

COMP3411/9814 Artificial Intelligence

Term 1, 2021

Assignment 2 – Machine Learning

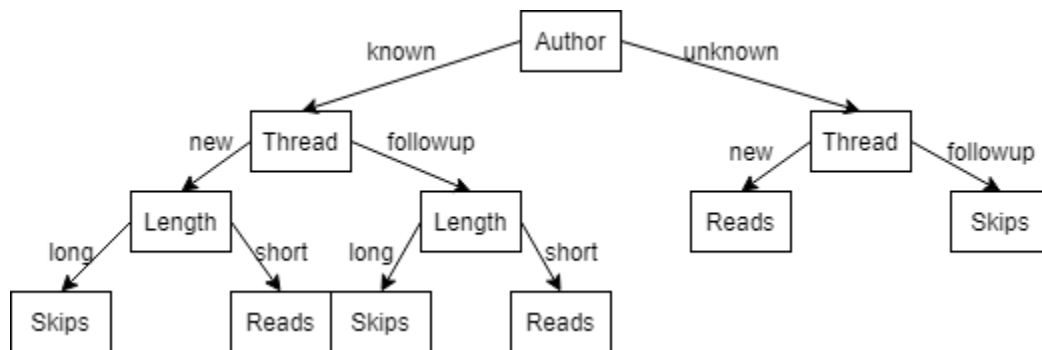
By John Dao (z5258962)

Part 1: Decision trees

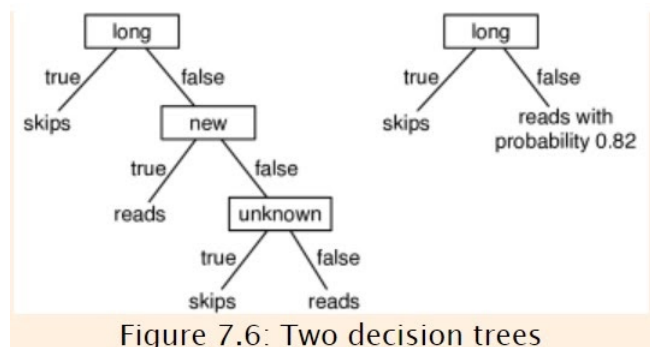
Question 1.1

a)

When features are in the order [Author, Thread, Length, WhereRead] the following tree is created.



However, this tree is different when we compare the one found with the maximum information gain split. We can deduce this through the reference of Figure 7.6 from Poole & Mackworth shown below.



In this figure, we can take the example case of

1. Author = unknown
2. Thread = new
3. Length = long
4. WhereRead = home*

* Not included in stepping

In the model created, we are given the path

Author -> Unknown -> Thread -> Long -> Skips.

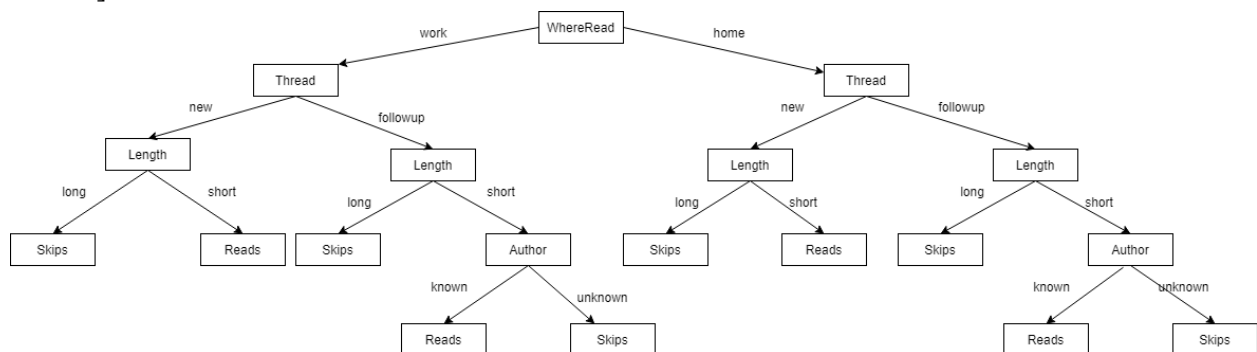
Given the same case, for figure 7.6 we are provided the path

Long -> true -> skips.

This means that given the same input/case, there are different outputs between the tree representing the order [Author, Thread, Length, WhereRead] and the maximum information based tree given in the figure 7.6. Therefore, through the different outputs determined by both trees, we can prove how a different function is found compared to the maximum information tree.

b)

The following tree is found when the features are in the order [WhereRead, Thread, Length, Author].



Upon observation, this is the same tree that could be found with the maximum information gain split as it is able to account for paths that figure 7.6 accounts to as it is structured much the same when in comparison, the original tree with order [Author, Thread, Length, WhereRead] cannot. This can be simply explained in the aforementioned path where again we take the case:

1. Author = unknown
2. Thread = new
3. Length = long
4. WhereRead = home

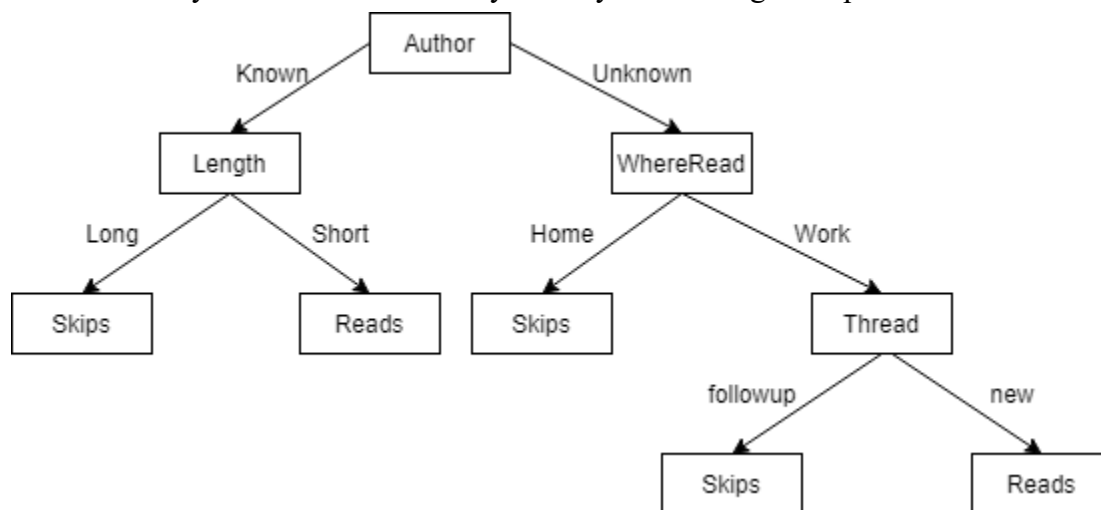
In this tree, we get the following

WhereRead -> home -> thread -> new -> length -> long -> skips.

