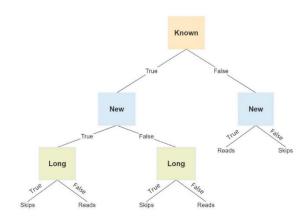
Assignment 2 - Machine Learning

Part 1: Decision Trees

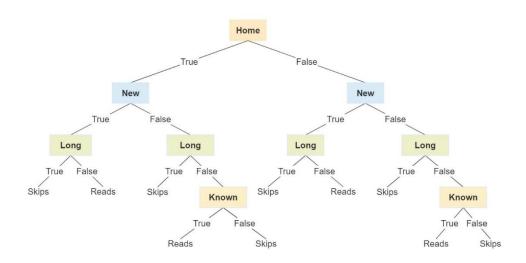
Question 1.1

a)



Yes, this tree represents a different function than that found with the maximum information gain splits. If we use e_{19} on this tree, the result we get is "reads" instead of "skips" if we were to use the maximum information gain splits.

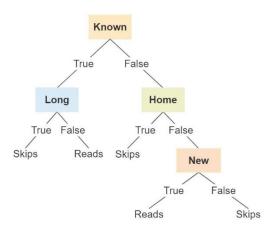
b)



This tree represents the same function as the tree found with the maximum information gain split. However, this tree represents a different function than the one in part (a). As mentioned in part (a), example such as $e_{19} = [unknown, new, long, work]$ will result in a

different answer if we follow the order [Author; Thread, Length, WhereRead] than [WhereRead; Thread, Length, Author]

c) Below is an example of a tree representing a different function than those found by the preceding algorithms but correctly classifies the training examples



Question 1.2

I created an ARFF file that contains both the training and test and loaded it to Weka. In the ARFF file, I included the 15 attributes: age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country, and **salary** which determined whether it was <=50K or >50K based on the inputs given. After the ARFF file is loaded to Weka, I applied the **Filter -> ReplaceMissingValues** to replace all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data.

I then went to Classify tab and ran J48 classifier on the dataset. The different settings I applied to the dataset are:

- Pruned the tree with cross-validation test: 86.0038% in accuracy and the size of tree is 873.
- Pruned the tree with 66% training data and 34% test data: **85.8184%** in accuracy and the size of the tree is **1064.**
- Unpruned the tree with cross-validation test: 84.454 % in accuracy and the size
 of the tree is 9038.
- Unpruned the tree with 66% training data and 34% test data: **83.9215%** in accuracy and the size of the tree is **7693**.

- Pruned the tree with subtree not raising and cross-validation test: **85.9363**% in accuracy and the size of the tree is **1497.**
- Pruned the tree with 90% training data and 10% test data: **85.258%** in accuracy and the size of the tree is **763.**

Conclusion:

As I was experimenting with different settings, these are the few things I noticed:

- The size of the decision tree is significantly increased but the percentage accuracy decreased by 2% when the tree is unpruned.
- With the subtree not raising, the size of the pruned tree is significantly increased, and the percentage of accuracy decreased a little
- The size of the tree is decreased when the percentage of the training data is being used increased.

Overall, I found that with the tree is pruned in the cross-validation test setting gave the highest percentage in accuracy and smallest in tree size. The result of the pruned tree is showed below (I took out Classifier models for fold 1 to 9).

```
=== Run information ===
```

