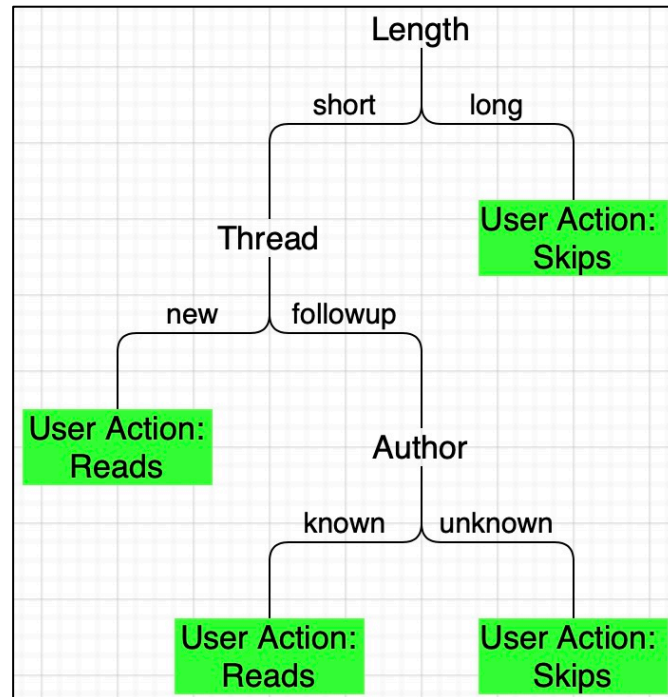


## Assignment 2 – Machine learning

### Question 1- Decision Trees

Tree when maximum gain split is used:



This tree is made in accordance with the maximum information gain split rules and formulas. After doing the calculations which are attached at the end of the assignment, we get:

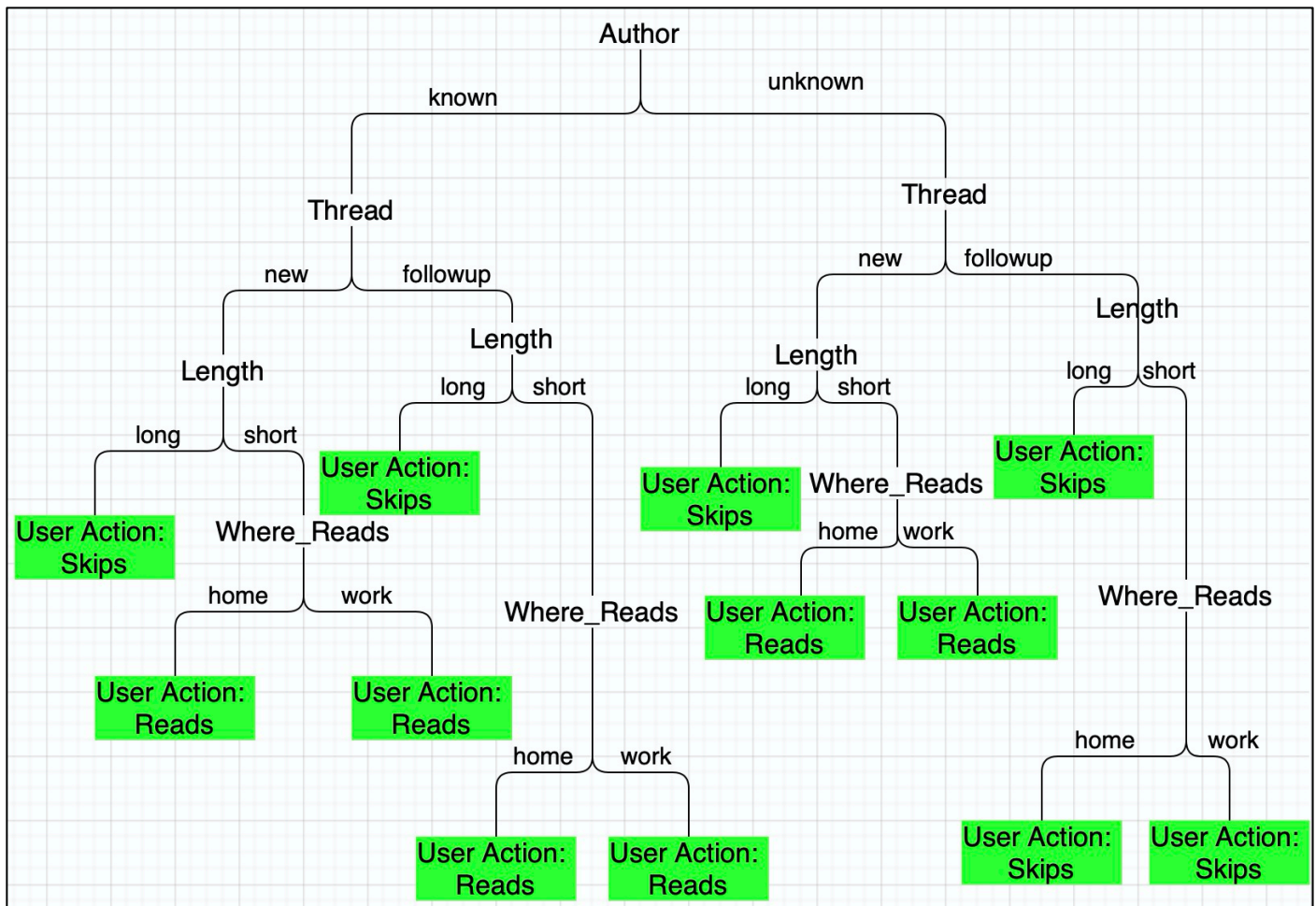
- $\text{Entropy}(\text{User\_Action}) = 1$
- $\text{Entropy}(\text{Author}) = \text{Entropy}(\text{Where\_Read}) = 1$
- $\text{Entropy}(\text{Thread}) = 0.85$
- $\text{Entropy}(\text{Length}) = 0.418$