This skips output is the same as the one in figure 7.6, with other examples following also matching allowing us to deduce the fact that this tree represents the maximum information gain split. Furthermore, this difference in output in comparison with the preceding tree, allows us to distinguish the fact that this tree is in fact different.

c)

There are many trees that can correctly classify the training examples. One of these is below.



In comparison to the other two trees, this tree represents a different function than those found in the preceding algorithms. This can be explained through the use of one example with the case;

1. Author = unknown
2. Thread = new
3. Length = short
4. Whereread = home

In this case, the following path is achieved in the above tree

Author -> unknown -> whereread -> home -> skips

In contrast, the two other trees in questions a) and b) both provide results of read. This means that this tree represents a different function than those found in preceding algorithms.

Question 1.2

Overview

The goal of this task is to master the workflow of model building (via Weka) through the creation of a model of predicted income based upon directed census data (census income dataset).

With the required utilisation of Weka, I have been able to convert both adult.data and adult.test, into a model. Before my exploration into how such steps were taken, there is one assumption that is made throughout my report.

Both data sets were appended into one file named adult.arff. This method was decided upon as at the current time, it was the quickest to get running through Weka and allowed myself to experiment more with the GUI and ultimately led to a higher level of capability during my exposure to Weka.

Method

After retrieving the files adult.data and adult.test, I then with vscode, opened adult.data and took all of the required attributes and data headers and appended them into the top of the file as shown in figure 1.2.1.

```
@relation 'adult'

@attribute Age numeric
@attribute 'Workclass' {' State-gov',' Self-emp-not-inc',' Private',' Federal-gov',' Local-gov',' ?',' Self-emp-inc',' Without-pay',' Never-worked'}
@attribute 'fnlwgt' numeric
@attribute 'Education' {' Bachelors',' HS-grad',' 11th',' Masters',' 9th',' Some-college',' Assoc-acdm',' Assoc-voc',' 7th-8th',' Doctorate',' Prof'}
@attribute 'Education-num' numeric
@attribute 'Marital-status' {' Never-married',' Married-civ-spouse',' Divorced',' Married-spouse-absent',' Separated',' Married-AF-spouse',' Widowed'}
@attribute 'Occupation' {' Adm-clerical',' Exec-managerial',' Handlers-cleaners',' Prof-specialty',' Other-service',' Sales',' Craft-repair',' Trans'}
@attribute 'Relationship' {' Not-in-family',' Husband',' Wife',' Own-child',' Unmarried',' Other-relative'}
@attribute 'Race' {' White',' Black',' Asian-Pac-Islander',' Amer-Indian-Eskimo',' Other'}
@attribute 'Sex' {' Male',' Female'}
@attribute 'Capital-gain' numeric
@attribute 'Capital-loss' numeric
@attribute 'Hours-per-week' numeric
@attribute 'Native-country' {' United-States',' Cuba',' Jamaica',' India',' ?',' Mexico',' South',' Puerto-Rico',' Honduras',' England',' Canada',''}
@attribute 'Predicted_Income' {' <=50K',' >50K'}

@data
39, State-gov, 77516, Bachelors, 13, Never-married, Adm-clerical, Not-in-family, White, Male, 2174, 0, 40, United-States, <=50K
50, Self-emp-not-inc, 83311, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 13, United-States, <=50K
38, Private, 215646, HS-grad, 9, Divorced, Handlers-cleaners, Not-in-family, White, Male, 0, 0, 40, United-States, <=50K
53, Private, 234721, 11th, 7, Married-civ-spouse, Handlers-cleaners, Husband, Black, Male, 0, 0, 40, United-States, <=50K
28, Private, 338409, Bachelors, 13, Married-civ-spouse, Prof-specialty, Wife, Black, Female, 0, 0, 40, Cuba, <=50K
37, Private, 284582, Masters, 14, Married-civ-spouse, Exec-managerial, Wife, White, Female, 0, 0, 40, United-States, <=50K
49, Private, 160187, 9th, 5, Married-spouse-absent, Other-service, Not-in-family, Black, Female, 0, 0, 16, Jamaica, <=50K
52, Self-emp-not-inc, 209642, HS-grad, 9, Married-civ-spouse, Exec-managerial, Husband, White, Male, 0, 0, 45, United-States, >50K
31, Private, 45781, Masters, 14, Never-married, Prof-specialty, Not-in-family, White, Female, 14884, 0, 50, United-States, >50K
42, Private, 159449, Bachelors, 13, Married-civ-spouse, Exec-managerial, Husband, White, Male, 5178, 0, 40, United-States, >50K
37, Private, 288464, Some-college, 10, Married-civ-spouse, Exec-managerial, Husband, Black, Male, 0, 0, 80, United-States, >50K
```

Figure 1.2.1

After the subsequent addition of the headers, I then opened adult.test and removed all '.' to clean up the data and appended it onto the original adult.test. This allowed me to create a file valid for direct conversion to .arff format (that Weka allowed for import). What followed was a simple rename of the file and open file into Weka that then proceeded to show all of the existing data in its entirety (a total of 48842 instances) as shown in figure 1.2.2.

