

CITS3401 Project Report

Hu CHEN, Xinhao HAO
22045072, 21890522

Before doing all data mining, it is required to clean the data first. For association rule mining, it can only be applied to categorical(nominal) data.

Attribute Fnlwgt is removed because it will not be considered important in this example and education is removed because it can be represented by education number. Attributes Age, capital gain, capital loss and hours per week are all modified into categorical attributes with 3 bins using weka. Education number has been changed into categorical, with the same value (1-16).

1. Association Rule Mining

According to the data given, it's difficult to find income bracket >50k since the size of samples is small. However, it is possible to mine a rule with income bracket > 50k with a very small minMetric (0.3 is used) and a large numRules (364 rules mined). Marital status is removed since it is not

Top 5 rules mined with income bracket >50k:

1. age='(41.333333-65.666667]' sex= Male hours_per_week='(33.666667-66.333333]' 3568 ==> income_bracket= >50K. 1669 conf:(0.47)
2. age='(41.333333-65.666667]' sex= Male capital_gain='(-inf-33333]' 3979 ==> income_bracket= >50K. 1748 conf:(0.44)
3. age='(41.333333-65.666667]' hours_per_week='(33.666667-66.333333]' 4915 ==> income_bracket= >50K. 1893 conf:(0.39)
4. age='(41.333333-65.666667]' capital_gain='(-inf-33333]' hours_per_week='(33.666667-66.333333]' 4866 ==> income_bracket= >50K. 1844 conf:(0.38)
5. age='(41.333333-65.666667]' capital_loss='(-inf-1256.666667]' hours_per_week='(33.666667-66.333333]' 4630 ==> income_bracket= >50K. 1708 conf:(0.37)

In plain English:

1. If a man is at middle age and he works 33.666667-66.333333 hours per week, he has a confidence of 0.47 to earn >50k.
2. If a man is at middle age and he has a capital gain of (-inf-33333), he has a confidence of 0.44 to earn >50k.
3. If a person is at middle age and works 33.666667-66.333333 hours per week, he has a confidence of 0.39 to earn >50k.
4. If a person is at middle age and has a capital gain of (-inf-33333) and works 33.666667-66.333333 hours per week, he has a confidence of 0.38 to earn >50k.
5. If a person is at middle age and has a capital loss of (-inf-1256.666667) and works

33.666667-66.333333 hours per week, he has a confidence of 0.37 to earn >50k.
A recommendation to increase income:
There are 4 out of 5 rules including: hours_per_week='(33.666667-66.333333]'. So to increase one's income, it is a good way to work for a longer period.

2. Classification:

a) select all attributes

Use the dataset creating **after cleaning and discretization**, choose **GainRatioAttributeEval** as the Attribute Evaluator. Get ranked attributes as below.

The screenshot shows the WEKA Attribute Evaluator window. The 'Attribute Evaluator' dropdown is set to 'GainRatioAttributeEval'. The 'Search Method' is set to 'Ranker -T -1.7976931348623157E308 -N -1'. The 'Attribute Selection Mode' is set to 'Use full training set'. The 'Result list' shows the ranked attributes for the 'income_bracket' class.

Attribute Evaluator

Choose GainRatioAttributeEval

Search Method

Choose Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode

☒ Use full training set
☐ Cross-validation Folds 10 Seed 1

No class

Start Stop

Attribute selection output

income_bracket
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 13 income_bracket):
Gain Ratio feature evaluator

Ranked attributes:

Gain Ratio	Attribute
0.22468	9 capital_gain
0.08594	4 marital_status
0.07688	6 relationship
0.04057	10 capital_loss
0.03892	8 gender
0.03523	11 hours_per_week
0.03054	3 education_num
0.02865	1 age
0.02563	5 occupation
0.01444	2 workclass
0.01039	7 race
0.00994	12 native_country

Selected attributes: 9,4,6,10,8,11,3,1,5,2,7,12 : 12

Result list (right-click for options)

15:21:21 - Ranker + GainRatioAttributeEval

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % 66
 More options...

(Nom) income_bracket

Start Stop

Result list (right-click for options)

15:21:41 - trees.J48

Classifier output

Number of Leaves : 115
 Size of the tree : 138
 Time taken to build model: 0.53 seconds

=== Stratified cross-validation ===
 === Summary ===

Correctly Classified Instances	13509	82.974 %
Incorrectly Classified Instances	2772	17.026 %
Kappa statistic	0.4725	
Mean absolute error	0.2427	
Root mean squared error	0.3517	
Relative absolute error	67.2507 %	
Root relative squared error	82.8096 %	
Total Number of Instances	16281	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.936	0.513	0.855	0.936	0.894	0.485	0.836	0.924	<=50K.
	0.487	0.064	0.701	0.487	0.575	0.485	0.836	0.627	>50K.
Weighted Avg.	0.830	0.407	0.819	0.830	0.818	0.485	0.836	0.854	