Hence the corresponding Information gains are:

- $\text{Information gain}(\text{Author}) = \text{Information gain}(\text{Where\_Read}) = 1 - 1 = 0$
- $\text{Information gain}(\text{Thread}) = 1 - 0.85 = 0.15$
- $\text{Information gain}(\text{Length}) = 1 - 0.418 = 0.582$

Since we see that length has highest IG, it comes as the root node in the decision tree and then followed by Thread and then Author. Author can be replaced by Where\_Reads and the tree will look same, since they both have equal IG.

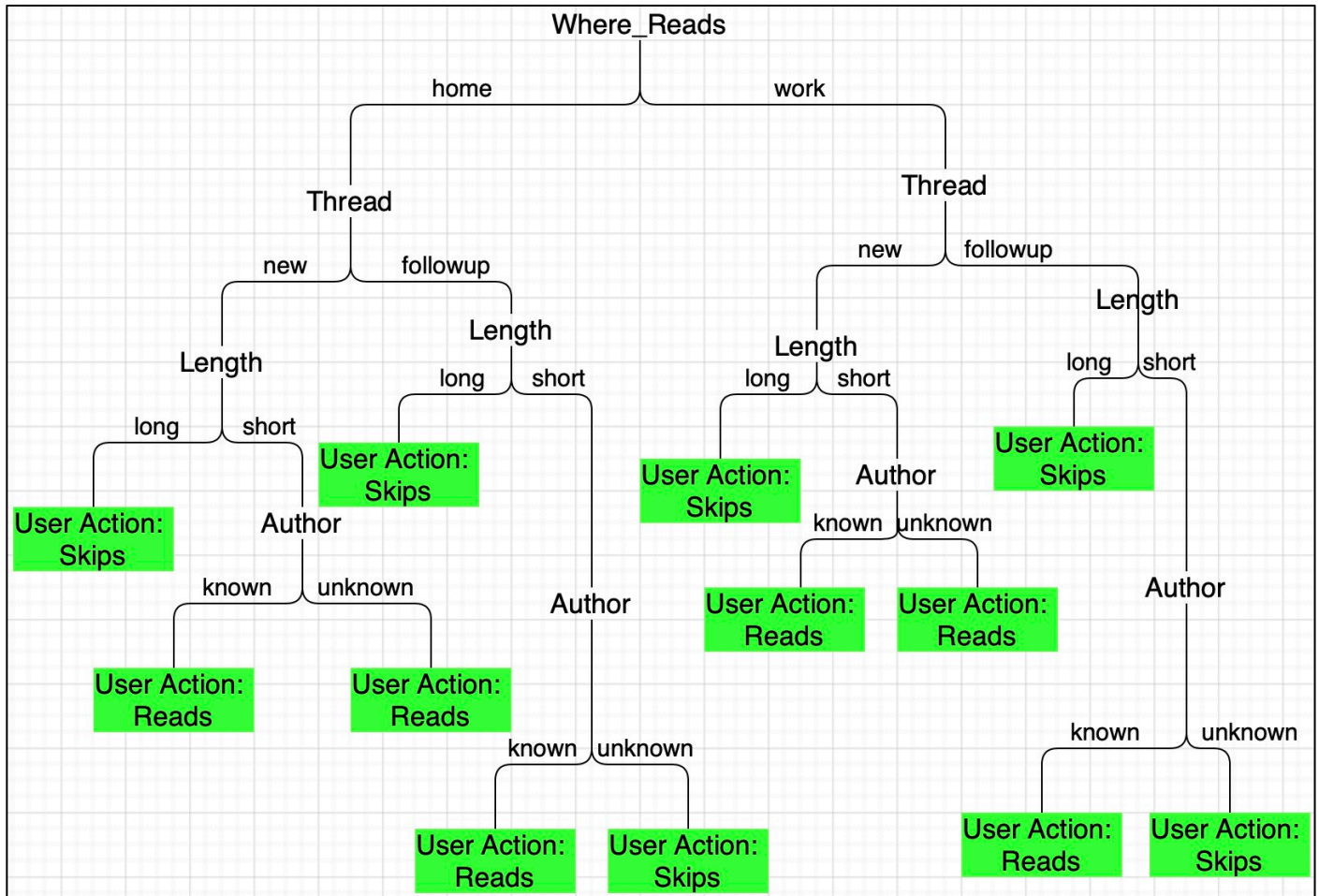
- a. If we change the algorithm to select the features in order (Author, Thread, Length, Where\_Read), the tree would look like this:



The tree is very different from what we say in information gain split.