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

Relation: adults-weka.filters.unsupervised.attribute.ReplaceMissingValues

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

salary

```
Test mode: 10-fold cross-validation
=== Classifier model for fold 10 ===
J48 pruned tree
capital-gain <= 6849
| marital-status = Married-civ-spouse
| | capital-loss <= 1844
  | | education-num <= 11
    | | capital-gain <= 5060
        | age <= 29: <=50K (1808.0/220.0)
             age > 29
             hours-per-week <= 34: <=50K (1115.0/143.0)
             | hours-per-week > 34
                | education-num <= 9: <=50K (6276.0/1651.0)
                  education-num > 9
                  | capital-gain <= 4416
                     | capital-loss <= 1510
                       occupation = Tech-support
                             age <= 37: <=50K (54.0/18.0)
                          | age > 37: >50K (101.0/33.0)
                          occupation = Craft-repair
                             race = White: <=50K (637.0/248.0)
                             race = Asian-Pac-Islander
                               age <= 47: <=50K (17.0/5.0)
                               age > 47: >50K (6.0)
                             race = Amer-Indian-Eskimo
                               hours-per-week <= 46: <=50K (3.0)
                             hours-per-week > 46: >50K (3.0/1.0)
                             race = Other: <=50K (2.0/1.0)
                             race = Black
                               age <= 45
                                hours-per-week <= 47: <=50K (15.0/4.0)
                                | hours-per-week > 47: >50K (3.0)
                            | age > 45: >50K (10.0/1.0)
                          occupation = Other-service: <=50K (130.0/30.0)
                          occupation = Sales
                            workclass = Private: <=50K (297.0/143.0)
                        | | workclass = Self-emp-not-inc
```

```
education = Bachelors: <=50K (0.0)
     education = Some-college
     fnlwgt <= 345734: <=50K (47.0/13.0)
       fnlwgt > 345734: >50K (6.0)
     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
     | age <= 46: <=50K (4.0/1.0)
     | age > 46: >50K (3.0)
     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 (3.0/1.0)
     hours-per-week > 39: <=50K (11.0/2.0)
  hours-per-week > 42: >50K (38.0/11.0)
  workclass = Federal-gov: <=50K (0.0)
  workclass = Local-gov: <=50K (1.0)
  workclass = State-gov: <=50K (2.0/1.0)
  workclass = Without-pay: <=50K (0.0)
  workclass = Never-worked: <=50K (0.0)
occupation = Exec-managerial
  workclass = Private
     relationship = Wife
       hours-per-week <= 47
          age <= 64: >50K (26.0/2.0)
          age > 64: <=50K (2.0)
       hours-per-week > 47
         fnlwgt <= 107897: >50K (4.0/1.0)
       | fnlwgt > 107897: <=50K (6.0)
     relationship = Own-child: <=50K (3.0/1.0)
     relationship = Husband
```

```
hours-per-week <= 41
        education = Bachelors: >50K (0.0)
        education = Some-college: >50K (85.0/39.0)
        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: \leq 50K (20.0/7.0)
        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)
     hours-per-week > 41: >50K (159.0/51.0)
  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 (62.0/20.0)
workclass = Self-emp-inc
  fnlwgt <= 74389: <=50K (6.0/1.0)
fnlwgt > 74389: >50K (76.0/24.0)
workclass = Federal-gov: >50K (22.0/5.0)
workclass = Local-gov
  capital-loss <= 653
     education = Bachelors: <=50K (0.0)
     education = Some-college: <=50K (23.0/10.0)
     education = 11th: <=50K (0.0)
     education = HS-grad: \leq 50K (0.0)
     education = Prof-school: <=50K (0.0)
     education = Assoc-acdm: <=50K (0.0)
     education = Assoc-voc: >50K (8.0/3.0)
     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)
     capital-loss > 653: >50K (2.0)
  workclass = State-gov
     relationship = Wife: >50K (3.0/1.0)
     relationship = Own-child: <=50K (0.0)
     relationship = Husband
     | age <= 45: <=50K (8.0/1.0)
       age > 45: >50K (11.0/3.0)
     relationship = Not-in-family: <=50K (0.0)
     relationship = Other-relative: <=50K (0.0)
     relationship = Unmarried: <=50K (0.0)
  workclass = Without-pay: >50K (0.0)
  workclass = Never-worked: >50K (0.0)
occupation = Prof-specialty
  workclass = Private
     education = Bachelors: <=50K (0.0)
     education = Some-college: <=50K (176.0/83.0)
     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 (60.0/25.0)