=== Confusion Matrix ===

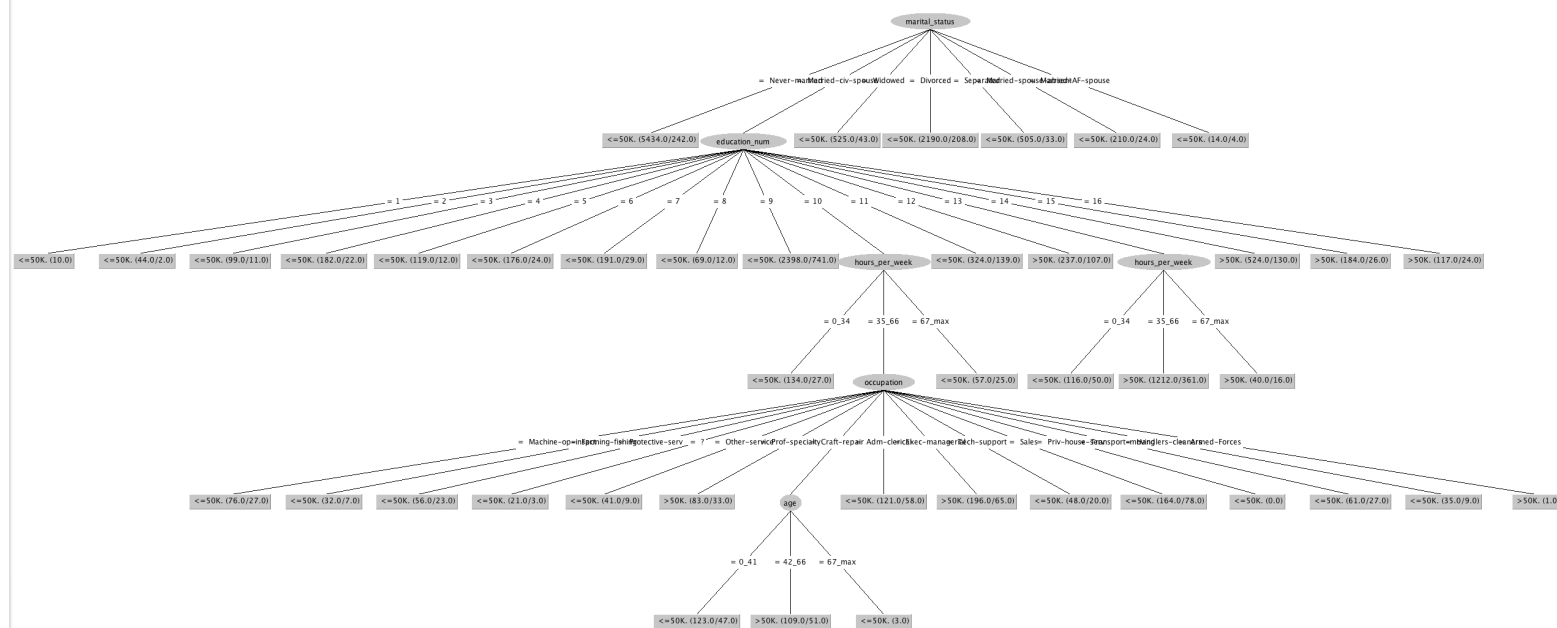
a	b	← classified as
11637	798	a = <=50K.
1974	1872	b = >50K.

10-fold cross-validation 82.97%

The precision is 85.5%, better than random guess (50%). This measures the percentage of tuples/instances the model thinks are "<=50K" are actually labelled "<=50K" in the test set. The recall is 93.6%, which is pretty good, this measures the percentage of "<=50K" tuples/instances the model "recalled" out of all the "<=50K" tuples/instances in the test set.

The precision is 70.1%, which is not great, but better than random guess (50%). This measures the percentage of tuples/instances the model thinks are ">50K" are actually labelled ">50K" in the test set. The recall is 48.7%, which is not good as it is worse than random guess, this measures the percentage of "<=50K" tuples/instances the model "recalled" out of all the "<=50K" tuples/instances in the test set.