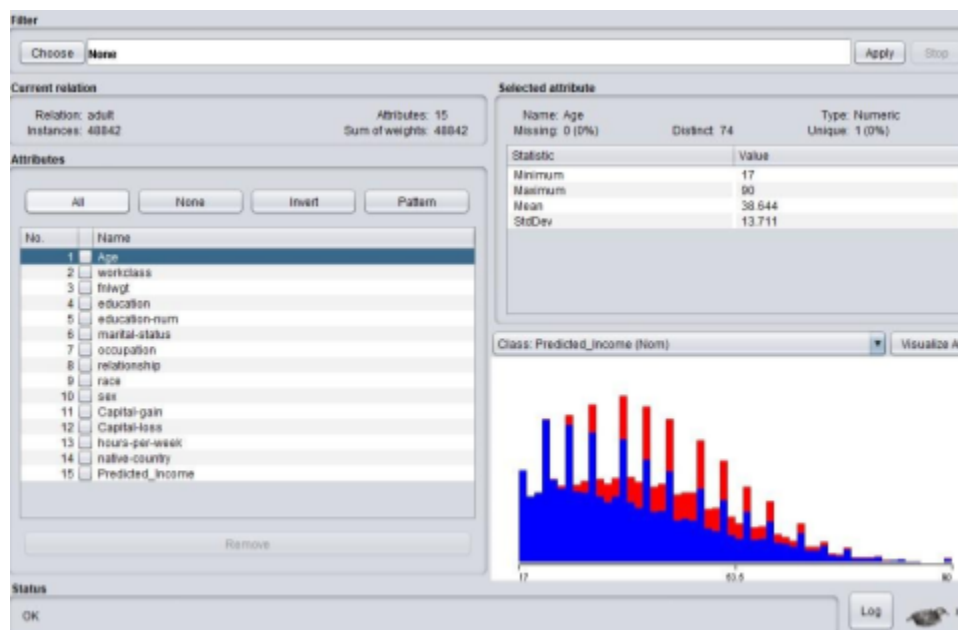


Figure 1.2.2

My attention was then directed to classify and as per the requirements, chose the J48 classifier and ran a basic cross validation of 10 folds on a pruned tree (default setting and one that could possibly increase accuracy) the results of which is shown in figure 1.2.3.

```
Time taken to build model: 0.95 seconds

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

Correctly Classified Instances      42051           86.096 %
Incorrectly Classified Instances    6791           13.904 %
Kappa statistic                    0.5869
Mean absolute error                 0.1962
Root mean squared error            0.3203
Relative absolute error             53.8831 %
Root relative squared error        75.0717 %
Total Number of Instances         48842
```

Figure 1.2.3

I then ran the test again (with the same options but this time with it unpruned (figure 1.2.4), expecting a decrease in accuracy, the results in which indicated a reduction of 1.337% in accuracy. The difference in leaves and nodes is also shown in figure 1.2.5 and 1.2.6 and dictates the overall change in the tree when switching from both pruned and unpruned settings.


```

Time taken to build model: 0.94 seconds

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

Correctly Classified Instances      41398           84.789 %
Incorrectly Classified Instances    7444           15.241 %
Kappa statistic                    0.566
Mean absolute error                 0.1839
Root mean squared error             0.34
Relative absolute error             50.5044 %
Root relative squared error        79.6923 %
Total Number of Instances          48842

=== Detailed Accuracy By Class ===

```

Figure 1.2.4

```

Number of Leaves :    696   Number of Leaves :    9129
Size of the tree :    911   Size of the tree :   10657

```

Figure 1.2.5. Pruned

Figure 1.2.6. Unpruned

This led me to believe that my expectation was correct in assuming that pruning increases the accuracy of the tree and that unpruned trees were deemed to be less accurate than pruned counterparts. Thus I began experimenting with various settings to find out what provided the highest level of accuracy for this particular dataset when running creating a model in Weka. Initially, I thought running a higher number of cross-validation folds was sufficient in increasing accuracy but that was not to be the case. In fact a drop in accuracy had been experienced when utilising 30 folds with a pruned tree shown in figure 1.2.7.

```

Time taken to build model: 0.95 seconds

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

Correctly Classified Instances      42025           86.0428 %
Incorrectly Classified Instances    6817           13.9572 %
Kappa statistic                    0.5846
Mean absolute error                 0.1969
Root mean squared error             0.321
Relative absolute error             54.0915 %
Root relative squared error        75.2335 %
Total Number of Instances          48842

```

Figure 1.2.7

Consequently I referred to other methods such as percentage split and subtree raising, with one instance of which I achieved an accuracy higher than the one indicated in 1.2.3 that utilised a percentage split of 87% with subtreeRaising set to false shown in figure 1.2.8.

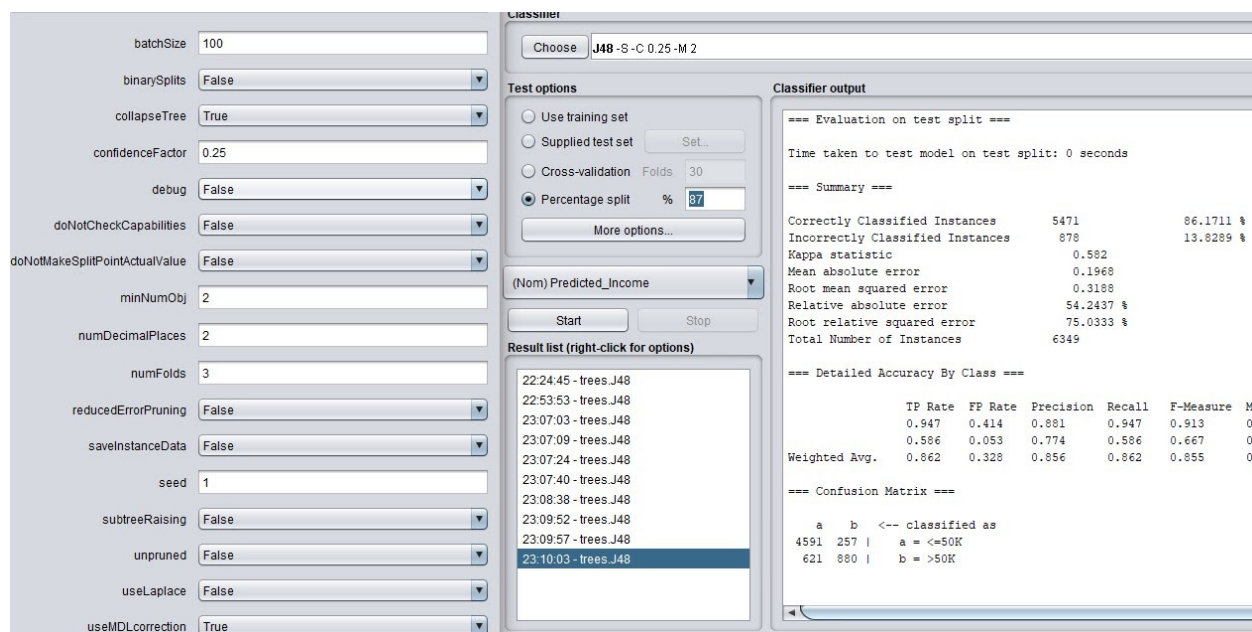


Figure 1.2.8

This experimental sequence led me to discover my highest accuracy model with the below settings.