     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
     relationship = Wife: >50K (4.0)
     relationship = Own-child: <=50K (0.0)
     relationship = Husband: <=50K (17.0/6.0)
     relationship = Not-in-family: <=50K (0.0)
     relationship = Other-relative: <=50K (0.0)
     relationship = Unmarried: <=50K (0.0)
  workclass = Self-emp-inc
```

```
age <= 53: >50K (4.0)
     age > 53: <=50K (2.0)
  workclass = Federal-gov: >50K (9.0/2.0)
  workclass = Local-gov
    hours-per-week <= 42
     | age <= 50: >50K (4.0/1.0)
     | age > 50: <=50K (3.0/1.0)
     hours-per-week > 42: <=50K (2.0)
  workclass = State-gov
     age <= 46: >50K (9.0/2.0)
     age > 46
     | age <= 51: <=50K (4.0/1.0)
     | age > 51: >50K (2.0)
  workclass = Without-pay: >50K (0.0)
  workclass = Never-worked: >50K (0.0)
occupation = Handlers-cleaners: <=50K (79.0/15.0)
occupation = Machine-op-inspct: <=50K (169.0/60.0)
occupation = Adm-clerical
  workclass = Private
     race = White: <=50K (146.0/51.0)
     race = Asian-Pac-Islander: >50K (4.0/1.0)
     race = Amer-Indian-Eskimo: <=50K (0.0)
     race = Other: <=50K (0.0)
     race = Black
        education = Bachelors: >50K (0.0)
        education = Some-college: >50K (16.0/5.0)
        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 (4.0)
        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 (3.0)
```

```
workclass = Self-emp-inc: >50K (1.0)
  workclass = Federal-gov
     race = White
     fnlwgt <= 166606: <=50K (18.0/7.0)
     | fnlwgt > 166606: >50K (25.0/5.0)
     race = Asian-Pac-Islander: >50K (3.0/1.0)
     race = Amer-Indian-Eskimo: >50K (0.0)
     race = Other: >50K(0.0)
     race = Black: >50K (5.0/1.0)
  workclass = Local-gov
     relationship = Wife: <=50K (14.0/5.0)
     relationship = Own-child: <=50K (0.0)
     relationship = Husband: >50K (7.0/2.0)
     relationship = Not-in-family: <=50K (0.0)
     relationship = Other-relative: <=50K (0.0)
     relationship = Unmarried: <=50K (0.0)
  workclass = State-gov
     relationship = Wife: >50K (4.0/1.0)
     relationship = Own-child: <=50K (0.0)
     relationship = Husband: <=50K (8.0/2.0)
     relationship = Not-in-family: <=50K (0.0)
     relationship = Other-relative: <=50K (0.0)
     relationship = Unmarried: <=50K (0.0)
  workclass = Without-pay: <=50K (0.0)
  workclass = Never-worked: <=50K (0.0)
occupation = Farming-fishing: <=50K (129.0/27.0)
occupation = Transport-moving
  workclass = Private
     capital-gain <= 1173: <=50K (138.0/50.0)
     capital-gain > 1173
     | capital-gain <= 2964: <=50K (2.0)
        capital-gain > 2964: >50K (5.0/1.0)
  workclass = Self-emp-not-inc
     education = Bachelors: <=50K (0.0)
     education = Some-college: <=50K (11.0/2.0)
     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 (2.0)
     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 (7.0/1.0)
  workclass = Federal-gov
     race = White: >50K (2.0)
     race = Asian-Pac-Islander: <=50K (0.0)
     race = Amer-Indian-Eskimo: <=50K (0.0)
     race = Other: <=50K (0.0)
    race = Black: <=50K (2.0)
  workclass = Local-gov: <=50K (13.0/3.0)
  workclass = State-gov: <=50K (4.0)
  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 (15.0/1.0)
     workclass = Self-emp-not-inc: >50K (1.0)
     workclass = Self-emp-inc: >50K (1.0)
     workclass = Federal-gov
     | age <= 49: <=50K (2.0)
       age > 49: >50K (2.0)
     workclass = Local-gov
        education = Bachelors: >50K (0.0)
        education = Some-college
          age <= 54
             age <= 40: <=50K (32.0/13.0)
             age > 40
              | fnlwgt <= 138403
                  fnlwgt <= 42617: >50K (3.0/1.0)
               | fnlwgt > 42617: <=50K (5.0)
               fnlwgt > 138403: >50K (20.0/1.0)
        | age > 54: <=50K (4.0)
        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
                           | fnlwgt <= 118025: <=50K (2.0)
                           | fnlwgt > 118025: >50K (12.0/2.0)
                           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 (7.0)
                        | age > 41: >50K (14.0/5.0)
                        workclass = Without-pay: <=50K (0.0)
                        workclass = Never-worked: <=50K (0.0)
                     race = Asian-Pac-Islander: <=50K (1.0)
                     race = Amer-Indian-Eskimo: <=50K (1.0)
                     race = Other: >50K (2.0)
                     race = Black
                      | age <= 45: <=50K (8.0/3.0)
                     | age > 45: >50K (4.0)
