

## Week 2

### Ass. 1

Output of classifier:

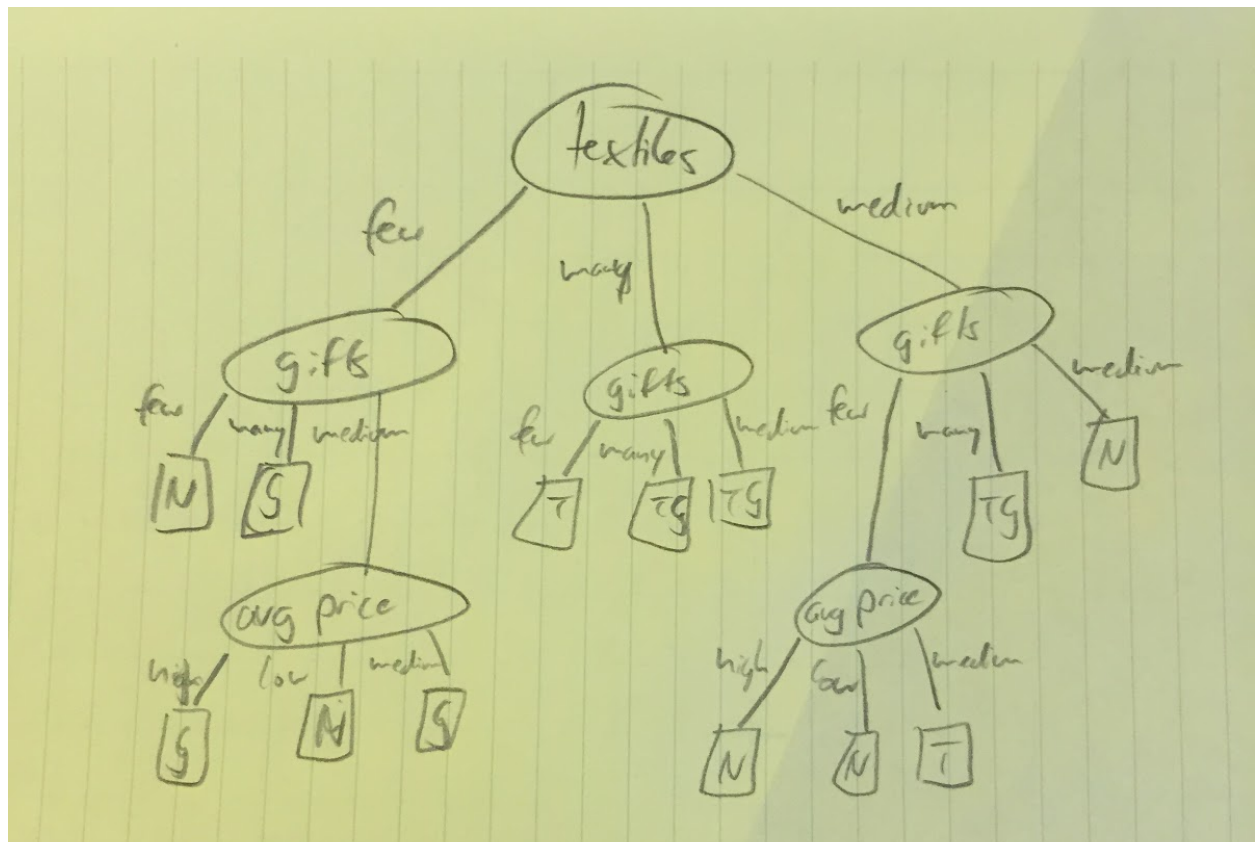
```
> str(tree)
List of 2
 $ splitFeature: chr "textiles"
 $ children    :List of 3
  ..$ :List of 3
  .. ..$ splitFeature: chr "gifts"
  .. ..$ children    :List of 3
  .. .. ..$ :List of 2
  .. .. .. ..$ label      : chr "N"
  .. .. .. ..$ edgeValue: chr "few"
  .. .. .. ..$ :List of 2
  .. .. .. ..$ label      : chr "G"
  .. .. .. ..$ edgeValue: chr "many"
  .. .. .. ..$ :List of 3
  .. .. .. ..$ splitFeature: chr "avgprice"
  .. .. .. ..$ children    :List of 3
  .. .. .. .. ..$ :List of 2
  .. .. .. .. .. ..$ label      : chr "G"
  .. .. .. .. .. ..$ edgeValue: chr "high"
  .. .. .. .. .. ..$ :List of 2
  .. .. .. .. .. ..$ label      : chr "N"
  .. .. .. .. .. ..$ edgeValue: chr "low"
  .. .. .. .. .. ..$ :List of 2
  .. .. .. .. .. ..$ label      : chr "G"
  .. .. .. .. .. ..$ edgeValue: chr "medium"
  .. .. .. .. .. ..$ edgeValue : chr "medium"
  .. .. .. .. .. ..$ edgeValue : chr "few"
  .. .. .. .. .. ..$ :List of 3
  .. .. .. .. .. ..$ splitFeature: chr "gifts"
  .. .. .. .. .. ..$ children    :List of 3
  .. .. .. .. .. .. ..$ :List of 2
  .. .. .. .. .. .. ..$ label      : chr "T"
  .. .. .. .. .. .. ..$ edgeValue: chr "few"
  .. .. .. .. .. .. ..$ :List of 2
  .. .. .. .. .. .. ..$ label      : chr "TG"
  .. .. .. .. .. .. ..$ edgeValue: chr "many"
  .. .. .. .. .. .. ..$ :List of 2
```