1. Percentage split: 87%
2. subtreeRaising: true
3. Unpruned: false

The model itself contained an accuracy value of 86.5333% and is dictated below in figure 1.2.9.

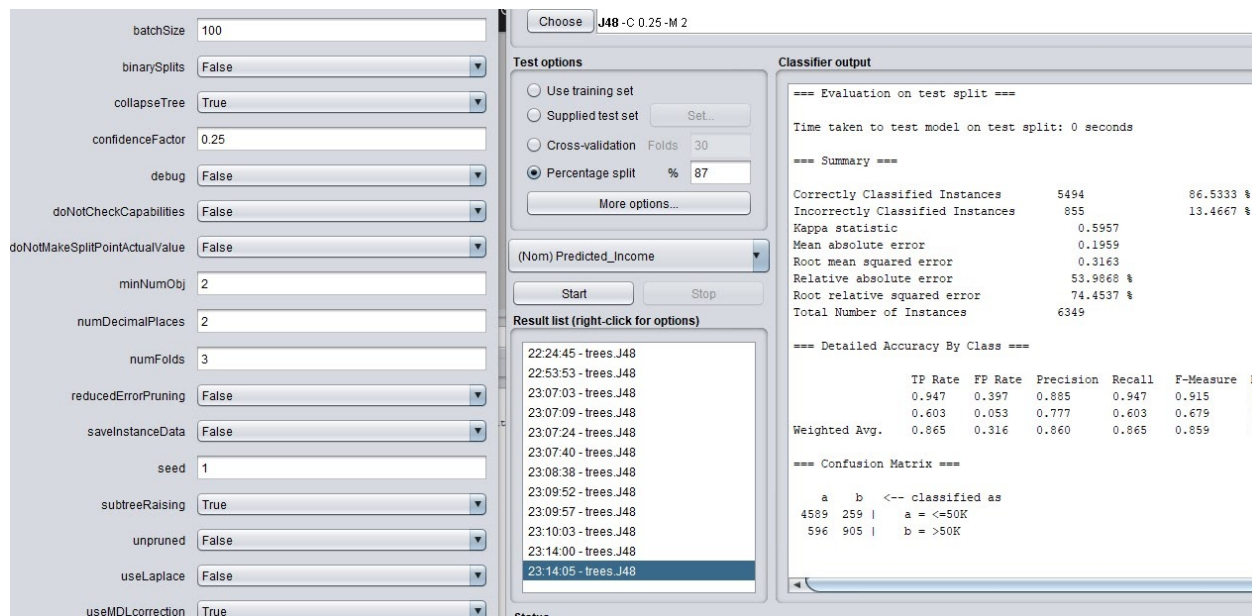


Figure 1.2.9

Ultimately, through this experimentation, I believe to have achieved, to a high degree, a good understanding of the process and workflow of model building and have mastered the critical steps of Importing data, cleaning data, splitting into train/test or cross-validation sets, Pre-processing, transformations and feature engineering all indicated in the figures stated in this question. Through Weka and the J48 implementation, I have been able to successfully create a model based upon adult data indicating predicted income with a relatively high accuracy rate. The model has also been included below for clarification on the structure of my highest accuracy result.

=== Classifier model (full training set) ===

J48 pruned tree

Capital-gain <= 6849

```
| marital-status = Married-civ-spouse
| | Capital-loss <= 1844
| | | education-num <= 11
| | | | Capital-gain <= 5060
| | | | | Age <= 29: <=50K (1999.0/241.0)
| | | | | Age > 29
| | | | | | hours-per-week <= 34: <=50K (1243.0/155.0)
| | | | | | hours-per-week > 34
| | | | | | | education-num <= 9: <=50K (6969.0/1820.0)
| | | | | | | education-num > 9
| | | | | | | | Capital-loss <= 1510
| | | | | | | | | occupation = Tech-support
| | | | | | | | | | Capital-gain <= 3103: >50K (169.69/71.36)
| | | | | | | | | | Capital-gain > 3103: <=50K (12.05/2.0)
| | | | | | | | | | occupation = Craft-repair
| | | | | | | | | | | race = White
| | | | | | | | | | | workclass = Private
| | | | | | | | | | | | fnlwgt <= 152968
| | | | | | | | | | | | Age <= 49
| | | | | | | | | | | | Capital-gain <= 1506: <=50K (153.71/41.51)
| | | | | | | | | | | | Capital-gain > 1506
| | | | | | | | | | | | Capital-gain <= 3273: >50K (4.0)
| | | | | | | | | | | | Capital-gain > 3273: <=50K (7.0/2.0)
| | | | | | | | | | | | Age > 49
```


[illegible]

[illegible]

[illegible]

		hours-per-week > 51:	<=50K (11.37/3.27)
		workclass = Self-emp-not-inc:	<=50K (19.37/6.13)
		workclass = Self-emp-inc:	<=50K (4.08/2.03)
		workclass = Federal-gov:	>50K (12.23/3.15)
		workclass = Local-gov	
		hours-per-week <= 42:	>50K (6.1/2.07)
		hours-per-week > 42:	<=50K (2.06/0.03)
		workclass = State-gov:	>50K (14.27/5.18)
		workclass = Without-pay:	>50K (0.0)
		workclass = Never-worked:	>50K (0.0)
		relationship = Not-in-family:	>50K (0.0)
		relationship = Other-relative:	<=50K (1.0)
		relationship = Unmarried:	>50K (0.0)
		occupation = Handlers-cleaners:	<=50K (92.91/18.68)
		occupation = Machine-op-inspct:	<=50K (197.06/64.45)
		occupation = Adm-clerical	
		race = White	
		workclass = Private	
		education = Bachelors:	<=50K (0.0)
		education = Some-college:	<=50K (137.77/53.75)
		education = 11th:	<=50K (0.0)
		education = HS-grad:	<=50K (0.0)
		education = Prof-school:	<=50K (0.0)
		education = Assoc-acdm:	<=50K (0.0)
		education = Assoc-voc	
		fnlwgt <= 143312	
		hours-per-week <= 42	
		fnlwgt <= 110213:	<=50K (4.12/0.06)
		fnlwgt > 110213:	>50K (3.12/0.06)
		hours-per-week > 42:	>50K (4.17/1.06)
		fnlwgt > 143312:	<=50K (23.4/3.23)
		education = 9th:	<=50K (0.0)
		education = 7th-8th:	<=50K (0.0)
		education = 12th:	<=50K (0.0)
		education = Masters:	<=50K (0.0)
		education = 1st-4th:	<=50K (0.0)
		education = 10th:	<=50K (0.0)
		education = Doctorate:	<=50K (0.0)
		education = 5th-6th:	<=50K (0.0)
		education = Preschool:	<=50K (0.0)