- This tree is very complex and complicated compared to the IG split tree. It has a lot of repeated nodes and even repeated sub-trees.
- The sub-tree Where\_Reads is practically serving no purpose, because it further splits into home and work which gives the same decision 'reads' or 'skips'.
- The branch 'length' appears 4 times in this tree behaving the nearly same at every node. When the length is long, the user skips irrelevant of the Thread or Author. Similarly, when the length is short, the user always reads when the author is known and skips if the it's a follow-up from an unknown author.
- Though this tree is less compact and seems less efficient compared to the IG split tree, it represents an equivalent algorithm since the output for a given input is same in this tree.

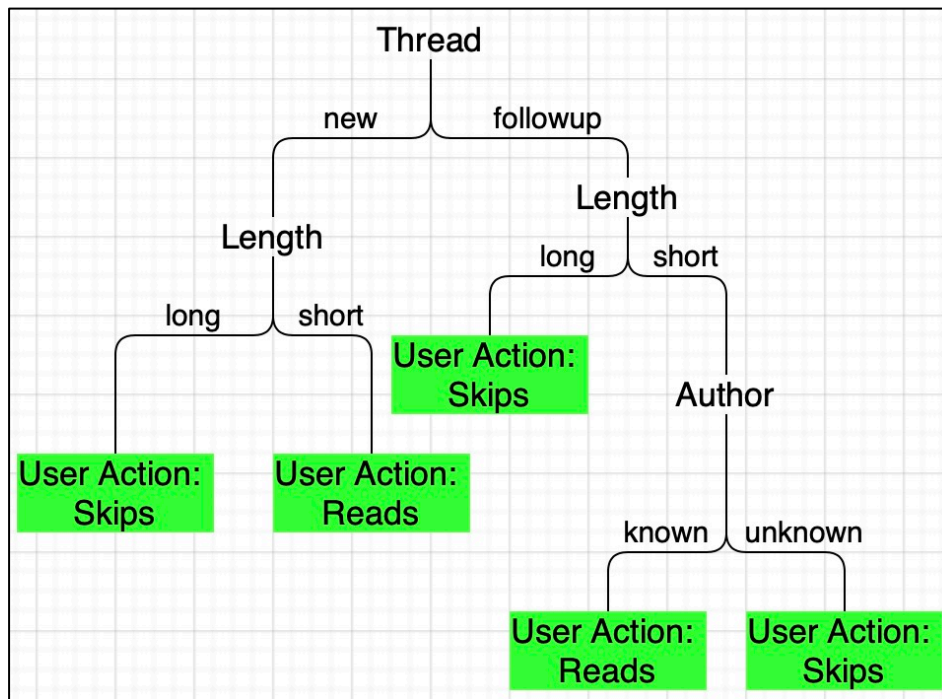
- b. When the features are in order (Where\_Reads, Thread, Length, Author) the tree would like this:



There are a lot of differences between this and IG split tree just like the one in previous question, but the tree in the previous question is similar to this tree in a lot of ways.

- The number of nodes and sub-trees are exactly same as the tree from previous question. It even looks similar to it. The reason being, the  $IG(\text{Author}) = IG(\text{Where\_Reads})$ . Swapping Author and Where\_Reads doesn't affect the tree as their IGs are equal. Even in the IG split tree, only Author is there because it is sufficient to represent.
- The sub-tree of Thread is equal in both 'home' and 'work' branches.
- Just like the 'Where\_Reads' in previous question, 'Author' sub tree serves no purpose when the 'length' is short and it's a new 'thread'. When the 'length' is short in a 'follow-up' thread 'Author' sub-tree gives same decisions in both 'home' and 'work' cases.
- If the IGs of Author and Where\_Reads were to be different, then this tree would've been different from the one before.

- Just like the previous tree, this tree is less compact and seems less efficient compared to the IG split tree, it represents an equivalent algorithm since the output for a given input is same in this tree.
- c. This tree can be used to correctly classify the training examples:



- Though this tree is not perfect like the one made using the IG split, it serves the purpose of making a decision tree much better than the ones used in previous questions.
- It is small, not complicated and contains very less repetition of nodes and sub-trees.
- The only drawback of this tree is that length appears twice while behaving the same when its long in both the 'new' and 'followup' branches.

## Question 2- Decision Trees

In cross validation, the best result is given when 10 folds are used. The following is the output WEKA produced:

Skip to page 28 for the end of this output.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: adult.data

Instances: 48842

Attributes: 15

- age
- workclass
- fnlwgt
- education
- education-num
- marital-status
- occupation
- relationship
- race
- sex
- capital-gain
- capital-loss
- hours-per-week
- native-country
- class