Norman Rzepka, 754644

Markus Dücker, 779867

```
.. .. .. ..$ label      : chr "TG"
.. .. .. ..$ edgeValue: chr "medium"
.. ..$ edgeValue      : chr "many"
..$ :List of 3
.. ..$ splitFeature: chr "gifts"
.. ..$ children      :List of 3
.. .. ..$ :List of 3
.. .. .. ..$ splitFeature: chr "avgprice"
.. .. .. ..$ children      :List of 3
.. .. .. .. ..$ :List of 2
.. .. .. .. ..$ label      : chr "N"
.. .. .. .. ..$ edgeValue: chr "high"
.. .. .. .. ..$ :List of 2
.. .. .. .. ..$ label      : chr "N"
.. .. .. .. ..$ edgeValue: chr "low"
.. .. .. .. ..$ :List of 2
.. .. .. .. ..$ label      : chr "T"
.. .. .. .. ..$ edgeValue: chr "medium"
.. .. .. ..$ edgeValue      : chr "few"
.. .. ..$ :List of 2
.. .. .. ..$ label      : chr "TG"
.. .. .. ..$ edgeValue: chr "many"
.. .. ..$ :List of 2
.. .. .. ..$ label      : chr "N"
.. .. .. ..$ edgeValue: chr "medium"
.. ..$ edgeValue      : chr "medium"
```

Graphical representation of the tree:



## Ass. 2

### Explanation of Adaption:

The features of this data set are made up of continuous values. One way to use tree classifiers with these sort of values is to discretize them beforehand.

We used an equal-frequency binning with 10 bins to discretize the feature values. The results are 10 discrete values per feature which represent the intervals the values belong to. An example: 7.0 -> "(6.82,7.03]"

### Statistics:

All Nodes	8611
Leaves	7750
Max Depth	12
Min Depth	4

Norman Rzepka, 754644

Markus Dücker, 779867

### Ass. 3

mit  $\text{nodeErrors} < \text{subtreeErrors}$

	Original	Pruned tree
All Nodes	6651	4901
Leaves	5986	4411
Max Depth	7	7
Min Depth	4	3

mit  $\text{nodeErrors} \leq \text{subtreeErrors}$

	Original	Pruned tree
All Nodes	8611	8401
Leaves	7750	7561
Max Depth	12	7
Min Depth	4	4

### Ass. 4

**Original tree:**

	acc	prec	rec	f1
1	0.5807771	0.5754372	0.3660131	0.4091673
2	0.6134969	0.7568745	0.6076472	0.6447421
3	0.5725971	0.5779841	0.3604154	0.4074531
4	0.6032720	0.5747248	0.3728006	0.4168078
5	0.5991820	0.6476218	0.4593354	0.5022900
6	0.6216769	0.5973619	0.3853260	0.4334867
7	0.5521472	0.6218217	0.4435422	0.4793953
8	0.5541922	0.5534577	0.3518321	0.3899811
9	0.5725971	0.6449281	0.4545403	0.4894051
10	0.5644172	0.5670236	0.3527002	0.3938888

Norman Rzepka, 754644

Markus Dücker, 779867

	Mean	Standard deviation
Accuracy	0.583436	0.024598
Precision	0.611724	0.060539
Recall	0.415415	0.079838
F1	0.456662	0.077628

#### Pruned tree:

	acc	prec	rec	f1
1	0.5562372	0.5657527	0.3548739	0.3960312
2	0.6053170	0.7660890	0.6041629	0.6382817
3	0.5439673	0.5504054	0.3484037	0.3910088
4	0.5521472	0.5357918	0.3485232	0.3896758
5	0.6032720	0.6431935	0.4617212	0.5016440
6	0.5807771	0.5657410	0.3635602	0.4017751
7	0.5357873	0.5904877	0.4356395	0.4729454
8	0.5296524	0.5332537	0.3401776	0.3722002
9	0.5378323	0.6256615	0.4399798	0.4720816
10	0.5623722	0.5540897	0.3526349	0.3900074

	Mean	Standard deviation
Accuracy	0.560736	0.027263
Precision	0.593047	0.070874
Recall	0.404968	0.083408
F1	0.442565	0.082048

P-value for paired T-Test of F1-Scores: 0.001655

Based on the observed metrics, especially the very low p-value, we cannot say that one tree is significantly better than the other.