Problem Statement : Personality Profile Prediction.

Objectives :

Predicting personality based on a simple dataset which includes nature, intuitiveness, decision making, character. We mainly deals with one is perceiving or judging mentally or not.

Methodology :

To obtain the expected gain we have chosen two classifier, those are given below –

- ✓ K-STAR
- ✓ PART

K-Star : K* is an instance-based classifier, that is the class of a test instance is based upon the class of those training instances similar to it, as determined by some similarity function.

PART : PART uses partial decision trees to generate the decision list that is shown in the output, but only this final list is what is used to make classifications. So there is no need to consider the partial trees that are generated during the learning process, just use the list of rules that is presented by WEKA.

Result Analysis :

K-Star Algorithm :

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.941	0.042	0.941	0.941	0.941	0.900	0.995	0.993	perceiving
	0.958	0.059	0.958	0.958	0.958	0.900	0.995	0.997	judging
Weighted Avg.	0.951	0.052	0.951	0.951	0.951	0.900	0.995	0.995	

Figure : Training Set Analysis.

```
KStar options : -B 20 -M a

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances          39                95.122 %
Incorrectly Classified Instances         2                 4.878 %
Kappa statistic                      0.8995
Mean absolute error                  0.1456
Root mean squared error              0.2043
Relative absolute error             29.9482 %
Root relative squared error         41.4618 %
Total Number of Instances               41
```

Figure : Training Set Accuracy.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 16  1 |  a = perceiving
  1 23 |  b = judging
```

Figure : Training Set Confusion Matrix.

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.556	0.400	0.714	0.556	0.625	0.149	0.733	0.876	perceiving
	0.600	0.444	0.429	0.600	0.500	0.149	0.733	0.631	judging
Weighted Avg.	0.571	0.416	0.612	0.571	0.580	0.149	0.733	0.789	

Figure : Testing Set Analysis.

```
KStar options : -B 20 -M a

Time taken to build model: 0 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.01 seconds

=== Summary ===

Correctly Classified Instances        8                57.1429 %
Incorrectly Classified Instances      6                42.8571 %
Kappa statistic                       0.1429
Mean absolute error                   0.4101
Root mean squared error               0.532
Relative absolute error              78.3687 %
Root relative squared error         100.5585 %
Total Number of Instances            14
```

Figure : Testing Set Accuracy.

```
=== Confusion Matrix ===


a b    <-- classified as
5 4 |  a = perceiving
2 3 |  b = judging
```

Figure : Testing Set Confusion Matrix.