Tree View (J48) with minNumObj=100



This decision tree first makes a decision according to one's marital status, next by their education number, then hours per week, occupation, finally age. As shown in the tree, income bracket > 50k will only appear when (marital status = married civ, education num = 10 and 13, hours per week = 35_66, occupation = craft_repair and age = 42_66).

b) select attributes based on the information gain

Get ranked attributes based on *InfoGainAttributeEval* attribute evaluator and first, we selected **the first 6 attributes (Selected attributes: 6,4,5,3,8,1).**

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator: Choose **InfoGainAttributeEval**

Search Method: Choose Ranker -T -1.7976931348623157E308 -N -1

Attribute Selection Mode: ☒ Use full training set ☐ Cross-validation Folds: 10 Seed: 1

No class

Start Stop

Result list (right-click for options)

15:21:21 - Ranker + GainRatioAttributeEval
00:02:52 - Ranker + InfoGainAttributeEval

Attribute selection output

Income Bracket
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 13 income_bracket):
Information Gain Ranking Filter

Ranked attributes:

0.16575	6 relationship
0.15809	4 marital_status
0.09056	5 occupation
0.08943	3 education_num
0.03572	8 gender
0.03293	1 age
0.02923	11 hours_per_week
0.02423	2 workclass
0.01123	10 capital_loss
0.01098	9 capital_gain
0.00901	12 native_country
0.00819	7 race

Selected attributes: 6,4,5,3,8,1,11,2,10,9,12,7 : 12

Status: OK

Log x 0

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % 66

More options...

(Nom) income_bracket

Start Stop

Result list (right-click for options)

15:21:41 - trees.J48
15:24:01 - trees.J48
15:26:55 - trees.J48
22:00:47 - trees.REPTree
22:01:13 - trees.REPTree
22:02:40 - trees.J48
22:03:28 - trees.J48
22:04:11 - trees.J48
00:10:35 - trees.J48
02:06:10 - trees.J48
04:33:42 - trees.J48
04:34:08 - trees.J48
04:38:12 - trees.J48
12:00:34 - trees.J48
12:01:14 - trees.J48

Classifier output

Number of Leaves : 85

Size of the tree : 99

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	13508	82.9679 %
Incorrectly Classified Instances	2773	17.0321 %
Kappa statistic	0.4718	
Mean absolute error	0.245	
Root mean squared error	0.3519	
Relative absolute error	67.8849 %	
Root relative squared error	82.8437 %	
Total Number of Instances	16281	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.936	0.515	0.855	0.936	0.894	0.484	0.836	0.925	<=50K.
	0.485	0.064	0.702	0.485	0.574	0.484	0.836	0.624	>50K.
Weighted Avg.	0.830	0.408	0.819	0.830	0.818	0.484	0.836	0.854	

=== Confusion Matrix ===

a	b	← classified as
11641	794	a = <=50K.
1979	1867	b = >50K.

10-fold cross-validation 82.97%

By comparison, the accuracy difference is very small, therefore we chose **7 attributes** and processed again (**Selected attributes: 6,4,5,3,8,1,11**). (Shown as below)

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % 66

More options...

(Nom) income_bracket

Start Stop

Result list (right-click for options)

15:24:01 - trees.J48
15:26:55 - trees.J48
22:00:47 - trees.REPTree
22:01:13 - trees.REPTree
22:02:40 - trees.J48
22:03:28 - trees.J48
22:04:11 - trees.J48
00:10:35 - trees.J48
02:06:10 - trees.J48
04:33:42 - trees.J48
04:34:08 - trees.J48
04:38:12 - trees.J48
12:00:34 - trees.J48
12:01:14 - trees.J48
12:03:37 - trees.J48

Classifier output

Number of Leaves : 81

Size of the tree : 98

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	13494	82.8819 %
Incorrectly Classified Instances	2787	17.1181 %
Kappa statistic	0.4722	
Mean absolute error	0.2447	
Root mean squared error	0.3522	
Relative absolute error	67.8205 %	
Root relative squared error	82.9071 %	
Total Number of Instances	16281	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.933	0.509	0.856	0.933	0.893	0.483	0.836	0.925	<=50K.
	0.491	0.067	0.695	0.491	0.575	0.483	0.836	0.626	>50K.
Weighted Avg.	0.829	0.404	0.818	0.829	0.818	0.483	0.836	0.854	

=== Confusion Matrix ===

a	b	← classified as
11605	830	a = <=50K.
1957	1889	b = >50K.