                  occupation = Armed-Forces: >50K (1.0)
            | capital-loss > 1510: <=50K (39.0/1.0)
          | capital-gain > 4416: <=50K (27.0)
  capital-gain > 5060
     capital-gain <= 6612: >50K (101.0)
  | capital-gain > 6612: <=50K (5.0)
education-num > 11
  hours-per-week <= 30
     sex = Female: >50K (134.0/53.0)
  sex = Male: <=50K (351.0/102.0)
  hours-per-week > 30
    age <= 33
     | age <= 25: <=50K (112.0/27.0)
       age > 25
        | native-country = United-States
       | | education-num <= 12: <=50K (107.0/33.0)
```

```
| relationship = Wife: >50K (100.0/27.0)
           relationship = Own-child: <=50K (2.0)
           relationship = Husband
             occupation = Tech-support
                education-num <= 13
                 | fnlwgt <= 142049: >50K (4.0/1.0)
                 fnlwgt > 142049: <=50K (11.0/3.0)
                education-num > 13: >50K (2.0)
              occupation = Craft-repair
                age <= 32
                 | hours-per-week <= 42
                   | age <= 30: <=50K (9.0/2.0)
                     age > 30
                     | fnlwgt <= 170877: <=50K (4.0/1.0)
                     | fnlwgt > 170877: >50K (2.0)
                 hours-per-week > 42: >50K (13.0/4.0)
                age > 32: <=50K (6.0)
              occupation = Other-service
                fnlwgt <= 120676: >50K (2.0)
                fnlwgt > 120676: <=50K (8.0/1.0)
              occupation = Sales
                workclass = Private
                 fnlwgt <= 261437: >50K (68.0/21.0)
                 fnlwgt > 261437: <=50K (11.0/3.0)
                workclass = Self-emp-not-inc: <=50K (9.0/2.0)
                workclass = Self-emp-inc
                 | age <= 30: >50K (4.0)
                 | age > 30: <=50K (6.0/2.0)
                workclass = Federal-gov: >50K (0.0)
                workclass = Local-gov: <=50K (1.0)
                workclass = State-gov: >50K (0.0)
                workclass = Without-pay: >50K (0.0)
                workclass = Never-worked: >50K (0.0)
              occupation = Exec-managerial
                workclass = Private
                   age <= 28
                   | capital-gain <= 1506
                     | capital-loss <= 653: <=50K (26.0/10.0)
                      | capital-loss > 653: >50K (2.0)
                   | capital-gain > 1506: >50K (2.0)
```

```
| age > 28: >50K (80.0/17.0)
  workclass = Self-emp-not-inc
     fnlwgt <= 67671: <=50K (3.0)
     fnlwgt > 67671: >50K (4.0)
  workclass = Self-emp-inc
     age <= 29: <=50K (5.0/1.0)
     age > 29: >50K (4.0/1.0)
  workclass = Federal-gov: <=50K (1.0)
  workclass = Local-gov: <=50K (3.0/1.0)
  workclass = State-gov: >50K (5.0/1.0)
  workclass = Without-pay: >50K (0.0)
  workclass = Never-worked: >50K (0.0)
occupation = Prof-specialty
  race = White
     workclass = Private
       age <= 28: <=50K (30.0/12.0)
     | age > 28: >50K (104.0/35.0)
     workclass = Self-emp-not-inc
       hours-per-week <= 52: >50K (7.0)
        hours-per-week > 52
        | fnlwgt <= 151693: >50K (2.0)
          fnlwgt > 151693: <=50K (2.0)
     workclass = Self-emp-inc: >50K (4.0/1.0)
     workclass = Federal-gov
       age <= 32: >50K (9.0/2.0)
        age > 32: <=50K (2.0)
     workclass = Local-gov
        fnlwgt <= 191277: <=50K (17.0/3.0)
       fnlwgt > 191277: >50K (12.0/5.0)
     workclass = State-gov
       fnlwgt <= 192740: <=50K (10.0/3.0)
       fnlwgt > 192740: >50K (6.0/2.0)
     workclass = Without-pay: >50K (0.0)
     workclass = Never-worked: >50K (0.0)
  race = Asian-Pac-Islander: >50K (6.0)
  race = Amer-Indian-Eskimo: >50K (0.0)
  race = Other: <=50K (1.0)
  race = Black: <=50K (3.0/1.0)
occupation = Handlers-cleaners
  hours-per-week <= 45: <=50K (2.0)
  hours-per-week > 45: >50K (3.0)
```

```
occupation = Machine-op-inspct: <=50K (4.0/1.0)
          occupation = Adm-clerical
             age <= 31
                age <= 29
                | fnlwgt <= 202752: <=50K (5.0)
                | fnlwgt > 202752: >50K (3.0/1.0)
             | age > 29: >50K (8.0/1.0)
             age > 31: <=50K (8.0/3.0)
          occupation = Farming-fishing: <=50K (8.0/1.0)
          occupation = Transport-moving: <=50K (2.0/1.0)
          occupation = Priv-house-serv: >50K (0.0)
          occupation = Protective-serv
             workclass = Private: <=50K (1.0)
             workclass = Self-emp-not-inc: <=50K (0.0)
             workclass = Self-emp-inc: <=50K (0.0)
             workclass = Federal-gov: <=50K (0.0)