Test mode: 10-fold cross-validation

=== 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
| | | | | | | | | | | | | | hours-per-week <= 51
| | | | | | | | | | | | | | | age <= 59: >50K (40.36/13.19)
| | | | | | | | | | | | | | | age > 59: <=50K (10.19/1.34)
| | | | | | | | | | | | | | | hours-per-week > 51: <=50K (6.17/0.17)
| | | | | | | | | | | | | | | fnlwgt > 152968
| | | | | | | | | | | | | | | | hours-per-week <= 47
| | | | | | | | | | | | | | | | | education = Bachelors: <=50K (0.0)
| | | | | | | | | | | | | | | | | education = Some-college: <=50K (184.92/78.36)
| | | | | | | | | | | | | | | | | 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 <= 192344: <=50K (22.0/6.0)
| | | | | | | | | | | | | | | | | | fnlwgt > 192344: >50K (39.17/14.17)
| | | | | | | | | | | | | | | | | 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)

```







												hours-per-week <= 49: >50K (31.17/4.68)
												hours-per-week > 49
												fnlwgt <= 107897: >50K (4.0/1.0)
												fnlwgt > 107897: <=50K (6.29/0.1)
												relationship = Own-child: <=50K (3.1/1.0)
												relationship = Husband: >50K (299.57/117.61)
												relationship = Not-in-family: >50K (0.0)
												relationship = Other-relative: >50K (2.0/1.0)
												relationship = Unmarried: >50K (0.0)
												workclass = Self-emp-not-inc: <=50K (68.41/21.5)
												workclass = Self-emp-inc: >50K (91.89/32.22)
												workclass = Federal-gov: >50K (25.53/6.34)
												workclass = Local-gov
												race = White
												education = Bachelors: <=50K (0.0)
												education = Some-college: <=50K (26.55/12.15)
												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 <= 234428: >50K (7.12/2.06)
												fnlwgt > 234428: <=50K (2.05/0.03)
												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)
												race = Asian-Pac-Islander: >50K (0.05/0.01)
												race = Amer-Indian-Eskimo: >50K (3.01/1.01)
												race = Other: <=50K (0.01)
												race = Black: >50K (2.06/0.03)
												workclass = State-gov
												relationship = Wife: >50K (3.1/1.06)
												relationship = Own-child: <=50K (0.01)
												relationship = Husband
												age <= 45: <=50K (9.11/1.03)
												age > 45: >50K (11.27/3.17)









						race = Amer-Indian-Eskimo: <=50K (1.0)
						race = Other: <=50K (1.0)
						race = Black: <=50K (6.0/1.0)
						sex = Male: <=50K (386.0/111.0)
						hours-per-week > 30
						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
| | | race = White: <=50K (49.0/22.0)
| | | race = Asian-Pac-Islander: <=50K (0.0)
| | | race = Amer-Indian-Eskimo: <=50K (0.0)
| | | race = Other: <=50K (0.0)
| | | race = Black: >50K (3.0)
| | age > 29: >50K (84.69/22.79)
| 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 (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: >50K (17.24/2.0)
| education = 1st-4th: >50K (0.0)
| education = 10th: >50K (0.0)
| education = Doctorate: <=50K (1.24/0.24)
| education = 5th-6th: >50K (0.0)
| education = Preschool: >50K (0.0)
occupation = Prof-specialty
| workclass = Private
| | age <= 28
| | | education = Bachelors
| | | | hours-per-week <= 56: <=50K (20.79/6.0)
| | | | hours-per-week > 56: >50K (2.0)
| | | education = Some-college: <=50K (0.0)
| | | education = 11th: <=50K (0.0)
| | | education = HS-grad: <=50K (0.0)
| | | education = Prof-school
| | | | hours-per-week <= 55: >50K (2.0)
| | | | hours-per-week > 55: <=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

```





											occupation = Farming-fishing: <=50K (8.03/1.03)
											occupation = Transport-moving: >50K (2.01/1.0)
											occupation = Priv-house-serv: >50K (0.0)
											occupation = Protective-serv
											hours-per-week <= 69
											race = White
											fnlwgt <= 225339: <=50K (10.07/2.07)
											fnlwgt > 225339: >50K (4.0)
											race = Asian-Pac-Islander: <=50K (0.0)
											race = Amer-Indian-Eskimo: <=50K (0.0)
											race = Other: <=50K (0.0)
											race = Black: >50K (4.0/2.0)
											hours-per-week > 69: >50K (4.0)
											occupation = Armed-Forces: >50K (0.0)
											relationship = Not-in-family: >50K (0.0)
											relationship = Other-relative: <=50K (6.0/1.0)
											relationship = Unmarried: >50K (0.0)
											native-country = Cambodia: <=50K (1.02/0.01)
											native-country = England
											age <= 32: >50K (2.1/0.06)
											age > 32: <=50K (3.01)
											native-country = Puerto-Rico: <=50K (1.02/0.01)
											native-country = Canada: <=50K (4.09/1.03)
											native-country = Germany: <=50K (4.09/1.03)
											native-country = Outlying-US(Guam-USVI-etc): >50K (0.0)
											native-country = India
											workclass = Private: <=50K (10.16/2.07)
											workclass = Self-emp-not-inc: <=50K (0.06/0.01)
											workclass = Self-emp-inc: <=50K (0.0)
											workclass = Federal-gov: >50K (0.01)
											workclass = Local-gov: <=50K (0.01/0.0)
											workclass = State-gov: >50K (2.03/0.03)
											workclass = Without-pay: <=50K (0.0)
											workclass = Never-worked: <=50K (0.0)
											native-country = Japan: <=50K (6.14/1.05)
											native-country = Greece: >50K (1.02/0.01)
											native-country = South: <=50K (1.02/0.01)
											native-country = China
											fnlwgt <= 315203: <=50K (2.09/0.03)
											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

```



									hours-per-week > 67: <=50K (6.0/1.0)
									capital-loss > 0: <=50K (3.0)
									education = 1st-4th: <=50K (0.0)
									education = 10th: <=50K (0.0)
									education = Doctorate: >50K (4.0/1.0)
									education = 5th-6th: <=50K (0.0)
									education = Preschool: <=50K (0.0)
									race = Asian-Pac-Islander: <=50K (3.0/1.0)
									race = Amer-Indian-Eskimo: <=50K (1.0)
									race = Other: <=50K (0.0)
									race = Black
									fnlwgt <= 171306: >50K (5.0)
									fnlwgt > 171306: <=50K (4.0/1.0)
									capital-loss > 1485: >50K (10.0)
									occupation = Prof-specialty
									education-num <= 14: <=50K (159.81/32.0)
									education-num > 14
									capital-gain <= 1086
									relationship = Wife: >50K (0.0)
									relationship = Own-child: >50K (1.0)
									relationship = Husband: >50K (0.0)
									relationship = Not-in-family
									hours-per-week <= 52
									hours-per-week <= 47: >50K (6.0)
									hours-per-week > 47: <=50K (7.0/1.0)
									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)
```