```
train.csv | crossvalidation.arff | Test.arff | Train.arff | kstar.arff
 1 @relation personality-weka.filters.unsupervised.instance.Resample-S1-Z60.0-no-replacement-V-weka.filters.unsupervis
 2
 3 @attribute id numeric
 4 @attribute nature {introvert,extrovert,ambivert}
 5 @attribute intuitive {yes,no}
 6 @attribute decision {feeling,thinking}
 7 @attribute 'prediction margin' numeric
 8 @attribute 'predicted character' {perceiving,judging}
 9 @attribute character {perceiving,judging}
10
11 @data
12 1103001,introvert,yes,feeling,0.921257,perceiving,perceiving
13 1103029,ambivert,yes,feeling,0.564326,perceiving,perceiving
14 1103082,extrovert,yes,thinking,0.968308,judging,judging
15 1103057,ambivert,no,thinking,-0.690218,judging,perceiving
16 1103006,introvert,no,feeling,-0.223481,perceiving,judging
17 1103086,ambivert,yes,thinking,-0.928745,judging,perceiving
18 1103047,ambivert,yes,thinking,0.700327,judging,judging
19 1103025,extrovert,yes,thinking,-0.49132,judging,perceiving
20 1103053,introvert,no,thinking,0.50684,perceiving,perceiving
21 1103015,ambivert,no,thinking,-0.120614,perceiving,judging
22 1103017,introvert,no,thinking,0.79149,perceiving,perceiving
23 1103103,introvert,no,feeling,0.674363,perceiving,perceiving
24 1103093,ambivert,no,thinking,0.727415,judging,judging
25 1103037,ambivert,no,thinking,-0.881867,judging,perceiving
```

Figure : Result (K-Star).
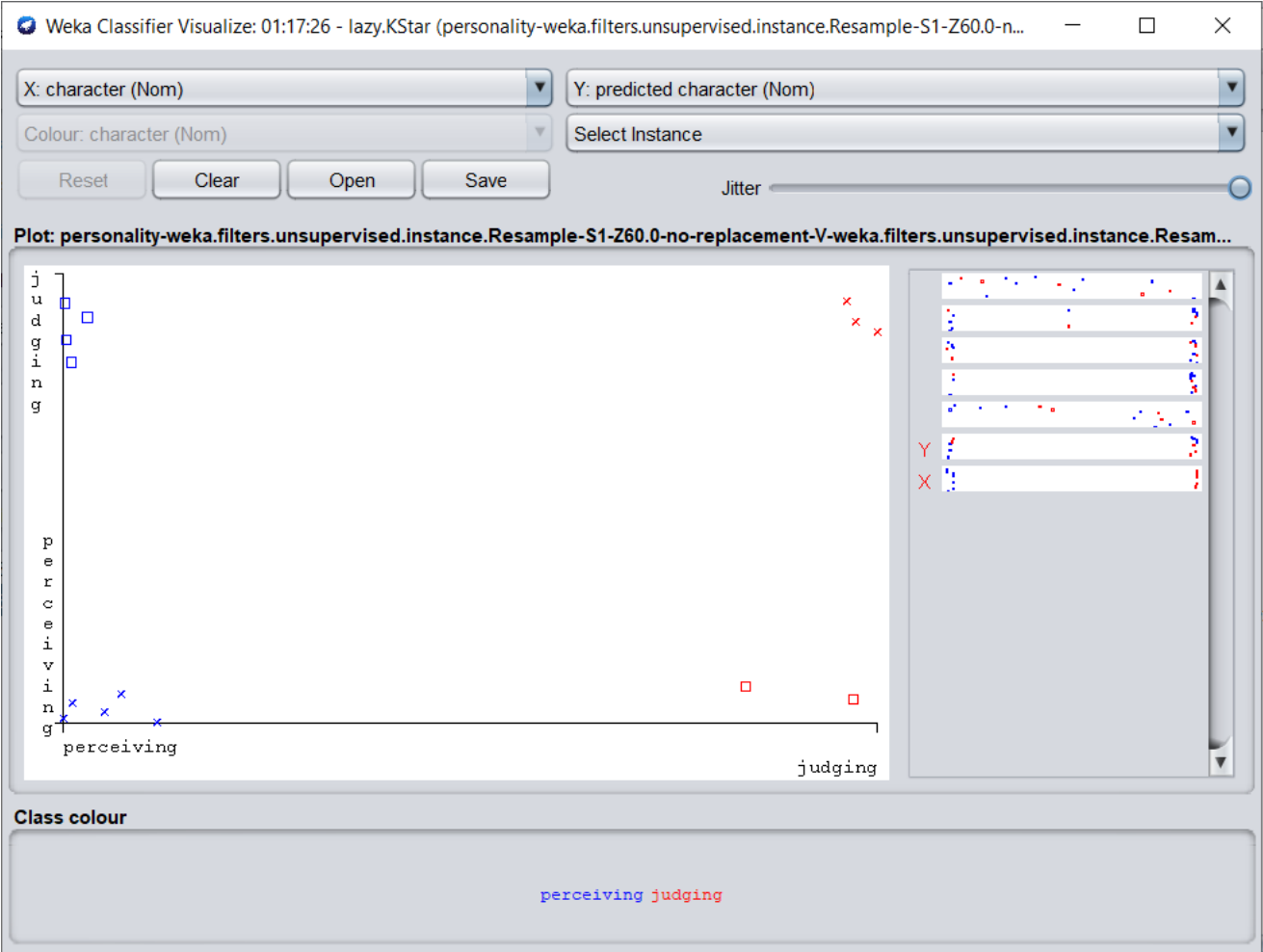
Figure : Classifier visualization (K-Star).

PART Algorithm :

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.882	0.250	0.714	0.882	0.789	0.623	0.875	0.786	perceiving
	0.750	0.118	0.900	0.750	0.818	0.623	0.875	0.897	judging
Weighted Avg.	0.805	0.173	0.823	0.805	0.806	0.623	0.875	0.851	

Figure : Training Set Analysis.