[illegible]

[illegible]

[illegible]


```

| | | | | Age <= 33
| | | | | Age <= 25: <=50K (121.0/28.0)
| | | | | Age > 25
| | | | | education-num <= 12: <=50K (129.0/40.0)
| | | | | education-num > 12
| | | | | native-country = United-States
| | | | | relationship = Wife: >50K (109.69/29.9)
| | | | | relationship = Own-child: <=50K (2.0)
| | | | | relationship = Husband
| | | | | occupation = Tech-support
| | | | | education-num <= 13
| | | | | hours-per-week <= 46
| | | | | fnlwgt <= 142049: >50K (4.0)
| | | | | fnlwgt > 142049
| | | | | Age <= 28: <=50K (2.0)
| | | | | Age > 28
| | | | | fnlwgt <= 168030: <=50K (2.0)
| | | | | fnlwgt > 168030: >50K (7.0/2.0)
| | | | | hours-per-week > 46: <=50K (3.0)
| | | | | education-num > 13: >50K (2.06)
| | | | | occupation = Craft-repair
| | | | | hours-per-week <= 42: <=50K (20.0/6.0)
| | | | | hours-per-week > 42: >50K (16.11/7.0)
| | | | | occupation = Other-service
| | | | | fnlwgt <= 120676: >50K (2.0)
| | | | | fnlwgt > 120676: <=50K (7.93/1.03)
| | | | | occupation = Sales
| | | | | workclass = Private: >50K (89.18/33.9)
| | | | | workclass = Self-emp-not-inc: <=50K (8.82/2.03)
| | | | | workclass = Self-emp-inc
| | | | | Age <= 30: >50K (4.03)
| | | | | Age > 30: <=50K (6.0/2.0)
| | | | | workclass = Federal-gov: >50K (0.0)
| | | | | workclass = Local-gov: <=50K (1.0/0.0)
| | | | | workclass = State-gov: >50K (0.0)
| | | | | workclass = Without-pay: >50K (0.0)
| | | | | workclass = Never-worked: >50K (0.0)
| | | | | occupation = Exec-managerial
| | | | | education = Bachelors
| | | | | Age <= 29

```


[illegible]

[illegible]

[illegible]

		fnlwgt > 315203: >50K (2.0/0.0)
		native-country = Cuba: >50K (3.07/1.04)
		native-country = Iran: >50K (2.05/0.03)
		native-country = Honduras: >50K (1.02/0.01)
		native-country = Philippines: <=50K (6.14/3.05)
		native-country = Italy: <=50K (1.02/0.01)
		native-country = Poland: >50K (0.0)
		native-country = Jamaica: <=50K (1.02/0.01)
		native-country = Vietnam: >50K (1.02/0.01)
		native-country = Mexico: <=50K (8.19/1.07)
		native-country = Portugal: >50K (0.0)
		native-country = Ireland: >50K (0.0)
		native-country = France: >50K (3.07/0.04)
		native-country = Dominican-Republic: <=50K (2.05/1.02)
		native-country = Laos: >50K (0.0)
		native-country = Ecuador: <=50K (2.05/0.02)
		native-country = Taiwan: <=50K (3.07/1.03)
		native-country = Haiti: <=50K (1.02/0.01)
		native-country = Columbia: <=50K (3.07/1.03)
		native-country = Hungary: >50K (0.0)
		native-country = Guatemala: >50K (0.0)
		native-country = Nicaragua: <=50K (1.02/0.01)
		native-country = Scotland: >50K (0.0)
		native-country = Thailand: >50K (1.02/0.01)
		native-country = Yugoslavia: <=50K (2.05/1.02)
		native-country = El-Salvador: <=50K (1.02/0.01)
		native-country = Trinidad&Tobago: >50K (0.0)
		native-country = Peru: >50K (0.0)
		native-country = Hong
		Age <= 29: >50K (2.05/0.02)
		Age > 29: <=50K (3.07/1.01)
		native-country = Holand-Netherlands: >50K (0.0)
		Age > 33
		occupation = Tech-support: >50K (128.28/38.19)
		occupation = Craft-repair
		education = Bachelors
		workclass = Private: >50K (90.34/37.74)
		workclass = Self-emp-not-inc: <=50K (24.36/12.16)
		workclass = Self-emp-inc: <=50K (10.15/3.07)
		workclass = Federal-gov: >50K (3.05/1.02)


```

| | | | | | | | | | workclass = Local-gov
| | | | | | | | | |   hours-per-week <= 44: <=50K (8.11/1.03)
| | | | | | | | | |   hours-per-week > 44: >50K (3.06/0.02)
| | | | | | | | | | workclass = State-gov: >50K (3.05/1.02)
| | | | | | | | | | workclass = Without-pay: >50K (0.0)
| | | | | | | | | | workclass = Never-worked: >50K (0.0)
| | | | | | | | | | education = Some-college: >50K (0.0)
| | | | | | | | | | education = 11th: >50K (0.0)
| | | | | | | | | | education = HS-grad: >50K (0.0)
| | | | | | | | | | education = Prof-school: <=50K (5.39/2.22)
| | | | | | | | | | education = Assoc-acdm
| | | | | | | | | |   Age <= 37: <=50K (15.06/3.0)
| | | | | | | | | |   Age > 37
| | | | | | | | | |   fnlwgt <= 153052: <=50K (18.11/5.06)
| | | | | | | | | |   fnlwgt > 153052: >50K (26.06/5.06)
| | | | | | | | | | education = Assoc-voc: >50K (0.0)
| | | | | | | | | | education = 9th: >50K (0.0)
| | | | | | | | | | education = 7th-8th: >50K (0.0)
| | | | | | | | | | education = 12th: >50K (0.0)
| | | | | | | | | | education = Masters: >50K (20.11/8.5)
| | | | | | | | | | education = 1st-4th: >50K (0.0)
| | | | | | | | | | education = 10th: >50K (0.0)
| | | | | | | | | | education = Doctorate: <=50K (1.22/0.11)
| | | | | | | | | | education = 5th-6th: >50K (0.0)
| | | | | | | | | | education = Preschool: >50K (0.0)
| | | | | | | | | | occupation = Other-service: <=50K (57.01/10.49)
| | | | | | | | | | occupation = Sales
| | | | | | | | | | relationship = Wife: <=50K (18.92/7.4)
| | | | | | | | | | relationship = Own-child: >50K (2.0)
| | | | | | | | | | relationship = Husband: >50K (519.69/175.48)
| | | | | | | | | | relationship = Not-in-family: >50K (0.0)
| | | | | | | | | | relationship = Other-relative: <=50K (1.0)
| | | | | | | | | | relationship = Unmarried: >50K (0.0)
| | | | | | | | | | occupation = Exec-managerial
| | | | | | | | | |   Age <= 62: >50K (1132.64/264.89)
| | | | | | | | | |   Age > 62
| | | | | | | | | |   workclass = Private
| | | | | | | | | |   hours-per-week <= 46
| | | | | | | | | |   hours-per-week <= 42: >50K (14.21/4.56)
| | | | | | | | | |   hours-per-week > 42: <=50K (2.13)