Number of Leaves : 696

Size of the tree : 911

Time taken to build model: 2.56 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	42050	86.0939 %
Incorrectly Classified Instances	6792	13.9061 %
Kappa statistic	0.587	
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

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.600	0.057	0.768	0.600	0.674	0.594	0.890	0.758	>50K
	0.943	0.400	0.882	0.943	0.912	0.594	0.890	0.951	<=50K
Weighted Avg.	0.861	0.318	0.855	0.861	0.855	0.594	0.890	0.905	

=== Confusion Matrix ===

```

a  b  <-- classified as
7009 4678 |  a = >50K
2114 35041 |  b = <=50K

```

- This Method is fairly accurate as you can see 86.09% of the results are correctly classified.
- I took 10 fold validation because it gave the best accuracy while taking less time than others.
- The following is the accuracy output for 9 fold validation:

=== Summary ===

Correctly Classified Instances	42049	86.0919 %
Incorrectly Classified Instances	6793	13.9081 %
Kappa statistic	0.5869	
Mean absolute error	0.1958	
Root mean squared error	0.3204	
Relative absolute error	53.7745 %	
Root relative squared error	75.086 %	
Total Number of Instances	48842	

- This takes a few milliseconds more time that 10 fold while giving nearly the same accuracy for classification.
- The following is the accuracy output for 11 fold validation:

=== Summary ===

Correctly Classified Instances	42026	86.0448 %
Incorrectly Classified Instances	6816	13.9552 %
Kappa statistic	0.5846	
Mean absolute error	0.1964	
Root mean squared error	0.3205	
Relative absolute error	53.9435 %	
Root relative squared error	75.1315 %	
Total Number of Instances	48842	

- Though this takes a few milliseconds less than the 10 fold, it gives off the accuracy by a very few numbers.

So, in my opinion, 10 fold validation is the best pick while using cross validation in WEKA.

For the percentage split, 70% split (70% data for training, rest for testing) is the most appropriate. The following is the output for 70% data split:

Skip to page 52 for the end of this output.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: adult.data

Instances: 48842

Attributes: 15

age  
workclass  
fnlwgt  
education  
education-num  
marital-status  
occupation  
relationship  
race  
sex  
capital-gain  
capital-loss  
hours-per-week  
native-country  
class