             workclass = Local-gov
                race = White
                | age <= 28: <=50K (4.0/1.0)
                  age > 28
                   hours-per-week <= 42: <=50K (3.0/1.0)
                    hours-per-week > 42: >50K (2.0)
                race = Asian-Pac-Islander: >50K (0.0)
                race = Amer-Indian-Eskimo: >50K (0.0)
                race = Other: >50K(0.0)
                race = Black: >50K (3.0)
             workclass = State-gov
                race = White
                | fnlwgt <= 238574: <=50K (2.0)
                  fnlwgt > 238574: >50K (2.0)
                race = Asian-Pac-Islander: <=50K (0.0)
                race = Amer-Indian-Eskimo: <=50K (0.0)
                race = Other: <=50K (0.0)
                race = Black: <=50K (2.0)
             workclass = Without-pay: <=50K (0.0)
             workclass = Never-worked: <=50K (0.0)
        occupation = Armed-Forces: >50K (0.0)
        relationship = Not-in-family: >50K (0.0)
        relationship = Other-relative: <=50K (5.0/1.0)
        relationship = Unmarried: >50K (0.0)
| native-country = Cambodia: <=50K (1.0)
```

```
| native-country = England
     age <= 32: >50K (3.0)
  | age > 32: <=50K (3.0)
  native-country = Puerto-Rico: <=50K (1.0)
  native-country = Canada: <=50K (4.0/1.0)
  native-country = Germany: <=50K (4.0/1.0)
  native-country = Outlying-US(Guam-USVI-etc): >50K (0.0)
  native-country = India
     workclass = Private: <=50K (10.0/2.0)
     workclass = Self-emp-not-inc: <=50K (0.0)
     workclass = Self-emp-inc: <=50K (0.0)
     workclass = Federal-gov: <=50K (0.0)
     workclass = Local-gov: <=50K (0.0)
     workclass = State-gov: >50K (2.0)
     workclass = Without-pay: <=50K (0.0)
     workclass = Never-worked: <=50K (0.0)
  native-country = Japan
     fnlwgt <= 138790: >50K (2.0)
   | fnlwgt > 138790: <=50K (2.0)
  native-country = Greece: >50K (1.0)
  native-country = South: <=50K (1.0)
  native-country = China: <=50K (3.0/1.0)
  native-country = Cuba: >50K (3.0/1.0)
  native-country = Iran: >50K(3.0/1.0)
  native-country = Honduras: >50K (1.0)
  native-country = Philippines: <=50K (6.0/3.0)
  native-country = Italy: <=50K (2.0)
  native-country = Poland: >50K (0.0)
  native-country = Jamaica: <=50K (1.0)
  native-country = Vietnam: >50K (1.0)
  native-country = Mexico: <=50K (8.0/1.0)
  native-country = Portugal: <=50K (1.0)
  native-country = Ireland: >50K (0.0)
  native-country = France: >50K (3.0)
  native-country = Dominican-Republic: <=50K (2.0/1.0)
  native-country = Laos: >50K (0.0)
  native-country = Ecuador: <=50K (2.0)
  native-country = Taiwan: <=50K (3.0/1.0)
  native-country = Haiti: <=50K (2.0)
  native-country = Columbia: <=50K (3.0/1.0)
  native-country = Hungary: >50K (0.0)
```

```
| | native-country = Guatemala: >50K (0.0)
        native-country = Nicaragua: <=50K (2.0)
        native-country = Scotland: >50K (0.0)
        native-country = Thailand: >50K (2.0)
        native-country = Yugoslavia: <=50K (2.0/1.0)
        native-country = El-Salvador: <=50K (3.0)
        native-country = Trinadad&Tobago: >50K (0.0)
        native-country = Peru: >50K (0.0)
        native-country = Hong
        | fnlwgt <= 141957: <=50K (2.0)
      | | fnlwgt > 141957: >50K (4.0)
        native-country = Holand-Netherlands: >50K (0.0)
   age > 33
     occupation = Tech-support: >50K (116.0/35.0)
     occupation = Craft-repair
      | education = Bachelors
        | workclass = Private: >50K (78.0/31.0)
           workclass = Self-emp-not-inc: <=50K (24.0/12.0)
           workclass = Self-emp-inc: <=50K (10.0/3.0)
           workclass = Federal-gov: >50K (3.0/1.0)
           workclass = Local-gov
           | hours-per-week <= 43: <=50K (8.0/1.0)
           | hours-per-week > 43: >50K (3.0)
           workclass = State-gov: >50K(3.0/1.0)
          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 (5.0/2.0)
        education = Assoc-acdm
           capital-loss <= 653
            age <= 37: <=50K (15.0/3.0)
             age > 37
              | fnlwgt <= 153052
                | hours-per-week <= 47: <=50K (11.0/2.0)
              | hours-per-week > 47: >50K (4.0/1.0)
              | fnlwgt > 153052: >50K (25.0/5.0)
     | | capital-loss > 653: <=50K (3.0)
      | education = Assoc-voc: >50K (0.0)
```

```
| | education = 7th-8th: >50K (0.0)
        education = 12th: >50K (0.0)
        education = Masters: >50K (18.0/8.0)
        education = 1st-4th: >50K(0.0)
        education = 10th: >50K (0.0)
        education = Doctorate: <=50K (1.0)
