figure 1: Snapshot 1.

```
.utuattus.euu./peopte/cs/m/mur120030/MLIW1
Main.java  Node.java  Readme.txt  test_set.csv  training_set.csv  validation_set.csv
{cslinux1:~/MLHW1} javac Main.java Node.java
{cslinux1:~/MLHW1} java Main 5 5 training_set.csv validation_set.csv test_set.csv yes
============Accuracy Information====================
Accuracy for Information Gain: 49.7
Accuracy for Variance Impurity: 49.9
Accuracy for Information Gain with pruned tree: 49.85
Accuracy for Variance Impurity with pruned tree: 51.4
============Print Tree================================================================
**********************Decision Tree for Information Gain***************
XO = 0: 0

XO = 1 :

|  XI = 0: 0

|  XI = 1 :

|  |  XT = 0: 0

|  |  XT = 1 :

|  |  |  XS = 0: 0

|  |  |  XS = 1 :

|  |  |  |  XL = 0 :

|  |  |  |  |  XU = 0: 0

|  |  |  |  |  XU = 1 :

|  |  |  |  |  |  XN = 0 :

|  |  |  |  |  |  |  XF = 0 :

|  |  |  |  |  |  |  |  XC = 0 :

|  |  |  |  |  |  |  |  |  XK = 0 :

|  |  |  |  |  |  |  |  |  |  XM = 0 :

|  |  |  |  |  |  |  |  |  |  |  XJ = 0 :

|  |  |  |  |  |  |  |  |  |  |  |  XR = 0: 0

|  |  |  |  |  |  |  |  |  |  |  |  XR = 1: 0

|  |  |  |  |  |  |  |  |  |  |  XJ = 1: 0

|  |  |  |  |  |  |  |  |  |  XM = 1: 1

|  |  |  |  |  |  |  |  |  XK = 1: 0

|  |  |  |  |  |  |  |  XC = 1: 0
```

3

Figure 2: Snapshot 2.

```
{cslinux1:~/MLHW1} java Main 5 20 training_set.csv validation_set.csv test_set.csv no
===========Accuracy Information====================
Accuracy for Information Gain: 49.7
Accuracy for Variance Impurity: 49.9
Accuracy for Information Gain with pruned tree: 51.4
Accuracy for Variance Impurity with pruned tree: 49.85
===========Print Tree=============================================================================
{cslinux1: /MLHW1}
```