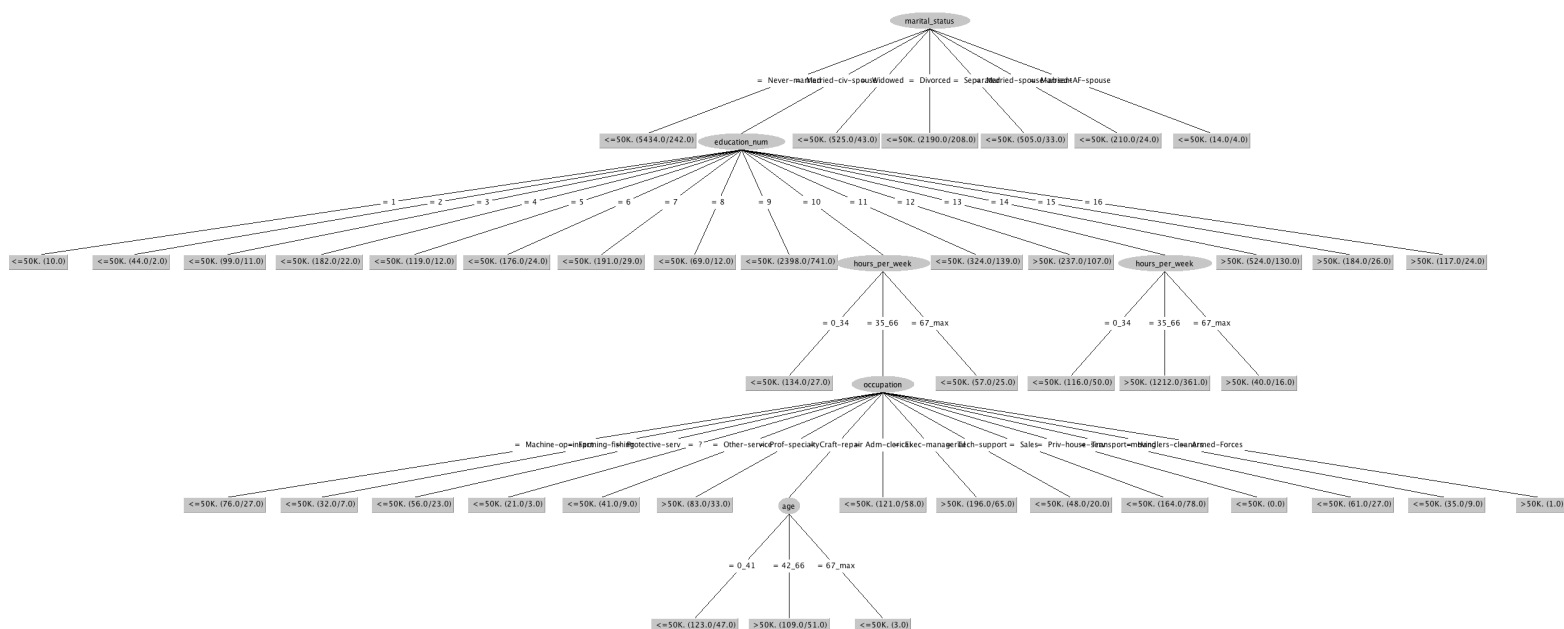
10-fold cross-validation 82.70%

Thus, it can be seen that even **7 attributes** make a **small difference** in accuracy.

The precision is 85.6%, better than random guess (50%). This measures the percentage of tuples/instances the model thinks are "<=50K" are actually labelled "<=50K" in the test set. The recall is 93.3%, which is pretty good, this measures the percentage of "<=50K" tuples/instances the model "recalled" out of all the "<=50K" tuples/instances in the test set.

The precision is 69.5%, which is not great, but better than random guess (50%). This measures the percentage of tuples/instances the model thinks are ">50K" are actually labelled ">50K" in the test set. The recall is 49.1%, which is not good as it is worse than random guess, this measures the percentage of "<=50K" tuples/instances the model "recalled" out of all the "<=50K" tuples/instances in the test set.

Tree View (J48) with minNumObj=100



As two decision trees appear to be mostly the same, thus the same interpreting for this tree model.

3. Clustering:

First, **remove *fnlwgt* and *education* attributes** from the data set. Apply the filter **NumericToNominal** to the *age*, *education_num*, *hours_per_week* attributes.

The screenshot shows the Weka GUI with the 'Preprocess' tab selected. The 'Filter' dropdown is set to 'NumericToNominal -R 11'. The 'Current relation' is 'adult-test-weka.filters.unsupervised.attribute.Remove-R3-4-w...'. The 'Attributes' list on the left shows the following attributes with their selection status:

No.	Name	Selected
1	age	<input checked="" type="checkbox"/>
2	workclass	<input checked="" type="checkbox"/>
3	education_num	<input checked="" type="checkbox"/>
4	marital_status	<input type="checkbox"/>
5	occupation	<input type="checkbox"/>
6	relationship	<input type="checkbox"/>
7	race	<input type="checkbox"/>
8	gender	<input type="checkbox"/>
9	capital_gain	<input type="checkbox"/>
10	capital_loss	<input type="checkbox"/>
11	hours_per_week	<input checked="" type="checkbox"/>
12	native_country	<input type="checkbox"/>
13	income_bracket	<input type="checkbox"/>