```
Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances        33              80.4878 %
Incorrectly Classified Instances       8              19.5122 %
Kappa statistic                        0.6114
Mean absolute error                    0.2685
Root mean squared error                0.3664
Relative absolute error               55.2305 %
Root relative squared error           74.3667 %
Total Number of Instances             41
```

Figure : Training Set Accuracy.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 15  2 |  a = perceiving
  6 18 |  b = judging
```

Figure : Training Set Confusion Matrix.

```
=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
               0.556    0.200    0.833      0.556   0.667      0.344  0.722     0.739     perceiving
               0.800    0.444    0.500      0.800   0.615      0.344  0.722     0.614     judging
Weighted Avg.  0.643    0.287    0.714      0.643   0.648      0.344  0.722     0.694
```

Figure : Testing Set Analysis.

```
Time taken to build model: 0.01 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances          9               64.2857 %
Incorrectly Classified Instances        5               35.7143 %
Kappa statistic                         0.3137
Mean absolute error                     0.401
Root mean squared error                 0.5408
Relative absolute error                76.6306 %
Root relative squared error           102.2175 %
Total Number of Instances              14
```

Figure : Testing Set Accuracy.

```
=== Confusion Matrix ===

a b   <-- classified as
5 4 | a = perceiving
1 4 | b = judging
```

**Figure : Testing Set Confusion Matrix.**

```
@relation personality-weka.filters.unsupervised.instance.Resample-S1-Z60.0-no-replacement-V-weka.filters.unsupervis

@attribute id numeric
@attribute nature {introvert,extrovert,ambivert}
@attribute intuitive {yes,no}
@attribute decision {feeling,thinking}
@attribute 'prediction margin' numeric
@attribute 'predicted character' {perceiving,judging}
@attribute character {perceiving,judging}

@data
1103001,introvert,yes,feeling,1,perceiving,perceiving
1103029,ambivert,yes,feeling,0.272727,perceiving,perceiving
1103082,extrovert,yes,thinking,1,judging,judging
1103057,ambivert,no,thinking,-0.75,judging,perceiving
1103006,introvert,no,feeling,-1,perceiving,judging
1103086,ambivert,yes,thinking,-0.75,judging,perceiving
1103047,ambivert,yes,thinking,0.75,judging,judging
1103025,extrovert,yes,thinking,0.5,perceiving,perceiving
1103053,introvert,no,thinking,0.333333,perceiving,perceiving
1103015,ambivert,no,thinking,0.75,judging,judging
1103017,introvert,no,thinking,1,perceiving,perceiving
1103103,introvert,no,feeling,-0.333333,judging,perceiving
1103093,ambivert,no,thinking,0.75,judging,judging
1103037,ambivert,no,thinking,-0.75,judging,perceiving
```