Test mode: split 70.0% train, remainder test

=== 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
| | | | | | | | | | | | hours-per-week <= 51
| | | | | | | | | | | | | age <= 59: >50K (40.36/13.19)
| | | | | | | | | | | | | age > 59: <=50K (10.19/1.34)
| | | | | | | | | | | | | hours-per-week > 51: <=50K (6.17/0.17)
| | | | | | | | | | | | | fnlwgt > 152968
| | | | | | | | | | | | | hours-per-week <= 47
| | | | | | | | | | | | | | education = Bachelors: <=50K (0.0)
| | | | | | | | | | | | | | education = Some-college: <=50K (184.92/78.36)
| | | | | | | | | | | | | | 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 <= 192344: <=50K (22.0/6.0)
| | | | | | | | | | | | | | | fnlwgt > 192344: >50K (39.17/14.17)
| | | | | | | | | | | | | | | 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)

```

[illegible]

										occupation = Other-service: <=50K (144.99/32.06)
										occupation = Sales
										workclass = Private
										fnlwgt <= 89259: <=50K (67.76/17.19)
										fnlwgt > 89259
										age <= 40: <=50K (119.89/49.47)
										age > 40: >50K (147.25/56.46)
										workclass = Self-emp-not-inc
										education = Bachelors: <=50K (0.0)
										education = Some-college
										fnlwgt <= 345734: <=50K (52.01/13.27)
										fnlwgt > 345734: >50K (7.08/0.02)
										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
										age <= 46: <=50K (5.11/1.08)
										age > 46: >50K (3.19/0.1)
										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)
										workclass = Self-emp-inc
										hours-per-week <= 42
										hours-per-week <= 39: >50K (4.1/1.05)
										hours-per-week > 39: <=50K (11.69/2.16)
										hours-per-week > 42: >50K (42.41/12.2)
										workclass = Federal-gov: <=50K (0.0)
										workclass = Local-gov: <=50K (1.02/0.01)
										workclass = State-gov: <=50K (2.04/1.01)
										workclass = Without-pay: <=50K (0.0)
										workclass = Never-worked: <=50K (0.0)
										occupation = Exec-managerial
										workclass = Private
										relationship = Wife
										hours-per-week <= 49: >50K (31.17/4.68)



												hours-per-week > 49
												fnlwgt <= 107897: >50K (4.0/1.0)
												fnlwgt > 107897: <=50K (6.29/0.1)
												relationship = Own-child: <=50K (3.1/1.0)
												relationship = Husband: >50K (299.57/117.61)
												relationship = Not-in-family: >50K (0.0)
												relationship = Other-relative: >50K (2.0/1.0)
												relationship = Unmarried: >50K (0.0)
												workclass = Self-emp-not-inc: <=50K (68.41/21.5)
												workclass = Self-emp-inc: >50K (91.89/32.22)
												workclass = Federal-gov: >50K (25.53/6.34)
												workclass = Local-gov
												race = White
												education = Bachelors: <=50K (0.0)
												education = Some-college: <=50K (26.55/12.15)
												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 <= 234428: >50K (7.12/2.06)
												fnlwgt > 234428: <=50K (2.05/0.03)
												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)
												race = Asian-Pac-Islander: >50K (0.05/0.01)
												race = Amer-Indian-Eskimo: >50K (3.01/1.01)
												race = Other: <=50K (0.01)
												race = Black: >50K (2.06/0.03)
												workclass = State-gov
												relationship = Wife: >50K (3.1/1.06)
												relationship = Own-child: <=50K (0.01)
												relationship = Husband
												age <= 45: <=50K (9.11/1.03)
												age > 45: >50K (11.27/3.17)
												relationship = Not-in-family: <=50K (0.0)

[illegible]

		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)
		workclass = Self-emp-not-inc: >50K (4.08/1.06)
		workclass = Self-emp-inc: >50K (1.02/0.01)
		workclass = Federal-gov: >50K (45.95/16.63)
		workclass = Local-gov
		sex = Female: ≤50K (10.07/3.03)
		sex = Male: >50K (8.31/2.21)
		workclass = State-gov: ≤50K (11.23/4.08)
		workclass = Without-pay: ≤50K (0.0)
		workclass = Never-worked: ≤50K (0.0)
		race = Asian-Pac-Islander: >50K (8.34/2.08)
		race = Amer-Indian-Eskimo: ≤50K (4.08/2.0)
		race = Other: ≤50K (0.08)
		race = Black
		workclass = Private
		hours-per-week ≤ 42
		fnlwgt ≤ 285200: ≤50K (13.3/4.12)
		fnlwgt > 285200: >50K (5.0/1.0)
		hours-per-week > 42: >50K (6.0)
		workclass = Self-emp-not-inc: >50K (0.0)
		workclass = Self-emp-inc: >50K (0.0)
		workclass = Federal-gov: >50K (5.06/1.04)
		workclass = Local-gov: ≤50K (3.04/1.01)
		workclass = State-gov: ≤50K (2.02/0.01)
		workclass = Without-pay: >50K (0.0)
		workclass = Never-worked: >50K (0.0)
		occupation = Farming-fishing: ≤50K (146.01/30.07)
		occupation = Transport-moving
		workclass = Private: ≤50K (158.26/60.16)
		workclass = Self-emp-not-inc
		education = Bachelors: ≤50K (0.0)
		education = Some-college: ≤50K (11.21/2.06)
		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: >50K (2.06/0.03)
											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)
											workclass = Self-emp-inc: >50K (7.15/1.09)
											workclass = Federal-gov
											race = White: >50K (2.07/0.05)
											race = Asian-Pac-Islander: >50K (0.0/0.0)
											race = Amer-Indian-Eskimo: <=50K (0.0)
											race = Other: <=50K (0.0)
											race = Black: <=50K (2.01/0.0)
											workclass = Local-gov: <=50K (14.29/3.1)
											workclass = State-gov: <=50K (4.08/0.03)
											workclass = Without-pay: <=50K (0.0)
											workclass = Never-worked: <=50K (0.0)
											occupation = Priv-house-serv: <=50K (0.0)
											occupation = Protective-serv
											race = White
											workclass = Private: <=50K (19.39/2.13)
											workclass = Self-emp-not-inc: >50K (1.02/0.01)
											workclass = Self-emp-inc: >50K (1.02/0.01)
											workclass = Federal-gov
											age <= 46: <=50K (2.03/0.01)
											age > 46: >50K (2.06/0.03)
											workclass = Local-gov
											education = Bachelors: >50K (0.0)
											education = Some-college
											age <= 53
											age <= 38: <=50K (32.35/13.03)
											age > 38
											hours-per-week <= 36: <=50K (2.0)
											hours-per-week > 36
											fnlwgt <= 121253