```



```

| | | | | | | | | | hours-per-week > 46: >50K (5.52/1.13)
| | | | | | | | | | workclass = Self-emp-not-inc: <=50K (6.9/0.33)
| | | | | | | | | | workclass = Self-emp-inc
| | | | | | | | | | education-num <= 13: <=50K (10.15/4.33)
| | | | | | | | | | education-num > 13: >50K (3.66/0.33)
| | | | | | | | | | workclass = Federal-gov: >50K (1.15/0.1)
| | | | | | | | | | workclass = Local-gov: <=50K (2.3/0.11)
| | | | | | | | | | workclass = State-gov: <=50K (2.3/0.11)
| | | | | | | | | | workclass = Without-pay: <=50K (0.0)
| | | | | | | | | | workclass = Never-worked: <=50K (0.0)
| | | | | | | | | | occupation = Prof-specialty
| | | | | | | | | | education-num <= 14
| | | | | | | | | | hours-per-week <= 53
| | | | | | | | | | Capital-gain <= 1506: >50K (853.23/231.2)
| | | | | | | | | | Capital-gain > 1506
| | | | | | | | | | Capital-gain <= 4687: >50K (19.7/3.35)
| | | | | | | | | | Capital-gain > 4687
| | | | | | | | | | Capital-gain <= 5060: <=50K (7.0)
| | | | | | | | | | Capital-gain > 5060: >50K (10.35)
| | | | | | | | | | hours-per-week > 53
| | | | | | | | | | education = Bachelors
| | | | | | | | | | relationship = Wife: <=50K (6.35/1.35)
| | | | | | | | | | relationship = Own-child: >50K (0.0)
| | | | | | | | | | relationship = Husband: >50K (85.11/29.7)
| | | | | | | | | | relationship = Not-in-family: <=50K (1.0)
| | | | | | | | | | relationship = Other-relative: >50K (0.0)
| | | | | | | | | | relationship = Unmarried: >50K (0.0)
| | | | | | | | | | education = Some-college: >50K (0.0)
| | | | | | | | | | education = 11th: >50K (0.0)
| | | | | | | | | | education = HS-grad: >50K (0.0)
| | | | | | | | | | education = Prof-school: >50K (0.0)
| | | | | | | | | | education = Assoc-acdm: <=50K (7.35/1.0)
| | | | | | | | | | education = Assoc-voc: >50K (0.0)
| | | | | | | | | | education = 9th: >50K (0.0)
| | | | | | | | | | education = 7th-8th: >50K (0.0)
| | | | | | | | | | education = 12th: >50K (0.0)
| | | | | | | | | | education = Masters
| | | | | | | | | | workclass = Private
| | | | | | | | | | hours-per-week <= 65
| | | | | | | | | | Age <= 38: >50K (5.0/1.0)

```



```

| | | | | | | | | | | Age > 38: <=50K (20.3/8.3)
| | | | | | | | | | | hours-per-week > 65: <=50K (2.0)
| | | | | | | | | | | workclass = Self-emp-not-inc
| | | | | | | | | | | fnlwgt <= 194924: >50K (2.02)
| | | | | | | | | | | fnlwgt > 194924: <=50K (2.02/0.02)
| | | | | | | | | | | workclass = Self-emp-inc: >50K (3.03/1.0)
| | | | | | | | | | | workclass = Federal-gov: >50K (0.0)
| | | | | | | | | | | workclass = Local-gov: >50K (19.21/6.0)
| | | | | | | | | | | workclass = State-gov
| | | | | | | | | | | Age <= 40: <=50K (4.06/1.06)
| | | | | | | | | | | Age > 40: >50K (7.06)
| | | | | | | | | | | workclass = Without-pay: >50K (0.0)
| | | | | | | | | | | workclass = Never-worked: >50K (0.0)
| | | | | | | | | | | education = 1st-4th: >50K (0.0)
| | | | | | | | | | | education = 10th: >50K (0.0)
| | | | | | | | | | | education = Doctorate: >50K (0.0)
| | | | | | | | | | | education = 5th-6th: >50K (0.0)
| | | | | | | | | | | education = Preschool: >50K (0.0)
| | | | | | | | | | | education-num > 14: >50K (391.87/62.76)
| | | | | | | | | | | occupation = Handlers-cleaners: <=50K (31.56/10.27)
| | | | | | | | | | | occupation = Machine-op-inspct: <=50K (51.92/22.44)
| | | | | | | | | | | occupation = Adm-clerical
| | | | | | | | | | | race = White
| | | | | | | | | | | relationship = Wife
| | | | | | | | | | | Age <= 57: >50K (32.18/7.0)
| | | | | | | | | | | Age > 57: <=50K (3.12)
| | | | | | | | | | | relationship = Own-child: <=50K (1.0)
| | | | | | | | | | | relationship = Husband
| | | | | | | | | | | Age <= 58
| | | | | | | | | | | Age <= 41: <=50K (54.29/22.06)
| | | | | | | | | | | Age > 41
| | | | | | | | | | | Capital-gain <= 3942: >50K (84.11/24.47)
| | | | | | | | | | | Capital-gain > 3942
| | | | | | | | | | | Capital-gain <= 5060: <=50K (3.0)
| | | | | | | | | | | Capital-gain > 5060: >50K (2.0)
| | | | | | | | | | | Age > 58: <=50K (14.16/4.99)
| | | | | | | | | | | relationship = Not-in-family: >50K (0.0)
| | | | | | | | | | | relationship = Other-relative: >50K (1.0)
| | | | | | | | | | | relationship = Unmarried: >50K (0.0)
| | | | | | | | | | | race = Asian-Pac-Islander: >50K (31.12/15.12)