        education = 5th-6th: >50K (0.0)
        education = Preschool: >50K (0.0)
     occupation = Other-service: <=50K (48.0/8.0)
     occupation = Sales: >50K (483.0/162.0)
     occupation = Exec-managerial
      workclass = Private: >50K (645.0/125.0)
      | workclass = Self-emp-not-inc
          education = Bachelors: <=50K (62.0/27.0)
           education = Some-college: <=50K (0.0)
           education = 11th: <=50K (0.0)
           education = HS-grad: <=50K (0.0)
           education = Prof-school: <=50K (1.0)
           education = Assoc-acdm: <=50K (9.0/2.0)
           education = Assoc-voc: <=50K (0.0)
           education = 9th: <=50K (0.0)
           education = 7th-8th: <=50K (0.0)
           education = 12th: <=50K (0.0)
           education = Masters
           | age <= 54: >50K (13.0/2.0)
           | age > 54: <=50K (2.0)
           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 <= 67
             age <= 49: >50K (65.0/10.0)
             age > 49
               fnlwgt <= 135315: >50K (12.0)
                fnlwgt > 135315
                  fnlwgt <= 282643: <=50K (15.0/5.0)
                   fnlwgt > 282643
                     fnlwgt <= 373263: >50K (5.0)
                  | fnlwgt > 373263: <=50K (3.0/1.0)
```

```
| education-num <= 13: <=50K (7.0/1.0)
      | | education-num > 13: >50K (2.0)
      workclass = Federal-gov: >50K (54.0/15.0)
      workclass = Local-gov
      | capital-loss <= 625: >50K (78.0/14.0)
      | capital-loss > 625: <=50K (4.0)
      workclass = State-gov
        hours-per-week <= 46
          age <= 41: <=50K (12.0/3.0)
           age > 41
             age <= 53: >50K (22.0/2.0)
              age > 53
                 education = Bachelors: <=50K (9.0/3.0)
                 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 (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 (4.0)
                 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)
        hours-per-week > 46: >50K (16.0)
     workclass = Without-pay: >50K (0.0)
     workclass = Never-worked: >50K (0.0)
   occupation = Prof-specialty: >50K (1339.0/357.0)
   occupation = Handlers-cleaners: <=50K (28.0/9.0)
   occupation = Machine-op-inspct
     workclass = Private: <=50K (40.0/15.0)
      workclass = Self-emp-not-inc: >50K (3.0/1.0)
      workclass = Self-emp-inc: >50K (1.0)
      workclass = Federal-gov: <=50K (0.0)
      workclass = Local-gov: <=50K (0.0)
   | workclass = State-gov: <=50K (0.0)
```

```
workclass = Without-pay: <=50K (0.0)
     workclass = Never-worked: <=50K (0.0)
   occupation = Adm-clerical
   | relationship = Wife
   | age <= 57: >50K (39.0/12.0)
      | age > 57: <=50K (3.0)
     relationship = Own-child: <=50K (1.0)
     relationship = Husband
      | education = Bachelors
        | age <= 58
           | age <= 41
                workclass = Private
                | fnlwgt <= 286487: >50K (22.0/7.0)
                 fnlwgt > 286487: <=50K (6.0/1.0)
                workclass = Self-emp-not-inc: <=50K (1.0)
                workclass = Self-emp-inc: <=50K (0.0)
                workclass = Federal-gov: <=50K (5.0/1.0)
                workclass = Local-gov: >50K(3.0/1.0)
                workclass = State-gov: <=50K (3.0/1.0)
                workclass = Without-pay: <=50K (0.0)
                workclass = Never-worked: <=50K (0.0)
              age > 41
                capital-gain <= 3942: >50K (58.0/11.0)
                capital-gain > 3942
                | capital-gain <= 5060: <=50K (3.0)
                | capital-gain > 5060: >50K (2.0)
        | age > 58: <=50K (8.0/1.0)
        education = Some-college: >50K (0.0)
        education = 11th: >50K (0.0)
        education = HS-grad: >50K (0.0)
        education = Prof-school: >50K (6.0/1.0)
        education = Assoc-acdm
           workclass = Private
             age <= 36: >50K (2.0)
             age > 36
              | age <= 50: <=50K (6.0)
           | age > 50: >50K (2.0)
           workclass = Self-emp-not-inc: <=50K (1.0)
           workclass = Self-emp-inc: <=50K (0.0)
           workclass = Federal-gov
        | age <= 49: >50K (6.0/1.0)
```

```
| | | | age > 49: <=50K (2.0)
           workclass = Local-gov: <=50K (3.0)
           workclass = State-gov: >50K (1.0)
           workclass = Without-pay: <=50K (0.0)
         | workclass = Never-worked: <=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
         | workclass = Private: <=50K (16.0/6.0)
           workclass = Self-emp-not-inc: <=50K (0.0)