The 'Selected attribute' panel on the right shows the details for the 'hours_per_week' attribute, which is of type 'Nominal' and has 89 distinct values. The 'Class' is set to 'income_bracket (Nom)'. A bar chart at the bottom right visualizes the distribution of the 'hours_per_week' attribute, showing a peak at 40 hours per week.

```

=====
Number of iterations: 5
Within cluster sum of squared errors: 78612.89899348503

Initial starting points (random):

Cluster 0: 39, ' Private', 9, ' Married-civ-spouse', ' Craft-repair', ' Husband', ' White', ' Male', 0, 1848, 40, ' United-States'
Cluster 1: 43, ' Private', 5, ' Married-civ-spouse', ' Adm-clerical', ' Husband', ' White', ' Male', 0, 0, 84, ' United-States'

```

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (16281.0)	Cluster#	
		0 (5699.0)	1 (10582.0)
age	35	39	23
workclass	Private	Private	Private
education_num	9	9	10
marital_status	Married-civ-spouse	Married-civ-spouse	Married-civ-spouse
occupation	Prof-specialty	Craft-repair	Adm-clerical
relationship	Husband	Husband	Husband
race	White	White	White
gender	Male	Male	Male
capital_gain	1081.9051	608.5099	1336.8549
capital_loss	87.8993	199.396	27.852
hours_per_week	40	40	40
native_country	United-States	United-States	United-States

Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

```

0      5699 ( 35%)
1     10582 ( 65%)

```

Class attribute: income_bracket
Classes to Clusters:

```

    0    1  <-- assigned to cluster
4514 7921 | <=50K.
1185 2661 | >50K.

```

```

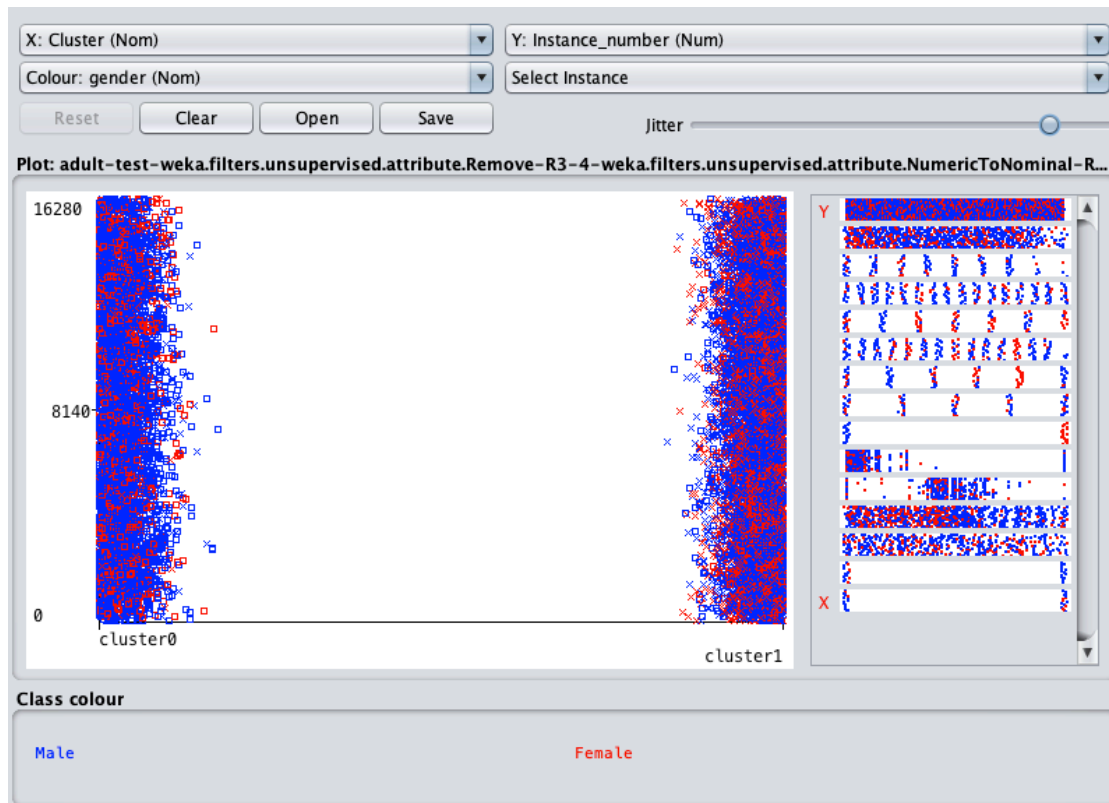
Cluster 0 <-- >50K.
Cluster 1 <-- <=50K.