Figure : Result (PART).
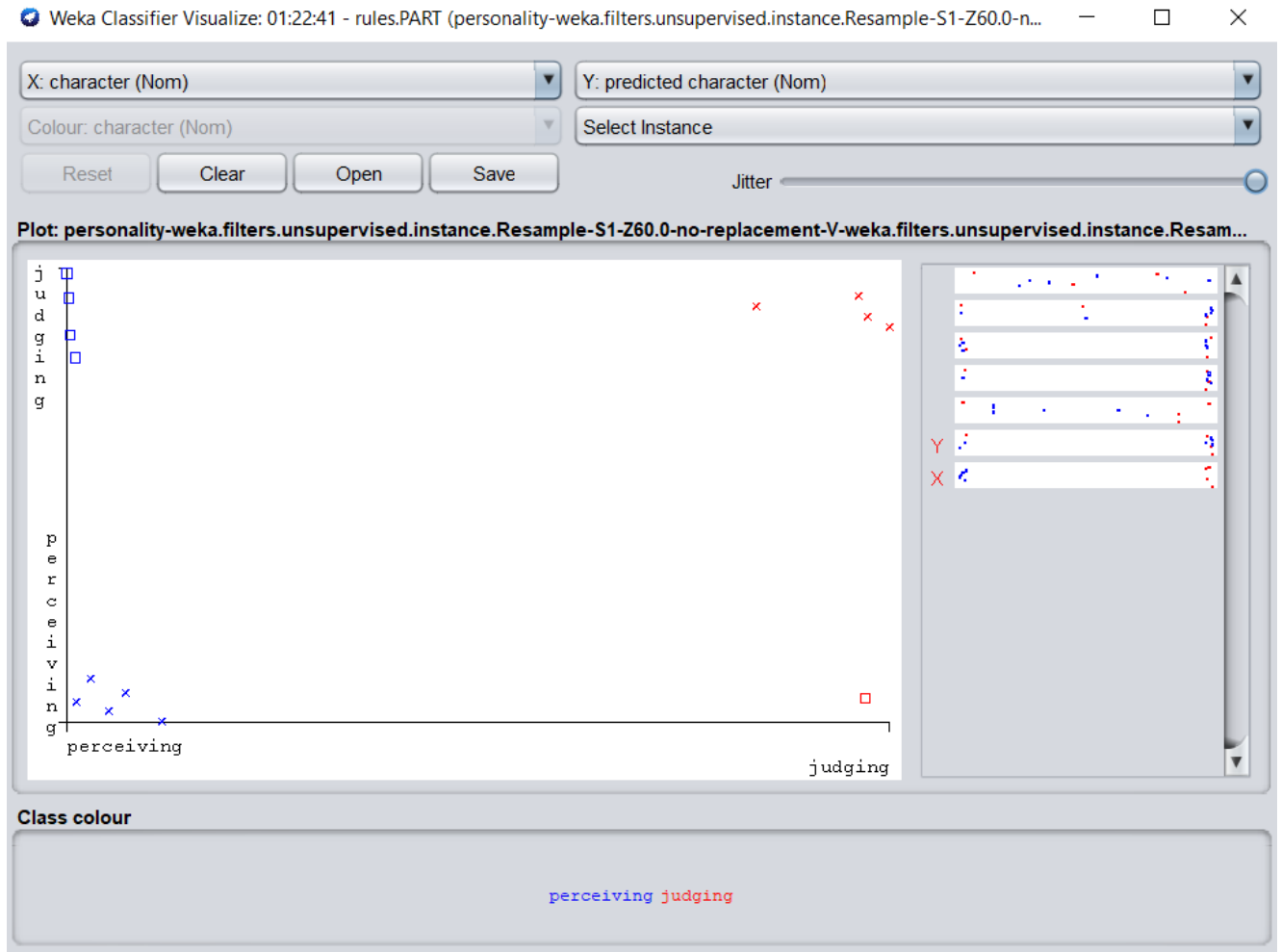
Figure : Classifier visualization (PART).

Comparison :

In our dataset the dominating portion of dataset consists of ambivert in nature, not intuitive , decision making approach is thinking based and characters are based on judging mentality.

After the analysis we find that in PART algorithm we get better result in testing process.

	K-Star	PART
Correctly Classified	57.122%	64.2857%
Incorrectly Classified	42.857%	35.7143%

Discussion :

        Since we have used and checked several classifiers we reach in a decision that for this particular dataset PART gives better result. We also add that the data used to train and test the model does not represent the overall amount of personalities worldwide. We have just used the student types of CSE 3rd batch of Bangladesh Army international University of Science & Technology. The outcomes are the flashes of personality based on perceiving persona and judging persona.