```



```

| | | | | | | race = Amer-Indian-Eskimo: >50K (1.06)
| | | | | | | race = Other: >50K (1.0)
| | | | | | | race = Black: >50K (12.23/3.12)
| | | | | | | occupation = Farming-fishing
| | | | | | | workclass = Private: <=50K (20.36/6.17)
| | | | | | | workclass = Self-emp-not-inc: <=50K (41.74/10.36)
| | | | | | | workclass = Self-emp-inc: >50K (6.11/0.06)
| | | | | | | workclass = Federal-gov: <=50K (2.04/1.02)
| | | | | | | workclass = Local-gov: <=50K (0.0)
| | | | | | | workclass = State-gov: >50K (1.02/0.01)
| | | | | | | workclass = Without-pay: <=50K (1.02/0.01)
| | | | | | | workclass = Never-worked: <=50K (0.0)
| | | | | | | occupation = Transport-moving: <=50K (54.98/20.47)
| | | | | | | occupation = Priv-house-serv: <=50K (2.04/0.02)
| | | | | | | occupation = Protective-serv
| | | | | | | Age <= 42
| | | | | | | Capital-gain <= 1506
| | | | | | | | race = White: <=50K (25.15/8.06)
| | | | | | | | race = Asian-Pac-Islander: >50K (1.0)
| | | | | | | | race = Amer-Indian-Eskimo: <=50K (0.0)
| | | | | | | | race = Other: <=50K (0.0)
| | | | | | | | race = Black
| | | | | | | | Age <= 39: <=50K (2.0)
| | | | | | | | Age > 39: >50K (2.0)
| | | | | | | Capital-gain > 1506: >50K (2.02/0.02)
| | | | | | | Age > 42: >50K (44.19/7.59)
| | | | | | | occupation = Armed-Forces: >50K (1.02/0.01)
| | Capital-loss > 1844
| | | Capital-loss <= 1980: >50K (857.0/18.0)
| | | Capital-loss > 1980
| | | | Capital-loss <= 2163: <=50K (104.0)
| | | | Capital-loss > 2163
| | | | | Capital-loss <= 2415
| | | | | Capital-loss <= 2377
| | | | | Age <= 64: <=50K (38.0/4.0)
| | | | | Age > 64: >50K (30.0/3.0)
| | | | | Capital-loss > 2377: >50K (82.0)
| | | | | Capital-loss > 2415: <=50K (14.0)
| marital-status = Divorced
| | education-num <= 12: <=50K (5220.0/225.0)

```



```

| | education-num > 12
| | | hours-per-week <= 42
| | | | Capital-loss <= 1726: <=50K (722.0/87.0)
| | | | Capital-loss > 1726
| | | | | Capital-loss <= 2415: <=50K (21.0/5.0)
| | | | | Capital-loss > 2415: >50K (7.0/1.0)
| | | hours-per-week > 42
| | | | occupation = Tech-support: <=50K (4.07/1.0)
| | | | occupation = Craft-repair: <=50K (12.2/2.0)
| | | | occupation = Other-service: <=50K (13.22/6.0)
| | | | occupation = Sales: <=50K (72.19/25.0)
| | | | occupation = Exec-managerial
| | | | | Capital-loss <= 1485
| | | | | race = White
| | | | | education = Bachelors
| | | | | | workclass = Private
| | | | | | relationship = Wife: <=50K (0.0)
| | | | | | relationship = Own-child: <=50K (2.0/1.0)
| | | | | | relationship = Husband: <=50K (0.0)
| | | | | | relationship = Not-in-family
| | | | | | sex = Female: >50K (17.67/6.67)
| | | | | | sex = Male: <=50K (22.45/10.0)
| | | | | | relationship = Other-relative: <=50K (1.0)
| | | | | | relationship = Unmarried
| | | | | | sex = Female: <=50K (9.0/2.0)
| | | | | | sex = Male: >50K (5.0/1.0)
| | | | | | workclass = Self-emp-not-inc
| | | | | | Age <= 33: >50K (2.03/0.03)
| | | | | | Age > 33: <=50K (6.13)
| | | | | | workclass = Self-emp-inc
| | | | | | hours-per-week <= 57
| | | | | | sex = Female: <=50K (2.06)
| | | | | | sex = Male
| | | | | | | Age <= 44: <=50K (2.06)
| | | | | | | Age > 44: >50K (2.0)
| | | | | | | hours-per-week > 57: >50K (2.03/0.03)
| | | | | | | workclass = Federal-gov
| | | | | | | hours-per-week <= 52: <=50K (2.06)
| | | | | | | hours-per-week > 52: >50K (2.02/0.02)
| | | | | | | workclass = Local-gov: <=50K (0.0)

```


[illegible]


```

| | | | | | | | hours-per-week > 52: >50K (9.0)
| | | | | | | | relationship = Other-relative: >50K (0.0)
| | | | | | | | relationship = Unmarried
| | | | | | | | sex = Female: <=50K (4.0)
| | | | | | | | sex = Male: >50K (5.0/1.0)
| | | | | | | | Capital-gain > 1086: <=50K (2.4)
| | | | | | | | occupation = Handlers-cleaners: <=50K (2.03/1.0)
| | | | | | | | occupation = Machine-op-inspct: <=50K (2.03)
| | | | | | | | occupation = Adm-clerical: <=50K (13.22/1.0)
| | | | | | | | occupation = Farming-fishing: <=50K (4.07)
| | | | | | | | occupation = Transport-moving: <=50K (7.12/1.0)
| | | | | | | | occupation = Priv-house-serv: <=50K (1.02)
| | | | | | | | occupation = Protective-serv
| | | | | | | | race = White: <=50K (6.17/2.0)
| | | | | | | | race = Asian-Pac-Islander: >50K (0.0)
| | | | | | | | race = Amer-Indian-Eskimo: >50K (0.0)
| | | | | | | | race = Other: >50K (0.0)
| | | | | | | | race = Black: >50K (4.0)
| | | | | | | | occupation = Armed-Forces: <=50K (0.0)
| | | | | | | | marital-status = Never-married
| | | | | | | | education-num <= 12
| | | | | | | | Capital-loss <= 2206: <=50K (12461.0/146.0)
| | | | | | | | Capital-loss > 2206
| | | | | | | | Capital-loss <= 2377: <=50K (17.0/3.0)
| | | | | | | | Capital-loss > 2377: >50K (11.0/1.0)
| | | | | | | | education-num > 12
| | | | | | | | education-num <= 14
| | | | | | | | hours-per-week <= 43: <=50K (2266.0/117.0)
| | | | | | | | hours-per-week > 43
| | | | | | | | Capital-loss <= 1485: <=50K (887.0/155.0)
| | | | | | | | Capital-loss > 1485
| | | | | | | | Capital-loss <= 1573: >50K (11.0/2.0)
| | | | | | | | Capital-loss > 1573
| | | | | | | | Capital-loss <= 2377: <=50K (35.0/3.0)
| | | | | | | | Capital-loss > 2377: >50K (6.0)
| | | | | | | | education-num > 14
| | | | | | | | Age <= 32: <=50K (82.0/9.0)
| | | | | | | | Age > 32
| | | | | | | | Capital-loss <= 1485
| | | | | | | | hours-per-week <= 26: <=50K (7.0)