```

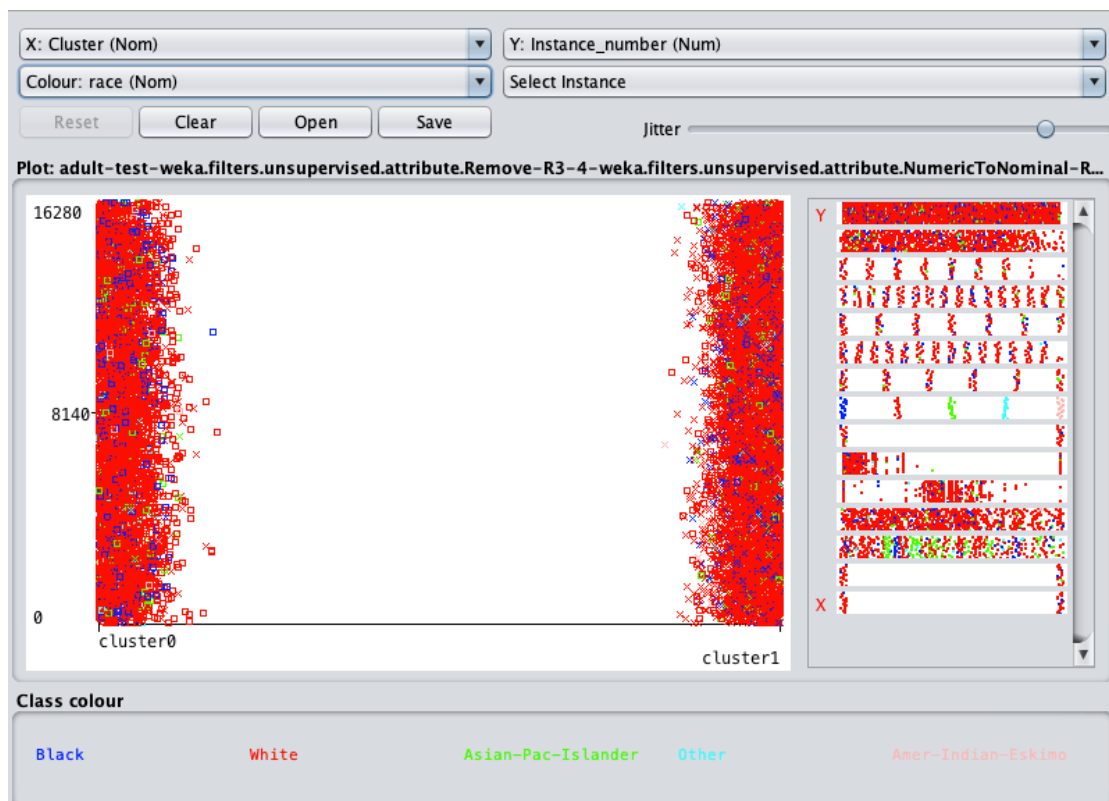
Incorrectly clustered instances : 7175.0 44.0698 %

As we can see from the above screenshots, **cluster 0** is centred around **married white Male (Husband) aged 39**, native country is the **United States**, private workclass: work as **Craft-repair**, education level 9 (HS-grad) with 0.61K capital gain and 0.20K capital loss.

Cluster 1 is mainly **married white Male (Husband) aged 23**, native country is the **United States**, private workclass: work as **Adm-clerical**, education level 10 (Some-college) with 1.34K capital gain and 0.03K capital loss.



Choose **Cluster** as the x-axis, **Instance number** as the y-axis, and **gender** attribute as the colour, we can see *most of cluster 0 and cluster 1 are male*.



Choose **Cluster** as the x-axis, **Instance number** as the y-axis, and **race** attribute as the colour, we can see *most of cluster 0 and cluster 1 are white*.

native_country= Puerto-Rico	0.0043	0.0034	0.005
native_country= Vietnam	0.0012	0.0012	0.0011
native_country= South	0.0021	0.002	0.0022
native_country= Columbia	0.0016	0.0019	0.0013
native_country= Japan	0.0018	0.0023	0.0015
native_country= India	0.0031	0.0034	0.0029
native_country= Cambodia	0.0006	0.0004	0.0007
native_country= Poland	0.0017	0.0018	0.0016
native_country= Laos	0.0003	0.0003	0.0003
native_country= England	0.0023	0.002	0.0025
native_country= Cuba	0.0026	0.0033	0.0021
native_country= Taiwan	0.0009	0.001	0.0008
native_country= Italy	0.002	0.0033	0.0009
native_country= Canada	0.0037	0.0041	0.0035
native_country= Portugal	0.0018	0.0027	0.0011
native_country= China	0.0029	0.0042	0.0018
native_country= Nicaragua	0.0009	0.0005	0.0012
native_country= Honduras	0.0004	0.0003	0.0006
native_country= Iran	0.001	0.0012	0.0008
native_country= Scotland	0.0006	0.0005	0.0006
native_country= Jamaica	0.0015	0.001	0.002
native_country= Ecuador	0.001	0.001	0.0011
native_country= Yugoslavia	0.0004	0.0004	0.0004
native_country= Hungary	0.0004	0.0004	0.0003
native_country= Hong	0.0006	0.0008	0.0004
native_country= Greece	0.0012	0.0019	0.0007
native_country= Trinidad&Tobago	0.0005	0.0007	0.0003
native_country= Outlying-US(Guam-USVI-etc)	0.0006	0.0003	0.0008
native_country= France	0.0006	0.0005	0.0006