           workclass = Self-emp-inc: <=50K (1.0)
           workclass = Federal-gov: >50K (4.0)
           workclass = Local-gov: \leq 50K (6.0/2.0)
           workclass = State-gov: >50K(3.0/1.0)
           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 (3.0/1.0)
        education = 5th-6th: >50K (0.0)
      education = Preschool: >50K (0.0)
      relationship = Not-in-family: >50K (0.0)
     relationship = Other-relative: >50K (3.0/1.0)
     relationship = Unmarried: >50K (0.0)
   occupation = Farming-fishing
      workclass = Private: <=50K (18.0/5.0)
      workclass = Self-emp-not-inc: <=50K (37.0/9.0)
      workclass = Self-emp-inc: >50K (6.0)
      workclass = Federal-gov: <=50K (2.0/1.0)
      workclass = Local-gov: <=50K (0.0)
      workclass = State-gov: >50K (1.0)
     workclass = Without-pay: <=50K (1.0)
   | workclass = Never-worked: <=50K (0.0)
   occupation = Transport-moving: <=50K (51.0/19.0)
   occupation = Priv-house-serv: <=50K (2.0)
   occupation = Protective-serv
   | age <= 42
      | hours-per-week <= 41: >50K (13.0/4.0)
| | hours-per-week > 41: <=50K (17.0/4.0)
```

```
| | | | occupation = Armed-Forces: >50K (1.0)
| capital-loss > 1844
| | capital-loss <= 1980: >50K (766.0/17.0)
  | capital-loss > 1980
     | capital-loss <= 2163: <=50K (91.0)
     | capital-loss > 2163
       | capital-loss <= 2415
       | | capital-loss <= 2377
         | age <= 64: <=50K (33.0/1.0)
       | | age > 64: >50K (28.0/2.0)
       | | capital-loss > 2377: >50K (72.0)
marital-status = Divorced
| education-num <= 12: <=50K (4668.0/197.0)
| education-num > 12
  | hours-per-week <= 42
  | | capital-loss <= 1726: <=50K (647.0/74.0)
    | capital-loss > 1726
       | age <= 42: <=50K (7.0)
  | hours-per-week > 42
    occupation = Tech-support: <=50K (4.0/1.0)
     occupation = Craft-repair: <=50K (11.0/2.0)
     occupation = Other-service
       relationship = Wife: <=50K (0.0)
        relationship = Own-child: <=50K (0.0)
         relationship = Husband: <=50K (0.0)
        relationship = Not-in-family
          | hours-per-week <= 52: >50K (3.0/1.0)
         | hours-per-week > 52: <=50K (4.0)
          relationship = Other-relative: <=50K (0.0)
       relationship = Unmarried: >50K (6.0/2.0)
       occupation = Sales: <=50K (65.0/24.0)
       occupation = Exec-managerial
        | capital-loss <= 1485
         | race = White
           relationship = Wife: >50K (0.0)
         |  relationship = Own-child: <=50K (2.0/1.0)
            relationship = Husband: >50K (0.0)
  | | | | relationship = Not-in-family
```

```
education = Some-college: >50K (0.0)
          education = 11th: >50K (0.0)
          education = HS-grad: >50K (0.0)
          education = Prof-school: <=50K (1.0)
          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
           hours-per-week <= 63: >50K (22.0/7.0)
           hours-per-week > 63: <=50K (6.0/1.0)
          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)
        relationship = Other-relative: <=50K (2.0)
        relationship = Unmarried
        | hours-per-week <= 56
          | sex = Female: <=50K (16.0/4.0)
          sex = Male
            | hours-per-week <= 47: >50K (4.0)
          | hours-per-week > 47: <=50K (4.0/1.0)
          hours-per-week > 56: >50K (4.0)
     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 (3.0)
     | fnlwgt > 171306: <=50K (4.0/1.0)
  capital-loss > 1485: >50K (10.0)
occupation = Prof-specialty
   education-num <= 14: <=50K (146.0/27.0)
   education-num > 14
  | capital-gain <= 1086
     sex = Female
     relationship = Wife: <=50K (0.0)
          relationship = Own-child: <=50K (0.0)
| | | relationship = Husband: <=50K (0.0)
```

```
| | | | relationship = Not-in-family
                | | fnlwgt <= 256062: >50K (4.0)
                | | fnlwgt > 256062: <=50K (2.0)
                  relationship = Other-relative: <=50K (0.0)
                  relationship = Unmarried: <=50K (4.0)
                sex = Male
                  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 (4.0)
                   | hours-per-week > 47: <=50K (4.0)
                   | hours-per-week > 52: >50K (7.0)
             | | relationship = Other-relative: >50K (0.0)
             | | relationship = Unmarried: >50K (4.0)
        | | capital-gain > 1086: <=50K (3.0)
        occupation = Handlers-cleaners: <=50K (2.0/1.0)
        occupation = Machine-op-inspct: <=50K (2.0)
       occupation = Adm-clerical: <=50K (11.0)
       occupation = Farming-fishing: <=50K (4.0)