```



```

| | | | | | | | | | hours-per-week > 26
| | | | | | | | | | occupation = Tech-support: >50K (0.0)
| | | | | | | | | | occupation = Craft-repair: >50K (1.02/0.01)
| | | | | | | | | | occupation = Other-service: <=50K (2.04/1.02)
| | | | | | | | | | occupation = Sales: <=50K (1.02/0.01)
| | | | | | | | | | occupation = Exec-managerial
| | | | | | | | | | Age <= 36: <=50K (2.13/0.13)
| | | | | | | | | | Age > 36: >50K (12.13/1.13)
| | | | | | | | | | occupation = Prof-specialty
| | | | | | | | | | Age <= 52
| | | | | | | | | | workclass = Private
| | | | | | | | | | sex = Female: >50K (10.0/3.0)
| | | | | | | | | | sex = Male
| | | | | | | | | | hours-per-week <= 43: >50K (10.0/1.0)
| | | | | | | | | | hours-per-week > 43: <=50K (7.32/1.32)
| | | | | | | | | | workclass = Self-emp-not-inc
| | | | | | | | | | sex = Female
| | | | | | | | | | hours-per-week <= 52: <=50K (3.0)
| | | | | | | | | | hours-per-week > 52: >50K (2.0)
| | | | | | | | | | sex = Male: >50K (13.21/2.0)
| | | | | | | | | | workclass = Self-emp-inc: >50K (2.02/1.0)
| | | | | | | | | | workclass = Federal-gov: >50K (5.06/1.0)
| | | | | | | | | | workclass = Local-gov
| | | | | | | | | | Age <= 46: >50K (3.06/1.0)
| | | | | | | | | | Age > 46: <=50K (2.0)
| | | | | | | | | | workclass = State-gov: <=50K (11.13/2.13)
| | | | | | | | | | workclass = Without-pay: >50K (0.0)
| | | | | | | | | | workclass = Never-worked: >50K (0.0)
| | | | | | | | | | Age > 52: <=50K (17.79/3.0)
| | | | | | | | | | occupation = Handlers-cleaners: >50K (0.0)
| | | | | | | | | | occupation = Machine-op-inspct: >50K (1.02/0.01)
| | | | | | | | | | occupation = Adm-clerical: >50K (1.02/0.01)
| | | | | | | | | | occupation = Farming-fishing: <=50K (2.04/0.02)
| | | | | | | | | | occupation = Transport-moving: >50K (0.0)
| | | | | | | | | | occupation = Priv-house-serv: >50K (0.0)
| | | | | | | | | | occupation = Protective-serv: >50K (0.0)
| | | | | | | | | | occupation = Armed-Forces: >50K (0.0)
| | | | | Capital-loss > 1485: >50K (16.0/2.0)
| marital-status = Separated
| | education-num <= 12: <=50K (1318.0/36.0)

```



```

| | education-num > 12
| | | education = Bachelors: <=50K (129.0/20.0)
| | | education = Some-college: <=50K (0.0)
| | | education = 11th: <=50K (0.0)
| | | education = HS-grad: <=50K (0.0)
| | | education = Prof-school
| | | | race = White: >50K (7.0/1.0)
| | | | race = Asian-Pac-Islander: >50K (0.0)
| | | | race = Amer-Indian-Eskimo: >50K (0.0)
| | | | race = Other: >50K (0.0)
| | | | race = Black: <=50K (2.0)
| | | education = Assoc-acdm: <=50K (0.0)
| | | education = Assoc-voc: <=50K (0.0)
| | | education = 9th: <=50K (0.0)
| | | education = 7th-8th: <=50K (0.0)
| | | education = 12th: <=50K (0.0)
| | | education = Masters
| | | | sex = Female: <=50K (24.0/2.0)
| | | | sex = Male
| | | | | Age <= 45: >50K (5.0)
| | | | | Age > 45: <=50K (10.0/3.0)
| | | education = 1st-4th: <=50K (0.0)
| | | education = 10th: <=50K (0.0)
| | | education = Doctorate
| | | | fnlwgt <= 192286: <=50K (5.0)
| | | | fnlwgt > 192286: >50K (5.0/1.0)
| | | education = 5th-6th: <=50K (0.0)
| | | education = Preschool: <=50K (0.0)
| marital-status = Widowed
| | Capital-loss <= 2205: <=50K (1460.0/82.0)
| | Capital-loss > 2205
| | | race = White: >50K (23.0/9.0)
| | | race = Asian-Pac-Islander: >50K (0.0)
| | | race = Amer-Indian-Eskimo: >50K (0.0)
| | | race = Other: >50K (0.0)
| | | race = Black: <=50K (2.0)
| marital-status = Married-spouse-absent: <=50K (613.0/44.0)
| marital-status = Married-AF-spouse: <=50K (35.0/12.0)
Capital-gain > 6849
| hours-per-week <= 35

```



```
| | Age <= 27
| | | Capital-gain <= 22040: >50K (3.0)
| | | Capital-gain > 22040: <=50K (6.0)
| | Age > 27: >50K (165.0/6.0)
| hours-per-week > 35: >50K (1881.0/16.0)
```