Time taken to build model (full training data) : 1.32 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	7363 (45%)
1	8918 (55%)

Class attribute: income_bracket
Classes to Clusters:

0	1	<-- assigned to cluster
4079	8356	<=50K.
3284	562	>50K.

Cluster 0 <-- >50K.
Cluster 1 <-- <=50K.

Incorrectly clustered instances : 4641.0 28.5056 %

Then, apply the **NorminalToBinary** filter to **all the attributes** and apply **SimpleKmeans** clustering with **Classes to clusters** evaluation, the result shows **better** than before.

4. Data reducing

Numerosity reduction using sampling by using **StratifiedRemoveFolds** filter, which makes the number smaller but keeps a similar ratio.

Filter: Choose **StratifiedRemoveFolds -S 0 -N 10 -F 1** Apply Stop

Current relation: Relation: adult-test-weka.filters.unsupervised.attribute.R... Attributes: 13 Instances: 16281 Sum of weights: 16281

Attributes: All None Invert Pattern

No.	Name
1	age
2	workclass
3	education_num
4	marital_status
5	occupation
6	relationship
7	race
8	gender
9	capital_gain
10	capital_loss
11	hours_per_week
12	native_country
13	income_bracket

Remove

Selected attribute: Name: gender Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	Male	10860	10860.0
2	Female	5421	5421.0

Class: income_bracket (Nom) Visualize All

As shown above, before applying the filter, the number of instances is 16281 and for attribute for example: in the gender attribute, Male 10860 and Female 5421, the ratio is 2.003. After applying, the number of instances reduces to 1629. In gender attributes, Male 1090 and Female 539 which keeps the similar ratio 2.022 but reduces the number.

Filter: Choose **StratifiedRemoveFolds -S 0 -N 10 -F 1** Apply Stop

Current relation: Relation: adult-test-weka.filters.unsupervised.attribute.Re... Attributes: 13 Instances: 1629 Sum of weights: 1629

Attributes: All None Invert Pattern

No.	Name
1	age
2	workclass
3	education_num
4	marital_status
5	occupation
6	relationship
7	race
8	gender
9	capital_gain
10	capital_loss
11	hours_per_week
12	native_country
13	income_bracket