```

| | | | | | | | | | | | | | | | | fnlwgt <= 39411: >50K (3.0/1.0)
| | | | | | | | | | | | | | | | | fnlwgt > 39411: <=50K (4.09/0.03)
| | | | | | | | | | | | | | | | | fnlwgt > 121253: >50K (27.21/3.15)
| | | | | | | | | | | | | | | | | age > 53: <=50K (5.77/0.27)
| | | | | | | | | | | | | | | | | 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 <= 112772: <=50K (2.12/0.06)
| | | | | | | | | | | | | | | | | fnlwgt > 112772: >50K (15.29/2.12)
| | | | | | | | | | | | | | | | | 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)
| | | | | | | | | | | | | | | | | workclass = State-gov
| | | | | | | | | | | | | | | | | age <= 41: <=50K (9.13/0.03)
| | | | | | | | | | | | | | | | | age > 41: >50K (15.35/6.22)
| | | | | | | | | | | | | | | | | workclass = Without-pay: <=50K (0.0)
| | | | | | | | | | | | | | | | | workclass = Never-worked: <=50K (0.0)
| | | | | | | | | | | | | | | | | race = Asian-Pac-Islander: <=50K (2.18/0.14)
| | | | | | | | | | | | | | | | | race = Amer-Indian-Eskimo: <=50K (1.05)
| | | | | | | | | | | | | | | | | race = Other: >50K (2.05/0.05)
| | | | | | | | | | | | | | | | | race = Black
| | | | | | | | | | | | | | | | | fnlwgt <= 154897: <=50K (3.14/0.05)
| | | | | | | | | | | | | | | | | fnlwgt > 154897: >50K (11.09/3.05)
| | | | | | | | | | | | | | | | | occupation = Armed-Forces: >50K (1.02/0.01)
| | | | | | | | | | | | | | | | | capital-loss > 1510: <=50K (43.0/1.0)
| | | | | | | | | | | | | | | | | capital-gain > 5060
| | | | | | | | | | | | | | | | | capital-gain <= 6612: >50K (111.0)
| | | | | | | | | | | | | | | | | capital-gain > 6612: <=50K (5.0)
| | | | | | | | | | | | | | | | | education-num > 11
| | | | | | | | | | | | | | | | | hours-per-week <= 30
| | | | | | | | | | | | | | | | | sex = Female
| | | | | | | | | | | | | | | | | race = White: >50K (145.0/54.0)
| | | | | | | | | | | | | | | | | race = Asian-Pac-Islander: >50K (6.0/3.0)
| | | | | | | | | | | | | | | | | race = Amer-Indian-Eskimo: <=50K (1.0)

```

						race = Other: <=50K (1.0)
						race = Black: <=50K (6.0/1.0)
						sex = Male: <=50K (386.0/111.0)
						hours-per-week > 30
						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



```
| | | | hours-per-week > 40: >50K (10.04/2.0)
| | | | education = 1st-4th: <=50K (8.94/1.0)
| | | | education = 10th: <=50K (8.94/1.0)
| | | | education = Doctorate: >50K (10.04/2.0)
| | | | education = 5th-6th: <=50K (8.94/1.0)
| | | | education = Preschool: <=50K (8.94/1.0)
| | | age > 28: >50K (117.38/3.0)
| | workclass = Self-emp-not-inc: <=50K (8.94/1.0)
| | workclass = Self-emp-inc: >50K (10.04/2.0)
| | workclass = Federal-gov: >50K (10.04/2.0)
| | workclass = Local-gov: <=50K (8.94/1.0)
| | workclass = State-gov: <=50K (8.94/1.0)
| | | | education = Bachelors: >50K (10.04/2.0)
| | | | education = Some-college: >50K (10.04/2.0)
| | | | education = 11th: >50K (10.04/2.0)
| | | | education = HS-grad: >50K (10.04/2.0)
| | | | education = Prof-school: >50K (10.04/2.0)
| | | | education = Assoc-acdm: >50K (10.04/2.0)
| | | | education = Assoc-voc: >50K (10.04/2.0)
| | | | education = 9th: >50K (10.04/2.0)
| | | | education = 7th-8th: >50K (10.04/2.0)
| | | | education = 12th: >50K (10.04/2.0)
| | | | education = Masters: >50K (10.04/2.0)
| | | | fnlwgt <= 192740: <=50K (8.94/1.0)
| | | | fnlwgt > 192740: >50K (10.04/2.0)
| | | | education = 1st-4th: >50K (10.04/2.0)
| | | | education = 10th: >50K (10.04/2.0)
| | | | education = Doctorate: >50K (10.04/2.0)
| | | | education = 5th-6th: >50K (10.04/2.0)
| | | | education = Preschool: >50K (10.04/2.0)
| | workclass = Without-pay: >50K (10.04/2.0)
| | workclass = Never-worked: >50K (10.04/2.0)
| | occupation = Handlers-cleaners: >50K (10.04/2.0)
| | | | hours-per-week <= 43: <=50K (8.94/1.0)
| | | | hours-per-week > 43: >50K (10.04/2.0)
| | occupation = Machine-op-inspct: >50K (10.04/2.0)
| | occupation = Adm-clerical: <=50K (8.94/1.0)
| | | | age <= 31: <=50K (8.94/1.0)
| | | | age <= 29: <=50K (8.94/1.0)
| | | | age > 29: >50K (10.04/2.0)
| | | | age > 31: <=50K (8.0/3.0)
| | occupation = Farming-fishing: <=50K (8.94/1.0)
```



											occupation = Transport-moving: >50K (2.01/1.0)
											occupation = Priv-house-serv: >50K (0.0)
											occupation = Protective-serv
											hours-per-week <= 69
											race = White
											fnlwgt <= 225339: <=50K (10.07/2.07)
											fnlwgt > 225339: >50K (4.0)
											race = Asian-Pac-Islander: <=50K (0.0)
											race = Amer-Indian-Eskimo: <=50K (0.0)
											race = Other: <=50K (0.0)
											race = Black: >50K (4.0/2.0)
											hours-per-week > 69: >50K (4.0)
											occupation = Armed-Forces: >50K (0.0)
											relationship = Not-in-family: >50K (0.0)
											relationship = Other-relative: <=50K (6.0/1.0)
											relationship = Unmarried: >50K (0.0)
											native-country = Cambodia: <=50K (1.02/0.01)
											native-country = England
											age <= 32: >50K (2.1/0.06)
											age > 32: <=50K (3.01)
											native-country = Puerto-Rico: <=50K (1.02/0.01)
											native-country = Canada: <=50K (4.09/1.03)
											native-country = Germany: <=50K (4.09/1.03)
											native-country = Outlying-US(Guam-USVI-etc): >50K (0.0)
											native-country = India
											workclass = Private: <=50K (10.16/2.07)
											workclass = Self-emp-not-inc: <=50K (0.06/0.01)
											workclass = Self-emp-inc: <=50K (0.0)
											workclass = Federal-gov: >50K (0.01)
											workclass = Local-gov: <=50K (0.01/0.0)
											workclass = State-gov: >50K (2.03/0.03)
											workclass = Without-pay: <=50K (0.0)
											workclass = Never-worked: <=50K (0.0)
											native-country = Japan: <=50K (6.14/1.05)
											native-country = Greece: >50K (1.02/0.01)
											native-country = South: <=50K (1.02/0.01)
											native-country = China
											fnlwgt <= 315203: <=50K (2.09/0.03)
											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)

[illegible]