        occupation = Transport-moving: <=50K (7.0/1.0)
        occupation = Priv-house-serv: <=50K (1.0)
        occupation = Protective-serv
        | race = White: <=50K (6.0/2.0)
          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: <=50K (11241.0/141.0)
 education-num > 12
  | education-num <= 14
     hours-per-week <= 42: <=50K (2021.0/106.0)
       hours-per-week > 42
        | capital-loss <= 1485: <=50K (809.0/132.0)
          capital-loss > 1485
          | capital-loss <= 1573: >50K (11.0/1.0)
             capital-loss > 1573
          | | capital-loss <= 2377: <=50K (33.0/2.0)
```

```
| | | | | capital-loss > 2377: >50K (6.0)
      education-num > 14
      | age <= 32: <=50K (77.0/9.0)
        age > 32
         | capital-loss <= 1485
           relationship = Wife: <=50K (0.0)
           relationship = Own-child: <=50K (3.0)
              relationship = Husband: <=50K (0.0)
              relationship = Not-in-family
              | age <= 54
                | education = Bachelors: >50K (0.0)
                   education = Some-college: >50K (0.0)
                   education = 11th: >50K (0.0)
                   education = HS-grad: >50K (0.0)
                   education = Prof-school: >50K (41.0/14.0)
                   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 (0.0)
                   education = 1st-4th: >50K (0.0)
                   education = 10th: >50K (0.0)
                   education = Doctorate
                   | fnlwgt <= 144071: <=50K (16.0/4.0)
                   | fnlwgt > 144071: >50K (17.0/4.0)
                 | education = 5th-6th: >50K (0.0)
                education = Preschool: >50K (0.0)
             | age > 54: <=50K (18.0/4.0)
           relationship = Other-relative: <=50K (2.0)
        relationship = Unmarried: <=50K (2.0)
     | | capital-loss > 1485: >50K (16.0/2.0)
marital-status = Separated: <=50K (1345.0/64.0)
marital-status = Widowed
| education-num <= 11: <=50K (1144.0/45.0)
 | education-num > 11
  | sex = Female: <=50K (168.0/23.0)
   sex = Male
      | hours-per-week <= 32: <=50K (11.0)
        hours-per-week > 32
  | | age <= 33: <=50K (5.0)
```

```
| fnlwgt <= 219446
               | workclass = Private
                 | fnlwgt <= 129786: >50K (6.0/1.0)
                 | fnlwgt > 129786: <=50K (5.0/1.0)
                 workclass = Self-emp-not-inc: <=50K (5.0/1.0)
                 workclass = Self-emp-inc: >50K (2.0)
                 workclass = Federal-gov: <=50K (0.0)
                 workclass = Local-gov: <=50K (0.0)
                 workclass = State-gov: <=50K (1.0)
               workclass = Without-pay: <=50K (0.0)
               | workclass = Never-worked: <=50K (0.0)
  marital-status = Married-spouse-absent: <=50K (544.0/39.0)
  marital-status = Married-AF-spouse
| | education-num <= 12: <=50K (26.0/6.0)
| | education-num > 12: >50K (6.0/1.0)
capital-gain > 6849: >50K (1849.0/28.0)
Number of Leaves: 670
Size of the tree:
                    873
=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances
                              42006
                                           86.0038 %
Incorrectly Classified Instances
                               6836
                                           13.9962 %
Kappa statistic
                          0.5858
K&B Relative Info Score
                              44.8975 %
K&B Information Score
                             17408.0821 bits
                                               0.3564 bits/instance
Class complexity | order 0
                             38772.9212 bits
                                               0.7938 bits/instance
Class complexity | scheme
                             329255.0562 bits
                                                 6.7412 bits/instance
                         (Sf) -290482.135 bits -5.9474 bits/instance
Complexity improvement
Mean absolute error
                              0.1936
Root mean squared error
                                0.3213
Relative absolute error
                             53.1688 %
Root relative squared error
                               75.3068 %
Total Number of Instances
                              48842
```

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.941 0.397 0.883 0.941 0.911 0.592 0.887 0.945 <=50K 0.603 0.059 0.763 0.603 0.673 0.592 0.887 0.742 >50K Weighted Avg. 0.860 0.316 0.854 0.860 0.854 0.592 0.887 0.896

=== Confusion Matrix ===

a b <-- classified as 34962 2193 | a = <=50K 4643 7044 | b = >50K