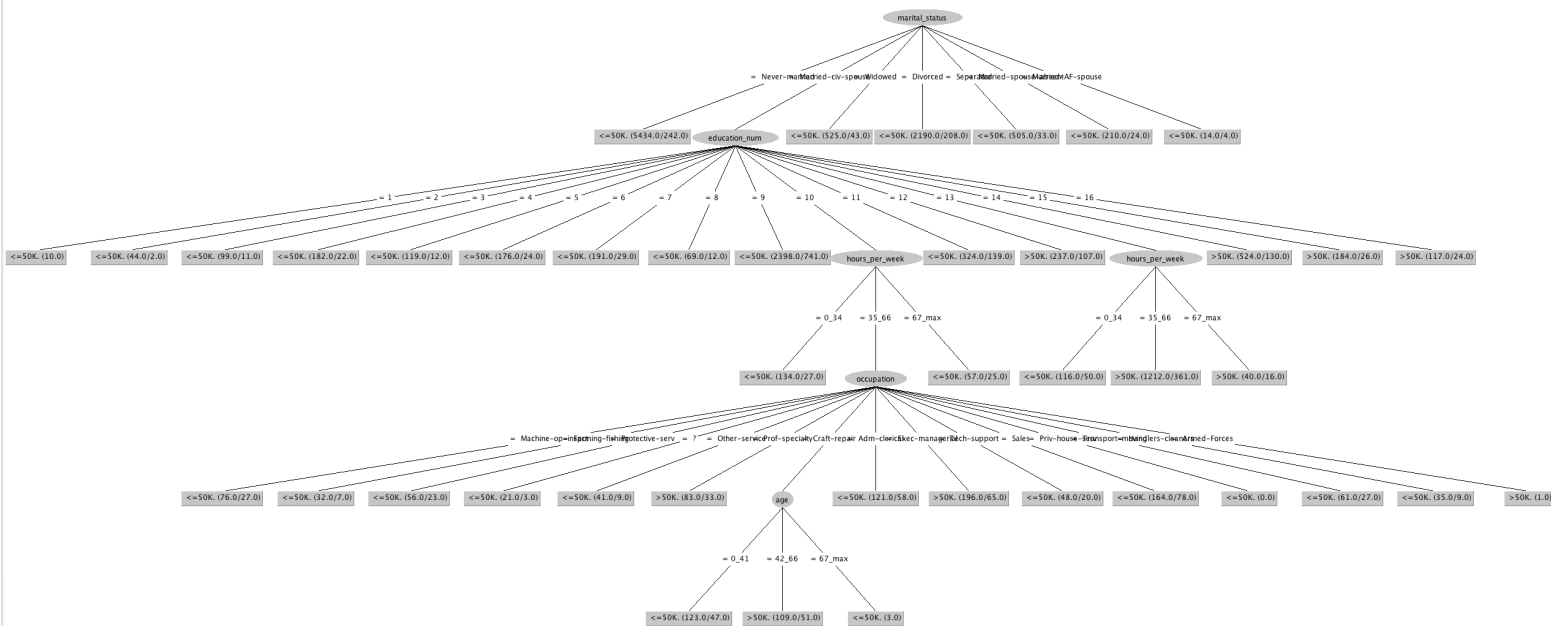
Remove

Selected attribute: Name: gender Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	Male	1090	1090.0
2	Female	539	539.0

Class: income_bracket (Nom) Visualize All

Tree View (J48) with minNumObj=100



For **the original data set**, the tree is huge and can be readable after modifying the minNumObj to 100. The 10-fold cross-validation 82.97%, the precision is 85.5% and recall 93.6% in class $\leq 50K$, and precision 70.1% and recall 48.7% in class $> 50K$.

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds **10**

☐ Percentage split % 66

More options...

(Nom) income_bracket

Start Stop

Result list (right-click for options)

15:21:41 - trees.J48

Classifier output

marital_status = married AF-spouse: <=50K (170/410)

Number of Leaves : 115

Size of the tree : 138

Time taken to build model: 0.53 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	13509	82.974 %
Incorrectly Classified Instances	2772	17.026 %
Kappa statistic	0.4725	
Mean absolute error	0.2427	
Root mean squared error	0.3517	
Relative absolute error	67.2507 %	
Root relative squared error	82.8096 %	
Total Number of Instances	16281	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.936	0.513	0.855	0.936	0.894	0.485	0.836	0.924	$\leq 50K$.
	0.487	0.064	0.701	0.487	0.575	0.485	0.836	0.627	$> 50K$.
Weighted Avg.	0.830	0.407	0.819	0.830	0.818	0.485	0.836	0.854	

=== Confusion Matrix ===

a	b	<- classified as	
11637	798	a =	$\leq 50K$.
1974	1872	b =	$> 50K$.



For **the reduced data set**, the tree is smaller which is readable under minNumObj=2, with 10-fold cross-validation 81.83%, the precision is 85.2% and recall 92.2% in class ≤50K, and precision 65.7% and recall 48.3% in class >50K.

As compared, the outputs are similar to the original dataset as there is little difference between those two models.

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

- ☐ Use training set
- ☐ Supplied test set Set...
- ☒ Cross-validation Folds 10
- ☐ Percentage split % 66

More options...

(Nom) income_bracket

Start Stop

Result list (right-click for options)

- 04:33:42 - trees.J48
- 04:34:08 - trees.J48
- 04:38:12 - trees.J48
- 12:00:34 - trees.J48
- 12:01:14 - trees.J48
- 12:03:37 - trees.J48
- 12:31:39 - trees.J48
- 12:38:14 - trees.J48
- 12:42:59 - trees.J48
- 12:47:02 - trees.J48
- 12:49:14 - trees.J48
- 12:49:22 - trees.J48
- 12:51:34 - trees.J48
- 12:53:41 - trees.J48
- 12:54:38 - trees.J48
- 13:45:49 - trees.J48

Classifier output

```

| | occupation= Sales <= 0: >50K. (3.0/1.0)
| | occupation= Sales > 0: <=50K. (3.0/1.0)
Number of Leaves : 34
Size of the tree : 67
Time taken to build model: 0.2 seconds
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances 1333 81.8293 %
Incorrectly Classified Instances 296 18.1707 %
Kappa statistic 0.4459
Mean absolute error 0.2425
Root mean squared error 0.3719
Relative absolute error 67.1377 %
Root relative squared error 87.537 %
Total Number of Instances 1629

```

Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.922	0.517	0.852	0.922	0.886	0.454	0.780	0.878	<=50K.
	0.483	0.078	0.657	0.483	0.557	0.454	0.780	0.564	>50K.

Confusion Matrix

```

a b <- classified as
1147 97 | a = <=50K.
199 186 | b = >50K.

```