```
| | | | | 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)
								workclass = State-gov: >50K (1.02/0.02)
								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 (1.31)
								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
								capital-loss <= 0
								hours-per-week <= 67: >50K (32.61/11.61)
								hours-per-week > 67: <=50K (6.0/1.0)

								capital-loss > 0: <=50K (3.0)
								education = 1st-4th: <=50K (0.0)
								education = 10th: <=50K (0.0)
								education = Doctorate: >50K (4.0/1.0)
								education = 5th-6th: <=50K (0.0)
								education = Preschool: <=50K (0.0)
								race = Asian-Pac-Islander: <=50K (3.0/1.0)
								race = Amer-Indian-Eskimo: <=50K (1.0)
								race = Other: <=50K (0.0)
								race = Black
								fnlwgt <= 171306: >50K (5.0)
								fnlwgt > 171306: <=50K (4.0/1.0)
								capital-loss > 1485: >50K (10.0)
								occupation = Prof-specialty
								education-num <= 14: <=50K (159.81/32.0)
								education-num > 14
								capital-gain <= 1086
								relationship = Wife: >50K (0.0)
								relationship = Own-child: >50K (1.0)
								relationship = Husband: >50K (0.0)
								relationship = Not-in-family
								hours-per-week <= 52
								hours-per-week <= 47: >50K (6.0)
								hours-per-week > 47: <=50K (7.0/1.0)
								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)
```

Number of Leaves : 696

Size of the tree : 911

Time taken to build model: 2.58 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	12631	86.2008 %
Incorrectly Classified Instances	2022	13.7992 %
Kappa statistic	0.584	
Mean absolute error	0.195	
Root mean squared error	0.3197	

Relative absolute error            53.7039 %  
 Root relative squared error        75.3124 %  
 Total Number of Instances        14653

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
>50K	0.593	0.055	0.769	0.593	0.669	0.592	0.881	0.744
<=50K	0.945	0.407	0.883	0.945	0.913	0.592	0.881	0.944
Weighted Avg.	0.862	0.324	0.856	0.862	0.855	0.592	0.881	0.897

=== Confusion Matrix ===

```

a   b  <-- classified as
2047 1407 |   a = >50K
615 10584 |   b = <=50K
    
```

I chose 70% split due to the following reasons:

- Though 70% doesn't give the best possible accuracy, it's still the best split.
- The best possible accuracy is achieved by 87% split. To be more precise by 87.1%.  
 The following is the summary for 87.1% split

=== Summary ===

Correctly Classified Instances        5454            86.5577 %  
 Incorrectly Classified Instances    847            13.4423 %  
 Kappa statistic                    0.5957  
 Mean absolute error                0.1957  
 Root mean squared error            0.3159  
 Relative absolute error            54.0071 %  
 Root relative squared error        74.4688 %  
 Total Number of Instances        6301

- As you can see the accuracy is not much different, its only 0.3% more than the 70% split. Here, the number of instances in the testing data changes significantly. For 70% it was 14653 while here its only 6301, nearly 8000 more testing cases.
- 8000 test cases might seem less, but in a database of 45000 cases, 8000 is a lot. If you remove them, 70% split will give nearly the same, if not better accuracy than the 87.1%
- Any split above 87% reduces the accuracy while the testing cases get reduced by a lot. The following is the summary for 90% split:

=== Summary ===

Correctly Classified Instances	4172	85.4218 %
Incorrectly Classified Instances	712	14.5782 %
Kappa statistic	0.5626	
Mean absolute error	0.205	
Root mean squared error	0.3269	
Relative absolute error	56.3692 %	
Root relative squared error	76.6959 %	
Total Number of Instances	4884	

- As you can see the accuracy is less and this is even less than a 65% split. Furthermore number of cases used for testing is only 10% of the data base provided.
- There is no point in taking time into consideration for this data base as all the splits only differ by a few milliseconds in giving results.
- So, looking at all the points and taking in consideration accuracy and amount of data used for testing, 70% gives the best possible split.

